

**T.C.  
SAKARYA ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ**

**İLK YIL ÖĞRENCİLERİNİN AKADEMİK PERFORMANSINA ETKİ  
EDEN FAKTÖRLERİN ARAŞTIRILMASI VE BU FAKTÖRLERE  
BAĞLI OLARAK BAŞARILARININ TAHMİNİNE YÖNELİK BİR  
KARAR DESTEK SİSTEMİ TASARIMI**

**YÜKSEK LİSANS TEZİ**

**Tuğrul Cabir HAKYEMEZ**

**Enstitü Anabilim Dalı : Yönetim Bilişim Sistemleri**

**Tez Danışmanı: Prof. Dr. Erman COŞKUN**

**TEMMUZ – 2015**

T.C.  
SAKARYA ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ




İLK YIL ÖĞRENCİLERİNİN AKADEMİK PERFORMANSINA ETKİ  
EDEN FAKTÖRLERİN ARAŞTIRILMASI VE BU FAKTÖRLERE  
BAĞLI OLARAK BAŞARILARININ TAHMİNİNE YÖNELİK BİR  
KARAR DESTEK SİSTEMİ TASARIMI

YÜKSEK LİSANS TEZİ

Tuğrul Cabir HAKYEMEZ

Enstitü Anabilim Dalı : Yönetim Bilişim Sistemleri

Bu tez 31.09.2015 tarihinde aşağıdaki jüri tarafından Oybirliği / Oyçokluğu ile kabul edilmiştir.

JÜRİ ÜYESİ	KANAATİ	İMZA
Prof. Dr. ERMAN COŞKUN	BAŞARILI	
Prof. Dr. ALPTEKİN ERKOLLAR	BAŞARILI	
Ord. Doç. Dr. ÇAĞLA EDİZ	BAŞARILI	

## **BEYAN**

Bu tezin yazılmasında bilimsel ahlak kurallarına uyulduđunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduđunu, kullanılan verilerde herhangi bir tahrifat yapılmadıđını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadıđını beyan ederim.

**Tuđrul Cabir HAKYEMEZ**

**31.07.2015**

## ÖNSÖZ

Tez çalışmamın tüm kritik aşamalarında yol gösteren, her daim desteğini hissettiren ve umudunu kesmeyen değerli danışmanım Prof. Dr. Erman COŞKUN hocama, hayatımın her döneminde daima yanımda olan herşeyimi borçlu olduğum babam ve annem ALİ-FATMA HAKYEMEZ'e, bu zorlu dönemlerimde sürekli yanımda olan kıymetli arkadaşlarım Fatih FAYDALI'ya, Faruk ERGİN' e, Ömer Sezai AYKAÇ'a, Kadir KURT'a, Akın ÖZDEMİR'e, Dorukcan PEHLİVAN'a, çok kıymetli oda arkadaşlarım, Mustafa-Tuğba KOÇ çiftine, değerli öğrenciler Samet ARSLAN ve Aykut ÖZDEMİR'e, çalışmamda kritik noktalarda önemli katkılarda bulunan Ayşegül KARATAŞ'a, Yusuf ARSLAN'a ve Güliz UĞUR'a ve Volkan GÖKTAŞ'a teşekkürleri bir borç bilirim.

**Tuğrul Cabir HAKYEMEZ**

**31.07.2015**



# İÇİNDEKİLER

<b>KISALTMALAR</b> .....	vii
<b>TABLO LİSTESİ</b> .....	viii
<b>ŞEKİL LİSTESİ</b> .....	ix
<b>ÖZET</b> .....	xii
<b>SUMMARY</b> .....	xiii
<b>GİRİŞ</b> .....	1
<b>BÖLÜM 1: AKADEMİK PERFORMANSA İLİŞKİN KAVRAMLAR VE LİTERATÜR TARAMASI</b> .....	8
1.1. Performansın tanımı .....	8
1.2 Performans Yönetim Sistemleri .....	10
1.3 Öğretimde performans ölçümü ve yönetimi .....	12
1.4 Türkiye’de Ölçme Ve Değerlendirme Faaliyetleri .....	14
1.4.1 Türkiye’de Üniversiteye Giriş Sınavları .....	14
1.4.2 Yükseköğretime Geçiş Sınavı (YGS) ve Lisans Yerleştirme Sınavı (LYS)..	16
1.5 Öğrencilerin Akademik Başarılarının Ölçme ve Değerlendirmesi .....	18
1. 6 Literatür .....	19
1.6.1 Akademik Başarı Üzerine Teorik Modeller .....	19
1.6.2 Akademik Başarıyı Etkileyen Faktörler Üzerine Uygulamalar .....	33
<b>BÖLÜM 2: AKADEMİK PERFORMANSIN TAHMİNİNE YÖNELİK BİR MODEL TASARIMI VE UYGULAMASI</b> .....	54
2.1 Türkiye’de Mevcut Durum .....	54
2.2 Metodoloji .....	59
2.2.1 Problem Tanımlanması ve Anlaşılması .....	62
2.2.2 Veriyi anlama .....	63
2.2.3. Veriyi Hazırlama .....	89
2.2.4. Modelleme .....	106
2.2.5 Değerlendirme .....	169

2.2.6 Yayılım.....	173
<b>SONUÇ VE ÖNERİLER.....</b>	<b>177</b>
<b>KAYNAKÇA .....</b>	<b>184</b>
<b>ÖZGEÇMİŞ.....</b>	<b>205</b>

## KISALTMALAR

- ÇKA** : Çok Katmanlı Algılayıcılar
- RF** : Random Forest (Rassal Ormanlar)
- NCES** : National Center For Education Statistics
- CIRP** : Cooperative Institutional Research Program
- ÖSYM** : Ölçme, Seçme ve Yerleştirme Merkezi
- CRISPDM**: Cross Industry Standart Process for Data Mining
- EVM** : Eğitimde Veri Madenciliği
- TP** : True Positive
- YÖK** : Yükseköğretim Kurulu
- DSO** : Doğru Sınıflandırma Oranı
- RGB** : Risk Grubunu Belirleyebilme



## TABLO LİSTESİ

<b>Tablo 1:</b> LYS ve YGS puan türlerinin ham puan oluşumuna katkı yüzdeleri (ÖSYM, 2015) .....	18
<b>Tablo 2 :</b> 6 yılda mezun olma oranına etki eden faktörler (Hosch, 2008).....	37
<b>Tablo 3:</b> Veri madenciliği Algoritmaları kategorizasyonu (Jing, 2002: s.13) .....	45
<b>Tablo 4:</b> Farklı yapıda Yapay Sinir Ağlarının tahminleme Doğrulukları (Fausett&Elwasif, 1994) .....	48
<b>Tablo 5:</b> Çeşitli veri madenciliği tekniklerinin sınıflandırma doğruluk oranları (Superby vd., 2006) .....	50
<b>Tablo 6:</b> Sakarya Üniversitesi Not Değerlendirme Sistemi (Sakarya Üniversitesi, 2011: s. 95).....	56
<b>Tablo 7:</b> CRISP-DM görevleri ve girdilerine genel bir bakış (Wirth&Hipp, 2000: s. 34).....	61
<b>Tablo 8:</b> Öğrencilerin üniversiteye gelme nedenleri (Higher Education Research Institute, 2015) .....	69
<b>Tablo 9:</b> Öğrencilerin üniversiteye gelme nedenlerine ait frekanslar .....	70
<b>Tablo 10:</b> Çevrilmiş ve revize edilmiş Genel Özyeterlik Ölçeği (Chen, 2001) .....	72
<b>Tablo 11:</b> Çalışmada kullanılan Akademik Özyeterlik Ölçeği (Yılmaz vd. , 2007).....	74
<b>Tablo 12:</b> Türkçe'ye çevrilmiş Kurumsal Entegrasyon Ölçeği (French&Oakes, 2004) 76	
<b>Tablo 13:</b> Tahminleme modelinde kullanılan değişkenler ve bağımlı değişken ile aralarındaki korelasyon katsayıları .....	102
<b>Tablo 14:</b> Oluşturulan modeller için doğru sınıflandırma oranları .....	134
<b>Tablo 15:</b> Mevcut durum ve Senaryo1 için modellerin doğru sınıflandırma performansları .....	147
<b>Tablo 16:</b> Modellerin performans puanları .....	167
<b>Tablo 17:</b> Modellere ait bilgisayar zamanları .....	169
<b>Tablo 18:</b> Operasyonel ve Bilgilendirici Sistemlerin karşılaştırılması (Hoffer vd.; 2002: s. 398).....	176
<b>Tablo 19:</b> Yakıştırma yöntemleri (Engels&Diehr, 2003: s. 970).....	179

## ŞEKİL LİSTESİ

Şekil 1: Performans ölçümü ve yönetimi arasındaki ilişki .....	11
Şekil 2 : Öğretim sürecinde Ölçme ve Değerlendirmenin Yeri .....	13
Şekil3 : Yükseköğretime geçiş sınavı (YGS) puanlarına göre tercih edilebilecek yükseköğretim programları .....	17
Şekil 4 : Okulu bırakma sürecinin açıklayıcı bir sosyolojik modeli .....	21
Şekil 5 : Üniversiteye devam karar modeli .....	22
Şekil 6 : Öğrenci ayrılık kararının revize edilmiş bir modeli .....	24
Şekil 7 : Örnek bir Öğrenen Topluluk Modeli .....	26
Şekil 8 : Kurumsal eylemin model yapısı ve bileşenleri .....	28
Şekil 9 : Türkiye’de yükseköğretim öğrenci sayısındaki değişim (1974-2013) .....	30
Şekil 10: Bir yükseköğretim programına yerleşen ve bir yükseköğretim programından mezun olan sayıları (1982-2011) .....	31
Şekil 11: Üniversite mezuniyetine etki eden nedensel değişkenler arasındaki ilişki ....	34
Şekil 12: Transkript grafiğinin BBN modeli .....	47
Şekil 13: Genel Özyeterlik Ölçeği Cronbach $\alpha$ değeri .....	73
Şekil 14: Akademik Özyeterlik Ölçeği Cronbach $\alpha$ değeri .....	74
Şekil 15: Birincil faktör analizi KMO örneklem yeterlilik sonuçları .....	79
Şekil 16: Birincil faktör analizi sonucunda veride açıklanan toplam varyans.....	79
Şekil 17: Birincil faktör analizi Rotated Component matrisi .....	80
Şekil 18: İkincil faktör analizi KMO örneklem yeterlilik sonuçları.....	81
Şekil 19: İkincil faktör analizi sonucunda veride açıklanan toplam varyans .....	82
Şekil 20: İkincil faktör analizi Rotated Component matrisi .....	83
Şekil 21: Çalışmada elde edilen alt ölçeklere ait iç tutarlılık oranları ( $\alpha$ ) .....	87
Şekil 22: İkinci Dönem Not Ortalaması ve Başarı Durumu Üzerine Teorik bir Model.	88
Şekil 23: Kümeleme aşamaları .....	94
Şekil 24: Kümeleme yaklaşımları taksonomisi .....	95
Şekil 25: Lise Başarı Durumu için eşik değerleri ve kümelerin üye sayıları .....	96

<b>Şekil 26:</b> İlk Dönem Başarı Durumu için eşik değerleri ve kümelerin üye sayıları .....	97
<b>Şekil 27:</b> İkinci Dönem Başarı Durumu için eşik değerleri ve kümelerin üye sayıları ..	97
<b>Şekil 28:</b> Genel Özyeterlik Derecesi için eşik değerleri ve kümelerin üye sayıları .....	98
<b>Şekil 29:</b> Akademik Özyeterlik Derecesi için eşik değerleri ve kümelerin üye sayıları	98
<b>Şekil 30:</b> Kurumsal entegrasyon boyutlarına yönelik algı sınıfları eşik değerleri ve her bir sınıf için üye sayıları.....	100
<b>Şekil 31:</b> Kurumsal Entegrasyon Ölçeği boyutları arasındaki korelasyon katsayıları .	106
<b>Şekil 32:</b> Bir sinir hücresi yapısı .....	108
<b>Şekil 33:</b> Yapay Sinir Hücresi yapısı .....	108
<b>Şekil 34:</b> YSA'da kullanılan aktivasyon fonksiyonları .....	111
<b>Şekil 35:</b> Yapay Sinir Ağlarına ait bir taksonomi.....	113
<b>Şekil 36:</b> Çok Katmanlı Algılayıcı yapısı .....	114
<b>Şekil 37:</b> Hastalara verilecek ilaçlar ile ilgili örnek bir Karar Ağacı .....	116
<b>Şekil 38:</b> Örnek bir CART yapısı .....	119
<b>Şekil 39 :</b> Çok Katmanlı Algılayıcı ile yapılan test çalışması sonuçları.....	124
<b>Şekil 40:</b> Rassal Ormanlar ile yapılan test çalışması sonuçları .....	125
<b>Şekil 41:</b> Çokterimli Lojistik Regresyon test aşaması sonuçları .....	126
<b>Şekil 42:</b> Korelasyon temelli özellik seçiminden sonra oluşan Veriseti1'de kullanılan değişkenler .....	127
<b>Şekil 43:</b> Orijinal verisetinin korunduğu verisetinde kullanılan değişkenler .....	128
<b>Şekil 44:</b> Model1 için örnek çıktılar sonuçlar.....	130
<b>Şekil 45:</b> Model2 için örnek çıktılar ve sonuçlar.....	131
<b>Şekil 46:</b> Model3 için örnek çıktılar ve sonuçlar.....	133
<b>Şekil 47:</b> Dengelenmemiş veriseti .....	136
<b>Şekil 48:</b> Dengelenmiş veriseti .....	136
<b>Şekil 49:</b> Senaryo1 için Model3'e ait örnek çıktılar ve sonuçlar.....	138
<b>Şekil 50:</b> Senaryo1 için Model1'e ait örnek çıktılar ve sonuçlar.....	139
<b>Şekil 51:</b> Senaryo1 için Model2'ye ait örnek çıktılar ve sonuçlar.....	141

<b>Şekil 52:</b> Senaryo1 için Model4' e ait örnek çıktılar ve sonuçlar.....	143
<b>Şekil 53:</b> Senaryo1 için Model5' e ait örnek çıktılar ve sonuçlar.....	144
<b>Şekil 54:</b> Senaryo1 için Model6'ya ait örnek çıktılar ve sonuçlar.....	146
<b>Şekil 55:</b> Senaryo2 için öğrencilerin başarı durumları dağılımları.....	147
<b>Şekil 56 :</b> Senaryo2 için Model1'e ait örnek çıktılar ve sonuçlar.....	149
<b>Şekil 57:</b> Senaryo2 için Model2'ye örnek çıktılar ve sonuçlar.....	150
<b>Şekil 58:</b> Senaryo2 için Model3'e ait örnek çıktılar ve sonuçlar.....	151
<b>Şekil 59:</b> Senaryo2 için Model4'e ait örnek çıktılar ve sonuçlar.....	153
<b>Şekil 60:</b> Senaryo2 için Model5'e ait örnek çıktılar ve sonuçlar.....	154
<b>Şekil 61:</b> Senaryo2 için Model6'ya ait örnek çıktılar ve sonuçlar.....	155
<b>Şekil 62:</b> Senaryo3 için yeni veri dağılımları .....	156
<b>Şekil 63:</b> Senaryo3 için Model1'e ait örnek çıktılar ve sonuçlar.....	157
<b>Şekil 64:</b> Senaryo3 için Model2'ye ait örnek çıktılar ve sonuçlar.....	158
<b>Şekil 65:</b> Senaryo3 için Model3'e ait örnek çıktılar ve sonuçlar.....	160
<b>Şekil 66:</b> Senaryo3 için Model4 için örnek çıktılar ve sonuçlar.....	161
<b>Şekil 67:</b> Senaryo3 için Model5 için örnek çıktılar ve sonuçlar.....	163
<b>Şekil 68:</b> Senaryo3 için Model6 için örnek çıktılar ve sonuçlar.....	164
<b>Şekil 69:</b> Mevcut durumda Model1'e ait duyarlılık analizi sonuçları .....	170
<b>Şekil 70:</b> Öğrencilerin akademik başarı durumlarının tahminine ve bu başarıya etki eden faktörlerin belirlenmesine yönelik bir sistem tasarımı.....	174

**Tezin Başlığı:** İlk Yıl Öğrencilerinin Akademik Performansına Etki Eden Faktörlerin Araştırılması Ve Bu Faktörlere Bağlı Olarak Başarılarının Tahminine Yönelik Bir Karar Destek Sistemi Tasarımı

**Tezin Yazarı:** Tuğrul C. HAKYEMEZ **Danışman:** Prof. Dr. Erman COŞKUN

**Kabul Tarihi:** 31.07.2015 **Sayfa Sayısı:** xii (ön kısım) + 205 (tez)

**Anabilim dalı:** Yönetim Bilişim Sistemleri **Bilim dalı:**

“Öğretici merkezli” eğitim anlayışından “Öğrenci merkezli” eğitim anlayışına geçişin yaşandığı yükseköğretim çevrelerinde öğrencilerin kayıtlı oldukları lisans programlarında göstermiş oldukları akademik performans önem kazanmaktadır. Bu bağlamda, öğrencilerin yeni akademik ve sosyal ortama “hassas” bir geçiş süreci yaşadığı ilk yıl (Milem&Berger,1997), yükseköğretim sürecinde gösterilecek akademik performansa temel teşkil etmesi bakımından pek çok teorik (Tinto,1975; Astin, 1984) ve uygulamaya dönük çalışmaya (Delen,2010 ; Mishra vd., 2014) konu olmuştur. Konu ile ilgili yapılan literatür taramasında, öğrencilerin ilk yıl akademik performanslarına etki eden faktörleri belirleyen ve bu faktörlere bağlı olarak yılsonu akademik başarılarını tahminleyen herhangi bir yerli “bütüncül” çalışmaya rastlanmamıştır. Bu bağlamda; çalışmada karmaşık “yükseköğretimde ilk yıl” sürecine yönelik bir teorik model ortaya konması ve bu modele paralel olarak veri madenciliği süreci yaklaşımı kullanılarak akademik performans önceden tahmin etmek ve etki eden faktörleri belirlemek suretiyle yöneticilere karar desteği sağlayacak genel bir sistem tasarlanması amaçlanmaktadır.

Mevcut çalışmada; 2014-2015 akademik yılında Sakarya Üniversitesi-İşletme Fakültesi lisans programlarında ilk yıllarını geçiren öğrencilere “Yeni Gelen Öğrenci” anketi uygulanmıştır. Bu şekilde tahminleme çalışmasında kullanılmak üzere geleneksel (yaş, cinsiyet, lise not ortalaması vs.) ve geleneksel olmayan (genel özyeterlik, akademik özyeterlik, kurumsal entegrasyon) değişkenler elde edilmiştir. Yöntem olarak; pek çok farklı alanda veri madenciliği uygulama süreçlerine yönelik standart bir metodoloji sunan (Wirth&Hipp, 2000) CRISP-DM benimsenmiştir. Bu kapsamda, veri hazırlama aşamasından (seçme, temizleme, türetme vs.) sonra tahminleme için makine öğrenme teknikleri (Çok Katmanlı Algılayıcılar ve Rassal Orman yöntemi) ve geleneksel istatistiksel sınıflandırıcı teknikleri (Çokterimli Lojistik Regresyon) ile farklı verisetleri (korelasyon katsayılarına göre seçilen bağımsız değişkenlerden oluşan veriseti ve tüm bağımsız değişkenlerin dahil edildiği orijinal veriseti) kullanılarak modeller oluşturulmuştur. Mevcut durum ve farklı senaryolarda denenilen bu modeller göstermiş oldukları performanslara göre değerlendirilmiş ve genel bir tasarımı sunulan karar destek sistemlerinde kullanılacak uygun modeller belirlenmiştir.

Araştırma sonucunda, Rassal Ormanlar yöntemini kullanan Model5 en iyi performansı sergilemiştir.. Öğrencilerin yılsonu performansı üzerinde etkisi en fazla olan faktör “İlk dönem not ortalaması” olarak bulunmuştur.

**Anahtar Kelimeler:** Akademik Performans, Eğitimde Veri Madenciliği, Çok Katmanlı Algılayıcı, Rassal Ormanlar, Çokterimli Lojistik Regresyon

**Sakarya University Institute of Social Sciences Abstract of Master's Thesis**

**Title of the Thesis:** A Research on the Factors Affecting Freshmen Students' Performance and a Decision Support System Design for Predicting Their Academic Achievement Based on Those Factors

**Author:** Tuğrul C. HAKYEMEZ **Supervisor:** Prof. Erman COŞKUN

**Date:** 31.07.2015 **Number of Pages:** xiii (pretext) + 205 (main body)

**Department:** Management Info. Systems **Subfield:**

A transition has been existing in the approaches to education from "Tutorial-centered" to "Student-centered" in the higher education environment, so the concept of academic performance for undergraduate students is gaining importance. In this context, the first year in which students experience "sensitive" period of transition to new academic and social environments (Milem&Berger, 1997), has been subject to many studies both practical (Delen, 2010; Mishra et al., 2014) and theoretical (Tinto,1975; Astin, 1984) to be the basis of the academic performance shown in the higher education process. In the native literature research on the subject, there observed no "holistic" study determining the factors affecting the academic performance of the first year and predicts the year-end achievements based on these factors. In this context, it is intended to design a general system to provide decision support to managers in predicting the academic performance and determining factors by using a theoretical model demonstrated to the process of complex "the first year in the higher education" and data mining process parallel to this model.

In the study, "Freshman Student" survey was conducted to the freshmen who registered in undergraduate programs of the Business Faculty of Sakarya University in 2014- 2015 academic year. In this way, to be used in prediction study, traditional (age, gender,average point of high school grade, etc.) and non-traditional (general self-efficacy, academic self-efficacy, instutional integration) variables were obtained. As the method of the study, CRISP- DM, which offers a standardized methodology for data mining application process in many different fields (Wirth&Hipp, 2000), is adopted. As a part of the process; following data preparation (selection, cleaning, construct etc.) stage, machine learning techniques (Multilayer Perceptron, Random Forests) and a traditional statistical classifier(Multinomial Logistic Regression) were employed in different datasets (original dataset and the one created with correlation-based feature selection) to build predictive models. The appropriate models which were examined in both current situation and different scenarios were evaluated in accordance with their performance and the ones which will be used in a decision support systems, roughly presented in the current study, are determined.

As a conclusion, using the Random Forest method Model5 showed the best performance. It is found that " grade point average (GPA) of the first semester" factor is the most influential variable on the year-end performance.

**Keywords:** Academic Performance, Educational Data Mining, Multilayer Perceptron, Random Forest, Multinomial Logistic Regression

## GİRİŞ

Üniversite, Türkçede “Evrenşehir” anlamına gelmektedir. ÖSYM’nin tanımına göre ise; yükseköğretim kurumlarının temel birimi olan üniversite, kendisine bağlı fakülte, yüksekokul, konservatuvar, meslek yüksekokulları ve enstitülerle bunların alt bölümlerinden oluşmaktadır. Üniversitelerin tarihsel gelişim sürecine göz atıldığında; geçmişten günümüze yenilikçi ve bilgi üreten kimliği bu kurumlara toplumsal anlamda bazı misyonlar yüklemiştir. Dünya çapında saygın üniversitelerden birisi olan Cambridge Üniversitesi misyonunu “ mükemmelliğin en üst uluslararası seviyelerinde eğitim, öğrenme ve araştırmayı takip etmek suretiyle topluma katkı sağlamak” olarak tanımlamıştır. Tanımdan da anlaşılacağı üzere üniversiteler topluma değer katmak ile mükellef olan kurumlardır. Bunu gerçekleştirmek için ise eğitim ve öğretim ile nitelikli eleman yetiştirmeli ve araştırma ile topluma yenilikler sunmalıdır.

Günümüzde, üniversitelerin bu amaçlara hizmet ederken ne ölçüde başarılı olduğu ile toplumdaki itibarı doğru orantılı olarak değişmektedir. Yükseköğretim camiası ve toplum nezdinde saygın bir yere sahip olmak ise, üniversiteler açısından yetenekli öğrencileri ve öğretim üyelerini cezbederek eğitim kalitesinin artırılması gibi avantajları beraberinde getirmektedir. Baird’e (1967) göre öğrenciler üniversite tercihlerinde öğretim elemanı kalitesi, yüksek okul itibarı ve özel bir müfredata önem vermektedirler. Ardıç (2011: s. 10) ise bunlardan farklı olarak; üniversitenin ulusal ve uluslararası düzeyde proje üretebilme kapasitesi, eğitim kalitesini belirleyen mezunların iş bulma kapasitesi, bilimsel toplantı ve kongre düzenleyebilme ve mediko sosyal hizmetler, spor, kültür ve sanat etkinliklerinin üniversite tercihlerinde etkili olabileceğini belirtmiştir. Bu noktada üniversiteler itibarlarını koruyabilmek adına yetenekli öğrencilerini ve öğretim üyelerini bünyelerinde tutmak durumundadırlar. Öyle ki bu duruma üniversitelerin performanslarının değerlendirilmesinde büyük bir önem atfedilmektedir. U.S. News’in 2013 dünya üniversiteler sıralamasına etki eden göstergelerde akademik itibar %22,5; devamlılık ve mezuniyet oranı %22,5 luk bir paya sahiptir (Morse, 2014). Bu oranlardan da anlaşılacağı üzere akademik saygınlık ve

devamlılık oranları toplamda sıralamaya etki eden faktör ağırlıklarının neredeyse yarısına sahiptir.

Akademik devamlılık, ya da bir başka deyişle, öğrenci devamlılık oranı, bir yükseköğretim kurumunda eğitimine devam oranlarını yüzde olarak ifade eden bir ölçüdür. Daha açık olarak ifade etmek gerekirse, lisans programlarında kayıt yenileyen öğrencilerin bir önceki dönemde kayıtlı olan tüm öğrencilere oranı olarak belirtilebilir (IPEDS, 2015). A.B.D. gibi yükseköğretim dinamiklerinin öğrencilerin kayıt yenilememe kararlarını kolaylaştırdığı ücretli bir eğitim anlayışının hakim olduğu ülkelerde, akademik devamlılık oranı, Türkiye gibi öğrencilerin okulu bırakma oranlarının çok daha düşük olduğu ücretsiz ya da daha makul seviyelerde ücret politikalarının uygulandığı ülkelere göre daha fazla önem arz eden bir performans göstergesi olarak düşünülmektedir. Öyle ki, 1991 yılı itibariyle A.B.D.’de yürürlüğe giren “ Federal Öğrenci Bilme Hakkı ve Kampüs Güvenliği” yasası, her bir üniversitenin eğitim programlarının kalitesini tamamlama ve mezuniyet oranları gibi göstergeler vasıtasıyla ifşa etmesini zorunlu tutmaktadır (Astin A. , 1997: s. 647). Dolayısıyla üniversiteler eğitim kalitelerini ve başarılarını üst düzeylerde tutmak mecburiyetindedirler. NCES (National Center for Education Statistics) web sitesi üzerinden kamuoyuyla paylaşılan bu verilerin üniversite tercihinde bulunacak adaylara seçimleri konusunda yardımcı olması öngörülmüştür. Bu oranların saygınlık ve üniversite sıralamalarında önemli bir paya sahip olduğunun bilincinde olan üniversite yöneticileri ve araştırmacılar ise, bu performans göstergelerine etki eden faktörleri nedensellik çerçevesinde inceleyerek, performansı üst seviyelere taşıyacak önlem ve faaliyetler üzerinde hummalı bir çalışma içerisindeyler. Akademik başarı ve devamlılığın sağlanmasının prestijin yanında üniversitenin finansal durumuna etkisi de göz ardı edilmemelidir. Georgia Üniversite sistemine bağlı Valdosta Devlet Üniversitesi’nde yapılan bir araştırmaya göre; 10 öğrencinin yeni dönemde kayıt yaptırmamasının üniversiteye maliyeti 326,811 \$ olmuştur (Barker vd., 2004: s.79). Ayrıca üniversitelerin kurumsal performanslarına göre ödeneğe sahip olduğu eğitim sistemlerinde, akademik başarı ve devam oranlarının istenen düzeyde tutulamamasının üniversitelere maddi yükü katlanarak artacaktır. A.B.D.’de 6 yılda okulu bitirme oranlarının %54’ te kalması kamoyunca “kriz” olarak nitelendirilmektedir (Burke, 2013). Ancak bu durumun dünya üzerinde her ülke için aynı olduğunu söylemek yanlış



olacaktır. Türkiye gibi örgün eğitimin ücretsiz olduğu yükseköğretim sistemlerinde, öğrencilerin akademik başarı ve devamlılığının sağlanamamasının gelir kaybından ziyade kaynak israfına neden olacağı savunulabilir. Dolayısıyla, bu tip performans göstergelerinin, üniversitenin bulunduğu eğitim sisteminin dinamiklerini göz önünde bulundurarak ele almak gerekmektedir.

Lisans mezunu bir bireyin kariyeri boyunca elde ettiği toplam gelirin, lise mezunu bir bireyin kariyeri boyunca elde edeceği toplam gelirin yaklaşık iki katı olduğu tahmin edilmektedir (Day&Newburger, 2002). Bu tahmin, son yıllarda yükseköğretim öğrenci sayılarındaki büyük artışın nedenlerinden biri olarak görülebilir. Örneğin, Türkiye’de 2005 yılında 2.309.918 olan yükseköğretim öğrenci sayısı yaklaşık %150 ‘lik bir artış oranıyla 2013 yılı itibariyle 5.449.961 e ulaşmıştır. Yükseköğretim öğrenci sayılarında yaşanan bu muazzam artış, üniversitelerin iş yükünü arttırmış ve kaynakların etkin ve verimli bir şekilde kullanılmasını zorunlu kılmıştır. Bu kısıtlamalar altında, talebi karşılamak üzere sayıları her geçen gün artan üniversitelerin yükseköğretim sürecinde performansı günümüzde her zamankinden çok daha fazla ilgi çekmektedir. Bu kapsamda üniversite idarecileri, araştırmacılar, akreditasyon kuruluşları ve ilgili devlet kurumları bu noktada etkinlik ve verimliliği ortaya koyan kurumsal performansı çok daha dikkatli bir şekilde takip etmektedirler. Zamanında mezuniyet ve devamlılık oranları bu noktada takip edilen performans ölçütleridir. A.B.D. de Federal Yükseköğretim yasası akademik devamlılık oranını, üniversitenin etkinlik göstergelerinden biri olarak kabul etmektedir (Fike&Fike, 2008: s. 69).

Öğrencilerin yükseköğretim sürecinde göstereceği yüksek akademik performansın zamanında mezuniyet ve devamlılık oranları üzerinde olumlu bir etkiye sahip olacağı açıktır. Örneğin; araştırmacılar üniversitede ilk dönem not ortalaması ile akademik devamlılık oranı üzerinde anlamlı ilişkiler tespit etmişlerdir (Delen, 2010; Porter, 2008; Vandamme vd., 2007; Slim vd., 2014). Yükseköğretim sürecinde, akademik performans ve devamlılık üzerinde en fazla etkiye sahip olan yıl, üniversiteye hassas bir geçiş ve entegrasyon süreci olarak düşünülen ilk yıldır ve bu yılda gösterilen akademik başarının sonraki sınıflarda gösterilecek performansa temel teşkil ettiği düşünülmektedir (Olani, 2009; Tinto, 1988; Barefoot, 2000). Dolayısıyla üniversite idarecileri, öğrencilerin yükseköğretimde yaşayacağı ilk yıl deneyimini en iyilemeye çalışmakta ve bu yönde

destek programları sunmaktadırlar. A.B.D.'de üniversitelerin yaklaşık %95 inde ilk yıl öğrencileri destek programları sunulmaktadır (Jamelske, 2008: s. 375).

### **Araştırma Problemi**

Türkiye’de 2014-2015 akademik yılı itibariyle yükseköğretim lisans programlarına yeni kayıt yaptırmış 1,382,571 öğrenci bulunmaktadır (YÖK Bilgi Sistemi, 2015). Yeni gelen öğrenciler için, Türkiye’de birkaç özel üniversite haricinde hiç bir üniversitede ilk yıl destek programına rastlanmamıştır. Mevcut çalışmalara baktığımızda, yeni gelen öğrencilere yönelik destek programlarının oryantasyon eğitimlerinden ileri gidemediği görülmektedir. Dolayısıyla Türkiye’de yeni gelen öğrencilere gereken ihtimam gösterilmemekte ve bu konudaki eksiklik göze çarpmaktadır. Modern öğrenci merkezli üniversite eğitimi anlayışı çerçevesinde düşünüldüğünde bu durumun büyük bir sorun teşkil ettiği ortadadır. Öyleki Türkiye’nin de 2002 yılında beri üyesi olduğu tüm Avrupa’da yüksek öğretim ve akademik konularda standartlar geliştirmek ve ayrılıkları en aza indirgeyerek eğitim sistemlerini bağdaştırmak ve Avrupa’da birbiriyle tam uyumlu bir yükseköğrenim alanı yaratmak amacıyla oluşturulmuş Bologna süreci “öğrenci merkezli” bir eğitim anlayışını vurgulamaktadır. Türkiye’de yer alan üniversitelerin kaliteyi arttırmak adına öğrenciyi “merkez”e koyan bir sürece uyum sağlama çalışmalarında, uzun yıllardır üzerine çalışmalar yapılan “Yeni Gelen Öğrenciler”e yönelik bir program bulunmaması düşündürücüdür.

Ölçülemeyen ya da takip edilemeyen performans geliştirilemez. Dolayısıyla öğrencilerin akademik performansının takip edilerek, öğrenciye ya da gruba yönelik önlemler alınmasını sağlayacak her hangi bir sisteme rastlanmamıştır. Bu bağlamda öğrencilerin akademik performanslarının takip edilmesi ve bu performansa etki eden faktörlerin araştırılmasına dönük araştırmalarda nitelik ve nicelik olarak eksiklik göze çarpmaktadır.

### **Çalışmanın Amacı**

Mevcut çalışmada, Türkiye’de yeni gelen öğrencilerin akademik performansına etki eden faktörlerin araştırılması ve bu faktörlere bağlı olarak yılsonu not ortalamalarının tahmin edilmesi amaçlanmaktadır. Çalışmadan elde edilecek sonuçlar, akademik performans anlamında risk grubunda yer alan öğrencilerin belirlenmesi ve hassas bir

geçiş dönemi yaşayan bu öğrencilere dönük alınacak tedbirlere yönelik genel bir karar destek sistemi tasarlanması düşünülmektedir. Ayrıca, elde edilecek sonuçlar ışığında, Türkiye’de henüz yükseköğretim bünyesinde var olmayan “yeni gelen” öğrencilere yönelik kapsamlı destek programlarının temellerinin atılması hedeflenmektedir.

### **Çalışmanın Konusu**

Bu doğrultuda, çalışmanın ikinci bölümünde konuyla ilgili temel kavramlara değinilerek, dünya çapında, özellikle “Yeni Gelen Öğrenci”lere yönelik olmak üzere yapılan betimleyici ve tahminleyici çalışmalara yer verilecektir. Uygulama bölümünde ise öncelikle, “Yeni Gelen Öğrenciler” in akademik performansına etki eden faktörlere dönük kavramsal bir model ortaya konulacaktır. Sonrasında ise, Sakarya Üniversitesi İşletme Fakültesi bünyesindeki lisans programlarına kayıtlı “yeni gelen” öğrenciler ile yürütülen çalışmanın metodolojisi, analiz sonuçları ve öğrencilerin yılsonu not ortalamalarına ilişkin tahminler paylaşılacaktır. Sonuç kısmında ise bulgulardan, önerilerden, çalışmanın kısıtlarından ve gelecekte yapılması öngörülen çalışmalardan bahsedilecektir.

### **Çalışmanın Önemi**

- Çalışma, yerli literatürde gözlemlenmeyen bütüncül (hem demografik hem de algısal değişkenlerin kullanıldığı) bir tahminleme çalışmasıdır.
- Çalışmada, Türkiye’deki mevcut akademik ve sosyal ortamda öğrencinin başarısını irdeleyen teorik bir model ortaya konmuştur. Üniversitelerin müdahale alanlarını göstermesi ve daha önce yerli literatürde rastlanmaması açısından önemlidir.
- Öğrencilerin algı ve başarı sınıflarının sınırlarının belirlenmesinde “kümeleme analizi” kullanılmıştır. Bu şekilde literatürde rastlanan bir çalışma yoktur.
- İlk yıl ya da lisans programlarının herhangi bir seviyesinde Türkiye’deki öğrencilerin akademik ve sosyal ortama adaptasyonlarını ölçmeye yönelik Türkçe’ye çevrilmiş bir ölçek yoktur. Bu çalışmada Terenzini&Pascarella (1980) tarafından geliştirilen French&Oakes (2004) tarafından revize edilen Institutional Integration Scale, “Kurumsal Entegrasyon Ölçeği” adıyla Türkçe’ye

çevrilmiş ve alt boyutlarının tahminleme çalışmasında kullanılabilmesi için değişkenlere dönüştürmek üzere faktör analizi uygulanmıştır.

- Chen vd. (2001) tarafından geliştirilen “Genel Özyeterlik Ölçeği” Türkçe’ye çevrilerek uygulanmıştır.
- Tahminlemede kullanılan yöntemler, verisetleri ve farklı senaryolar bakımından özgündür.
- Türk Yükseköğretim çevrelerinde öğrencilerin performansına etki eden faktörlerin belirlenmesine ve bunlara bağlı olarak tahminlenmesine yönelik bir karar destek sistemi tasarımına rastlanmamıştır.

### **Çalışmanın Yöntemi**

Mevcut çalışma, her bir sektör için standart bir “Veri madenciliği” metodolojisi sunan CRISP-DM (Problem tanımlama - Veriyi anlama- Veri hazırlama- Modelleme- Değerlendirme- Yayılım)’e uygun olarak ilerlemiştir. Öğrencilerin akademik performansı üzerinde etkili olabilecek demografik değişkenler ve öğrencinin akademik ve sosyal çevresine olan algısına yönelik değişkenleri elde edebilmek üzere “Yeni Gelen Öğrenci” anketi uygulanmıştır. Bu verileri tahminleme modelinde kullanabilmek için, demografik değişkenlerde olduğu gibi algısal değişkenler de nominalleştirilmiştir. Bu doğrultuda, katılımcıları algıları ve başarılarına göre sınıflara ayırırken, sınıflar arası eşik değerleri “kümeleme analizi” kullanılarak belirlenmiştir. Çok boyutlu ölçeklere “Faktör Analizi” uygulanarak her bir boyuta karşı tutum derecesi bir bağımsız değişken olarak kabul edilmiştir. Tüm değişkenler elde edildikten sonra, bu değişkenlerin bağımlı değişken ile aralarındaki Pearson  $r$  ve Kendall’s tau korelasyon katsayıları hesaplanarak, anlamlı düzeyde korelasyon gözlenmeyen bağımsız değişkenler çıkarılmıştır. Böylece, yeni bir veriseti oluşturulmuş ve bu yeni veriseti ile orijinal veriseti farklı yöntemler kullanılarak, amaca uygunluk yönünden değerlendirilmiştir. Öğrencilerin başarı durumlarını sınıflandırmak için makine öğrenme yöntemlerinden Yapay Sinir Ağları ve Karar Ağaçları ve klasik istatistiksel sınıflandırıcılardan Çokterimli Lojistik Regresyon Analizi kullanılmıştır. Verisetleri ve sınıflandırıcılar kullanılarak kurulan modeller, mevcut durum ve farklı senaryolarda denenmiştir. Tüm modeller, sonrasında çalışmaya özgü “performans puan”larına göre karşılaştırılmıştır.

Analizler sonucu en yüksek performansa sahip modellerin yayılımı için genel bir karar destek sistemi tasarımı önerilmiştir.

## **BÖLÜM 1: AKADEMİK PERFORMANSA İLİŞKİN KAVRAMLAR VE LİTERATÜR TARAMASI**

Çalışmayı kavramsal açıdan sağlam temellere oturtabilmek için, bazı temel kavramları açıklamak yerinde olacaktır. Bu kapsamda öncelikle performans kavramı açıklığa kavuşturulacak ve performans ölçümü konusunda temel bilgiler sağlanacaktır.

### **1.1. Performansın tanımı**

Performans, en genel anlamıyla, bir kişinin, makinenin vs. bir aktivite ya da işi ne oranda yapabildiği olarak tanımlanabilir (Cambridge, 2015).

Şimdiye dek yapılan çalışmalarda ise performans kavramı araştırmacılar tarafından genel anlamda farklı şekillerde tanımlanmıştır. Kahraman'ın aktardığına göre (2014) performans; görev çerçevesinde önceden belirlenen ölçütleri karşılayacak biçimde, görevin yerine getirilmesi ve amacın gerçekleştirilmesi yönünde ortaya konan mal, hizmet ya da düşüncedir (Pugh, 1991)

Bununla birlikte performans kavramı, kullanıldığı terminolojiye, belirlenen kriterlere veya sürece bağlı olarak farklı anlamlara gelebilmesinden dolayı, genel olarak bir tanımının yapılabilmesi oldukça zordur. Örneğin; Kasnaklı (2002: s.131) performansı; işletme perspektifinden, işletmeyi oluşturan tüm bileşenlerin – makineler, çalışanlar, yönetim, varlıklar, çevre etkileşimi- birlikte ortak çaba ile elde edilen toplam sonuç olarak ifade etmektedir. Bir diğer tanımda Çakmak ve Ocaklı (2006: s. 214) performansı, örgütsel bir bakış açısıyla, örgütün maddi ve manevi unsurlarından en üst düzeyde yararlanma derecesi olarak belirtmiştir. Organizasyonları oluşturan bireylere yönelik bakış açısıyla ise, “görev” performansı, sorumluların, ya gerekli malzeme ya da hizmet ile dolaylı olarak ya da teknolojik sürecin bir kısmını yürüterek direkt olarak organizasyonun teknik çekirdeğine katkı sağlayan faaliyetlerde göstermiş oldukları etkinlik olarak ifade edilebilir. Hooley vd. (2005: s. 20) pazarlama alanında yaptığı çalışmasında daha iyi bir müşteri performansını, göreceli olarak hizmet edilmesi daha verimli olacak tatmin edilmiş ve sadık müşteri kitlesi oluşturma; daha iyi bir pazar performansını ise satış hacminin ve pazar payının arttırılması olarak tanımlamıştır. Daha spesifik bir alanda örnek vermek gerekirse; Neely vd. (2015: s. 80) performansın kalite ile ilintili bir boyutu olan ürün güvenilirliğini ele almıştır ve etkinlik bakımından, daha yüksek bir ürün güvenilirliği seviyesi yakalanması daha fazla memnun edilmiş müşteri;

verimlilik bakımından ise daha az kusurlu ve garanti dışı kalan ürünler ile maliyetin düşürülmesi anlamına geleceğini belirtmiştir. İstenen performans seviyelerine ulaşılmasının katkıları etkinlik ve verimlilik yönleriyle vurgulanmıştır. Lembas (1995) yaptığı çalışmada performans kavramını irdelemiştir. Bu çalışmada, diğer çalışmalardan farklı olarak, performansın özellikle yönetim açısından, geçmiş başarılarından ziyade geleceğe yönelik olduğunu ve değerlendirilmekte olan birimin kabiliyetleriyle ilgili olduğu vurgulanmıştır. Aynı çalışmada; performansın belirlenmiş ölçütlere bağlı olarak değişkenlik gösterdiği varsayılarak, ölçütlerin doğru bir şekilde belirlenebilmesi için performansın ölçüldüğü sürecin iyi anlaşılması gerektiği savunulmuş ve tüm bu ifadeler nedensel bir modelde belirtilmiştir.

Verilen farklı tanımlardan da anlaşılacağı üzere performans ifade edildiği her bir alan ve süreç için farklı bir karşılık bulmaktadır. Ancak performans kavramını tanımlamak her ne kadar zor olsa da; performansın ölçülmesi, değerlendirilmesi ve geliştirilmesi, alana bağlı olmaksızın; üzerinde çalışılan ortak konu başlıklarıdır.

Bugüne dek performans ölçümü ve değerlendirilmesi üzerine çalışmalar yapılagelmiştir. Performansı oluşturan unsurlar nicel ve nitel olarak ölçüldüğünde, elde edilen sonuç önceden belirlenen misyon, amaç ve hedeflerle paralellik gösteriyorsa, istenilen performans düzeyine yaklaşılmış ya da erişilmiş demektir (Çakmak&Ocaklı, 2006: s. 214). Performansın ölçülmesine paralel olarak değerlendirilebilmesi için, bazı hedef değerler belirlenmektedir. Bu hedefler belirlenirken bazı hususlar göz önüne alınmalıdır. Diğer bir deyişle her bir hedef SMART özelliklere sahip olmalıdır (Doran, 1981).

**S- Specific:** Hedefler yapılan iş ile ilgili olmalı, önceden uzlaşılmalı, belirli ve somut olmalı ve performans aktörleri kendilerinden ne beklenildiğini bilmelidir.

**M- Measurable:** Standartlar/hedefler objektif ve “ölçülebilir” olmalıdır. Hedefin başarısının nasıl ölçüleceği belirlenmelidir. Standartlar nicel ise ölçüm daha kolaydır, ancak nitel olduğunda biraz daha dikkat etmek gerekir.

**A-Achievable:** Standartlar zorlayıcı olmakla birlikte “ulaşılabilir” olmalıdır. Hedefler çalışanın asla başaramayacağı zorlukta olmamalı ve çok kolay başarılabilecek düzeyde de olmamalıdır.

**R-Reasonable:** Hedefler “gerçekleşebilir” (uygun, makul) olmalıdır.

**T-Time-Bound:** Hedeflerin gerçekleştirilmesi “zaman sınırlı” (altı aylık veya yıllık) olmalıdır.

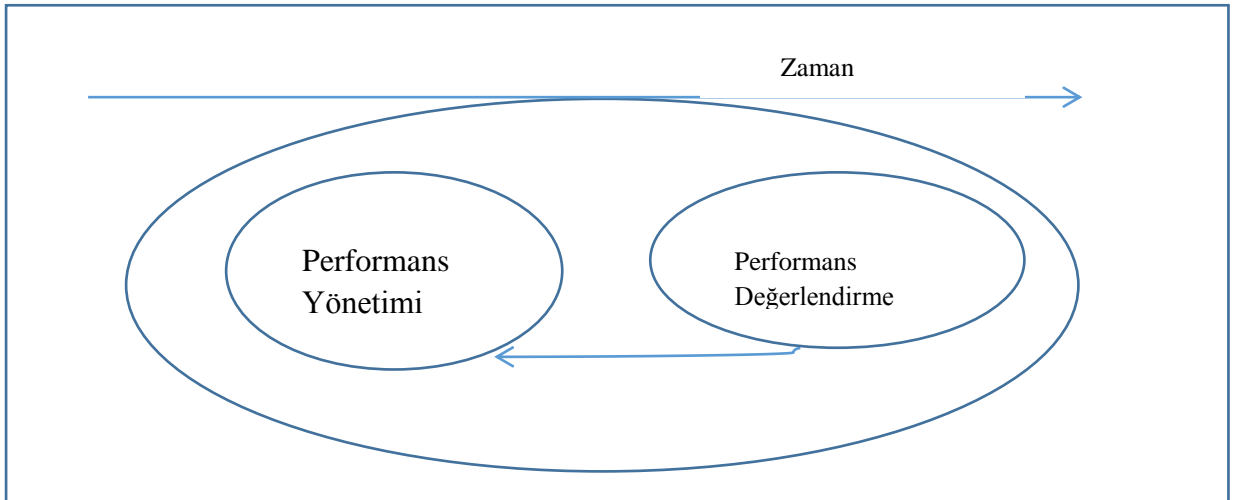
Üniversitelerin bilime katkı sağlamayı ve alanında yetkin bireyler yetiştirmeyi amaçlayan organizasyonlar olduğu göz önüne alınırsa, örgütsel açıdan performansı oluşturan temel unsurların tanımlanması performans kavramının sağlam temellere oturtulması açısından önem arz etmektedir. Özmutaf (2007: s. 44-50) yaptığı çalışmada performansın 3 temel unsuru belirlemiştir. Bu unsurlar, yönetsel unsurlar (misyon ve vizyonun belirlenmesi, liderlik, katılımcılık, iletişim motivasyon, stres yönetimi ve performans değerlendirme) bireyden kaynaklanan unsurlar (sosyodemografik özellikler: yaş, cinsiyet, medeni hal, eğitim düzeyi, bireyin kültürel yapısı, uzmanlık alanı ve ücret düzeyi algısı) ve diğer unsurlar (bireyin spesifik alanlardaki mesleklerde çalışması, dış çevre şartlarına uyum bağlamında gelişen teknolojilerin örgüte aktarılması, sosyal güvence konusundaki yetersizlikler, kriz dönemlerinde ağırlaşan koşullar, işyeri ile ikamet edilen yerin uzaklığı, madde bağımlılığı) olarak belirtilmiştir.

## 1.2 Performans Yönetim Sistemleri

Ölçülemeyen bir şeyin geliştirilemeyeceği gerçeği, performansın ölçülmesi ihtiyacını doğurmuştur. Performansın doğru bir şekilde ölçülmesi her zaman araştırmacıların ve yöneticilerin ilgi alanları arasında yer almıştır. Performans ölçümünün zaman içerisindeki değişimi 4 kısımda incelenmiştir. Öneriler, çerçeveler, sistemler organizasyonlar arası performans ölçümleri bu evrimin aşamaları olarak düşünülebilir (Folan&Browne, 2005: s.663). Bu silsile yorumlanacak olursa; performans ölçümü öncelikle performansa etki eden ölçütlerin belirlenmesi ile başlamıştır denebilir. Bu süreç çoğunlukla nesnel ve çalışanlar ve yöneticilerin bulunduğu toplantılar (beyin fırtınası, DELPHI yöntemi vs.) sonrasında süreç için önemli görülen ölçütlerin belirlenmesi ile sona erer. Ölçütlerin belirlenmesi sürecinde; yöneticilerin şu temel iki soruyu cevaplaması gerekmektedir: 1- Niçin ölçmek istiyoruz? 2- Neyi ölçmek istiyoruz? (Lembas; 1995: s.24) Ancak bu sorulara açık ve herkesin mutabık kalacağı cevaplar verildiğinde önemli ölçütler belirlenmiş olabilir. Bu ölçütlerin belirlenmesi performansın takibini sağlayacak sistemin tasarımı ve kurulması açısından büyük önem arz etmektedir. Bu ölçütlerin belirlenmesi ve buna bağlı olarak performans ölçümünün



sistematiik hale getirilmesi performans ölçüm sistemleri kavramını ortaya çıkarmıştır. Performans ölçüm sistemleri; eylemlerin hem etkinliğini hem de verimliliğini ölçmek üzere kullanılan bir metrikler kümesi olarak tanımlanabilir (Neely, 1994). Bu noktada etkinlik ve verimlilik kavramlarını açıklayacak olursak; etkinlik, en genel anlamıyla; bir faaliyet, hareket ya da davranışın, olanaklı olduğu kadar, yöneltilmiş bulunduğu amaca ulaşma derecesi; verimlilik ise; geniş anlamıyla amaçlara ulaşmada araçların duyarlılık ve etkinliğini ölçen soyut bir kavram olarak düşünülebilir (ekodialog: 2015). Etkinlik ve verimliliği ölçmek için gerekli olan metrik ise; belirli bir sürece ait durumun takibi ve değerlendirilmesi için kullanılan bir ölçü birimi olarak tanımlanır. Performans ölçüm sistemleri her ne kadar organizasyonun mevcut durumunu ortaya koymaya büyük katkı sağlamış olsa da, gelecekteki durum hakkında bilgi sağlama ve mevcut performansın iyileştirilmesi konusunda yetersiz kalmıştır. Bu da performans yönetimi kavramının ortaya çıkmasına önayak olmuştur. Performans yönetimi, performans ölçümü tarafından desteklenen bir felsefedir (Lembas, 1995: s.34). Aralarındaki bağıntı şu şekilde betimlenebilir.



**Şekil 1: Performans ölçümü ve yönetimi arasındaki ilişki (Lembas, 1995: s.34)**

Şekilden de anlaşılacağı üzere performans ölçümü ve performans yönetimi birbirlerini sürekli takip eden iki süreçtir. Bu noktada performans yönetimi tanımlanacak olursa; performans yönetimi, üzerinde mutabık kalınan performans hedeflerinin belirlenmesine yardımcı olarak, kaynakları tahsis ederek ve önceliklendirerek, bu hedeflere ulaşmak amacıyla mevcut politikaları veya programın yönünü değiştirmek ya da onayına sunmak amacıyla yöneticiye ileterek ve bu hedeflerin peşinden giderken performans sonuçlarının paylaşılmasını sağlayarak, organizasyon kültüründe pozitif değişim

yaratmak amacıyla performans ölçüm bilgilerinin kullanımınıdır (Amaratunga& Baldry; 2002: s.218). Daha basit bir diğer tanıma göre performans yönetimi; faaliyet yöneticilerinin sahibi olduğu ve yönlendirdiği ve organizasyon, takım ve bireysel performansı iyileştirmek üzere tasarlanmış bir süreçtir (Armstrong, 2000: s.69).

Lembas (1995: s.35) a göre güçlü bir performans sistemi, şu ölçüler üzerine inşa edilendir:

- Kendi kontrol sınırları içerisinde bireylere otonomi vermesi
- Neden ve sonuç ilişkisini yansıtması
- Bireyleri yetkilendirmesi ve içermesi
- Tartışmaya bir zemin oluşturmak suretiyle sürekli iyileştirmeyi desteklemesi
- Karar vermeyi desteklemesi

Belirtilen özelliklerdeki performans kriterleriyle mevcut durumun karşılaştırılabilmesi için ise performans ölçüm sistemlerine ihtiyaç duyulmaktadır. Eylemlerin etkinlik ve verimliliğini ölçmek için gerekli olan metrik ise; belirli bir sürece ait durumun takibi ve değerlendirilmesi için kullanılan bir ölçü birimi olarak tanımlanır.

Performans ölçüm sistemleri ve performans yönetim sistemlerinin tanımlanması, üniversitelerde bu kapsamda yapılan çalışmaların daha iyi anlaşılabilmesi için yerinde olmuştur. Bu amaç doğrultusunda eğitim çevrelerinde performans ölçümü ve performans yönetimi faaliyetleri açıklanmaya çalışılacaktır.

### **1.3 Öğretimde performans ölçümü ve yönetimi**

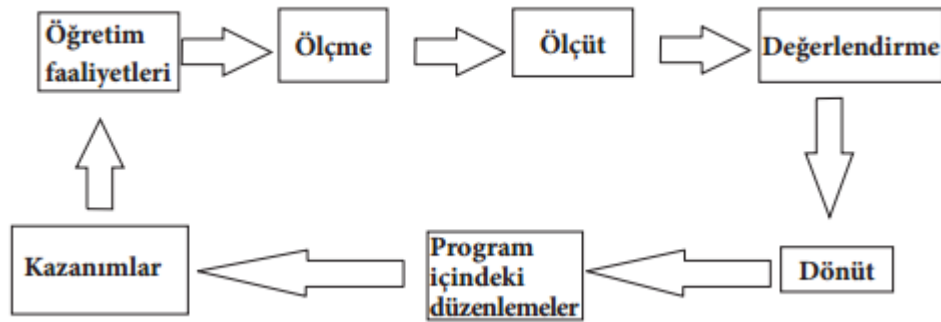
Eğitim ve öğretim, birbirlerine çok yakın kavramlar olmakla birlikte; birbirlerinin yerine kullanılabilen ifadeler değildir. Ertürk (1997) eğitimi, “bireyin davranışında kendi yaşantısı yoluyla ve kasıtlı olarak istenen değişimi meydana getirme süreci olarak tanımlamıştır. Öğretim ise; öğrenmenin gerçekleşmesi için planlanan, kasıtlı ve sistematik bir eğitim olarak tanımlanabilir (Demirel, 2003). Bu noktada “örgün eğitim” kavramı ile “öğretim” birbirlerinin yerlerine kullanılacak kavramlar olarak ele alınabilir (Laska, 1984: s.205). Tanımlardan da anlaşılacağı üzere eğitim, yaşam boyu devam ederken, öğretim belirli bir müfredat dâhilinde ve belirli bir zaman diliminde devam etmektedir. Bu durum, performans kriterlerinin sahip olması gereken özelliklerden “Zaman sınırlı (Time Bound)” ile uyum göstermektedir. Dolayısıyla

performans ölçme ve yönetiminin öğretim faaliyetlerinde de gerçekleştirildiğini belirtmek yanlış olmayacaktır.

Uygulanan bir öğretim programının başarılı olup olmadığına karar verilmesi ve başarının derecesinin ortaya konması, performansın geliştirilmesinde önemli bir yere sahiptir. Zira mevcut öğretim programında risk grubunda yer alan öğrencilerin belirlenmesi ve güncel ihtiyaçlara göre programın yeniden tasarlanması ancak başarılı bir ölçme ve değerlendirme sistemiyle mümkün olacaktır.

Sınavlar, öğretmen görüşleri ve portfolyolar gibi ölçme tekniklerinin bazı özelliklere sahip olması beklenmektedir. Temel (2010: s.3) bu özellikleri şu şekilde sıralamıştır.

- 1- **Geçerlilik:** Ölçme aracının ölçmeyi amaçladığı özelliği, başka bir özellikle karıştırmadan, doğru olarak ölçebilme derecesidir
- 2- **Güvenilirlik:** Ölçme aracının, aynı özelliği ölçmek için her uygulandığında yaklaşık aynı sayısal sonucu vermesidir
- 3- **Nesnellik:** Bir testi kim, ne zaman, nerede, hangi koşullar altında puanlarsa puanlasın sonuç değişmez ise test nesnelidir.
- 4- **Kullanışlılık:** Hazırlanması, çoğaltılması, uygulanması, puanlanması, yorumlanması için harcanan emek, zaman ve para bakımlarından ekonomik olan test kullanışlıdır.
- 5- **Örnekleyicilik:** Test, dersin konularını ve hedef davranışlarını örneklemelidir.
- 6- **Ayrırtedicilik:** Sorular, ölçülecek davranışlara sahip olanlarla olmayanları ayıracak nitelikte olmalıdır.



Şekil 2 : Öğretim sürecinde Ölçme ve Değerlendirmenin yeri (Başol, 2012: s.4)

Şekilden de anlaşılacağı üzere, değerlendirme ölçme sonuçlarının belirli bir ölçüte vurularak bir yargıya ulaşılmasıdır (Temel, 2010: s.4). Öğretim sürecinde dikkat edilmesi gereken bir diğer nokta değerlendirmenin ölçüm sonuçlarını kullanmasıdır. Doğru bir ölçüm sonucuna dayanmayan değerlendirmeler mesnetsiz olacaktır.

#### **1.4 Türkiye’de Ölçme Ve Değerlendirme Faaliyetleri**

Türkiye’de de, dünyanın pek çok ülkesinde olduğu gibi, eğitim-öğretimin her aşamasında ölçme ve değerlendirme faaliyetleri uygulanmaktadır. Mevcut çalışmada tüm aşamalardaki ölçme ve değerlendirme sistemlerinden bahsetmek mümkün olmayacağından, çalışmanın kapsamına paralel olarak, üniversiteye giriş sınavlarından ve üniversitedeki ölçme ve değerlendirme faaliyetlerinden bahsedilecektir.

##### **1.4.1 Türkiye’de Üniversiteye Giriş Sınavları**

Günümüz dünyasında öğrenci sayılarının hızlı bir şekilde artması ve eğitim-öğretim faaliyetlerindeki nesnellik kaygısı, ölçme ve değerlendirme faaliyetlerinin, özellikle üniversiteye geçiş sınavlarında geçerli olmak üzere, tek merkezden yapılması ihtiyacını da beraberinde getirmiştir. Ülkemizde ise bu faaliyetler Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından organize edilmektedir.

Türkiye’de üniversiteye geçiş sınavlarının zaman içerisinde yaşadığı değişime kısaca bir göz atılacak olursa;

- Cumhuriyet döneminden 1960’lı yıllara kadar olan süreçte; lise mezunlarının sayısının az olmasından dolayı, pek çok fakülte başvuran adayları sınavsız olarak kabul etmiştir. Zaman içerisinde başvuran öğrenci sayısını kontenjanları aşması üzerine çeşitli önlemler (başvuru sırasını dikkate alma ve kontenjan kadar adayı kabul ettikten sonra kayıtları durdurma, lise bitirme derecesine göre adayları kabul etme vs.) almak zorunda kalmışlardır.
- 1960’lı yıllarda üniversiteler kendilerine özel giriş sınavları düzenlemeye başlamıştır. Bu durum farklı üniversitelerin giriş sınavlarının aynı gün ve saate denk gelmesi gibi farklı olumsuz durumları ortaya çıkarmıştır. Üniversiteler bu gibi durumların önüne geçebilmek için işbirliğine gitmiştir. Ancak yapılan bu işbirliği istenen sonuçları verme konusunda yetersiz kalmıştır.

- 1974 yılında, Üniversitelerarası Kurul, üniversiteye giriş sınavının tek merkezden yapılmasını uygun görmüş ve 1750 sayılı Üniversiteler Kanunu'nun 52'nci Maddesine dayanarak 19 Kasım 1974 tarihinde Üniversitelerarası Öğrenci Seçme ve Yerleştirme Merkezini (ÜSYM) kurmuştur. Üniversitelere öğrenci seçme ve yerleştirme işlemleri, 1981 yılına kadar bu merkez tarafından yürütülmüştür.
- 1981 yılında Ölçme, Seçme ve Yerleştirme Merkezi kurulmuş ve bununla birlikte üniversite sınavları Nisan ayında Öğrenci Seçme Sınavı (ÖSS) ve Haziran ayında Öğrenci Yerleştirme Sınavı (ÖYS) olarak iki basamaklı sınavlar olarak uygulanmaya başlamıştır.
- 1982 yılından itibaren adayların ortaöğretim başarı puanları üniversite sınavları puanlarına Ortaöğretim Başarı Puanı (OBP) olarak yansıtılmaya başlanmıştır.
- 1999 yılında iki basamaklı sınavın ikinci basamağı kaldırılmış, sınav ÖSS adı altında tek basamaklı bir sınav hâline getirilmiştir.
- Mevcut üniversiteye geçiş sınavları son halini 2010 yılında almıştır. Buna göre üniversiteye geçiş sınavları 2 aşamalı olarak gerçekleştirilmektedir (ÖSYM, 2015).

**Yükseköğretime Geçiş Sınavı (Birinci Aşama)**; Ortaöğretimden yükseköğretime geçişte iki aşamalı sınavın birinci aşaması “Yükseköğretime Geçiş Sınavı” olarak adlandırılan ortak ve tek bir sınavdır. “Yükseköğretime Geçiş Sınavı”, ortaöğretimi başarı ile tamamlayan ve yükseköğrenim görmek isteyen kişilerin tabi tutulacağı, yükseköğretime geçiş için yeterliliği ölçen bir sınav olup bu sınavla,

- a) Açık öğretim programları ile örgün önlisans programlarına yerleştirmede esas alınacak olan başarı puanı,
- b) Lisans programlarına yerleştirme amacıyla yapılacak “Lisans Yerleştirme Sınavları'na” girebilmek için aranan asgari başarı puanı belirlenir.

**Lisans Yerleştirme Sınavları (İkinci Aşama)**; “Lisansa Yerleştirme Sınavları”, adayların ders düzeyindeki bilgi ve yeteneklerini ölçen ve açık öğretim dışındaki örgün lisans programlarına yerleştirmede esas alınacak başarı puanını belirleyen sınavlardır.

## **Lisansa Yerleřtirme Sınavları;**

1. Matematik, Geometri Sınavı (LYS 1),
2. Fen Bilimleri (Fizik, Kimya, Biyoloji) Sınavı (LYS 2),
3. Türk Dili ve Edebiyatı, Coğrafya 1 Sınavı (LYS 3),
4. Sosyal Bilimler (Tarih, Coğrafya 2, Felsefe grubu) Sınavı (LYS 4),
5. Yabancı Dil Sınavı (LYS 5)

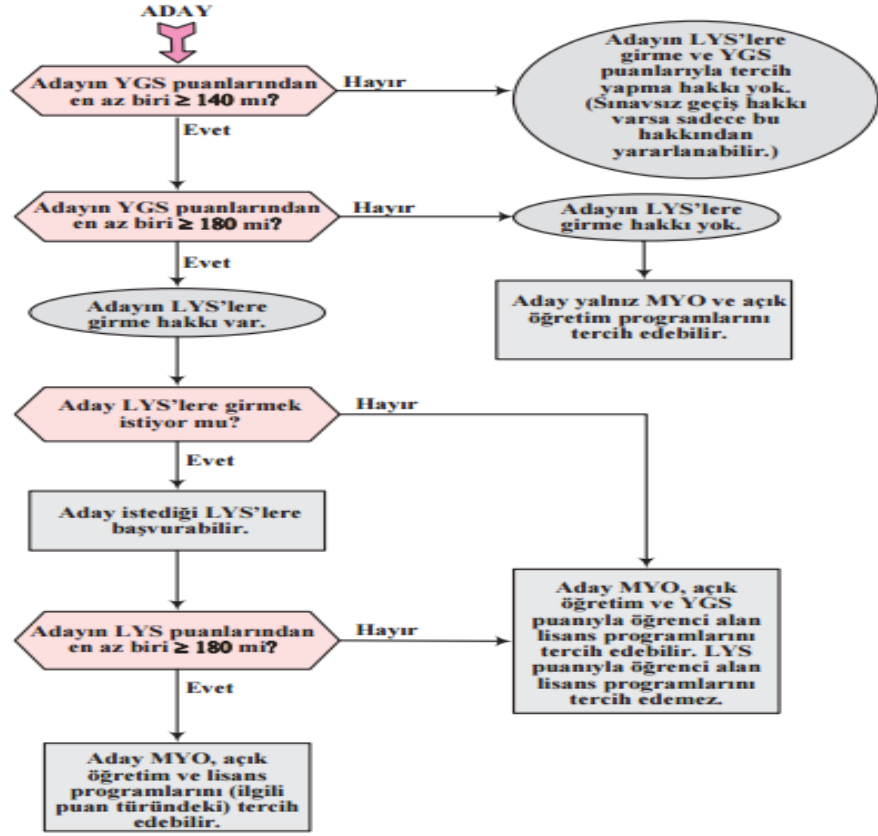
olmak üzere beř alanda, Haziran ayında iki hafta sonunda ve ayrı ayrı 4 oturumda yapılır.

### **1.4.2 Yükseköğretime Geçiř Sınavı (YGS) ve Lisans Yerleřtirme Sınavı (LYS)**

Türkiye’de yükseköğretim programlarına yerleřebilmek için tek ölçüt iki aşamalı geçiř sınavlarından alınan puanlardır. Dolayısıyla bu sınavların mevcut yapısını anlamak, üniversiteye geçiřte mevcut ölçme ve deęerlendirme faaliyetlerine ışık tutması açısından önemlidir.

#### **1.4.2.1 Yükseköğretime Geçiř Sınavı (YGS)**

Üniversiteye girmek isteyen öğrencilerin girmesi gereken sınavların ilk ayağıdır. Bu sınavda Türkçe, Sosyal Bilimler, Temel Matematik ve Fen Bilimleri testleri yer almaktadır. Bu sınavda her alandan 40’ar olmak üzere toplam 160 soru yer almaktadır. Sınav süresi 160 dakikadır. Bu sınavdan alınabilecek minimum puan 100,000, maksimum puan ise 500.000 dür. 140 puan barajının altında kalan adaylar; YGS puanları ile herhangi bir yükseköğretim programını tercih etme hakkı elde edemezler. 140,000-179,999 arası puan alan adaylar ise yalnızca meslek yüksekokullarını ve önlisans ve açıköğretim programlarını tercih edebilirler. Bu sınavdan 180,000 puan ve üzeri alan adaylar ise, üniversiteye geçiř sınavlarının ikinci ayağı olan Lisans Yerleřtirme Sınavlarına girmeye hak kazanırlar. Bu sistemi görselleřtirmek anlamında, ařağıdaki akıř řeması faydalı olacaktır.



**Şekil 3 : Yükseköğretime geçiş sınavı (YGS) puanlarına göre tercih edilebilecek yükseköğretim programları (ÖSYM, 2015)**

#### 1.4.2.2. Lisans Yerleştirme Sınavı (LYS)

LYS puan türleri ile öğrenci alan yükseköğretim programlarına girmek isteyen adayların, YGS'ye ek olarak tercih edecekleri yükseköğretim programları için gerekli olan LYS'lere de girmeleri zorunludur. LYS sınavları 4 oturumda gerçekleştirilmektedir. Bu sınavlar: Matematik Sınavı (LYS-1), Fen Bilimleri Sınavı (LYS-2), Edebiyat – Coğrafya Sınavı (LYS-3), Sosyal Bilimler Sınavı (LYS-4) ve Yabancı Dil Sınavı (LYS-5) olmak üzere 5 farklı alandadır.

Adaylar, YGS ve LYS sınavlarına girdikten sonra bu sınavlarda göstermiş oldukları performansa göre ham puanları hesaplanır. YGS ve LYS sınavlarından alınan puanların puan türlerine katkıları aşağıdaki tablolarda sunulmuştur.

**Tablo 1****LYS ve YGS puan türlerinin ham puan oluşumuna katkı yüzdeleri (ÖSYM, 2015)**

Tablo 1E										
LYS Puan Türleri										
Testlerin Ağırlıkları (%)										
YGS					LYS (LYS-3+LYS-4)					
Puan Türü	Türkçe	Sosyal Bilimler	Temel Matematik	Fen Bilimleri	Türk Dili ve Edebiyatı	Coğrafya-1	Tarih	Coğrafya-2	Felsefe Grubu Din Kul. ve Ahlak Bil.	
TS-1	13	12	10	5	15	8	15	7		15
TS-2	18	11	6	5	25	5	15	5		10

Tablo 1E					
LYS Puan Türleri					
Testlerin Ağırlıkları (%)					
YGS					LYS (LYS-5)
Puan Türü	Türkçe	Sosyal Bilimler	Temel Matematik	Fen Bilimleri	Yabancı Dil
DİL-1	15	9	6	5	65
DİL-2	25	13	7	5	50
DİL-3	48	20	7	5	20

Tablo 1E										
LYS Puan Türleri										
Testlerin Ağırlıkları (%)										
YGS					LYS (LYS-1+LYS-2)					
Puan Türü	Türkçe	Sosyal Bilimler	Temel Matematik	Fen Bilimleri	Matematik	Geometri	Fizik	Kimya	Biyoloji	
MF-1	11	5	16	8	26	13	10	6	5	
MF-2	11	5	11	13	16	7	13	12	12	
MF-3	11	7	11	11	13	5	13	14	15	
MF-4	11	6	14	9	22	11	13	9	5	

Tablo 1E									
LYS Puan Türleri									
Testlerin Ağırlıkları (%)									
YGS					LYS (LYS-1+LYS-3)				
Puan Türü	Türkçe	Sosyal Bilimler	Temel Matematik	Fen Bilimleri	Matematik	Geometri	Türk Dili ve Edebiyatı	Coğrafya-1	
TM-1	14	5	16	5	25	8	20	7	
TM-2	14	7	14	5	22	6	24	8	
TM-3	15	10	10	5	18	5	28	9	

Adayların ham puanları ilgili alanlardaki ağırlıklara göre hesaplandıktan sonra; nihai puanın oluşturulması için ortaöğretimde göstermiş olduğu performansa bağlı olarak, Ortaöğretim Başarı Puanı (OBP) hesaplanarak ham puanın üzerine eklenir. Türkiye’de bu uygulama 1982 yılından itibaren yapılmaktadır. Ortaöğretim başarı puanı lise diploma notundan, en az 250 ve en fazla 500 olmak üzere hesaplanır. Daha sonra bu puan 0,06 katsayısıyla çarpılarak adayın hampuanının üzerine eklenir ve nihai puan elde edilir. Adaylar bu puanı kullanarak, her yıl ÖSYM tarafından yayınlanan tercih kılavuzundan faydalanarak yerleşmek istedikleri bölümleri tercih ederler. Yapılan tercihler arasında adaylar ilgili yükseköğretim programının kabul ettiği puan türünden sahip oldukları puanlara göre yerleştirilirler.

### 1.5 Öğrencilerin Akademik Başarılarının Ölçme ve Değerlendirmesi

Üniversiteler, bilimsel özerkliğe ve kamu tüzel kişiliğine sahip, yüksek düzeyde eğitim, öğretim, bilimsel araştırma ve yayın yapan fakülte, enstitü, yüksekokul vb. kuruluş ve



birimlerden oluşan öğretim kurumlarıdır (Türk Dil Kurumu, 2015). Farklı bir tanıma göre ise; üniversitelerin temel amacı bilginin peşinden gitmek ve bilgiyi aktarmaktır (Higgins, 1989: s. 358).

Tanımlardan da anlaşılacağı üzere üniversitelerin temel iki fonksiyonundan birisi: öğretimdir. Bu süreç sonunda üniversiteler, öğrencilerin dönemlik ve genel olarak sergiledikleri performansa göre, kişinin alanında ihtisas sahibi olduğunu gösteren belgeler olan diplomaları vermektedirler. Dönemlik ve genel performansın tüm dünyada kabul görmüş yegane göstergeleri ise, öğrencinin her bir yarıyılıda elde ettiği dönemlik not ortalaması ve dönemler ilerledikçe kümülatif bir şekilde elde edilen genel not ortalamasıdır. Bu not ortalamaları öğrencinin kayıtlı olduğu programdaki başarısını ortaya koymaktadır. Döneme ait not ortalamaları ve genel not ortalaması dünyanın farklı yerlerinde yer alan üniversitelerde tek akademik başarı göstergesi olarak kabul edilse de; ölçme ve değerlendirme yöntemleri üniversiteden üniversiteye farklılık gösterebilmektedir. Öğrencinin akademik performansının araştırılmasına yönelik çalışmalarda bu noktayı göz önünde bulundurmak oldukça önemlidir.

## **1. 6 Literatür**

Akademik performans üniversite kapsamında düşünülecek olursa; öğrencinin bulunduğu sınıf/dönem ve derslere göre belirlenmiş sonuçlara ulaşmada göstermiş olduğu ilerlemedir (Ateşoğlu, 2013: s. 6). Öğretim sürecinde öğrencilerin akademik başarısı ise dönemlik ve genel not ortalamalarıyla ölçülmektedir. Araştırmacılar bugüne dek, öğretim sürecinde akademik başarının göstergesi olan not ortalamalarını bağımlı değişken olarak kabul etmişler (Jamelske, 2008; French vd., 2005; Komarraju vd., 2013; Slim vd., 2014) ve bu performans göstergesini ortaya çıkaran süreç üzerine yoğunlaşan teorik çalışmalar yapmışlar ve bu teorik çalışmaları kullanarak bağımlı değişkeni betimlemeyi ve tahminlemeyen uygulama çalışmaları ortaya koymuşlardır. Bu çalışmalar arasındaki mantıksal ilişkinin daha iyi gözlemlenebilmesi için literatürü oluşturan bu iki sacayağını ayrı başlıklarda ele almak yerinde olacaktır.

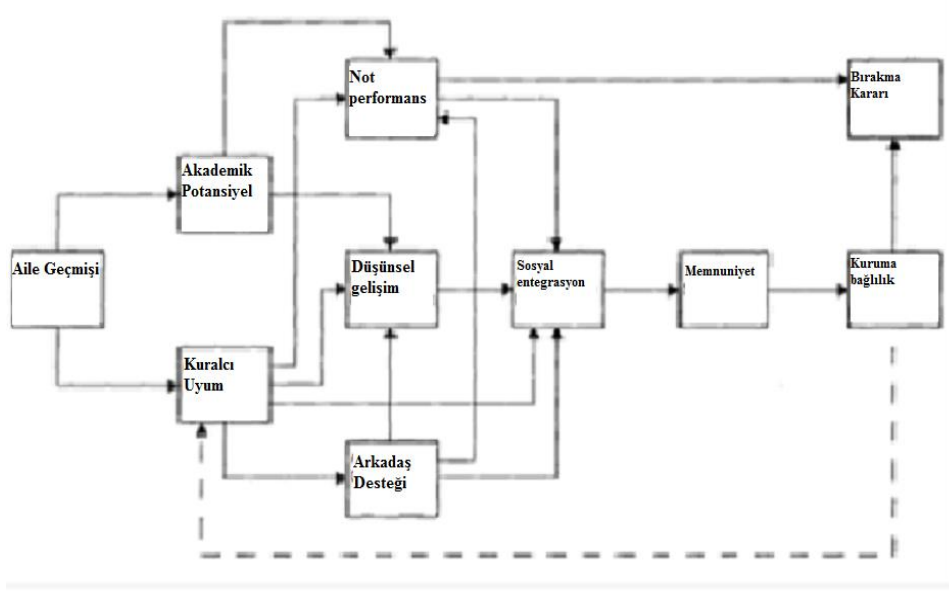
### **1.6.1 Akademik Başarı Üzerine Teorik Modeller**

Literatürde bugüne dek yapılan pek çok çalışmada, öğrencinin akademik başarısı ve bir sonraki sene okula devam etme oranı birlikte ele alınmıştır (Jamelske, 2008; DeBerard vd. , 2004; Besterfield-Sacre vd., 1997). Bunun nedeni; öğrencinin içerisinde bulunduğu

sürece bağı olarak deęişen deęerler olarak düşünölmelidir. Bir başka deyişle; akademik başarı ve bir sonraki sene okula devam etme oranı aynı sürecin çıktılarıdır. Dolayısıyla, teorik araştırmacılar ister öęrencinin akademik başarısı üzerine, ister öęrencinin bir sonraki sene okula devam etme kararı üzerine çalışmalarını yapsınlar; araştırılan süreç aynıdır. Bu sebeple öęrencinin akademik performansı ve bir sonraki sene okula devam etme kararı benzer unsurlara dayanmaktadır. Bu iki kavram için de ortak bir literatürden bahsetmek yanlış olmayacaktır.

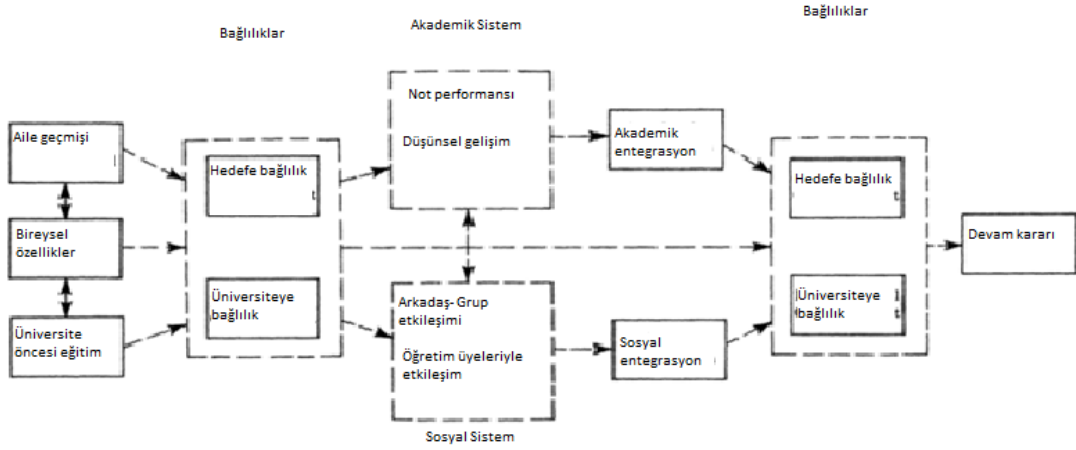
Bu karmaşık sürecin daha iyi anlaşılabilmesi için araştırmacılar, kavramsal olarak yetersiz kalan daha çok betimleyici çalışmalar yapmışlardır. Ailenin sosyoekonomik durumu (Astin, 1964; Eckland, 1964), öęrencinin üniversite eğitimi süresince aileden aldığı tavsiye, övgü ve ilgilenme (Trent&Ruyle; 1965; akt: Tinto, 1975: s. 100), lise not ortalaması (Blanchfield, 1971; Lavin; 1965; akt: Tinto, 1975: s. 100; Lawhorn, 1971), kişilik ve yaklaşım farklılıkları (Pervin, 1968) gibi faktörlerin devam etme kararı ve akademik ve sosyal entegrasyon derecesi üzerinde önemli derecede etkiye sahip olduğu ortaya konmuştur.

William G. Spady (1970) okula devam etme kararı süreci üzerine yapılan çalışmaları incelemiş ve o zamana kadar ilgili literatürün (Knoell, 1960, 1966; Marsh, 1966) aile yapısı, yetenekler ve akademik performans ile üniversiteye devam etme kararı arasındaki ilişki ile ilgili birkaç sağlam genelleme dışında, yetersiz olduğunu belirtmiştir. Spady, bu karmaşık süreci, ünlü sosyolog Emile Durkheim'in "İntihar Teorisi" ile benzeştirerek kavramsal olarak modellemiştir. Durkheim'in teorisinden esinlenilerek ortaya konan modele göre; belirli bir sosyal sistemi terketme kararı, aile ve önceki eğitim geçmişi, akademik potansiyel, kurallara uyum, arkadaş desteęi, düşünsel gelişim, akademik başarı, sosyal entegrasyon, memnuniyet ve üniversiteye olan baęlılık faktörlerini içeren karmaşık bir sosyal sürecin sonucudur. (Spady, 1971: s. 38)



**Şekil 4: Okulu bırakma sürecinin açıklayıcı bir sosyolojik modeli (Spady W. G., 1970: s.79)**

Öğrencilerin okulu bırakma karar süreci ile ilgili çalışmalarıyla öncü araştırmacılardan biri olan Tinto bu süreci teorik olarak modellemeyi amaçladığı makalesinde, düşük akademik başarının okulu bırakma kararı üzerinde etkili olabileceğini savunmuştur (Tinto, 1975:s. 92). Aynı çalışmada, bireylerin üniversitede eğitime devam davranışlarının zaman içerisindeki değişimini teorik bir modelle açıklamak isteyen araştırmacılara çalışmalarında kişilerin sadece özgeçmiş değişkenlerini (sosyal statü, lise performansı, oturulan yer ve cinsiyet, yetenek, ırk ve uyruk gibi kişisel özellikler) değil, beklentiye dönük ve motivasyonel değişkenlerini de (kariyer ve eğitime dair beklentiler ve akademik başarı için motivasyon) göz önünde bulundurmaları tavsiye edilmiştir (Tinto, 1975: s. 93) . Bu önerilerden sonra Tinto önerdiği modeli şu şekilde sunmuştur.



**Şekil 5: Üniversiteye devam kararı modeli (Tinto, 1975: s. 95)**

Ortaya konan bu model, sonraki yıllarda pek çok araştırmacıya ilham kaynağı olmuş ve akademik başarı ve öğrencinin devam kararının tahmini üzerine yapılan uygulama çalışmalarına rehberlik yapmıştır. 1975 yılında yapılan bu çalışmaya bugüne dek yapılan 5294 atıf bu durumun göstergesi olarak düşünülebilir.

Alanında bir diğer öncü araştırmacı Astin, Tinto'nun temel modeline ek olarak, öğrencinin davranışlarının ve çevreyi algılama şeklinin de kişinin üniversiteye devam kararı üzerinde etkili olduğunu öne sürmüştür ve aidiyet (involvement) kavramını ortaya atmıştır. Aidiyet, öğrencinin akademik deneyim için harcadığı fiziksel ve psikolojik enerji miktarı olarak tanımlanabilir. Buna göre ; aidiyet seviyesi yüksek olan öğrencilerin, bir başka deyişle daha fazla çalışan, kampüste çok zaman geçiren, öğrenci organizasyonlarına aktif olarak katılan ve sık sık öğretim üyeleri ve diğer öğrencilerle etkileşim halinde olanların akademik başarısının ve okula devam etme olasılığının yüksek olacağı öngörülebilir (Astin, 1984: s. 297). Aidiyeti belirleyen ise, kişinin hissettikleri veya düşündüklerinden ziyade, ne yaptığı ve nasıl davrandığıdır. Bu bağlamda aidiyet teorisinin 5 temel varsayımı vardır.

- 1- Aidiyet, fiziksel ve psikolojik enerjiyi çeşitli alanlara harcamayı ifade eder.
- 2- Alanı ne olursa olsun, aidiyet sürekli bir süreçte devam eder, yani her hangi bir alanda değişik öğrenciler değişik aidiyet derecelerine sahip olabilirler ve aynı

öğrenci değişik alanlarda ve değişik zamanlarda değişik aidiyet dereceleri sergileyebilir.

- 3- Aidiyetin hem nicel hem de nitel olan özellikleri vardır.
- 4- Öğrencinin herhangi bir eğitim programına ilişkin öğrenmesi ve kişisel gelişimi, öğrencinin o programa aidiyet niceliği ve niteliğiyle doğru orantılıdır.
- 5- Herhangi bir eğitsel politikanın etkinliği, politikanın kapasitesi veya öğrenci aidiyetini arttıracak uygulamalara direkt olarak bağlıdır (Astin, 1984: s. 519).

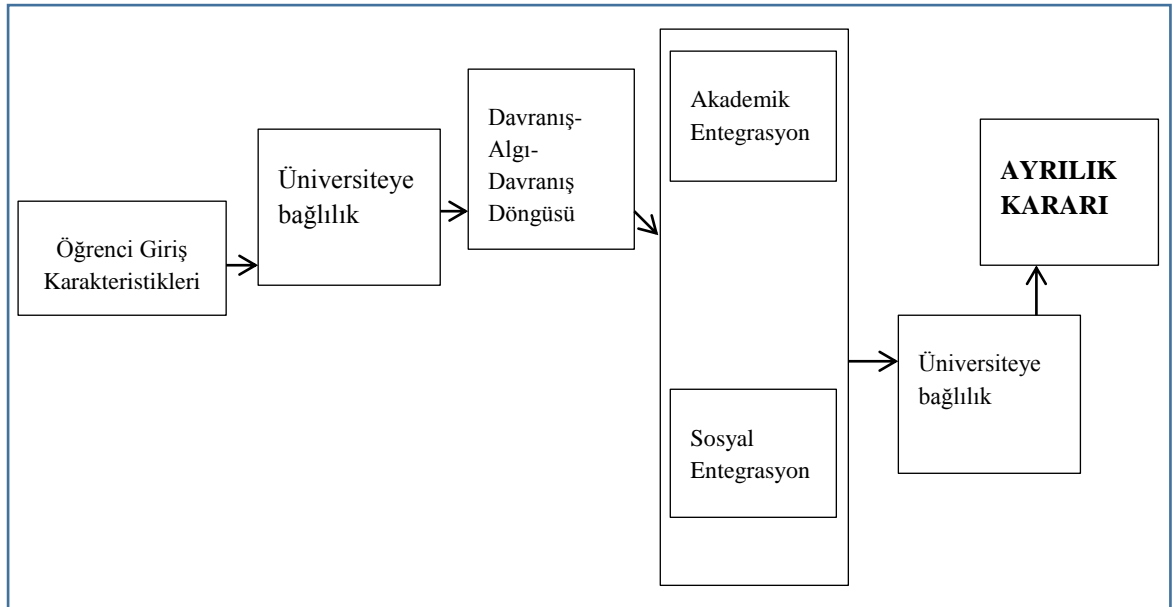
Öğrenciler, kısıtlı zamanları içerisinde hayatlarını oluşturan konulara (aile, arkadaşlar, eğlence vs.) belirli bir enerji harcamak durumundadırlar. Kurum yöneticilerinin bu kısıtlı zaman içerisinde belirli bir paya sahip olarak aidiyet derecesini yükseltmek ve böylece daha başarılı ve devamlılık sağlayacak öğrenciler ortaya çıkarmak için politikalarını düzenlerken göz önüne alması gereken bazı hususlar vardır. Öncelikle; yöneticiler ve öğretim üyeleri neredeyse tüm kurumsal politika ve uygulamayı (ders programı, derse katılım ile ilgili düzenlemeler, akademik staj, öğretim elemanı haftalık görüşme saatleri ile ilgili düzenlemeler, öğrenci oryantasyonu ve yönlendirme) tanımlamalıdır. Ayrıca, yönetimin akademik olmayan konular (kafeterya ve yurtların yerlerinin belirlenmesi, yurt kuralları, dinlenme alanlarının tasarımı, öğrenci personel olma fırsatları, ders dışı aktivite sayısı ve tipleri ve katılıma dair yönetmelikler: sıklık, tipi ve kültürel faaliyetlerin maliyetleri, oda arkadaşı atamaları, finansal destek politikaları, yemekhanelerin cezbediciliği ve park kuralları) hakkında aldığı kararlar öğrencinin okul ile arasındaki etkileşimin kalitesini etkileyecek niteliktedir (Astin, 1984: s.523). Sonuç olarak; öğrencilerin aidiyet boyutları şu boyutlarda özetlenmiştir: Yaşanılan yer, Ödüllendirme programları, Akademik dahil olma, Öğrenci- öğretim üyesi etkileşimi, bedensel aktivitelere katılım, Öğrenci konseylerine katılma, Bilişsel gelişim üzerine araştırma.

Bu çalışma, öğrencinin okula entegre olma sürecinde bireyin psikolojik ve fiziksel boyutunun da önemli olduğunu göstermesi bakımından büyük önem arz etmektedir. Dolayısıyla bu model, Tinto'nun modelini hem tamamlayıcı nitelikte olduğu hem de Tinto'nun modelinde geçen akademik ve sosyal entegrasyon kavramlarını daha sağlam bir zemine oturttuğu savunulabilir. Öğrencinin ilk yıl üniversiteye geçiş sürecini araştıran bu iki temel çalışma; dinamikleri açısından birbirlerine benzer niteliktedir.

Tinto (1988); ortaya koyduğu etkileşim süreci modelinin revize edilmiş versiyonu mahiyetinde olan bir diğer çalışmada, öğrencinin üniversiteye entegre olmasını temel olarak 3 aşamalı bir süreç olarak ele almıştır. Buna göre;

- 1- **Ayrılık aşaması:** Bu aşamada öğrenci kendisini geçmişte içerisinde bulunduğu topluluğa (lise, aile, üniversite öncesi) ait normlardan belirli ölçülerde soyutlar.
- 2- **Geçiş aşaması:** Kişi geçmişte içerisinde bulunduğu topluluklara ait normlardan sıyrılmış, ancak içerisinde bulunduğu topluluk ve çevrenin değerlerini tam anlamıyla benimseyememiştir.
- 3- **Özümseme aşaması:** Üniversite topluluğunun ve çevresinin benimsemiş olduğu değerler içselleştirildiğinde kişi tam anlamıyla kuruma entegre olmuştur.

Milem ve Berger, Tinto'nun revize edilmiş modeli ve Astin'in öğrenci dahil olması ile ilgili teorisi arasındaki ilişkiyi vurgulamış ve geçiş aşamasında öğrencilerin davranış ve algılarının üniversite yönetiminin akademik ve sosyal deneyimlerine ne derece destek olduğuna bağlı olarak şekilleneceğini ve bu destek derecesine bağlı olarak başarılı bir özümseme gerçekleşeceğini savunmaktadır. (Milem&Berger, 1997: s. 390) Buna göre çalışmada sunulan kavramsal model şu şekilde oluşmuştur:



**Şekil 6: Öğrenci ayrılık kararının revize edilmiş bir modeli (Milem&Berger, 1997: s. 389)**

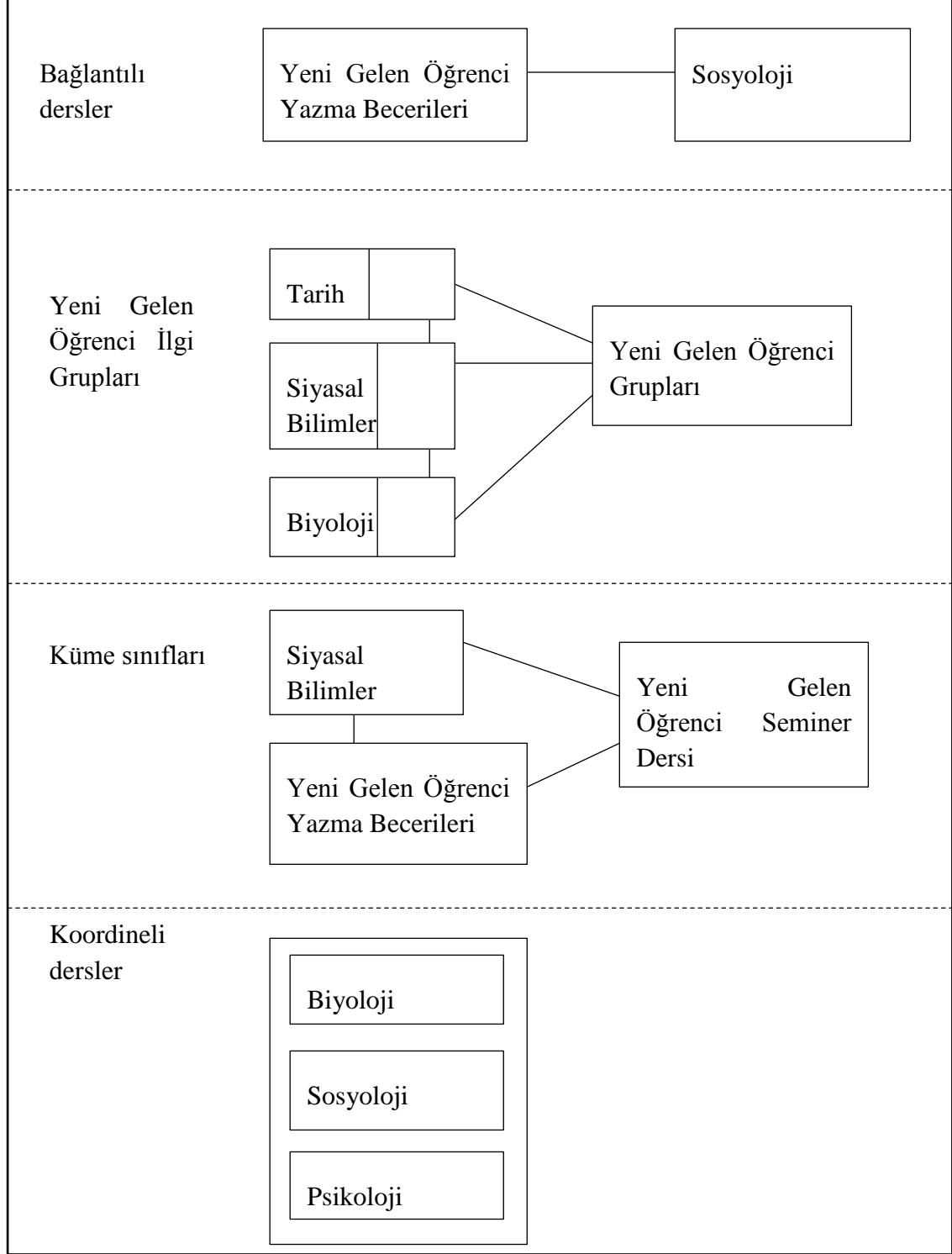
Tinto'nun etkileşim esaslı kavramsal modellerinin analizine ve revize edilmesine yönelik pek çok çalışma yapılmıştır. Üniversite tipi (Chapman&Pascarella, 1983) organizasyonel özellikler (Berger&Braxton, 1998), öğrenci davranışları ve algısı (Milem&Berger, 1997) (Astin, 1984), üniversiteye gitme motivasyonları (Stage, 1989) akt: (Braxton vd., 2000), finansal yardım (Cabrera vd. 1992), üniversiteden beklentilerin karşılık bulması (Braxton vd., 1995), öğrenci yurtlarındaki topluluk algısı (Berger&Milem,1999), yaşam görev üstünlüğü (Brower, 1992) ve öz yeterlilik (Peterson, 1993) faktörlerinin öğrencinin okula devam etme kararında etkili olduğu savunulmuştur.

Tüm bu çalışmalara rağmen A.B.D.'de okula devam oranları ve akademik başarı anlamında beklenen sıçramanın gerçekleşmemesi üzerine; öğrencilerin devamlılığının sağlanması konusunun üniversite yönetimlerinde yeterince ciddiye alınmadığı yönünde bir algı oluşmuştur. Bu noktada, üniversitelerin müdahale alanlarının sınırları gündeme gelmiştir ve üniversite yönetiminin müdahale edebileceği alanlar 4 ana başlık altında toplanmıştır. (Tinto, 1999: s. 6)

- 1- Bilgi/ Danışma:** Öğrencilere kurumsal gereklilikler konusunda anlaşılabilir ve tutarlı bilgilerin sağlandığı bir ortamda, devam etme ve mezuniyet oranlarının yüksek olacaktır.
- 2- Destek:** Akademik, sosyal ve kişisel destek sağlayan üniversitelerde devamlılık oranları daha yüksek olacaktır.
- 3- Aidiyet:** Kendilerini kayıtlı oldukları üniversitelerin değerli birer üyesi olarak gören öğrencilerin eğitimlerine aynı üniversitede devam etme olasılıkları daha yüksek olacaktır.
- 4- Öğrenme:** Şüphesiz en önemli unsurdur. Basit olarak düşünülecek olursa, öğrenen öğrenciler , eğitimlerine devam ederler.

Çalışmada, öğrenme aktivitelerine aktif olarak katılımın (aidiyet), okulda kalma oranlarının artırılmasının anahtarı olduğu düşünülmüştür. Bu süreçte öğrencilerin birbirleriyle ve öğretim üyeleriyle etkileşim halinde bulunabildiği tek ortamın sınıf olduğu ve sınıfların aidiyeti arttıracak doğrultuda tasarlanması gerektiği vurgulanmıştır. Tinto (1999: s. 6)'ya göre “Öğrenen Topluluklar” kavramı bu ihtiyaç doğrultusunda ortaya çıkmıştır. Bu toplulukların içerik değişse bile 3 amacı vardır: paylaşımlı

öğrenme, paylaşılan öğrenme, paylaşılan sorumluluk. Öğrenen topluluklar; öğrencilerin öğrenme aktivitelerinde diğer öğrencilerle ve öğretim üyeleriyle etkileşimi artırma potansiyeli nedeniyle önemlidir.

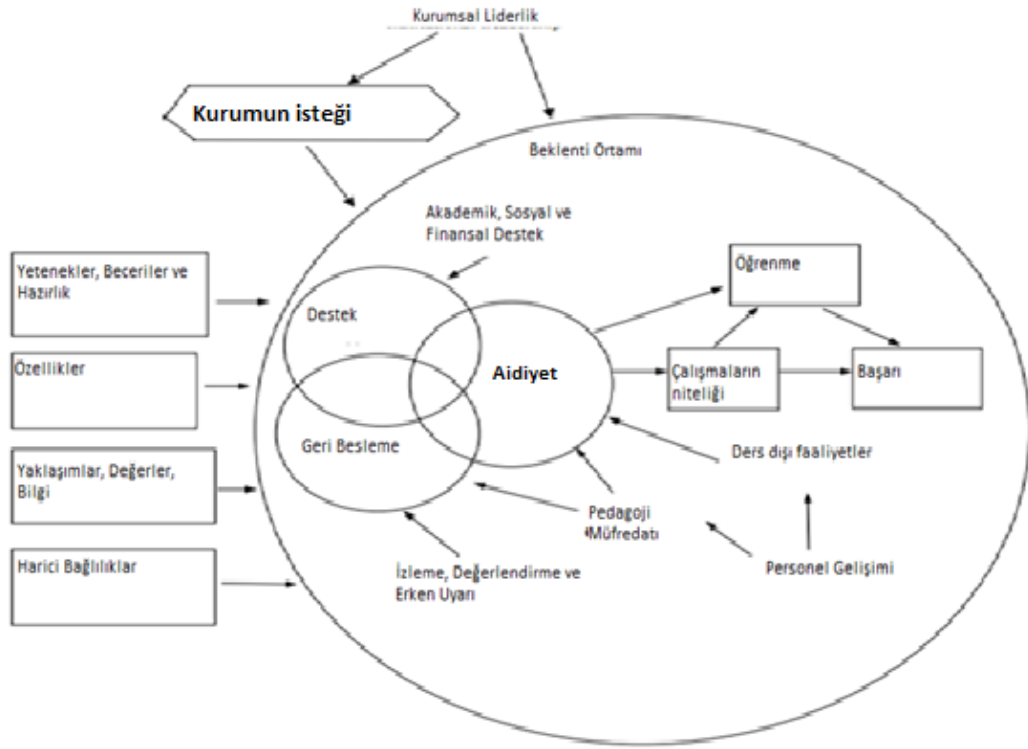


Şekil 7: Örnek bir Öğrenen Topluluk modeli (Tinto V. , 1999: s.6)



Şekil 7’ de görülen Öğrenen Topluluk Modelini kısaca izah etmek gerekirse; öğrenen topluluklar en temel haliyle, öğrencilerin diğer programlardan ilk yıl öğrencileriyle birlikte ders almalarını sağlayan bir çeşit yan kayıt veya blok ders programı ile başlar. Bazı durumlarda, öğrenen topluluklar, dersleri birbirine bağlamak şeklinde öğrencileri birleştirirler. Örneğin; yazma becerileri ile seçilen literatürde herhangi bir ders tipik olarak birlikte bulunabilir (Bağlantılı dersler). Bazı durumlarda bu derslerin müfredatları tamamen aynı olabilir. Bu durumda bu dersleri alan öğrencilerin dönem boyu kullandıkları ders materyalleri aynı olacaktır. Bazı büyük üniversitelerde, öğrenen topluluk içinde yer alan öğrenciler derslere 200-300 diğer öğrenci ile birlikte katılırlar, ama sıklıkla dersle ilgili ufak bir tartışma için bir arada kalırlar. Bu gruba sıklıkla “Yeni Gelen Öğrenci İlgi Grupları” adı verilmektedir. Bu gruplara üst sınıflardan ya da mezun öğrencilerden biri öncülük etmektedir. “Küme Sınıfları” durumunda ise, öğrenen topluluk içerisinde yer alan tüm öğrenciler bağlantılı ya da bağlantısız tüm dersleri birlikte almaktadırlar. “Koordineli Dersler” ise haftada 4 ila 6 saat arasında bir araya gelen büyük bir sınıfı ifade etmektedir.

Beklenen okula devam oranlarının yakalanamamasının ortaya çıkardığı bir diğer tartışma konusu, araştırmacılar tarafından yapılan teorik çalışmalar ile uygulamalar arasındaki uyumsuzluktur. Yapılan bir araştırmada, A.B.D.’de 1984-2004 yılları arasında mezuniyet oranlarında önemli bir değişimin yaşanmadığı belirtilmiştir (Carey, 2004). Bu bahsedilen uyumsuzluğun bir göstergesi olarak düşünülebilir. Tinto ve Pusser (2006) çalışmalarında, bu oranların istenen seviyeye ulaşabilmesi için üniversitelerin müdahale sınırlarını, üniversitenin içerisinde bulunduğu eğitim çevresini ve öğrencilere ait özellikleri (yaklaşımlar, değerler, yetenekler, yaş, hedeflere olan bağlılık vs.) göz önünde bulunduran kavramsal bir model ortaya koymuşlardır. Bu model, üniversite yöneticilerinin içerisinde bulunduğu ortamı ve süreci daha iyi anlamaları ve öğrenci başarı ve devamlılığını arttırmaya yönelik müdahalelerini buna göre şekillendirmeleri bakımından önemlidir.



**Şekil 8: Kurumsal Eylemin model yapısı ve bileşenleri (Tinto&Pusser, 2006: s.9)**

Buna göre, bu modelde belirtilen ana argüman; öğrenemeyen öğrencinin başarılı ya da devamlı olamayacağıdır. Diğer argüman ise; başarılı öğrenmenin gerçekleşebileceği tek yer sınıf ortamıdır. Bu noktada öğretim üyelerine özel bir görev düşmektedir. Modeli oluşturan bileşenleri kısaca açıklamak gerekirse;

- **Kurumun isteği:** Temel ve belki de en önemli etmen olarak düşünülebilir. Öğrencilerin akademik performansını yükseltme konusunda bağlılık ve ciddiyetle çalışan üniversiteler bu amaca bir şekilde ulaşabilmektedir.
- **Beklenti ortamı:** Hiçbir öğrenci düşük beklentiye sahip değildir. Ancak üniversitelerin ilk yıl öğrencilerinden beklentileri, maalesef düşük olmaktadır. Ya da üniversite yönetimleri değişik öğrenci grupları için değişik beklentilere sahiplerdir. Öğrencilerin başarı durumları öğrencinin beklentisi ve yönetimin öğrenciden beklentisine bağlı olarak, öğrencinin beklentilerinin karşılanabilme oranından etkilenmektedir. Beklentiler, resmi ve gayri resmi danışma yoluyla

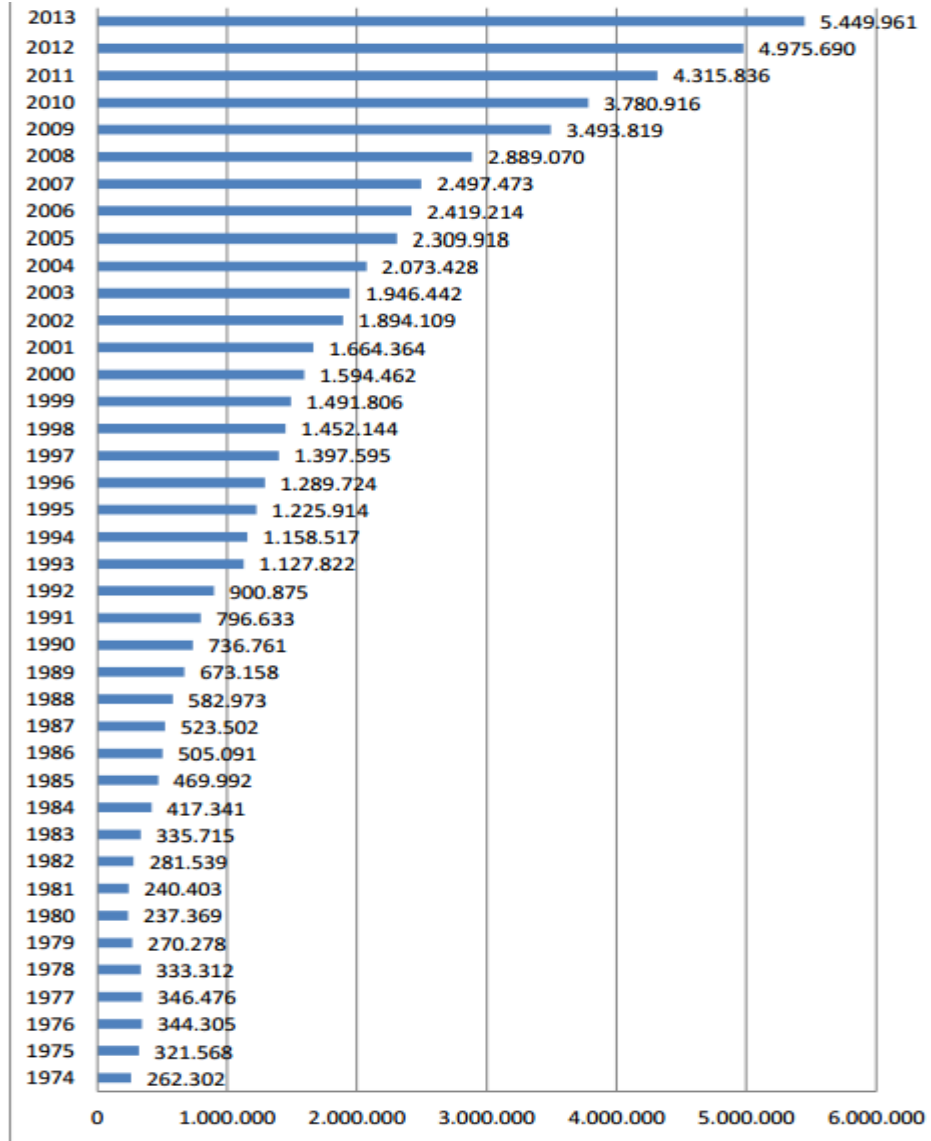
ifade edilebilmekte; bazıları beklentilerini karşılayabilirken diğerleri karşılayamamaktadır.

- **Destek:** Akademik destek (gelişmeye yönelik eğitim dersleri, özel dersler, çalışma grupları ve ek dersler), Sosyal destek (Danışma, mentorluk ve etnik öğrenci merkezleri)
- **Geri besleme:** Öğrenciler, performansları hakkında öğretim üyelerinden veya arkadaşlarından ne kadar fazla geri dönüş alırlarsa, başarılı olma ihtimalleri o derece artmaktadır. Bu sadece erken uyarı sistemleri ile değil sınıf içerisinde farklı değerlendirme teknikleriyle de mümkün olacaktır.
- **Aidiyet:** Öğrencinin aidiyet duygusunun muallakta olduğu ve kritik bir seviyede yer aldığı ilk yılda, akademik başarı için önemli bir anahtardır. Bu noktada akademik ve sosyal entegrasyon tabirini kullanmak yanlış olmayacaktır.

Üniversite yönetimlerinin pek çok faktörün sarmal bir yapıda yer aldığı bu karmaşık süreci yönetebilme (liderlik etme) kabiliyetine paralel olarak başarı ve devamlılık seviyesinin artacağı savunulabilir.

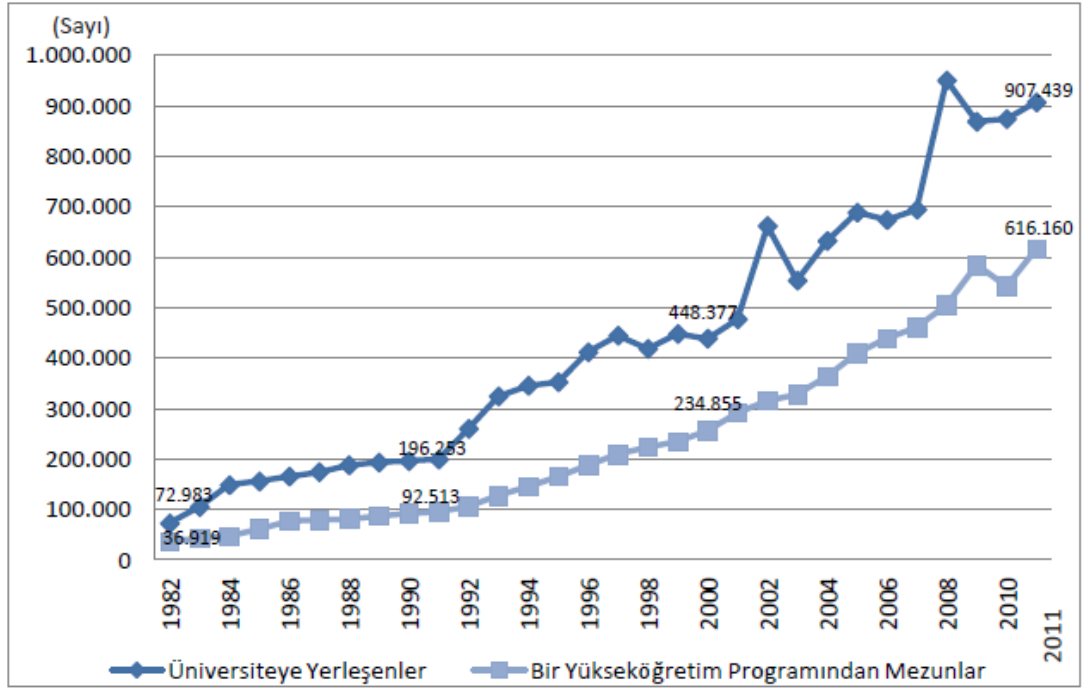
Bu teorik modellerin ve üniversite yönetimlerinin müdahale alanlarının içinde bulunulan yükseköğretim ortamına bağlı olarak değişebileceği bir gerçektir. Ancak bahsedilen faktörlerin birçoğunun (aidiyet, sosyal destek, akademik destek vs.) evrensel nitelik taşıdığı düşünülebilir. Bu durumdan yola çıkarak, Türkiye’de bu tarz bir modelin uygulanma ya da eğitim sistemine uygun olarak geliştirilme ihtiyacını değerlendirebilmek için bazı göstergelere bakmak faydalı olacaktır.

Türkiye’de 80’li yıllardan itibaren yükseköğretim programlarına kayıtlı öğrenci sayılarında hızlı bir artış gözlemlenmektedir. Örnek vermek gerekirse 2000 yılında 1.594.462 olan yükseköğretim öğrenci sayısı, 2010 yılı itibariyle %121 seviyesinde bir artış göstererek 3.529.334 seviyesine ulaşmıştır ve bu artış oranıyla dünyada 6. Sırada yer almaktadır. (UNESCO, 2013)



**Şekil 9: Türkiye’de yükseköğretim öğrenci sayısındaki değişim (1974-2013)**  
(Çetinsaya, 2014)

Bu durumda, yükseköğretim kurumlarına ait bir performans göstergesi olarak, üniversitelerin öğrencilerinin ne kadarını mezun edebildiği önem kazanmıştır. Örneğin, en son 2010 yılında üniversiteye yerleşen öğrenci sayısı 907.439 iken mezun olan öğrenci sayısı ise 616.160 olarak gerçekleşmiştir.



**Şekil 10: Bir yükseköğretim programına yerleşen ve bir yükseköğretim programından mezun olan sayıları (1982-2011) (Çetinsaya, 2014: s.58)**

Bu rakamlar aynı zaman diliminde yorumlanacak olursa; 1982 yılında yükseköğretim programlarına kayıtlı öğrenci sayısı 281,539 iken 2010 yılı itibariyle bu sayı yaklaşık %1200 oranında artarak 3,780,916 olmuştur. Aynı süreçte mezun olan öğrenci sayısının yeni kayıt olan öğrenci sayısına oranı 1982 yılında %50.5 iken; 2010 yılı itibariyle bu oran %67 seviyesine çıkmıştır. Mezuniyet oranlarındaki bu iyileşme her ne kadar kabul edilebilir seviyede olsa da, öğrenci sayılarındaki değişime paralel olarak değişmediği ve yeterli olamayacağı açıktır. Ayrıca bu iyileşmenin bir kısmının belirli zamanlarda çıkan öğrenci afları sayesinde gerçekleştiğini de göz önünde bulundurmak gerekmektedir. Mezun öğrenciler ile yeni kayıt yaptıran öğrenci sayısı arasında açılan makas, yükseköğretim dinamikleri hakkında fikir vermektedir. Bu farkın temel sebepleri arasında, sınav kazanıldığı halde kayıt yaptırılmaması, üniversitede yerleşilen bölümden sonra tekrar sınava girilerek başka bölümlere yerleşilmesi ve yerleşilen lisans programlarından mezun olmadan ayrılması gösterilebilir. Yükseköğretim öğrencisiyken tekrar sınava girilerek başka bölümlere yerleşilmesi ve lisans programlarını tamamlayamadan bırakma akademik devamlılığı etkileyen unsurlardır. Çalışmanın perspektifinden değerlendirilecek olursa, yıllar içerisinde mezun olan

öğrenciler ve yeni kayıt yaptıranlar arasındaki farkın açılması akademik devamlılık oranlarının düştüğü şeklinde yorumlanabilir.

Şekil 9'dan da anlaşılacağı üzere, Türkiye'de zaman içerisinde düşüş gösteren akademik devamlılık oranlarının iyileştirilmesine yönelik araştırmalar önceki bölümlerde belirtildiği üzere dünyada uzun yıllardır yapılagelmektedir.

Tinto (2012) bu istatistikleri yükseköğretimde öğrenci performanslarının göstergesi olarak düşünmüş ve bu performansın artırılmasına yönelik bazı önerilerde bulunmuştur. Çalışmada öğrencilerin başarılarının etkin sınıf ortamıyla doğru orantılı olduğu vurgulanmıştır ve etkin bir sınıfın özellikleri şu şekilde belirtilmiştir:

- **Beklentiler:** Sınıf performansı temel olarak öğretim üyelerinin öğrencilerden beklentileri ve öğrencilerin de kendilerinden beklentileri tarafından yönlendirilir. Bu noktada başarı, bu beklentilerin açıklığı ve tutarlılığının yanısıra seviyesinden de etkilenebilmektedir.
- **Destek:** Beklentiler kadar, bu beklentileri gerçekleştirebilmeleri için öğrenciye destek olmak da başarıya giden yolda çok önemlidir.
- **Değerlendirme ve geri bildirim:** İlk yılında akademik ve sosyal ortama ayak uydurmaya çalışan öğrencilerin performansını sık sık değerlendirmek ve geri dönüşlerde bulunmak, öğrencilerin başarı şansını arttıracaktır.
- **Dahil olma (Aidiyet):** Özellikle sınıf aktivitelerinde öğretim üyeleri ve arkadaşlarıyla yakın ilişkiler kurabilen öğrencilerin başarılı olma ihtimalleri çok daha fazladır. (Tinto, 2012: s. 4-5)

Bu özellikleri sağlayarak öğrencilerin akademik başarılarını arttırmak için ise bazı önerilerde bulunulmuştur: Eğitim takımları (temel eğitmen ve öğretim üyeleri) akademik destek, otomatikleştirilmiş sınıf değerlendirme, geri dönüş ve erken uyarı, sınıf içerisinde derse katılım oranını artırma ve öğretim üyelerinin pedagojik eğitim alması bu yönde katkı sağlayacak düzenlemeler olarak belirtilebilir (Tinto, 2012: s. 5-7).

### 1.6.2 Akademik Başarıyı Etkileyen Faktörler Üzerine Uygulamalar

Çoğunlukla kavramsal modellerden (Astin, 1984; Spady, 1971; Tinto, 1975) yola çıkılarak öğrencilerin akademik performansına ve okula devam etme kararlarına yönelik yapılan araştırmalar temel olarak iki alanda yoğunlaşmıştır: Betimleyici çalışmalar, tahminleyici çalışmalar.

Bu tip çalışmalarda; akademik performans ve okula devam etme kararlarına etki eden faktörlerin belirlenmesi (betimleyici çalışmalar) ve bu faktörlere bağlı olarak bağımlı değişkenin tahmin edilmesi (tahminleyici çalışmalar) için kavramsal modellerden faydalanılarak geleneksel (yaş, uyruk, finansal durum vs.) ve geleneksel olmayan (kariyer beklentileri, akademik özyeterlik vs.) değişkenler kullanılmaktadır.

**Geleneksel Değişkenler:** Genellikle nümerik ve kategorik veri tiplerinden oluşmaktadırlar. Öğrencinin eğitim geçmişi (Lise Not Ortalaması (nümerik), LYS puanı (nümerik) vs.), kişisel özellikleri (yaş (nümerik), cinsiyet (kategorik), medeni hal (kategorik), kardeş sayısı (nümerik), sigara kullanım durumu (kategorik), kaldığı yer (kategorik), ailesinin ikamet ettiği şehir (kategorik)) gibi değişkenler bu tip verilere örnek olarak sunulabilir. Geleneksel değişkenler çoğunlukla bu tip değişkenlerden oluşsa da; içerisinde anne eğitim düzeyi, ailenin aylık ortalama geliri sosyo-demografik ve sosyo-ekonomik değişkenler ordinal tipinde değişkenler de barındırır. Bu tip değişkenler, çoğunlukla, üniversite yönetiminin müdahale alanının dışarısında yer alırlar.

**Geleneksel olmayan değişkenler:** Genel anlamıyla, öğrencinin bulunduğu akademik ve sosyal çevreye olan tutumu, kişilik özellikleri ve psikolojik durumu ile ilgili değişkenlerdir. Geliştirilen hazır ölçekler, bu değişkenler hakkında bir yargıya varma amacıyla kullanılan araçlardır. Genel özyeterlik, akademik özyeterlik, kaygı seviyesi, üniversite ortamına ayak uydurma, duygusal zekâ gibi değişkenler bu tip değişkenlere örnek olarak gösterilebilir.

(Barker vd., 2004: s.80) literatürde yer alan değişkenleri iki ana gruba ayırmıştır: Demografik değişkenler (cinsiyet (Fike&Fike, 2008; Jamelske, 2008; Delen, 2010; Astin A. , 1997), ailenin ikamet ettiği şehir (Mishra vd., 2014;Aylesworth&Bloom, 1976), finansal durum (Thammasiri vd. 2014; Mary Besterfield-Sacre, 1997), lise

mezuniyet notu (Caison, 2007; Komarraju vd., 2013), üniversite giriş sınavı puanı (Komarraju vd., 2013; French vd., 2005) vs.) ve tutumsal faktörler (motivasyon (French vd. , 2005; Mishra vd. ,2014; Mckenzie&Schweitzer, 2001), özyeterlik (DeTure, 2004; Chemers vd., 2001) vs.)

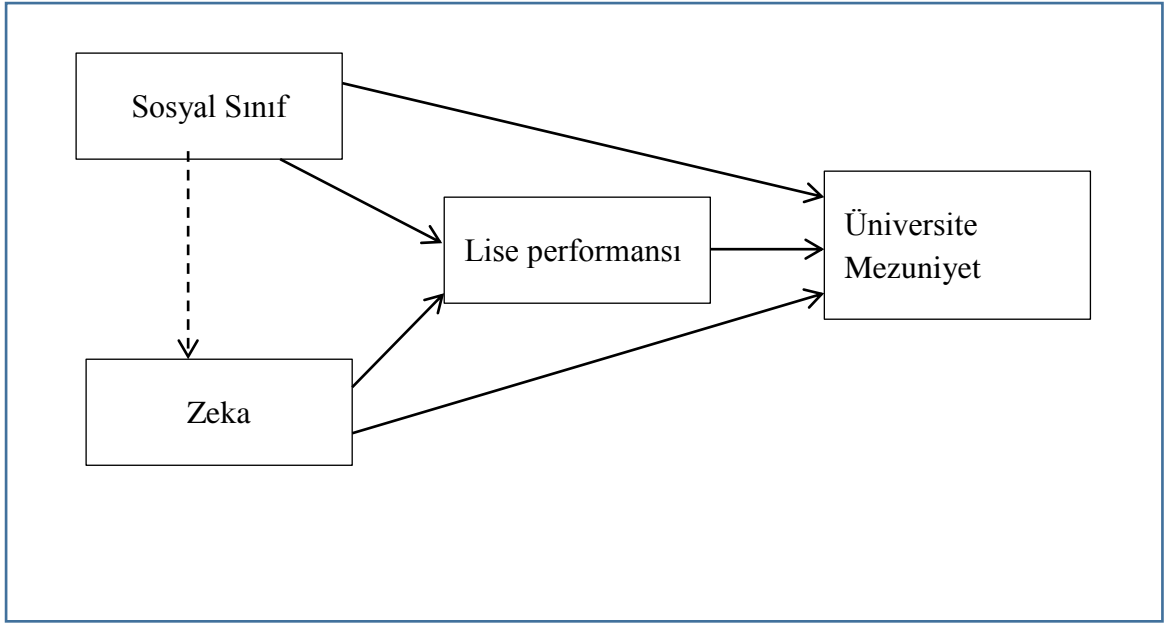
Parmentier (1994) öğrencilerin akademik performans sürecine dair ortaya koyduğu modelinde, ara ve nihai akademik performansın birbirleriyle etkileşim halinde olan 3 çeşit faktörden etkilendiğini öne sürmüştür: sabit ve yapısal faktörler (kimliği, sosyo-ailevi geçmişi, akademik geçmişi vs.), öğrencinin çalışmalarına olan bağlılığı (opsiyonel aktivitelere katılım, öğretim üyeleriyle görüşme vs.), öğrencinin algıları (akademik içerik, öğretim üyeleri ve derslere yönelik algılar) (akt. Superby vd. ,2006)

### **1.6.2.1 Betimleyici Çalışmalar**

Betimsel istatistiklerin sunulduğu bu çalışmalarda çoğunlukla değişkenler (cinsiyet, lise not ortalaması, ailenin sosyoekonomik durumu vs.) bakımından farklı değerlere sahip olan gruplar arasında belirli önem düzeylerinde anlamlı farklılıklar vurgulanarak akademik başarı ve devam üzerinde etkili olan faktörler belirlenmeye çalışılmaktadır.

Eckland (1964), farklı sosyal statülere sahip 1180 öğrenciyle yaptığı çalışmasında akademik başarı göstergelerinden birisi olarak kabul edilen mezuniyet oranı için farklı statülere sahip öğrencilerin baba mesleğinin, ebeveynlerin eğitim durumlarının ve lise performanslarının  $p=0.01$  anlam düzeyinde önemli derecede farklılaştığını ortaya koymuştur.





**Şekil 11: Üniversite mezuniyetine etki eden nedensel değişkenler arasındaki ilişki (Eckland, 1964: s.46)**

Gurin vd., (1968) ; (akt: Spady W. G., 1970: s. 69) Michigan Üniversitesi sanat ve bilim yükseköğretim programlarına kayıtlı olan 4400 yeni gelen öğrenciyle yaptıkları çalışmalarında, kozmopolitanizm (ailenin dini eğilimi ve eğitim düzeyi, kırsal – şehir durumu ve lisenin büyüklüğü) ve akademik hazırlık (lise not ortalaması ve SAT skoru) başlıkları altında bağımsız değişkenleri almış ve kozmopolitanizmin nispeten daha yüksek bir öneme sahip olduğunu gözlemlemiştir. Ancak her iki değişken ayrı ayrı ele alındığında belirgin bir ilişkiye rastlanamamıştır.

Yapılan çalışmalarda gerçekçi ve açık performans hedefleri olmayan öğrencilerin akademik başarılarının düşük olması beklenen bir durumdur. (Pervin vd., 1966; Knoell, 1966; Wood&Locke, 1987; Bjørnebekk vd., 2013) çalışmalarında bu görüşü destekleyen bulgulara ulaşmışlardır.

Terenzini&Pascarella (1980), yaptıkları boylamsal çalışmalarında Tinto'nun (1975) kavramsal modelini ele almış ve bu modelin gerçek verilerle geçerliliğini test etmişlerdir. Buna göre; Tinto'nun da belirttiği üzere okulu bırakanlar ve devam edenler arasında akademik ve sosyal entegrasyon değişkenleri yönüyle anlamlı farklılıklar bulunduğunu ifade etmişlerdir. Aynı yıl yayınlanan bir diğer makalelerinde araştırmacılar öğrencilerin akademik ve sosyal entegrasyon derecesini belirleyebilmek adına “Kurumsal Entegrasyon Ölçeği” ni geliştirmişlerdir. İlk aşamada 1457, sonraki

aşamada 773 öğrenciyle gerçekleştirilen bu çalışmada; akademik ve sosyal entegrasyonun hazır ölçekleri ve faktör analizi vasıtasıyla boyutları ortaya konulmuştur: Arkadaş- Grup etkileşimi ( $\alpha$ : 0.84), Öğretim Üyeleriyle Etkileşim ( $\alpha$ : 0.83), Öğretim Üyelerinin Öğrencilerle İlgilenmesi ( $\alpha$ : 0.82), Akademik ve Düşünsel Gelişim ( $\alpha$ : 0.74) ve Üniversiteye ve Hedefe Bağlılık ( $\alpha$ : 0.71) . French&Oakes (2004) bu geliştirilen ölçeğin güvenilirliğini ve geçerliliğini test etmişler ve çeşitli değişiklikler yaparak güvenilirlik oranlarında iyileştirmeler elde etmişlerdir. İlk aşamada 773, ikinci aşamada 1734 öğrenciyle gerçekleştirdikleri çalışmalarında doğrulayıcı faktör analizi kullanarak ilk aşamada geliştirilen hazır ölçeğin problemlili olabileceğini savunmuşlardır. Çalışmada sırasıyla; ifade analizi, altölçekler arası korelasyon, iç tutarlılık güvenilirliği ve doğrulayıcı faktör analizi kullanılmış ve yapılan çalışma sonucunda tutarlılık oranları ve uygunluk konusunda iyileştirmede bulunmuşlardır. Buna göre yeni  $\alpha$  değerleri şu şekilde olmuştur. Arkadaş- Grup etkileşimi ( $\alpha$ : 0.84), Öğretim Üyeleriyle Etkileşim ( $\alpha$ :0.89), Öğretim Üyelerinin Öğrencilerle İlgilenmesi ( $\alpha$ : 0.88), Akademik ve Düşünsel Gelişim ( $\alpha$ : 0.82) ve Üniversiteye ve Hedefe Bağlılık ( $\alpha$ : 0.76) Bu iyileştirilen Kurumsal Entegrasyon Ölçeği daha sonra Mühendislik öğrencilerinin başarı ve devamlılıklarının incelendiği bir diğer çalışmada da kullanılmıştır. Hiyerarşik çoklu regresyon analizi, kullanılan bağımsız değişkenlerin (Cinsiyet, Lise sıralaması, SAT puanı, Kurumsal Entegrasyon, Motivasyon, Oryantasyon Sınıfı) bağımlı değişken (Akademik Not Ortalaması) üzerinde etkilerini araştırmak için kullanılmıştır. Buna göre, SAT puanları, Lise Sıralaması ve Cinsiyet  $p = 0.05$  anlam düzeyinde önemli değişkenler olarak ortaya çıkmıştır (French vd. ,2005) .

Hosch (2008) 397 üniversiteden ilk yıl öğrencilerinin CIRP (Cooperative Institutional Research Program) anketlerini kullanarak yaptığı çalışmasında aşamalı regresyon kullanarak mezuniyet oranlarının üniversiteye göre değişimini ortaya koymaya çalışmıştır. Sonuç olarak ortaya 2 model konulmuş ve bu modellere ait  $R^2$  değerleri sunulmuştur. 6 yılda mezun olma oranının bağımsız değişken olarak kabul edildiği her iki modelde kullanılan değişkenler şu şekildedir:

**Tablo 2**

**6 yılda mezun olma oranına etki eden faktörler (Hosch, 2008)**

	<b>Değişkenler</b>	<b>R<sup>2</sup></b>
<b>Model1</b>	SAT puanı, Kampüste yaşayan öğrenci oranı, İlk dönem not ortalaması 2.0 in altında olanların oranı	0.76
<b>Model2</b>	Lisede derece ile mezun olanların oranı, Kampüste yaşayanların oranı, Bursla okuyan öğrencilerin oranı ve İlk dönem not ortalaması 2.0 in altında olanların oranı	0.73

Bu çalışmada dikkat çeken nokta; kampüste yaşayanların oranının her iki modelde anlamlı değişkenler arasında yer almasıdır. Bu durum (Astin A., 1997: s.654) ve (Nandeshwar vd. 2011) çalışmaları ile bağdaşmaktadır. Bu çalışmalardan yola çıkarak öğrencinin kampüste yaşamasının akademik performansa olumlu anlamda etki ettiği savunulabilir. Dolayısıyla “yaşayan” kampüsler oluşturmak bu bakımdan önem arz etmektedir.

Akademik performansın ve devamlılığın üniversite türüne göre değişip değişmediğini araştıran bir çalışma Kerkvliet&Nowell (2005) tarafından yapılmıştır. Çalışmaya sadece lisans düzeyinde eğitim veren Weber State University (WSU) ile Yüksek Lisans ve Doktora seviyesinde de eğitim veren Oregon State University (OSU)’de Mikro ve Makro Ekonomi derslerini alan 1200 öğrenci katılmıştır. Algılanan üniversiteden mezun olma olasılığını hesaplamak için Negatif Binom Dağılımı’ndan faydalanılmıştır. Sonuçlara göre; fırsat maliyetleri ve finansal yardım ile ilgili değişkenler bu iki üniversitede farklılık göstermiştir. WSU da hem çalışıp hem okuma, OSU da ise burslar okula devamlılığı desteklemektedir. Yeniden kayıt olma niyeti ise sadece WSU için önemli bir değişkendir. Bu noktada, değişkenlerin önem derecelerinin üniversite tipine göre değiştiği yargısına varılabilir.

Üniversitedeki ilk yıl kritik bir geçiş sürecidir. Çünkü, öğrenciler geriye kalan yıllarda akademik başarılarının ve devamlılıklarının temellerini bu yılda atarlar (Olani, 2009: s.1056). Üniversite yönetimleri bu durumun farkındadır ve çeşitli destek programlarıyla öğrencilere bu hassas süreçte yardımcı olmaya çalışmaktadırlar. Bu tür bir programın istenilen ölçüde faydalı olup olmadığına ölçülmesine yönelik bir araştırmada, A.B.D.'nin ortabatısında yer alan orta ölçekli bir üniversitede (University of Wisconsin-Eau Claire) 1,997 öğrenci ile yapılan bir araştırmada ilk yıl deneyim programı (FYE: First Year Experience) katılım durumu ile not ortalamaları arasındaki ilişki incelenmiştir. Çoklu regresyon ve tahminci logit fonksiyonları kullanarak ilk yıl not ortalaması ile ACT (American College Testing) puanı, kampüste yaşama ve hedefe uyumlu FYE ye katılım durumu arasında pozitif bir ilişki bulunmuştur (Jamelske, 2008). Bu noktadan hareket ile ilk yıl öğrencilere yönelik destek programlarının önemi vurgulanabilir.

Öğrenci başarısı ve devamlılığını sağlamak üzere, ilk yıl destek programları olduğu gibi yapılan projeler de mevcuttur. Ocak 2012 – Ekim 2012 arasında yürütülen ATTRACT (Enhance the Attractiveness of Studies in Science and Technology) bu projelere bir örnek teşkil etmektedir. Avrupa çapında çeşitli ülkelerde yürütülen bu projenin bulgularının paylaşıldığı çalışmada, geri besleme ve ileri besleme sistemlerinin anketlerle yapılması, sosyal becerinin öğrenilmesi, matematik desteği, E-öğrenme, öğrenci izleme uygulamaları, danışmanlık ve rehberlik, gelecek kariyer fırsatları gibi konular vurgulanmıştır. Uygulama aşamasında üniversitelelerin mevcut yapılarına uygun programlar izlemesi gerekliliği belirtilmiştir (Rintala vd. , 2012).

Geleneksel ve geleneksel olmayan değişkenlerin bir arada kullanıldığı bir diğer çalışmada (Byrne&Flood, 2008) Finans ve Muhasebe lisans programına kayıtlı 129 ilk yıl öğrencisine yapılan anketlerde; korelasyon ve regresyon analizleri ile genel ilk yıl performansı ve bağımsız değişkenler arasında kurulan hipotezler test edilmiştir. Buna göre; akademik performans geçmişi, muhasebe ile ilgili önceden bilgi sahibi olma, öğrencinin bölüme uygun olduğunu düşünmesi ilk yıl genel not ortalaması üzerinde önemli etkilere sahiptir.

Mckenzie&Schweitzer (2001) çalışmalarında Standart Regresyon Model vasıtasıyla öğrencilerin not ortalamaları ve öğrencilerin akademik geçmiş, kişilik ve psikoloji ile

ilgili deęişkenlerini kullanmışlardır. Büyük çaplı bir üniversitede Bilim ve Bilişim Teknolojileri Fakültesi'ne kayıtlı 197 ilk yıl öğrencisiyle gerçekleştirilen çalışmada sonuç olarak; öğrencinin üniversiteye entegrasyonu, özyeterlik, üniversiteye giriş puanı öğrencinin not ortalaması üzerinde önemli deęişkenleri olarak ortaya çıkmıştır. Bu çalışmanın dikkat çekici yönü; geleneksel olmayan deęişkenlerin temini için farklı disiplinlerde geliştirilmiş hazır ölçeklerden yararlanılmış olmasıdır. Çalışma becerileri için College Adjustment Scales Inventory ( $\alpha$ : 0.87) den bir alt ölçek, üniversiteye baęlılık, öğrenci- üniversite entegrasyonu, üniversiteden memnuniyet, kariyer odaklılık, finansal zorluklar ve sosyal destek için Himelstein'in (1992) 'in 4'lü Likert ölçeęini, öğrencinin psikolojik deęerlendirmesi için ise Lovibond&Lovibond (1995) tarafından geliştirilen DASS (Depression, Anxiety, Stress Scale) 'in kısa versiyonu kullanılmıştır. Bu durum akademik başarı ve devamlılık araştırmalarının disiplinlerarası yapısını ortaya koymasından önemlidir.

Jawitz&Scott (1997), University of Cape Town Mühendislik Fakültesi'nde ilk yıllarını geçiren 2201 öğrenci ile gerçekleştirdikleri çalışmada; çok deęişkenli analizden faydalanmıştır.  $p = 0.05$  anlam düzeyinde yapılan analizlerde; siyah öğrencilerin beyaz öğrencilere göre devam oranlarını düşük olduęu, beyaz kadın öğrencilerin ise beyaz erkeklere göre devam oranlarının daha düşük olduęu sonucuna ulaşılmıştır. Bu çalışma; ırk unsurunun çalışmanın yapıldığı ülkeye göre önemli olabilmesi açısından önemlidir. Dolayısıyla kullanılan deęişkenlerin uygulamanın yapıldığı üniversitenin bulunduęu ülkenin demografik yapısına göre farklılık gösterebileceęi savunulabilir. Bir dięer çarpıcı sonuç ise ASPECT (Academic Support Program for Engineering in Cape Town) programına katılan adaylardan, Elektrik Mühendisliği programına kayıtlı öğrencilerin, Kimya Mühendisliği programında okuyanlara göre daha başarılı olduęu belirtilmiştir. Bu bulgu ilk yıl programlarının akademik performansa olumlu etkisini göstermesi açısından Jamelske (2008) ile birbirini desteklemektedir. Ayrıca bu noktada; deęişkenlerin önem derecesinin ülkeden ülkeye olduęu gibi programdan programa deęişebileceęi hususu göz önünde bulundurulmalıdır.

Belanger vd. (2007) çeşitli üniversitelerden 626 öğrenciyle yaptıkları on-line anketlerden elde edilen verileri kullanarak hipotez testleri ve çoklu regresyon analizleri uygulamıştır. Baęımlı deęişken olarak not ortalaması belirlenmiştir. Duygusal zeka ve

mücadele stratejileri ve not ortalaması arasındaki ilişkiler teknik ve teknik olmayan lisans programlarına kayıtlı öğrenciler için ayrı ayrı ele alınmıştır. Analizler sonucunda ; teknik ve teknik olmayan lisans programlarına kayıtlı öğrencilerin, duygusal zeka kapsamında özyeterlik, duygusal zeka ve kabullenme stratejileri yönünden farklılık gösterdiği bulgusuna ulaşılmıştır. Ayrıca teknik olmayan bölümlerde okuyan öğrenciler için, içerisinde bulunulan duruma ya da problemi tanımlamaya yönelik aktif adımların atıldığı, bir başka deyişle aktif, adaptif ve problem odaklı bir mücadele stratejisi olan “durumu değiştir” akademik başarının anahtarı olarak sunulmuştur.

Choi (2005) araştırmasında havuz dersleri alan 205 öğrenci ile özyeterlik ve özbenlik kavramlarının akademik performansa olan etkileri araştırmıştır. Genel Özyeterlik Ölçeği (Sherer&Adams, 1983) ve Akademik Özyeterlik Ölçeği’nden (Owen&Froman, 1988) faydalanılmıştır. Regresyon analizi yapılan bu çalışmada derslerden alınan notlar ile akademik özbenlik, spesifik özbenlik ve spesifik özyeterlik arasında anlamlı pozitif bir ilişki ortaya konmuştur. Bireyin kişilik özelliklerinin akademik performansa etkisini vurgulayan çalışmalardan bir diğeri de akademik disipline vurgu yapmıştır (Komarraju vd., 2013). Bilişsel ve bilişsel olmayan değişkenlerin akademik performanstaki rolü üzerine 540 öğrenci ile yapılan bu çalışmada, katılımcılara psikososyal ve akademik becerileri ölçmek üzere geliştirilen 108 ifadeli Student Readiness Inventory (Le, 2005) uygulanmıştır. Araştırma sonucunda, lise not ortalaması, ACT puanları ve akademik disiplin değişkenlerinin, not ortalaması üzerinde önemli bir rol oynadığı ortaya konmuştur.

Faktörlerin akademik performans üzerinde dolaylı etkilerinin olduğu çalışmalar literatürde yer almaktadır (Belanger vd., 2007). Furnham vd. (2003) İngiltere’de bir üniversitede 93 öğrenci ile gerçekleştirdiği çalışmalarında; 5 büyük kişilik tipi, bilişsel yetenek ve zeka hakkındaki inançların akademik performansa (derslerde alınan notlar) ve seminer dersi (sınıftaki davranışlar, deneme notları, derse devam) performansına olan etkilerini incelemiştir. Hiyerarşik regresyon analizi kullanılarak akademik performans üzerinde sorumluluk sahibi olmanın pozitif, dışadönüklüğün negatif etkileri sonuçlarla desteklenmiş ve akademik performans düzeylerinde cinsiyet bakımından farklılıklar tespit edilmiştir. Kişilik özelliklerinin deneme notları, sınıf davranışları ve devamsızlık

üzerinde anlamlı etkilerinin direkt olmasa da akademik performans üzerine dolaylı olarak etkileri olduğu bu çalışmada da savunulabilir.

Öğrencilerin, bir lisans programı boyunca bölüm ile direkt olarak ilişkili derslerden alınan notların diploma notuna etkisini araştıran bir çalışmada anlamlı sonuçlara ulaşılmıştır (Golding&Donaldson, 2006). Jamaika’da yapılan araştırmada, 2001-2002 akademik yılında Bilgisayar ve Bilişim Teknolojileri programına kayıt yaptıran öğrenciler ile yaptığı uzun soluklu çalışmada, alan ile ilgili geçiş derslerinden ilk yıl alınan (C programlama, Ağ Teknolojileri, Bilgisayar Mantık ve Dijital Tasarım) ve 2. yıl alınan (Algoritma Analizi ve Yazılım Mühendisliği) başarı notları ile diploma notu arasında pozitif yönde bir ilişki gözlemlenmiştir. Öğrencinin seçtiği alan ile ilgili derslerdeki performansının diploma notuna olan pozitif etkisi, üniversiteye geçiş sürecinde öğrencilerin ilgi alanlarına uygun bölümler seçmelerinin akademik performansı olumlu yönde etkilediğini gözler önüne sermektedir. Dolayısıyla, öğrencinin okuyacağı lisans programını tercih aşamasında ilgi alanları doğrultusunda yönlendirilmesi faydalı olacaktır. Bu noktadan hareket ile, öğrencilerin akademik performansına ve devamlılığına etki eden hassas geçiş sürecinin tercih aşamasından başladığı iddia edilebilir.

#### **1.6.2.2 Tahminleyici Çalışmalar**

Her alanda olduğu gibi, günümüzde eğitim alanında da bilgi teknolojileri ve bilişim sistemleri vasıtasıyla her gün daha fazla veri elde edilmekte ve bu veriler anlamlı bilgilere ulaşmak amacıyla kullanılmaktadır. Çağdaş eğitimde değişen faktörler, günümüzde, etkin ve verimli bir şekilde öğrenci performans takibini geleneksel ölçüm yöntemlerinden hayati derecede önemli örtülü ve gizli bilgilere ulaşmak amacıyla, çeşitli veri anlama ve araştırma yöntemleri kullanan veri vadenciliği tekniklerine geçiş yapmakta olan üniversiteler için önemli bir görev haline getirmiştir (Ogor, 2007: s. 354). Bu görevi yerine getirmek üzere eğitim alanında karar vericiler ise; ellerinde bulunan büyük miktarlarda ham veriyi eğitsel araştırma meselelerini çözüme kavuşturmak amacıyla kullanmaktadırlar (Romero&Ventura, 2010: s. 601). Son yıllarda Eğitim Bilişim Sistemlerinde saklanan ham verilerin kullanılabilir faydalı bilgilere dönüştürülmesi yönündeki çabalar literatürde “Eğitimde Veri Madenciliği” başlığı

altında karşılık bulmuştur. Bu noktada “Veri madenciliği” kavramına kısaca değinmek, çalışmaların daha iyi anlaşılabilmesi bakımından yararlı olacaktır.

Literatürde her ne kadar “Veri madenciliği” adıyla anılsa da; bu kavramı “Veri tabanlarında bilgi keşfi” süreci olarak adlandırmak daha uygun bir tanımlama olacaktır (Fayyad vd. ,1996). Bu tanıma göre; veritabanlarında bilgi keşfi, sürecin tümünü kapsarken; veri madenciliği bu süreçte bir adım olarak düşünülebilir. Sürecin aşamaları ise Fayyad vd. (1996) tarafından şu şekilde tanımlanmıştır.

- 1- Anlaşılabilir bir uygulama kapsamı geliştirmek ve KDD sürecinin amacını müşteri bakış açısıyla belirlenmesi.
- 2- Hedef bir veri kümesi oluşturulması
- 3- Veri temizleme ve ön işleme
- 4- Veri indirgeme ve yansıtma, işin amacına uygun olarak kullanılabilir özellikler bulunması
- 5- KDD sürecinin amacını uygun bir veri madenciliği yöntemiyle eşleştirmek. Örneğin, özetleme, sınıflandırma, regresyon, kümeleme vs...
- 6- Keşifçi analiz, model ve analiz seçimi, bu süreç hangi parametreler ve modellerin uygun olduğuna karar verilmesini içerir.
- 7- Veri madenciliği, istenen formatta örüntü araştırma
- 8- Çıkarılmış örüntülerin yorumlanması ve çıkarılan örüntüleri görselleştirilmesi
- 9- Keşfedilmiş bilgi doğrultusunda harekete geçme: dokümantasyon ve ilgili birime iletme

Bir diğer bakış açısına göre; veri madenciliği örüntüleri ve basit analizlerin ötesine geçen trendleri keşfetmek amacıyla büyük veri yığınlarının otomatik olarak aranması uygulamasıdır. Veri madenciliği gelecekte yaşanacak olayların olasılıklarını değerlendirmek ve verileri bölümlendirmek üzere sofistike matematik algoritmaları kullanır ve “Veride Bilgi Keşfi” olarak da bilinir (Oracle, 2015). Bu ifade (Fayyad vd., 1996) ile ayrılmaktadır. Ancak literatürde bu iki kavramı birbirinin yerine kullanan araştırmacılar yadsınamayacak kadar fazladır (Romero&Ventura, 2010; Ogor, 2007; Bhardwaj&Pal, 2011; Baker, 2010) .

Kavram karmaşasının önüne geçmek adına, eğitim çevrelerinde bilgi keşfi sürecinden bundan sonraki kısımlarda “Eğitimde Veri Madenciliği (EVM)” olarak bahsedilecektir.



EVM, tipik veri madenciliği tekniklerinden sınıflandırma, kümeleme, özdeşlik kuralı çıkarımı, ardışık madencilik ve metin madenciliği ile çalışsa da; regresyon, korelasyon ve görselleştirme gibi diğer yaklaşımları da tanımlamaktadır (Romero&Ventura, 2010: s. 601). EVM çeşitli amaçlar doğrultusunda pek çok kullanıcı ve paydaş (öğretim üyeleri, öğrenciler, eğitim araştırmacıları, üniversitede yöneticiler) tarafından kullanılmaktadır. Çalışmamızın amacında belirtildiği üzere; öğrencilerin akademik performanslarına etki eden faktörlerin belirlenmesi ve bu faktörlere bağlı olarak genel not ortalamalarının tahmin edilmesi, olası risk gruplarını belirleyerek ilgili önlemleri alması gereken üniversite yöneticilerinin ilgi alanına girmektedir. Yöneticiler EVM' yi:

- Yükseköğretim kurumlarında karar sürecinin zenginleştirilmesi
- Karar verme sürecinde verimliliğin düzene konması
- Spesifik hedeflerin gerçekleştirilmesi
- Devamlılık ve derslerden alınan notları iyileştirecek maliyet etkin yolların bulunması
- Başvuranlar arasında en kalifiye olanın seçilmesi
- Üniversitede iyi performans sergileyecek öğrencilerin seçilebilmesi

amaçları doğrultusunda kullanılmaktadırlar (Romero&Ventura, 2010: s. 602).

Veri madenciliği, eğitim ortamında kendisine pek çok noktada uygulama alanı bulmuştur. Romero&Ventura (2010: s. 604-610), EVM kavramını ve mevcut durumunu inceledikleri çalışmalarında bu alanları şu şekilde başlıklandırmışlardır:

- 1- **Verinin analizi ve görselleştirilmesi:** Amaç, kullanışlı bilgilerin özetlenmesi ve karar sürecine destek sağlanmasıdır.
- 2- **Öğretim üyelerine destek için geri dönüş sağlanması:** Öğretim üyelerinin öğrenmenin iyileştirilmesi, kaynakların etkin kullanılması gibi konularda karar almalarının kolaylaştırılması ve gerekli proaktif önlemleri almalarının sağlanması amaçlanır.
- 3- **Öğrenciler için öneriler:** Her bir öğrenciye özel olarak faaliyetleri (ödev, sınav, sınıf performansı) gibi konularda önerilerde bulunulması amaçlanır.
- 4- **Öğrencinin performansının tahmin edilmesi:** Öğrenciyi tanımlayan bir bilinmeyen değişkenin tahminine yönelik çalışmalardır.

- 5- **Öğrenci modelleme:** Öğrencinin bilişsel açıdan modellenmesi amaçlanır.
- 6- **İstenmeyen öğrenci davranışlarının tespit edilmesi:** Düşük motivasyon, yanlış faaliyetler, kopya çekme, okulu bırakma, istismar, akademik başarısızlık gibi istenmeyen durumların belirlenmesine yönelik çalışmalardır.
- 7- **Öğrenci gruplandırma:** Benzer özellik ve karakterdeki öğrencilerin belirlenerek kişiselleştirilmiş bir eğitimin mümkün kılınabilmesi için yapılan uygulamalardır.
- 8- **Sosyal ağ analizi:** Sosyal ortamın analizine yönelik çalışmaları barındırır.
- 9- **Kavram haritalarının geliştirilmesi:** Kavramlar arası ilişkiler ve bilginin hiyerarşik yapısını görselleştiren kavram haritalarının geliştirilmesi için yapılır.
- 10- **Eğitim yazılımı geliştirme:** Yazılım geliştirme sürecinde öğretim üyelerine ve yazılımcılara destek sağlanması istenmektedir.
- 11- **Planlama ve çizelgeleme:** Gelecekteki derslerin planlanması, ders programı oluşturmaya yardımcı olma, kaynak tahsislerinin planlanması ve kabul ve danışma süreçlerinde yardımcı olunması ve müfredat geliştirme gibi konularda EVM ‘ ye sık sık başvurulmaktadır.

Öğrencinin performansının tahmin edilmesi ve istenmeyen öğrenci davranışlarının tahmini alanlarında yapılan çalışmaların; hassas bir geçiş dönemi yaşayan ilk yıl öğrencilerinin entegrasyonu ile ilgili olduğu savunulabilir. Bu çalışmalarda, daha çok veri madenciliği tekniklerinden öğrencileri daha önceden belirlenmiş sınıflara (Mükemmel, İyi, Kötü vs.) atayarak tahminde bulunmayı amaçlayan “Sınıflandırma’dan” (Classification) yararlanılmıştır.

Araştırma problemine bağlı olarak uygun veri madenciliği algoritmasını seçmek, çalışmanın makul sonuçlara ulaşabilmesi için önem arz etmektedir. Bu noktada veri madenciliği teknikleri ve uygun algoritmalar şu şekilde tanımlanabilir :

**Tablo 3****Veri madenciliği algoritmaları kategorizasyonu (Jing, 2002: s.13)**

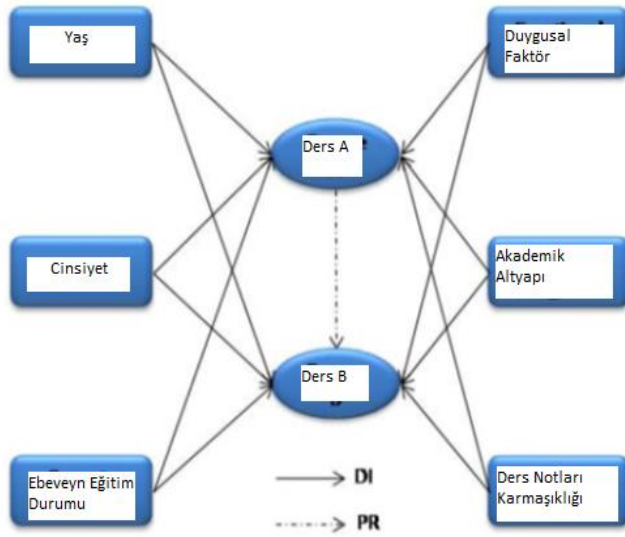
Görevler	Öğreticili	Öğreticisiz
Sınıflandırma	Hafıza temelli mantık, Genetik Algoritmalar, C&RT, Bağıntı Analizi, C 5.0, Yapay Sinir Ağları	Kohonen Ağları
Tahminleme	Yapay Sinir Ağları, C&RT	
Kümeleme/ Özdeşlik	Sepet Analizi, Hafıza temelli mantık, Bağıntı Analizi, Kural Çıkarımı	Küme tespiti, K-ortalamalar, Genelleştirilmiş kural çıkarımı, APRIORI, İki adım
Betimleme	Kural Çıkarımı, Sepet Analizi	Uzamsal görselleştirme

Romero vd. (2008) 7 Moodle (Açık kaynak kodlu online bir öğrenme platformu) dersine kayıtlı 437 Cordoba Üniversitesi öğrencilerine ait gerçek verilerle yaptıkları sınıflandırma çalışmalarında; öncelikle her bir öğrenciye ait nümerik özellikleri (Ders numarası, yapılan ödev sayısı, girilen quiz sayısı, geçilen quiz sayısı vs.) manuel olarak kategorik hale getirmişlerdir (Değer<5 ise KALDI, Değer >=5 ve Değer<7 ise GEÇER vs.). Elde edilen bu verileri değişik veri madenciliği teknikleri (Karar Ağaçları, Kural Çıkarımı, Bulanık Kural Öğrenme ve Yapay Sinir Ağları) kullanarak sınıflandırmışlardır. Elde edilen sonuçların yeniden dengelenmiş hallerine göre; kural çıkarım algoritmalarından olan AprioriC algoritması en isabetli tahmini sunmuştur (%61.64). Bu durumun, sadece nümerik değişkenlerin kullanılması ve değişkenlerin az sayıda olmasından kaynaklandığı öne sürülebilir. Sadece geleneksel değişkenleri kullanan bir diğer çalışmada (Delen, 2010); araştırmacı, 2004-2008 yılları arasında çeşitli üniversitelere kayıt yaptırmış 16066 öğrenci ile akademik devamlılığı tahmin etmeye çalışmıştır. Öğrencilerin akademik, demografik ve finansal özellikleri ile ilgili 39 geleneksel değişkenin kullanıldığı çalışmada, öğrencilerin ikinci yıl aynı okula kayıt yaptırma durumu bağımlı değişken olarak kabul edilmiştir. Sınıflandırma için, Karar

Ağacı (C5), Yapay Sinir Ağları (MLP), Lojistik Regresyon, Support Vector Machines, Bagging, Busting ve Information Fusion tekniklerinden faydalanılmıştır. Dengelenmiş veri setlerinden elde edilen sınıflandırma sonucunda doğru sınıflandırma oranları : Yapay Sinir Ağları (%79.85), Karar Ağacı (%80.65), Support Vector Machines (%81.18), Lojistik Regresyon (%74.26), Busting (%80.21), Bagging (%81.80) ve Information Fusion (%82.10) olmuştur. Her bir değişkenin modelden çıkartılarak, göreceli önem derecesinin belirlenmesine yönelik uygulanan duyarlılık analizi sonucunda ise, en önemli 3 değişken şu şekilde sıralanmıştır: Geçilen Kredi/ Kaydolunan Kredi, Kredi alma durumu, Güz dönemi not ortalaması. Bu çalışmada çok daha fazla sayıda değişkenin kullanılmasının, öğrenci performansını ve davranışını çok daha kapsamlı olarak modelleyebildiği savunulabilir. Romero vd.'nin (2008) çalışmasına göre daha isabetli sonuçlar ortaya konması bu görüşü destekler niteliktedir. Benzer bir diğer çalışmada Zaidah&Rusli (2007) 206 öğrenci ile gerçekleştirdiği araştırmasında, öğrencilere ait geleneksel değişkenleri (Bilişim teknolojileri uygulama bilgisi, önceki okul tipi (yatılı/ yatılı olmayan), programlama bilgisi ve ailenin finansal durumu) kullanarak diploma notunu tahmin etmeye çalışmıştır. Karar Ağaçları, YSA ve Lineer regresyon analizinden yararlanan bu çalışmada, tekniklerin her biri %80'in üzerinde doğru tahminlemede bulunmuşlardır. Akademik devamlılığın tahminine yönelik şimdiye dek en fazla geleneksel değişken (n=145) kullanılarak yapılan çalışmada ise Jing (2002) çok katmanlı YSA, C 5.0 ve C&RT kullanmıştır. Sonuç olarak, çok katmanlı YSA için ortalama doğru tahmin oranı %78.8, C 5.0 algoritması için %65.8 ve C&RT için ise ortalama %32.4 olarak gerçekleşmiştir. Mishra vd. (2014) bu çalışmaları bir adım öteye götürerek, öğrencinin akademik performansını tahmin etmeye çalıştıkları 250-öğrencilik araştırmalarında, geleneksel olmayan değişkenlerden de (özgüven, empati, karar verme yeteneği, liderlik, motivasyon, stres yönetimi becerileri) faydalanmışlardır. WEKA veri madenciliği paket programı ile değişik Karar Ağacı Elgoritmaları (J48, Random Tree) kullanılarak, 3. yarıyıl not ortalama aralıkları tahmin edilmiştir. Sonuçlara göre; J48 algoritması ile %88.37, Random Tree algoritması ile ise %94.41 doğru sınıflandırma oranlarına ulaşılmıştır. Bu çalışmada, bahsi geçen diğer çalışmalarda ortaya çıkan doğruluk oranlarına göre çok daha isabetli sınıflandırma oranlarına ulaşılmıştır. Elde edilen bulgulara göre, geleneksel ve geleneksel olmayan değişkenlerin akademik başarı ve devamlılığın tahmininde kullanıldığı çalışmalarda,

değişken sayısı ve geleneksel olmayan değişkenlerin modele dahil edilmesinin sınıflandırma doğruluğuna olumlu bir etki yaptığı savunulabilir.

Gerçekleşmiş olanlardan yola çıkarak henüz gerçekleşmemiş olanı tahmin etmeye çalışmak, veri madenciliğinin temelinde yatan düşüncelerden birisidir. Slim vd. (2014) bir önceki performansı baz alarak sonraki performansı tahmin etmeye çalışmışlardır. 115746 öğrenci kullanılarak oluşturulan Durumsal Olasılık Tablosu, New Mexico University'den rastgele seçilen 400 öğrenci ile test edilmiştir. Çalışmada, öğrencilerin bir dersten alacağı not Bayes İnanç Ağları (BBN) kullanarak tahmin edilmeye çalışılmış ve her bir dersten alınan notlardan yola çıkılarak öğrencinin tahmini genel performansı ortaya konmuştur.



**Şekil 12: Transkript grafiğinin BBN modeli (Slim vd. , 2014: s. 3)**

Şekilde görselleştirilen transkript grafiğinde derslerin öncül-ardıl ilişkisi görselleştirilmiştir. Buna göre, her bir öğrencinin bir sonraki dönem not ortalaması Maximum a Posteriori Probability (MAP) ve Expected Grade (EG) olasılık fonksiyonları ile hesaplanmış ve bu fonksiyonlar için performanslar MSE (hataların karelerinin ortalaması) kullanılarak karşılaştırılmıştır. Olasılık fonksiyonları ise; BBN kullanan Yapı 1 ve önceki dönemin sonraki döneme etkisi olmadığını varsayan Yapı2 ile test edilmiştir. Sonuca göre önceki performansın sonrakine etkisi olduğunu kabul

eden BBN Yapı1, Yapı 2 ye göre çok daha doğru sonuçlar üretmiştir. Ayrıca olasılık fonksiyonları karşılaştırıldığında MAP fonksiyonu çok daha doğru sonuçlar vermiştir. Bu çalışmada, bir önceki dönemin performansının, bir sonraki dönem üzerinde etkili olabileceğini göstermesi açısından önemlidir. Benzer şekilde gerçekleştirilen bir çalışmada, 116 öğrencinin Matematik (Calculus 1) dersi seviye belirleme testinde verdikleri cevaplara göre; derslerden alacakları notlar iki farklı yapıda (Geriye Yayılımlı ve Hibrit yayılımlı) Yapay Sinir Ağları tarafından tahmin edilmiştir. Orijinal verisetinin miktarının yetersiz olmasından ötürü eğitim için veri türetilmiştir (Noisy Data) (Fausett&Elwasif, 1994). Bu iki farklı yapının Calculus 1 dersinden alınan notları doğru tahmin etme oranları ise şu şekilde gerçekleşmiştir:

**Tablo 4**

**Farklı Yapıda Yapay Sinir Ağlarının tahminleme doğrulukları (Fausett&Elwasif, 1994)**

	<b>Orijinal veriseti</b>	<b>Türetilmiş veriseti</b>
<b>Hibrit Yapılı YSA</b>	%80	%78
<b>Geriye Yayılımlı YSA</b>	%98.3	%97.7

Yüksek doğru tahminleme oranlarından anlaşılacağı üzere, öğrencinin geçmişte göstermiş olduğu performans, geleceği hakkında bir şeyler anlatmaktadır (Slim vd., 2014: s. 2). Bu görüşü destekleyen bir diğer çalışmada; MOODLE (Modular Object-Oriented Dynamic Learning Environment) loglarını (kayıtlarını) kullanarak öğrencilerin bu sistem üzerinde aldığı derslerde dönem sonu performansını tahmin edilmeye çalışılmıştır. Araştırma, University of Cordoba’da Metodoloji ve Programlama Teknolojisi Moodle dersine kayıtlı 240 öğrenci ile gerçekleştirilmiştir. Öğrencilerin öncelikle sistemde tutulan log verileri (IP adresi, tam kullanıcı adı, sitede dolaştığı kısımlar vs.) kullanılarak idari verilere (tam adı, öğrencinin dersi alttan alma durumu vs.) ulaşılmıştır. Buradan elde edilen her bir veriyle bir Yapay Sinir Ağı çeşidi olan Radyal Temelli Fonksiyon (RBF)’a besleme yapılmıştır. Sonuç olarak; artan gizli katman yöntemiyle Yapay Sinir Ağı tahminlerinde % 80 doğruluk oranı tutturulmuştur (Calvo-Flores vd., 2006). Bu çalışma üniversitelerin veritabanlarında her geçen gün boyutları artmasına rağmen atıl kalmakta olan logların kullanımın sağlayabileceği faydalı bilgilere önemli bir örnek teşkil etmektedir. Çalışmalardan anlaşılacağı üzere,

Yapay Zeka tekniklerinden birisi olan Yapay Sinir Ağları, EVM uygulamalarında sıklıkla kullanılmakta ve arařtırmacılara faydalı bilgiler sağlamaktadır. Ancak, Yapay Sinir Ağları'na yönelik en büyük eleřtirilerden birisi, "kara kutu" ya benzeyen alıřma prensibi ve olası nedensel iliřkileri aık bir řekilde ortaya koyamamasıdır (Tu, 1996: s. 1229). Arařtırmacılar bu sorunun üstesinden gelebilmek üzere YSA'nın ürettiđi sonuçları kurallara evirmek üzerinde alıřmıřlardır. Konuyla ilgili bir makalede, YSA ıktıları doktorların teřhis koyma sürecine benzetilerek, ıktıların muhtemel girdileri ile birlikte sunulmasına yönelik bir prosedür geliřtirilmiřtir (Gedeon&Turner, 1993). řüphesiz ki, bu yöntem karmařık bir yapıya ve ıktılara sahip olan Yapay Sinir Ağları'nın eđitim evrelerinde ok daha geniř alanlarda daha fazla paydař tarafından kullanılmasına ön ayak olacak özelliğindedir.

Akademik anlamda iyi performans gösteren öđrencilerin profilini belirlemeyi amalayan bir alıřma Romanya'da bir üniversitede İktisat ve İşletme Fakültesi lisans programlarına kayıtlı 50 öđrenciyle gerekleřtirilmiřtir. İlk ařamada öđrenciler, K-ortalamalar algoritması kullanarak kümelene miř ve bařarılı olanların profilleriyle ilgili kural ıkarımında bulunabilmek üzere J48 Karar Ađacı algoritmasından faydalanılmıřtır. Cinsiyet, Lise Diploma Notu gibi geleneksel deđiřkenlerin yanısıra, öđrencilerin görüşlerinin yer aldıđı geleneksel deđiřkenlerin de kullanıldıđı bu alıřmada, geerleme işle mi sonucunda %76 dođru sınıflandırma oranına ulařılmıřtır. Sonuç olarak akademik bařarı durumuyla ilgili "Eđer bir öđrenci lisansüstü eđitime istekliyse ve tüm beklentilerinin karřılıandıđına inanıyorsa; o halde öđrenci hibir dersten kalmayacaktır." tarzında kurallar oluřutulmuřtur. Bu řekilde ıkarılan kurallar odaklanılması gereken boyutlar üzerinde karar vericilere yardımcı olacak niteliktedir (Bresfelean, V. P. vd. , 2008).

Biliřim teknolojilerinin yaygınlařması ile birlikte; yükseköđretimde son yıllarda yeni eđitim řekilleri ortaya ıkmaktadır. Uzaktan eđitim, Karma eđitim, Aıköđretim bu yeni eđitim řekillerine birer örnek teřkil etmektedir. Her geen yıl üniversiteler, uzaktan eđitim, aıköđretim ve karma eđitim lisans programları amakta ve bu programlarda okuyan öđrenci sayısında hızlı bir artıř yařanmaktadır. Bu alanda kabul edilebilir derecelerde akademik performansın ve devamlılıđın sađlanabilmesi için arařtırmacılar alıřmalarını yođun bir řekilde sürdürmektedirler (Kotsiantis&Panayiotis, 2005;

Gazza&Hunker, 2014; Mancini vd. , 2015; DeTure, 2004; Hoskins&Van Hooff, 2005). Bu çalışmalarda dikkat çeken nokta; örgün öğretim programlarında çoğunlukla anlamlı düzeyde etkileri olmayan değişkenlerin (medeni hal, yaş, çocuk sayısı vs.) bu tip eğitimleri tercih eden aday profiline bağlı olarak anlamlı farklar oluşturabilmesidir. Bu durumun nedeninin açıköğretim, online eğitim gibi lisans programlarına kayıtlı olan öğrencilerin çoğunun eğitimine örgün olarak devam etmelerini engelleyecek başka sorumlulukları olduğu öne sürülebilir. Dolayısıyla akademik performans ve devamlılığın tahminine yönelik yürütülen araştırmalarda değinilmesi gereken bir diğer husus, öğretim türüdür.

Superby vd. (2006), 3 farklı üniversiteden 533 öğrenciyle yaptıkları çalışmalarında kişisel geçmiş (okul geçmişi, sosyo- ailevi yapı), öğrenci davranış göstergeleri (devam oranı, ders dışı aktivitelere katılım) ve öğrenci algıları (özgüven, akademik içerik algısı) gibi konularda 42 soruluk bir anket uygulamışlardır. Anketlerden elde edilen veriler değişik veri madenciliği teknikleri ile irdelenmiş ve öğrenciler bağımsız değişkenin (not ortalaması) değerlerine göre 3 ayrı risk grubuna ayrılmıştır. (Yüksek-risk, Orta-risk, Düşük-risk) Bu sınıflandırma çalışmasının sonuçları her bir yöntem için şu şekilde tablolastırılabilir.

**Tablo 5**

**Çeşitli veri madenciliği tekniklerinin sınıflandırma doğruluk oranları (Superby vd., 2006)**

	Karar Ağacı	Random Forest	Yapay Sinir Ağı	Lineer Diskriminant Analizi
Yüksek-risk	%48.65	%22.92	%%45.95	%45.95
Orta-Risk	%18.46	%68.72	%47.06	%57.35
Düşük-risk	%60.34	%55.06	%61.82	%67.27

Bu tabloya göre, öğrencileri farklı risk gruplarına ayırmada Lineer Diskriminant Analizi'nin ortalama olarak en isabetli sonuçları verdiği belirtilebilir. Bu çalışmayı diğer çalışmalardan ayıran özellik hem betimleyici hem de tahminleyici çalışmaların bir arada yapılmış olmasıdır. Çalışmanın diğer temel sonucuna göre; öğrencinin son yıl lise



not ortalaması, derslere devam oranı, öğrencinin yeteneklerine olan güveni, akademik başarıyla pozitif yönde korelasyon göstermektedir.

Yapılan teorik ve uygulamaya dönük literatür taramasını özetlemek gerekirse; akademik başarı ve devamlılık üzerine yapılan araştırmalar; yeni bir akademik ve sosyal ortama geçiş dönemindeki öğrencilerin içerisinde buldukları sürecin daha iyi anlaşılmasını sağlayacak kavramsal modellerin geliştirilmesi ve bu modeller üzerine kurulmuş uygulamalar üzerine yoğunlaşmıştır.

Teorik modeller sıklıkla, bazı faktörlerin öğrencinin akademik başarısı ve devamlılığında etkili olabileceğini iddia eden hipotez testleriyle yapılan çalışmaların derlenmiş sonuçlarını kullanmaktadır. Örneğin, şimdiye dek akademik başarı ve devamlılık üzerine yapılan uygulamalarda en sık faydalanılan çalışmalardan biri olan (Tinto, 1975) o zamana kadar olan literatüre yönelik bir taramadır. Yazar yaptığı araştırma sonrasında ortaya öğrencilerin akademik devam kararını etkileşim perspektifinden (Milem&Berger, 1997) teorik olarak modellemiştir. Bir başka deyişle, öğrencilerin akademik başarısına dönük hipotez testleri ile bağımsız değişkenler belirlenmeye çalışılmakta, bu değişkenlerin mantıksal olarak bir araya getirilmesiyle teorik modellerin temelleri atılmaktadır.

Kavramsal çalışmalarda yıllar içerisinde yaş, cinsiyet, ırk, lise not ortalaması gibi geleneksel değişkenlerden öğrencinin içerisinde bulunduğu akademik ve sosyal ortama dönük algı ve tutumlarına yönelik geleneksel değişkenlere bir geçiş yaşanmıştır. (Pascarella&Terenzini, 1991; Terenzini&Pascarella, 1980; French vd., 2005) Çoğunlukla Tinto'nun (1975,1993) ortaya koymuş olduğu modellerden yola çıkılarak yapılan betimleyici çalışma sonuçları, tahminleyici çalışmalarda bağımsız değişken olarak ele alınmıştır. Ancak bu etkileşimli modellerin öğrencinin karakteristik özellikleri ve içerisinde bulunduğu finansal durumu yeterince vurgu yapmaması nedeniyle eleştirilere maruz kalmıştır. Bu eksiklikten yola çıkılarak kişinin karakteristik özellikleri (genel özyeterlik, akademik özyeterlik, duygusal zeka, kişilik tipi vs.) ve finansal durumunu bağımsız değişkenler olarak ele alan araştırmalar yapılmıştır (Cabrera vd., 1992; Choi, 2005; Belanger vd., 2007; Chemers, 2001; Furnham vd., 2003; Komarraju&Ramsey, 2013; Zaidah&Rusli, 2007). Sonuç olarak kavramsal

çalışmalar ile betimleyici çalışmaların birbirlerinden beslendiğini savunmak yanlış olmayacaktır.

İlgili literatürün diğer bir sacayağı olan tahminleyici çalışmalarda ise kullanılan yöntemler zaman içerisinde farklılık göstermiştir. Geleneksel istatistiksel sınıflandırıcılar (Park&Kerr, 1990; Gamoran, 1996; Mallette&Cabrera, 1991) veri toplama, depolama ve işleme teknolojilerinin günden güne gelişmesi ve veri erişilebilirliğinin artmasıyla birlikte hacmi devasa boyutlara ulaşan “big data” ile analiz yapabilme konusunda yetersiz kalabilmektedirler. Dolayısıyla araştırmacılar bu noktada, pek çok alanda büyük miktarlarda verilerle analiz konusunda yeteneklerini ispatlamış yöntemlere yönelmişlerdir. “makine öğrenme yöntemleri” olarak da bilinen bu yöntemler ile yapılan çalışmalarda Yapay Sinir Ağları (Delen, 2010; Mishra vd., 2014; Cripps, 1996; Hardgrave vd., 1994) Karar Ağaçları (Ramaswami&Bhaskaran, 2010; Superby vd., 2006; Baradwaj&Pal, 2011; Vandamme vd. , 2007; Delen, 2010) başlıca faydalanılan modern sınıflandırıcılar olarak göze çarpmaktadır. Süreci tahminleme amaçları doğrultusunda modelleyebilmek üzere “veri temelli” ve “anket temelli” yaklaşımlar mevcuttur. Ancak akademik performansın tahmin edilmesine yönelik güncel çalışmalarda daha çok veri temelli geleneksel değişkenler ve anket temelli geleneksel olmayan değişkenlerin birlikte kullanıldığı “hibrit yaklaşım” lardan faydalanılmaktadır. (Caison, 2007; Mckenzie&Schweitzer, 2001)

Yapılan çalışmalarda anketler aracılığıyla elde edilen algıya yönelik verilerin tahminleme modellerinde kullanılmak üzere kategorik değişkenlere dönüştürülmesine dair bir metodoloji eksikliğinden söz edilebilir. Bu noktada mevcut çalışmada, tek ve çok ölçekli anketlerden elde edilen verilerin kesikli hale getirilmesine dönük bir metodoloji ortaya konmaya çalışılacaktır. Literatürde göze çarpan bir diğer eksiklik ise, tahminleme modellerine dahil edilecek bağımsız değişkenlerin seçimine yönelik yaklaşımlardır. Mevcut çalışmada, bağımsız değişkenlerin modele dahil edilmesi ile ilgili olarak korelasyon temelli bir yaklaşım benimsenmiştir.

Yerli literatürde, öğrencilerin akademik performansına etki eden faktörlerin tahmini ile ilgili betimleyici çalışmalar mevcut olmakla birlikte, bu faktörlere bağlı olarak akademik performansın tahminine ve genel bir karar destek sistemi tasarımı için “hibrit” yaklaşım benimseyen herhangi bir çalışmaya rastlanmamıştır. Bu bağlamda üniversite

yöneticilerine risk grubunda yer alan öğrencilerin önceden belirleyerek genel bir karar destek sistemi tasarlamak suretiyle eksiklik giderilmeye çalışılacaktır.

## **BÖLÜM 2: AKADEMİK PERFORMANSIN TAHMİNİNE YÖNELİK BİR MODEL TASARIMI VE UYGULAMASI**

Son yıllarda “Öğretici merkezli” eğitim anlayışından “öğrenci merkezli” eğitim anlayışına doğru bir geçiş yaşanmaktadır. Böyle bir ortamda, öğrencilerin akademik anlamda göstermiş oldukları performans, bir başka deyişle öğrenme derecesi, her zamankinden fazla önem arz etmektedir. Toplumun nitelikli işgücü ihtiyacını karşılamak misyonunu üstlenen yükseköğretim kurumları için ise, öğrencilerin yüksek akademik performans göstermesi, misyonunu gerçekleştirme derecesinin yükselmesiyle toplumsal kalkınmaya sağlayacağı büyük katkıdan ötürü önemli bir yere sahiptir. Bu bağlamda yükseköğretim kurumları, gelişen teknolojiye paralel olarak kullanımı günden güne yaygınlaşan bilişim sistemleri vasıtasıyla öğrencilerin olası akademik performansları konusunda ipuçları sunacak verileri toplamakta ve faydalı bilgilere ulaşabilmek amacıyla depolamaktadır. Ancak üniversitelerin veritabanlarında öğrenciler ile ilgili depolanan verilerden yola çıkan bilgi keşfi sürecinde her yükseköğretim kurumu aynı derecede başarılı olamamaktadır. Bu durum yükseköğretim kurumunun yer aldığı eğitim anlayışından kaynaklanmaktadır. Örneğin, bugün A.B.D.’de üniversiteler kamuoyunu aydınlatmak amacıyla genel performansına dair dünya çapında kabul görmüş performans göstergelerini paylaşmak zorunda iken, Türkiye’de yer alan üniversitelerde bu performans göstergelerinin hesaplanmasına yönelik bir çalışmaya rastlanmamıştır. Öte yandan, A.B.D.’de bu performans göstergelerinin toplumla paylaşılması zorunluluğu, üniversitelerin ilgisini öğrencilerin akademik performanslarının yükseltilmesine yönelik çalışmalara çevirmiştir. Üniversitelerin %95’inin ilk yıl gelen öğrencilere yönelik bir destek programına sahip olması bu duruma örnek olarak gösterilebilir (Jamelske, 2008: s. 375). Bu noktada araştırmanın probleminin daha iyi anlaşılabilmesi için, öncelikle mevcut durumun tasvir edilmesi faydalı olacaktır.

### **2.1 Türkiye’de Mevcut Durum**

“Akademik Performans” kavramı Türkiye’de çoğunlukla akademisyenlere ait performans değerlerinde (makale sayısı, atıf sayısı, bildiri sayısı, yer alınan projeler vs.) kendisine karşılık bulmaktadır. Öyle ki; akademik performans bir akademisyenin değişik ölçütler göz önüne alınarak belirlenen değeri olarak belirtilmiştir

(Kaptanoğlu,&Özok, 2006: s. 193). Dolayısıyla bu alanda yapılan çalışmalar akademisyenlerin bilimsel çalışmalarındaki performanslarına dönük olmaktadır. Nitelikli bilimsel çalışmalarını arttırmayı amaçlayan “Akademik Teşvik Ödeneği” 6564 nolu Yükseköğretim Personel Kanununun 2. Maddesinde yapılan değişiklik ile birlikte 14/11/2014 tarihinde yürürlüğe girmiştir. Bu kanun her ne kadar üniversitelerin temel misyonlarından olan “bilgi üretme” alanında performansı arttırmayı hedefliyor olsa da, bir diğer ana görev olan “bilgiyi aktarma” konusunda herhangi bir teşvik içermemektedir. Akademisyen performansının değerlendirilmesi ve iyileştirilmesine yönelik bir diğer uygulama ise çeşitli üniversiteler tarafından öğrencilere uygulanan “Öğretim Elemanı Değerlendirme” anketleridir. Örneğin; Sakarya Üniversitesi’nde her dönem sonunda öğrenciler aldıkları derslerden ilan edilen başarı notlarını öğrenebilmek için SABİS (Sakarya Üniversitesi Bilgi Sistemi) üzerinden dersin öğretim üyesi ile ilgili görüşlerini belirten anketi doldurmak zorundadırlar. Bu uygulama, akademik personelin performans boyutlarından biri olan öğretme hakkında bilgi sağlamayı amaçlamaktadır. Bu gibi örneklerden de anlaşılacağı üzere Türkiye’de yapılan mevcut performans değerlendirme ve iyileştirme çalışmaları ekseriyetle akademisyenlere yöneliktir. Öğrenci performansının geliştirilmesine yönelik uygulamalar güdük kalmıştır.

Makro boyutta düşünülecek olursa; Türkiye’de yer alan üniversiteleri sıralamak için ODTÜ Enformatik Enstitüsü bünyesinde kurulan URAP (University Ranking by Academic Performance) laboratuvarı sıralamayı belirlerken şu kriterleri kullanmaktadır:

- Makale sayısı
- Öğretim üyesi başına düşen makale sayısı
- Atıf sayısı
- Öğretim üyesi başına düşen atıf sayısı
- Toplam bilimsel doküman sayısı
- Öğretim üyesi başına düşen bilimsel doküman sayısı
- Doktora öğrenci sayısı
- Doktora öğrenci oranı
- Öğretim üyesi başına düşen öğrenci sayısı (ÖSYM2015, 2015)

Görüldüğü üzere üniversite sıralamalarında kıstas olarak kabul edilen maddelerin çoğunluğu “bilgi üretme” performansına dayandırılmıştır. U.S. News’in 2013 dünya

üniversiteler sıralamasına etki eden göstergelerde akademik itibar %22.5, devamlılık ve mezuniyet oranı %22.5'luk bir orana sahiptir (Morse, 2014). Dünyada durum böyleyken, üniversite sıralamalarının Türkiye’de sadece bir alana sıkıştırılması Türkiye’de henüz “öğretici merkezli” eğitimden “öğrenci merkezli” eğitim anlayışına geçişin tamamlanamadığı şeklinde yorumlanabilir.

Yükseköğretimde öğrencilerin akademik performansı kavramının daha iyi anlaşılması ve tahmin edilmesine yönelik resmi anlamda çalışmalar mevcut olmamakla birlikte, Türkiye’de araştırmacılar bu konunun bazı boyutları üzerinde durmuşlardır. Öğrenci devamsızlığı ve nedenleri (Altınkurt, 2008), sosyal ilişkiler (Gültekin& Dereboy, 2011), kaygı gibi psikolojik etkenler (Silah, 2003) kurumsal faktörler (Memduhoğlu& Tanhan, 2013), zaman yönetimi becerileri (Alay&Koçak, 2003), işbirliğine dayalı öğrenme (Sezer&Tokcan; 2003) ve akademik başarı arasındaki ilişki araştırılan konular arasında yer almaktadır. Bu çalışmalar her ne kadar akademik performansı ortaya çıkaran sürecin bazı boyutlarına ışık tutsalar da süreci bütüncül olarak ele alma konusunda yetersiz kalmışlardır.

Ayrıca yapılan literatür taramasında, hassas bir geçiş süreci yaşayan ilk yıl öğrencilerine yönelik bir çalışmaya rastlanmamıştır. Bu konuda teorik model ve bütüncül uygulamalar bakımından büyük bir eksiklikten bahsetmek yanlış olmayacaktır.

Çalışmanın uygulandığı üniversitede akademik performansı oluşturan sürecin tanımlanması, bağımlı değişkenin yapısının anlaşılması açısından önem arz etmektedir. Dolayısıyla Sakarya Üniversitesi’nin dönem ve genel not ortalamalarının belirlenme şeklinin ortaya konması faydalı olacaktır.

Sakarya Üniversitesi bünyesinde herhangi bir yükseköğretim programına kayıtlı bir öğrenci her dönem maksimum 30 AKTS (Avrupa Kredi Transfer Sistemi) lik ders alabilmektedir. Her bir öğrencinin başarı notu; o derse ait yarıyıl içi notu (ara sınav, kısa sınav ve ödev vb.) ile yarıyıl sonu (final sınavı) notunun dersin hocasının belirlediği oranlarda ağırlıklarından oluşmaktadır. Yapılacak değerlendirmede sınıfın başarı düzeyi, notların istatistiksel dağılımı ve sınıf ortalaması göz önünde bulundurularak belirlenir. Bağlı değerlendirme olarak adlandırılan bu değerlendirme sonucunda, o derse ait başarı notu, aşağıda karşılıkları verilmiş olan harf takdir olunarak verilir. Mutlak notu % 40’ın altında olan öğrenciler, bağlı notlarına bakılmaksızın FF notu alırlar.

**Tablo 6**

**Sakarya Üniversitesi Not Değerlendirme Sistemi (Sakarya Üniversitesi, 2011: s. 95)**

<b>Başarı Derecesi</b>	<b>Başarı Notu</b>	<b>Katsayı Puanı</b>
<b>Pekiyi (AA)</b>	<b>4.00</b>	<b>90-100</b>
<b>İyi-Pekiyi (BA)</b>	<b>3.50</b>	<b>85-89</b>
<b>İyi (BB)</b>	<b>3.00</b>	<b>80-84</b>
<b>Orta-İyi (CB)</b>	<b>2.50</b>	<b>75-79</b>
<b>Orta (CC)</b>	<b>2.00</b>	<b>65-74</b>
<b>Zayıf-Orta (DC)</b>	<b>1.50</b>	<b>58-64</b>
<b>Zayıf (DD)</b>	<b>1.00</b>	<b>50-57</b>
<b>Başarısız( FF)</b>	<b>0.00</b>	<b>49 ve aşağı</b>
<b>Devamsız (DZ )</b>	<b>0.00 --</b>	<b>--</b>
<b>Sınava Girmeme (GR)</b>	<b>0.00 --</b>	<b>--</b>
<b>Yeterli (YT)</b>	<b>--</b>	<b>--</b>
<b>Yetersiz (YZ)</b>	<b>--</b>	<b>--</b>
<b>Muaf (MU)</b>	<b>--</b>	<b>--</b>

Tabloya göre notların yorumlanması ise şu şekilde yapılmaktadır:

- Bir dersten AA, BA, BB, CB ve CC notlarından herhangi birini alan öğrenci o dersi başarmış sayılır.
- Bir dersten alınan DC ve DD notları bu dersin "şartlı" olarak başarılı olduğunu gösterir.
- FF notu öğrencinin dersten başarısız olduğunu belirtir.
- Derse devam yükümlülüklerini yerine getirmeyen veya ders uygulamalarına ilişkin koşulları sağlamadığı için başarısız olan öğrencilere DZ notu verilir. DZ notu, not ortalamaları hesabında FF notu işlemi görür.
- Dönem sonu sınavına girmeyen öğrenciye, dönem içi çalışmalarına bakılmaksızın GR notu verilir. GR notu, not ortalamaları hesabında FF notu işlemi görür.

- MU notu, SAÜ lisans ve ön lisans bölüm/programlarına yatay/dikey geçiş yapan veya daha önce devam ettiği herhangi bir Yükseköğretim Kurumunda başarılı olduğu derslerin eşdeğerlerine ve muafiyet sınavı uygulanan derslerden başarılı olan öğrencilere verilir. MU notu, not ortalamaları hesabına katılmaz.
- YT notu, ortalamaya girmeyen ve başarılımış derslerin başarı notu olarak verilir.
- YZ notu, ortalamaya girmeyen ve başarılmamış derslerin başarı notu olarak verilir.

Bir döneme ait not ortalaması ise:

$$DNO = \sum_{k=0}^n \frac{\text{Derse ait başarı notu} \times \text{Dersin AKTS si}}{\text{Dönemde alınan toplam AKTS}}$$

(n= ders sayısı)

formülünden elde edilen değer olarak belirlenmektedir.

Bu durumda genel not ortalaması:

$$GNO = \sum_{k=1}^i \sum_{k=1}^n \frac{\text{Derse ait başarı notu} \times \text{Dersin AKTS si}}{\text{Dönemde alınan toplam AKTS}}$$

(i= dönem sayısı n= ders sayısı)

Bu formüllerde yer alan i dönem sayısını, n ise bir dönemde alınan ders sayısını göstermektedir. Hesaplanan Genel Not Ortalamasının ilgili programdan mezun olabilmek için son dönem itibarıyla 2.00 ve üzerinde olması gerekmektedir. (Sakarya Üniversitesi, 2011: s. 93-97)

Bu noktada mevcut çalışma, ilk yıl öğrencilerine ait geleneksel ve geleneksel olmayan değişkenleri kullanarak teorik bir model oluşturmayı ve bu modele bağlı olarak öğrencilerin akademik performanslarına etki eden faktörleri belirlemeyi ve akademik performanslarının tahmin edilmesini sağlayacak bir metodolojiyi ortaya koyacaktır. Böylece, akademik performans yönüyle risk grubunda yer alan öğrencilerin önceden belirlenmesini sağlayarak, risk faktörlerine karşı proaktif önlemler alınmasını



sağlayacak ve bu faktörler ile ilgili istendiği anda raporlama sağlayacak bir karar destek sistemi tasarlanması amaçlanmaktadır.

## 2.2 Metodoloji

Çalışmada, farklı çalışma alanları için standart bir veri madenciliği sürecinin tanımlandığı popüler bir yöntem olan CRISP-DM (Cross Industry Standard Process for Data Mining) den yararlanılmıştır.(Mishra vd. 2014; Delen, 2010; Catley vd., 2009; Venter vd. , 2007). “Veri madenciliği için sektörler arası standart bir süreç” olarak Türkçe’ye çevrilebilecek olan CRISP-DM, genelden özele 4 seviyeden oluşan hiyerarşik bir süreç modeli olarak tanımlanabilir.

- Aşamalar – Veri madenciliği sürecinin her bir bölümüne verilen addır.
- Genel görev- Her bir aşama pek çok ikincil seviye genel görevlerden oluşmaktadır. “Genel” denmesinin nedeni tüm veri madenciliği durumlarını göz önünde bulunduracak kadar kapsayıcı olmasıdır. Bu görevler mümkün olduğunca eksiksiz (tüm veri madenciliği sürecinin ve uygulamaların kapsanması) ve stabil (yeni modelleme teknikleri için de geçerli olma durumu) olmalıdır.
- Özelleştirilmiş görev- özel durumlarda genel görevlerde yer alan eylemlerin nasıl uygulanması gerektiğini belirtmektedir. Örneğin, “model kur” genel bir görev iken “yanıt modeli kur” özelleştirilmiş bir görev olarak düşünülebilir.
- Süreç örnekleri - gerçek veri madenciliği çalışmalarının eylemleri, kararları ve sonuçlarıdır. (Wirth&Hipp, 2000: s. 31).

Buna göre CRISP – DM ‘nin aşamaları şu şekilde belirtilmiştir:

- 1- **Problem tanımlanması ve anlaşılması:** Bu başlangıç aşaması, problem perspektifinden, projenin amaçlarının ve gereksinimlerinin anlaşılmasına odaklanır ve bu bilgiyi bir veri madenciliği problem tanımına dönüştürülür ve bu projeyi gerçekleştirmek üzere öncül proje planı hazırlanır.
- 2- **Veriyi anlama:** Öncül veri toplama ile başlar ve veriyi daha iyi anlamak, veri kalitesi problemlerini belirlemek, veri hakkında görüş sahibi olmak veya gizli bilgilere yönelik hipotezleri oluşturmak üzere ilginç (önemli) alt kümeleri tespit etmek şeklinde devam eder.

- 3- **Veri hazırlama:** Veri hazırlama aşaması, modelleme araçlarında kullanılacak nihai verisetini oluşturmak adına yapılan tüm faaliyetlerden (özellik seçimi, veri temizleme, yeni özelliklerin oluşturulması, verilerin dönüştürülmesi) oluşur.
- 4- **Modelleme:** Verisetine uygun çeşitli modelleme teknikleri seçilir ve uygulanır ve parametreleri optimal değerlere uygun olacak şekilde ayarlanır.
- 5- **Değerlendirme:** Denenen modellerin yayılımı sağlanmadan önce kapsamlı analizini içerir ve modelin amacı ne ölçüde karşılayabildiği belirlenir.
- 6- **Yayılım:** Kazanılan bilgi kullanıcılar tarafından rahatça kullanılabilmelidir. Sonuçlar rapor şeklinde olabilir ya da tekrarlanan veri madenciliği süreçleri şeklinde uygulanabilir (Wirth&Hipp, 2000: s. 33-35).

Tablo 7

CRISP-DM görevleri ve girdilerine genel bir bakış (Wirth&Hipp, 2000: s. 34)

1-Problem Tanımlanması ve Anlaşılması	2-Veriyi Anlama	3-Veri hazırlama	4-Modelleme	5-Değerlendirme	6-Yayımlı
<p><b><u>Amacı belirle:</u></b> -Altyapı -Amaçlar -Başarı kriteri</p> <p><b><u>Durumu değerlendir</u></b> -Kaynakların durumu -Gereksinimler, Varsayımlar, Kısıtlar -Riskler ve ihtimaller -Terminoloji -Fayda ve Maliyetler</p> <p><b><u>Verimadenciliği hedeflerini belirle</u></b> -Veri madenciliği hedefleri -Veri madenciliği başarı kriterleri</p> <p><b><u>Proje planı oluşturun</u></b> -Proje planı -Araç ve tekniklerin öncül değerlendirilmesi</p>	<p><b><u>Öncül veri toplama</u></b> -Öncül veri toplama -Raporlama</p> <p><b><u>Veriyi tanımla</u></b> -Veri tanımlama raporu</p> <p><b><u>Veriyi keşfetme</u></b> -Veri keşif raporu</p> <p><b><u>Veri kalitesini onayla</u></b> -Veri kalite raporu</p>	<p>-Veriseti -Veriseti tanımı</p> <p><b><u>Veri seç</u></b> -Veriyi dahil etme kriteri</p> <p><b><u>Veri temizle</u></b> -Veri temizleme raporu</p> <p><b><u>Veri üret</u></b> -Türetilmiş değişkenler -Üretilmiş raporlar</p> <p><b><u>Veri entegrasyonu:</u></b> -Birleştirilmiş veri</p> <p><b><u>Veriyi formatla</u></b> -Yeniden formatlanmış veri</p>	<p><b><u>Modelleme tekniği seç</u></b> -Modelleme tekniği -Modelleme varsayımları</p> <p><b><u>Tasarımı test et</u></b> -Tasarım testi</p> <p><b><u>Model kur</u></b> -Parametre ayarlama -Modeller -Model tanımı</p> <p><b><u>Modeli değerlendir</u></b> -Revize edilmiş parametre ayarları</p>	<p><b><u>Sonuçları değerlendir</u></b> -Veriyi değerlendir -Madencilik sonuçları -Problem başarı kriteri -Onaylanan modeller</p> <p><b><u>Gözden geçir</u></b> -Sürecin gözden geçirilmesi</p> <p><b><u>Sonraki adımları belirle</u></b> -Olası eylemlerin listesi -Karar</p>	<p><b><u>Plan yayımlı</u></b> -Yayımlı planı</p> <p><b><u>Plan izleme ve sürdürme</u></b> -İzleme ve sürdürme planı</p> <p><b><u>Nihai raporu hazırla</u></b> -Nihai rapor -Nihai sunum</p> <p><b><u>Projeyi gözden geçir</u></b> -Deneyim -Dokümantasyon</p>

### 2.2.1 Problem Tanımlanması ve Anlaşılması

Türkiye’de 2014-2015 akademik yılı itibariyle yükseköğretim lisans programlarına yeni kayıt yaptırmış 1,382,571 öğrenci bulunmaktadır. (YÖK Bilgi Sistemi, 2015) Sayı bu kadar fazla iken halihazırda, yeni gelen öğrenciler için, Türkiye’de birkaç özel üniversite hariç hiç bir üniversitede ilk yıl destek programına rastlanmamıştır. Mevcut çalışmalara baktığımızda, yeni gelen öğrencilere yönelik destek programlarının oryantasyon eğitimlerinden ileri gidemediği görülmektedir. Karşılaştırma yapmak gerekirse, günümüzde A.B.D.’de bulunan üniversitelerin %95.4 ü ilk yıl öğrencilerine yönelik destek programları yürütmektedir (Jamelske, 2008: s. 375). Dolayısıyla Türkiye’de yeni gelen öğrencilere gereken ihtimam gösterilmemekte ve bu konudaki eksiklik göze çarpmaktadır.

Öğrencilerin akademik performansları ve devamlılıklarına etki eden faktörlerin belirlenmesi ve bu parametrelere dayalı olarak önceden tahmin edilebilmesi, risk grubunda yer alan öğrencilerin önceden tespit edilerek gerekli önlemler alınmasına yardımcı olmaktadır. Bu konuda yapılan çalışmalar genel anlamda veri temelli çalışmalar ve anket temelli çalışmalar olmak üzere 2 ana grupta toplanmaktadır (Delen, 2010: s. 499). Caison (2006) yaptığı çalışmasında, bu iki temel yaklaşımı akademik devamlılığın araştırılmasında kullanmıştır. Kurulan modellerde verilerden ve anketlerden elde edilen değişkenlerin tümü kullanılmıştır. Modeller, bir modelin lojistik regresyon analizi için veriye uygunluğunu test etme yöntemlerinden biri olan (Ellison, 2013) Hoshmer&Lemeshow testi ve pek çok alanda kullanımı olan ve ikame modeller arasında seçim yapmada kullanılan bir indeks olan (Everitt, 1998) Akaike Information Criterion (AIC) kullanılarak ayrı ayrı ele alınmıştır. Tüm değişkenlerin dahil edildiği modelin çok daha iyi sonuçlar verdiğini saptamıştır. Bu noktada geleneksel ve geleneksel olmayan değişkenlerin çalışmaya birlikte dahil edildiği “hibrit çalışma” kavramı ortaya çıkmıştır. Literatürde son yıllarda “hibrit çalışma” kapsamında değerlendirilebilecek pek çok çalışma mevcuttur (French vd. , 2005; Mishra vd. , 2014; Superby vd. ,2006; Furnham vd., 2003; Belanger vd. , 2007). Hibrit modeller öğrencinin akademik performansını ve devamlılık kararlarına temel teşkil eden süreci çok boyutlu ele aldığından ötürü daha isabetli tahmin oranlarına ulaşmaktadırlar. Örnek vermek

gerekirse; Mishra vd. (2014) öğrencilerin akademik başarı aralıklarını tahminlemeye yönelik çalışmasında öğrencinin empati yeteneği, karar verme kabiliyeti gibi öğrencinin algı, tutum ve kişilikleriyle ilgili geleneksel olmayan değişkenleri kullanarak çalışmasında %94.41 doğru tahminleme oranı yakalamışlardır. Benzer şekilde Ramaswami&Bhaskaran (2010) öğrencilerin başarı aralıklarını tahmin etmeye çalışmışlar ve modellerinde herhangi bir geleneksel olmayan değişkene yer vermemişlerdir. Bu çalışmada ise, genel doğru tahminleme oranı %44.69 seviyesinde kalmıştır. Bu durumda hibrit yaklaşımın her ne kadar tahmin oranları arasındaki farkın tamamından sorumlu olduğu düşünülse de, hibrit yaklaşımın avantajlarını ortaya koyması açısından önemlidir. Bu noktada Türkiye’de mevcut duruma bakılacak olursa, tahminlemeye yönelik anket temelli, veri temelli ya da hibrit yaklaşımlı herhangi bir bütüncül çalışmaya rastlanmamıştır.

Mevcut çalışmada öğrencilerin akademik performanslarına etki eden faktörleri ortaya koymak ve bu faktörlere bağlı olarak tahminlemede bulunmak üzere bütüncül bir çalışma yapılması öngörülmüştür. Bu suretle öğrencilerin Türkiye’de yükseköğretim aşamasında ilk yılda içerisinde buldukları sürecin daha iyi anlaşılabilmesi ve bulgulara dayanarak uygulanabilecek destek programlarına dönük önerilerde bulunulması amaçlanmaktadır.

### **2.2.2 Veriyi anlama**

Çalışmada süreci modelleyebilmek üzere kapsamlı literatür taraması sonucunda belirlenen veriler “Yeni Gelen Öğrenci Anketi” ve üniversiteye ait öğrenci veritabanı olmak üzere iki temel kaynaktan sağlanmıştır.

Mevcut çalışma; 2014-2015 akademik yılında İşletme Fakültesi’ndeki lisans programlarında ilk yıl öğrencisi olan 520 kişiye uygulanmıştır. Mayıs-2015’te uygulanan anketler içerisinden çeşitli nedenlerle (öğrenci numarasının yanlış beyan edilmesi veya hiç beyan edilmemesi, anketin rastgele doldurulması, katılımcının örneklem dışında olması vs.) eleme işleminin ardından kullanılabilir 445 adet anket elde edilmiştir. Çalışmaya katılan her bir öğrencinin beyanına dayalı verileri şu şekilde listenebilir.

- 1- Öğrenci Numarası:** Anketlerde genellikle dolduran kişinin gizliliği esastır. Ancak mevcut çalışmada, öğrencilerin akademik performanslarının takip edilmesi amacıyla öğrencilerden numaralarını beyan etmeleri istenmiştir. Buna benzer uygulamalara öğrencilerin akademik performanslarına yönelik tahminleme çalışmalarında sıklıkla karşılaşılmaktadır (Higher Education Research Institute, 2015; Calvo-Flores vd. , 2006; Caison, 2007). Çalışmada kişilerin sağladığı bilgilerin gizliliği esastır ve katılımcılara gerekli açıklamalar yapılmıştır.
- 2- Bölüm:** Akademik performansın kayıtlı olunan bölüme göre değişiklik gösterebileceği ile ilgili çalışmalar mevcuttur (Belanger vd. 2007). Örneğin; Jawitz&Scott (1997) Mühendislik Fakültesinde eğitim gören öğrencilerle yaptıkları çalışmalarında bölümler arasında performans boyutunda anlamlı farklılıklar elde gözlemlemişlerdir. Çalışma, İşletme Fakültesi'nde İşletme, Turizm İşletmeciliği, İnsan Kaynakları Yönetimi, Sağlık Yönetimi, Uluslararası Ticaret, Yönetim Bilişim Sistemleri ve Leeds-İşletme lisans programlarına kayıtlı ilk yıl öğrencilerine uygulanmıştır.
- 3- Hazırlık durumu:** Çalışmanın uygulandığı üniversitede, hazırlık okumak öğrencilerin tercihlerine bırakılmıştır. Tinto'nun (1975; 1988) akademik ve sosyal entegrasyonla ilgili ortaya koyduğu teorik modellere dayanarak hazırlık okumayı seçen öğrencilerin, kayıtlı oldukları bölümlerde lisans eğitimlerine başlamadan önce sosyal ve akademik ortama adaptasyon anlamında mesafe katettiği savunulabilir. Bu noktadan yola çıkarak, dezavantajlarının (akademik anlamda boşluk) yanında hazırlık okuma durumu, akademik performans üzerinde etkisi araştırılmak üzere çalışmaya tahminleyici değişken olarak dahil edilmiştir. Hazırlık okuyan öğrenciler "1", hazırlık okumayan öğrenciler ise "2" olarak gruplanmıştır.
- 4- Yaş:** Literatürde yer alan pek çok çalışmada yaş, akademik performansın tahminine yönelik tahminleyici değişkenlerden biri olarak ele alınmıştır(Jing, 2002; Bresfelean vd., 2008; Thammasiri vd., 2014; Delen, 2010; Fike&Fike, 2008; Kerkvliet&Nowell, 2005; Barker vd., 2004). Çalışmaların uygulandığı örnekleme göre yaş olgunluğun bir göstergesi olarak düşünülmüş ve akademik performans ile pozitif ilişkiler saptanmıştır (Houltram, 1996;

Salamonson&Andrew, 2006). Mevcut çalışmada, Yeni Gelen Öğrenci anketinde öğrencilerden yaşlarının beyan edilmesi istenmiş ve dağılıma uygun olarak 3 yaş grubu altında nominalleştirilmiştir. Buna göre her bir numara ilgili grubu temsil etmek üzere; 17-19 yaş arası öğrenciler “1”, 20-22 yaş arasındaki öğrenciler “2”, 23 yaş ve üzeri öğrenciler ise “3” şeklinde kodlanmıştır.

- 5- Cinsiyet:** Akademik performansın ve devamlılığın cinsiyete göre farklılık gösterdiği şeklindeki hipotezler pek çok çalışmada test edilmiş ve tahminleyici bir değişken olarak ele alınmıştır (Astin, 1997; Tinto, 1975; Delen, 2010; Byrne&Flood, 2008). Bu doğrultuda akademik performans ile cinsiyet arasında anlamlı ilişkilere ulaşılmış çalışmalara rastlanmıştır (Allen&Haniff, 1991; Ravenscroft&Buckless, 1992). Çalışmada; “1” kadınları, “2” ise erkekleri temsil etmektedir.
- 6- Kardeş sayısı:** Bu değişken ailenin sosyoekonomik durumuna ve kaynaklarına dair ipuçları sunması yönüyle önemlidir. Öğrenciye ailesi tarafından sağlanan kaynak (maddi ve manevi) miktarıyla okuldaki başarısının pozitif bir ilişki içerisinde olduğu belirtilmiştir (Teachman, 1987: s. 551). Aile tarafından sağlanan kaynakların kardeş sayısı ile orantılı olacağı ve buna bağlı olarak akademik performansın değişebileceği savunulabilir (Şen vd., 2012).
- 7- Ailenin ikamet ettiği şehir:** Öğrencilerin lisans programlarında göstereceği akademik performans üzerinde ailenin ikamet ettiği tahminleyici bir değişken olabileceği düşünülmüştür (Mishra vd., 2014). Üniversite öncesi aileyle birlikte ikamet edilen şehirdeki olanakların (sınıflardaki ortalama öğrenci sayısı, laboratuvar olanakları vs.) eğitim hayatında gösterilen performans üzerinde etkili olduğu savunulabilir. Çalışmada şehirler arasındaki olanak farklarının o şehrin nüfusuna bağlı olarak değişeceği varsayılmıştır. Bu noktada öğrencilerin ailelerinin ikamet ettiği şehirler; Kırsal “1”, Kent “2” ve Metropol “3” olarak kodlanmıştır.
- 8- Anne –Baba eğitim durumları:** Literatürde ebeveynlerin eğitim düzeylerinin öğrencilerin akademik performansları üzerinde bir etkiye sahip olduğu görüşü oldukça yaygındır (Stevenson&Baker, 1987; Fike&Fike, 2008; Astone&McLanahan, 1991). Özellikle annelerin eğitim durumlarının öneminin vurgulandığı bu değişkenler, ailelerin çocuklarının eğitim sürecine olan

ilgilerinin bir ölçüsü olarak kabul edilebilmektedir. Bu durumda, üniversite mezunu ebeveynlerin çocuklarının daha yüksek akademik performans göstereceği öne sürülebilir. Mevcut çalışmada, Anne ve Baba eğitim durumları ordinal değişkenler olarak ele alınmıştır. İlköğretim “1” , Lise “2”, Üniversite “3”, Yüksek lisans “4” ve Doktora “5” olarak sayısallaştırılmıştır.

**9- Ailenin aylık ortalama geliri:** Ailenin aylık (ya da yıllık) ortalama gelirinin öğrencilerin akademik performansı ve devamlılığı üzerine etkisi tek boyutlu (Stampen&Cabrera, 1986; Hill, 2001) ya da bütüncül çalışmalarda (Fike&Fike, 2008; Kerkvliet&Nowell, 2005) araştırılmıştır. Bu çalışmalara göre yükseköğretim sürecinde maddi zorluklarla mücadele etmek zorunda kalmayan öğrencilerin akademik anlamda daha iyi performans gösterdiği saptanmıştır (Nandeshwar vd., 2011). Mevcut çalışmada ortalama gelir için alt sınır Türkiye’de asgari ücret olarak belirlenmiş ve ailenin ortalama aylık geliri şu şekilde ordinal olarak tanımlanmıştır: 1000 TL ve altı “1”, 1001 TL ve 2500 TL arası “2”, 2501-4000 TL arası “3” , 4001 TL ve 5500 TL arası “4” ve 5500 TL ve üzeri aylık ortalama gelir “5”

**10- Ekstra gelir:** Ebeveynlerinin öğrencilere tahsis ettiği maddi kaynaklar dışında elde edilen gelirin akademik performans üzerine olumlu etki ettiği varsayılmaktadır (Kerkvliet&Nowell, 2005). Ancak eğer öğrenci, part time çalışarak fazladan bir gelir elde ediyorsa, iş yerindeki sorumlulukları ve yoğunluğunun akademik performans üzerinde değişken etkilere sahip olabileceği de savunulmaktadır.. Örneğin; Ehrenberg&Sherman (1987) öğrencinin kayıtlı olduğu üniversite bünyesinde istihdam edilmesinin akademik performans üzerine pozitif, okul dışında bir yerde istihdam edilmesinin ise negatif etkilere sahip olduğunu belirtmişlerdir. Kerkvliet&Nowell (2005) ise iki farklı tipte üniversitede gerçekleştirdikleri çalışmalarında, araştırma üniversitesinde akademik devamlılık üzerinde part-time çalışmanın, sadece lisans düzeyinde eğitim veren üniversitede ise bursların devamlılık üzerinde olumlu anlamda etkiye sahip olduğunu ortaya koymuşlardır. Mevcut çalışmada sahip olunan ekstra gelir kaynaklarındaki düzensizlik dolayısıyla net rakamlar elde edilememiş ve veriler analizde kullanılmak üzere nominalleştirilmiştir. Buna göre; Burs “1”, Kredi “2”, Part Time çalışma “3”, Herhangi bir gelire



sahip olmama “4”, Birden fazla gelir kaynağına sahip olma ise “5” olarak sayısal değerler almıştır.

**11- Maddi kaygı:** Bu çalışmada, öğrencinin içerisinde bulunduğu finansal durumun bir sonucu olarak oluşan kaygı durumunun akademik performans üzerine etkisi araştırılmıştır. Bunun için ise, her yıl A.B.D.’de yeni gelen öğrencilere uygulanan ve bu yıl 50.si yapılan CIRP Freshman Survey 2015’in ilgili kısmından yararlanılmıştır (Higher Education Research Institute, 2015: s. 2). Buna göre kaygı durumu ile ilgili: “*Üniversite masraflarını (harç ücretleri, kira, yurt ücreti vs.) karşılamada maddi anlamda kaygılarınız var mı?*” sorusuna verilen yanıtlara göre öğrencinin finansal kaygı seviyesine yönelik bir fikir edinilmeye çalışılmıştır. Cevaplar ise aslına uygun olarak çevrilerek 3 farklı seçenek şeklinde sunulmuş ve katılımcılardan tek bir şıkkı işaretlemeleri istenmiştir. Bu seçenekler:

- Hayır; kaygım yok. (Yeterli maddi güce sahibim.)
- Bazen; kaygılanıyorum (Ama genel olarak imkanlarım yeterlidir.)
- Evet; kaygılıyım. (Üniversiteyi bitirebilmek için yeterli maddi güce sahip olduğumdan emin değilim.)

Bu seçenekler ise sırasına göre “1”, “2” ve “3” şeklinde kodlanarak analize dahil edilmiştir.

**12- Kalınan yer:** Konuyla ilgili yapılan çalışmalarda, öğrencinin kayıtlı olduğu üniversitenin ailesinin yaşadığı şehire uzaklığı ve öğrencinin kampüse ne kadar uzakta ikamet ettiği araştırılmış ve bu mesafelerin entegrasyon süreciyle ters orantılı olarak değişeceği öne sürülmüştür. Örneğin; kampüs içerisinde yaşayan öğrencilerin kampüs dışarısında yaşayanlara göre daha kısa sürede akademik ve sosyal ortama adapte olabildiği böylece daha başarılı olduğu ile ilgili bulgular mevcuttur (Astin, 1997; Jamelske, 2008; Hosch, 2008). Bu durum Tinto’nun (1975) ortaya koyduğu teorik modelle de uyumludur. Benzer olarak, üniversiteye ve sosyal ortama adaptasyon sürecinde aile desteğinin daha rahat alınabilmesi bakımından, öğrencinin ailesinin yanında ya da yakınında olması tahminleyici bir değişken olarak düşünülebilir. Tüm bunlara ek olarak kalınan yerin öğrencinin finansal durumu ile ilgili ipuçları sunabileceği savunulabilir.

Mevcut çalışmada; kalınan yer değişkeni “1- Devlet yurdu”, “2-Özel yurt”, “3- Öğrenci evi” ve “4-Aile/Akraba yanı ” olarak nominal bir değişken olarak belirlenmiştir.

**13- Sigara kullanımı:** Zararlı alışkanlıkların kişinin sağlığını olumsuz anlamda etkilediği bilinmektedir. Bu noktada, fiziksel sağlık durumuna ve zararlı alışkanlıklara bağlı olarak gösterilen akademik performans ilgi çeken konulardan biridir (DeBerard vd., 2004). İlginç bir çalışmada ise, sigara içen öğrencilerin içmeyen öğrencilere göre daha başarılı bir performans sergiledikleri sonucuna ulaşılmıştır (Warburton vd., 1984). Çalışmamızda, sigara içenler “1” ve içmeyenler ise “2” olarak kodlanmıştır.

**14- Lise mezuniyet not ortalaması:** Öğrencilerin lisede göstermiş olduğu akademik performans ile üniversitede gösterdiği performans arasında güçlü bir ilişki olduğu pek çok çalışmada vurgulanmıştır (DeBerard, 2004; Byrne&Flood, 2008; Jamelske, 2008; Gazza&Hunker, 2014; Nandeshwar vd., 2011; French vd., 2005). Bu durum, öğrencilerin eğitimin hangi seviyesinde olursa olsun akademik başarı ile özdeşleştirilebilecek bazı kişisel özelliklere (sorumluluk, özgüven, öz yeterlilik vs.) sahip olunması ve başarı arasındaki ilişkinin olumlu yönde olduğu şeklinde yorumlanabilir. Örneklemimizde katılımcıların lise not ortalamalarının ortalaması 2,9 olarak bulunmuştur.

**15- Lise türü:** Mezun olunan lisenin türü ve kalitesi, öğrencinin sahip olduğu lise not ortalamasının yorumlanmasına olanak sağlayacaktır. Mezun olunan liseye göre lisans programlarında başarı durumu, üniversite tercih sürecinde gözardı edilmemesi gereken hususlardan biridir. Bu çalışmada, “1- Düz lise”, “2- Anadolu lisesi”, “3-Çok programlı lise”, “4- Meslek lisesi”, “5-Fen lisesi” ve “6- Diğer lise türleri” şeklinde bir kodlama uygun görülmüştür.

**16- Yükseköğretime geçiş sınavı başarı durumu:** Yükseköğretime geçiş için uygulanan sınavların öğrencinin yerleşeceği lisans programındaki başarısını etkilediği yaygın olarak savunulmaktadır (Hosch, 2008; Ting, 2000; Jencks&Phillips, 1998). Türkiye’de, Kuran (1988) yaptığı çalışmasında yeni gelen öğrencilerin ÖSS puanları ile ilk dönem not ortalamaları arasındaki ilişkiyi incelemiş ve anlamlı bir sonuca ulaşamamıştır. Bu değişkenin modele katılması, akademik performansın tahminine yönelik olduğu kadar ÖSYM’nin her yıl

düzenlediği üniversiteye giriş sınavlarının yordayıcılığının irdelenmesi yönüyle de önemlidir. Mevcut çalışmada, Üniversitelerin her yıl artan kontenjanlarına ya da öğrencilerin sınavlarda göstermiş oldukları performansa bağlı olarak dinamik bir yapı sergileyen LYS puanları yerine, analizlerde daha güvenilir sonuçlar verecek olan nispeten çok daha az dinamik olan LYS başarı sıralamaları kullanılmıştır. Aralık üzerinden tanımlanan değişken şu şekilde sayısallaştırılmıştır: “1- İlk 50000” “2- 50001 -100000 arası” “3- 100001 - 150000 arası” “4-150001- 200000 arası ” ve “5- 200001 ve üzeri”

**17- Önem sırası:** CIRP Freshmen Survey’ nin ilgili kısmından esinlenerek Türkçeye çevrilen bu değişken, öğrencilerin üniversiteye gelme motivasyonlarına yönelik bir fikir vermesi açısından modele dahil edilmiştir. Uygulanan ankette yer alan bu kısımda, öğrencilerden belirtilen ifadelerden kendileri için en önemli olanının karşısına “1” yazılması önemli olan seçeneğin karşısına ise “2” yazılması istenmiştir. 6 seçenekli kısım uygulanan ankette şu şekilde yer almaktadır.

**Üniversiteye gelmenizden aşağıda belirtilen ifadelerden; en önemli olan iki tanesini belirtiniz.**

(En önemli ifade için ilgili alana “1”, diğer önemli ifade için ilgili alana “2” yazınız.)

**Tablo 8**

**Öğrencilerin üniversiteye gelme nedenleri (Higher Education Research Institute, 2015)**

	Önem sırası
<b>Daha iyi bir işe sahip olmak</b>	
<b>Genel bir eğitime sahip olmak ve fikir edinme</b>	
<b>Daha kültürlü bir insan olmak</b>	
<b>Daha fazla para kazanabilmek</b>	
<b>İlgi alanlarım hakkında daha fazla bilgiye sahip olmak</b>	
<b>Hedeflediğim kariyer üzerine hazırlık yapmak</b>	

Bu seçenekler 1 den 6 ya kadar sıralanmış ve Onem\_sirası1 ve Onem\_Sirası2 değişkenleri altında toplanmıştır. Buna göre, seçeneklere göre, en önemli ifade için “1” ifadesinin yazıldığı alanlar şu şekilde dağılmıştır:

**Tablo 9**

**Öğrencilerin üniversiteye gelme nedenlerine ait frekanslar**

	Kişi sayısı
<b>Daha iyi bir işe sahip olmak</b>	<b>160</b>
<b>Genel bir eğitime sahip olmak ve fikir edinme</b>	<b>44</b>
<b>Daha kültürlü bir insan olmak</b>	<b>62</b>
<b>Daha fazla para kazanabilmek</b>	<b>65</b>
<b>İlgi alanlarım hakkında daha fazla bilgiye sahip olmak</b>	<b>21</b>
<b>Hedeflediğim kariyer üzerine hazırlık yapmak</b>	<b>93</b>

**18- Hedeflenen eğitim seviyesi:** Tinto (1975) akademik başarı ve devamlılık üzerine ortaya koyduğu modelinde kişilerin eğitim planlarına “hedefe bağlılık” başlığı altında yer vermiştir. Bu noktadan yola çıkarak, akademik anlamda yüksek seviyeleri hedefleyen kişilerin akademik motivasyonlarının dolayısıyla performanslarının yüksek olacağı savunulabilir (Zimmerman vd., 1992). Ankette yeni gelen öğrencilerden hedefledikleri en üst eğitim seviyelerini belirtmeleri istenmiştir. Verilen cevaplar ise analizde kullanılmak üzere şu şekilde kodlanmıştır: “1- Lisans” “2- Yüksek lisans” ve “3- Doktora ”

**19- İlk dönem not ortalaması:** Öğrencinin geçmişte ortaya koymuş olduğu akademik kabiliyetin gelecekte göstereceği performansı üzerinde etkili olduğu savunulabilir (Slim vd. , 2014). Bu nedenle öğrencinin ilk dönem sonunda elde etmiş olduğu not ortalaması sonraki dönemlerdeki not ortalamalarının tahmin edilmesinde yaygın olarak kullanılan bir değişken olarak karşımıza çıkmaktadır (Slim vd., 2014a; Mishra vd., 2014; Caison, 2007; Belch vd., 2001). Mevcut çalışmada, İşletme Fakültesi’ndeki lisans programlarında ilk yıllarını geçiren

445 öğrencinin ilk dönem not ortalamalarının ortalaması 2,022 olarak bulunmuştur.

**20- Giriş dersi başarı notu:** Genellikle, lisans programına ait 8 dönemde üzerinde durulacak olan konuların özetlendiği bir içeriğe sahip olan bölüme giriş dersinden alınan not, öğrencinin kayıtlı olduğu lisans programı ile ilgi alanlarının ne denli uyumlu olduğunun bir göstergesi olarak kabul edilebilir. Bir başka deyişle, bölüme ait giriş derslerinden başarılı olan öğrencilerin, bu başarısını genele yayabileceği düşünülmektedir. Yapılan çalışmalarda benzer şekilde, bir muhasebe lisans programında Finansal Muhasebe dersinden geçme notu (Byrne&Flood, 2008), temel derslerden (Temel Matematik ve Dil) alınan geçme notları (Fike&Fike, 2008) bağımsız değişkenler olarak düşünülmüştür. Çalışmada, katılımcılardan kendi bölümlerine ait giriş derslerinden alınan notları beyan etmeleri istenmiştir. Analizde kullanabilmek üzere alınan notlar ise şu şekilde numaralandırılmıştır: “0- Giriş dersi almadı”, “1-AA”, “2-BA”, “3- BB”, “4-CB”, “5-CC”- “6-DC”, “7-DD”, “8-FF”

**21- Devamsızlık:** Tinto (1999: s. 6) ya göre, için sınıf, çoğu öğrencinin öğretim üyeleri ve diğer öğrencilerle etkileşim sağlayarak sosyal ve akademik ortama ayak uydurabileceği yegâne alandır. Dolayısıyla derslere katılımı üst seviyede olan öğrencilerin akademik başarıya olumlu katkıda bulunabileceği düşünülmüştür. Bu doğrultuda devamsızlık oranının, akademik başarı ve devamlılık üzerine etkileri pek çok araştırmaya konu olmuştur (Bresfelean vd. ,2008; Superby vd. , 2006; Ullah&Wilson, 2007; Altinkurt, 2008). Mevcut çalışmamızda öğrencilerin 2014- 2015 akademik yılı güz döneminde yapmış oldukları tahmini devamsızlıklarını beyan etmeleri istenmiştir. Çalışmada seçenekler şu şekilde kodlanmıştır: “1: 0-2 hafta” “2: 2-4 hafta” “3: 4-6 hafta” “4: 4- 6 hafta ve üzeri”

**22- Genel özyeterlik:** Özyeterlik, bireyin bir hedefe ulaşmak için yapması gereken bir görevi gerçekleştirmede algılanan kabiliyetini ifade etmektedir (Bandura, 1997). Araştırmalar özyeterliğin, işe yaklaşımını (Saks, 1995), meslek eğitimini (Martocchio&Judge, 1997) ve iş performansını (Stajkovic&Luthans, 1998) tahmin edebildiğini göstermiştir (akt. Chen vd., 2001: s. 62). Bu yönleriyle özyeterlik akademik performans ve devamlılık üzerine yapılan araştırmalarda

tahminleyici bir deęişken olarak kullanılmıřtır (Multon vd., 1991; Pintrich&De Groot, 1990; Choi, 2005). alıřmada, katılımcıların kendi yeteneklerine dđnđk tutumları ve algılarının ölçülmesinde Chen vd. (2001) tarafından revize edilen tek boyutlu ve 8 ifadeli Genel Özyeterlik Ölçeęi Türke'ye evrilerek kullanılmıřtır. Kullanılan ölçek tabloda sunulmuřtur.

**Tablo 10**

**evrilmiř ve revize edilmiř Genel Özyeterlik Öleęi (Chen, 2001)**

	Tamamen katılıyorum	Kısmen Katılıyorum	Ne katılıyorum ne katılmıyorum	Kısmen katılmıyorum	Tamamen katılmıyorum
Kendim için koymuř olduęum hedeflerin çoęunu başaracaęım.	5	4	3	2	1
Zor iřlerle karřılařtıęımda, o iři tamamlayacaęımdan řüphem olmaz.	5	4	3	2	1
Genelde, istedięim sonuları elde edeceęimi düşünürüm.	5	4	3	2	1
Kafama koyduęum iřte başarılı olacaęıma inanıyorum.	5	4	3	2	1
Pek çok zorluęun üstesinden başarılı bir şekilde geleceęim.	5	4	3	2	1
Pek çok deęiřik iřte etkin bir performans göstereceęimden eminim.	5	4	3	2	1
Dięer insanlarla karřılařtırıldıęında; verilen çoęu iři gayet iyi yapıyorum.	5	4	3	2	1
Herřeyin zorlařtıęı zamanlarda bile gayet iyi bir performans sergileyebilirim.	5	4	3	2	1

Genel özyeterlik ölçeğinin kullanılabilir olduğuna karar verilen 472 anketten elde edilen İçsel tutarlılık ölçüsü, bir başka deyişle ölçeğin güvenilirlik ölçütü olan Cronbach  $\alpha$  değeri IBM SPSS Statistics 20 kullanılarak 0,897 olarak bulunmuştur. Çoğu sosyal araştırmalarda bu değer 0,7'den büyük olması ölçeğin “Kabul edilebilir” seviyede olduğunu göstermektedir (UCLA : Statistical Consulting Group, 2007). Dolayısıyla kullanılan ölçeğin gayet güvenilir olduğu söylenebilir.

**Güvenilirlik İstatistiği**

Cronbach's Alpha	İfade sayısı N
,897	8

**Şekil 13: Genel Özyeterlik Ölçeği Cronbach  $\alpha$  değeri**

**23- Akademik özyeterlik:** Akademik özyeterlik, bireyin, belirli seviyelerde kendisine verilen akademik görevleri başarıyla yerine getirebileceğine dair sahip olduğu kanaatlerdir (Bong&Skaalvik, 2003: s. 6). Kişinin bu yöndeki algısı ve akademik performansı arasındaki ilişki ise yaygın olarak incelenmiştir (Chemers vd. , 2001; Lent vd. , 1986; Gore, 2006). Çalışmamızda, katılımcıların akademik özyeterliklerine dair algıları hakkında fikir sahibi olabilmek üzere, Jerusalem&Schwarzer (1981) tarafından geliştirilen ve Yılmaz vd. (2007) tarafından Türkçe'ye uyarlanan tek boyutlu ve 7 ifadeli Akademik Özyeterlik Ölçeği kullanılmıştır.

**Tablo 11**

**Çalışmada kullanılan Akademik Özyeterlik Ölçeği (Yılmaz vd. , 2007)**

Akademik olarak;	Tamamen katılıyorum	Kısmen Katılıyorum	Ne katılıyorum ne katılmıyorum	Kısmen katılmıyorum	Tamamen katılmıyorum
Herşeyin zorlaştığı zamanlarda bile gayet iyi bir performans sergileyebilirim.	5	4	3	2	1
Yeterince hazırlandığım zaman sınavlarda daima yüksek başarı elde ederim.	5	4	3	2	1
İyi not almak için ne yapmam gerektiğini çok iyi biliyorum.	5	4	3	2	1
Bir yazılı sınav çok zor olsa bile, onu başaracağımı biliyorum.	5	4	3	2	1
Başarısız olacağım herhangi bir sınav düşünmüyorum	5	4	3	2	1
Sınav ortamlarında rahat bir tavır sergilerim, çünkü zekama güveniyorum	5	4	3	2	1
Sınavlara hazırlanırken öğrenmem gereken konularla nasıl başa çıkmam gerektiğini genellikle bilirim.	5	4	3	2	1

Genel Özyeterlik Ölçeği kullanılabilir 472 anketten elde edilen içsel tutarlılık ölçüsü Cronbach  $\alpha$  değeri 0,809 olarak bulunmuştur. Orijinal çalışmada bu değer 0,79 olarak belirtilmiştir (Yılmaz vd. , 2007: s. 257).

**Güvenilirlik İstatistiği**

Cronbach's Alpha	İfade sayısı N
,809	7

**Şekil 14: : Akademik Özyeterlik Ölçeği Cronbach  $\alpha$  değeri**

**24- Kurumsal Entegrasyon Ölçeği (KEÖ) :** Akademik başarı ve devamlılık süreci üzerine yaptığı kavramsal çalışmasında Tinto (1975) öğrencinin içerisinde



bulunduğu akademik ve sosyal ortama adaptasyon durumunun akademik devamlılık ve başarı üzerinde etkili olduğunu öne sürmüştür.

Bu teorik modelden yola çıkan Pascarella&Terenzini (1980) ise öğrencilerin adaptasyon derecelerini ölçmek üzere ölçek geliştirme çalışmalarında bulunmuşlardır. Araştırmacılar 773 ilk yıl öğrencisiyle gerçekleştirdikleri boylamsal çalışmalarında, öğrencilerin akademik ve sosyal entegrasyonları üzerinde etkili olan 5 boyut elde etmişlerdir. Buna göre öğrencilerin adaptasyon sürecinde etkili olan özellikler ve örnek ifadeler şu şekilde belirtilmiştir.

**a. Arkadaş- Grup Etkileşimi ( $\alpha =0,84$ )**

- Diğer öğrencilerle yakın arkadaşlık bağları kurdum.
- Karşı cins ile iletişim kurmakta zorlanmıyorum.

**b. Öğretim Üyeleriyle Etkileşim ( $\alpha =0,83$ )**

- Öğretim üyeleriyle tanışma ve etkileşim halinde olma fırsatlarından memnunum.
- En az bir öğretim üyesiyle yakın ve kişisel bir ilişki kurdum.

**c. Öğretim Üyelerinin Öğrenci Gelişimi ve Öğretme ile İlgisi ( $\alpha =0,82$ )**

- İrtibat halinde olduğum pek çok öğretim üyesi, öğrenciler için önemli ve ilgi uyandıran meseleleri ders dışında tartışmak üzere zaman ayırmaktan çekinmezler.
- Şimdiye kadar irtibat kurduğum öğretim üyeleri gerçekten öğrencilerle ilgileniyorlar.

**d. Akademik ve Düşünsel Gelişim ( $\alpha =0,74$ )**

- Üniversiteye başlamadan öncesine oranla kültürel etkinliklere katılmaya daha istekliyim.
- Derslerim bana yeni bakış açıları kazandırıyor.

**e. Üniversiteye ve Hedeflere Bağlılık ( $\alpha =0,71$ )**

- Bu üniversiteden mezun olmak benim için önemlidir.
- Bu üniversiteye gelmek ile doğru bir karar verdiğimden eminim.

French&Oakes (2004) KEÖ'nün psikometrik özelliklerini, ifade analizi, alt ölçekler arasındaki korelasyonlar, iç tutarlılık oranı ve önceden oluşturulan bir model aracılığıyla gözlenen değişkenlerden yola çıkarak gizil değişken (faktör) oluşturmaya yönelik bir yöntem (Aytaç&Öngen, 2012: s. 16) olan, Doğrulayıcı Faktör Analizi ile değerlendirmişlerdir. Bu çalışma sonucunda iyileştirilmiş iç tutarlılık oranı ve ölçekler arası korelasyon katsayıları elde edilmiştir. Mevcut çalışmada, 34 ifadeli bu yenilenmiş ölçekten yararlanılmıştır.

Yeni gelen öğrencilerin akademik ve sosyal ortama entegrasyonlarını ölçmede kullanılan bu ölçek 5 alt ölçekten oluşmaktadır. Katılımcıların bu alt ölçeklerin her birindeki algı ve tutumların mevcut çalışmada ayrı değişkenler olarak kullanılması düşünülmüştür. Türkçe'ye çevrilen bu ölçeğin kültüre ve sosyolojik yapıya bağlı olarak boyutlarının değişebileceği göz önünde bulundurularak çevrilen ölçeğe faktör analizi uygulanmıştır.

**Aşama 1-Çeviri:** Çevrilen ölçeğin güvenilir, geçerli ve aslına uygun olup olmadığını anlamak için uygulanan geleneksel yöntem ve seri yöntem mevcuttur (Hançer, 2003: s. 47). Çalışmada ilk aşamada geleneksel yönteme uygun olarak KEÖ İngilizceden Türkçe'ye çevrilmiş ve uzman ve ön katılımcı değerlendirmeleri göz önünde bulundurularak ifadelerdeki anlatım bozuklukları ve düşüklükleri giderilmiştir.

**Tablo 12**

**Türkçe'ye çevrilmiş Kurumsal Entegrasyon Ölçeği (French&Oakes, 2004)**

<b>Bu üniversitede şimdiye kadar:</b>	<b>Tamamen katılıyorum</b>	<b>Kısmen katılıyorum</b>	<b>Ne katılıyorum ne katılmıyorum</b>	<b>Kısmen katılmıyorum</b>	<b>Tamamen katılmıyorum</b>
Derslerim bana yeni bakış açıları kazandırıyor.	5	4	3	2	1
Bu üniversiteden akademik anlamda memnunum.	5	4	3	2	1
Üniversiteye başlamadan öncesine oranla kültürel etkinliklere katılmaya daha istekliyim.	5	4	3	2	1
Düşünsel gelişim seviyemden memnuniyet duyuyorum.	5	4	3	2	1
İstenen okuma ödevlerine ek olarak, derslerde tavsiye edilen pek çok kitap okuyorum.	5	4	3	2	1
Üniversiteye başlamamdan bu yana fikirler ve düşünsel meselelere olan ilgim arttı.	5	4	3	2	1
Gelecekte ne yapmak istediğim hakkında fikir sahibiyim.	5	4	3	2	1
İyi notlar almak benim için önemli.	5	4	3	2	1
Bu yıl şimdiye dek edindiğim akademik deneyimim, düşünsel gelişimimi ve fikirlere olan ilgimi olumlu anlamda etkiliyor.	5	4	3	2	1
Akademik anlamda tahmin ettiğim performans seviyesini yakaladım.	5	4	3	2	1
Okuldan arkadaşlarımla aramdaki ilişkiler düşünsel gelişimimi ve fikirlere olan ilgimi olumlu anlamda etkiliyor.	5	4	3	2	1
Diğer öğrencilerle yakın arkadaşlık bağları kurdum.	5	4	3	2	1
Kurduğum arkadaşlıklardan memnunum.	5	4	3	2	1

**Tablo 12- devamı**

Diğer öğrencilerle aramdaki kişisel ilişkiler kişisel gelişimimi, değer yargılarımı ve yaklaşımlarımı olumlu anlamda etkiliyor.	5	4	3	2	1
Diğer öğrencilerle tanışmak ve arkadaş olmak benim için çok kolay.	5	4	3	2	1
Karşı cins ile iletişim kurmakta zorlanmıyorum.	5	4	3	2	1
Tanıdığım çoğu öğrenci arkadaşım, kişisel bir problemim olduğunda dinlemeye ve yardım etmeye hazırdır.	5	4	3	2	1
Bu üniversitede çoğu öğrenciyle benzer değer yargıları ve yaklaşımları paylaşıyorum.	5	4	3	2	1
Bu üniversitede ders dışı organizasyonlara katılma fırsatlarından memnunuz.	5	4	3	2	1
Yaşadığım yerde mutluyum.	5	4	3	2	1
Öğretim üyeleriyle tanışma ve etkileşim halinde olma fırsatlarından memnunuz.	5	4	3	2	1
İrtibat halinde olduğum pek çok öğretim üyesi, öğrenciler için önemli ve ilgi uyandıran meseleleri ders dışında tartışmak üzere zaman ayırmaktan çekinmezler.	5	4	3	2	1
En az bir öğretim üyesiyle yakın ve kişisel bir ilişki kurdum.	5	4	3	2	1
Öğretim üyeleriyle ders dışı etkileşimlerim düşünsel gelişimimi ve fikirlere olan merakımı olumlu anlamda etkiliyor.	5	4	3	2	1
Öğretim üyeleriyle ders dışı etkileşimlerim kişisel gelişimimi, değer yargılarımı ve yaklaşımlarımı olumlu anlamda etkiliyor.	5	4	3	2	1
Öğretim üyeleriyle ders dışı etkileşimlerim kariyer hedeflerimi ve isteğimi olumlu anlamda etkiliyor.	5	4	3	2	1
Şimdiye kadar irtibata geçtiğim öğretim üyeleri gerçekten seçkinler ya da üstün öğretimcilerdir.	5	4	3	2	1
Şimdiye kadar irtibat kurduğum öğretim üyeleri gerçekten öğrencilerle ilgileniyorlar.	5	4	3	2	1
Şimdiye kadar irtibat kurduğum öğretim üyeleri gerçekten öğretmekle ilgileniyorlar	5	4	3	2	1
Şimdiye kadar irtibat kurduğum öğretim üyeleri ders dışı konularda da gelişmeme yardımcı olmak konusunda ilgilidirler.	5	4	3	2	1
Herhangi bir üniversiteden mezun olmak benim için önemlidir.	5	4	3	2	1
Sakarya Üniversitesi'nden mezun olmak benim için önemlidir.	5	4	3	2	1
Bu üniversiteye gelmek ile doğru bir karar verdiğimden eminim.	5	4	3	2	1

**Aşama 2-Pilot Çalışma:** 116 öğrencinin katılımıyla gerçekleştirilen çalışmada mevcut çeviride değiştirilmesi veya çıkarılması gereken herhangi bir ifadeye rastlanmamıştır. Çünkü ölçekte yer alan 34 ifadeden hiçbirisinin anti-image korelasyon matrisindeki değeri çalışma için kabul edilebilir düzey olan 0,45 in altında kalmamıştır. Verisetinin faktör analizi için uygunluk ve yeterliliğini ölçen ve örneklem yeterlilik göstergesi olarak kabul edilen KMO (Kaiser-Mayer-Olkin) değeri ise 0,808 olarak bulunmuştur. Literatürde bir verisetine faktör analizi yapılabilmesi için en düşük değer 0,6 olarak belirtilmiştir (Pallant, 2001). Ön araştırma sonucunda, ölçeğin anti-image korelasyon matrisi ve KMO örneklem yeterlilik değerleri göz önünde bulundurularak kabul edilebilir seviyede olduğuna ve örneklemin arttırılarak çalışmanın devam ettirilmesine karar verilmiştir.

**Aşama 3–Ana çalışma :** Başka bir dilden ölçek çevirme veya uyarlama çalışmalarında, ölçeğin çevrildiği ülkenin kültürel ve sosyolojik yapısı önemli bir rol oynamaktadır. Bu amaçla ölçeğin orijinal çalışmada (French&Oakes, 2004) belirtilen boyutlarla ne denli uyumlu sonuçlar ürettiğini belirlemek amacıyla öncelikle Faktör Analizi Yöntemi uygulanmıştır. Bu noktada Faktör Analizini tanımlamak gerekirse, başlıca amacı aralarında ilişki bulunduğu düşünülen çok sayıdaki değişken arasındaki ilişkilerin anlaşılmasını ve yorumlanmasını kolaylaştırmak için daha az sayıdaki temel boyuta indirgemek veya özetlemek olan bir grup çok değişkenli analiz tekniğine verilen genel bir isimdir (istatistikanaliz, 2015). Ön çalışmanın ardından “Yeni Gelen Öğrenci Anketi” toplamda 520 öğrenciye uygulanmış ve bu anketlerden KEÖ alanı kullanılabilir olan 472 tanesi analiz edilmiştir. Ölçekte yer alan ifade sayısının 5 katı kadar büyük bir örneklem büyüklüğü literatürde yeterli görülmektedir (Child, 2006). Buna göre 34 ifadeli mevcut çalışma için kabul edilebilir en az sayı 170’dir. 445 katılımcı sayısının bu koşulu sağladığı söylenebilir. IBM SPSS Statistics 20 paket programı kullanılarak yapılan faktör analizinde; verilerin orijinal çalışmada olduğu gibi 5 boyuta ayrılması istenmiştir ve temel bileşenler yöntemi uygulanmıştır. Temel Bileşenler Yöntemi (PCA: Principle Component Analysis) bir veri indirgeme tekniğidir ve keşifçi faktör analizi ile pek çok benzerlikler barındırır. Bu tekniğin amacı, büyük boyutlu değişken kümelerini orijinal değişkenlerdeki varyansın önemli bir kısmını açıklayacak şekilde daha küçük yapay değişken kümelerine indirgemektir. Bu yapay değişkenlere “Temel Bileşenler” adı verilmektedir (Laerd Statistics, 2015).

Döndürme yöntemi olarak ise, faktörlerin birbirinden bağımsız olduğu durumlarda kullanılan varimax tipi döndürme (Tonta, 2007) kullanılmıştır. Bu işlem sonucunda birincil faktör analizi sonuçları şekillerde gösterilmektedir:

Kaiser-Meyer-Olkin örneklem yeterlilik ölçüsü		,937
Yaklaşık Ki-kare		7725,149
Bartlett'in küresellik	df	561
	Sig.	,000

**Şekil 15: Birincil faktör analizi KMO örneklem yeterlilik sonuçları**

Bu tablo yorumlanacak olursa, örneklem yeterlilik göstergesi mevcut veriseti için 0,937 olarak belirlenmiştir. Mevcut bir verisetine faktör analizi uygulanması için literatürde minimum değerin 0,6 olarak belirtildiği (Pallant, 2001) göz önüne alındığında örneklemin faktör analizine gayet elverişli olduğu savunulabilir.

#### **AÇIKLANAN TOPLAM VARYANS**

<b>Bileşen</b>	<b>Başlangıç Özdeğerleri</b>	<b>Yüklerin karesinin çıkarım toplamları</b>	<b>Yüklerin karelerinin dönel toplamları</b>
	% Kümülatif	%Kümülatif	%Kümülatif
1	32,89	32,89	18,22
2	40,92	40,92	30,08
3	46,113	46,11	40,44
4	50,359	50,35	49,52
5	53,650	53,65	53,65

**Şekil 16: Birincil faktör analizi sonucunda veride açıklanan toplam varyans**

Bu tabloda yorumlanması gereken rakam, verilerden elde edilen boyutların örneklemdaki varyansı yüzdesel olarak açıklama oranıdır. Dolayısıyla bu boyutların kümülatif açıklama yüzdesine bakıldığında; %53,650 oranında bir açıklayıcılığa sahip

olduğu görülmektedir. Kline (1994), faktör analizinde kabul edilebilir minimum açıklayıcılığın 0,41 olması gerektiğini belirtmiştir.

	Bileşen				
	1	2	3	4	5
IIS18	,789				
IIS16	,761				
IIS27	,727				
IIS28	,725				
IIS22	,721				
IIS13	,707				
IIS30	,698				
IIS10	,689				
IIS32	,673				
IIS31					
IIS19					
IIS20		,734			
IIS15		,716			
IIS17		,698			
IIS26		,653			
IIS29		,595			
IIS1		,591			
IIS34		,545			
IIS23		,498			
IIS4					
IIS21			,700		
IIS11			,684		
IIS33			,674		
IIS3			,642		
IIS2			,551		
IIS6				,720	
IIS9				,687	
IIS8				,597	
IIS7				,577	
IIS5				,492	
IIS14				,486	
IIS12					
IIS25					,656
IIS24	,529				,559

Çıkarım Yöntemi: Temel Bileşen Analizi.

### Şekil 17: Birincil faktör analizi Rotated Component matrisi

Tabloda ifade numaralarının önündeki IIS Kurumsal Entegrasyon Ölçeği'ni ifade etmektedir. Örneğin, IIS1 anketteki “Derslerim bana yeni bakış açıları kazandırıyor.” İfadesine karşılık gelmektedir. Birbirine yakın ifadelerin gruplandığı bu matriste

yorumlanması gereken rakamlar, ifadelerin ilgili faktörlere yükleme oranlarıdır. İfadenin ilgili faktörlere yüklenmesi için eşik değerleri Tabachnik&Fidell (2007) tarafından şu şekilde sınıflandırılmıştır: 0,32- Yetersiz, 0,45- Geçerli, 0,55- İyi, 0,63- Çok iyi ve 0,71- Mükemmel. Örneklem yapısı ve büyüklüğü gözönüne alınarak faktör yükleme eşik değeri 0,45 olarak belirlenmiştir. Dolayısıyla 0,45 den az faktör yükleme değerlerine sahip olan ifadeler analiz dışı bırakılmıştır. Rotated Component Matrisi'ne göre 31. “ *Akademik anlamda tahmin ettiğim performans seviyesini yakaladım*”, 19. İfade “*Bu üniversitede ders dışı organizasyonlara katılma imkanlarından memnunum*”, 4. “ *Düşünsel gelişim seviyemden memnuniyet duyuyorum*” ve 12. “*Üniversiteye başlamadan öncesine oranla kültürel etkinliklere katılmaya daha istekliyim*” ifadelerinin faktör yükleme oranları belirlenen eşik değeri olan 0,45 in altında kalmıştır. Dolayısıyla bu ifadeler analiz dışı bırakılarak ikincil bir faktör analizi uygulamasına gerek görülmüştür.

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Örneklem yeterlilik ölçüsü		,931
	Yaklaşık Ki-Kare	6863,266
Bartlett's küresellik testi	df	435
	Sig.	,000

#### Şekil 18: İkincil faktör analizi KMO örneklem yeterlilik sonuçları

İkincil faktör analizi için yeni KMO-Örneklem yeterlilik değeri 0,931 olarak elde edilmiştir.

### AÇIKLANAN TOPLAM VARYANS

Bileşen	Başlangıç Özdeğerleri	Yüklerin karesinin çıkarım toplamları	Yüklerin karelerinin dönel toplamları
	% Kümülatif	%Kümülatif	%Kümülatif
1	33,471	33,471	19,514
2	42,332	42,332	32,266
3	48,072	48,072	43,464
4	52,780	52,780	52,375
5	56,388	56,388	56,388

#### Şekil 19: İkincil faktör analizi sonucunda veride açıklanan toplam varyans

Bu şekilde dikkat edilmesi gereken nokta, herhangi bir faktöre yüklenemeyen ifadelerin analiz dışı bırakılması kümülatif varyans açıklama oranını %53,650 den %56,388 e yükseltmiştir. Bu noktada ilgisiz ifadelerin, kullanılan anketin algı ve tutum ölçme yeteneğini olumsuz anlamda etkilediği söylenebilir.



Rotated Component Matrix<sup>a</sup>

	Component				
	1	2	3	4	5
IIS18	,789				
IIS16	,762				
IIS27	,733				
IIS28	,723				
IIS22	,717				
IIS13	,709				
IIS30	,700				
IIS10	,692				
IIS32	,681				
IIS24	,564				,474
IIS20		,745			
IIS15		,716			
IIS17		,704			
IIS26		,663			
IIS1		,597			
IIS29		,587			
IIS34		,556			
IIS23		,509			
IIS21			,699		
IIS11			,695		
IIS33			,685		
IIS3			,647		
IIS2			,563		
IIS6				,721	
IIS9				,684	
IIS7				,584	
IIS8				,577	
IIS5				,534	
IIS14				,481	
IIS25					,784

Çıkarım Yöntemi: Temel Bileşen Analizi.

Döndürme Yöntemi: Kaizer Normalizasyonlu Varimax

Şekil 20: İkincil faktör analizi Rotated Component matrisi

Elde edilen matriste yer alan her bir ifade belirlenen eşik değeri üzerinde yükleme oranına sahiptir. Dolayısıyla örneklemin nihai boyutları aşağıda belirtilen şekilde oluşmuştur:

### **1. boyutta yer alan ifadeler**

- İrtibat halinde olduğum pek çok öğretim üyesi, öğrenciler için önemli ve ilgi uyandıran meseleleri ders dışında tartışmak üzere zaman ayırmaktan çekinmezler.
- Şimdiye kadar irtibat kurduğum öğretim üyeleri gerçekten öğrencilerle ilgileniyorlar.
- Öğretim üyeleriyle ders dışı etkileşimlerim kariyer hedeflerimi ve isteğimi olumlu anlamda etkiliyor.
- Şimdiye kadar irtibata geçtiğim öğretim üyeleri gerçekten seçkinler ya da üstün öğreticilerdir.
- Öğretim üyeleriyle tanışma ve etkileşim halinde olma fırsatlarından memnunum.
- Öğretim üyeleriyle ders dışı etkileşimlerim kişisel gelişimimi, değer yargılarımı ve yaklaşımlarımı olumlu anlamda etkilemektedir.
- Şimdiye kadar irtibat kurduğum öğretim üyeleri gerçekten öğretmekle ilgileniyorlar.
- Şimdiye kadar irtibat kurduğum öğretim üyeleri ders dışı konularda da gelişmeye yardımcı olmak konusunda ilgilidirler.
- Öğretim üyeleriyle ders dışı etkileşimlerim düşünsel gelişimimi ve fikirlere olan merakımı olumlu anlamda etkiliyor.
- En az bir öğretim üyesiyle yakın ve kişisel bir ilişki kurdum.

### **2. boyutta yer alan ifadeler**

- Diğer öğrencilerle yakın arkadaşlık bağları kurdum.
- Diğer öğrencilerle tanışmak ve arkadaş olmak benim için çok kolay.
- Tanıdığım çoğu öğrenci arkadaşım, kişisel bir problemim olduğunda dinlemeye ve yardım etmeye hazırdır.
- Kurduğum arkadaşlıklardan memnunum.
- Diğer öğrencilerle aramdaki kişisel ilişkiler kişisel gelişimimi, değer yargılarımı ve yaklaşımlarımı olumlu anlamda etkiliyor.

- Karşı cins ile iletişim kurmakta zorlanmıyorum.
- Okuldan arkadaşlarımla aramdaki ilişkiler düşünsel gelişimimi ve fikirlere olan ilgimi olumlu anlamda etkiliyor.
- Bu üniversitede çoğu öğrenciyle benzer değer yargıları ve yaklaşımları paylaşıyorum.

### **3. boyutta yer alan ifadeler**

- Bu üniversitede eğitimime devam etmek istiyorum.
- Bu üniversiteye gelmek ile doğru bir karar verdiğimden eminim.
- Sakarya Üniversitesi'nden mezun olmak benim için önemlidir.
- Yaşadığım yerde mutluyum.
- Bu üniversiteden akademik anlamda memnunum.

### **4. boyutta yer alan ifadeler**

- Üniversiteye başlamamdan bu yana fikirler ve düşünsel meselelere olan ilgim arttı.
- Bu yıl şimdiye dek edindiğim akademik, düşünsel gelişimimi ve fikirlere olan ilgimi olumlu anlamda etkiliyor.
- Gelecekte ne yapmak istediğim hakkında fikir sahibiyim.
- İyi notlar almak benim için önemli.
- Derslerim bana yeni bakış açıları kazandırıyor.
- İstenen okuma ödevlerine ek olarak, derslerde tavsiye edilen pek çok kitap okuyorum.

### **5. boyutta yer alan ifadeler**

- Herhangi bir üniversiteden mezun olmak benim için önemlidir.

KEÖ' ne ait 5 alt ölçeğe ulaşılmıştır. Ancak dikkat edilmesi gereken nokta; bir ölçek ya da alt ölçeğin tek bir ifadeden oluşamayacağıdır. Dolayısıyla faktör analizleri sonucunda oluşan 5. boyut çalışmada göz ardı edilmiştir.

Elde edilen boyutlar orijinal ölçekte (French&Oakes, 2004) yer alan alt boyutlarla karşılaştırıldığında, 1. boyut, orijinal ölçekte “Öğretim Üyeleriyle Etkileşim” ve “Öğretim Üyelerinin Öğretmek ve Gelişim Konusundaki İlgisi” alt ölçeklerinde yer alan

ifadeleri barındırmaktadır. Dolayısıyla orijinal çalışmanın uygulandığı A.B.D.'de yükseköğretim gören öğrenciler için farklı iki boyutta belirtilen ifadeler Türkiye'de ki öğrenciler için tek boyutta yer almaktadır. Sonuç olarak yapılan faktör analizleri sonucunda öğrencilerin ilk yılda edindikleri sosyal ve akademik deneyim 4 faktör altında toplanmaktadır. Çalışma sonucunda elde edilen boyutlara içerdikleri ifadelerle göre şu şekilde isimlendirilmiştir.

- 1. boyut : Öğretim Üyeleriyle Etkileşim
- 2. boyut: Arkadaş- Grup Etkileşimi
- 3. boyut: Üniversiteye Bağlılık ve Memnuniyet
- 4. boyut: Akademik ve Düşünsel Gelişim

Yeni elde edilen alt ölçeklerin güvenilirlik dereceleri olan iç tutarlılık oranları kabul edilebilir seviye (Nunnally, 1978) olan 0,7'nin üzerindedir.

### **Ölçek: Öğretim Üyeleriyle Etkileşim**

#### **Güvenilirlik istatistikleri**

Cronbach's Alpha	İfade sayısı N
,914	10

### **Ölçek: Arkadaş-Grup Etkileşimi**

#### **Güvenilirlik istatistikleri**

Cronbach's Alpha	İfade sayısı N
,846	8

### **Ölçek: Üniversiteye Bağlılık ve Memnuniyet**

#### **Güvenilirlik istatistikleri**

Cronbach's Alpha	İfade sayısı N
,811	5

## Ölçek: Akademik ve Düşünsel Gelişim

### Güvenilirlik istatistikleri

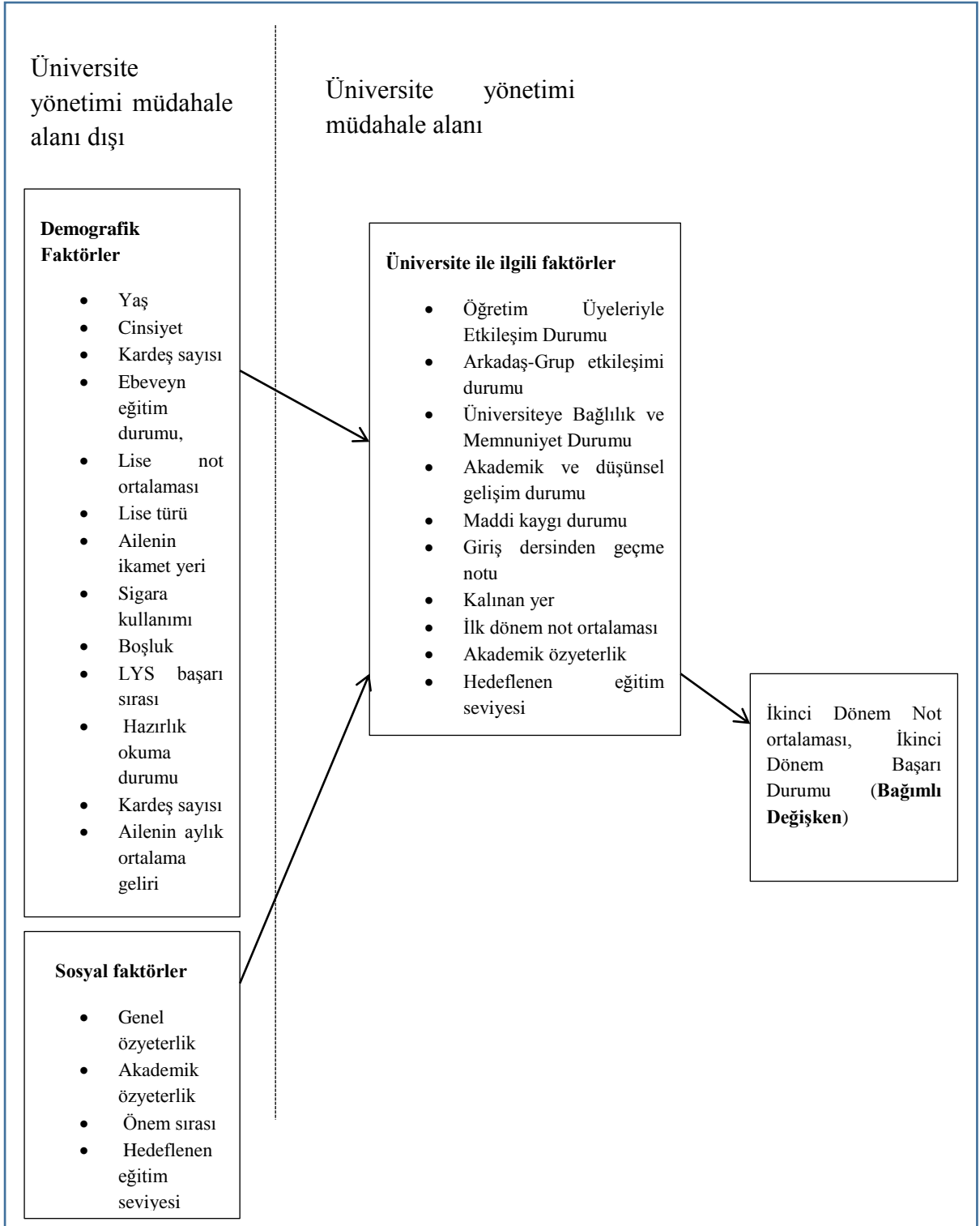
Cronbach's Alpha	İfade sayısı N
,756	6

### Şekil 21: Çalışmada elde edilen alt ölçeklere ait iç tutarlılık oranları ( $\alpha$ )

Elde edilen sonuçlara göre KEÖ'den (Kurumsal Entegrasyon Ölçeği) elde edilen değişkenler şu şekildedir:

- Öğretim Üyeleriyle Etkileşim Durumu
- Arkadaş-Grup Etkileşim Durumu
- Üniversiteye Bağlılık ve Memnuniyet Durumu
- Akademik ve Düşünsel Gelişim Durumu

Elde edilen bu veriler ışığında toplanan verileri daha iyi anlayabilmek üzere çalışmaya temel teşkil eden teorik model şu şekilde görselleştirilebilir.



Şekil 22: İkinci Dönem Not Ortalaması ve Başarı Durumu Üzerine Teorik bir Model

Ortaya konan teorik model; ilgili literatürde çok tartışılan konulardan birisi olan üniversitenin müdahale alanı (Tinto&Pusser, 2006; Astin, 1997; Pascarella&Terenzini, 1991) içerisinde bağımlı ve bağımsız değişkenleri konumlandırmıştır. Örnek vermek gerekirse, öğrencinin lise not ortalaması üniversite yönetimlerinin müdahale edemeyeceği bir değişken iken, ilk dönem not ortalamasına, alınacak bazı iyileştirici tedbirler (öğretim üyelerine yönelik pedagojik eğitimler, sosyal, akademik ve finansal destek sağlama, etkin bir izleme ve geri besleme sistemi, ders dışı aktiviteler, oryantasyon vs.) vasıtasıyla müdahale edilebilir. Modelde vurgulanması gereken bir diğer nokta ise; bazı değişkenlerin üniversitelerin hem müdahale alanı dışında hem de müdahale alanı içerisinde yer almalarıdır. Öğrencinin akademik özyeterlik algısı ve hedeflenen eğitim seviyesi modelde yer alan bu tip değişkenlerdir. Bunun sebebi, öğrencilerin yükseköğretime başlamadan önce sahip oldukları akademik özyeterlik algısı ve hedeflenen eğitim seviyesinin akademik ortama ve yönlendirmelere bağlı olarak değişebilmesidir. Model ile ilgili son olarak belirtilmesi gereken husus, üniversite müdahale alanı içerisinde ve dışarısında yer alan değişkenler, öğrencinin hassas bir geçiş dönemi yaşadığı ilk yıl sonundaki not ortalamasına etki etmektedir.

### **2.2.3. Veriyi Hazırlama**

Veri hazırlama; veri toplama, veri entegrasyonu, veri dönüştürme, veri temizleme, veri indirgeme ve veri ayrıştırma gibi ham veriden kaliteli verilere ulaşabilmek üzere kullanılan analiz tekniklerinden oluşmaktadır (Zhang vd.; 2010: s.376). Pek çok işlemde oluşan bu karmaşık aşamanın daha iyi anlaşılabilmesi için izlenen yol Tablo-7'de (s.55) (Wirth&Hipp, 2000: s.34) belirtilen veri hazırlama aşamalarına uygun bir şekilde sunulacaktır.

#### **2.2.3.1 Veriseti tanımı**

Çalışmada Sakarya Üniversitesi İşletme Fakültesi'nde bulunan lisans programlarında ilk yıllarını geçiren 520 öğrenciye uygulanan "Yeni Gelen Öğrenci Anketi" vasıtasıyla veri toplama işlemi gerçekleştirilmiştir. Bu anketlerden elde edilen veriler şu şekildedir:

- OGRENCI\_NUMARASI
- EGITIM (Örgün Öğretim- 2. Öğretim- Leeds)
- YAS
- CINSIYET
- KARDES\_SAYISI
- SEHIR (Ailenin ikamet ettiği şehir)
- BABA\_EGITIM (Baba eğitim durumu)
- ANNE\_EGITIM (Anne eğitim durumu)
- GELIR (Ailenin aylık ortalama geliri)
- EKSTRA\_GELIR (Öğrencinin ailesinin sağladığının dışında dışında sahip olduğu gelir)
- KALINAN\_YER
- SIGARA\_KULLANIM
- MADDI\_KAYGI (Öğrencinin hayatını idame ettirebilme konusundaki kaygı seviyesi)
- LISE\_MEZUNİYET\_YILI
- LISE\_TURU (Mezun olunan lise türü)
- LISE\_NOT\_ORTALAMASI
- LYS\_SIRA
- ONEM\_SIRA\_1 (Öğrencinin üniversiteye gelmesindeki en önemli neden)
- ONEM\_SIRA\_2 (Öğrencinin üniversiteye gelmesindeki diğer önemli neden)
- HEDEF\_EGITIM\_SEVIYESI
- GIRIS\_DERSI (Öğrencinin kayıtlı olduğu bölümde aldığı giriş dersinden aldığı not)
- DEVAMSIZLIK



### **2.2.3.2 Veri seçimi**

520 katılımcıyla uygulanan “Yeni Gelen Öğrenci Anketi”nden elde edilen veriler için ilk eleme kriteri, çalışmanın amacına uygun olarak akademik performansın takibi ve değerlendirilmesine yönelik talep edilen öğrenci numarasının uygunluğudur. Öğrenci numarasını yanlış beyan eden veya hiç beyan etmeyen 15 adet katılımcının anketleri örneklem dışı bırakılmıştır. Sonraki eleme aşamasında, çalışmanın örnekleminde yer almayan yatay geçiş ile gelen öğrenciler, kayıtlı oldukları lisans programlarında ilk yıllarını geçirmeyen öğrenciler (anketlerin uygulandığı dersleri alttan öğrenciler) analiz için kullanılacak verisetinin dışında tutulmuştur. Öğrenci numaralarının doğruluğunun ve uygunluğunun kontrolünde Sakarya Üniversitesi Öğrenci Veritabanı’ndan sağlanan verilerden yararlanılmıştır. Verisetine dahil olabilmek için son kriter ise, anket alanlarındaki ifadelerle katılımı belirten 5-li likert ölçeğine verilen cevapların tutarlılığıdır. Daha anlamlı sonuçlara ulaşabilmek üzere Genel Özyeterlik, Akademik Özyeterlik ve Kurumsal Entegrasyon ölçeklerinde yer alan tüm 5-li Likert tipi ifadelerle aynı katılım düzeyinde cevap veren katılımcıların anketleri elenmiştir. Böylelikle analizde kullanılabilir anket sayısı 445 olarak ortaya çıkmıştır.

### **2.2.3.3. Veri temizleme**

Veri temizleme , verideki tutarsızlıkları ve hataları (geçersiz veri, yanlış girilmiş veri, tekrarlanan veri)belirlemek ve düzeltmek suretiyle veri kalitesini arttırmayı amaçlar (Erhard&Do, 2000: s. 3). Bu aşamada öncelikle, Üniversitenin Öğrenci Veritabanından elde edilen bazı veriler ile beyan edilen veriler karşılaştırılarak yanlış veri girişinin önüne geçilmesi amaçlanmıştır. Yapılan karşılaştırma işlemlerinde katılımcıların yaşları, lise mezuniyet yılları, LYS başarı sıralamaları, ilk dönem not ortalamaları, bölüme ait giriş derslerinden alınan notları öğrenci veritabanından elde edilen veriler ile karşılaştırılmış ve çeşitli nedenlerle (hatırlayamama, çekinme vs.) yanlış beyan edilen alanlar değiştirilmiştir. Bu çalışmaya örnek vermek gerekirse, 70 öğrencinin ilk dönem not ortalamalarını yanlış beyan ettiği görülmüştür. Dikkat çeken nokta ise; notlarını yanlış beyan eden öğrencilerin büyük kısmının not ortalamalarının 2,00’in altında olmasıdır. Sonraki aşama ise, analizde kullanılmak üzere kayıp verilerin örneklem dağılımına uygun olarak doldurulmasıdır. Uygulanan anketlerde kayıp alanlar nominal değerler alıyorsa, tüm örneklemin ilgili alanının modu bu alanın alacağı değere uygun

olarak düşünülmüştür. Örneğin, yaşadığı şehiri beyan etmeyen bir katılımcının ilgili alanı dağılıma uygun olarak doldurulmuştur (Örnekleme “Kırsal”da yaşayan 45, “Kent” te yaşayan 251 ve “Metropol” de yaşayan 159 öğrenci bulunmaktadır.). Lise not ortalaması, Genel Özyeterlik, Akademik Özyeterlik, Kurumsal Entegrasyon ölçeklerinde yer alan boş bırakılmış alanlar ise ilgili değişkenin örneklem ortalama değeri ile doldurulmuştur.

#### 2.2.3.4. Veri türetme

En genel anlamıyla mevcut veriler kullanılarak yeni bir değişken oluşturma işlemidir. Çalışmada kullanılması düşünülen tahminleme modelinin doğruluk oranını arttırması öngörülen değişkenler türetilmiştir. Bu bağlamda türetilen değişkenler şu şekildedir:

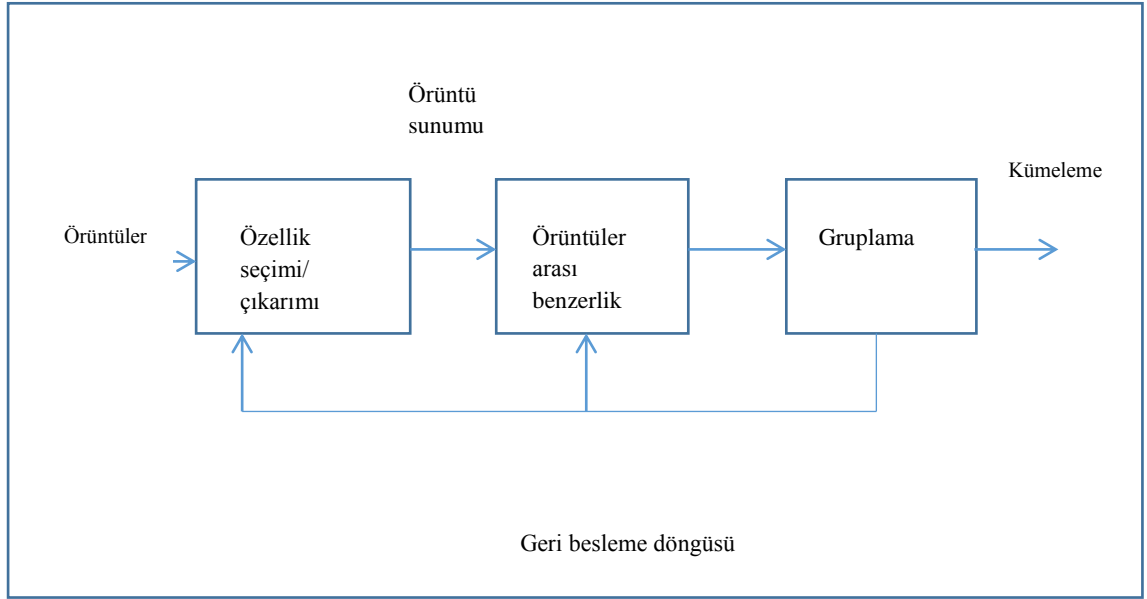
- **Lise mezuniyet yılı – Liseden sonra boşluk:** Öğrencinin liseden sonra yükseköğretime geçişte ara vermesi, akademik anlamda bir soğuma yaşanması ihtimalini beraberinde getirmektedir. Dolayısıyla yapılan çalışmalarda, bu durum ile akademik performans ve devamlılık arasındaki ilişki incelenmiştir (Mishra vd., 2014; Delen, 2010). Buna göre, öğrencilerin beyan ettiği lise mezuniyet yılı ile üniversiteye kayıt yılı arasında fark olması “1-Boşluk var” , olmaması ise “2- Boşluk yok” olarak belirtilmiştir.
- **Hazırlık durumu:** Çalışmanın uygulandığı üniversitede, hazırlık okumak öğrencilerin tercihlerine bırakılmıştır. Tinto'nun (1975; 1988) akademik ve sosyal entegrasyonla ilgili ortaya koyduğu teorik modellere dayanarak hazırlık okumayı seçen öğrencilerin, kayıtlı oldukları bölümlerde lisans eğitimlerine başlamadan önce sosyal ve akademik ortama adaptasyon anlamında mesafe katettiği savunulabilir. Bu noktadan yola çıkarak, dezavantajlarının (akademik anlamda boşluk) yanında hazırlık okuma durumu, akademik performans üzerinde etkisi araştırılmak üzere çalışmaya bağımsız değişken olarak dahil edilmiştir. Hazırlık okuyan öğrenciler “1”, hazırlık okumayan öğrenciler ise “2” olarak gruplanmıştır. Bu verinin türetilmesinde katılımcının öğrenci numarasından faydalanılmıştır. Sakarya Üniversitesi öğrenci numaralarında yer alan ilk iki rakam öğrencinin giriş yılını temsil etmektedir. (Örneğin b1313.060033 numaralı bir öğrenci üniversiteye 2013 yılında kayıt yaptırmıştır.) Buna göre, Üniversiteye 2013-2014 akademik yılında kaydolun ancak 2014-

2015 yılından itibaren kayıtlı oldukları bölümlerde ilk yıllarını geçiren öğrencilerin hazırlık okumuş oldukları varsayılmıştır. Bu çıkarım, Üniversite Öğrenci Veritabanı'ndan teyit edilmiştir.

Analizlerde kullanılmak üzere, katılımcıların memnuniyet ve başarı durumlarının göreceli olarak değerlendirilebilmesi için ilgili değerlerin sınıflandırılması ile yeni bir ordinal değişken türetilmesi düşünülmüştür. Sınıflar arasındaki uygun eşik değerlerinin belirlenmesi için başarı ve memnuniyetin örnekleme göre değişebileceği gözönünde bulundurularak kümeleme analizinden yararlanılmıştır. Burada yöntemin daha iyi anlaşılabilmesi adına kümeleme analizini tanımlamak faydalı olacaktır.

**Kümeleme Analizi:** Öğrencilerin akademik performanslarına yönelik çalışmalarda sıklıkla kullanılan (Oyelade vd., 2010; Rasmani&Shen, 2006; Taylan&Karagözoğlu, 2009; Shovon, 2012) kümeleme analizi, bir diğer deyişle doğal gruplama, örüntülerin (gözlem, her bir veri) aralarındaki benzerliklere dayanarak kümelere toplanma işlemidir. Öğreticisiz bir sınıflandırma tekniği olan kümeleme gruplama, belge tarama, görüntü bölümlendirme ve örüntü sınıflandırma gibi pek çok konuda kendisine kullanım alanı bulmuştur. Standart bir kümeleme işleminin adımları şu şekilde belirtilmiştir:

- Örüntü sunumu (Özellik çıkarımı ve/veya seçimi işlemleri)
- Verinin yapısına uygun olarak noktalar arası uygun uzaklık ölçüsü tanımı
- Kümeleme veya gruplama
- Veri özetleme
- Çıktının değerlendirilmesi (Jain vd. ,1999: s.265-266).

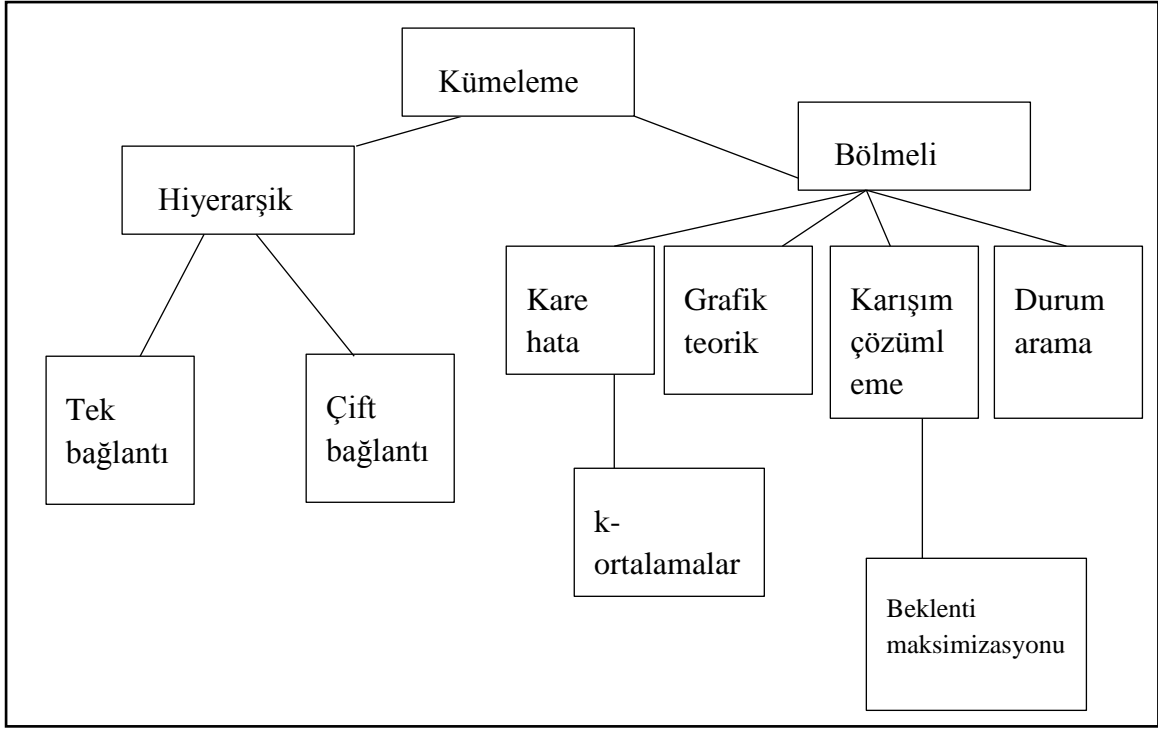


**Şekil 23: Kümeleme aşamaları (Jain vd. , 1999: s. 267)**

Kümeleme prosedüründe örüntüler arasındaki benzerlik ölçüsü büyük önem arz etmektedir. İki örüntü arasındaki benzersizliği belirleyebilmek için en sık kullanılan ölçü özellik uzayında tanımlanan mesafedir. Bilinen en popüler mesafe ölçüsü ise iki örüntü arasındaki dik uzaklığı hesaplayan Öklid mesafesidir:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Verilerin kümelenmesine yönelik pek çok yaklaşım mevcuttur. Veri analizi yapacak olanlara yol gösterecek bu yaklaşımların taksonomisi şekilde sunulmuştur:



**Şekil 24: Kümeleme yaklaşımları taksonomisi (Jain vd. , 1999: s.275)**

Mevcut çalışmada, sınıflar arasındaki eşik değerleri taksonomide yer alan k-ortalamalar yöntemi kullanılarak belirlenecektir.

Kümeleme yaklaşımlarından biri olan k-ortalamalar yöntemi Jain (2010) tarafından şu şekilde özetlenmiştir: k-ortalamalar algoritması, kümenin deneysel ortalaması ve o kümede yer alan her bir nokta arasındaki kareli hatayı minimize etmeye çalışır ve hata formülü şu şekildedir:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2.$$

$\mu_k$ ,  $c_k$  kümesinin ortalaması,  $x_i$  ise  $c_k$  kümesinde yer alan n- boyutlu noktalardan her biridir.. k-ortalamalar algoritmasının genel işleyiş adımları şu şekildedir.

- 1- Verisetinin bölünmesi istenen sayı belirtilir.
- 2- Her bir noktayı en yakın küme merkezine atayarak yeni bir bölümlendirme oluştur.
- 3- Yeni küme merkezleri belirlenir (Jain, 2010: s. 653-654).

k-ortalamlar yönteminde küme merkezi ve noktalar arasındaki 2 boyutlu mesafe hesabında tipik olarak Öklid'in uzaklık formülünden yararlanılmaktadır.

Mevcut çalışmada, k-ortalamlar yöntemiyle kümeler arası eşik değerlerin belirlenmesinde IBM SPSS Statistics 20 paket programı kullanılmıştır.

**Lise Başarı Durumu:** Katılımcıların lise not ortalamalarına göre 3 gruba ayrılması istenmiştir: “1- Düşük” “2- Ortalama” “3- Başarılı” Buna göre lise not ortalamalarının dağılımına göre k-Ortalamlar yöntemiyle belirlenen küme merkezleri şu şekilde oluşmuştur:

	Cluster		
	1	2	3
LİSE_NOT_ORTALAMASI	2,37	3,39	2,90

**Her bir kümedeki örnek sayısı**

1	73,000
Cluster 2	78,000
3	294,000
Valid	445,000
Missing	,000

**Şekil 25: Lise Başarı Durumu için eşik değerleri ve kümelerin üye sayıları**

**İlk dönem ve ikinci dönem not ortalamaları başarı durumları:** Katılımcıların ilk ve ikinci dönemde göstermiş oldukları akademik performansları tahminleyici analizde kullanılmak üzere “1-Düşük” “2- Ortalama” “3- Başarılı” şeklinde kümelenebilir istenmiştir ve sınıflandırma işleminde referans olarak kullanılacak eşik değerleri k-ortalamlar yöntemiyle bulunmuştur.

**Nihai küme merkezleri**

	Cluster		
	1	2	3
İLK_DÖNEM_NOT_ORTAL AMASI	1,29	2,89	2,04

**Her bir kümedeki örnek sayısı**

1	152,000
Cluster 2	124,000
3	169,000
Valid	445,000
Missing	,000

**Şekil 26: İlk Dönem Başarı Durumu için eşik değerleri ve kümelerin üye sayıları**

**Nihai küme merkezleri**

	Cluster		
	1	2	3
İKINCI_DONEM_NOT_ORT ALAMASI	1,0198	2,8715	1,8566

**Her bir kümedeki örnek sayısı**

1	131,000
2	124,000
3	190,000
Valid	445,000
Missing	,000

**Şekil 27: İkinci Dönem Başarı Durumu için eşik değerleri ve kümelerin üye sayıları**

**Genel Özyeterlik- Akademik Özyeterlik durumu:** Katılımcıların Genel Özyeterlik ve Akademik Özyeterlik algı seviyelerinin ayırdedilebilmesi amacıyla verilen cevapların “1-Düşük” “2-Orta” “3-Yüksek” olarak sınıflandırılması düşünülmüştür. Bu anlamda ifadeler katılım derecelerinin “1- Hiç katılmıyorum ..... 5- Tamamen katılıyorum” skalasında yer aldığı göz önüne alınarak olumsuz herhangi bir ifadenin yer almadığı ölçeklerde algı seviyelerinin, ifadelerden elde edilen puanların toplamı ile doğru orantılı olarak değiştiği varsayılmıştır. k-ortalama kümeleme algoritması uygulanması

sonucu ortaya çıkan küme merkezleri ve her bir derece için üye sayıları şekilde belirtildiği gibidir.

**Nihai küme merkezleri**

	Cluster		
	1	2	3
GEN_OZYET	20,03	29,77	36,22

**Her bir kümedeki örnek sayısı**

	1	37,000
Cluster	2	181,000
	3	254,000
Valid		472,000
Missing		,000

**Şekil 28: Genel Özyeterlik Derecesi için eşik değerleri ve kümelerin üye sayıları**

**Nihai küme merkezleri**

	Cluster		
	1	2	3
AKADEMIK_OZYETERLIK_DERECE	16,29	23,23	29,52

**Her bir kümedeki örnek sayısı**

	1	42,000
Cluster	2	205,000
	3	225,000
Valid		472,000
Missing		,000

**Şekil 29: Akademik Özyeterlik derecesi için eşik değerleri ve kümelerin üye sayıları**

**Kurumsal entegrasyon durumu:** Katılımcıların yapılan faktör analizi için elde edilen 4 boyuta yönelik algı seviyelerinin “1- Düşük” “2-Orta” “3-Yüksek” şeklinde kümelere ayrılması uygun görülmüştür. Yapılan kümeleme analizi sonucu her bir boyuta dönük küme merkezleri ve üye sayıları Şekil 30’ da belirtilmiştir.



**Nihai küme merkezleri**

	Cluster		
	1	2	3
OGRETIM_UYELERIYLE_ETKILESIM_DERECE	19,71	32,20	42,68

**Her bir kümedeki örnek sayısı**

	1	69,000
Cluster	2	223,000
	3	180,000
Valid		472,000
Missing		,000

**Nihai küme merkezleri**

	Cluster		
	1	2	3
ARKADAS_GRUP_ETKILE	17,47	28,47	34,87

**Her bir kümedeki örnek sayısı**

	1	43,000
Cluster	2	177,000
	3	252,000
Valid		472,000
Missing		,000

**Nihai küme merkezleri**

	Cluster		
	1	2	3
UNIVERSITEYE_BAGLILIK_VE_MEMNUNİYET_DERECE	18,09	11,23	22,95

**Her bir kümedeki örnek sayısı**

	1	233,000
Cluster	2	52,000
	3	187,000
Valid		472,000
Missing		,000

**Nihai küme merkezleri**

	Cluster		
	1	2	3
AKADEMIK_DUSUNSEL_G	14,43	26,25	21,24
ELISIM_DERECE			

**Her bir kümedeki örnek sayısı**

	1	44,000
Cluster	2	235,000
	3	193,000
Valid		472,000
Missing		,000

**Şekil 30: Kurumsal entegrasyon boyutlarına yönelik algı sınıfları eşik değerleri ve her bir sınıf için üye sayıları**

Oluşturulan bu değişkenlerden çalışmanın amacına uygun bağımsız değişkenlerin korelasyon katsayılarına göre tahminleme modellerine dahil edilmiştir. Korelasyon analizi, bağımlı değişkenin tahmin edilmesine yönelik yapılan çalışmalarda sınıflandırma modelinin doğruluk oranını arttırabilmek amacıyla uygun bağımsız değişken seçimi işlemlerinde kullanılmaktadır. Buna göre bağımlı değişken ile anlamlı bir ilişki içerisinde bulunan bağımsız değişkenlerin modelin tahmin gücünü arttıracağı varsayılmaktadır (Yu&Liu, 2003; Hall, 1999).

Bu noktada korelasyon analizini tanımlamak gerekirse, iki değişken arasındaki ilişkinin yönünü ve şiddetini gösterebilen bir istatistiksel tekniktir. Korelasyon analizi sonucu elde edilen katsayı ( $r$ )  $[-1,+1]$  aralığında değerler alır. Bu katsayı şu şekilde yorumlanmaktadır:

$-1 \leq r \leq -0,5$  ise negatif yönde kuvvetli bir ilişki,

$-0,5 \leq r < 0$  ise negatif yönde zayıf bir ilişki

$r = 0$  ise nötr ilişki

$0 < r \leq 0,5$  ise pozitif yönde zayıf bir ilişki

$0,5 \leq r \leq 1$  ise pozitif yönde kuvvetli bir ilişkiden söz edilebilir (SurveySystem, 2015).

Korelasyon katsayısı hesaplanmasına yönelik değişik yaklaşımlar olmakla birlikte değişkenler arasındaki ilişkiyi yorumlayabilmek adına en sık başvurulan yöntem Pearson Product Moment korelasyon katsayısıdır ve x ve y değişkenleri arasındaki korelasyon katsayısını hesaplama formülü şu şekildedir:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Pearson korelasyon katsayısı, rakamlar arasındaki mesafenin eşit olacağı varsayımında bulunur. Dolayısıyla kategorik değişkenler arasındaki ilişkilerin yönünü ve derecesini ortaya koymada yetersiz olduğu düşünülse de kullanılabilir olduğu düşünülmektedir (Bollen&Barb, 1981). Bahsedilen yetersizliğin üstesinden gelebilmek için kategorik (nominal ya da ordinal) değişkenler arasındaki ilişkilerin yorumlanmasına olanak sağlayan Pearson korelasyon katsayısının özelleşmiş formları olan Spearman sıralama korelasyonları (rho) ve Kendall'ın tau korelasyon katsayıları da kullanılmaktadır. Bu korelasyon göstergelerinden, Kendall'ın tau korelasyon katsayısının Spearman'ın sıralama katsayısına göre avantajları şu şekilde belirtilmiştir:

- Kendall'ın tau dağılımı daha iyi istatistiksel özelliklere sahiptir.
- Kendall'ın tau dağılımının uyumlu ve uyumsuz değişken çiftlerini gözlemleyebilme olasılığı bakımından yorumlanması çok açıktır.
- Kendal'ın tau katsayısı ve Spearman'ın sıralama korelasyonu katsayıları çok benzerdir ve çoğunlukla benzer çıkarımlarda bulunurlar (StatisticsSolutions, 2015).

Mevcut çalışmanın kategorik değişkenler içerdiği düşünülerek modelde kullanılacak verilere ait değişken isimleri ve tanımları ile birlikte bağımlı değişken ile olan Pearson ve Kendall's tau korelasyon katsayıları tabloda sunulmuştur.

**Tablo 13**

**Tahminleme modelinde kullanılan değişkenler ve bağımlı değişken ile aralarındaki korelasyon katsayıları**

Değişken adı	Tanımı (Frekans)	Bağımlı değişken ile arasındaki Pearson katsayısı (Anlam düzeyi)	Bağımlı değişken ile arasındaki Kendall's tau katsayısı (Anlam düzeyi)
OGRENCI_NUMARASI	Öğrenci numarası	-	-
BOLUM	1="İnsan Kaynakları Yönetimi" (130) 2="İşletme" (159) 3= "LEEDS" (10) 4="Turizm İşletmeciliği" (68) 5="Sağlık Yönetimi" (33) 6="Uluslararası Ticaret" (17) 7="Yönetim Bilişim Sistemleri" (28)	0,211 (% 1)	0,182 (% 1)
EGITIM	1="Öğün öğretim" (250) 2="İkinci öğretim"(195)	-0,311 (% 1)	-0,294 (% 1)
HAZIRLIK	1="Okudu" (57) 2="Okumadı" (388)	Anlamlı bir ilişki gözlenmedi	Anlamlı bir ilişki gözlenmedi
YAS	1= "17-19 yaş arası" (254) 2= "20-22 yaş arası" (185) 3= "23 yaş ve üzeri" (5)	Anlamlı bir ilişki gözlenmedi	Anlamlı bir ilişki gözlenmedi
CINSIYET	1= "Kadın" (267) 2= "Erkek" (178)	-0,135 (% 1)	-0,128 (% 1)
KARDES_sAYISI	Katılımcının kendisi dahil kardeş sayısı	Anlamlı bir ilişki gözlenmedi	Anlamlı bir ilişki gözlenmedi

**Tablo 13 - devamı**

SEHIR	1= "Kırsal" (45) 2= "Kent" (241) 3="Metropol" (159)	0,045 (% 1)	0,029 (% 1)
BABA_EGITIM	1="İlköğretim (216) 2= "Lise" (161) 3="Üniversite" (62) 4="Yükseklisans" (4) 5="Doktora"(2)	-0,158 (% 1)	-0,152 (% 1)
ANNE_EGITIM	1= "İlköğretim (305) 2= "Lise" (105) 3= "Üniversite" (33) 4="Yükseklisans"(2) 5="Doktora"(0)	-0,139 (% 1)	-0,138 (% 1)
GELIR	1="1000 TL ve altı"(48) 2="1001 -2500 TL arası"(209) 3="2501-4000 TL arası" (138) 4="4001-5500 TL arası"(28) 5="5501 TL ve üzeri" (22)	-0,128 (% 5)	-0,123 (% 1)
EKSTRA_GELIR	1="Burs" (60) 2="Kredi"(185) 3="Part-Time" (14) 4="Yok" (178) 5="Birden fazla gelir" (8)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi
KALINAN_YER	1="Devlet yurdu" (39) 2="Özel yurt" (205) 3="Öğrenci evi" (65) 4="Aile/Akraba yanı"(136)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi
HEDEF_EGITIM_SEVIYESI	1="Lisans" (112) 2="Yüksek lisans" (240) 3="Doktora" (93)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi
ILK_DONEM_NOT_ORTALAMAS I	0,0-4,00 aralığında yer almaktadır.	0,913 (% 1)	0,812 (% 1)
GIRIS_DERSI	0="ALMADI" (191) 1="AA" (31) 2="BA" (26) 3="BB" (33) 4="CB" (39) 5="CC" (77) 6="DC" (26)	-0,369 (% 1)	-0,286 (% 1)

**Tablo 13 - devamı**

	7="DD" (14) 8="FF" (8)		
DEVAMSIZLIK	1="0-2 hafta" (217) 2="2-4 hafta" (169) 3="4-6 hafta" (50) 4="6 hafta ve üzeri" (9)	-0,277 (% 1)	-0,252 (% 1)
GOY_DERECE	Genel Özyeterlik Derecesi 1="Düşük" (35) 2="Orta" (167) 3="Yüksek" (243)	0,118 (% 5)	0,111 (% 5)
AOY_DERECE	Akademik Özyeterlik Derecesi 1="Düşük" (38) 2="Orta" (192) 3="Yüksek" (215)	0,152 (% 1)	0,125 (% 1)
SIGARA_KULLANIM	1="Evet" (105) 2="Hayır" (340)	0,226 (% 1)	0,214 (% 1)
MADDI_KAYGI	1="Hayır; kaygım yok" (147) 2="Bazen; kaygılanıyorum" (247) 3="Evet; kaygılıyım" (51)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi
BOSLUK	1="Var" (200) 2="Yok" (245)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi
LISE_TURU	1="Düz lise" (169) 2="Anadolu lisesi" (170) 3="Çok programlı lise" (11) 4="Meslek lisesi" (83) 5="Fen lisesi" (0) 6="Diğer" (12)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi
ONEM_SIRA_1	Üniversiteye gelmede en önemli motivasyon 1="Daha iyi bir işe sahip olmak" (160) 2="Genel bir eğitime sahip olma ve fikir edinme" (44) 3="Daha kültürlü bir insan olmak" (62) 4="Daha fazla para kazanabilmek" (65) 5="İlgi alanlarım hakkında daha fazla bilgiye sahip olmak" (21)	Anlamli bir ilişki gözlenmedi	Anlamli bir ilişki gözlenmedi

**Tablo 13 - devamı**

	6-“Hedeflediğim kariyer üzerine hazırlık yapmak”(93)		
BOYUT_1_DERECE	Öğretim Üyeleriyle Etkileşim Durumu 1=“Düşük” (66) 2=“Orta”(205) 3=“Yüksek”(174)	0,110 (%1)	0,101 (%1)
BOYUT_2_DERECE	Arkadaş-Grup etkileşimi durumu 1=“Düşük” (37) 2=“Orta”(166) 3=“Yüksek”(242)	Anlamli bir ilişki gözlemlenmedi	Anlamli bir ilişki gözlemlenmedi
BOYUT_3_DERECE	Üniversiteye Bağlılık ve Memnuniyet Durumu 1=“Düşük” (51) 2=“Orta”(217) 3=“Yüksek”(177)	Anlamli bir ilişki gözlemlenmedi	Anlamli bir ilişki gözlemlenmedi
BOYUT_4_DERECE	Akademik ve Düşünsel Gelişim Durumu 1=“Düşük” (43) 2=“Orta”(180) 3=“Yüksek”(222)	0,211 (%1)	0,182 (%1)
LISE_NOT_ORTALAMASI	Akademik not ortalamasına uygun olarak (0,00-4,00) aralığında tanımlanmıştır.	0,175 (%1)	0,199 (%1)
LISE_BASARI_DURUMU	Lise diploma notundan elde edilen başarı durumu 1=“Düşük” (73) 2=“Orta”(294) 3=“Yüksek”(78)	0,175 (%1)	0,159 (%1)
ILK_DONEM_BASARI_DURUMU	1=“Düşük” (138) 2=“Orta”(179) 3=“Yüksek”(128)	0,603 (%1)	0,548 (%1)

İncelenmesi gereken bir diğer dikkat çeken husus ise, Kurumsal Entegrasyon Ölçeğinden elde edilen boyutlar arasındaki korelasyon katsayılarıdır.

Correlations						
			BOYUT1_DE RECE	BOYUT2_DE RECE	BOYUT3_DE RECE	BOYUT4_DE RECE
Kendall's tau_b	BOYUT1_DERECE	Correlation Coefficient	1,000	,391**	,402**	,442**
		Sig. (2-tailed)	.	,000	,000	,000
		N	445	445	445	445
	BOYUT2_DERECE	Correlation Coefficient	,391**	1,000	,396**	,368**
		Sig. (2-tailed)	,000	.	,000	,000
		N	445	445	445	445
	BOYUT3_DERECE	Correlation Coefficient	,402**	,396**	1,000	,410**
		Sig. (2-tailed)	,000	,000	.	,000
		N	445	445	445	445
	BOYUT4_DERECE	Correlation Coefficient	,442**	,368**	,410**	1,000
		Sig. (2-tailed)	,000	,000	,000	.
		N	445	445	445	445

\*\* . Correlation is significant at the 0.01 level (2-tailed).

**Şekil 31: Kurumsal Entegrasyon Ölçeği boyutları arasındaki korelasyon katsayıları**

Buna göre, Öğrencilerin Öğretim Üyeleriyle Etkileşim Durumu ve Akademik ve Düşünsel Gelişim Durumu arasındaki Kendall's tau korelasyon katsayısı = 0.442 ( $p=0.01$ ) olarak bulunmuştur. Dolayısıyla Öğretim Üyeleriyle Etkileşim Durumu yüksek olan öğrencilerin akademik ve düşünsel gelişim olarak olumlu etkilendiği söylenebilir. Tablodan yorumlanacak bir diğer katsayı ise Boyut3 (Üniversiteye Bağlılık ve Memnuniyet) ve Boyut4 (Akademik ve düşünsel gelişim) arasındaki korelasyon katsayısıdır.  $r$   $p=0.01$  anlam düzeyinde 0.410 olarak bulunmuştur. Dolayısıyla Üniversiteye Bağlılık ve Memnuniyet Durumu ile akademik ve düşünsel gelişimin paralellik gösterdiği savunulabilir.

Modelleme aşamasında kurulacak modellerde bağımsız değişkenler ise bağımsız değişkenler arasındaki korelasyon katsayıları göz önünde bulundurularak veriler modele dahil edilecektir. Diğer modelde ise tüm değişkenler modele dahil edilerek, korelasyon temelli özellik seçimi yaklaşımının tahminleme modeline olan etkisi araştırılmıştır.

## 2.2.4. Modelleme

### 2.2.4.1. Modelleme tekniği seç

Çalışmada, elde edilen veriler kullanılarak şimdiye dek akademik başarı ve devamlılık alanında pek çok tahminleme çalışmasında faydalanılmış Yapay Sinir Ağları (YSA) (Hardgrave vd. ,1994; Naik&Ragothaman, 2004; Gorr vd. ,1994; Delen, 2010; Wilson&Hardgrave, 1995; Mishra, Kumar,&S., 2014), Karar Ağaçları

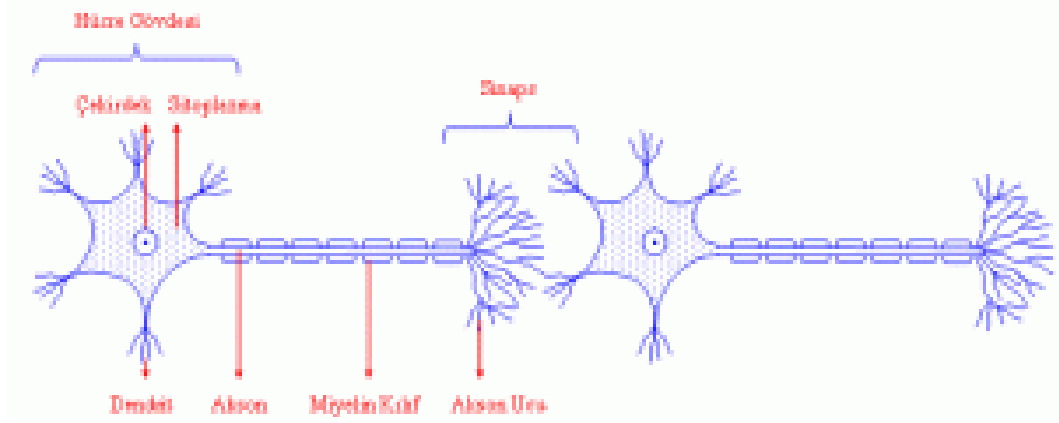


(Dekker&Pechenizkiy, 2009; Kovacic, 2010; Kabra&Bichkar, 2011; Ramaswami&Bhaskaran, 2010; Quadri&Kalyankar, 2010) ve Çokterimli Lojistik Regresyon (DeBerard vd. , 2004; Gore, 2006; Scales vd., 2006; Allen&Robbins, 2008; Chomitz vd., 2009) teknikleri kullanılacaktır. Model ortaya konmadan önce bu tekniklerin kısaca tanımlanması yerinde olacaktır.

### **a. Yapay Sinir Ağları**

Yapay Sinir Ağları (YSA) insan beyninin en temel özelliği olan öğrenme fonksiyonunu gerçekleştiren bilgisayar sistemleridir (Öztemel, 2012: s. 29) ve kollektif bir sistem olarak işleyen basit nöronlar arası çok büyük ölçekli bağlantılardır (Pal&Mitra, 1992: s. 683). Daha spesifik bir tanıma bakılacak olursa; bir YSA nöronları taklit eden ve sinyallerin ağlar üzerinde seri ve paralel olarak hareket edebileceği şekilde bir “ağırlıklar” (sinir sistemindeki sinaptik bağlantılara benzer olarak) kümesi vasıtasıyla birbirine bağlı olan işlem birimlerinden oluşmaktadır (Cross vd. ,1995: s. 1076). Uzaydan, eğlenceye pek çok sektörde kendisine uygulama alanı bulan Yapay Sinir Ağları genellikle şu amaçlar doğrultusunda kullanılmaktadır.

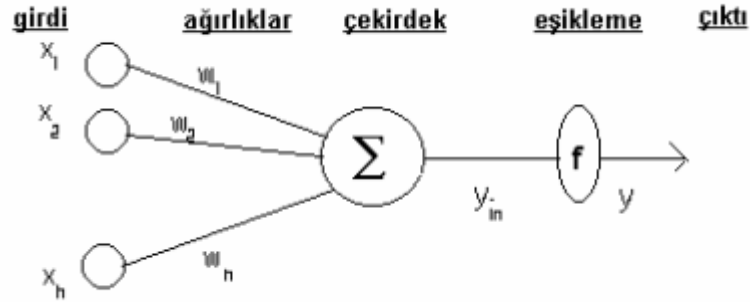
- Muhtemel fonksiyon kestirimleri
- Sınıflandırma
- İlişkilendirme veya örüntü eşleştirme
- Zaman serileri analizleri
- Sinyal filtreleme
- Veri sıkıştırma
- Örüntü tanıma
- Doğrusal olmayan sinyal işleme
- Doğrusal olmayan sistem modelleme
- Optimizasyon
- Kontrol (Çayiroğlu, 2015).



**Şekil 32: Bir sinir hücresi yapısı (Şen S. , 2011)**

Yapay Sinir Hücreleri dışarıdan gelen bilgileri toplama fonksiyonu ile toplar ve aktivasyon fonksiyonundan geçirerek ürettiği çıktıyı bağlantılar üzerinden diğer proses elemanlarına gönderir. Standart bir YSA temel olarak 3 bileşenden oluşmaktadır.

- Girdi katmanı (Bilgiler ağı girdi katmanından iletilir)
- Ara katman (Aktivasyon fonksiyonu doğrultusunda işlenerek çıktı katmanına iletilir)
- Çıktı katmanı (Üretilen değerler çıktı katmanında sunulur)



**Şekil 33: Yapay Sinir Hücresi yapısı (Cinsdikici, 2015: s. 3)**

Girdiler için doğru ağırlıklar vasıtasıyla doğru çıktılar üretilme sürecine ağı eğitilmesi adı verilmektedir. Ağı bir olayı öğrenebilmesi, o ağı karakterize eden parametrelerin doğru belirlenmesi ile mümkün olabilmektedir. Model olarak da bilinen bir ağı karakterize eden bu unsurlar şu şekildedir:

- **Ağ yapısı**

- **Tek katmanlı:** Yapay Sinir Ağları ile ilgili çalışmalar tek katmanlı algılayıcılar ile başlamıştır. Bu tip ağlar sadece girdi ve çıktı katmanlarından oluşmaktadırlar. Bu algılayıcıların en önemli özelliği problem uzayını bir doğru veya bir düzlem olarak sınıflara ayırmalarıdır. Problemin girdileri ağırlıklar ile çarpılıp toplandıktan sonra elde edilen değerlerin bir eşik değerinden büyük veya küçük olmasına göre girdinin sınıfı belirlenir. Sınıflar 1 veya -1 rakamları (bazen 1 ve 0) rakamları ile gösterilir. Doğrusal olmayan ilişkileri öğrenememekteler (Öztemel, 2012: s. 59-74).
- **Çok katmanlı:** Tek katmanlı ağların doğrusal olmayan ilişkileri öğrenememesi sorununu çözmek için geliştirilmiş ağlardır. 3 katmandan oluşmaktadırlar: Girdi katmanı (Bilgilerin alındığı katmandır), Ara katmanlar (Girdi katmanlarından gelen bilgileri işlerler. Girdi/Çıktı arasındaki ilişkinin karmaşıklığına göre katman sayısı artabilir.), Çıktı katmanı (Ara katmandan sağlanan bilgiye göre ilgili girdiye ait çıktıyı sunar) Bu katmanlara ait parametreler deneme-yanılma yöntemiyle bulunur (Öztemel, 2012: s. 75-113).

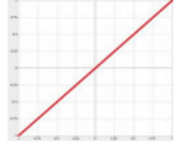
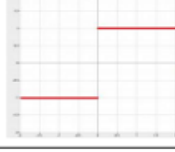
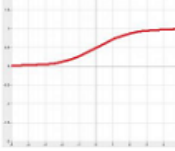
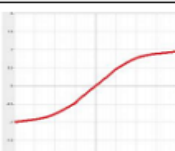
Ağlar verinin akış yönüne göre de yapısal farklılıklar göstermektedirler.

- **İleri beslemeli ağ:** Her bir katmandaki hücreler sadece bir önceki katmanın hücrelerince beslenir.
- **Geri beslemeli ağ:** En az bir hücre sonraki katmanlardaki hücrelerce de beslenir (Gulez, 2015).

- **Kullanılan toplama fonksiyonu**

- **Çarpım:** Ağırlık değerleri girdiler ile çarpılır ve daha sonra bulunan değerler birbirleri ile çarpılarak net girdi hesaplanır.
- **Maksimum:** N adet girdi içinden ağırlıklar ile çarpıldıktan sonra en büyük değer Yapay Sinir Hücresinin net girdisi olarak kabul edilir.
- **Minimum:** N adet girdi içinden ağırlıklar ile çarpıldıktan sonra en küçük değer Yapay Sinir Hücresi'nin net girdisi olarak kabul edilir

- **Çoğunluk:** N adet girdi içinden ağırlıklar ile çarpıldıktan sonra pozitif ve negatif olanların sayısı bulunur. Büyük olan sayı hücrenin net girdisi olarak kabul edilir.
- **Kümülatif toplam:** Hücreye gelen bilgiler ağırlıklı toplanır ve daha önce gelen bilgilere eklenerek hücrenin net girdisi bulunur (Uğurlu, 2015).
- **Kullanılan aktivasyon fonksiyonu**
  - **Lineer aktivasyon fonksiyonu:** Gerçek sayılar kümesinde yalnızca pozitif sayılar üretir. Sabit noktalarda kullanılamaz.
  - **Sigmoid aktivasyon fonksiyonu:** Yalnızca 0 ile 1 arasında pozitif rakamlar üretebilir. 0 ile 1 arasında yer alan eğitim verisinde gayet iyi sonuçlar verir.
  - **Sigmoid aşamalı aktivasyon fonksiyonu:** Bilinen sigmoid fonksiyonunun daha hızlı ancak biraz daha az kesin sonuçlar veren fonksiyonudur.
  - **Sigmoid simetrik aktivasyon fonksiyonu:** -1 ile +1 arasında sonuçlar üreten tanh (hiperbolik tanjant) sigmoid fonksiyonudur.
  - **Gauss aktivasyon fonksiyonu:** Aktivasyon aralığında daha yüksek kontrol istendiğinde kullanılır. 0-1 aralığında değerler almaktadır.
  - **Gauss simetrik aktivasyon fonksiyonu:** -1 - +1 aralığında değerler almaktadır.
  - **Elliot aktivasyon fonksiyonu:** Hiperbolik tanjant aktivasyon fonksiyonundan daha hızlı bir versiyonudur. 0 ile 1 arasında değerler alabilir.
  - **Elliot simetrik aktivasyon fonksiyonu:** -1 ile +1 arasında değerler almaktadır.
  - **Lineer parçalı aktivasyon fonksiyonu:** Doyurucu lineer fonksiyon olarak da bilinir ve hem ikili (0-1) hem de çift kutuplu aralıkta yer alır. 0-1 aralığındadır.
  - **Lineer parça simetrik aktivasyon fonksiyonu:** -1 ile +1 arasında değişen değerler alır (Sibi vd., 2013: s. 1266) .

Doğrusal (Lineer) Aktivasyon Fonksiyonu		$F(\text{NET})=A \cdot \text{NET}$ (A sabit bir sayı)	Doğrusal problemler çözmek amacıyla aktivasyon fonksiyonu doğrusal bir fonksiyon olarak seçilebilir. Toplama fonksiyonundan çıkan sonuç, belli bir katsayı ile çarpılarak hücrenin çıktısı olarak hesaplanır.
Adım (Step) Aktivasyon Fonksiyonu		$F(\text{Net})= \begin{cases} 1 & \text{if Net} > \text{Eşik Değer} \\ 0 & \text{if Net} \leq \text{Eşik Değer} \end{cases}$	Gelen Net girdinin belirlenen bir eşik değerinin altında veya üstünde olmasına göre hücrenin çıktısı 1 veya 0 değerini alır.
Sigmoid Aktivasyon Fonksiyonu		$F(\text{Net})= \frac{1}{1+e^{-\text{Net}}}$	Sigmoid aktivasyon fonksiyonu sürekli ve türevi alınabilir bir fonksiyondur. Doğrusal olmayışı dolayısıyla yapay sinir ağı uygulamalarında en sık kullanılan fonksiyondur. Bu fonksiyon girdi değerlerinin her biri için 0 ile 1 arasında bir değer üretir.
Tanjant Hiperbolik Aktivasyon Fonksiyonu		$F(\text{Net})= \frac{e^{\text{Net}} + e^{-\text{Net}}}{e^{\text{Net}} - e^{-\text{Net}}}$	Tanjant hiperbolik fonksiyonu, sigmoid fonksiyonuna benzer bir fonksiyondur. Sigmoid fonksiyonunda çıkış değerleri 0 ile 1 arasında değişirken hiperbolik tanjant fonksiyonunun çıkış değerleri -1 ile 1 arasında değişmektedir.
Eşik Değer Fonksiyonu		$F(\text{Net})= \begin{cases} 0 & \text{if Net} \leq 0 \\ \text{Net} & \text{if } 0 < \text{Net} < 1 \\ 1 & \text{if Net} \geq 1 \end{cases}$	Gelen bilgilerin 0 dan küçük-eşit olduğunda 0 çıktısı, 1 den büyük-eşit olduğunda 1 çıktısı, 0 ile 1 arasında olduğunda ise yine kendisini veren çıktılar üretilebilir.
Sinüs Aktivasyon Fonksiyonu		$F(\text{Net}) = \text{Sin}(\text{Net})$	Öğrenilmesi düşünülen olayların sinüs fonksiyonuna uygun dağılım gösterdiği durumlarda kullanılır.

Şekil 34: YSA' da kullanılan aktivasyon fonksiyonları (Çayiroğlu, 2015)

- **Öğrenme stratejisi**

- **Denetimli öğrenme:** Ağa belirli girdi ve çıktı değerleri verilir. Nöronlar arasındaki ağırlıklar verilen giriş ve çıkış değerlerine göre yarılanır. Delta öğrenme kuralı ve geri yayımlı (backpropagation) öğrenme denetimli öğrenmedir.
- **Denetimsiz öğrenme:** Çıktı değerlerinin girilmesine gerek yoktur. Öğrenme süresince sadece örnek giriş değerleri verilir ve örnekler arasındaki ilişkiler ile sistemin kendi kendine öğrenmesi sağlanır.
- **Destekleyici öğrenme:** Denetimli öğrenmenin farklı bir türüdür. Çıkış değerleri girmek yerine verilen giriş değerlerine göre çıkış değerlendirerek öğrenmeyi sağlar (Akyol, 2015).

- **Öğrenme kuralı**

- **Hebb kuralı:** Eğer nöron (A) başka bir nörondan (B) girdi alıyorsa ve her ikisi de aktifse, (A) ve (B) arasındaki ağırlık artar.

$$\Delta w_{jk} = \alpha \cdot y_j \cdot x_k$$

Bu formülde  $w_{jk}$  nöron  $u_k$  den  $u_j$  ya olan ağırlık,  $y_j$ ,  $u_j$  nöronunun çıktısı ve  $x_k$  ise  $u_k$  nöronunun çıktısıdır.  $\alpha$  “öğrenme katsayısı” olarak adlandırılmaktadır. 0 ile 1 arasında değerler almaktadır. Bu değer büyüklüğü ile ağırlık öğrenme hızı arasında pozitif, genelleme yeteneği arasında negatif bir ilişki mevcuttur.

- **Delta kuralı:** Bu kural, nöronun gerçek çıktı ile istenilen çıktı değerleri arasındaki farkı azaltan, giriş bağlantılarını güçlendiren ve sürekli olarak değiştiren bir düşünceye dayanmaktadır. Delta kuralı, ortalama karesel hatayı, bağlantı ağırlık değerlerinin değiştirilmesi ile düşürme prensibine dayanır. Hata aynı anda bir katmandan bir önceki katmanlara geri yayılarak azaltılır. Ağırlık hatalarının düşürülmesi işlemi, çıkış katmanından giriş katmanına ulaşıncaya kadar devam eder. Bir nöron çıktısının,  $y_j$ , hedeflenen çıktı  $d_j$  den farkı (hata),  $\delta_j$  nöron  $j$  için şu şekilde hesaplanır:

$$\delta_j = d_j - y_j$$

Bu algoritma, hata karelerinin ortalamasını alarak, bu değer en küçük olduğu çözümü bulmayı amaçlar. Ağ için hata karelerinin ortalaması  $\xi$

$$\xi = E \left[ \frac{1}{2} \sum_j \delta_j^2 \right]$$

formülü yardımıyla hesaplanır.  $E$  istatistiksel olarak beklenen değeri (ortalamayı) ifade etmektedir.

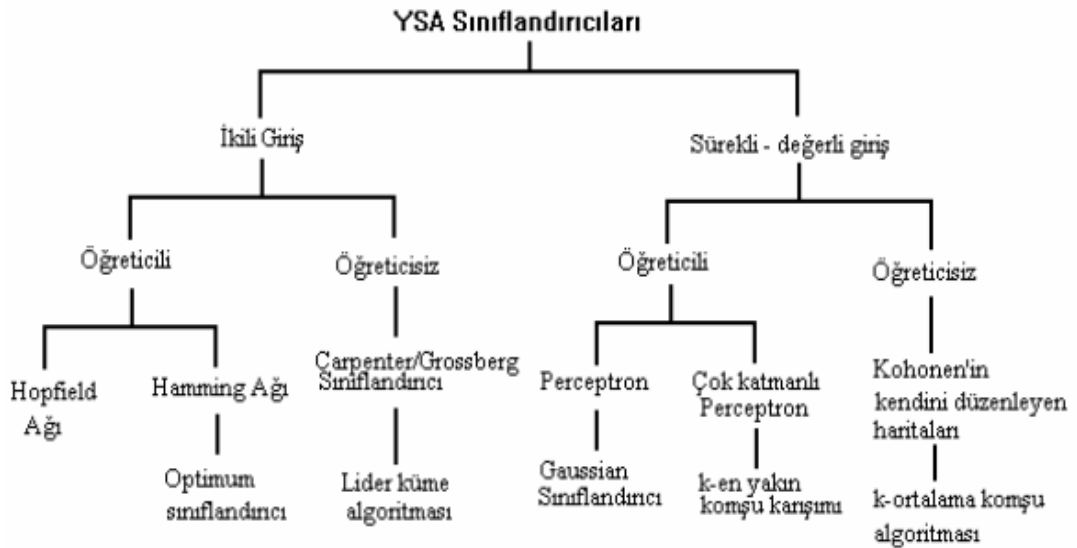
Delta algoritması hataların karesinin en küçük olduğu noktayı bulurken bir fonksiyonun lokal minimumunu bulma yöntemlerinden biri olan (OnMyPhD, 2015) dereceli azaltma yöntemini (gradient descent) kullanmaktadır.

- **Kohonen kuralı:** Nöronlar öğrenmek için yarışır. Kazanan nöronun ağırlıkları güncellenir. Bu kural “kazanan tamamını alır” olarak da

bilinir. En büyük çıkışa sahip işlemci nöron kazanır. Bu nöron, komşularını uyarma ve men etme hakkını elde eder. Kohonen kuralı, hedef çıkışa gereksinim duymaz. Bu nedenle danışmansız bir öğrenme kuralıdır.

**Hopfield kuralı:** Eğer istenilen çıkış ve girişin her ikisi aktif veya her ikisi de aktif değilse öğrenme oranı tarafından bağlantılı ağırlığı artırılır. Diğer durumlarda ise azaltılır. Birçok öğrenme katsayısı oranı veya sabiti vardır. Genellikle bu terim 0 ile 1 arasında değerler alır (Bilecik Üniversitesi- Bilgisayar Mühendisliği, 2015: s. 58-61) .

Bahsedilen özelliklerden yola çıkılarak Yapay Sinir Ağları şu şekilde sınıflandırılabilir.

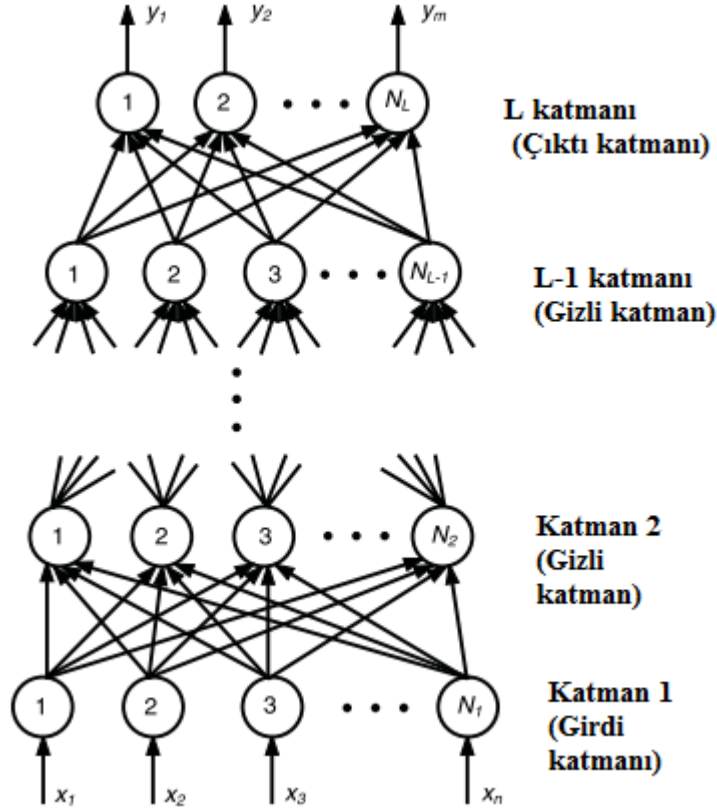


**Şekil 35: Yapay Sinir Ağları'na ait bir taksonomi (Gulez, 2015: s.6)**

Geliştirilen modeller arasında en yaygın olarak kullanılan tek ve çok katmanlı (MLP) algılayıcılar, LVQ (Learning Vector Quantization), ART (Adaptive Resonance Theory) ağları, SOM (Self-Organizing Maps) gibi ağlardır (Öztemel, 2012: s.45-57).

Çalışmada Yapay Sinir Ağı modellerinden yaygın olarak kullanılan (Oladokun vd. ,2008; Delen, 2010; Osmanbegovic&Suljic, 2012) Çok Katmanlı Algılayıcılar (ÇKA) ile öğrencilerin başarı aralıkları tahmin edilmeye çalışılmıştır. ÇKA, ileri beslemeli bir

ağ yapısına sahiptir ve sürekli ve bütünleşebilir değişkenler içeren fonksiyonların genel sınıflarının tahmin edebilir (Scarselli&Tsoi, 1998) .



Şekil 36: Çok Katmanlı Algılayıcı yapısı (Zhang&Gupta, 2000)

ÇKA'lar ve genel olarak YSA'lar, girdi-çıkı arasındaki doğrusal olmayan karmaşık ilişkileri tespit edebilmekte ve bu ilişkiler doğrultusunda tahminlerde bulunmaktadır. Ayrıca yapılarında tuttukları ağırlıklar değiştirilerek ağı eğitilmesi ve sonraki örnekler için daha isabetli tahminlerde bulunabilmeleri dikkat çekici bir diğer yöndür. Bu yönleriyle düşünüldüğünde ÇKA, öğrencilerin yükseköğretimde ilk yıl deneyimlerinin sarmal halde bulunan değişkenlere (kişilik özellikleri, finansal durum, akademik ortam vs.) bağlı olarak değişmesiyle karmaşıklaşan bu süreci modellemede uygun bir araç olarak düşünülmüştür.

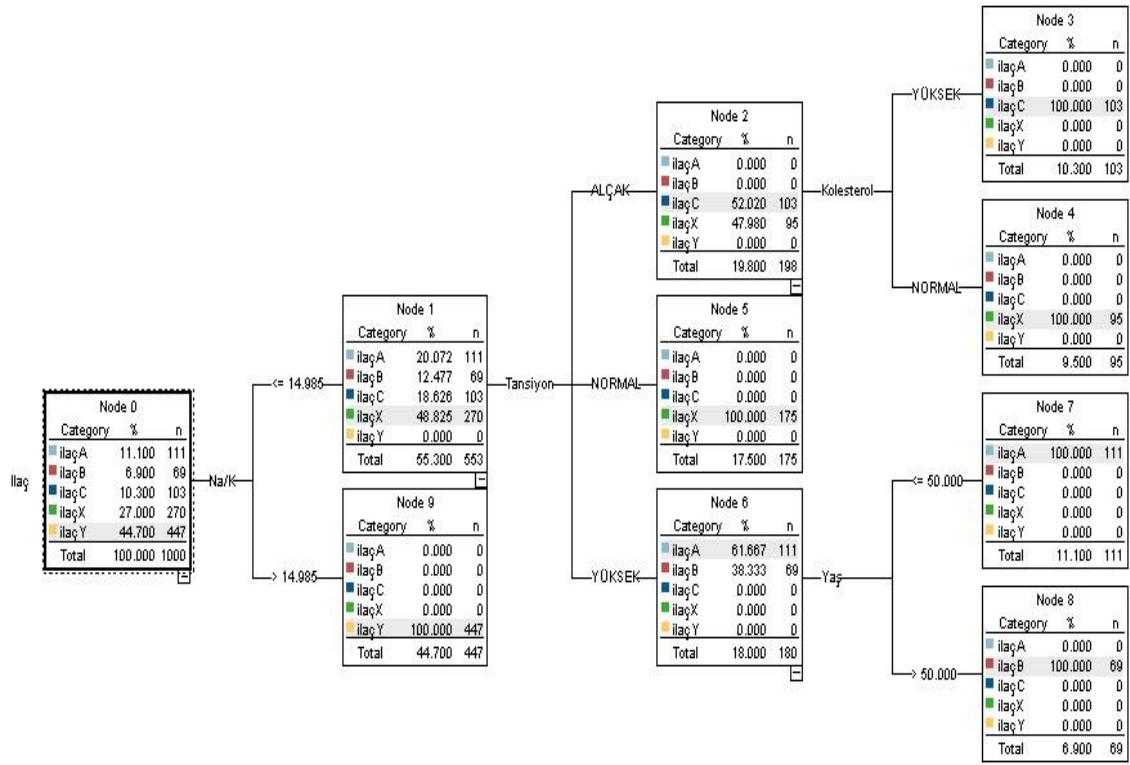
## b. Karar Ağaçları



Makine öğrenme yöntemlerinden biri olarak gösterilen Karar Ağaçları, çoklu değişken analizinin basit ama güçlü bir şeklidir (Statistical Analysis System, 2015). Özyinelemeli olarak verinin bölünmesine ve hiyerarşik ve sıralı yapıları verilerin altında yatan kuralların sunulmasına yönelik bir yöntemdir. Karar Ağaçları belirtilen amaçlar doğrultusunda kullanılabilir:

- **Betimleme:** Büyük miktarlarda veriyi, özkarakteristikleri kaybetmeden ve doğru bir özet sunarak daha kompakt (yoğun) bir verisetine dönüştürebilir.
- **Sınıflandırma:** Verinin iyi bir şekilde ayrıştırılabilir sınıfları içerip içermediğini keşfedebilir.
- **Genelleme:** Bağımsız değişkenlerden yola çıkarak bağımlı değişkenlere giden yollar ortaya koyar ve bağımlı değişkenin gelecekteki değerinin tahmin edilmesinde kullanılabilir (Murthy, 1998: s. 345).

Karar Ağaçları akış şemalarına benzeyen yapılardır. Her bir nitelik (özellik) bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının elemanlarıdır. En son yapı “yaprak”, en üst yapı “kök” ve bunların arasında kalan yapılar ise “ dal” olarak isimlendirilir (Quinlan, 1993: s. 5).



**Şekil 37: Hastalara verilecek ilaçlar ile ilgili örnek bir Karar Ağacı (Soner, 2011)**

Karar Ağaçlarının işleyiş mantığı basit bir örnek üzerinde anlatılacak olursa, şekilde görselleştirilmiş olan Karar Ağacında bir hastaya ait bağımsız değişkenler olan Yaş, Cinsiyet, Tansiyon (Yüksek, Normal, Alçak), Kolesterol (Normal, Yüksek), Na (Kandaki sodyum miktarı) ve K (Kandaki potasyum miktarı) kullanılarak hastaya verilmesi uygun olan ilacın belirlenmesi amaçlanmaktadır. Buna göre dallanmalar Karar Ağacına göre sırasıyla kazanım oranları en yüksek NA/K (Kandaki Sodyumun Potasyuma oranı), Tansiyon, Kolesterol ve Yaş değişkenlerine göre olmuştur. Uygulanan modelin hastaya doğru ilacı tavsiye etme oranı ise %99,5 olarak gerçekleşmiştir. Bu gösterimden basit bir örnek bir kural şu şekilde ifade edilebilir:

- Eğer NA/K ≤ 14,985 ve Eğer Tansiyon= Alçak ve Eğer Kolesterol= Yüksek ise İlaçC

Buna göre, gelecek yeni bir hasta için belli dallanma kriterleri göz önünde bulundurularak bir kural oluşturmak mümkündür. Bu noktada sınıflandırma ve tahmin

yapabilmek için dallanma kriterinin belirlenmesi büyük önem taşımaktadır. Dallanma kriterlerinin belirlenmesine yönelik geliştirilen algoritmalar 3 ana başlık altında toplanmıştır:

**1- Entropiye dayalı algoritmalar:** Bir sistemdeki belirsizliğin ölçüsüne “entropi” adı verilir. Daha açık bir şekilde belirtmek gerekirse,  $S$  bir kaynak için, bu kaynağın  $\{m_1, m_2, \dots, m_n\}$  olmak üzere  $n$  mesaj ürettiğini varsayılır. Bu mesajlar birbirinden bağımsız olmak olarak üretilmektedir ve  $m_i$  mesajlarının üretilme olasılığı  $p_i$  ‘dir.  $P = \{p_1, p_2, \dots, p_n\}$  olasılık dağılımına sahip mesajları üreten  $S$  kaynağının entropisi şu şekildedir (Shannon, 2001: s. 23) :

$$H(S) = H(S) = - \sum_{i=1}^n (p_i * \log_2(p_i))$$

Karar Ağaçlarının oluşturulması esnasında dallanmaya hangi özellikten başlanacağı önem taşımaktadır. Öyle ki, 5 özellikli 20 elemanlı bir eğitim kümesindeki verilerden yararlanılarak çizilebilecek ağaçların sayısı  $10^6$  dan daha büyüktür (Kantardzic, 2003). Dolayısıyla bu anlamda başlangıç için özellik seçiminde entropiden yararlanılabileceği öngörülmüştür. Buna göre; eğitim kümesinin  $\{C_1, C_2, \dots, C_k\}$  olmak üzere  $k$  sınıfa ayrıldığını varsayıldığında,  $T$  sınıf değerlerini içeren küme için  $P_T$  olasılık dağılımıdır:

$$P_T = \left( \left| \frac{C_1}{T} \right|, \left| \frac{C_2}{T} \right|, \dots, \left| \frac{C_k}{T} \right| \right)$$

Bu durumda  $T$  için ortalama bilgi miktarı veya bir başka deyişle entropi şu şekilde ifade edilebilir:

$$H(T) = - \sum_{i=1}^n (p_i * \log_2(p_i)) \text{ (Quinlan, 1993)}$$

Hedef niteliğini ifade eden  $T$ , hedef niteliği olmayan (yani sınıf niteliği olmayan) bir  $X$  değerine bağlı olarak  $T_1, T_2, \dots, T_n$  alt kümelerine ayrılırsa,  $T$  nin bir elemanının sınıfını belirlemek için gerekli bilgi,  $T_i$  nin bir elemanının sınıfının belirlenmesinde gerekli olan bilginin ağırlıklı ortalaması olarak kabul edilir. O halde tanıma bağlı olarak  $T$  nin bir elemanının sınıfını belirlemek için gerekli bilgi şu şekilde hesaplanır (Quinlan, 1993):

$$H(S) = - \sum_{i=1}^n \left| \frac{T_1}{T} \right| * H(T_i)$$

T veritabanını  $X$  testine uygun olarak bölmekle elde edilen bilgileri ölçmek için “kazanç ölçütü” adı verilen bir ifadeye başvurulmuştur. Bu ölçüt şu şekilde tanımlanır:

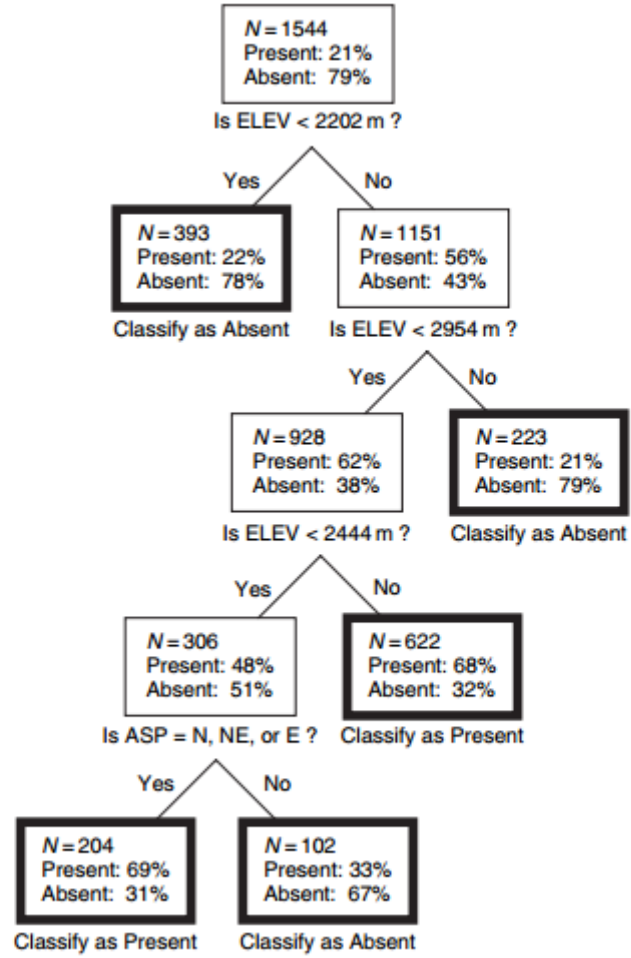
$$Kazanç (X,T) = H(T) - H(X,T)$$

Burada ayırma işlemi yapılırken  $Kazanç (X,T)$  değerini maksimize etmek amaçlanır. En yüksek bilgi kazancını sağlayan  $X$  testi seçilir.

Kategorik değişkenlerle çalışan ID3 ve sayısal değerlerinde kullanılabildiği C4.5 algoritmaları entropiye dayalı algoritmalara iyi birer örnektir (Özkan, 2008: s. 51-76).

2- **Sınıflandırma ve Regresyon Ağaçları (CART)** : Hem kategorik hem de sürekli değişkenleri analizde kullanılabilen CART veri setlerindeki değişkenler arası ilişkileri keşfetmede kullanılan analitik araçlardır (Speybroeck, 2012: s. 243). CART’ın sahip olduğu en önemli özelliklerden biri eksik verilerle çalışabilmesidir. CART, açıklayıcı değişkenleri olası çıktılarının en homojen bileşkesini içerecek şekilde yapraklara bölütleme yapmaktadır. Buna göre, her bir düğüm nihai düğüme kadar bir sonraki düğüme ikili düzende bölünmektedir. Bu noktada saflığın korunması esas alınmaktadır. Saflığın korunmasına yönelik geliştirilen ölçütlerden en yaygın kullanılan 4 tanesi:

- Yanlış sınıflandırma hatası
- Gini indeksi
- Entropi indeksi
- Twoing indeksi (Moisen, 2008: s. 583-584)



Şekil 38: Örnek bir CART yapısı (Moisen, 2008: s. 584)

**3- Bellek tabanlı sınıflandırma (IBL) algoritmaları:** k-en yakın komşu algoritması olarak da bilinen IBL, sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır. Örnek kümedeki her bir gözlem bir nokta olarak ele alınır ve noktalar arası uzaklıkların hesaplanmasında Öklid uzaklık formülü kullanılır (Özkan, 2008: s. 117). Breiman vd. (1984) en yakın komşu algoritmalarına yönelik pek çok problem tespit etmişlerdir.

- 1- Tüm eğitim örneklerini saklamalarından ötürü hesaplama açısından maliyetlidirler.
- 2- Özellik kirliliğine karşı toleransları yoktur.
- 3- İlgisiz özelliklere karşı toleransları yoktur.
- 4- Algoritmanın benzerlik fonksiyonunun seçimine karşı hassastırlar.

5- Nominal değerli ya da boş bırakılmış alanlarla (özellikler) ilgilenebilmek üzere doğal bir yol mevcut değildir.

6- Verinin yapısı göz önüne alındığında nispeten az faydalı bilgi sunarlar.

Bu problemleri alanları göz önüne alan araştırmacılar, farklı Karar Ağaçları modellerini bir araya getirerek “topluluk” tekniklerini ortaya koymuşlardır. “Rassal Ormanlar“ (RF) (Breiman, 2001) bu tekniklerden biridir.  $\{\Theta_k\}$  birbirinden bağımsız eş dağılmış rassal vektörler olmak üzere ve her bir ağacın  $x$  girdisi için en popüler sınıf lehine bir birim oluşturduğu durumda bulunan ağaç yapılı sınıflayıcılar  $\{h(x, \Theta_k), k = 1, \dots\}$  topluluğundan oluşan bir sınıflama algoritmasıdır. Bu algoritma, aynı deneyin çok fazla tekrar edilmesiyle elde edilen değerlerin ortalamalarının beklenen değere yaklaşacağını savunan bir olasılık teoremi olan Büyük Sayılar Yasasına dayanarak aşırı uyum sorununu ortadan kaldırmaktadır. Temel çalışma mantığı ise şu şekilde işlemektedir: mevcut ağaç için eğitim seti değiştirmeye örneklendiğinde, verilerin yaklaşık 3 te 1 i örneklem dışı bırakılmaktadır. Bu örneklem dışı (out-of-bag) veri ağaçlar ormana eklendikçe sınıflandırma hatalarını yansız tahminleyicisi olarak kullanılmaktadır. Ayrıca değişkenlerin önemlerinin tahminlenmesinde de faydalanılmaktadır. Her bir ağaç oluşturulduktan sonra, tüm veriler ağaçtan aşağı yönde ilerletilir ve her bir veri çifti için yakınlıklar hesaplanır. Eğer iki veri aynı nihai düğümde yer alıyorsa, yakınlıkları bir arttırılır. Sürecin sonunda, yakınlıklar ağaç sayısına bölünerek normalleştirilir. Yakınlıklar kayıp verilerin değiştirilmesinde, aykırı verilerin konumlandırılmasında ve verinin aydınlatıcı düşük-boyutsal görünüşünün üretilmesinde kullanılır (Breiman, 2001). Daha basit bir şekilde açıklanacak olursa; RF yönteminde, karar ormanını oluşturan Karar Ağaçları orijinal verisetinden bootstrap yöntemiyle seçilen farklı örneklerden oluşturulmaktadır. Her Karar Ağacında verisetindeki tüm değişkenlerden rastgele seçilen az sayıda değişken kullanılmaktadır. Her ağaç bir sınıf için oy vermektedir ve orman sınıflayıcısı bütün ağaçların verdiği oyları toplayarak bir sınıf için son tahminini yapmaktadır (Akman vd., 2011: s.36). RF nin özellikleri ise şu şekilde belirtilebilir:

- Mevcut algoritmalar içerisinde doğruluk oranı yönünden eşsizdir.
- Büyük veri setlerinde etkin olarak çalışır.
- Değişken silmeden binlerce girdi değişkenini kullanabilir.

- Sınıflandırmada hangi deęişkenlerin önemli olabileceğine dair tahminler sunar.
- Ormanın oluşturulması esnasında genelleme hatasına yönelik yansız bir iç tahminleyici üretir.
- Kayıp verilerin tahmin edilmesine yönelik etkin bir metodu vardır ve verisetinin büyük bir kısmının kayıp olması halinde bile doğruluk oranını sürdürebilir.
- Hatanın sınıf içerisinde dengelenmesi için yöntemleri vardır.
- Üretilen ormanlar gelecekteki kullanım için saklanabilir.
- Sınıflandırma ve deęişkenler arasındaki ilişkiler hakkında bilgi sağlayan yakınlıkları hesaplar.
- Veri çiftlerinin arasındaki yakınlığı hesaplaması yönüyle kümeleme, aykırı veri belirleme amacıyla kullanılabilir.
- Veriler arasındaki etkileşimi tespit etmek için deneysel bir yöntem sunar (University of California, Berkeley, 2015).

RF akademik performans ve devamlılığın tahmin edilmesine yönelik ideal bir yapı ortaya koymaktadır ve bu yönde kullanılmaktadır (Delen, 2010; Dekker vd., 2009; Yadav , 2012). Tüm bu belirtilenler doğrultusunda RF çalışmanın yapısına uygun bir yöntem olarak düşünülmüştür.

**c. Çokterimli Lojistik Regresyon:** Geleneksel istatistiksel sınıflandırma yöntemlerinden olan Çok terimli lojistik regresyon, kategorik yerleşimi tahmin etmede veya çoklu bağımsız deęişkenlere baęlı olarak baęımlı deęişkenlerin kategorik üyeliklerinin olasılıklarının tahmininde kullanılır. Sadece 2 sınıfta çıktı üreten lojistik regresyonun basit bir uzantısıdır. Çokterimli Lojistik Regresyon analizinde birden çok çıktı sınıfı tanımlanabilmektedir. Kategorik üyeliklerin olasılıklarının deęerlendirilebilmesi için ikili lojistik regresyonda olduęu gibi maksimum benzerlik tahmini kullanılır (Starkweather&Moske, 2015). Çokterimli Lojistik Regresyon için önerilen örneklem büyüklüğü her bir bağımsız deęişken için en az 10 örnektir (Schwab, 2015). Örneğin, mevcut çalışmada 32 bağımsız deęişken için analize girmesi gereken minimum örneklem büyüklüğü 320 dür. Çalışmanın örneklem büyüklüğü N=445 dir. Dolayısıyla Çokterimli Lojistik Regresyonuna elverişli olduęu savunulabilir.

Öğrencilerin akademik ve kişisel özgeçmişlerinin, finansal durumlarının, genel ve akademik özyeterliklerinin ve lisans programlarında geçirdikleri ilk yılda içerisinde

buldukları akademik ve sosyal ortama entegre olabilmelerinin akademik başarıya etki ettiğini varsayan tahminleme modelinde, Çok Katmanlı Algılayıcılar, Rassal Ormanlar ve Çokterimli Lojistik Regresyon tekniklerinin kullanılması ve farklı senaryolar için sonuçlarının karşılaştırılarak uygun olan yöntemin seçilmesi öngörülmüştür.

#### **2.2.4.2. Tasarımı test et**

Elde edilen verileri kullanarak seçilen tahminleme yöntemlerinin her biri Waikato Üniversitesi tarafından geliştirilen açık kaynak kodlu WEKA 3.6 veri madenciliği paket programı vasıtasıyla test edilmiştir. Bu aşamada tekniklerin sınıflandırma doğruluklarının daha gerçekçi sayılarla ortaya konması ve güvenilirliğin artırılması amacıyla k- katlamalı çapraz doğrulama tekniğinden yararlanılmıştır. Bu teknikte, veriseti k eşit parçaya bölünür. Elde edilen k adet alt veri setlerinden biri test seti olarak kullanılırken, diğerleri eğitim seti olarak kullanılır. Sırayla her bir alt verisetinin test seti olarak ayrıldığı bu yöntem k kez tekrarlanır. Bu denemeler sonucunda her bir tekrarın hatalarının ortalamaları alınarak genel bir doğru sınıflandırma oranına erişilir. Bu yöntemde, k sayısı arttıkça tahmin oranlarındaki sapma azalmakta ve daha tutarlı sonuçlar elde edilmektedir. Böylelikle modelin yeni verileri genelleme yeteneği kontrol edilmektedir (Schneider, 1997). Bu teknikte kullanılan k sabiti örneklem büyüklüğünden yola çıkılarak farklı şekillerde belirlenebiliyor olsa da, akademik performans ve devamlılığa yönelik tahminleme çalışmalarında genellikle k=10 şeklinde belirlenmektedir (Mishra vd. 2014; Delen, 2010; Bresfelean vd., 2008).

Test aşamasında modelin performansının değerlendirilmesi için kullanılan göstergeler şu şekildedir:

- **Doğru sınıflandırılan örnek sayısı (Correctly Classified Instances)**
- **Yanlış sınıflandırılan örnek sayısı (Incorrectly Classified Instances)**
- **Kappa istatistiği (Cohen's Kappa coefficient):** Kategorik değişkenlerin, değerlendirici uzlaşımını ölçmek için kullanılan istatistiksel bir ölçümdür. Kappa katsayısı, N tane birimi C tane ayrı kategori altında sınıflandıran iki değerlendirici arasındaki uzlaşım düzeyini ölçmektedir. Genellikle 0 ile 1 arasında değerler alan Kappa istatistiği, nadiren de olsa negatif değerler alabilmektedir. Bu katsayının 1 olması mükemmel uyumu göstermektedir (TÜİK, 2011: s.38).



- **Ortalama mutlak hata (Mean absolute error)**
- **Standart Sapma (Root mean squared error):** Tahmin edilen ile gözlemlenen değerler arasındaki farkların örneklem standart sapmasını ifade etmektedir.
- **Göreceli mutlak hata (Relative absolute error):** Mutlak sapmanın ortalama değere bölümünün 100 ile çarpılması sonucu elde edilmektedir (Demir, 2009: s.15).
- **Göreceli hataların karelerinin karekökü (Root relative squared error):** Standart sapmanın , sadece hedef değerlerin tahmin edilmesinden elde edilen standart sapmaya bölünmesinden elde edilen değer 100 ile çarpılmış halidir. İdeal değeri 0 dır. Bir modele ait RRSE nin %100 den büyük olması, ortalamayı tahmin etmekten ziyade kötüleştiğini göstermektedir.
- **Pozitif doğru oranı (TP Rate) :** Belirli bir sınıf için doğru tahmin edilen örnek sayısının, gerçekte o sınıfta yer alan örneklem sayısına oranıdır.
- **Yanlış pozitif oranı (FP Rate):** Tahminleme sonucunda bir sınıf içerisinde yanlış sınıflandırılmış örneklerin gerçekte ilgili sınıfta yer alan örnek sayısına oranıdır.
- **Hassasiyet (Precision):** Gerçekte x sınıfına mensup örnek sayısının, tahminleme sonucunda x sınıfında gösterilen örneklere oranıdır.
- **F – ölçüsü (F-measure) :**  $\frac{2 * \text{Hassasiyet} * \text{Pozitif Doğru oranı}}{\text{Hassasiyet} + \text{Pozitif doğru oranı}}$  şeklinde tanımlanmış özel bir ölçüdür. (Waikato University, 2015)
- **Karışıklık matrisi (Confusion matrix) :**  $\begin{vmatrix} x & y \\ z & t \end{vmatrix}$  ifade edilen matriste diagonal değişkenlerin toplamı (x+t) doğru sınıflandırılan örneklerin sayısını, y ve z değişkenleri ise yanlış sınıflandırılan örnekleri temsil etmektedir. Kullanılan yöntemin performansının değerlendirilmesinde önemli bir göstergedir.
- **Alıcı işlem karakteristiği alanı (ROC Area):** Basit anlamıyla doğru pozitiflerin yanlış pozitiflere oranı olarak ifade edilebilir (Swets, 1996). Bu değer 1'e yaklaştıkça pozitiflerin negatiflerden ayrılma oranı mükemmelleşmektedir.

Buna göre tasarıma ait test sonuçları şu şekilde gerçekleşmiştir. Bağımsız değişkenlerden (BOLUM, SEHIR, GENEL\_OZYETERLIK vs.) yola çıkarak bağımlı değişkenin (İkinci Dönem Başarı durumu) tahmin edilmesinde ilk olarak ÇKA kullanılmıştır. Tahminleme

çalışması sonucu başarı durumları için sınıflandırma performans göstergeleri şu şekilde oluşmuştur.

```
Time taken to build model: 19.67 seconds
```

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

```
Correctly Classified Instances      237          53.2584 %
Incorrectly Classified Instances    208          46.7416 %
Kappa statistic                    0.2826
Mean absolute error                 0.3119
Root mean squared error             0.5049
Relative absolute error             71.5973 %
Root relative squared error        108.1772 %
Total Number of Instances          445
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.55	0.175	0.567	0.55	0.558	0.805	1
	0.516	0.384	0.5	0.516	0.508	0.59	2
	0.54	0.171	0.549	0.54	0.545	0.768	3
Weighted Avg.	0.533	0.263	0.533	0.533	0.533	0.703	

```
=== Confusion Matrix ===
```

```
 a  b  c  <-- classified as
72 48 11 | a = 1
48 98 44 | b = 2
 7 50 67 | c = 3
```

### Şekil 39 : Çok Katmanlı Algılayıcı ile yapılan test çalışması sonuçları

Kurulan tahmin modeli geçerli bir sonuç üretmiş ve test aşamasından geçmiştir.

ÇKA 'nın test edilmesinden sonra Rassal Ormanlar Yöntemi kullanılarak tahmin modeli oluşturulmuştur.

```

Time taken to build model: 0.25 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      282          63.3708 %
Incorrectly Classified Instances    163          36.6292 %
Kappa statistic                    0.4316
Mean absolute error                 0.3513
Root mean squared error             0.409
Relative absolute error             80.6317 %
Root relative squared error         87.634 %
Total Number of Instances          445

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.618   0.092   0.736     0.618   0.672     0.876    1
                0.663   0.365   0.575     0.663   0.616     0.659    2
                0.605   0.128   0.647     0.605   0.625     0.856    3
Weighted Avg.   0.634   0.218   0.643     0.634   0.635     0.778

=== Confusion Matrix ===

  a  b  c  <-- classified as
81 46  4 |  a = 1
27 126 37 |  b = 2
 2  47 75 |  c = 3

```

#### Şekil 40: Rassal Ormanlar ile yapılan test çalışması sonuçları

Çapraz geçерleme yöntemiyle oluşturulan model tahmin sonuçları üretebilmiştir. Dolayısıyla test aşamasında yapısal bir soruna sahip olmadığı savunulabilir.

Test aşamasında son olarak Çokterimli Lojistik Regresyon analizi ile tahminleme modeli oluşturulmuştur.

Time taken to build model: 0.44 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	244	54.8315 %
Incorrectly Classified Instances	201	45.1685 %
Kappa statistic	0.3065	
Mean absolute error	0.3092	
Root mean squared error	0.4789	
Relative absolute error	70.9821 %	
Root relative squared error	102.6219 %	
Total Number of Instances	445	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.565	0.162	0.592	0.565	0.578	0.821	1
	0.516	0.388	0.497	0.516	0.506	0.593	2
	0.581	0.159	0.585	0.581	0.583	0.794	3
Weighted Avg.	0.548	0.258	0.55	0.548	0.549	0.716	

=== Confusion Matrix ===

```
 a b c <-- classified as
74 50 7 | a = 1
48 98 44 | b = 2
3 49 72 | c = 3
```

### Şekil 41: Çokterimli Lojistik Regresyon test aşaması sonuçları

Bu sonuçlarla beraber çalışmada akademik başarı durumunu tahmin etmek amacıyla seçilen tüm yöntemler mevcut verilerle belirli tahminlerde bulunmuşlardır. Böylelikle seçilen yöntemlerin modelleme aşamasında kullanılmasına karar verilmiştir.

#### 2.2.4.3. Model kur

Başarı durumunun tahmin edilmesine yönelik modelde özellik seçimi, bağımsız değişkenlerin bağımlı değişken ile aralarında olan korelasyon katsayısına göre yapılacaktır. Buna göre bağımlı değişken ile arasında anlamlı bir ilişki gözlemlenmeyen değişkenlerin veriseti dışında bırakılmasına karar verilmiştir. Bu korelasyon temelli özellik seçimi yöntemini test edebilmek için ise, tüm bağımsız değişkenlerin parametre olarak kabul edildiği orijinal veriseti, karşılaştırma amaçlı korunmuştur. Böylelikle korelasyon katsayısına göre özellik seçiminin tahminleme modelinin performansına olumlu anlamda katkı yapıp yapmadığı araştırılmıştır. Buna göre:

**Veriseti1:** Yalnızca bağımlı deęişken ile aralarında anlamlı bir ilişki gözlemlenen bağımsız deęişkenler tahminleme modelinde kullanılacaktır. Bu doğrultuda bağımlı deęişken İkinci Dönem Başarı Durumu ile aralarında anlamlı ilişkiler saptanamayan YAS, KARDES\_SAYISI, HAZIRLIK, EKSTRA\_GELIR, KALINAN\_YER, MADDI\_KAYGI, BOSLUK, HEDEF\_EGITIM SEVIYESI, LISE\_TURU, ONEM\_SIRA, BOYUT\_2\_DERECE ve BOYUT\_3\_DERECE deęişkenleri verisetinden çıkarılmıştır.

```
Relation:    academic_performance-weka.filters.unsupervised.attribute.Remove-R4-5,7,12-13,15-17,21-22,30-31,33
Instances:   445
Attributes:  21
            ID
            BOLUM
            EGITIM
            CINSIYET
            SEHIR
            BABA_EGITIM
            ANNE_EGITIM
            GELIR
            SIGARA_KULLANIM
            LISE_NOT_ORTALAMASI
            LISE_BASARI_DURUMU
            LYS_SIRA
            ILK_DONEM_NOT_ORTALAMASI
            ILK_DONEM_BASARI_DURUMU
            GIRIS_DERSI
            DEVAMSIZLIK
            GOY_DERECE
            AOY_DERECE
            BOYUT1_DERECE
            BOYUT4_DERECE
            BASARI_DURUMU
Test mode:10-fold cross-validation
```

**Şekil 42: Korelasyon temelli özellik seçiminden sonra oluşan Veriseti1’de kullanılan deęişkenler**

**Veriseti2 :** Tüm deęişkenler tahminleme modeline dahil edilmiştir.

```
Relation:      academic_performance-weka.filters.unsupervised.attribute.Remove-R1,33
Instances:    445
Attributes:   32
              BOLUM
              EGITIM
              HAZIRLIK
              YAS
              CINSIYET
              KARDES_SAYISI
              SEHIR
              BABA_EGITIM
              ANNE_EGITIM
              GELIR
              EKSTRA_GELIR
              KALINAN_YER
              SIGARA_KULLANIM
              MADDI_KAYGI
              BOSLUK
              LISE_TURU
              LISE_NOT_ORTALAMASI
              LISE_BASARI_DURUMU
              LYS_SIRA
              ONEM_SIRA_1
              HEDEF_EGITIM_SEVIYESI
              ILK_DONEM_NOT_ORTALAMASI
              ILK_DONEM_BASARI_DURUMU
              GIRIS_DERSI
              DEVAMSIZLIK
              GOY_DERECE
              AOY_DERECE
              BOYUT1_DERECE
              BOYUT2_DERECE
              BOYUT3_DERECE
              BOYUT4_DERECE
              BASARI_DURUMU
```

**Şekil 43: Orijinal verisetinin korunduęu Veriseti2 de kullanılan deęişkenler**

Buna göre çalışmanın amacına uygun olarak seçilen yöntemler ve veri setlerine göre modeller şu şekilde oluşturulmuştur:

- **Model1 : Veriseti1 + Çok Katmanlı Algılayıcılar (ÇKA) (Öğrenme oranı  $\alpha$  : 0,3)**
- **Model2 : Veriseti1 + Rassal Ormanlar (RF) (Ağaç sayısı: 100)**
- **Model3: Veriseti1 + Çokterimli Lojistik Regresyon (Benzerlik Ceza puanı: 1.0E-8)**
- **Model4: Veriseti2 + Çok Katmanlı Algılayıcılar**
- **Model5: Veriseti2 + Rassal Ormanlar**
- **Model6: Veriseti2 + Çokterimli Lojistik Regresyon**

Sınıflandırıcı olarak seçilen yöntemlerin kullandığı verisetlerinin %66 'sı sınıflandırma modelini oluşturabilmek üzere “eğitim seti”, kalan %34 ise kurulan modelin geçerliliğini test edebilmek için “test seti” olarak ayrılmıştır.

#### 2.2.4.4. Model değerlendirme

Bu aşamada öncelikle korelasyon katsayılarına göre modele dahil edilen değişkenleri içeren Veriseti1 kullanılarak sınıflandırıcılar (Model1, Model2, Model3) test edilmiştir. Model1'e ait sonuçlar şu şekildedir:

```

Time taken to build model: 8.38 seconds

=== Predictions on test data ===

inst#,      actual, predicted, error, probability distribution
  1         3:3      3:3      0      0.005 *0.995
  2         3:3      2:2      + 0      *0.826 0.174
  3         3:3      3:3      0      0.001 *0.999
  4         3:3      3:3      0      0.122 *0.878
  5         3:3      3:3      0      0.014 *0.986
  6         3:3      2:2      + 0      *0.999 0
  7         3:3      2:2      + 0      *0.999 0.001
  8         3:3      2:2      + 0.398 *0.558 0.044
  9         3:3      3:3      0      0      *1
 10         3:3      3:3      0      0.108 *0.892
 11         3:3      2:2      + 0      *0.857 0.143
 12         3:3      3:3      0      0.138 *0.862
 13         3:3      2:2      + 0      *0.96 0.039
 14         2:2      2:2      0.01   *0.952 0.037
 15         2:2      2:2      0.001  *0.858 0.141
 16         2:2      2:2      0.064  *0.936 0
 17         2:2      2:2      0      *0.995 0.005
 18         2:2      1:1      + *0.915 0.085 0
 19         2:2      3:3      + 0      0.096 *0.904
 20         2:2      2:2      0.362  *0.638 0
 21         2:2      2:2      0      *0.962 0.038
 22         2:2      2:2      0      *1 0
 23         2:2      2:2      0.095  *0.905 0
 24         2:2      1:1      + *0.749 0.251 0

```

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      252          56.6292 %
Incorrectly Classified Instances    193          43.3708 %
Kappa statistic                    0.3291
Mean absolute error                 0.3023
Root mean squared error            0.4943
Relative absolute error             69.384 %
Root relative squared error        105.9072 %
Total Number of Instances          445
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.58	0.14	0.633	0.58	0.606	0.79	1
	0.589	0.392	0.528	0.589	0.557	0.609	2
	0.516	0.153	0.566	0.516	0.54	0.805	3
Weighted Avg.	0.566	0.251	0.57	0.566	0.567	0.717	

```
=== Confusion Matrix ===
```

```
  a  b  c  <-- classified as
76 48  7 |  a = 1
36 112 42 |  b = 2
 8  52 64 |  c = 3
```

#### Şekil 44: Model1 için örnek çıktılar ve sonuçlar

Öncelikle ÇKA'nın örneklerin sınıflarına yönelik yaptığı tahmin değerlerinin karşılaştırıldığı örnek bir tablo sunulmuştur. Buna göre tahminleme modeli ilk örneğin sınıfını doğru tahmin etmiştir. Model1 toplamda akademik başarı durumu üzerine yapılan tahminleme çalışmasında 445 veriden 252 tanesini (%56,62) doğru olarak sınıflandırmıştır. En yüksek doğru sınıflandırma oranı (TPrate) BASARI\_DURUMU =”Orta” olan öğrenciler için gerçekleştirilmiştir. (%58,9)



Time taken to build model: 0.24 seconds

=== Predictions on test data ===

inst#	actual	predicted	error	probability distribution
1	3:3	3:3		0.04 0.169 *0.791
2	3:3	2:2	+	0.046 *0.571 0.383
3	3:3	3:3		0.005 0.24 *0.755
4	3:3	3:3		0.016 0.208 *0.776
5	3:3	3:3		0.02 0.276 *0.704
6	3:3	2:2	+	0.06 *0.649 0.291
7	3:3	3:3		0.024 0.285 *0.691
8	3:3	2:2	+	0.163 *0.574 0.264
9	3:3	3:3		0.044 0.132 *0.824
10	3:3	3:3		0.03 0.282 *0.688
11	3:3	2:2	+	0.077 *0.56 0.363
12	3:3	3:3		0.03 0.461 *0.509
13	3:3	3:3		0.092 0.424 *0.483
14	2:2	1:1	+	*0.561 0.429 0.011
15	2:2	2:2		0.109 *0.469 0.423
16	2:2	1:1	+	*0.381 0.318 0.302
17	2:2	3:3	+	0.032 0.387 *0.581
18	2:2	1:1	+	*0.605 0.381 0.014
19	2:2	2:2		0.15 *0.77 0.08
20	2:2	1:1	+	*0.654 0.316 0.03
21	2:2	2:2		0.044 *0.524 0.432
22	2:2	2:2		0.058 *0.706 0.236
23	2:2	2:2		0.323 *0.623 0.053
24	2:2	2:2		0.344 *0.554 0.102

=== Stratified cross-validation ===  
=== Summary ===

Correctly Classified Instances	270	60.6742 %
Incorrectly Classified Instances	175	39.3258 %
Kappa statistic	0.3931	
Mean absolute error	0.3336	
Root mean squared error	0.4096	
Relative absolute error	76.5783 %	
Root relative squared error	87.7651 %	
Total Number of Instances	445	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.595	0.118	0.678	0.595	0.634	0.865	1
	0.605	0.365	0.553	0.605	0.578	0.639	2
	0.621	0.14	0.631	0.621	0.626	0.856	3
Weighted Avg.	0.607	0.229	0.612	0.607	0.608	0.766	

=== Confusion Matrix ===

a	b	c	<-- classified as
78	48	5	a = 1
35	115	40	b = 2
2	45	77	c = 3

Şekil 45: Model2 için örnek çıktılar ve sonuçlar

Her biri 5 rassal özellik seçilerek oluşturulan 100 ağaçla uygulanan RF yöntemine ait tahmin edilen değerler ile gerçekleşmiş değerler sunulmuştur. “error” sütununun altında yer alan “+” işareti örneğin tahmininin gerçek değerle uyuşmadığını göstermektedir. Model2 için tahminlemede kullanılan Rassal Ormanlar yöntemi 445 adet veriden 270 tanesini (%60,67) doğru sınıflandırabilmiştir. Karışıklık matrisinden de anlaşılacağı üzere bu yöntem en başarılı tahminlerini “BASARI\_DURUMU = “YÜKSEK” olanlar için yapmıştır. Bu sınıf için doğru sınıflandırma oranı %62,1 olarak gerçekleşmiştir.

```
Time taken to build model: 0.34 seconds
```

```
=== Predictions on test data ===
```

inst#,	actual,	predicted,	error,	probability	distribution
1	3:3	3:3	0.067	0.191	*0.741
2	3:3	3:3	0	0.461	*0.539
3	3:3	3:3	0	0.048	*0.952
4	3:3	3:3	0	0.068	*0.932
5	3:3	3:3	0	0.049	*0.951
6	3:3	3:3	0	0.287	*0.713
7	3:3	3:3	0	0.06	*0.94
8	3:3	2:2	+	0.139	*0.668 0.193
9	3:3	3:3	0.006	0.063	*0.932
10	3:3	3:3	0	0.079	*0.921
11	3:3	3:3	0	0.472	*0.528
12	3:3	3:3	0.003	0.192	*0.806
13	3:3	2:2	+	0.029	*0.718 0.252
14	2:2	1:1	+	*0.513	0.478 0.009
15	2:2	2:2	0.097	*0.796	0.107
16	2:2	2:2	0.414	*0.536	0.05
17	2:2	2:2	0.002	*0.618	0.38
18	2:2	1:1	+	*0.546	0.433 0.021
19	2:2	2:2	0.139	*0.664	0.198
20	2:2	1:1	+	*0.588	0.399 0.013
21	2:2	2:2	0.002	*0.708	0.29
22	2:2	2:2	0	*0.818	0.182
23	2:2	2:2	0.278	*0.716	0.005
24	2:2	2:2	0.178	*0.691	0.131

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      259           58.2022 %
Incorrectly Classified Instances    186           41.7978 %
Kappa statistic                    0.3541
Mean absolute error                 0.3075
Root mean squared error            0.4389
Relative absolute error            70.5828 %
Root relative squared error        94.0527 %
Total Number of Instances          445

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.573   0.146    0.62     0.573   0.595     0.833    1
                0.584   0.388    0.529    0.584   0.555     0.62     2
                0.589   0.128    0.64     0.589   0.613     0.846    3
Weighted Avg.   0.582   0.244    0.587    0.582   0.583     0.746

=== Confusion Matrix ===

  a  b  c  <-- classified as
75  51  5 |  a = 1
43 111 36 |  b = 2
 3  48 73 |  c = 3

```

#### Şekil 46: Model3 için örnek çıktılar ve sonuçlar

Çokterimli Lojistik Regresyon analizi tahminlere ait örnek çıktıda 24 örnekten 19 unun sınıfını doğru tahmin etmiştir. Toplamda ise 445 örneğin 259 tanesini doğru sınıflandırmıştır. Buna göre Veriset1 için, Çokterimli Lojistik Regresyon Analizi %60,67 doğru sınıflandırma oranına ulaşmıştır. Diğer bir deyişle, Çokterimli Lojistik Regresyon örnekleme yer alan her 10 öğrenciden 6 sınıf başarı durumunu korelasyon temelli Veriset1 için doğru tahmin etmiştir.

Model4, Model5 ve Model6 için tahminleme sonuçları, test aşamasında sunulan Şekil. 39, Şekil.40 ve Şekil.41 de sunulmuştur.

Bağımlı değişken ile aralarındaki korelasyon katsayısına göre özelliklerin seçildiği Veriset1 ile oluşturulan modeller ile orijinal verisetinin kullanıldığı modellerin doğru sınıflandırma oranlarının karşılaştırılması akademik performans tahmininde korelasyon temelli özellik seçiminin geçerliliği konusunda fikir vermesi açısından önemlidir. Oluşturulan modeller için doğru sınıflandırma performansları şu şekildedir:

**Tablo 14**

**Oluşturulan modeller için doğru sınıflandırma oranları**

	<b>Doğru sınıflandırma oranı (%)</b>
<b>Model1</b>	56,62
<b>Model2</b>	60,67
<b>Model3</b>	58,20
<b>Model4</b>	53,25
<b>Model5</b>	63,37
<b>Model6</b>	54,83

Model2 diğer korelasyon temelli verilere göre daha başarılı sonuçlar üretmiştir. Doğru sınıflandırma oranlarından yola çıkarak, ÇKA ve Çokterimli Lojistik Regresyon yöntemlerinin korelasyon temelli özellik seçimi yaklaşımıyla elde edilen verilerle daha isabetli sonuçlar ürettiği gözlemlenmiştir. Özellikle korelasyon temelli yaklaşımla ÇKA'nın birlikte kullanıldığı Model1'de doğru sınıflandırma oranında yaklaşık %4 lük bir iyileşme sağlanmıştır. Dikkat edilmesi gereken bir diğer husus Rassal Ormanlar yönteminin tüm değişkenlerin modele dahil olduğu Model5 de daha doğru tahminlerde bulunduğudır. Korelasyon temelli özellik seçimi yaklaşımının önemli özellikleri kendisi belirleyen (Breiman, 2001) RF için uygun bir seçim olmayacağı öne sürülebilir.

İlgili literatürde incelenen çalışmalarda doğru sınıflandırma oranları %38.46 (Al-Radaideh vd., 2006) ve %94.41 (Mishra vd., 2014) arasında değişmektedir. 445 öğrenciyle yapılan mevcut çalışmada en yüksek doğru sınıflandırma oranı Model2 için %60,67 olarak bulunmuştur. Bu oran “kabul edilebilir” aralıkta yer almakla birlikte yeterli olarak düşünülmemektedir. Doğru sınıflandırma oranlarının mevcut çalışma için iyi bir modelin alt sınırı olarak belirlenen %70 değerinin altında kalmasının veri madenciliği çalışmalarında sıkça rastlanan verisetindeki sınıf dengesizliği ve bağımlı değişkenin 2 den fazla değer almasından kaynaklanabileceği düşünülerek modeller iyileştirilmeye çalışılmıştır. Bu noktada veri dengesizliği problemini ve geliştirilen çözüm yollarını tanımlamak yerinde olacaktır.

**Sınıf dengesizliği problemi:** Akademik başarı durumlarının tahminine yönelik çok sınıflı değişkenlerin tahmininde çok sık rastlanan (Al-Radaideh vd., 2006; Ramaswami&Bhaskaran, 2010) Sınıf dengesizliği problemi, önceden belirlenmiş

sınıflara mensup örneklem sayısının diğer sınıflarda yer alan örnek sayılarından daha fazla sayıda olması durumunda ortaya çıkmaktadır (Japkowicz&Stephen, 2000: s. 1). Gerçek dünya uygulamalarında sıklıkla karşılaşılan bu problem makine öğrenmesi ve veri madenciliği çalışmalarında yetersiz sınıflandırma performanslarına neden olmaktadır (Guo vd., 2008: s.192). Bu durumun nedenlerinden ilki, modern sınıflandırıcıların, kendilerinden sınıflandırılmaları istenen görünmeyen veri noktalarının eğitim verileriyle aynı dağılımı gösterdiklerini kabul etmeleridir (Japkowicz, 2000). Eğer eğitim ve geçerleme verileri farklı dağılımlardan seçilirse, bu durumda da eğitilmiş sınıflandırıcı kusurlu modelden dolayı yetersiz sonuçlar verecektir (Forman&Cohen, 2004). İkinci neden ise, çoğu modern sınıflandırıcı eğitim verilerinde spesifik bir kayıp fonksiyonunun değerini optimize etmeye çalışmaktadır. Bu kayıp fonksiyonları, azınlık sınıflarındaki doğru tahminleme oranını genel doğruluk oranına genellemektedir. Bu nedenle, bir sınıflandırıcı yalnızca ifade edilen amaca yönelik doğru tahminleme oranını maksimize etmektedir (Japkowicz, 2000).

Bu doğrultuda sınıf dengesizliği probleminin üstesinden gelebilmek için yapılan çalışmalar 3 ana başlık altında toplanmıştır: Sınıf dağılımını değiştirme, Maliyete duyarlı öğrenme, Değerlendirme metrikleri.

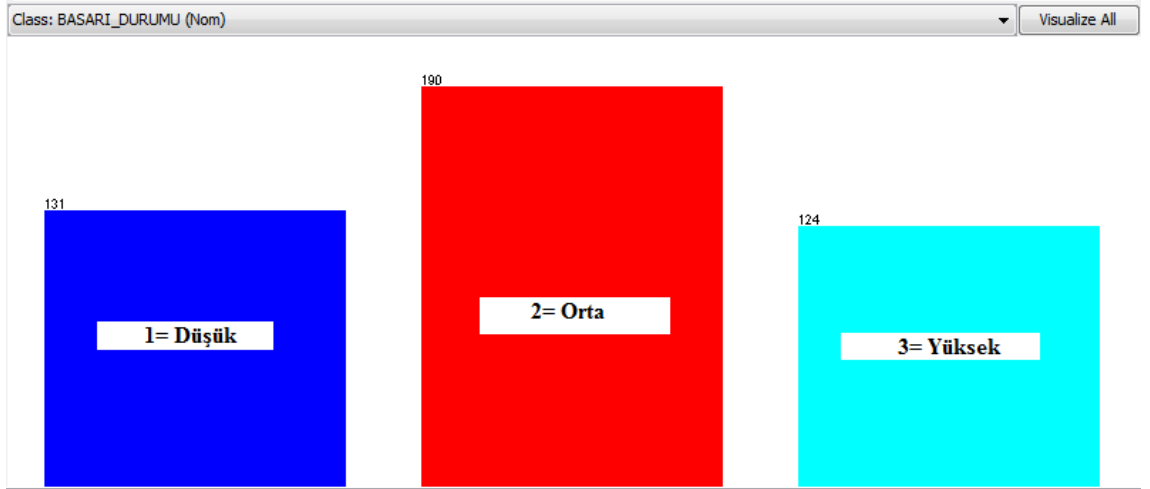
Mevcut çalışmada sınıf dağılımını değiştirme stratejisiyle verisetini dengeleme yoluna gidilmiştir. Akademik performans tahminleme çalışmalarında (Thammasiri vd. ,2014; Jayaprakash&Lauria, 2014). Chawla vd.’nin (2002) geliştirmiş oldukları sınıf dağılımını değiştirme suretiyle veri dengelemeyi amaçlayan SMOTE (Synthetic Minority Over-Sampling) başvurulan yöntemlerden biridir. SMOTE “veri uzayından ziyade “özellik uzayı” nda sentetik örnekler üretmektedir. SMOTE’nin işleyişi genel hatlarıyla şu şekildedir:

- Arttırılmak istenen örnek miktarı, artışın yapılacağı sınıfın mevcut sayısının yüzdesi şeklinde belirtilir.
- İstenen miktara bağlı olarak k adet (mevcut çalışma için k=5) en yakın komşu rassal olarak seçilir.
- Sentetik olarak arttırılması istenen özellik vektörü ile en yakın komşusu arasındaki vektörün farkı alınır.
- Bu fark 0 ile 1 arasında yer alan rassal bir sayı ile çarpılır.

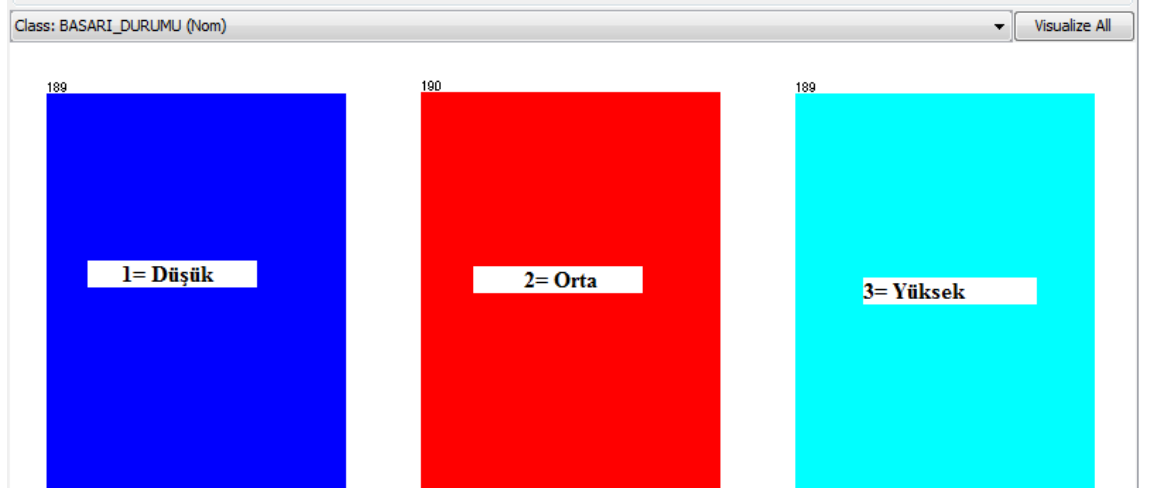
- Seçilen vektörün üzerine eklenen bu değer yeni bir sentetik veri oluşturmaktadır (Chawla vd. ,2002: s. 328) .

Hangi aralıkta ve miktarda veri üretileceğinin analizi yapan araştırmacı tarafından belirlenmesi yönüyle bu yöntem “Öğreticili” bir yöntem olarak düşünülebilir.

Mevcut çalışmada yararlanılan verisetleri için verinin dengelenmeden önceki ve sonraki halleri şekillerde belirtilmiştir.



Şekil 47: Dengelenmemiş veriseti



Şekil 48: Dengelenmiş veriseti

Yeni verisetinde BASARI\_DURUMU değişkenleri “1= Düşük” ve “3=Yüksek” sınıfına ait örneklem sayısı “2= Orta “ sınıfında yer alan örnek sayısına eşit olacak şekilde “sentetik” olarak arttırılmıştır.

Tüm sınıflarda yer alan örnek sayılarının yaklaşık olarak birbiriyle aynı sayıda olduğu varsayımı yapılarak veri dengeleme yapılması durumu “Senaryo1” olarak düşünülebilir.

Kurulan modeller SMOTE ile dengelenmiş verisetinin kullanıldığı Senaryo1 için test edilmiştir. İlk olarak, geleneksel istatistiksel sınıflandırıcılardan Çokterimli Lojistik Regresyon yöntemini kullanan Model3 test edilmiştir. Bu modele ait sonuçlar şekilde gösterilmiştir.

```
Time taken to build model: 0.22 seconds
```

```
=== Predictions on test data ===
```

```
inst#,      actual, predicted, error, probability distribution
  1         3:3       3:3          0.004 0.018 *0.977
  2         3:3       3:3          0      0.362 *0.638
  3         3:3       2:2          + 0.073 *0.717 0.21
  4         3:3       3:3          0      0.002 *0.998
  5         3:3       3:3          0      0.028 *0.972
  6         3:3       3:3          0      0.039 *0.961
  7         3:3       2:2          + 0.022 *0.722 0.256
  8         3:3       3:3          0.002 0.029 *0.969
  9         3:3       2:2          + 0      *0.888 0.112
 10         3:3       3:3          0      0.382 *0.618
 11         3:3       2:2          + 0.037 *0.787 0.176
 12         3:3       2:2          + 0.075 *0.606 0.319
 13         3:3       2:2          + 0.127 *0.606 0.267
 14         3:3       3:3          0.004 0.018 *0.978
 15         3:3       3:3          0.013 0.124 *0.863
 16         3:3       3:3          0.005 0.06  *0.935
 17         3:3       3:3          0      0.147 *0.853
 18         3:3       3:3          0.065 0.33  *0.605
 19         3:3       3:3          0      0.006 *0.994
 20         2:2       1:1          + *0.803 0.185 0.012
 21         2:2       2:2          0.048 *0.591 0.361
 22         2:2       2:2          0.005 *0.62 0.375
 23         2:2       2:2          0.058 *0.697 0.245
 24         2:2       3:3          + 0      0.383 *0.617
```

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      374          65.8451 %
Incorrectly Classified Instances    194          34.1549 %
Kappa statistic                     0.4877
Mean absolute error                 0.2656
Root mean squared error             0.4012
Relative absolute error             59.7673 %
Root relative squared error         85.1002 %
Total Number of Instances          568
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.73	0.145	0.715	0.73	0.723	0.884	1
	0.521	0.246	0.516	0.521	0.518	0.678	2
	0.725	0.121	0.749	0.725	0.737	0.897	3
Weighted Avg.	0.658	0.171	0.66	0.658	0.659	0.819	

```
=== Confusion Matrix ===
```

```
  a  b  c  <-- classified as
138 44  7 |  a = 1
 52 99 39 |  b = 2
  3 49 137 |  c = 3
```

### Şekil 49: Senaryo1 için Model3'e ait örnek çıktılar ve sonuçlar

Korelasyon katsayılarına göre bağımsız değişkenlerin dahil edildiği dengelenmiş verisiyle yapılan Çokterimli Lojistik Regresyon analizi sonucunda, 568 örnekten 374 tanesi doğru sınıflara atanmıştır. Dengelenmemiş verisinde %58,20 olan doğru sınıflandırma oranı, veriler dengelendikten sonra %65,84 olarak gerçekleşmiştir. Dolayısıyla sınıfların dağılımının değiştirilmesi yoluyla dengeleme işlemi yapan SMOTE, Çokterimli Lojistik Regresyona ait doğru sınıflandırma oranını %7,64 oranında arttırmıştır. Dengelenmiş sınıflara ait doğru sınıflandırma oranlarına bakıldığında; Çokterimli Lojistik Regresyon başarısı düşük öğrencileri dengelenmemiş verisinde %57,3 ile doğru sınıflandırırken, “sentetik” örneklerle dengelenmiş verisi için bu oran %73 e yükselmiştir. Benzer şekilde başarılı öğrencilerin tahmininde %58,9 doğru sınıflandırma oranı, dengelenmiş veri için %72,5 a ulaşmıştır.

Model1'nin dengelenmiş verisinde göstermiş olduğu doğru sınıflandırma performansı ise Çokterimli Regresyon Analizine benzer bir artış oranı göstermiştir.



Time taken to build model: 10.72 seconds

=== Predictions on test data ===

inst#	actual	predicted	error	probability distribution
1	3:3	3:3	0	0 *0.999
2	3:3	3:3	0	0.005 *0.995
3	3:3	2:2	+	0.029 *0.963 0.008
4	3:3	3:3	0	0.002 0 *0.997
5	3:3	3:3	0	0.009 0 *0.991
6	3:3	3:3	0	0 *1
7	3:3	2:2	+	0 *0.988 0.012
8	3:3	3:3	0	0 *1
9	3:3	3:3	0	0.496 *0.504
10	3:3	3:3	0	0.067 *0.933
11	3:3	2:2	+	0.005 *0.791 0.203
12	3:3	3:3	0	0.019 0.002 *0.979
13	3:3	2:2	+	0.009 *0.991 0
14	3:3	3:3	0	0.006 0 *0.994
15	3:3	3:3	0	0 *1
16	3:3	3:3	0	0.001 0.001 *0.999
17	3:3	3:3	0	0.003 *0.997
18	3:3	2:2	+	0 *0.994 0.006
19	3:3	3:3	0	0 *1
20	2:2	1:1	+	*0.97 0.03 0
21	2:2	3:3	+	0.01 0 *0.99
22	2:2	2:2	0	*0.928 0.072
23	2:2	1:1	+	*0.67 0.318 0.012
24	2:2	2:2	0	0.001 *0.844 0.154

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	372	65.493 %
Incorrectly Classified Instances	196	34.507 %
Kappa statistic	0.4824	
Mean absolute error	0.2407	
Root mean squared error	0.4392	
Relative absolute error	54.1587 %	
Root relative squared error	93.1728 %	
Total Number of Instances	568	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.73	0.142	0.719	0.73	0.724	0.878	1
	0.489	0.222	0.525	0.489	0.507	0.675	2
	0.746	0.153	0.709	0.746	0.727	0.871	3
Weighted Avg.	0.655	0.173	0.651	0.655	0.652	0.808	

=== Confusion Matrix ===

a	b	c	<-- classified as
138	41	10	a = 1
49	93	48	b = 2
5	43	141	c = 3

Şekil 50: Senaryo1 için Model1'e ait örnek çıktılar ve sonuçlar

ÇKA ile dengelenmiş Veriseti1 de yapılan analiz sonucunda, dengelenmemiş verisetinde 445 örnek için 252 adet doğru sınıflandırılan örnek sayısı, 568 örnek için 372' ye yükselmiştir. Örneklem kütle büyüklüklerindeki farklılık göz önüne alındığında doğru sınıflandırma oranlarının karşılaştırılması ile daha mantıklı yargılara ulaşılabacaktır.

Buna göre, dengelenmemiş verisetinde doğru sınıflandırılan örneklerin tüm örneklere oranı %56,62 iken bu rakam dengelenmiş verisetinde %65,49 olmuştur. Bir başka deyişle SMOTE yöntemi ÇKA'ya %8,87'lik bir tahminleme gücü kazandırmıştır. Daha spesifik değerler incelenecek olursa, dengelenmemiş veride ÇKA BASARI\_DURUMU = "1- Düşük" olanların dengelenmemiş veride Doğru Pozitif oranı %58 iken dengelenmiş veriyle bu oran %73'e ulaşmıştır. Benzer şekilde BASARI\_DURUMU= "3-Yüksek" olan örnekler için %51,6 olan doğru sınıflandırılma oranı ise dengeleme işleminden sonra %74,6 olarak gerçekleşmiştir. Bu noktada başarılı olan öğrencilerin tahmin edilmesinde SMOTE Model1 e %23'lük ciddi bir katkı sağlamıştır.

En yüksek performansı gösteren Model2 için dengelenmiş verilerle yapılan analiz sonuçları şu şekilde gerçekleşmiştir.

```
Time taken to build model: 0.13 seconds

=== Predictions on test data ===

inst#,      actual, predicted, error, probability distribution
  1         3:3      3:3          0.023  0.17 *0.807
  2         3:3      3:3          0      0.117 *0.883
  3         3:3      2:2          +  0.05 *0.553  0.397
  4         3:3      3:3          0.04  0.12 *0.84
  5         3:3      3:3          0.043  0.164 *0.793
  6         3:3      3:3          0      0.02 *0.98
  7         3:3      3:3          0.092  0.326 *0.582
  8         3:3      3:3          0      0.01 *0.99
  9         3:3      3:3          0.021  0.36 *0.619
 10         3:3      3:3          0      0.127 *0.873
 11         3:3      2:2          +  0.127 *0.619  0.254
 12         3:3      3:3          0.16  0.28 *0.56
 13         3:3      2:2          +  0.32 *0.368  0.311
 14         3:3      3:3          0.058  0.057 *0.886
 15         3:3      3:3          0      0.04 *0.96
 16         3:3      3:3          0.043  0.172 *0.785
 17         3:3      3:3          0      0.122 *0.878
 18         3:3      3:3          0.275  0.325 *0.4
 19         3:3      3:3          0.088  0.156 *0.755
 20         2:2      1:1          + *0.501  0.448  0.051
 21         2:2      2:2          0.332 *0.376  0.293
 22         2:2      3:3          +  0.045  0.387 *0.568
 23         2:2      2:2          0.3 *0.548  0.152
 24         2:2      2:2          0.049 *0.584  0.367
```

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	394	69.3662 %
Incorrectly Classified Instances	174	30.6338 %
Kappa statistic	0.5406	
Mean absolute error	0.2738	
Root mean squared error	0.3616	
Relative absolute error	61.6139 %	
Root relative squared error	76.7032 %	
Total Number of Instances	568	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.788	0.137	0.741	0.788	0.764	0.924	1
	0.484	0.183	0.571	0.484	0.524	0.765	2
	0.81	0.14	0.743	0.81	0.775	0.927	3
Weighted Avg.	0.694	0.153	0.685	0.694	0.687	0.872	

=== Confusion Matrix ===

a	b	c	<-- classified as
149	34	6	a = 1
51	92	47	b = 2
1	35	153	c = 3

### Şekil 51: Senaryo1 için Model2'ye ait örnek çıktıları ve sonuçları

Dengelenmemiş korelasyon temelli özelliklerin seçildiği veri setiyle %60,67'lik bir doğru sınıflandırma performansı gösteren RF, SMOTE yöntemiyle dengelenmiş veriyle yapılan analiz sonucunda %69,36 doğru sınıflandırma oranına ulaşmıştır. Doğru sınıflandırılan örnek oranlarında sağlanan bu %8,69'luk artış diğer modellerde yaşanan artışa benzerdir. Dengelenmiş sınıflardaki performansa yakından bir göz atılacak olursa; başarısı düşük öğrencilerin doğru sınıflandırılma oranı dengeleme işlemi sonrasında %59,5 dan %78,8 e ulaşmıştır. %19,3 lük bu artış; risk grubundaki öğrencilerin önceden belirlenerek önlemler alınmasını amaçlayan mevcut çalışma için dikkat çekicidir. Başarılı öğrencilerin tahmin edilmesinde doğru sınıflandırma performansı ise %62,1' den %81' e ulaşmıştır. Ancak dengelenmiş sınıflarda yaşanan bu muazzam artış, dengeleme işleminden önce en yüksek örnek sayısına sahip "2=Orta" sınıfındaki doğru

tahminleme oranını %60,5 den %48,4 e düşürmüştür. Ancak diğer sınıflarda yaşanan artışların yarısı kadar bir düşüş ana amaca hizmet açısından bakıldığında tolere edilebilirdir.

Tüm değişkenlerin tahminleme modeline dahil edildiği verisetini kullanan ve sınıflandırıcı olarak ÇKA nın belirlendiği Model4 için dengelenmiş veriseti ile yapılan analiz sonuçları şu şekildedir.

```
Time taken to build model: 24.9 seconds
```

```
=== Predictions on test data ===
```

```
inst#,      actual, predicted, error, probability distribution
  1         3:3         3:3          0.003  0      *0.997
  2         3:3         3:3          0      0.019 *0.981
  3         3:3         2:2          + 0.022 *0.714  0.263
  4         3:3         3:3          0      0.186 *0.814
  5         3:3         3:3          0      0.491 *0.509
  6         3:3         3:3          0      0      *1
  7         3:3         3:3          0      0.267 *0.733
  8         3:3         2:2          + 0      *0.675  0.325
  9         3:3         3:3          0      0.118 *0.882
 10         3:3         3:3          0      0.001 *0.999
 11         3:3         3:3          0.222  0.002 *0.776
 12         3:3         3:3          0.003  0.49  *0.507
 13         3:3         2:2          + 0.093 *0.907  0.001
 14         3:3         3:3          0.001  0      *0.999
 15         3:3         3:3          0.001  0      *0.999
 16         3:3         3:3          0.001  0.001 *0.998
 17         3:3         3:3          0      0.004 *0.996
 18         3:3         3:3          0      0.094 *0.906
 19         3:3         3:3          0      0      *1
 20         2:2         1:1          + *0.756  0.244  0
 21         2:2         3:3          + 0.001  0.002 *0.998
 22         2:2         3:3          + 0.003  0.018 *0.979
 23         2:2         3:3          + 0.008  0.1    *0.893
 24         2:2         2:2          0.428 *0.537  0.035
```

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      375          66.0211 %
Incorrectly Classified Instances    193          33.9789 %
Kappa statistic                     0.4904
Mean absolute error                 0.2432
Root mean squared error             0.438
Relative absolute error             54.7098 %
Root relative squared error         92.9236 %
Total Number of Instances          568
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.794	0.145	0.732	0.794	0.761	0.885	1
	0.447	0.206	0.521	0.447	0.482	0.679	2
	0.741	0.158	0.7	0.741	0.72	0.863	3
Weighted Avg.	0.66	0.17	0.651	0.66	0.654	0.809	

```
=== Confusion Matrix ===
```

```
  a  b  c  <-- classified as
150 32  7 |  a = 1
 52 85 53 |  b = 2
  3 46 140 |  c = 3
```

### Şekil 52: Senaryo1 için Model4'e ait örnek çıktılar ve sonuçlar

Sınıfların dengeli dağıldığını varsayan Senaryo1'e göre tüm değişkenlerin dahil edildiği ve sınıflandırıcı olarak ÇKA'yı kullanan Model4 te yapılan tahmin çalışmasında doğru sınıflandırma oranı 568 adet örnek için %66,0211 olarak bulunmuştur. Sınıfların dengesiz dağıldığı mevcut durumda Model4'ün doğru sınıflandırma performansının %53,25 olduğu düşünüldüğünde, bu oranda gerçekleşen %12,77 'lik artış veri madenciliği çalışmalarında sınıf dengesizliği probleminin önemini ortaya koymaktadır.

Tüm değişkenlerin dahil edildiği ve sınıflandırıcı olarak Rassal Ormanlar (RF) yöntemi kullanan Model5 için analiz sonuçları şu şekilde oluşmuştur.

Time taken to build model: 0.12 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability distribution
1	3:3	3:3		0.022 0.042 *0.936
2	3:3	3:3		0.013 0.103 *0.884
3	3:3	2:2	+	0.145 *0.47 0.385
4	3:3	3:3		0.054 0.183 *0.763
5	3:3	3:3		0.015 0.112 *0.872
6	3:3	3:3		0.003 0.038 *0.96
7	3:3	3:3		0.207 0.242 *0.551
8	3:3	3:3		0.021 0.16 *0.819
9	3:3	3:3		0.123 0.267 *0.611
10	3:3	3:3		0.011 0.156 *0.833
11	3:3	3:3		0.123 0.4 *0.476
12	3:3	3:3		0.151 0.283 *0.565
13	3:3	1:1	+	*0.346 0.322 0.332
14	3:3	3:3		0.101 0.072 *0.827
15	3:3	3:3		0.01 0.101 *0.889
16	3:3	3:3		0.149 0.403 *0.448
17	3:3	3:3		0.038 0.18 *0.782
18	3:3	3:3		0.225 0.376 *0.399
19	3:3	3:3		0.137 0.131 *0.733
20	2:2	1:1	+	*0.48 0.434 0.086
21	2:2	3:3	+	0.335 0.327 *0.338
22	2:2	3:3	+	0.236 0.292 *0.472
23	2:2	2:2		0.317 *0.488 0.196
24	2:2	2:2		0.06 *0.645 0.295

=== Stratified cross-validation ===  
=== Summary ===

Correctly Classified Instances	401	70.5986 %
Incorrectly Classified Instances	167	29.4014 %
Kappa statistic	0.5591	
Mean absolute error	0.3019	
Root mean squared error	0.3679	
Relative absolute error	67.9336 %	
Root relative squared error	78.0387 %	
Total Number of Instances	568	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.831	0.156	0.727	0.831	0.775	0.931	1
	0.474	0.151	0.612	0.474	0.534	0.776	2
	0.815	0.135	0.751	0.815	0.782	0.92	3
Weighted Avg.	0.706	0.147	0.697	0.706	0.697	0.875	

=== Confusion Matrix ===

a	b	c	<-- classified as
157	27	5	a = 1
54	90	46	b = 2
5	30	154	c = 3

Şekil 53: Senaryo1 için Model5'e ait örnek çıktılar ve sonuçlar

Dengelenmemiş Veriseti2 için doğru sınıflandırma performansı %63,37 olarak bulunan Rassal Ormanlar yöntemi, dengeleme işleminden sonra yaklaşık %8 lik bir artışla performansını %70,59 seviyesine çıkartmıştır. Buna paralel olarak başarısı düşük olan öğrencilerin doğru tahmin edilebilmesi %61,8 seviyesinden %83,1 seviyesine yükselmiştir. Bu durumda dengelenmiş veride risk grubundaki öğrencilerin doğru tahmin oranının dikkat çekici bir şekilde artış gösterdiği görülmektedir.

Senaryo1 için son olarak mevcut durumdaki analizlerde en iyi sonuçları veren Çokterimli Lojistik Regresyon analizi yöntemi kullanan Model6 için sonuçlar şu şekilde gerçekleşmiştir.

Time taken to build model: 0.36 seconds

=== Predictions on test data ===

inst#	actual	predicted	error	probability distribution
1	3:3	3:3	0	0.001 *0.999
2	3:3	2:2	+	0 *0.535 0.465
3	3:3	2:2	+	0.029 *0.917 0.053
4	3:3	3:3	0	0.001 *0.999
5	3:3	2:2	+	0 *0.525 0.475
6	3:3	3:3	0	0.063 *0.937
7	3:3	2:2	+	0.018 *0.599 0.383
8	3:3	2:2	+	0.02 *0.519 0.461
9	3:3	3:3	0	0.481 *0.519
10	3:3	3:3	0	0.251 *0.749
11	3:3	2:2	+	0.073 *0.833 0.094
12	3:3	2:2	+	0.008 *0.752 0.24
13	3:3	2:2	+	0.282 *0.605 0.113
14	3:3	3:3	0	0 *1
15	3:3	3:3	0	0.011 *0.989
16	3:3	3:3	0	0.001 *0.999
17	3:3	3:3	0	0.009 *0.991
18	3:3	2:2	+	0.01 *0.775 0.215
19	3:3	3:3	0	0 *1
20	2:2	1:1	+	*0.986 0.014 0
21	2:2	3:3	+	0.011 0.291 *0.699
22	2:2	3:3	+	0.003 0.11 *0.886
23	2:2	2:2	0.02	*0.87 0.11
24	2:2	2:2	0	*0.69 0.31

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      345          60.7394 %
Incorrectly Classified Instances    223          39.2606 %
Kappa statistic                     0.4111
Mean absolute error                 0.2722
Root mean squared error            0.4413
Relative absolute error             61.2527 %
Root relative squared error        93.6044 %
Total Number of Instances          568
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.688	0.166	0.674	0.688	0.681	0.866	1
	0.432	0.262	0.453	0.432	0.442	0.645	2
	0.704	0.161	0.686	0.704	0.695	0.864	3
Weighted Avg.	0.607	0.196	0.604	0.607	0.605	0.792	

```
=== Confusion Matrix ===
```

```
  a  b  c  <-- classified as
130 49 10 |  a = 1
 57 82 51 |  b = 2
  6 50 133 |  c = 3
```

#### Şekil 54: Senaryo1 için Model6'ya ait örnek çıktılar ve sonuçlar

Tüm değişkenlerin dahil edildiği dengelenmemiş verisetinde %54,83 lük bir doğru sınıflandırma performansı gösteren Model6, verilerin dengelenmesiyle birlikte doğru sınıflandırılan örneklerin oranını %60,73 e çıkarmıştır. Başarılı öğrencilerin tahmininde %70,4'lük oran dikkat çekicidir. Verilerin dengeli dağılması halinde veri madenciliği yöntemlerinin performansının arttığı gözlemlenmiştir. Bu durum (Japkowicz , 2000; Forman&Cohen, 2004; Thammasiri vd. , 2014) çalışmalarında savunulan görüşleri destekler niteliktedir. Mevcut durum ve Senaryo1 için modellerin doğru sınıflandırma performansları şu şekilde özetlenebilir.

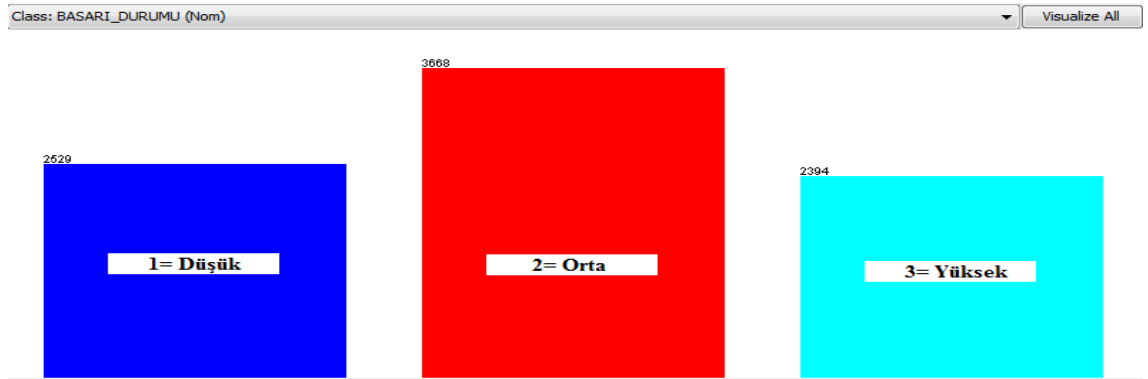


**Tablo 15**

**Mevcut durum ve Senaryo1 için modellerin doğru sınıflandırma performansları**

	<b>Mevcut Durum(%)</b>	<b>Senaryo1(%)</b>
<b>Model1</b>	56,62	65,49
<b>Model2</b>	60,67	69,36
<b>Model3</b>	58,20	65,84
<b>Model4</b>	53,25	66,02
<b>Model5</b>	63,37	70,59
<b>Model6</b>	54,83	60,73

2014-2015 Akademik yılında Sakarya Üniversitesi İşletme Fakültesi'nde yer alan lisans programlarında ilk yıllarını geçiren öğrencilere uygulanan "Yeni Gelen Öğrenci" anketinin ilerleyen çalışmalarda üniversitede ilk yıllarını geçiren tüm öğrencilere uygulanması düşünülmektedir. Böylesine geniş çaplı bir uygulamaya geçilme kararı verilmeden önce araştırmanın yönünü belirleyebilmek için bir simülasyon çalışması yapılması uygun görülmüştür. Buna göre, sadece İşletme Fakültesi'nde yer alan öğrencilere uygulanan anketin tüm üniversiteye uygulanması durumunda alınacak olası sonuçlar "Senaryo2" başlığı altında toplanmıştır. Sakarya Üniversitesi Bilişim departmanından edinilen bilgiye göre Sakarya Üniversitesi lisans programlarına 2014-2015 akademik yılında 8592 öğrenci kayıt yaptırmıştır. Buna göre elde edilen başarı dağılımının sabit kalacağı varsayılarak, her bir sınıfta yer alan örnek sayısı SMOTE yöntemiyle artırılmıştır. Her bir sınıftaki örnek sayısının 8592/445  $\approx$ 19,30 kat artırılmasıyla ortaya çıkan Senaryo2 için sınıf dağılımları şu şekilde oluşmuştur.



**Şekil 55: Senaryo2 için öğrencilerin başarı durumları dağılımları**

Elde edilen yeni veriler için modellerin doğru sınıflandırma performansları yeniden test edilerek üniversite çapında bir çalışmada olası sonuçlara yönelik ipuçları elde edilmeye çalışılmıştır.

Senaryo2 için ilk olarak sınıflandırıcı olarak Çok Katmanlı Algılayıcı tekniğini kullanan Model1'in performansı analiz edilmiştir.

```
Time taken to build model: 164.18 seconds
```

```
=== Predictions on test data ===
```

```
inst#,      actual, predicted, error, probability distribution
  1         2:2      2:2      0      *1      0
  2         2:2      2:2      0      *1      0
  3         2:2      2:2      0      *0.999 0.001
  4         2:2      2:2      0      *0.997 0.003
  5         2:2      2:2      0.001 *0.999 0
  6         2:2      2:2      0.001 *0.999 0
  7         2:2      2:2      0      *1      0
  8         2:2      2:2      0      *0.999 0.001
  9         2:2      2:2      0      *1      0
 10         2:2      2:2      0      *1      0
 11         2:2      2:2      0      *1      0
 12         2:2      2:2      0      *1      0
 13         2:2      2:2      0      *1      0
 14         2:2      2:2      0      *0.999 0.001
 15         2:2      2:2      0.001 *0.999 0
 16         2:2      2:2      0.006 *0.994 0
 17         2:2      2:2      0      *0.865 0.135
 18         2:2      2:2      0      *1      0
 19         2:2      2:2      0.002 *0.998 0
 20         2:2      2:2      0      *1      0
 21         2:2      2:2      0      *1      0
 22         2:2      2:2      0      *1      0
 23         2:2      2:2      0      *1      0
 24         2:2      2:2      0.001 *0.999 0
```

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      8380          97.578 %
Incorrectly Classified Instances    208           2.422 %
Kappa statistic                     0.9629
Mean absolute error                 0.0177
Root mean squared error             0.1154
Relative absolute error             4.0636 %
Root relative squared error         24.7267 %
Total Number of Instances          8588
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.98	0.009	0.979	0.98	0.98	0.995	1
	0.971	0.019	0.974	0.971	0.973	0.99	2
	0.978	0.01	0.974	0.978	0.976	0.995	3
Weighted Avg.	0.976	0.013	0.976	0.976	0.976	0.993	

```
=== Confusion Matrix ===
```

```
  a   b   c  <-- classified as
2478  45   5 |   a = 1
  49 3561  57 |   b = 2
   3   49 2341 |   c = 3
```

### Şekil 56 : Senaryo2 için Model1'e ait örnek çıktılar ve sonuçlar

Tüm öğrencilerin “Yeni Gelen Öğrenci Anketi”ne katıldığı varsayıldığında doğru sınıflandırma oranı Model1 için %97,578 olarak gerçekleşmiştir. Bu oran göz önüne alındığında büyük verisetleri için ÇKA'nın uygun bir yöntem olduğu öne sürülebilir.

Korelasyon katsayılarına göre değişkenlerin verisetinde dahil edildiği Model2 nin Senaryo2 için doğru sınıflandırma performansı şu şekilde gerçekleşmiştir.

Time taken to build model: 1.78 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability	distribution
1	1:1	1:1	*1	0	0
2	1:1	1:1	*1	0	0
3	1:1	1:1	*1	0	0
4	1:1	1:1	*1	0	0
5	1:1	1:1	*1	0	0
6	1:1	1:1	*1	0	0
7	1:1	1:1	*1	0	0
8	1:1	1:1	*0.503	0.327	0.169
9	1:1	1:1	*0.99	0.01	0
10	1:1	1:1	*1	0	0
11	1:1	1:1	*0.99	0.01	0
12	1:1	1:1	*1	0	0
13	1:1	1:1	*1	0	0
14	1:1	1:1	*1	0	0
15	1:1	1:1	*1	0	0
16	1:1	1:1	*1	0	0
17	1:1	1:1	*1	0	0
18	1:1	1:1	*1	0	0
19	1:1	1:1	*1	0	0
20	1:1	1:1	*1	0	0
21	1:1	1:1	*1	0	0
22	1:1	1:1	*0.933	0.057	0.01
23	1:1	1:1	*1	0	0
24	1:1	1:1	*1	0	0

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8454	98.4053 %
Incorrectly Classified Instances	137	1.5947 %
Kappa statistic	0.9756	
Mean absolute error	0.0188	
Root mean squared error	0.0867	
Relative absolute error	4.3191 %	
Root relative squared error	18.5745 %	
Total Number of Instances	8591	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.004	0.99	0.986	0.988	1	1
	0.985	0.015	0.98	0.985	0.982	0.999	2
	0.981	0.006	0.984	0.981	0.983	1	3
Weighted Avg.	0.984	0.009	0.984	0.984	0.984	1	

=== Confusion Matrix ===

a	b	c	<-- classified as
2493	33	3	a = 1
22	3612	34	b = 2
3	42	2349	c = 3

### Şekil 57: Senaryo2 için Model2'ye ait örnek çıktılar ve sonuçlar

Tüm yeni gelen öğrencilerin “Yeni Gelen Öğrenci Anketi” ne katıldığı varsayılan durumda Rassal Ormanlar yöntemiyle tahminlemede bulunan Model2'nin doğru sınıflandırma performansı %98,40 olarak elde edilmiştir. Bir başka deyişle 8591 öğrenciden 8454 tanesinin göstereceği muhtemel performanslar doğru bir şekilde tahmin edilebilmiştir.

Çokterimli Lojistik Regresyonun tüm öğrencilerin ankete katılımı durumunda göstereceği muhtemel doğru sınıflandırma performansı ise şu şekilde oluşmuştur.

```

Time taken to build model: 3.4 seconds

=== Predictions on test data ===

inst#,      actual, predicted, error, probability distribution
  1         1:1       1:1      *0.936 0.064 0
  2         1:1       1:1      *1      0      0
  3         1:1       1:1      *0.853 0.147 0
  4         1:1       1:1      *0.978 0.022 0
  5         1:1       1:1      *1      0      0
  6         1:1       1:1      *0.997 0.003 0
  7         1:1       2:2      + 0.484 *0.516 0
  8         1:1       1:1      *0.893 0.107 0
  9         1:1       1:1      *0.709 0.163 0.128
 10         1:1       1:1      *0.772 0.228 0
 11         1:1       1:1      *1      0      0
 12         1:1       1:1      *0.999 0.001 0
 13         1:1       1:1      *1      0      0
 14         1:1       1:1      *0.789 0.138 0.073
 15         1:1       1:1      *0.826 0.174 0
 16         1:1       1:1      *1      0      0
 17         1:1       1:1      *0.943 0.057 0
 18         1:1       1:1      *0.899 0.101 0
 19         1:1       1:1      *0.851 0.149 0
 20         1:1       1:1      *0.999 0.001 0
 21         1:1       1:1      *0.999 0.001 0
 22         1:1       1:1      *1      0      0
 23         1:1       1:1      *1      0      0
 24         1:1       1:1      *0.962 0.038 0

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8036          93.5398 %
Incorrectly Classified Instances    555           6.4602 %
Kappa statistic                     0.9011
Mean absolute error                  0.0765
Root mean squared error              0.1856
Relative absolute error              17.5605 %
Root relative squared error          39.7604 %
Total Number of Instances           8591

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.94    0.024    0.943     0.94    0.941     0.988     1
          0.928    0.056    0.924     0.928   0.926     0.974     2
          0.942    0.021    0.945     0.942   0.943     0.99      3
Weighted Avg.  0.935    0.037    0.935     0.935   0.935     0.982

=== Confusion Matrix ===

  a   b   c  <-- classified as
2377 143   9 |  a = 1
 141 3404 123 |  b = 2
   4  135 2255 |  c = 3

```

Şekil 58: Senaryo2 için Model3'e ait örnek çıktılar ve sonuçlar

Görüldüğü üzere yeni gelen tüm öğrencilerin ankete katıldığı varsayılırsa mevcut durumda korelasyon katsayısına bağlı özellik seçimini kullanan Model3'ün %58,20 olan doğru sınıflandırma oranı %93,53'e yükselmektedir. Risk grubunda yer alacağı varsayılan 2529 öğrencinin 2377 sinin doğru sınıflandırılması çalışma açısından önemlidir.

Tüm bağımsız değişkenlerin İkinci Dönem Başarı Durumu üzerinde etkili olduğunu kabul eden Veriseti2 ile çalışan ÇKA (Model4)' nın Senaryo2 için ürettiği sonuçlar şu şekildedir.

Time taken to build model: 374.49 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability	distribution
1	1:1	1:1	*0.999	0	0.001
2	1:1	1:1	*0.998	0.001	0.001
3	1:1	1:1	*0.999	0.001	0
4	1:1	1:1	*0.999	0	0.001
5	1:1	1:1	*0.998	0.002	0
6	1:1	1:1	*0.999	0	0.001
7	1:1	1:1	*1	0	0
8	1:1	1:1	*0.936	0.064	0
9	1:1	1:1	*0.997	0.001	0.002
10	1:1	1:1	*0.999	0	0.001
11	1:1	1:1	*1	0	0
12	1:1	1:1	*1	0	0
13	1:1	1:1	*0.999	0	0.001
14	1:1	1:1	*0.998	0.001	0.002
15	1:1	1:1	*0.999	0	0.001
16	1:1	1:1	*1	0	0
17	1:1	1:1	*0.995	0.002	0.003
18	1:1	1:1	*1	0	0
19	1:1	1:1	*0.999	0.001	0
20	1:1	1:1	*0.999	0	0.001
21	1:1	1:1	*0.999	0.001	0
22	1:1	1:1	*1	0	0
23	1:1	1:1	*0.998	0.002	0
24	1:1	1:1	*0.999	0	0.001

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8426           98.0794 %
Incorrectly Classified Instances    165           1.9206 %
Kappa statistic                    0.9706
Mean absolute error                0.0138
Root mean squared error            0.1054
Relative absolute error            3.1675 %
Root relative squared error        22.5784 %
Total Number of Instances          8591

=== Detailed Accuracy By Class ===

                TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
                0.981     0.008     0.981     0.981     0.981     0.996     1
                0.979     0.016     0.979     0.979     0.979     0.994     2
                0.982     0.006     0.984     0.982     0.983     0.996     3
Weighted Avg.   0.981     0.011     0.981     0.981     0.981     0.995

=== Confusion Matrix ===

  a   b   c  <-- classified as
2482  40   7 |   a = 1
  44 3592  32 |   b = 2
   4   38 2352 |   c = 3

```

### Şekil 59: Senaryo2 için Model4'e ait örnek çıktılar ve sonuçlar

Tüm ilk yıl öğrencilerine anketin uygulandığı varsayılan durumda ÇKA 'nın doğru sınıflandırma oranı %98.07 olarak gerçekleşmiştir.

Mevcut verisetindeki tüm bağımsız değişkenlerin tahminleme modeline dahil edildiği verisetini kullanan Model5'in analiz sonuçları şu şekilde ortaya çıkmıştır.

```

Time taken to build model: 1.75 seconds

=== Predictions on test data ===

inst#,      actual, predicted, error, probability distribution
  1         1:1         1:1         *1         0         0
  2         1:1         1:1         *0.99      0         0.01
  3         1:1         1:1         *1         0         0
  4         1:1         1:1         *1         0         0
  5         1:1         1:1         *0.99      0.01      0
  6         1:1         1:1         *1         0         0
  7         1:1         1:1         *1         0         0
  8         1:1         1:1         *0.531     0.328     0.142
  9         1:1         1:1         *1         0         0
 10         1:1         1:1         *1         0         0
 11         1:1         1:1         *1         0         0
 12         1:1         1:1         *1         0         0
 13         1:1         1:1         *1         0         0
 14         1:1         1:1         *0.99      0         0.01
 15         1:1         1:1         *1         0         0
 16         1:1         1:1         *1         0         0
 17         1:1         1:1         *1         0         0
 18         1:1         1:1         *1         0         0
 19         1:1         1:1         *1         0         0
 20         1:1         1:1         *1         0         0
 21         1:1         1:1         *1         0         0
 22         1:1         1:1         *0.762     0.154     0.084
 23         1:1         1:1         *1         0         0
 24         1:1         1:1         *1         0         0

```

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      8471          98.6032 %
Incorrectly Classified Instances    120           1.3968 %
Kappa statistic                     0.9786
Mean absolute error                 0.0172
Root mean squared error             0.0825
Relative absolute error             3.9547 %
Root relative squared error         17.6725 %
Total Number of Instances          8591
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.985	0.004	0.99	0.985	0.987	1	1
	0.987	0.013	0.983	0.987	0.985	1	2
	0.986	0.005	0.987	0.986	0.987	1	3
Weighted Avg.	0.986	0.008	0.986	0.986	0.986	1	

```
=== Confusion Matrix ===
```

```

 a   b   c  <-- classified as
2490 34   5 |   a = 1
 23 3620 25 |   b = 2
   3   30 2361 |   c = 3
```

### Şekil 60: Senaryo2 için Model5'e ait örnek çıktılar ve sonuçlar

Sınıflandırma aracı olarak Rassal Ormanlar (RF) yöntemini kullanan Model5 8591 öğrenciden 8471 (%98,6032) tanesinin sınıflarını doğru olarak tayin etmiştir. Risk grubunda yer alan öğrencilerin yer aldığı başarı durumu düşük olan öğrencilerin ise %98,5 oranında doğru tahmin edilmiştir.

Senaryo2 için son olarak Model6 ya ait analiz sonuçları ise şu şekildedir.

```
Time taken to build model: 9.55 seconds
```

```
=== Predictions on test data ===
```

inst#	actual	predicted	error	probability distribution
1	1:1	1:1		*0.986 0.014 0
2	1:1	1:1		*1 0 0
3	1:1	1:1		*1 0 0
4	1:1	1:1		*0.996 0.004 0
5	1:1	1:1		*0.995 0.005 0
6	1:1	1:1		*1 0 0
7	1:1	1:1		*0.777 0.223 0
8	1:1	1:1		*1 0 0
9	1:1	1:1		*0.598 0.338 0.065
10	1:1	1:1		*0.644 0.356 0
11	1:1	1:1		*1 0 0
12	1:1	1:1		*1 0 0
13	1:1	1:1		*0.999 0.001 0
14	1:1	1:1		*0.826 0.161 0.013
15	1:1	1:1		*0.99 0.01 0
16	1:1	1:1		*1 0 0
17	1:1	1:1		*0.691 0.309 0
18	1:1	1:1		*0.989 0.011 0
19	1:1	1:1		*0.974 0.026 0
20	1:1	1:1		*0.996 0.004 0
21	1:1	1:1		*1 0 0
22	1:1	1:1		*1 0 0
23	1:1	1:1		*1 0 0
24	1:1	1:1		*0.993 0.007 0



```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8395           97.7185 %
Incorrectly Classified Instances    196            2.2815 %
Kappa statistic                     0.9651
Mean absolute error                 0.0375
Root mean squared error             0.1293
Relative absolute error             8.606 %
Root relative squared error        27.7072 %
Total Number of Instances          8591

=== Detailed Accuracy By Class ===

                TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
                0.98     0.009     0.978     0.98     0.979     0.991     1
                0.973     0.017     0.976     0.973     0.975     0.982     2
                0.98     0.009     0.978     0.98     0.979     0.992     3
Weighted Avg.   0.977     0.013     0.977     0.977     0.977     0.987

=== Confusion Matrix ===

  a   b   c   <-- classified as
2478  43   8 |   a = 1
  52 3570  46 |   b = 2
   4   43 2347 |   c = 3

```

### Şekil 61: Senaryo2 için Model6'ya ait örnek çıktılar ve sonuçlar

Sınıflandırıcı olarak Çokterimli Lojistik Regresyon analizini kullanan Model6'nın gösterdiği doğru sınıflandırma performansı tüm ilk yıl öğrencilerinin ankete katıldığı varsayıldığında %97,715 olarak hesaplanmıştır. Bu analizde dikkat çekilmesi gereken nokta; mevcut durum ve verilerin dengelenmiş olduğu Senaryo1 de korelasyon temelli özellik seçimiyle oluşan Veriseti1 ile daha yüksek sınıflandırma performansı elde eden Çokterimli Lojistik Regresyon analizi, Senaryo2'de tüm bağımsız değişkenlerin modele dahil edildiği Veriseti2 ile daha yüksek bir performans ortaya koymuştur. Mevcut senaryoda Model3'ün doğru sınıflandırma oranı %93,53 olarak gerçekleşmiştir.

Tüm öğrencilerin ankete katıldığı durumda başarı sınıflarının dengeli olarak dağıldığı varsayımı çalışmada "Senaryo3" olarak test edilmiştir. Buna göre düşük, orta ve başarılı öğrencilere ait yeni veri dağılımı şu şekilde gerçekleşmiştir.



**Şekil 62: Senaryo3 için yeni veri dağılımları**

Senaryo3 için Model1 e ait performans şu şekilde gerçekleşmiştir.

Time taken to build model: 144.96 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability	distribution
1	3:3	2:2	+	0	*1 0
2	3:3	3:3		0.001	0 *0.999
3	3:3	3:3		0	0 *1
4	3:3	3:3		0.001	0 *0.999
5	3:3	3:3		0	0 *1
6	3:3	3:3		0	0 *1
7	3:3	3:3		0	0 *1
8	3:3	3:3		0.001	0 *0.999
9	3:3	3:3		0	0.001 *0.999
10	3:3	3:3		0	0 *1
11	3:3	3:3		0	0 *1
12	3:3	3:3		0.001	0 *0.999
13	3:3	3:3		0.019	0 *0.981
14	3:3	3:3		0	0 *1
15	3:3	3:3		0	0 *1
16	3:3	3:3		0	0 *1
17	3:3	3:3		0.004	0.079 *0.917
18	3:3	3:3		0	0 *1
19	3:3	3:3		0	0.004 *0.996
20	3:3	3:3		0	0.001 *0.999
21	3:3	3:3		0	0 *1
22	3:3	3:3		0.001	0 *0.999
23	3:3	3:3		0	0 *1
24	3:3	3:3		0	0.001 *0.999

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      8393          97.7066 %
Incorrectly Classified Instances    197           2.2934 %
Kappa statistic                    0.9656
Mean absolute error                 0.0172
Root mean squared error            0.1143
Relative absolute error            3.8594 %
Root relative squared error        24.2487 %
Total Number of Instances         8590
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.981	0.007	0.987	0.981	0.984	0.995	1
	0.964	0.015	0.97	0.964	0.967	0.992	2
	0.986	0.013	0.974	0.986	0.98	0.996	3
Weighted Avg.	0.977	0.011	0.977	0.977	0.977	0.994	

```
=== Confusion Matrix ===
```

```
  a   b   c  <-- classified as
2811  47   6 |   a = 1
  35 2760  68 |   b = 2
   3   38 2822 |   c = 3
```

### Şekil 63: Senaryo3 için Model1'e ait örnek çıktılar ve sonuçlar

Korelasyon temelli özellik seçimi yaklaşımında sınıflandırıcı olarak ÇKA kullanan Model1 %97.70 lik doğru sınıflandırma oranı yakalamıştır. En yüksek doğru sınıflandırma performansını ise başarılı öğrencilerin tahmininde (TPrate3 = %98.6) göstermiştir.

Öğrencilerin başarı durumlarının eşit dağıldığı varsayılan Senaryo3 için Model2 nin analiz sonuçları şu şekildedir.

Time taken to build model: 1.44 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability distribution
1	3:3	3:3	0.077	0.434 *0.489
2	3:3	3:3	0	0 *1
3	3:3	3:3	0	0 *1
4	3:3	3:3	0	0 *1
5	3:3	3:3	0	0 *1
6	3:3	3:3	0	0 *1
7	3:3	3:3	0	0 *1
8	3:3	3:3	0	0 *1
9	3:3	3:3	0	0.01 *0.99
10	3:3	3:3	0	0.01 *0.99
11	3:3	3:3	0	0 *1
12	3:3	3:3	0	0 *1
13	3:3	3:3	0	0.02 *0.98
14	3:3	3:3	0	0 *1
15	3:3	3:3	0	0 *1
16	3:3	3:3	0	0 *1
17	3:3	3:3	0	0.04 *0.96
18	3:3	3:3	0	0 *1
19	3:3	3:3	0	0 *1
20	3:3	3:3	0	0.01 *0.99
21	3:3	3:3	0	0 *1
22	3:3	3:3	0	0 *1
23	3:3	3:3	0	0 *1
24	3:3	3:3	0	0 *1

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8464	98.5332 %
Incorrectly Classified Instances	126	1.4668 %
Kappa statistic	0.978	
Mean absolute error	0.0188	
Root mean squared error	0.0865	
Relative absolute error	4.2195 %	
Root relative squared error	18.3543 %	
Total Number of Instances	8590	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.004	0.992	0.986	0.989	1	1
	0.981	0.012	0.977	0.981	0.979	0.999	2
	0.988	0.006	0.987	0.988	0.988	1	3
Weighted Avg.	0.985	0.007	0.985	0.985	0.985	0.999	

=== Confusion Matrix ===

a	b	c	<-- classified as
2824	35	5	a = 1
21	2810	32	b = 2
2	31	2830	c = 3

Şekil 64: Senaryo3 Model2'ye ait örnek çıktıları ve sonuçları

Rassal Ormanlar yöntemiyle örnek çıktılarda gösterilen tüm örneklerin sınıflarını doğru tahmin eden Model2, %98,5332 lik bir doğru sınıflandırma performansı göstermiştir. Model2 ye ait 0.999 olarak hesaplanan ROC değeri modelin veriyle mükemmel yakın bir uyum içinde olduğunu göstermektedir.

Korelasyon katsayısına bağlı verisiyle yapılan son tahminleme çalışması olan Model3 için tahmin çıktıları ve sonuçları şu şekilde gerçekleşmiştir.

```
Time taken to build model: 3.18 seconds
```

```
=== Predictions on test data ===
```

```
inst#,      actual, predicted, error, probability distribution
  1         3:3       2:2      + 0      *0.719 0.281
  2         3:3       3:3       0      0.121 *0.879
  3         3:3       3:3       0      0.026 *0.974
  4         3:3       3:3       0      0.047 *0.953
  5         3:3       3:3       0      0.068 *0.932
  6         3:3       3:3       0      0.002 *0.998
  7         3:3       3:3       0      0.022 *0.978
  8         3:3       3:3       0      0      *1
  9         3:3       3:3       0      0.167 *0.833
 10         3:3       3:3       0      0.013 *0.987
 11         3:3       3:3       0      0.003 *0.997
 12         3:3       3:3       0      0      *1
 13         3:3       3:3       0      0.082 *0.918
 14         3:3       3:3       0      0.001 *0.999
 15         3:3       3:3       0      0.001 *0.999
 16         3:3       3:3       0      0      *1
 17         3:3       3:3       0      0.233 *0.767
 18         3:3       3:3       0      0.125 *0.875
 19         3:3       3:3       0      0.071 *0.929
 20         3:3       3:3       0      0.006 *0.994
 21         3:3       3:3       0      0      *1
 22         3:3       3:3       0      0      *1
 23         3:3       3:3       0      0.053 *0.947
 24         3:3       3:3       0      0      *1
```

```
=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances      8071          93.9581 %
Incorrectly Classified Instances    519           6.0419 %
Kappa statistic                     0.9094
Mean absolute error                 0.0757
Root mean squared error             0.1859
Relative absolute error             17.0288 %
Root relative squared error         39.4331 %
Total Number of Instances          8590
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.963	0.029	0.943	0.963	0.953	0.989	1
	0.896	0.036	0.925	0.896	0.91	0.971	2
	0.96	0.025	0.95	0.96	0.955	0.991	3
Weighted Avg.	0.94	0.03	0.939	0.94	0.939	0.984	

```
=== Confusion Matrix ===
```

```
  a   b   c  <-- classified as
2759  96   9 |   a = 1
 164 2564 135 |   b = 2
   4  111 2748 |   c = 3
```

### Şekil 65: Senaryo3 için Model3'e ait örnek çıktılar ve sonuçlar

Veriseti1 ile yapılan sınıflandırma çalışmalarında en düşük doğru sınıflandırma oranına (%93,95) sahip Model3, sınıflandırıcı olarak Çokterimli Lojistik Regresyon yöntemini kullanmıştır. Bu modelde dikkat çeken bir diğer husus “orta” başarılı öğrencilerin tahminlenmesinde diğer modellere oranla yetersiz kalınmasıdır. (%89,6)

Senaryo3 için tüm değişkenlerin modele dahil edildiği Veriseti2 ile yapılan Model4'e göre tahminler şu şekilde oluşmuştur:

Time taken to build model: 368.78 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability	distribution
1	3:3	3:3	0.017	0	*0.983
2	3:3	3:3	0.001	0.003	*0.996
3	3:3	3:3	0	0	*1
4	3:3	3:3	0.001	0.002	*0.996
5	3:3	3:3	0	0	*1
6	3:3	3:3	0	0	*1
7	3:3	3:3	0	0	*1
8	3:3	3:3	0	0	*1
9	3:3	3:3	0.002	0.001	*0.997
10	3:3	3:3	0	0	*1
11	3:3	3:3	0	0	*1
12	3:3	3:3	0	0	*1
13	3:3	3:3	0	0.001	*0.999
14	3:3	3:3	0	0	*1
15	3:3	3:3	0.001	0	*0.999
16	3:3	3:3	0	0	*1
17	3:3	3:3	0	0.002	*0.998
18	3:3	3:3	0	0	*1
19	3:3	3:3	0.002	0.001	*0.997
20	3:3	3:3	0	0.002	*0.998
21	3:3	3:3	0	0	*1
22	3:3	3:3	0	0.001	*0.999
23	3:3	3:3	0	0.001	*0.999
24	3:3	3:3	0	0	*1

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8450	98.3702 %
Incorrectly Classified Instances	140	1.6298 %
Kappa statistic	0.9756	
Mean absolute error	0.012	
Root mean squared error	0.0969	
Relative absolute error	2.7033 %	
Root relative squared error	20.5525 %	
Total Number of Instances	8590	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0.008	0.985	0.988	0.986	0.999	1
	0.975	0.01	0.98	0.975	0.977	0.995	2
	0.988	0.007	0.986	0.988	0.987	0.998	3
Weighted Avg.	0.984	0.008	0.984	0.984	0.984	0.997	

=== Confusion Matrix ===

a	b	c	<-- classified as
2830	27	7	a = 1
40	2791	32	b = 2
4	30	2829	c = 3

Şekil 66: Senaryo3 için Model4'e ait örnek çıktılar ve sonuçlar

ÇKA ile tüm öğrencilerin başarı yönünden eşit dağıldığı varsayılan Senaryo3'e göre Model4 %98,37'lik bir doğruluk oranına ulaşmıştır. Karışıklık matrisi yorumlandığında ise, en fazla öğrencinin başarısının doğru olarak tahmin edildiği sınıf olarak 1="Düşük" görülmektedir.

Senaryo3 için, korelasyon temelli özellik seçimiyle elde edilen Veriseti1 ile yapılan çalışmada en yüksek performansı gösteren Rassal Ormanlar için analiz sonuçları şu şekildedir.

```
Time taken to build model: 1.39 seconds
```

```
=== Predictions on test data ===
```

```
inst#,      actual, predicted, error, probability distribution
  1         3:3       3:3       0.076  0.367 *0.557
  2         3:3       3:3       0       0       *1
  3         3:3       3:3       0       0       *1
  4         3:3       3:3       0.01   0       *0.99
  5         3:3       3:3       0       0.03   *0.97
  6         3:3       3:3       0       0       *1
  7         3:3       3:3       0       0       *1
  8         3:3       3:3       0       0       *1
  9         3:3       3:3       0       0       *1
 10         3:3       3:3       0       0       *1
 11         3:3       3:3       0       0       *1
 12         3:3       3:3       0       0       *1
 13         3:3       3:3       0       0       *1
 14         3:3       3:3       0       0       *1
 15         3:3       3:3       0.01   0.01   *0.98
 16         3:3       3:3       0       0       *1
 17         3:3       3:3       0       0       *1
 18         3:3       3:3       0       0.01   *0.99
 19         3:3       3:3       0       0       *1
 20         3:3       3:3       0       0       *1
 21         3:3       3:3       0       0       *1
 22         3:3       3:3       0       0       *1
 23         3:3       3:3       0       0       *1
 24         3:3       3:3       0       0       *1
```



```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8503           98.9872 %
Incorrectly Classified Instances     87            1.0128 %
Kappa statistic                     0.9848
Mean absolute error                  0.0176
Root mean squared error              0.0764
Relative absolute error              3.9498 %
Root relative squared error          16.2083 %
Total Number of Instances           8590

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.995   0.006   0.988     0.995   0.992     1         1
                0.983   0.005   0.989     0.983   0.986     1         2
                0.992   0.004   0.993     0.992   0.992     1         3
Weighted Avg.   0.99    0.005   0.99     0.99    0.99     1

=== Confusion Matrix ===

  a  b  c  <-- classified as
2851 12  1 |  a = 1
  30 2813 20 |  b = 2
   5  19 2839 |  c = 3

```

### Şekil 67: Senaryo3 için Model5'e ait örnek çıktılar ve sonuçlar

Tüm değişkenleri kullanan Veriseti2 yi kullanan Model5, öğrencileri %98,55 doğruluk oranında başarı sınıflarına ayırmıştır. Bu model ayrıca risk grubunda yer alan öğrencileri yaklaşık %99 oranında tahmin edebilmiştir.

Son olarak geleneksel istatistiksel sınıflandırma yöntemlerinden Çokterimli Lojistik Regresyon Analizine göre ise sonuçlar şu şekilde ortaya çıkmıştır.

Time taken to build model: 10.73 seconds

=== Predictions on test data ===

inst#,	actual,	predicted,	error,	probability distribution
1	3:3	3:3	0	0.033 *0.967
2	3:3	2:2	+	0.041 *0.623 0.336
3	3:3	3:3	0	0.002 *0.998
4	3:3	2:2	+	0.053 *0.654 0.293
5	3:3	3:3	0	0.007 *0.993
6	3:3	3:3	0	0.081 *0.919
7	3:3	3:3	0	0.032 *0.968
8	3:3	3:3	0	0.06 *0.94
9	3:3	3:3	0.005	0.022 *0.973
10	3:3	3:3	0	0 *1
11	3:3	3:3	0	0 *1
12	3:3	3:3	0	0.213 *0.787
13	3:3	3:3	0	0.25 *0.75
14	3:3	3:3	0	0.006 *0.994
15	3:3	3:3	0.001	0.022 *0.977
16	3:3	3:3	0	0 *1
17	3:3	3:3	0	0.413 *0.587
18	3:3	3:3	0	0.001 *0.999
19	3:3	3:3	0.181	0.001 *0.818
20	3:3	3:3	0	0.2 *0.799
21	3:3	3:3	0	0.001 *0.999
22	3:3	3:3	0	0.001 *0.999
23	3:3	3:3	0	0.128 *0.872
24	3:3	3:3	0.004	0.002 *0.993

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8075	94.0047 %
Incorrectly Classified Instances	515	5.9953 %
Kappa statistic	0.9101	
Mean absolute error	0.0718	
Root mean squared error	0.1851	
Relative absolute error	16.1646 %	
Root relative squared error	39.2727 %	
Total Number of Instances	8590	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.971	0.018	0.964	0.971	0.968	0.991	1
	0.907	0.038	0.923	0.907	0.915	0.971	2
	0.942	0.034	0.933	0.942	0.937	0.986	3
Weighted Avg.	0.94	0.03	0.94	0.94	0.94	0.983	

=== Confusion Matrix ===

a	b	c	<-- classified as
2781	60	23	a = 1
95	2597	171	b = 2
8	158	2697	c = 3

Şekil 68: Senaryo3 için Model6'ya ait örnek çıktılar ve sonuçlar

Tüm bağımsız değişkenlerin dahil edildiği Veriseti2 ile yapılan analizlerde doğru sınıflandırma oranı en düşük model, %94.0047 ile Model6 olmuştur. Ancak düşük öğrencileri tahmin etmede göstermiş olduğu performans itibariyle (%97,1) dikkat çekicidir.

Tüm bu analizlerden sonra, modelleri değerlendirebilmek ve en yüksek performansa sahip modeli belirleyebilmek üzere 3 performans kriteri belirlenmiştir:

- **Doğru Sınıflandırma Oranı (DSO):** Öğrencilerin başarı durumlarına bağlı profillerini ortaya çıkarması açısından önem arz etmektedir. Bu durum, üniversite yönetimlerinin alacağı tedbirlerin optimal bir şekilde spesifikleştirilmesini sağlayacaktır. Pek çok çalışmada birincil performans göstergesi olarak kullanılmaktadır (Vandamme vd., 2007; Delen, 2010; Thammasiri vd., 2014).
- **Başarısı düşük öğrencileri doğru sınıflandırma oranı- Risk Grubu Belirleyebilme gücü (RGB):** Muhtemel olarak başarısı düşük öğrencilerin önceden tahmin edilebilmesi, ilgili tedbirlerin zamanında alınabilmesi yönünden önemlidir. Bu sınıfa ait doğru sınıflandırma oranının kriter olarak belirlenmesinin nedeni, veri dağılımına ve kullanılan sınıflandırıcıya bağlı olarak genel sınıflandırma performansında önemli düzeylerde farklılık gösterebilmesidir. Örneğin, Senaryo3 için Model6'nın genel doğru sınıflandırma oranı %87,95 iken, aynı modelin başarısı düşük öğrencileri doğru sınıflandırma performansı %93,1 olarak gerçekleşmiştir. Çalışmamızın amacına uygun olarak bir performans göstergesi olarak kabul edilmiştir.
- **Modeli oluşturmak üzere harcanan bilgisayar zamanı:** Modellerin genel ve özel doğru sınıflandırma oranlarının eşit veya birbirine yakın olması durumunda karar vermek amacıyla kullanılacak bir performans göstergesi olarak belirlenmiştir. Her ne kadar gelişen teknoloji ile birim zamanda işlem yapma yetenekleri gelişen bilgisayarlarla farklı yöntemler için bu süreler arasındaki fark azalsa da, finansal boyut göz önüne alınarak, tahminleme çalışmasının hemen her bilgisayarda uygulanabilmesi açısından önem arz etmektedir. Zira, “doğaçlama” raporlar üretmesi istenen karar destek sisteminin en kısa zamanda sonuç vermesi kullanıcılar tarafından arzu edilen bir özellik olarak düşünülebilir.

Modellerin deęişik daęılım ve örneklem sayıları altında göstermiş oldukları performansları deęerlendirebilmek için Aęırlıklı Performans kriteri belirlenmiştir. Bu açıdan tüm öğrencilerin başarı durumlarının doęru olarak belirlenebilmesi birincil performans göstergesi olarak kabul edilmiş ve keyfi olarak 0,75 katsayısıyla aęırlıklandırılmıştır. Model deęerlendirme aşamasının ikincil performans göstergesi ise, başarı durumu düşük öğrencilerin doęru sınıflandırılma oranıdır. Aęırlık olarak 0,25 katsayısı uygun görülmüştür. Bu iki göstergenin mevcut durum ve senaryolardaki oranlarının ortalamalarının aęırlık katsayılarıyla çarpılıp toplanması sonucu modellerin performans puanları elde edilmiştir. Modellere ait performans puanları şu şekilde hesaplanmıştır.

**Tablo 16**  
**Modellerin performans puanları**

	Model1		Model2		Model3		Model4		Model5		Model6	
	DSO(%)	RGB(%)	DSO(%)	RGB(%)	DSO(%)	RGB(%)	DSO(%)	RGB(%)	DSO(%)	RGB(%)	DSO(%)	RGB(%)
<b>Mevcut durum</b>	56,62	58	60,67	59,5	58,20	57,3	53,25	55	63,37	61,8	54,83	56,5
<b>Senaryo1</b>	65,49	73	69,36	78,8	65,84	73	66,02	79,4	70,59	83,1	60,73	68,8
<b>Senaryo2</b>	97,57	98	98,40	98,6	93,53	94	98,07	98,1	98,60	98,5	97,71	98
<b>Senaryo3</b>	97,70	98,1	98,53	98,6	93,95	96,3	98,37	98,8	98,98	99,5	94,00	97,1
<b>Performans puanı (0,75*ODSO+0,25* ORGB)</b>	79,95		82,27		78,44		79,90		<b>83,595</b>		77,6	

Tablodan da görüldüğü üzere hesaplanan performans puanlarına göre; mevcut durum ve farklı örneklem sayıları ve dağılımları için en yüksek performansı, 83,595 puanla Model5 göstermiştir. Model5, tüm bağımsız değişkenlerin bağımlı değişken olan BASARI\_DURUMU üzerinde etkili olduğunu varsayan Veriseti2 üzerinde performansı geliştirmek adına Böl-Fethet stratejisini benimsemiş (Benyamin, 2012) “topluluk” yöntemlerinden biri olan Rassal Ormanlar (RF) yöntemini kullanmıştır. Burada dikkat edilecek bir diğer nokta, bu yöntemin korelasyon temelli özelliklerin seçildiği Veriseti1 de (Model2) aynı performansı gösterememesidir. Bu durum, Rassal Ormanlarda yer alan ağaçlar arasındaki korelasyonun artmasıyla hata oranının artacağını belirten Breiman’ın (2001) görüşünü desteklemektedir. Zira, korelasyon katsayısına göre özellik seçimi yapılarak, oluşturulabilecek ağaç sayısı azaltılmış ve kendi içlerinde korelasyonları yüksek olabilecek değişkenler modele dahil edilmiştir. Örneklem sayısının artmasıyla DSO ve RGB deki gözle görülür artış ise, Büyük Sayılar Yasası (Law of Large Numbers) ile izah edilebilir.

Çokterimli Lojistik Regresyon analizi, örneklem sayılarının diğer durumlara göre az olduğu mevcut durum ve Senaryo1 de Çok Katmanlı Algılayıcıya göre çok daha iyi performans sergilemesine rağmen, örneklem sayısının artmasıyla performans anlamında diğer yöntemlerin gerisinde kalmıştır. Buna dayanarak, klasik bir istatistiksel sınıflandırma tekniği olan Çokterimli Lojistik Regresyonun hacmi gün geçtikçe büyüyen “Big Data” ile sınıflandırma yapma konusunda makine öğrenme tekniklerine göre yetersiz kaldığı savunulabilir.

Mevcut çalışmada en başarılı 2. Performansı 82,27 puanla Model2 sergilemiştir. Sınıflandırıcı olarak, yine Rassal Ormanlar Yöntemini kullanan model, “topluluk” tekniklerinin sınıflandırma konusundaki muazzam yeteneğini ortaya koymuştur. Sınıflandırıcı olarak karmaşık veya belirgin olmayan verilerden anlam çıkarma konusunda üstün yeteneğe sahip olan ve eğitildiği alanda “uzman” olarak nitelendirilebilecek (Massachusetts Institute of Technology, 2015) Çok Katmanlı Sinir Ağları kullanan Model1 en iyi 3. doğru sınıflandırma performansını sergilemiştir. Korelasyon temelli özellik seçiminin yapıldığı Veriseti1 ile çalışan ÇKA, genel olarak tüm değişkenlerin tahminleme modeline dahil edildiği Model4 e göre daha başarılı

performans sergilemiştir. Buradan yola çıkarak girdi/çıkış arasındaki korelasyonun ÇKA'nın performansını olumlu anlamda etkilediği savunulabilir.

Performans olarak, 83,593 puanla en yüksek puana sahip Model5'e ait performans zamanları şu şekildedir.

**Tablo 17**  
**Modellere ait bilgisayar zamanları**

	<b>Model5 (sn.)</b>
<b>Mevcut durum</b>	0,25
<b>Senaryo1</b>	0,12
<b>Senaryo2</b>	1,75
<b>Senaryo3</b>	1,39
<b>Ortalama süre</b>	0,87

Bilgisayar zamanları kısa bir süre içerisinde "Ad Hoc" raporlar elde edebilmek açısından önemlidir. Örneğin; ÇKA kullanan Model1, süre ve sistem gereksinimleri düşünüldüğüne büyük ölçekli çalışmalar ve veriler için elverişli bir yöntem olarak kabul edilemez. Kronometre ile yapılan ölçüm sonucunda Model1'in Weka Explorer üzerinde yapılan analiz ile ilgili sonuçları üretmesi 60 dakika almaktadır. Dolayısıyla tüm bu noktalar göz önüne alındığında, ortalama olarak 0,87 saniyede model kurarak, 5 saniye içerisinde rapor verebilen Model5'in çalışmanın amacına en uygun model olduğu söylenebilir.

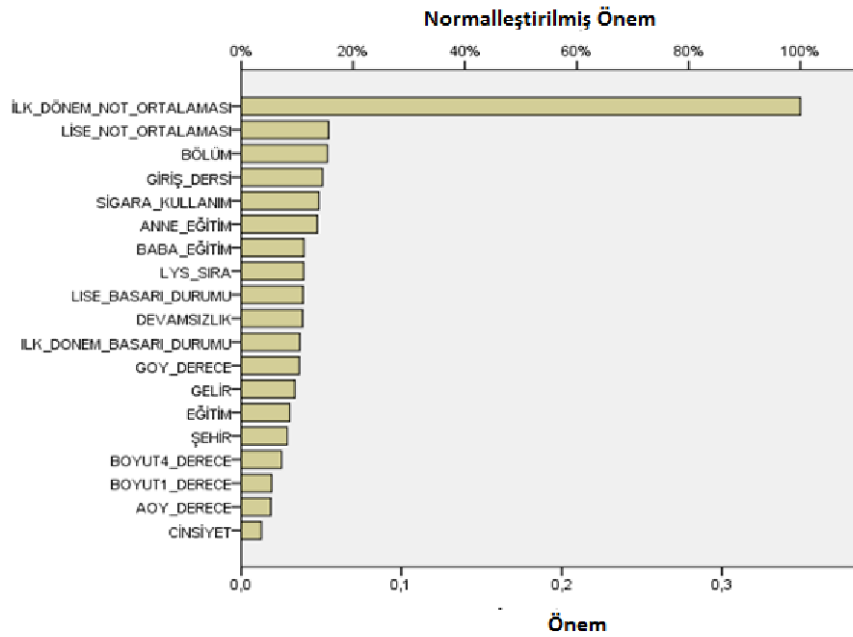
### **2.2.5 Değerlendirme**

Son yayılım aşamasına geçmeden önce, seçilen modellerin daha kapsamlı olarak ele alındığı aşamadır. Bu aşamada, modelin beklenen hedefi ne denli karşılayabildiği ortaya konur ve nihai karara varmadan önce önemli bir noktanın göz ardı edilip edilmediği kontrol edilir. (Wirth&Hipp, 2000: s. 6)

#### **2.2.5.1. Sonuçları değerlendir**

Yapılan analizler ve değerlendirme sonucunda Model5'in (Tüm bağımsız değişkenlerin modele dahil edildiği Veriseti2 + Rassal Ormanlar Yöntemi) en yüksek performansı gösterdiği görülmüştür. Model5'in doğru sınıflandırma oranlarında örneklem sayısının

artmasıyla birlikte gözlenen olağanüstü artış göze çarpmaktadır. Bu durum çalışmanın kapsamının genişletilmesi için olumlu bir işaret olarak yorumlanabilir. Farklı durumlar için denenen modeller arasında kullanılacak en iyi ikinci model ise 82,27 performans puanıyla Model2 olmuştur. Bu modelde ise, Model5 de olduğu gibi, korelasyon katsayısına bağlı olarak seçilen bağımsız değişkenlerden oluşan Veriseti1 ve “topluluk” yöntemlerinden “Rassal Ormanlar” sınıflandırıcı olarak kullanılmıştır. Çalışmada en yüksek 3. puana sahip Model1 için, IBM SPSS Statistics 20 paket programı “duyarlılık analizi” yapılmasına olanak vermektedir. Duyarlılık analizi, tahmin modelinde girdi/çıktı arasındaki nedensellik ilişkisinin saptanabilmesine yönelik bir yöntemdir (Davis Jr., 1989). Bu analizin ardında yatan temel düşünce, herhangi bir bağımsız değişkenin modele dahil edilmemesi durumunda eğitilmiş modelin hata oranındaki değişimi ölçebilmektir. Buna göre, bir model bir değişkene karşı ne kadar duyarlı ise, o değişken model için o denli önemlidir (Delen, 2010: s. 503). Buna göre, mevcut durumda Model1 için elde edilen duyarlılık analizi sonuçları şu şekildedir.



**Şekil 69: Mevcut durumda Model1'e ait duyarlılık analizi sonuçları**

y-ekseninde korelasyon temelli yaklaşımla tahminleme modeline dahil edilmiş değişkenler (önem sırasına göre büyükten küçüğe doğru sıralanmış şekilde) , x-ekseninde ise önem oranları yer almaktadır. Buna göre alt ekseninde yer alan bağımsız



değişkenin önemini, üst x-ekseninde yer alan değişken ise göreceli önemi göstermektedir.

Önem analizine göre, Model1'in tahmin gücü üzerinde en etkili değişken olarak ILK\_DONEM\_NOT\_ORTALAMASI bulunmuştur. Bu durum (Mishra vd. , 2014; Slim vd.,2014) çalışmalarındaki bulgularla uyumludur.

Yapılan analiz sonucu, ÇKA ile sınıflandırma analizinde en önemli ikinci değişken olarak LISE\_NOT \_ORTALAMASI bulunmuştur. Bu durumda aileden alınan akademik ve sosyal destek seviyesinin annenin eğitim seviyesiyle doğru orantılı olarak değiştiği görüşünü desteklemektedir. Spera vd. (2009) etnik gruplara yönelik yaptıkları çalışmalarında benzer şekilde anne eğitim durumu ile akademik performans arasında anlamlı bir ilişki saptamışlardır.

Tahmin modeli açısından önemi yaklaşık 0,10 olarak hesaplanan BÖLÜM değişkeni (1="İnsan Kaynakları Yönetimi", 2= "İşletme", 3="LEEDS", 4="Turizm İşletmeciliği", 5="Sağlık Yönetimi", 6="Uluslararası Ticaret", 7="Yönetim Bilişim Sistemleri") başarı durumunun ÇKA ile sınıflandırılmasında 3. Önemli değişken olarak belirtilmiştir. Sakarya Üniversitesi-İşletme Fakültesi'nde yer alan bölümlerin kodlanmasında 5,6 ve 7 değerlerine sahip olanlar 2011 yılında açılan nispeten yeni bölümlerdir. Bölüm ve başarı durumu arasındaki korelasyon katsayısı ve tahminleme modeli açısından BÖLÜM değişkeninin önemi düşünüldüğünde daha yeni olan ve LYS sonuçlarına göre daha yüksek sıralama ile öğrenci alan bölümlerde okuyan öğrencilerin diğer bölümlerde okuyanlara göre daha başarılı olduğu savunulabilir. Dolaylı olarak LYS puanının başarıya pozitif etki ettiği söylenebilir.

Geleneksel olmayan değişkenlerden en önemli değişken olarak sıralanan değişken bireyin bir hedefe ulaşmak için yapması gereken bir görevi gerçekleştirmede algılanan kabiliyetini ifade eden (Bandura, 1997) GOY\_DERECE (Genel Özyeterlik Derecesi) olmuştur. Bu sonuç Choi'nin (2005) çalışmasında elde edilen bulguyu desteklemektedir.

Kurumsal Entegrasyon Ölçeği'nden elde edilen alt boyutlardan Model1 açısından en önemli değişken BOYUT4\_DERECE olarak belirlenmiştir. Terenzini&Pascarella (1980) tarafından geliştirilen ve French&Oakes (2004) tarafından revize edilen Kurumsal Entegrasyon Ölçeğine (KEÖ) mevcut örneklem üzerinde uygulanan faktör

analizi sonucu elde edilen boyutlardan biri olan BOYUT4\_DERECE öğrencinin akademik ve düşünsel gelişimine yönelik algı durumu ile ilgili bir değişkendir. Buna göre Akademik ve Düşünsel Gelişim Durumu öğrencilerin başarı durumunun tahmin edilmesinde kullanılan önemli bir değişkendir.

Pek çok çalışmada (Astin, 1997; DeBerard, 2004; Noble&Sawyer, 2002) ilk yıl akademik başarı ile pozitif olarak ilişkilendirilen LİSE\_NOT\_ORTALAMASI, korelasyon temelli özellik seçimi ile modele dahil edilen değişkenler içerisinde en önemli 5. değişkendir. Geliştirilen teorik modele göre üniversitelerin müdahale alanı dışarısında kalan bu değişken, öğrencilerin başarı sınıflarından yola çıkarak bir profil oluşturmak için kullanılabilir.

Maalesef duyarlılık analizini, en başarılı tahminleyici Model5'e uygulamak mümkün değildir. Zira, Rassal Ormanlar yönteminde bağımsız değişkenler rastgele seçilmektedir (Breiman,2001).

Sonuç olarak; çalışmanın ana amacı olan öğrencilerin başarı durumlarının doğru sınıflandırılmasında göstermiş olduğu performans, risk grubunu belirleyebilme ve bilgisayar zamanı yönünden en iyi performansı ortaya koyması ve yorumlanmasının kolay olması nedeniyle Model5, çalışmanın bir diğer amacı olan akademik başarıya etki eden faktörlerin belirlenebilmesine yönelik duyarlılık analizi uygulanabilmesi nedeniyle ise Model1 kabul edilebilir modeller olarak belirlenmiştir.

#### **2.2.5.2. Gözden geçir**

Süreç gözden geçirilmiş ve şu problemlerle karşılaşmıştır:

- Çeşitli nedenlerden ötürü (devamsızlık, çalışmaya katılmayı kabul etmeme, uygun zaman ayarlayamama) tüm öğrencilere ulaşılamamıştır.
- Katılımcılardan bazıları çalışmanın ciddiyeti konusunda tereddüte düşmüş ve gerçekdışı beyanlarda (Öğrenci numarası, ilk dönem not ortalaması, finansal kaynaklar vs.) bulunmuşlardır. Bu durum çalışmanın güvenilirliğini olumsuz anlamda etkilemiştir.
- Katılımcılara ait öğrencilerin kişisel özelliklerine ve akademik-sosyal ortama yönelik algılarını ölçmeyi amaçlayan anketlerin bazıları istenen ciddiyetle doldurulmamıştır.

### 2.2.5.3. Sonraki adımları belirle

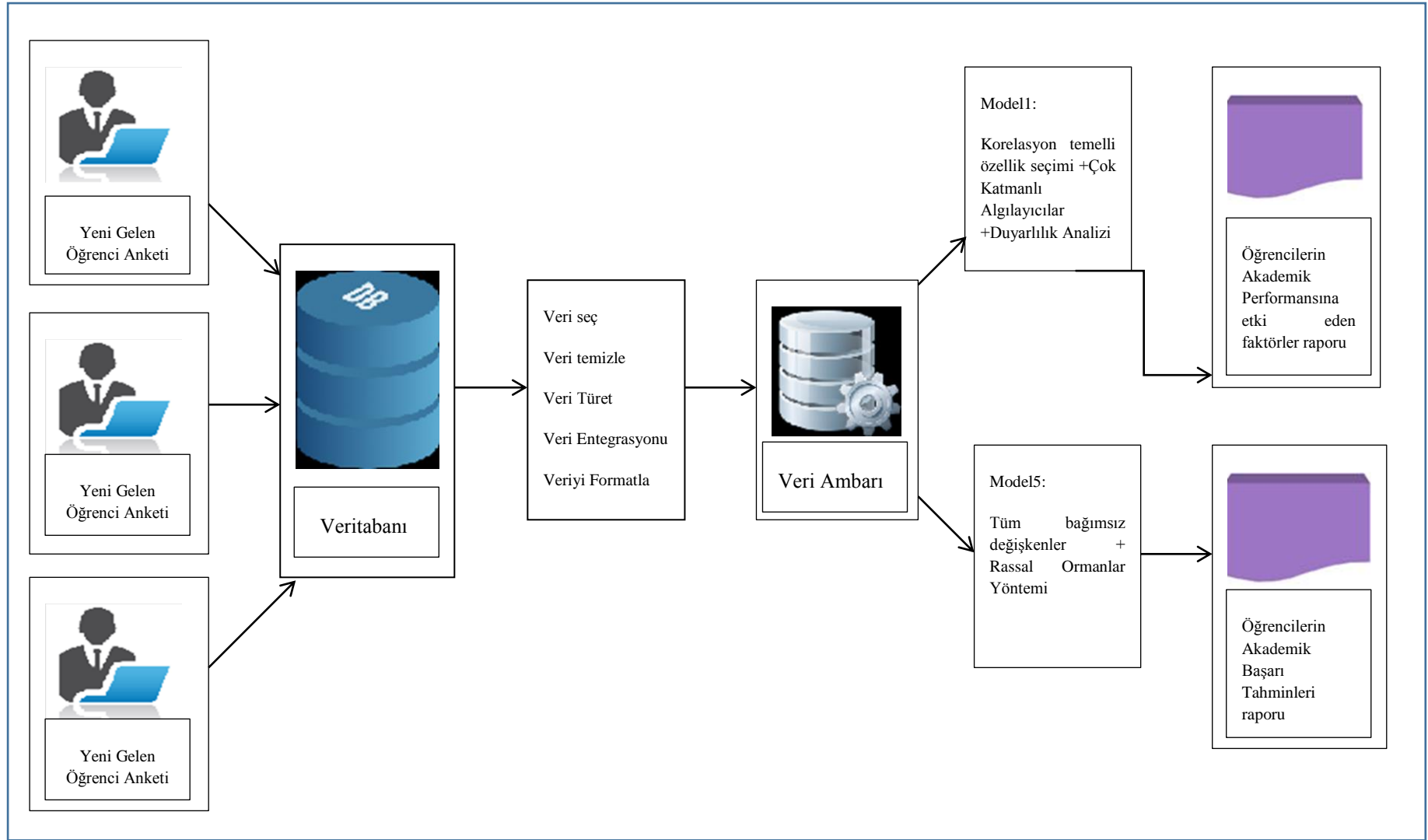
Süreçte yaşanan bazı sorunlara yönelik çözüm önerileri ve olası eylemler şu şekilde listelenebilir:

- Çalışmanın önemini daha iyi vurgulayabilmek açısından üniversite yönetiminden destek alınabilir.
- Uygulanan “Yeni Gelen Öğrenci Anketi” uygulama zamanları çok daha iyi planlanarak katılıma ve çalışmanın amacına olumsuz etki eden faktörler (sınav stresi, devamsızlık durumu vs.) en aza indirilebilir.
- Öğrencilerin ankete katılımını teşvik edecek önlemler (ücret, çekiliş vs.) alınabilir.
- Anket uygulanacak dersi veren öğretim üyesine çalışmanın amacı ve önemi, sunum vasıtasıyla çok daha iyi anlatılabilir.

Sonuç olarak, farklı senaryolardan da elde edilen veriler ışığında Model1 ve Model5’in üniversitede yeni gelen öğrencilerin akademik performansının tahmini ve bu performansa etki eden faktörlerin araştırılmasına yönelik çalışmada kullanılabilir olduğuna karar verilmiştir.

### 2.2.6 Yayılım

Mevcut çalışmanın kapsamının genişletilerek üniversite çapında uygulanması amaçlanmaktadır. Bu kapsamda sürecin yayılımını sağlamak üzere, akademik performansa etki eden faktörlerin zaman içerisindeki değişimini takip etmek, önem sıralarını güncelleyebilmek ve boylamsal (uzun soluklu) bir çalışma gerçekleştirebilmek amacıyla bir sistem tasarlanması düşünülmüştür. Çalışmada uygulanan süreç göz önünde bulundurularak şu şekilde genel bir tasarım önerilebilir.



**Şekil 70: Öğrencilerin akademik başarı durumlarının tahminine ve bu başarıya etki eden faktörlerin belirlenmesine yönelik bir sistem tasarımı**

Mevcut çalışmanın kapsamının genişletilerek uzun vadede kullanımını sağlayabilmek amacıyla genel hatlarıyla tasarlanan sistemin süreçleri aşamalı olarak planlanmıştır.

**1-** Sakarya Üniversitesi lisans programlarına kaydolan ilk yıl öğrencileri, her yıl bahar yarıyılı başlangıcında başında kullanıcı adları ile girecekleri sistem üzerinden “Yeni Gelen Öğrenci Anketi” ni doldurur. Böylece verilerin daha sistematik olarak elde edilmesi ve fazla iş yükünden kaçınılması amaçlanmıştır.

**2-** İlgili veritabanında toplanacak veriler sonrasında kolaylıkla erişilmek, yönetilmek ve güncellenebilmek amacıyla (Rouse&Leake, 2015) ilgili veritabanlarında saklanır.

**3-** ETL (Extract-Transform-Load) olarak da bilinen veri hazırlama sürecinde, Yeni Gelen Öğrenci anketinden elde edilen veriler çeşitli işlemlerden (veri temizleme, veri türetme, veri dönüştürme ve veri entegrasyonu) geçerek yüksek kalitede, entegre ve kurum çapında bilgi elde etmek (Hoffer vd.; 2002: s. 395) amacıyla veri ambarında depolanır.

**4-** Veri Ambarı, iş zekası ve yönetsel karar destek süreçlerini desteklemek amacıyla kullanılan özneye yönelik (Mevcut çalışma için özneler: Öğrenci, Öğretim Üyesi), entegre (dahili ve harici kaynaklardan elde edilen veriler için tutarlı kurallar, formatlar, yapısal kodlamalar ve ilgili karakteristiklerin belirlenmesi) zaman içerisinde değişen (zamansal boyut içerisindeki değişimlerin ve trendlerin gözlemlenmesi) ve güncellenemez (veri ambarları operasyonel sistemlerden elde edilen veriler doğrultusunda yüklenebilir ya da yenilenebilir, ancak son kullanıcılar tarafından güncellenemez) veri yığınıdır (Immon&Hackathorn, 1994: aktaran: Hoffer vd.; 2002: s. 394). Bilgilendirici bir sistem olan veri ambarlarının operasyonel bir sistem olarak düşünülebilecek veritabanlarından temel farkları şu şekilde sınıflandırılabilir.

**Tablo 18:**

**Operasyonel ve Bilgilendirici Sistemlerin karşılaştırılması (Hoffer vd.; 2002: s. 398)**

<b>Karakteristikler</b>	<b>Operasyonel Sistemler</b>	<b>Bilgilendirici Sistemler</b>
<i>Ana amaç</i>	İşler günlük bazda yürütülür.	Yönetimsel karar vermeyi destekler.
<i>Veri tipi</i>	İşin günlük durumunu sunar.	Tarihsel kesitler ve tahminler
<i>Temel Kullanıcılar</i>	Memurlar, Satış elemanları, alt yöneticiler	Üst yöneticiler, iş analistleri, müşteriler
<i>Kullanım kapsamı</i>	Dar, planlı ve basit güncellemeler ve sorgular	Geniş, doğaçlama, kompleks sorgular ve analizler
<i>Tasarım amacı</i>	Performans üretim oranı	Esnek erişim ve kullanımın kolaylaştırılması
<i>Veri boyutu</i>	Çok, tek ya da birkaç tablo satırında sabit güncellemeler ve sorgular	Periyodik yığın güncellemeler ve çok sayıda ya da bütün sütunların kullanımını gerektiren sorgular

Çalışmanın amacına paralel olarak, öğrencilerin başarı sınıflarının tahmin edilmesi ve bu sınıfların tahminine etki eden faktörlerin belirlenmesi için bilgilendirici bir sistem olan veri ambarları uygun bir seçim olacaktır. Zira, önerilen sistem ile öğrencilerin ilk yıl içerisinde bulunduğu karmaşık süreçle ilgili alınacak tedbirlere yönelik üniversite yöneticilerine karar destek sağlamayı amaçlanmaktadır.

**5-** Veri ambarlarında saklanan veriler, mevcut durum araştırmaları ve olası senaryolarda denenen modellerin performans puanları göz önüne alınarak, en yüksek doğru sınıflandırma oranını en kısa zamanda elde edebilmek üzere Model5'in, duyarlılık analizi vasıtasıyla performansa etki eden faktörlerin ifşa edilebilmesi için Model1'in kullanılması uygun görülmüştür.

**6-** Son aşamada ise öğrencilerin akademik başarılarının tahminleri ve akademik başarıya etki eden faktörler ile ilgili düzenli raporlar sunulur. Bu raporlar doğrultusunda, öğrencilerin akademik performanslarının iyileştirilmesine yönelik alınacak kararlara destek sağlanması amaçlanmaktadır.

## SONUÇ VE ÖNERİLER

Bu çalışmada öncelikle Sakarya Üniversitesi İşletme Fakültesi'nde bulunan lisans programlarında 2014-2015 akademik yılında ilk yıllarını geçiren öğrencilerin akademik performansına etki eden faktörler araştırılmıştır. Sonrasında ise bu faktörlere bağlı olarak öğrencilerin ilk yıl sonundaki başarı durumları tahmin edilmeye çalışılmıştır. Bu kapsamda toplamda 520 öğrenciye uygulanan anketlerden bazıları çeşitli nedenlerden ötürü elenmiştir. Bu nedenler şu şekilde belirtilebilir:

- Öğrencinin bireysel seviyede akademik performansın takip edilebilmesi için gerekli olan öğrenci numarasının yanlış beyan edilmesi veya hiç beyan edilmemesi
- Öğrencinin kayıtlı olduğu lisans programında ilk yılını geçiriyor olmaması
- Öğrencinin yatay geçiş ile gelmiş olması
- Çalışmanın bireysel tutum ve algıları ölçmeyi amaçlayan anket kısımlarında yer alan tüm ifadelerle aynı katılım düzeyinde cevaplar verilmesi

Öğrenci numarasının yanlış beyan edilmesi veya hiç beyan edilmemesinin nedeni, katılımcının öğrenci numarasını belirtmeyi özel alanına müdahale olarak algılaması olarak düşünülebilir. Bu noktada, her ne kadar uygulama aşamasında anketlerden temin edilen verilerin gizliliğinin çalışmanın esaslarından biri olduğu sözlü olarak belirtilmişse de, bu gizliliği garanti altına alacak bir belgenin öğrencilerin kafasındaki tereddütleri gidermede daha faydalı olabileceği söylenebilir. Bu konuda bir diğer öneri, ders saatleri içerisinde uygulanan bu anketlerde, dersi veren öğretim üyesinin “Yeni Gelen Öğrenci” anketine yönelik tutumudur. Öğretim üyesinin bu tip bir çalışmaya destek vermesi öğrencilerin kendilerini daha güvende hissederek anketleri ciddiyle doldurmalarını sağlayabilir. Bu noktada, öğretim üyesine anketi uygulamadan önce çalışmanın amacı ve olası sonuçları detaylı olarak anlatılmalı ve anket çalışması için kendisi ve sınıfı için en uygun zaman dilimi birlikte kararlaştırılmalıdır. Böylece çalışmaya olumsuz etki eden “zaman” kısıtı hafifletilmiş olacaktır.

Demografik ve algısal değişkenlere yönelik ifadelerle tutarsız ve yanlış yanıtlar verilmesi çalışmanın mantıklı sonuçlar üretmesi açısından büyük bir handikap teşkil etmektedir. Bu durumun önüne geçilebilmesi için, öncelikle çalışmanın uygulanacağı

fakülte daha sonra üniversite yönetiminin öğrencilerin akademik performansını geliştirmeyi amaçlayan bu çalışmaya destek vermesi sağlanmalıdır. Bu noktada yöneticilere “Eğitimde Veri Madenciliği” kavramından ve yapılan başarılı uygulamalardan bahsedilerek öğrencilerden elde edilen verilerin ne denli büyük bir öneme haiz olduğu belirtilebilir. Böylece üniversitenin tüm yönetim seviyelerinde bu tip çalışmalara yönelik olumlu bir algı oluşturarak, çalışmanın önemini öğrencilere yayılımı sağlanabilir. Mevcut çalışma da tam olarak bu amaca hizmet etmektedir. Yalnızca Sakarya Üniversitesi-İşletme Fakültesi bünyesinde uygulanan çalışma, üniversite genelinde olası bir çalışmanın temellerini ve uygulama esaslarını belirleme yönünden bir araştırma niteliğindedir. Bu noktada elde edilen sonuçlar, literatüre (Oyelade vd. 2010; Al-Radaideh vd., 2006; Taylan&Karagözoğlu, 2009) bakıldığında umut vericidir. Mevcut durumda hiçbir müdahalede bulunulmamış yalın verisetiyle yapılan sınıflandırma çalışmasında Model5 (Elde edilen tüm değişkenlerin kullanıldığı Veriseti1 + Rassal Ormanlar yöntemi) yaklaşık %63'lük bir doğru sınıflandırma oranına ulaşmıştır. Bir başka deyişle mevcut durumda her 10 öğrenciden 6'sının İkinci Dönem Başarı Durumu doğru olarak tahmin edilebilmiştir. Modern sınıflandırıcılarda en sık rastlanan problemlerden birisi olan “Sınıf dengesizliği” (Japkowicz&Stephen, 2000) giderildikten sonra bu oran, mevcut çalışma için de makul sınır olarak kabul edilen %70'in üzerine çıkmıştır. Mevcut çalışmada, “Sınıf dengesizliği” probleminin üstesinden gelebilmek için, örneklem sayısı azınlıkta olan sınıfları maksimum örnekleme sahip sınıfın örneklem sayısına “sentetik” olarak arttırmak suretiyle sınıf dengeleyen SMOTE (Chawla vd., 2002) yöntemi kullanılmıştır. Bu noktada, farklı “veri dengeleme” yöntemleri izlemek, tahminlerin doğruluk oranlarını geliştirebilme ihtimali göz önüne alındığında önemlidir. Bu anlamda Thammasiri vd. (2013) farklı yöntemler kullanarak ilgi çekici sonuçlara ulaşmışlardır.

Uygulama aşamasında, tutarsız yanıtlar kadar boş bırakılan alanlar da çalışmanın güvenilirliği ve genelleyiciliğini olumsuz anlamda etkilemektedir. Literatürde bu boş bırakılan alanları uygun bir şekilde doldurmayı amaçlayan “imputation methods”(yakıştırma yöntemleri) başlığı altında pek çok araştırma yapılmıştır (Li vd., 2004; Enders, 2001). Yakıştırma yöntemleri, tahmin etmeye çalıştıkları veri tiplerine göre 4 ana grupta toplanabilir (Engels&Diehr, 2003: s. 969).



- Popülasyon
- Dayanak
- Öncesi
- Önce/Sonra (B/A)

Mevcut çalışmada, popülasyon yöntemlerinden sütun ortalama/medyan yöntemi benimsenmiştir. Kayıp verilerin mümkün olduğunca gerçeğe yakın doldurularak eksiksiz bir veriseti elde edilmesi için farklı yöntemler şu şekilde tablolaştırılabilir.

**Tablo 19:**

**Yakıştırma yöntemleri (Engels&Diehr, 2003: s. 970)**

<b>Grup</b>	<b>Yöntem</b>	<b>Tanım</b>
Popülasyon	Sütun ortalaması	Herhangi bir yıl için verisetinde tüm kişilerin ortalaması
Popülasyon	Sütun medyanı	Herhangi bir yıl için verisetinde tüm kişilerin medyanı
Dayanak	Sınıf Ortalaması	Belirli bir sınıfta yer alan boş alan için diğer kişilerin ortalaması
Dayanak	Sınıf Medyan	Belirli bir sınıfta yer alan boş alan için diğer kişilerin medyanı
Dayanak	Hot Deck	İlgili sınıfta yer alan herhangi bir örneklemin değeri
Dayanak	Regresyon	Regresyon modeliyle tahmin edilen değer
Dayanak	Hatalı regresyon	Hata terimi eklenen regresyon değeri
Önce	Önceki satır ortalaması	Bir kişinin daha önceden bilinen değerlerinin ortalaması
Önce	Önceki satır medyanı	Bir kişinin daha önceden bilinen değerlerinin medyanı
Önce	LOCF (Last observation carried forward)	İleriye taşınan son gözlem
B/A	Satır ortalaması	Kişinin önce ve sonraki değerlerinin ortalaması
B/A	Satır medyanı	Kişinin önceki ve sonraki değerlerinin medyanı
B/A	NOCB (Next observation carried forward)	Geriye taşınmış sonraki gözlem
B/A	Son& İlk	En son bilinen ve sonraki bilinen değerlerin ortalaması

Mevcut çalışmada, kişiye özel herhangi bir bilgi içermeyen popülasyon yöntemlerinden sütun ortalaması ve medyanı uygulanmıştır. İleriki çalışmalarda akademik performansın tahmin edilmesinde daha yüksek doğru sınıflandırma oranlarına ulaşmada belirtilen farklı yöntemleri denemek faydalı olacaktır.

Öğrencilerin Sakarya Üniversitesinin sunduğu akademik ve sosyal ortama yönelik algılarını ölçmek amacıyla birden fazla alt ölçeğe sahip Kurumsal Entegrasyon Ölçeği'nden (Pascarella&Terenzini, 1980; French&Oakes, 2004) yararlanılmıştır. Bu kapsamda, katılımcıların verdiği yanıtlara göre ifadeleri gruplandırmak amacıyla faktör analizi uygulanmış ve orijinal çalışmadan farklı olarak Sakarya Üniversitesi'nde öğrencilerin akademik ve sosyal ortamda etkileşim halinde bulunduğu 4 boyut elde edilmiştir. Örnekleme, orijinal ölçekte "Öğretim Üyeleriyle Etkileşim" ve "Öğrenci gelişimine yönelik öğretim üyesi ilgisi" boyutlarında yer alan ifadelerin faktör analizleri sonrasında aynı grupta yer aldığı gözlemlenmiş ve bu iki altölçek " Öğretim Üyeleriyle Etkileşim" adı altında yeni bir boyut olarak belirlenmiştir. Literatürde Türkçe'ye ilk kez çevirilen ve uygulanan Kurumsal Entegrasyon Ölçeğinin orijinalinden farklı boyutlara sahip olması, orijinal çalışmanın uygulandığı A.B.D. ve mevcut çalışmanın yapıldığı Türkiye arasındaki eğitimsel, kültürel ve toplumsal anlamdaki farklılıkların bir göstergesi olarak düşünülebilir. Bu noktada, Türkiye Yükseköğretim ortamına uygun bir ölçek geliştirilmesinin öğrencilerin akademik performanslarına etki eden faktörlerin vurgulanması ve doğru tahminleme oranını arttırabilme potansiyeli nedeniyle önemli olduğu düşünülmektedir.

Öğrencileri başarı durumları (lise not ortalaması, ilk dönem not ortalaması, ikinci dönem not ortalaması) , kişisel özellikleri (genel özyeterlik, akademik özyeterlik) ve Kurumsal Entegrasyon Ölçeği ile elde edilen her bir boyuta karşı tutumları (Öğretim Üyeleriyle Etkileşim, Arkadaş-Grup Ekileşimi, Üniversiteye Bağlılık ve Memnuniyet, Akademik Ve Düşünsel Gelişim) derecelendirilerek kategorik değişkenlere dönüştürülmüştür. Başarı ve algı durumunun göreceli olduğu düşünülerek, yüksek, orta ve düşük kategorileri arasındaki sınır değerleri, örneklemin dağılımına göre belirlenmiştir. Bunun için veriler, küme üyeleri arasındaki benzerlik maksimum, kümeler arası benzerlik minimum olacak şekilde gruplandıran kümeleme analizi ile gruplandırılmıştır. Katılımcılar verinin dağılımına uygun olarak belirlenen merkezlere

yakınlıklarına göre kategorilere atanmışlardır. Bu noktada ilerideki araştırmalarda, sınırların kesin olarak çizildiği “crisp” küme tarzında belirlenen eşik değerlerin “bulanık” laştırılarak (Zadeh, 1975) daha hassas ve isabetli sonuçların elde edilmesi sağlanabilir.

Bağımlı bir değişkenin gelecekteki değerinin bağımsız değişkenlere bağlı olarak tahmin edilmeye çalışılan araştırmalarda, bağımsız değişken sayısı çok fazla ise, optimum verisini elde etmeye yönelik özellik seçimi yaklaşımları (Fisher kriteri, ki-kare, korelasyon, bilgi kazanım oranı) mevcuttur. Çalışmada sürekli değişkenler arasındaki ilişkiyi belirleyebilmek üzere kullanılan Pearson  $r$  ve aynı amaçla kategorik değişkenler için Kendall’s tau göstergelerine göre özellik seçimi yapılmıştır. Çok Katmanlı Algılayıcılar ve Çokterimli Lojistik Regresyon yönteminin bu yaklaşımla çok daha iyi sınıflandırma yapabildiği gözlemlenmiştir. Ancak, korelasyon katsayısının değişken ve hedef arasındaki yalnızca doğrusal ilişkileri saptayabileceği düşünülmektedir (Guyon&Elisseff, 2003: s. 1161). Dolayısıyla farklı özellik seçimi yöntemleri kullanılarak çalışmayı geliştirmek uygun olacaktır.

Yapılan sınıflandırma uygulamasında bugüne dek pek çok çalışmada faydalanılmış Yapay Sinir Ağları (Oladokun vd., 2008), öğrenen topluluk yöntemlerinden Rassal Ormanlar yöntemi (Delen, 2010) ve geleneksel sınıflandırıcılardan Çokterimli Lojistik Regresyon analizi (Park&Kerr, 1990) kullanılmıştır. En iyi model performansına erişmek için gelecekteki çalışmalarda değişik sınıflandırma tekniklerini (Genetik Algoritmalar, Naive-Bayes, Adaboost, Destek Vektör Makineleri vs.) denemek faydalı olacaktır.

“Öğrenci” merkezli bir eğitim anlayışına geçildiği günümüzde, her öğrencinin akademik durumunun doğru tahmin edilerek bireysel bazda proaktif önlemler alınması gerekmektedir. Bu noktada tahminleyici çalışmalarda kullanılan farklı modellerin değerlendirilmesinde “doğru sınıflandırma oranı” neredeyse tek seçim kriteri olarak sunulmaktadır. Bu doğrultuda, akademik performans anlamında “risk grubu”nda yer alan öğrencilerin doğru bir şekilde belirlenmesi pek çok çalışmada ana amaçlardan biri olarak belirtilmektedir. (Ramaswami&Bhaskaran, 2010; Calvo-Flores vd., 2006; Bresfelean vd., 2008). Ancak sınıflandırıcıların özel olarak risk grubunda yer alan öğrencileri belirleyebilme yeteneği genel doğru sınıflandırma yeteneğinden farklılık

gösterebilmektedir. Kullanılan sınıflandırma tekniğine ve verisetine bağlı olarak değişebilen aradaki fark, mevcut çalışmadan örnek vermek gerekirse, mevcut durumda Model4 için %1.98 olmuştur. Geniş çaplı katılım sağlanan uzun soluklu çalışmalarda bu oranın çok ciddi sayılara tekabül edeceği düşünüldüğünde, sınıflandırıcının “Risk Grubu Belirleyebilme” yeteneğinin değerlendirmede kullanılması gerektiği düşünülmüştür. Literatürde, her iki göstergiyi de kullanarak model değerlendiren bir yönteme rastlanmamıştır. Dolayısıyla, keyfi olarak, “Doğru Sınıflandırma Oranı” 0,75 ve “Risk Grubu Belirleyebilme Oranı” 0,25 ile ağırlıklandırılarak modele ait genel bir performans puanı elde edilmiştir. Keyfi olarak belirlenen değerlere göre yapılan bu model puanlama sistemi üzerinde çalışılması ve uygun yöntemler geliştirilmesi üzerinde durulması gereken noktalardan biridir.

Öğrencilerin akademik performanslarına etki eden faktörlerin belirlenmesi ve bu faktörlere bağlı olarak tahmin edilmesine yönelik çalışmalar, genellikle uzun soluklu (longitudinal) (Delen, 2010; Chamorro-Premuzic&Furnham, 2003) olmaktadır. Türkiye’de öğrencilerin akademik performansının takibi ve tahminine yönelik her hangi bir sistem ya da verilerin toplanması ile ilgili bir standart var olmadığından, çalışma tek bir akademik yıl içerisinde gerçekleştirilmiştir. Mevcut çalışma, Türkiye Yükseköğretim ortamında uzun vadeli ve geniş katımlı uygulamalara ön ayak olma amacını taşımaktadır. Bu kapsamda yapılan simülasyon çalışmalarından umut verici rakamlar elde edilmiştir. Örneğin; Model5, 2014-2015 akademik yılında Sakarya Üniversitesi’ndeki lisans programlarında ilk yıllarını geçiren tüm öğrencilerin ankete katıldığını varsayan bir senaryoda öğrencilerin yaklaşık %99’unu doğru sınıflandırmıştır. Bu oran her ne kadar bir simülasyon sonucu olsa da, geniş çaplı katılımın sağlandığı bir uygulama için cesaret vericidir. Uzun soluklu olarak elde edilen bu verilerin toplanması, saklanması ve raporlanmasına yönelik genel bir sistem tasarımı sunulmuştur. Buna göre; her yıl öğrenciler kullanıcı adı ve şifreleri ile girecekleri bu sistem üzerinden “Yeni Gelen Öğrenci” anketlerini dolduracaklardır. Sonrasında veritabanlarında toplanan bu verilerin ETL (Çıkar-Dönüştür-Yükle) işlemlerinden geçtikten sonra orta ve uzun vadeli kararlar almak için analiz yapma imkanı sunan veri ambarlarında saklanması öngörülmektedir. Böylece üniversite yöneticilerine hem öğrencilerin başarı durumlarını tahminlerini içeren hem de öğrencilerin genel olarak

akademik performanslarına etki eden faktörleri ortaya koyan raporlar sunmak suretiyle bir karar desteği sağlanması amaçlanmaktadır. Bu sistemin;

- Öğrencilerin akademik performanslarına etki eden faktörleri sıralamak
- Bu değişkenlerin yıllar içerisinde önem derecelerindeki değişimleri ve geçerlilik durumlarını gözlemleyebilmek
- Başarı durumları önceden tahmin edilen öğrencilere yönelik proaktif önlemler alabilmek
- İlk yıl öğrencilerine yönelik, oryantasyondan bir adım öteye giderek, “ilk yıl destek” programlarının temellerini atmak ve bu programların etkinliğini test edebilmek

gibi temel faydalar sağlaması amaçlanmaktadır.

Böyle bir sistemin kurulabilmesi için, geleneksel ve geleneksel olmayan değişkenlerin belirlenmesinde ve “Yeni Gelen Öğrenci” anketinin hazırlanması aşamasında, sosyal bilimler ile ilgilenen araştırmacıların, elde edilen bu değişkenlerin çeşitli analizlerde kullanılmasında ve bu sonuçların yayılımını sağlayacak sistemin oluşturulmasında fen bilimleri ile ilgilenen araştırmacıların bir arada bulunacağı bir proje ekibinin kurulması önerilebilir.

Çalışmanın bundan sonraki aşamalarda, çok daha hassas, bir başka deyişle çok daha fazla başarı sınıfı için yüksek doğru tahmin oranlarında bulunabilen, bir erken uyarı sistemine dönüştürülmesi planlanmaktadır. Böylece hem öğrencilerin hem de üniversitede karar vericilerin proaktif önlemler alarak, genel başarı durumunun artırılması amaçlanmaktadır.

## KAYNAKÇA

- Ahmad Slim, G. L. (2014). Predicting Student Success based on Prior Performance. *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on* (s. 410-415). Orlando, FL: IEEE.
- Ajansı, A. (2015, Şubat 27). *ensonhaber*. [www.ensonhaber.com: http://www.ensonhaber.com/kirsal-alanlar-yeniden-tanimlaniyor-2015-02-27.html](http://www.ensonhaber.com/kirsal-alanlar-yeniden-tanimlaniyor-2015-02-27.html) adresinden alınmıştır
- Akman, M., Genç, Y.,&Ankaralı, H. (2011). Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama. *Türkiye Klinikleri Biyoistatistik Dergisi*, 36-48.
- Akyol, M. A. (2015, Haziran 8). Zeki Optimizasyon Teknikleri. Ankara, Maltepe, Türkiye.
- Al- Radaideh, Q. A., Al-Shawakfa, E.,&Al-Najjar, M. I. (2006). Mining Student Data Using Decision Trees. *The 2006 International Arab Conference on Information Technology (ACIT'2006)* (s. 1-5). Jordan: Yarmouk University.
- Alay, S.,&Koçak, S. (2003). Üniversite Öğrencilerinin Zaman Yönetimleri ile Akademik Başarıları Arasındaki İlişki. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, 326-335.
- Allen, J.,&Robbins, S. B. (2008). Prediction of College Major Persistence Based on Vocational Interests, Academic Preparation, and First-Year Academic Performance. *Research in Higher Education*, 62-79.
- Allen, W. R.,&Haniff, N. Z. (1991). Race, gender, and academic performance in US higher education. W. R. Allen, E. G. Epss,&N. Z. Haniff içinde, *College in Black and White: African American students in predominantly White and in historically Black public universities* ((s. 95-109). Albany, NY: State University of New York Press.
- Altınkurt, Y. (2008). Öğrenci Devamsızlıklarının Nedenleri ve Devamsızlığın Akademik Başarıya Olan Etkisi. *Dumlupınar Üniversitesi-Sosyal Bilimler Dergisi*, 129-142.
- Ardıç, F. (2011, Şubat). Tercihler ve Üniversite Seçimine Etki Eden Faktörler. İstanbul, Büyükçekmece, Türkiye.
- Astin, A. (1997). How "Good" is your Institution's Retention Rate? *Research in Higher Education*, 647-658.

- Astin, A. W. (1964). Distribution of Students Among Higher Educational Institutions. *Journal of Educational Psychology*, 276-287.
- Astin, A. W. (1984). Student Involvement: A Developmental Theory for Higher Education. *Journal Of College Student Personnel*, 297-308.
- Astone, N. M.,&McLanahan, S. S. (1991). Family structure, parental practices and high school completion. *American sociological review*, 309-320.
- Ateşoğlu, M. (2013, Kasım 7). Sınavlar ve Ebeveyn Tutumları. Ankara, Çankaya, Türkiye.
- Aylesworth, L. S.,&Bloom, B. L. (1976). College Experiences and Problems of Rural and Small-Town Students. *Journal of College Student Personnel*, 236-242.
- Aytaç, M.,&Öngen, B. (2012). Doğrulayıcı Faktör Analizi İle Yeni Çevresel Paradigma Ölçeğinin Yapı Geçerliliğinin İncelenmesi. *İstatistikçiler Dergisi*, 14-22.
- Baird, L. L. (1967). *The Educational Tools of College Bound Youth, American College Testing Program Research Report*. Iowa.
- Baker, R. S. (2010). Data Mining for Education. *International Encyclopedia of Education* , 112-118.
- Bandura, A. (1997). Self- efficacy: The Exercise of Control. *New York: Freeman*.
- Baradwaj, B. K.,&Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications*, 63-69.
- Barefoot, B. O. (2000). The First-Year Experience. *About Campus*, 12-18.
- Barker, K., Trafalis, T.,&Rhoads, T. R. (2004). Learning From Student Data. *Proceedings of the 2004 Systems and Information Engineering Design Symposium* (s. 79-86). Charlottesville, VA: IEEE.
- Başol, G. (2012). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. Ltd. Şti.
- Belanger, F., Lewis, T., Kasper, G. M., Smith, W. J.,&Harrington, K. V. (2007). Are Computing Students Different? An Analysis of Coping Strategies and Emotional Intelligence. *Education, IEEE Transactions*, 188-196.
- Belch, H., Gebel, M.,&Maas, G. M. (2001). Relationship Between Student Recreation Complex Use, Academic Performance, and Persistence of First-Time Freshmen. *Journal of Student Affairs Research and Practice*, 220-234.

- Benyamin, D. (2012, Kasım 10). *Citizenet*. [www.citizenet.com:https://citizenet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/](http://www.citizenet.com:https://citizenet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/) adresinden alınmıştır
- Berger, J. B.,&Braxton, J. M. (1998). Revising Tinto's Interactionist Theory of Student Departure through Theory Elaboration: Examining the Role of Organizational Attributes in the Persistence Process. *Research in Higher Education*, 103-119.
- Berger, J. B.,&Milem, J. F. (1999). The Role of Student Involvement and Perceptions of Integration in a Causal Model of Student Persistence. *Research in Higher Education*, 641-664.
- Besterfield-Sacre, M., Atman, C. J.,&Shuman, L. (1997). Characteristics of Freshman Engineering Students: Models for Determining Student Attrition in Engineering. *Journal of Engineering Education*, 139-149.
- Bhardwaj, B. K.,&Pal, S. (2011). Data Mining: A Prediction for Performance Improvement Using Classification. *International Journal of Computer Science and Information Security*.
- Bilecik Üniversitesi- Bilgisayar Mühendisliği. (2015, Haziran 8). *Bilecik Üniversitesi*. [www.bm.bilecik.edu.tr: http://bm.bilecik.edu.tr/Dosya/Icerik/107/DosyaEki/nsg\\_ders\\_notu.pdf](http://www.bm.bilecik.edu.tr:www.bm.bilecik.edu.tr/http://bm.bilecik.edu.tr/Dosya/Icerik/107/DosyaEki/nsg_ders_notu.pdf) adresinden alınmıştır
- Bjørnebekk, G., Diseth, Å.,&Ulriksen, R. (2013). Achievement Motives, Self-Efficacy, Achievement Goals, and Academic Achievement at Multiple Stages of Education: A Longitudinal Analysis. *Psychological Reports*, 771-787.
- Blanchfield, W. C. (1971). College Dropout Identification: A Case Study. *The Journal of Experimental Education*, 1-4.
- Bollen, K. A.,&Barb, K. H. (1981). Pearson's R and Coarsely Categorized Measures. *American Sociological Review*, 232-239.
- Bong, M.,&Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review*, 1-40.
- Borman, W. C.,&Motowidlo, S. J. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research . *Human Performance*, 99-109.
- Braxton, J. M., Milem, J. F.,&Sullivan, A. S. (2000). The Influence of Active Learning on the College Student Departure Process: Toward a Revision of Tinto's Theory. *Journal of Higher Education*, 569-590.



- Braxton, J. M., Vesper, N., & Hossler, D. (1995). Expectations for College and Student Persistence. *Research in Higher Education*, 595-611.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 5-32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth.
- Bresfelean, V. P., Bresfelean, M., Ghisoiu, N., & Comes, C. A. (2008). Determining Students' Academic Failure Profile Founded on Data Mining Methods. *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference* (s. 317-322). Dubrovnik: IEEE.
- Brian F. French, J. C. (2005). An examination of indicators of Engineering Students' Success and Persistence. *The Research Journal for Engineering Education*, 419-425.
- Brower, A. M. (1992). The "Second Half" of Student Integration: The Effects of Life Task Predominance on Student Persistence. *The Journal of Higher Education*, 441-462.
- Burke, L. V. (2013, Ocak 24). *Politic365*. [www. politic365.com: http://politic365.com/2013/01/24/college-dropout-rate-called-national-crisis-in-new-report/](http://politic365.com/2013/01/24/college-dropout-rate-called-national-crisis-in-new-report/) adresinden alınmıştır
- Byrne, M., & Flood, B. (2008). University, Examining the relationships among background variables and academic performance of first year accounting students at an Irish. *Journal of Accounting Education*, 202-212.
- Cabrera, A. F., Nora, A., & Castanieda, M. B. (1992). The Role Of Finances in the Persistence Process: A Structural Model. *Research in Higher Education*, 571-593.
- Caison, A. L. (2007). Analysis of institutionally specific retention research : A comparison between survey and institutional database methods. *Research in Higher Education* , 435-451.
- Calvo-Flores, M. D., Galindo, E. G., Jimenez, M. P., & Pineiro, O. P. (2006). Predicting students' marks from Moodle logs using neural network models. *Current Developments in Technology-Assisted Education* , 586-590.
- Cambridge. (2015, Nisan 20). *Dictionary: Cambridge*. [www.cambridge.org: http://dictionary.cambridge.org](http://dictionary.cambridge.org) adresinden alınmıştır
- Carey, K. (2004). A matter of degrees: improving graduation rates in four-year colleges and universities. *The Education Trust*, -.

- Catley, C., Smith, K., McGregor, C., & Tracy, M. (2009). Extending CRISP-DM to incorporate temporal data mining of multidimensional medical data streams: A neonatal intensive care unit case stud. *Computer-Based Medical Systems* (s. 1-5). Albuquerque, New Mexico: IEEE Computer Society.
- Chamorro- Premuzic, T., & Furnham, A. (2003). Personality Predicts Academic Performance: Evidence from Two Longitudinal University Samples. *Journal of Research in Personality*, 319-338.
- Chapman, D. W., & Pascarella, E. T. (1983). Predictors of Academic and Social Integration of College Students. *Research in Higher Education*, 295-322.
- Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of AI Research*, 321-257.
- Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational psychology*, 55-64.
- Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. *Organizational Research Methods*, 62-83.
- Child, D. (2006). *The Essentials of Factor Analysis*. Edinburgh: A&C Black.
- Choi, N. (2005). Self-Efficacy and Self-Concept as predictors of college student's academic performance. *Psychology in the Schools*, 197-205.
- Chomitz, V. R., Slining, M. M., McGowan, R. J., Mitchell, S. E., Dawson, G. F., & Hacker, K. A. (2009). Is there a relationship between physical fitness and academic achievement? Positive results from public school children in the northeastern United States. *Journal of School Health*, 30-37.
- Cinsdikici, M. (2015, Haziran 8). Yapay Sinir Ağları. İzmir, Bornova, Türkiye.
- Cripps, A. (1996). "Using artificial neural nets to predict academic performance. *Proceedings of the 1996 ACM symposium on Applied Computing*. ACM (s. 33-37). Philadelphia, PA: ACM.
- Cross, S. S., Harrison, R. F., & Kennedy, M. (1995). Introduction to Neural Networks. *The Lancet*, 1075-1079.
- Çakmak, N., & Ocaklı, E. (2006). Performans değerlendirmesi gerekli midir ? Neden ? *ÜNAK'06 Bilimsel İletişim ve Bilgi Yönetimi Sempozyumu* (s. 212-230). Ankara: ÜNAK.
- Çayıroğlu, İ. (2015, Haziran 8). YAPAY SİNİR AĞLARI . Karabük, Karabük, Türkiye.

- ÇEKOPORT. (2014, Nisan 23). [www.cekoport.sakarya.edu.tr:  
http://www.cekoport.sakarya.edu.tr/Pages/Show.aspx?PAGEID=6a50f434-  
d9ed-4c95-98f7-88ce5537cb4d](http://www.cekoport.sakarya.edu.tr:8080/Pages/Show.aspx?PAGEID=6a50f434-d9ed-4c95-98f7-88ce5537cb4d) adresinden alınmıştır
- ÇETİNSAYA, G. (2014, Mayıs). Büyüme, Kalite, Uluslararasılaşma: Türkiye Yükseköğretimi İçin Bir Yol Haritası. *Büyüme, Kalite, Uluslararasılaşma: Türkiye Yükseköğretimi İçin Bir Yol Haritası*. Eskişehir, Türkiye: YÖK.
- David W. Chapman, E. T. (1983). Predictors of Academic and Social Integration of College Students. *Research in Higher Education*, 295-322.
- Davis Jr., G. W. (1989). Sensitivity Analysis in Neural Net Solutions. *Systems, Man and Cybernetics, IEEE Transactions*, 1078-1082.
- Day, J. C., & Newburger, E. C. (2002). *The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings. Special Studies. Current Population Reports*. Washington, DC.: Bureau of the Census.
- DeBerard, M. S. (2004). Predictors of Academic Achievement and Retention Among College Freshmen: A Longitudinal study. *College Student Journal*, 66-80.
- Dekker, G. B., & Pechenizkiy, M. V. (2009). Predicting Students Drop Out: A Case Study. *Proceedings of the International Conference on Educational Data Mining (EDM)* (s. 41-50). Cordoba: International Working Group on Educational Data Mining.
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 498-506.
- Demir, M. (2009). Analitik Verilerin Değerlendirilmesi. Aydın, Aydın, Türkiye.
- Demirel, O. (2003). *Eğitim sözlüğü*. Ankara: PeGem A Yayıncılık.
- DeTure, M. (2004). Cognitive style and self-efficacy: Predicting student success in online distance education. *American Journal of Distance Education*, 21-38.
- Doran, G. T. (1981). There's a SMART Way to Write Management's Goals and Objectives. *Management Review*, 35-36.
- Eckland, B. K. (1964). Social Class and College Graduation: Some Misconceptions Corrected. *American Journal of Sociology*, 36-50.
- Education, I. P. (2015, Mayıs 02). *NCES*. [www. nces.ed.gov:  
http://nces.ed.gov/ipeds/glossary/index.asp?searchtype=term&keyword=RETEN  
TION+RATE&Search=Search](http://nces.ed.gov/ipeds/glossary/index.asp?searchtype=term&keyword=RETENTION+RATE&Search=Search) adresinden alınmıştır

- Ehrenberg, R. G.,&Sherman, D. R. (1987). Employment While in College, Academic Achievement, and Postcollege Outcomes: A Summary of Results. *The Journal of Human Resources*, 1-23.
- Ellison, P. (2013, Mart 5). *Statistical Horizons*. [www.statisticalhorizons.com: http://statisticalhorizons.com/hosmer-lemeshow](http://www.statisticalhorizons.com/hosmer-lemeshow) adresinden alınmıştır
- Enders, C. K. (2001). The Performance of the Full Information Maximum Likelihood Estimator in Multiple Regression Models with Missing Data. *Educational and Psychological Measurement*, 713-740.
- Engels, J. M.,&Diehr, P. (2003). Imputation of Missing Longitudinal Data: A Comparison of Methods. *Journal of Clinical Epidemiology*, 968-976.
- Erhard, R.,&Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *EEE Data Eng. Bull*, 3-13.
- Ertürk, S. (1997). *Eğitimde Program Geliştirme*. Ankara: Yelkenstepe Yayınları.
- Everitt, B. S. (1998). *The Cambridge Dictionary of Statistics*. Cambridge: Cambridge University Press.
- Fausett, L. V. (1994). Predicting performance from test scores using backpropagation and counterpropagation. *IEEE International Conference* (s. 3398-3402). Orlando, FL: IEEE.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth,&P. (1996). From Data Mining to Knowledge Discovery in Databases. *The American Association for Artificial Intelligence*.
- Fike, D.,&Fike, R. (2008). Predictors of First-Year Student Retention in the community college. *Community College Review*, 68-88.
- Folan, P.,&Browne, J. (2005). A Review of Performance Measurement: Towards Performance Management. *Computers in Industry*, 663-680.
- Forman, G.,&Cohen, I. (2004). Learning from Little: Comparison of Classifiers Given Little Training. G. Forman,&I. Cohen içinde, *Knowledge Discovery in Databases: PKDD* (s. 161-172). Berlin: Springer.
- French, B. F., Immekus, J. C.,&Oakes, W. (2005). An examination of indicators of Engineering Students' Success and Persistence. *Journal of Engineering Education* , 419-425.
- French, B.,&Oakes, W. (2004). Reliability and Validity Evidence For the Institutional Integration Scale. *Educational and Psychological Measurement*, 88-98.

- Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, Cognitive Ability and Beliefs about Intelligence as Predictors of Academic Performance. *Learning and Individual Differences*, 49-66.
- Gamoran, A. (1996). Student Achievement in Public Magnet, Public Comprehensive, and Private City High Schools. *Educational Evaluation and Policy Analysis*, 1-18.
- Gazza, E. A., & Hunker, D. F. (2014). Facilitating student retention in online graduate nursing education programs: A review of the literature. *Nurse Education Today*, 1125-1129.
- Gedeon, T. D., & Turner, S. (1993). Explaining student grades predicted by a neural network. In *Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference* (s. 609-612). Nagoya: IEEE.
- Golding, P., & Donaldson, O. (2006). Predicting Academic Performance. *36th Annual Frontiers in Education Conference* (s. 21-26). San Diego: IEEE.
- Gore, P. A. (2006). Academic Self-Efficacy as a Predictor of College Outcomes: Two Incremental Validity Studies. *Journal of Career Assessment*, 92-115.
- Gorr, W. L., Nagin, D., & Szczypula, J. (1994). Comparative Study of Artificial Neural Network and Statistical Models for Predicting Student Grade Point Averages. *International Journal of Forecasting*, 17-34.
- Gulez, K. (2015, Haziran 8). Yapay Sinir Ağları. İstanbul, Beşiktaş, Türkiye.
- Guo, X., Yin, Y., Dong, C. Y., & Zhou, G. (2008). On the Class Imbalance Problem. *Fourth International Conference on Natural Computation* (s. 192-201). Washington, DC: IEEE Computer Society.
- Gurin, G., Newcomb, M. T., Cope, & G., R. (1968). *Characteristics of entering freshmen related to attrition in the Literary College of a large state university*. Final report.
- Guyon, I., & Elisseeff. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 1157-1182.
- Gültekin, B. K., & Dereboy, İ. F. (2011). Üniversite Öğrencilerinde Sosyal Fobinin Yaygınlığı ve Sosyal Fobinin Yaşam Kalitesi, Akademik Başarı ve Kimlik Oluşumu Üzerine Etkileri. *Türk Psikiyatri Dergisi*, 150-158.
- Hall, M. A. (1999, Nisan). Correlation-based feature selection for machine learning. *Doctoral dissertation, The University of Waikato*. Hamilton, Hamilton, Waikato, Yeni Zelanda: University of Waikato.

- Hançer, M. (2003). Ölçeklerin Yazım Dilinden Başka Bir Dile Çevirileri ve Kullanılan Değişik Yaklaşımlar. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 47-61.
- Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994). Predicting graduate student success: a comparison of neural networks and traditional techniques. *Computers & Operations Research*, 249-263.
- Higgins, J. (1989). Performance measurement in universities. *European Journal of Operational Research*, 358-368.
- Higher Education Research Institute. (2015, 25 5). Cooperative Institutional Research Program. *CIRP Freshman Survey*. Los Angeles, CA, Moore Hall, Amerika Birleşik Devletleri: Higher Education Research Institute.
- Hill, N. E. (2001). Parenting and academic socialization as they relate to school readiness: The roles of ethnicity and family income. *Journal of Educational Psychology*, 686.
- Himmelstein, H. C. (1992). Early Identification of High-Risk Students- Using Noncognitive Indicators. *Journal of College Student Development*, 89-90.
- Hoffer, J. A., Prescott, M. B., McFadden, & R., F. (2002). *Modern Database Management - Sixth Edition*. New Jersey: Pearson Education Inc.
- Hosch, B. J. (2008). Institutional and student characteristics that predict graduation and retention rates. *North East Association for Institutional Research Annual Meeting*. Rhode Island: North East Association for Institutional Research .
- Hoskins, S., & Van Hooff, J. C. (2005). Motivation and ability: which students use online learning and what influence does it have on their achievement? *British Journal of Educational Technology*, 177-192.
- Houltram, B. (1996). Entry age, entry mode and academic performance on a Project 2000 common foundation programme. *Journal of Advanced Nursing*, 1089-1097.
- Immon, W. H., & Hackathorn, R. D. (1994). *Using the Data Warehouse*. New York: John Wiley & Sons.
- istatistikanaliz. (2015, Mayıs 31). *istatistikanaliz*. [www.istatistikanaliz.com](http://www.istatistikanaliz.com): [http://www.istatistikanaliz.com/faktor\\_analizi.asp](http://www.istatistikanaliz.com/faktor_analizi.asp) adresinden alınmıştır
- Işığışık, E. (2004). *Bursa Ticaret ve Sanayi Odası (BTSO) Bursa Girişim Endeksi*. Bursa: Bursa Ticaret ve Sanayi Odası Basın Bülteni.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 651-666.

- Jain, A. K., Murty, M., & Flynn, P. J. (1999). Data Clustering: A Review. *ACM Computing Surveys (CSUR)*, 264-323.
- Jamelske, E. (2008). Measuring the Impact of a University First-Year Experience Program on Student GPA and Retention. *Higher Education*, 373-391.
- Japkowicz, N. (2000). Workshop Learning from Imbalanced Data Sets. *Proceedings of Association for the Advancement of Artificial Intelligence*. Austin: Association for the Advancement of Artificial Intelligence.
- Japkowicz, N., & Stephen, S. (2000). The Class Imbalance Problem: Significance and Strategies. *Proceedings of the International Conference on Artificial Intelligence*. Las Vegas: International Conference on Artificial Intelligence.
- Jawitz, J., & Scott, L. (1997). Who does not succeed in Engineering at the University of Cape Town? What can one tell from the retention rate? *Frontiers in Education Conference, 1997. 27th Annual Conference. Teaching and Learning in an Era of Change Proceedings* (s. 71-75). Pittsburgh: IEEE.
- Jayaprakash, S. M., & Lauria, E. J. (2014). Open Academic Early Alert System: Technical Demonstration. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (s. 267-268). New York, NY: ACM.
- Jeffrey F. Milem, J. B. (1997). A Modified Model of College Student Persistence: Exploring the Relationship Between Astin's Theory of Involvement and Tinto's Theory of Student Departure. *Journal of College Student Development*, 387-400.
- Jencks, C., & Phillips. (1998). *The Black-White Test Score Gap*. Washington D.C.: The Brookings Institution.
- Jerusalem, M., & Schwarzer, R. (1981). Fragebogen zur Erfassung von "Selbstwirksamkeit. Skalen zur Befindlichkeit und Persönlichkeit. *Forschungsbericht*.
- Jing, L. (2002). Data Mining and Knowledge Management in Higher Education - Potential Applications. *Annual Forum for the Association for Institutional Research* (s. 1-19). Ontario: ERIC.
- John M. Braxton, J. F. (2000). The Influence of Active Learning on the College Student Departure Process: Toward a Revision of Tinto's Theory. *The Journal of Higher Education*, 569-590.
- Joseph B. Berger, John M. Braxton. (1998). *Research in Higher Education*, 103-119.
- Kabra, R. R., & Bichkar, R. S. (2011). Performance Prediction of Engineering Students using Decision Trees. *International Journal of Computer Applications*, 36.

- Kantardzic, M. (2003). *Data Mining: Concepts, Methods and Algorithms*. New Jersey: John Wiley&Sons, Inc.
- Kaptanođlu, D.,&Özok, A. F. (2006). Akademik Performans Degerlendirmesi için Bir Bulanık Model. *İTÜDERGİSİ*, 193-204.
- Kerkvliet, J.,&Nowell, C. (2005). Does one size fit all? University differences in the influence of wages, financial aid and integration on student retention. *Economics of Education Review*, 85-95.
- Kline, P. (1994). *An Easy Guide to Factor Analysis*. Londra, Oxford, İngiltere: Routledge.
- Knoell, D. M. (1960). Institutional research on retention and withdrawa. *Research on college students*, 41-65.
- Knoell, D. M. (1966). A critical review of research on the college dropout. *The college dropout and the utilization of talent*, 63-81.
- Komarraju, M.,&Ramsey A., R. V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role of academic discipline. *Learning and Individual Differences*, 103-109.
- Kotsiantis, S. B.,&Panayiotis, E. .. (2005). Predicting students marks in hellenic open university. *Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference* (s. 664-668). Kaohsiung: IEEE.
- Kovacic, Z. (2010). Early Prediction of Student Success: Mining Students' Enrolment Data. *Proceedings of Informing Science&Information Technology Education Joint Conference* (s. 647-665). Cassino: Informing Science Institute.
- Kuran, K. (1988, Kasım). Çukurova Üniversitesi Eğitim Fakültesi Adana ve Hatay Eğitim Yüksekokullarındaki Öğrencilerin 1987 ÖSS Puanları ile Akademik Başarılarının Karşılaştırılması. Hatay, Türkiye: Yükseköğretim Kurulu Dökümantasyon Merkezi.
- Laerd Statistics. (2015, Mayıs 31). *SPSS Tutorials: Laerd Statistics*. www.statistics.laerd.com: <https://statistics.laerd.com/spss-tutorials/principal-components-analysis-pca-using-spss-statistics.php> adresinden alınmıştır
- Laska, J. A. (1984). The relationship between instruction and curriculum: A conceptual clarification. *Instructional Science*, 203-213.
- Lawhorn, J. T. (1971). A Study Of Persisters And Dropouts In The Secretarial Science Program At Miami-Dade Junior College. *Dissertations from ProQuest*. Miami, FL, USA.



- Le, H. (2005). Motivational and Skills, Social and Self- Management Predictors of College Outcomes: Constructing the Student Readiness Inventory. *Educational and Psychological Measurement* , 482-508.
- Lent, R. W., Brown, S. D.,&Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of counseling psychology*, 265-269.
- Li, D., Deogun, J., Spaulding, W.,&Shuart, B. (2004). Towards Missing Data Imputation: A Study of Fuzzy k-means Clustering Method. *Rough Sets and Current Trends in Computing*, 573-579.
- Lovibond, P. F.,&Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 335-343.
- Mallette, B. I.,&Cabrera, A. F. (1991). Determinants of Withdrawal Behavior: An Exploratory Study. *Research in Higher Education*, 179-194.
- Mancini, M. E., Ashwill, J.,&Cipher, D. J. (2015). A Comparative Analysis of Demographic and Academic Success Characteristics of On-line and On-Campus RN-to-BSN Students. *Journal of Professional Nursing* , 71-76.
- Marsh, L. M. (1966). College Dropouts: A Review. *The Personnel and Guidance Journal*, 475-481.
- Martocchio, J. J.,&Judge, A. T. (1997). Relationship between conscientiousness and learning in employee training: mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology*, 764.
- Mary Besterfield-Sacre, C. J. (1997). Characteristics of Freshman Engineering Students: Models for Determining Student Attrition in Engineering. *Journal of Engineering Education*, 139-149.
- Massachusetts Institute of Technology. (2015, Haziran 13). *Courses.Media. MIT*.  
www.mit.edu: [http://courses.media.mit.edu/2006fall/mas622j/Projects/manu-rita-MAS\\_Proj/MLP.pdf](http://courses.media.mit.edu/2006fall/mas622j/Projects/manu-rita-MAS_Proj/MLP.pdf) adresinden alınmıştır
- Mckenzie, K.,&Schweitzer, R. (2001). Who Succeeds At University? Factors Predicting Academic Performance in First Year Australian University Students. *Higher Education Research and Development*, 21-33.
- Memduhoğlu, H. B.,&Tanhan, F. (2013). Üniversite Öğrencilerinin Akademik Başarılarını Etkileyen Örgütsel Faktörler Ölçeğinin Geçerlik ve Güvenirlilik Çalışması. *YYÜ Eğitim Fakültesi Dergisi*, 106-124.

- Milem, J. F., & Berger, J. B. (1997). A Modified Model of College Student Persistence: Exploring the Relationship between Astin's Theory of Involvement and Tinto's Theory of Student Departure. *Journal of College Student Development*, 387-400.
- Mishra, T., Kumar, D., & Gupta, S. (2014). Mining Students' Data for Performance Prediction. *Fourth International Conference on Advanced Computing & Communication Technologies* (s. 255-262). Rohtak: IEEE.
- Moisen, G. G. (2008). Classification and Regression Trees. S. E. Jorgensen, & B. D. Fath içinde, *Encyclopedia of Ecology* (s. 582-588). Oxford: Elsevier.
- Monks, J., & G., E. R. (1999). US News & World Report's College Rankings: Why They Do Matter. *Change: The Magazine of Higher Learning*, 42-51.
- Morse, R. (2014, Eylül 8). *Education: US News*. US News: <http://www.usnews.com/education/best-colleges/articles/2014/09/08/best-colleges-ranking-criteria-and-weights> adresinden alınmıştır
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, 30-38.
- Murthy, K. S. (1998). Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. *Data Mining and Knowledge Discovery*, 345-389.
- Naik, B., & Ragothaman, S. (2004). Using Neural Networks to Predict MBA Student Success. *College Student Journal*, 143.
- Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning Patterns of University Student Retention. *Expert Systems with Applications*, 14984-14996.
- Neely, A. D. (1994). Performance Measurement System Design: A Literature Review and Research Agenda. *International Journal of Operations & Production Management*, 80-116.
- Noble, J., & Sawyer, R. (2002). *Predicting Different Levels of Academic Success in College Using High School GPA and ACT Composite Score*. ACT Research Report Series. Iowa City, IA: American Coll. Testing Program.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Ogor, E. N. (2007). Student Academic Performance Monitoring and Evaluation Using. *Fourth Congress of Electronics, Robotics and Automotive Mechanics* (s. 354-359). Morelos: CERMA.
- Oladokun, V. O., Adenbanjo, A. T., & Charles-Owaba, O. E. (2008). "Predicting Students' Academic Performance Using Artificial Neural Network: A Case

- Study of an Engineering Course. *The Pacific Journal of Science and Technology*, 72-79.
- Olani, A. (2009). Predicting First Year University Students' Academic Success. *Electronic Journal of Research in Educational Psychology*, 1053-1072.
- OnMyPhD. (2015, Haziran 8). *Contents. OnMyPhD.* www.onmyphd.com: <http://www.onmyphd.com/?p=gradient.descent> adresinden alınmıştır
- Oracle. (2015, Nisan 30). *Oracle* . www.oracle.com: [http://docs.oracle.com/cd/B28359\\_01/datamine.111/b28129/process.htm#DMC ON002](http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#DMC ON002) adresinden alınmıştır
- Osmanbegovic, E.,&Suljic, M. (2012). Data Mining Approach for Predicting Student Performance. *Journal of Economics and Business*, 3-12.
- Owen, S. V.,&Froman, D. R. (1988). Development of a College Academic Self-Efficacy Scale.
- Oyelade, O. J., Oladipupo, O. O.,&Obagbuwa, I. C. (2010). Application of k-Means Clustering Algorithm for Prediction of Students' Academic Performance. *International Journal of Computer Science and Information Security*, 292-295.
- ÖSYM. (2011). *ÖSYM Yükseköğretim İstatistikler*. Ankara: ÖSYM.
- ÖSYM. (2015, Nisan 23). www.osym.gov.tr: [http://dokuman.osym.gov.tr/pdfdokuman/2015/YGS/2015\\_OSYS\\_KILAVUZ.pdf](http://dokuman.osym.gov.tr/pdfdokuman/2015/YGS/2015_OSYS_KILAVUZ.pdf) adresinden alınmıştır
- ÖSYM. (2015, Haziran 27). *Hakkımızda- Tarihsel Gelişme: ÖSYM.* www.osym.gov.tr: <http://www.osym.gov.tr/belge/1-2706/tarihsel-gelisme.html> adresinden alınmıştır
- ÖSYM2015. (2015, Mayıs 12). *ÖSYM2015-INFO.* www.osym.2015.com: <http://www.osym2015.com/node/478> adresinden alınmıştır
- Özkan, Y. (2008). *Veri Madenciliğ Yöntemleri*. İstanbul: Papatya Yayıncılık.
- Özmutaf, N. M. (2007). Örgütlerde Bireysel Performans ve Çatışma. *C.Ü. İktisadi ve İdari Bilimler Dergisi*, 41-60.
- Öztemel, E. (2012). *Yapay Sinir Ağları*. Ankara: Altan Basım .
- Pal, S. K.,&Mitra, S. (1992). Multilayer Perceptron, Fuzzy Sets, and Classification. *IEEE Transactions on Neural Networks*, 683-697.

- Pallant, J. (2001). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows (versions 10 and 11): SPSS Student Version 11.0 for Windows*. Open University Press.
- Park, K. H., & Kerr, P. M. (1990). Determinants of academic performance: A multinomial logit approach. *The Journal of Economic Education*, 101-111.
- Parmentier, P. (1994). La réussite des études universitaires: facteurs structurels et processuels de la performance académique en première année en médecine. Leuven, Belgique: Université Catholique de Louvain.
- Pascarella, E. T., & Terenzini, P. T. (1980). Predicting Freshman Persistence and Voluntary Dropout Decisions from a Theoretical Model. *The Journal of Higher Education*, 60-75.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: A Third Decade of Research*. San Francisco, CA: Jossey-Bass.
- Pervin, L. A., Reik, L. E., & W., D. (1966). The College Dropout and the Utilization of Talent. *Princeton University Press*, 37-62.
- Peterson, D. B. (1993). Measuring Change: A Psychometric Approach to Evaluating Individual Coaching Outcomes. *Annual Conference of the Society for Industrial and Organizational Psychology*, (s. 1-20). San Francisco, CA.
- Pintrich, P. R., & De Groot, V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 33-40.
- Porter, K. B. (2008). Current trends in student retention: A literature review. *Teaching and Learning in Nursing*, 3-5.
- Quadri, M., & Kalyankar, N. (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology*.
- Quinlan, J. R. (1993). *C4. 5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.
- Ramaswami, M., & Bhaskaran, R. (2010). A CHAID Based Performance Prediction Model in Educational Data Mining. *International Journal of Computer Science Issues*, 10-18.
- Rasmani, K. A., & Shen, Q. (2006). Data-driven Fuzzy Rule Generation and its Application for Student Academic Performance Evaluation. *Applied Intelligence*, 305-319.

- Ravenscroft, S. P.,&Buckless, F. A. (1992). The effect of grading policies and student gender on academic performance. *Journal of Accounting Education*, 163-179.
- Rintala, U., Kairamo, A. K., Andersson, S., Strandas, S., Kelly, K., Gonalves, I.,&Lucas, A. (2012). Knowing our students- different approaches to student retention Experiences of the ATTRACT project. *SEFI 2012 Conference Proceedings: Engineering Education 2020: Meet the Future*. Thessaloniki: SEFI Conference.
- Romero, C.,&Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Systems, Man, and Cybernetics*, 601-618.
- Romero, C., Ventura, P. G.,&Hervas, C. (2008). Data Mining Algorithms to Classify Students. *The 1st International Conference on Educational Data Mining* (s. 8-17). Montreal: Educational Data Mining.
- Rouse, M.,&Leake, A. (2015, Haziran 16). *Data and Data Management Glossary: TechTarget SearchSQLServer*. [www.searchsqlserver.techtarget.com: http://searchsqlserver.techtarget.com/definition/database](http://searchsqlserver.techtarget.com/definition/database) adresinden alınmıřtır
- Sakarya Üniversitesi. (2011, Ağustos 27). Sakarya Üniversitesi Lisans ve Önlisans Eğitim- Öğretim ve Sınav Yönetmeliđi. *Derslerin Yapılma İlkeleri, Sınavlar ve Başarı Deđerlendirme, İtiraz Ders Tanım ve Uygulamaları*. Sakarya, Serdivan, Türkiye: Resmi Gazete.
- Sakarya Üniversitesi. (2015, Haziran 11). *Sayılarla SAÜ: Sakarya Üniversitesi*. [www.sakarya.edu.tr: http://sakarya.edu.tr/tr/sayilarla\\_sau](http://sakarya.edu.tr/tr/sayilarla_sau) adresinden alınmıřtır
- Saks, A. M. (1995). Longitudinal field investigation of the moderating and mediating effects of self-efficacy on the relationship between training and newcomer adjustments. *Journal of Applied Psychology*, 211-225.
- Salamonson, Y.,&Andrew, S. (2006). Academic performance in nursing students: influence of part-time employment, age and ethnicity. *Journal of Advanced Nursing*, 342-349.
- Scales, P. C., Benson, P. L., Roehlkepartain, E. C., Sesma, A.,&Van Dulmen, M. (2006). The role of developmental assets in predicting academic achievement: A longitudinal study. *Journal of Adolescence*, 691-708.
- Scarselli, F.,&Tsoi, A. C. (1998). Universal Approximation using Feed-forward Neural Networks: A Survey of Some Existing Methods and Some New Results. *Neural Networks*, 15-37.
- Schneider, J. (1997, Şubat 7). *Tutorial: Carnegie Mellon University*. [www.cmu.edu: http://www.cs.cmu.edu/~schneide/tut5/node42.html](http://www.cs.cmu.edu/~schneide/tut5/node42.html) adresinden alınmıřtır

- Schwab, J. A. (2015, Haziran 9). *University of Texas at Austin*. [www.utexas.edu: http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/](http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/) adresinden alınmıştır
- Sezer, A.,&H., T. (2003). İş birliğine dayalı öğrenmenin coğrafya dersinde akademik başarı üzerine etkisi. *Gazi Eğitim Fakültesi Dergisi*, 228-242.
- Shannon, C. E. (2001). A Mathematical Theoy of Communication. *CM SIGMOBILE Mobile Computing and Communications Review*, 3-55.
- Sherer, M.,&Adams, C. H. (1983). Construct validation of the self-efficacy scale. *Psychological Reports*, 899-902.
- Shovon, M., Islam, H.,&Haque, M. (2012). An Approach of Improving Students Academic Performance by Using k means Clustering Algorithm and Decision Tree. *nternational Journal of Advanced Computer Science and Applications*, 146-149.
- Sibi, P., Jones, S. A.,&Siddarth, P. (2013). Analysis of Different Activation Functions Using Back Propagation Neural Networks. *Journal of Theoretical and Applied Information Technology*, 1264-1268.
- Silah, M. (2003). Üniversite Öğrencilerinin Akademik Başarılarını Etkileyen Çeşitli Nedenler Arasında Süreksiz Durumluk Kaygının Yeri ve Önemi. *Eğitim Araştırmaları Dergisi* , 112-115.
- Sistemi, Y. B. (2015, Mayıs 3). Öğrenci Sayıları Özet Tablosu. Ankara, Ankara, Türkiye.
- Slim, A., Heileman, G. L., Kozlick, J.,&Abdallah, C. T. (2014). Employing Markov Networks on Curriculum Graphs to Predict Student Performance. *13th International Conference on Machine Learning and Applications* (s. 415-418). Detroit: IEEE.
- Slim, A., Heileman, G. L., Kozlick, J.,&Abdallah, C. T. (2014). Predicting Student Success based on Prior Performance. *Computational Intelligence and Data Mining (CIDM)*. Orlando, FL: IEEE.
- Solberg, V. S., O'Brien, K., Villareal, P., Kennel, R.,&Davis, B. (1993). Self-efficacy and Hispanic college students: Validation of the college self-efficacy instrument. *Hispanic Journal of Behavioral Sciences*, 80-95.
- Soner, V. (2011, 15 Eylül). *Volkan Soner Wordpress*. [www.volkansoner.wordpress.com](http://www.volkansoner.wordpress.com): <https://volkansoner.wordpress.com/2011/09/15/spss-clementine-uygulamasi/> adresinden alınmıştır

- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*, 64-85.
- Spady, W. G. (1971). Dropouts from Higher Education: Toward an Empirical Model. *Interchnge*, 38-62.
- Spera, C., Wentzel, K. R., & Matto, H. C. (2009). Parental Aspirations for Their Children's Educational Attainment: Relations to Ethnicity, Parental Education, Children's Academic Performance, and Parental Perceptions of School Climate. *Journal of Youth and Adolescence*, 1140-1152.
- Speybroeck, N. (2012). Classification and Regression Trees. *International Journal of Public Health*, 243-246.
- Stage, F. K. (1989). Motivation, Academic and Social Integration, and the Early Dropout. *American Educational Research Journal*, 385-402.
- Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 240-261.
- Stampen, J. O., & Cabrera, A. F. (1986). Exploring the Effects of Student Aid on Attrition. *Journal of Student Financial Aid*, 28-40.
- Starkweather, J., & Moske, A. K. (2015, Haziran 9). *Multinomial Logistic Regression*. [www.unt.edu: http://www.unt.edu/rss/class/Jon/Benchmarks/MLR\\_JDS\\_Aug2011.pdf](http://www.unt.edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011.pdf) adresinden alınmıştır
- Statistical Analysis System. (2015, Haziran 8). *Support: SAS*. [www.sas.com: http://support.sas.com/publishing/pubcat/chaps/57587.pdf](http://support.sas.com/publishing/pubcat/chaps/57587.pdf) adresinden alınmıştır
- StatisticsSolutions. (2015, Haziran 4). *Academic Solutions: Statistical solutions*. [www.statisticssolutions.com: http://www.statisticssolutions.com/kendalls-tau-and-spearman-rank-correlation-coefficient/](http://www.statisticssolutions.com/kendalls-tau-and-spearman-rank-correlation-coefficient/) adresinden alınmıştır
- Stergiou, C., & Siganos, D. (2015). *Neural Networks*. London: Imperial College.
- Stevenson, D. L., & Baker, D. P. (1987). The family-school relation and the child's school performance. *Child development*, 1348-1357.
- Superby, J. F., Vandamme, J. P., & N., M. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. *Workshop on Educational Data Mining*, 37-44.
- Survey System. (2015, Haziran 4). *Correlation: Survey System*. [www.surveysystem.com: http://www.surveysystem.com/correlation.htm](http://www.surveysystem.com/correlation.htm) adresinden alınmıştır

- Swets, J. A. (1996). *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics*. New Jersey: Lawrence Erlbaum Associates.
- Şen, B., Uçar, E.,&D., D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 9468-9476.
- Şen, S. (2011, Haziran 28). *Sümeyye Şen*. www.sumeyyesen.com: <http://www.sumeyyesen.com/2011/06/28/c-sharp-ile-yapay-sinir-aglarinerual-network-with-c/> adresinden alınmıştır
- Tabachnik, B. G.,&Fidell, L. S. (2007). *Using multivariate statistics Fifth Edition*. New York City, NY: Pearson Education INC.
- Taylan, O.,&Karagözoğlu, B. (2009). An Adaptive Neuro-Fuzzy Model for Prediction of Student's Academic Performance . *Computers&Industrial Engineering*, 732-741.
- Teachman, J. D. (1987). Family Background, Educational Resources, and Educational Attainment. *American Sociological Review* , 548-557.
- Temel, A. (2010). *Eğitimde Ölçme ve Değerlendime*. İstanbul: Maltepe Üniversitesi.
- Terenzini, P. T.,&Pascarella, E. T. (1980). Toward the Validation of Tinto's Model of College Student Attrition: A Review of Recent Studies. *Research in Higher Education*, 271-282.
- Thammasiri, D., Delen, D., Meesad, P.,&Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 321-330.
- Ting, S. R. (2000). Predicting Asian Americans' academic performance in the first year of college: An approach combining SAT scores and noncognitive variables. *ournal of College Student Development* , 442-449.
- Tinto, V. (1975). Dropout From Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 89-125.
- Tinto, V. (1988). Stages of Student Departure: Reflections on the Longitudinal Character of Student Leaving. *The Journal of Higher Education*, 438-455.
- Tinto, V. (1999). Taking Retention Seriously: Rethinking the First Year of College. *NACADA*, 5-9.
- Tinto, V. (2012). Enhancing Student Success: Taking the Classroom Success Seriously. *The International Journal of the First Year in Higher Education*, 1-8.



- Tinto, V.,&Pusser, B. (2006). *Moving from theory to action: Building a model of institutional action for student success*. Washington, D.C.: National Postsecondary Education Cooperative.
- Tonta, Y. (2007). Faktör Analizi. Ankara, Hacettepe, Türkiye.
- Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes . *Journal of Clinical Epidemiology*, 1225-1231.
- TÜİK. (2011, Kasım). Sorularla Resmi İstatistikler Dizisi-11. *İstatistiksel Kalite Kontrol* . Ankara, Çankaya, Türkiye: Türkiye İstatistik Kurumu.
- Türk Dil Kurumu. (2015, Temmuz 6). *Sözlük: Türk Dil Kurumu*. www.tdk.gov.tr: [http://tdk.gov.tr/index.php?option=com\\_gts&arama=gts&guid=TDK.GTS.5599d76d8dc731.44737183](http://tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.5599d76d8dc731.44737183) adresinden alınmıştır
- UCLA : Statistical Consulting Group. (2007, Kasım 24). *UCLA*. www.ats.ucla.edu: <http://www.ats.ucla.edu/stat/sas/notes2/> adresinden alınmıştır
- Uğurlu, B. (2015, Haziran 8). Slayt 7. Çanakkale, Çanakkale, Türkiye.
- Ullah, H.,&Wilson, M. A. (2007). Students' academic success and its association to student involvement with learning and relationships with faculty and peers. *College Student Journal*, 1192.
- UNESCO. (2013). UNESCO.
- University of California, Berkeley. (2015, Haziran 9). *Berkeley Statistics*. www.stat.berkeley.edu: [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm#features](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#features) adresinden alınmıştır
- Vandamme, J. P., Meskens, N.,&Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics*, 405-419.
- Venter, J., De Waal, A.,&Willers, C. (2007). Specializing CRISP-DM for Evidence Mining. P. Craiger,&S. Shenoï içinde, *Advances in Digital Forensics III* (s. 303-315). New York: Springer.
- Waikato University. (2015, Haziran 9). *Primer : Weka.Wikispaces*. www.weka.wikispaces.com: <https://weka.wikispaces.com/Primer> adresinden alınmıştır
- Warburton, D. M., Wesnes, K.,&Revell, A. (1984). Smoking and academic performance. *Current Psychology*, 25-31.

- Wilson, R. L.,&Hardgrave, B. C. (1995). Predicting Graduate Student Success in an MBA Program: Regression Versus Classification. *Educational and Psychological Measurement*, 186-195.
- Wirth, R.,&Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (s. 29-39). New York: The Association for the Advancement of Artificial Intelligence (AAAI).
- Wood, R. E.,&Locke, E. A. (1987). The Relation of Self-Efficacy and Grade Goals to Academic Performance. *Educational and Psychological Measurement*, 1013-1024.
- www.turkcebilgi.com.* (2015, Nisan 23). Türkçebilgi: <http://www.turkcebilgi.com/%C3%BCniversite#bilgi> adresinden alınmıştır
- Yadav, S. K.,&Pal, S. (2012). Data mining applications: A comparative study for predicting student's performance. *arXiv preprint arXiv*.
- Yılmaz, M., Gürçay, D.,&Ekici, G. (2007). Akademik Özyeterlik Ölçeğinin Türkçe'ye Uyarlanması . *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* , 253-259.
- Yu, L.,&Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)* (s. 856-863). Washington D. C.: International Conference on Machine Learning.
- Zadeh, L. A. (1975). Fuzzy Logic and Approximate Reasoning. *Synthese*, 407-428.
- Zaidah, I.,&Rusli, D. (2007). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. *21st Annual SAS Malaysia Forum*. Kuala Lumpur.
- Zhang, Q.,&Gupta, K. C. (2000). Neural Network Structures. Q. Zhang,&K. C. Gupta içinde, *Neural networks for RF and microwave design (Book+ Neuromodeler Disk)* (s. 61-103). Norwood, MA: Artech House, Inc.
- Zhang, S., Zhang, C.,&Yang, Q. (2010). Data Preparation For Data Mining. *Applied Artificial Intelligence: An International Journal*, 375-381.
- Zimmerman, B., Bandura, A.,&Martinez-Pons, M. (1992). Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Settin. *American Educational Research Journal*, 663-676.

## ÖZGEÇMİŞ

Tuğrul Cabir HAKYEMEZ, 25 Eylül 1988 tarihinde İstanbul'da doğdu. 2006 yılında Kartal Anadolu Lisesi'nden, 2011 yılında ise Sakarya Üniversitesi- Endüstri Mühendisliği Bölümü'nden mezun oldu. 2013 yılında Sakarya Üniversitesi- Yönetim Bilişim Sistemleri Bölümü'nde yüksek lisans eğitimine başladı. 2012 Eylül'den bu yana aynı bölümde araştırma görevlisi olarak görev yapmaktadır.