

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**VERİ MADENCİLİĞİ TEKNİKLERİ KULLANILARAK SOSYAL
AĞLAR ARACILIĞI İLE BİLGİSAYAR VE BİLİŞİM MÜHENDİSLİĞİ
MEZUN ÖĞRENCİ PROFİLLERİNİN BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

Esin Ayşe ZAIMOĞLU

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ**
Tez Danışmanı : Doç.Dr. Nilüfer YURTAY

Temmuz 2018

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ TEKNİKLERİ KULLANILARAK SOSYAL
AĞLAR ARACILIĞI İLE BİLGİSAYAR VE BİLİŞİM MÜHENDİSLİĞİ
MEZUN ÖĞRENCİ PROFİLLERİNİN BELİRLENMESİ

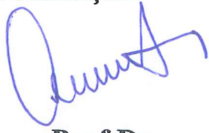
YÜKSEK LİSANS TEZİ

Esin Ayşe ZAIMOĞLU

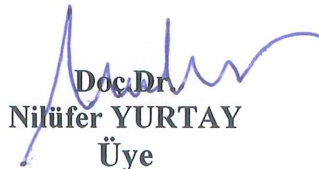
Enstitü Anabilim Dalı

: BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ

Bu tez 16 / 07 /2018 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.



Prof.Dr.
Cemil ÖZ
Jüri Başkanı



Doc.Dr.
Nilüfer YURTAY
Üye



Dr.Öğr. Üyesi
Barış KOÇER
Üye

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Esin Ayşe ZAIMOĞLU

16/07/2018

TEŐEKKÜR

Yüksek lisans eğitime başladığım andan itibaren, değerli bilgi ve deneyimlerinden yararlandığım, çalışmanın tüm aşamalarında yardımlarını esirgemeyen, teşvik eden, aynı titizlikte beni yönlendiren değerli danışman hocam Sayın Doç. Dr. Nilüfer YURTAY'a teşekkürlerimi sunarım.

Çalışma esnasında bilgi birikimi ve yardımlarını esirgemeyen Sayın hocam Öğr. Gör. Yüksel YURTAY'a verdiği destekten ötürü teşekkür ederim.

Tüm çalışma süreci boyunca anlayış ve yardımlarını esirgemeyen Sakarya Üniversitesi Öğrenci İşleri Daire Başkanı Sayın Fatma AYDIN'a, İstatistik ve Disiplin Şube Müdürü Sayın Serap MERT' e ve her türlü desteği sabırla gösteren değerli arkadaşım Sayın Niğar ÇELİK'e ayrıca teşekkür ederim.

Maddi manevi destekleriyle beni hiçbir zaman yalnız bırakmayan kıymetli eşim Fatih ZAIMOĞLU'na ve aileme de sonsuz teşekkür ederim.

İÇİNDEKİLER

TEŞEKKÜR	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	iv
ŞEKİLLER LİSTESİ	v
TABLOLAR LİSTESİ	vi
ÖZET	vii
SUMMARY	viii
BÖLÜM 1.	
GİRİŞ	1
BÖLÜM 2.	
LİTERATÜR ÖZETİ	4
BÖLÜM 3.	
YÖNTEM VE MATERYAL.....	8
3.1. Yöntem	8
3.1.1. Veri madenciliğine giriş	8
3.1.2. Veri madenciliğinin tanımı	8
3.1.3. Veri madenciliğinin tarihçesi	10
3.1.4. Veri tabanlarında bilgi keşfi.....	10
3.1.5. Veri madenciliğini oluşturan disiplinler	13
3.1.6. Veri madenciliği uygulama alanları	14
3.1.7. Veri madenciliği yöntemleri.....	15
3.1.7.1. Tanımlayıcı model	15

3.1.7.2. Öngörü (Tahmin) modeli	15
3.1.8. Birliktelik Kuralları	17
3.1.8.1. Güven ve destek kavramları	17
3.1.8.2. Birliktelik kuralları uygulaması	18
3.1.9. Apriori algoritması	20
3.1.10. Kullanılan araç gereçler	20
3.2. Materyal	21
3.2.1. Mezun öğrencilerin sosyal ağlardan tespiti	21
3.2.2. Mezun öğrenci bilgileri için veri ön işleme.....	21
3.2.3. Mezun öğrencilerin not bilgilerine ulaşılması	22
3.2.4. Veri indirgeme, dönüştürme ve eksik veri doldurma.....	23
3.2.5. Veriler ile ilgili istatistikî bilgiler.....	27
BÖLÜM 4.	
ARAŞTIRMA BULGULARI	31
4.1. Araştırma akış diyagramı	31
4.2. Modelleme	33
4.3. Regresyon ve korelasyon katsayısının bulunması.....	33
4.4. Apriori algoritması ile elde edilen sonuçlar.....	34
BÖLÜM 5.	
TARTIŞMA VE SONUÇ	41
5.1. Tartışma.....	41
5.2. Sonuç	44
KAYNAKLAR	46
EKLER	49
ÖZGEÇMİŞ	50

SİMGELER VE KISALTMALAR LİSTESİ

OİS : Öğrenci işleri sistemi
SABİS : Sakarya üniversitesi bilgi sistemi

ŞEKİLLER LİSTESİ

Şekil 3.1. Veri madenciliği süreci.....	12
Şekil 3.2. Veri madenciliği ile ilişkili disiplinler.....	14
Şekil 3.3. Tanımlayıcı ve öngörü veri madenciliği yöntemleri.....	16
Şekil 3.4. Eksik bilgi için kullanılan korelasyon matrisinin Rapidminer modeli	24
Şekil 3.5. Örneklemde yer alan mezun öğrencilerin sektörel dağılımı.....	28
Şekil 3.6. Mezun öğrenci ve çalıştıkları sektör dağılımı.....	29
Şekil 3.7. Sınıflandırılmış not durumlarının dağılımı.....	30
Şekil 4.1. Yapılan çalışmaya ait akış diyagramı.....	32
Şekil 4.2. Rapidminer programında hazırlanan modelin görüntüsü.....	33
Şekil 4.3. Örneklemde belirlenen dersler arasındaki regresyon katsayıları.....	34

TABLolar LİSTESİ

Tablo 3.1. Örnek alışveriş hareketleri	18
Tablo 3.2. Mezun öğrencilerin aldıkları dersler ve kısa adları.....	23
Tablo 3.3. Not durumlarının kategorik olarak ayrılması.....	24
Tablo 3.4. Mezun öğrencilerin ders notlarının kategorize edilmesi.....	25
Tablo 3.5. Mezun öğrencilerin çalıştıkları sektör tespiti.....	26
Tablo 3.6. Mezun öğrencilerin aldıkları dersler ve kategorileri.....	27
Tablo 4.1. Belirlenen sektörlere göre sıklık analizi.....	34
Tablo 4.2. Sektör ve bilgisayar bilimleri dersleri ilişkisi	36
Tablo 4.3. Sektör ve donanım dersleri ilişkisi	37
Tablo 4.4. Sektör ve yazılım dersleri ilişkisi.	38
Tablo 4.5. Sektör ve tüm dersler ilişkisi	40
Tablo 5.1. Fakülte seçmeli derslerin yıllara göre açık/kapalı olma durumu.....	43

ÖZET

Anahtar kelimeler: Veri madenciliği, apriori algoritması, sosyal ağlar.

Bilişim teknolojilerindeki gelişmeler ile birlikte, kişilerin şahsi ve mesleki yeterliliklerini göz önüne serdikleri sosyal iş ağları özellikle mesleki tecrübelerin öne çıktığı, yazılı kuralları olmayan ancak beyan edilen bilgilerin doğru olduğu kabul edilen bir nevi özgeçmiş yayınlama platformu haline gelmiştir.

Bu tez çalışmasında, lisans öğrenimini Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesinde tamamlayan 834 adet öğrencinin aktif olarak çalıştıkları kurum, kuruluş ve işletme bilgilerine sosyal iş ağlarından ulaşılarak mezun öğrencilerin sektör bilgileri etiketlenmiştir. Mezunların en çok bilişim, finans ve özel sektörde faaliyet gösterdiği, en az güvenlik ve medya sektörünü tercih ettiği gözlenmiştir. Etiketlenen sektör bilgileri ile öğrencilerin lisans eğitimleri sırasında aldıkları 10 adet dersin not ve başarı durumları temel istatistik yöntemleri ile işlenerek veri ambarı oluşturulmuştur. Apriori algoritması ve Rapidminer yazılımı kullanılarak mezun öğrencilerin sektörel eğilimleri ve öğrenim gördükleri dersler arasındaki başarı ilişkisi ortaya çıkarılmıştır. Ortaya çıkarılan “sektör-ders” ilişkileri göz önünde bulundurularak, Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi müfredatı için kullanılabilir faydalı bir model geliştirilmiştir.

ANALYSIS OF STUDENT PROFESSIONS GRADUATED FROM COMPUTER AND INFORMATION SYSTEMS ENGINEERING THROUGH SOCIAL NETWORKS USING DATA MINING TECHNIQUES

SUMMARY

Keywords: Data mining, apriori algorithm, social networks

Along with the developments in information technologies, the social business networks which the personal and professional competences of the people are taken into consideration, have become a platform for publishing resumes that do not have written rules and are especially prominent in their professional experience.

In this thesis, 834 students who completed their undergraduate studies at the Faculty of Computer and Information Sciences of Sakarya University were accessed from the social business networks and the sectoral information of the graduates was labeled. It has been observed that the graduates most preferred in the Information, Finance and Private Sector, the least in the Security and Media sectors. The data warehouse was formed by processing the sectoral information labels and the grades and achievements of the 10 courses that the students took during their undergraduate education using basic statistical methods. Apriori algorithm and RapidMiner software were used to reveal the relationship between the sectoral trends of graduate students and the courses they have studied. A useful model with taking into consideration the “sector-course” associations was developed for the Sakarya University Computer and Information Sciences Faculty curriculum in the following years.

BÖLÜM 1. GİRİŞ

Bilişim teknolojisindeki gelişmeler ile teknolojik hizmetlerin kullanımı ve erişimi gün geçtikçe artarken, insanların sosyal ağlarda bilgi paylaşımı da artarak bir nevi tanıtım alanı haline gelmiştir. Sosyal ağlar insanların dünyanın çeşitli yerlerinden yaş, ırk, cinsiyet farketmeksizin fikirlerinden deneyimlerine, hislerinden hobilerine ve dahi resim ve videolarına kadar paylaştıkları dijital ortamdır. Bu durum kamu kurumlarının ve özel kuruluşların sosyal medyada bulunan verilerden insanları analiz ederek kendi süreçlerini geliştirmesine imkân sağlamaktadır. Bu nedenle sosyal medyanın akademik ve endüstri tarafından yapılan araştırmalarda rolü büyüktür.

Sosyal ağların bir kolu olan kişisel iş ağları, mesleki tecrübelerini aktarmak ve kendi alanlarındaki kişiler ile daha kolay iletişim halinde bulunmak için kullanıldığı gibi aynı zamanda kişilerin mesleki evreleri de görebilmemizi sağlamaktadır. Sosyal iş ağları kişilerin özgeçmişleri ile mesleki yetkinliklerini ortaya koymak, benzer veya farklı iş ağlarında yeni iş ortamları bulabilmek ve dahası kurulan iş ilişkileri ile kariyer hedefi yaratabilmeyi sağlamaktadır.

Günümüzde sosyal ağlar kullanılarak kişilerin verilerine erişebilmek, bu verilerden yola çıkarak çeşitli değerlendirmeler yapmak oldukça kolaylaşmıştır. Sosyal ağlarda oluşan büyük miktardaki veriler, yönetici veya analistler gibi karar vericiler için ancak işlenmiş veriye dönüşebildiğinde anlamlandırılmaktadır. Bu verilerin toplanması ve anlamlı hale getirilmesi için kullanılan en temel yöntemlerden biri de veri madenciliğidir. Büyük ölçekli veriler kullanılarak “değerli olan” bilgiyi elde edip ondan faydalanma durumu ile tanımlanan veri madenciliği ile, verilerin birbiri ile ilişkilerini ortaya çıkarmak, bu ilişkiler kullanılarak çeşitli tahminler yapmak ve gizli kalmış örüntü ve eğilimleri keşfetmek mümkündür (Özkan, 2008).

Diğer mühendislik alanlarında olduğu gibi bilgisayar ve bilişim mühendisliği de mühendislik olarak, eldeki kaynakların en verimli şekilde kullanıp mevcut problemler için en iyi çözüm üretmeyi ve böylece insan hayatını kolaylaştırmayı amaçlar. Bilgisayar ve Bilişim Mühendisliği ise bilgi sistemlerini tasarlama, tasarlanan sistemlerin işlerliğini sağlama, çeşitli kaynaklardan veri elde etme, elde ettiği veriyi işleme ve değerlendirme çalışmalarında bilgisayarları kullanılır kılma konularında gereken desteği verir. İletişim şirketlerinde, finans kurumlarında, savunma endüstrisi kuruluşlarında, her türlü bilgi işlem merkezlerinde, tıbbi yazılım endüstrisinde, animasyon, ses ve görüntü işleme, güvenlik, film-eğlence-medya sektörü gibi birçok alanda bilgisayar mühendislerine ihtiyaç vardır. Bu yüzden bilgisayar ve bilişim mühendisliği bölümü mezunları 21. yüzyılda her iş kolunda çalışabilir.

Bilgisayar ve bilişim mühendisliği mezunları da diğer bilim dallarında olduğu gibi eğitim öğretim hayatları süresince aldıkları dersler ve yaptıkları çalışmalar ile iş hayatına hazırlanmaktadır. Mezuniyet sonrası çalıştıkları sektörlerdeki tercih sebepleri birçok etkene bağlı olup, bu etkenlerden bir tanesi de eğitim öğretim hayatında aldıkları dersler ve bu derslerdeki başarı yatkinlığı ile ilişkilendirilebilir.

Bu çalışmada sosyal ağlardan elde edilen Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi mezunu öğrencilerin çalıştıkları sektör analizi ile derslerdeki yatkinlıkları incelenmiş olup, neticesinde mezunların sektörel eğilimleri tespit edilmiştir.

Bu bölümde çalışmadaki temel motivasyon kaynağı açıklanırken, çalışmanın ikinci bölümünde literatür taraması yapılmıştır.

Üçüncü bölümünde veriden anlamlı bilgi elde etmeyi sağlayan veri madenciliği kavramı anlatılmış, alt başlıklarda veri madenciliği yöntem, süreç ve uygulama aşamaları anlatılarak çalışmaya genel bir bakış açısı kazandırılmıştır. Ayrıca sosyal ağlardan verilerin yani mezun öğrenci bilgilerinin toplanması, toplanan verilere uygulanan işlemler, kabul ve analizler, verilerden sektör tespiti, mezunların ders ve not bilgilerine erişilerek notların kategorize edilmesi ile uygulamanın materyal ve metodu anlatılmıştır. Sektör tercihleri tablolar halinde gösterilmiş ve yorumlanmıştır.

Dördüncü bölümde ise bu çalışmanın konusu, amacı, önemi, kabul ve varsayımları, uygulama sonuçları ayrıntılı şekilde anlatılmıştır. Nihai olarak elde edilen veriler kullanılarak mezunların sektör eğilimleri ve ders/not bilgilerine birliktelik kuralları uygulanmış, uygulama aşamaları detaylı şekilde anlatılarak konuyla ilgili çeşitli çıkarımlarda bulunulmuştur.

Son bölümde ise çalışmadan çıkarılan sonuçlar ve gelecek projeler için önerilerde bulunulmuştur.

BÖLÜM 2. LİTERATÜR ÖZETİ

Günümüz şartlarında üniversitelerin elektronik verilerindeki büyüme ile depolanan geniş hacimdeki veriler hızla artmaktadır. Veri madenciliği alanındaki ilerlemeler kullanılarak anlamlandırılan bu bilgiler ise çalışılan süreçlerin kalitesini artırmakta önemli rol oynamaktadır.

2017 yılında yapılan bir çalışmada, lisans öğrencilerinin performansını incelemek için veri madenciliği yöntemleri kullanılarak dört yıllık bir çalışma programının sonunda öğrencilerin akademik başarıları tahmin edilmiştir. Çalışma sonucunda başarısız öğrencilere zamanında uyarı ve destek sağlanarak, yüksek performans gösteren öğrencilere tavsiye ve fırsatlar sunarak performans arttırılmasının mümkün olduğunu göstermektedir (Asif ve ark., 2017).

Dokuz Eylül Üniversitesi'nde yapılan başka bir çalışmada uzaktan eğitim programlarına kayıtlı olan öğrencilerin tamamı olan 1952 öğrenciye kurum üzerinden e posta gönderilmiştir. Ankete katılan 285 öğrenci verileri kullanılarak veri madenciliği teknikleri olan kümeleme ve sınıflama analizleri ile “yüksek akademik fayda beklentisi” ve “teknolojik ve yeniliklere yakın-İlgili olma nedeniyle” sebepleri yüzünden uzaktan eğitimi tercih eden öğrencilerin en yüksek başarıya sahip olduğu ortaya çıkmıştır (Aksaraylı, 2017).

Öğrencilerin başarı durumlarının devamsızlıkla ilişkisinin analiz edildiği başka bir araştırmada, apriori algoritması kullanılarak birliktelik kuralları çıkartılmıştır. Araştırma sonucunda, cinsiyet, bölüm türü, öğretim türü, kayıtlı olunan yılın genel başarı durumu ile birlikte devamsızlık eğilimi üzerinde etkileri olduğu saptanmıştır (Dalkilic ve Aydın, 2017).

Başka bir çalışmada Facebook sosyal ağı vasıtasıyla elde edilen veri kümesi üzerine çok katmanlı algılayıcı ve bayes ağlar gibi sınıflandırıcı ağlar uygulanarak çalışma sonucunda sosyal medya madenciliği yapılabilmesi için uygun sınıflandırıcı yöntem önerilmeye çalışılmıştır (Ertam, 2017).

2010 yılında Ankara ilindeki trafik kazaları veri kümesi kullanılarak birliktelik kuralları oluşturulmuş ve apriori algoritması ile kural çıkarımı çalışması yapılmıştır. Çalışma sonucunda hava koşulları, yol türleri ve hasar durumları ile oluşturulan kurallar değerlendirilmiştir (Doğan ve ark., 2016).

Mehmet Akif Ersoy Üniversitesi'nde yapılan başka bir çalışmada Yönetim Bilişim Sistemleri Bölümü öğrencilerinin derslerdeki başarı durumları veri madenciliği yöntemleri ile incelenmiştir. Hazırlık sınıfı okuma durumu ve dersin kategorisinin öğrencinin başarı durumu ile ilişkili olduğu ortaya çıkmıştır (Çeşmeli ve ark., 2015).

Eskişehir Osmangazi Üniversitesi Bilgisayar Mühendisliği Bölümü'nde yapılan çalışma kapsamında apriori ve predictive apriori uygulanarak atılma politikasındaki değişiklik ile öğrencilerin not durumları arasındaki ilişki bulunmuş, öğrencilerin eğitim sürelerinin, burs veya kredi alınmasıyla değişiklik gösterdiği ve öğrencilerin parasal durumlarıyla annelerinin meslekleri arasında bağlantı olduğu görülmüştür (Kılınç, 2015).

Bir başka çalışmada apriori algoritması kullanılarak Bilecik Şeyh Edebali Üniversitesi Bilgisayar Mühendisliği Bölümü'nde okuyan öğrencilerin seçmeli ders tercihlerinin belirlenmesi amacıyla öğrencilerden anket yöntemi ile gerekli bilgiler alınmış ve ders seçimi yaparkenki kriterleri belirlenmiştir (Güngör ve ark., 2013).

Öğrencilerin akademik başarılarına etki eden faktörlerin belirlenmesi için yapılan bir diğer çalışmada veri madenciliği yöntemlerinin başarı ve başarısızlığa etki eden faktörlere farklı bir bakış açısı yarattığı tespit edilmiştir (Kurt ve Erdem, 2012).

Veri madenciliği tekniklerinden K-Means kullanılarak yapılan başka bir çalışmada Selçuk Üniversitesi öğrencilerinin ortak zorunlu derslerdeki başarı analizi yapılmıştır (Yavuz ve ark., 2011).

Selçuk Üniversitesinde yapılan başka bir çalışmada ise web üzerinden hazırlık sınıfında okuyan öğrencilere bir anket yapılmıştır. Anket verilerine veri madenciliği yöntemlerinden apriori ve karar ağacı algoritmaları uygulanmış olup, öğrencilerin başarılarını etkileyen faktörler bulunmaya çalışılmıştır. Öğrencinin başarısında en önemli faktörün aile eğitim seviyesi ve maddi durum olduğu ortaya çıkmıştır (Ekim, 2011).

Bir diğer çalışmada Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği bölümündeki yazılım dersleri için öğrencilere anket yapılmıştır. Anket verileri, yazılım geliştirme projelerinde süreç adımlarının takibini kolaylaştırmıştır (Cihan ve ark., 2013).

Uzaktan eğitim için oluşturulan öğretim materyali ve öğrenci başarısı üzerindeki etkisinin incelendiği başka bir çalışma sonucunda öğrenci başarısının hangi materyallerle daha olumlu etkilendiği belirlenmiştir (Savaş ve Arıcı, 2009).

2010 yılında sosyal ağ kullanımına yönelik yapılan çalışmada sosyal paylaşım sitesi facebook üzerindeki kullanıcıların genel özellikleri incelenmiştir. Facebook kullanım süresi ve erişim sıklığı incelenerek çeşitli tespitler yapılmaya çalışılmıştır (Bozkır ve ark., 2010).

Başka bir çalışmada web madenciliği yöntemi kullanılarak kişilerin yaş, cinsiyet, yaşadığı yer, lisans düzeyi gibi özelliklerine bakılıp, birbirleriyle olan benzerliklerinin bulunması hedeflenmiştir. Bunun için veri madenciliği yöntemlerinden apriori algoritması uygulanmış ve ortaya çıkan sonuçlar tartışılmıştır (Onat, 2008).

2014 yılında yapılan çalışmada ise öğrencilere ait çevrimiçi öğrenme ortamındaki veriler kullanılarak; dönem sonundaki akademik performansların tahmin edilmesi sağlanmıştır (Akçapınar, 2014).

Eđitim alanında yapılan veri madenciliđi alıřmalarının ortak zelliklerinin đrencinin başarı performansı ve đrenim hayatlarındaki eğilimlerin tespitine yönelik olduđu görölmüřtür.

BÖLÜM 3. MATERYAL VE YÖNTEM

3.1. Yöntem

3.1.1. Veri madenciliğine giriş

Günümüz teknoloji çağında insanların makinelerin ve hatta internetin bile veri kaynağı sayıldığı bilinmektedir. Gelişen teknoloji ile birlikte internet ortamında veri sürekli kendini güncelleyen, giderek artan ve çeşitlenen bir hal almıştır. Öyle ki marketten alışveriş yapmaktan, günlük hayatta zaman geçirmek için kullanılan sosyal ağlara kadar sistematik olarak yaptığımız her teknolojik hareket ile istemli veya istemsiz veri oluşturulmaktadır.

Bu verilerin katlanarak devasa boyutlara geldiği ve geleceği ise öngörülen bir gerçektir. İşte böyle büyük miktar veriden oluşan bir ortamda veri analizi için kullanılan en temel kavram veri madenciliğidir.

3.1.2. Veri madenciliğinin tanımı

Herhangi şekilde bir araya gelen anlamlı ve anlamsız veri kümelerinden teknolojinin sağladığı avantajlar ve çeşitli algoritmalar sayesinde farklı disiplinlerin de bir araya gelerek işlevsel veriler ortaya çıkarılmasına veri madenciliği denir. Mevcut bilgi sistemleri ile bu verileri işlemek, anlamlandırmak ve nitelikli bilgi haline dönüştürmek ise bu çalışma ve benzer tüm çalışmaların temel amacıdır. İşte bu noktada kullanılan veri madenciliği ise büyük miktardaki verilerden daha önce ortaya çıkarılmamış ve henüz keşfedilmemiş bilginin, çeşitli bilgisayar teknolojileri, istatistiki çalışmalar ve diğer disiplinler bir araya getirilerek anlamlı bilgi haline dönüşmesini sağlar. Veri madenciliği ile ortaya çıkarılan örüntülerin güvenilir, anlamlı, anlaşılabilir, kullanışlı

ve geçerli olması gerekmektedir (Holsheimer ve Siebes, 1994; Fayyad ve ark., 1996; Cabena ve ark., 1998)

Veri madenciliği literatürde çeşitli araştırmacılar tarafından aşağıdaki gibi tanımlanmıştır.

“Veri madenciliği; büyük veri setindeki, anlamlı, orijinalliği olan, kullanım potansiyeli bulunan ve sonuçta anlaşılabilir olan örüntülerin çıkarılmasıdır” (Fayyad ve ark., 1996).

“Tek başına ham verinin sunamadığı bilgiyi ortaya çıkaran veri analizi sürecine veri madenciliği denir “ (Jacobs, 1999).

“Mevcut bilgi sistemlerinin işleyemeyeceği geniş ve karmaşık bu veri kümeleri, büyük veri olarak tanımlanmaktadır ” (Gürsakal, 2014).

“Veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır” (Kittler ve Wang, 2000).

“Veri madenciliği istatistik, veri tabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veri tabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır” (Hand, 1998).

“Veri madenciliği, büyük ölçekli veriler arasından değeri olan bir bilgiyi elde etme işidir” (Özkan, 2008).

“Veri madenciliği, büyük miktarlardaki veriyi bilgiye dönüştürmektir ” (Han ve ark., 2011).

“Veri madenciliği, veri içerisindeki kullanışlı örüntülerin bulunması sürecidir ” (Roiger ve Geatz, 2013).

“Veri madenciliği veya veri tabanlarında bilgi keşfi, veri dizilerinden geçerli, yeni, mümkünse faydalı ve anlaşılır örüntülerin ortaya çıkartılabilmesi için gerçekleştirilen apaçık olmayan bir süreçtir ” (Akpınar, 2014).

3.1.3. Veri madenciliğinin tarihçesi

Veri madenciliği tanımları ilk olarak 1990’ların başında ortaya atılsa da, bu tarihten önce de veri madenciliği yapılmaktadır.

- 1950’li yıllarda bilgisayarlar matematikçiler tarafından sayma işlemleri için kullanılırken 1960’larda bilgisayarların şimdilerde yapay sinir ağları ile adlandırılan basit öğrenme yapabileceği fikri ortaya atılmıştır (Adriaans ve Zantinge,1996).
- 1970’li yıllarda ise veri tabanı yönetim sistemlerinin ortaya çıkarılması ile temel seviyede uzman sistem ve makine öğrenmesi işlemleri yapılabilmektedir.
- 1980’li yıllarda ise veri tabanı yönetim sistemleri kullanımı büyük ölçekli işletme ve kurumlarda artmıştır. Yaygınlaşan kullanım sebebiyle veri tabanlarında biriken verilerden faydalı bilgilerin nasıl bulunabileceği soruları ise veri madenciliğinin doğuşuna zemin hazırlamıştır.
- 1990’larda büyük veri tabanları için yaratılan veri ambarlarının oluşturulması ve veri madenciliği çalışmalarından ilki sayılan “Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop” makalesiyle veri tabanlarındaki bilgi keşfi sürecinin temel kavramları ortaya çıkmıştır (Piatetsky-Shapiro, 1990).
- İlerleyen yıllarda ilk veri madenciliği yazılımlarının gerçekleşmesi ile veri madenciliği süreci resmen başlamıştır.

3.1.4. Veri tabanlarında bilgi keşfi

Geniş veri tabanlarında var olan veri kullanılarak daha önceden bilinmeyen ve yeni keşfedilen bilginin çeşitli süreçlerde ve işletmelerde karar verilirken kullanılmasına

veri tabanlarında bilgi keşfi denilmektedir. Yapılan bu keşif ise veri madenciliği yöntemleri ile ortaya çıkarılmaktadır.

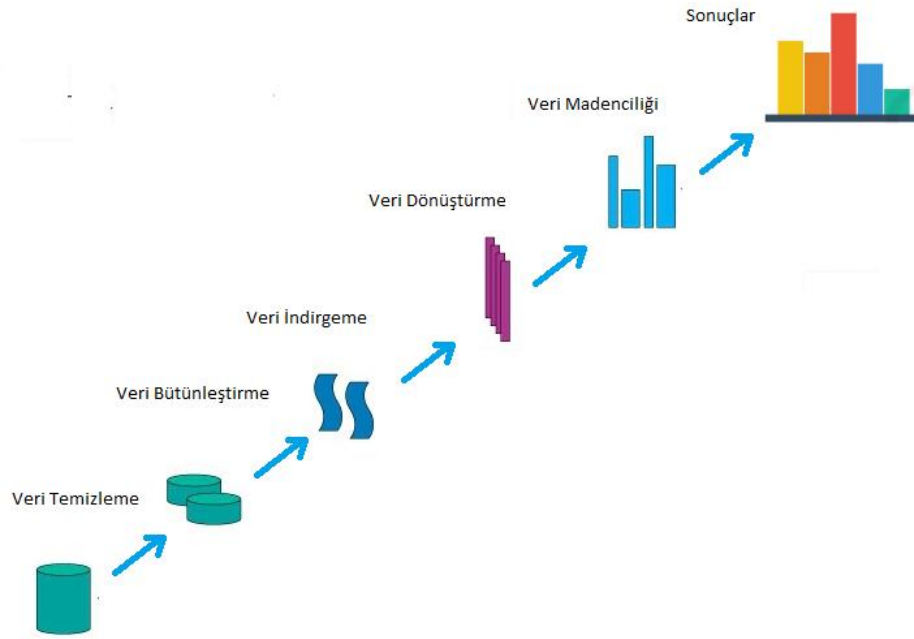
Veri madenciliği için birçok kavram kullanılmaktadır. Bu kavramlardan bazıları

- Veri tabanlarında bilgi madenciliği,
- Bilgi çıkarımı,
- Veri ve örüntü analizi,
- Veri arkeolojisi,
- Veri tabanlarında bilgi keşfi' dir.

Bu kavramların arasındaki en bilinen kullanım “veri tabanlarından bilgi keşfi”dir (Akgöbek Ömer, 2011).

Bir araya getirilen büyük ölçekli verilerin değerlendirilmesi amacıyla istatistik ve yapay zekâ tekniklerinin kullanılması sonucunda veri madenciliği ortaya çıkmıştır (Argüden ve Erşahin, 2008).

Veri madenciliği çalışmalarında Şekil 3.1.’de gösterilen veri ön işleme işlemleri uygulanmaktadır (Özkan, 2008).



Şekil 3.1. Veri madenciliği süreci (Özkan, 2008)

Veri temizleme: Veriler çeşitli kaynaklardan temin edilebildiği gibi mevcut bir sistemden de alınabilir. Bu noktada veriden anlamlı bilgiyi ortaya çıkarabilmek için yapılması gereken kayıp veya kirli verilerin tespit edilmesi ve tespit edilen verilere uygulanacak işlemi belirlemektir. Kayıp veya kirli veri için yapılabilecek çeşitli işlemler vardır. Bunlar:

- Kayıp verinin bulunduğu kaydı silmek,
- Kayıp verileri teker teker doldurmak,
- Tüm kayıp verilere aynı bilgiyi girmek,
- Tüm verilerin ortalama değerini girmek,
- Regresyon yöntemi ile kayıp verinin diğer bir değişken ile tahmin edilmesidir.

Veri bütünleştirme: Farklı veri kaynaklarından derlenen verilerin birlikte değerlendirilmesi amacıyla tek bir veri türüne dönüştürülmesi işlemidir.

Veri indirgeme: Uygulama yapılırken bazı veri öbeklerinin çözümleme aşamasında sonucu değiştirmedeği tespit edilmesi halinde veri sayısı ya da değişkenin azaltılması

işlemdir. Böylelikle veriler arasından seçme işlemi yaparak gereksiz veriler çıkarılarak boyut azaltılması yapılır (Özkan, 2008).

Veri dönüştürme: Veri içerdiği anlam bakımından algoritmayı çalıştırmaya uygun olmadığı durumlarda verinin belirlenen bir işlem ile dönüştürülüp kullanıma alınması sağlanmalıdır. Bu durumlarda normalleştirme veya standartlaştırma gibi işlemler ile veri dönüştürme yapılır (Özkan, 2008).

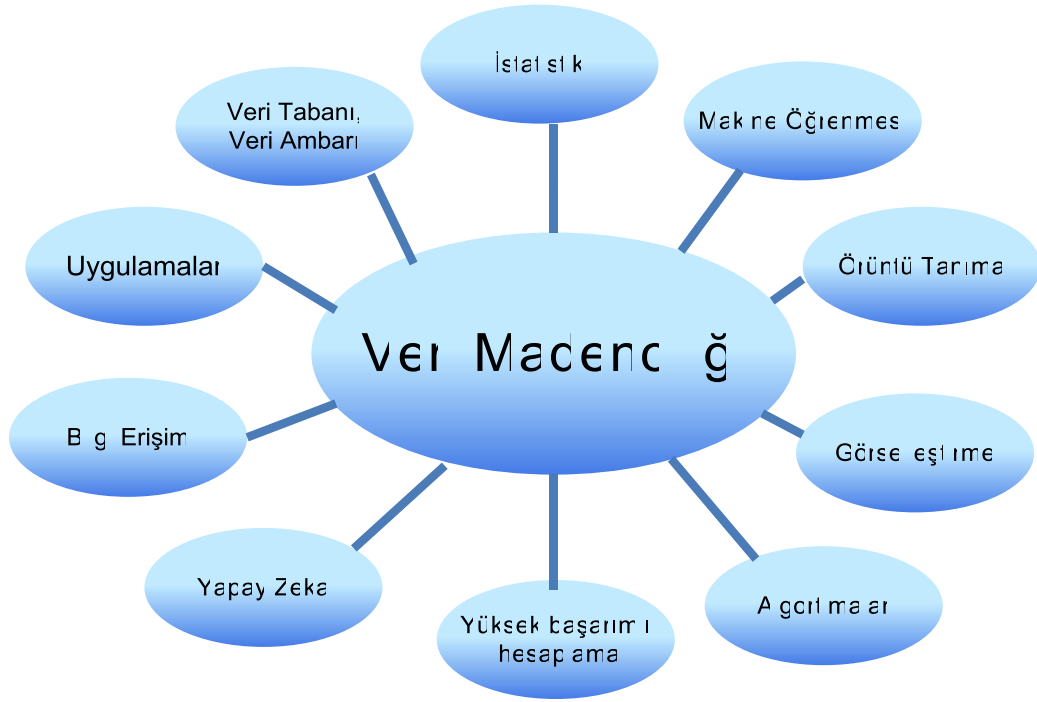
Veri madenciliği: Veri madenciliği veri üzerinde yapılan işlemlerden sonra veriler arasındaki ilişkinin ortaya çıkarılması amacıyla algoritma çalıştırma işlemidir.

Sonuçlar: İşlem sonucunda ortaya çıkan değerli bilginin geçerli, yararlı ve bilinmeyen bir bilgi olup olmadığının değerlendirilmesidir.

3.1.5. Veri madenciliğini oluşturan disiplinler

Veri madenciliği sadece veri analizi yapmak olmadığı için beraberinde farklı disiplinlerden de beslenmektedir. Farklı disiplinlerin birlikte çalışılmasıyla veri madenciliği sürekli gelişmekte ve uygulama alanını da büyötmektedir.

Veri madenciliği, Şekil 3.2.'de gösterildiği gibi makine öğrenmesi, veri tabanı, yapay zekâ gibi disiplinlerle ilişkilidir.



Şekil 3.2. Veri madenciliği ile ilişkili disiplinler (Han ve ark., 2011)

3.1.6. Veri madenciliği uygulama alanları

Son yıllarda bilgisayar ve internet kullanımının artması ve kişilerin her yerden bilgiye ulaşılabilmesiyle, bilgisayarlarda yapılan her bir işlemin ve dahası teknoloji ile gerçekleşen her bir hareketin veri haline dönüşmesine olanak sağlamıştır. Veri madenciliği ise bu kadar büyük bilgilerin faydalı sonuçlar üretebilmesi açısından doğru bir yöntem olarak gün geçtikçe daha fazla kullanılmaktadır. Veri madenciliği tekniklerinin uygulandığı alanlar aşağıda kısaca özetlenmiştir.

- Pazarlama
- Bankacılık ve Finans
- Sigortacılık
- Eğitim
- E-Ticaret
- Sağlık
- Devlet Uygulamaları

- Haberleşme ve İletişim
- Mühendislik Uygulamaları
- Üretim ve İmalat
- Biyomedikal ve DNA

3.1.7. Veri madenciliği yöntemleri

Veri madenciliği yapılması düşünülen verilerden istenilen bilgiyi elde etmek için uygun yöntem bulmak ve bu yöntemi veri setine uygularken veriyi de seçilen yönteme uygun şekilde değiştirmek gerekir. Bu sebeple seçilecek yöntem verinin doğru işlenmesi açısından önemlidir.

Veri madenciliği yöntemleri tanımlayıcı ve öngörü yöntemleri olmak üzere iki ana başlık altında toplanmıştır (Maimon ve Rokach, 2007).

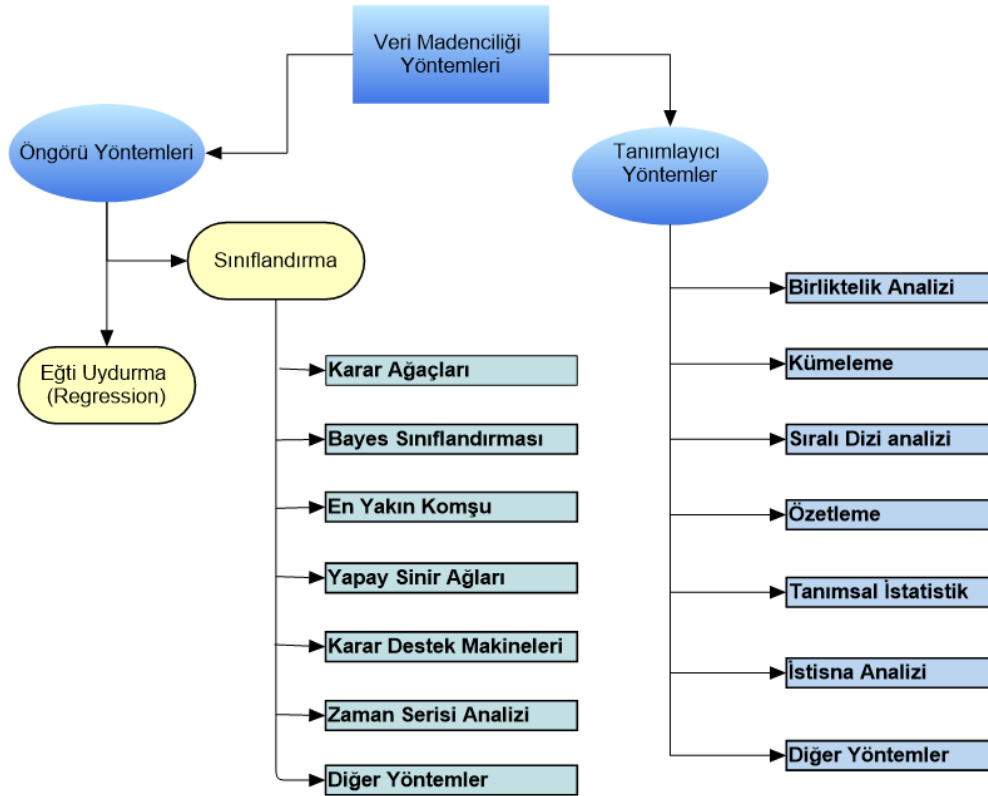
3.1.7.1. Tanımlayıcı model

Veri tabanındaki verilerin analizi sonucu verinin genel olarak neyi ifade ettiğini veya mevcut durumunu ortaya çıkarmaya yönelik yöntemleri içerir. Karar vermeye yönelik verilerden çıkarılacak örüntülerin tespiti mümkündür. Belirli bir aralıktaki veriyi ortaya çıkarmak, aynı anda birden fazla nitelikte aynı bilgiyi içerenlerin durumunu ve bu örüntünün veri kümesindeki benzerliklerini tespit etmek tanımlayıcı modele örnektir (Farboudi, 2009). Kısaca daha önceden tespit edilememiş, henüz ortaya çıkmamış bilgiyi belirlemekte kullanılır. En sık kullanılan tanımlayıcı model uygulamaları kümeleme, özetleme ve birliktelik kurallarıdır (Dunham ve Ming, 2003).

3.1.7.2. Öngörü (Tahmin) modeli

Verilerden geleceğe yönelik tahmin yapma veya sonuç çıkarma amacı güdülür. Sıklıkla kullanılan yöntemlerden biri danışmanlı öğrenme yöntemidir. Bu yöntemle sonuçları bilinen mevcut veriden sonuçlarla birlikte tahmin için kullanılacak bir model oluşturulur. Oluşturulan model sonucu bilinmeyen bir veri setine uygulanarak öngörü

yapılır. Bir diğer yöntem olan danışmansız öğrenme yönteminde ise mevcut verinin sonuçları hariç modele girdi olarak verilmesi, gerçek sonuçları kendi kuralları ile tahmin etmesi istenir. Tahmin sonucu mevcut sistemde iyileştirme yapmak veya geleceğe yönelik planlar yapmakta kullanılır. Örneğin bir önceki yıla ait sipariş bilgileri kullanılarak geleceğe yönelik üretim planlaması yapmak öngörü modeline örnektir. Öngörü modeli sınıflandırma ve eğri uydurma olmak üzere iki kısma ayrılmaktadır. Bu yaklaşımdan yola çıkarak geliştirilen en temel iki teknik karar ağaçları ve bayes sınıflandırmadır (Silahtaroglu, 2008). Şekil 3.3.'te öngörü ve tanımlayıcı veri madenciliğinde kullanılan yöntemler listelenmiştir.



Şekil 3.3. Tanımlayıcı ve öngörü veri madenciliği yöntemleri (Silahtaroglu, 2008; Han ve ark., 2011)

Veri madenciliğinde tanımlayıcı ve tahmin edici modeller kesin çizgilerle birbirlerinden ayrılmazlar. Eğer model anlaşılabilir ise tahmin edici model tanımlayıcı model yerine, tanımlayıcı modeller ise tahmin edici model yerine kullanılabilir (Velickov ve Solomatine, 2000).

3.1.8. Birliktelik kuralları

Veri madenciliğinde en sık kullanılan ilk teknik birliktelik kurallarıdır (Agrawal ve ark., 1993). Birliktelik kuralları belirli türdeki veri ilişkilerini tanımlayan modeldir (Silahtaroglu, 2008). Başka bir deyişle verilerin birlikte bulunma / gerçekleşme durumlarının tespitinde birliktelik kuralları kullanılmaktadır (Özkan, 2008).

Birliktelik kuralları uygulamalarında pazar sepet analizi örnek verilebilir (Piatetski ve Frawley, 1991). Yapılan alışverişlerde ürünler arasındaki birlikteliği bularak kişilerin alışveriş alışkanlıkları tespiti yapılır. Pazar sepeti analizinde örneğin “müşteriler bira satın aldığı anda %75 ihtimalle cipte satın alırlar“ sonucu çıkarılarak bu iki ürün arasında kuvvetli bir ilişki olduğu sonucu tespit edilir (Farboudi, 2009). Ortaya çıkan sonuçlar kullanılarak daha etkin satış stratejileri, market raf düzenlemesi veya kampanyalar oluşturulabilir.

Eğitim alanında yapılan veri madenciliği çalışmalarında birliktelik kurallarının kullanılmasıyla elde edilen verilerden oluşturulan öğrenci modellemesi ile başarı durumları ve ders eğilim tahminlerinde bulunularak eğitim sistemi geliştirilebilir (Agrawal ve ark., 1993).

3.1.8.1. Güven ve destek kavramları

Veri kümeleri arasındaki ilişkileri ortaya koymak için “destek” ve “güven” gibi iki kavram vardır. En sık kullanılan pazar sepet analizinden örneklemek gerekirse, Destek değeri bir ürünün tüm alışverişler içinde tekrar oranını belirtir. Güven değeri ise, A ürün grubunu alan müşterilerin B ürün grubunu da A ile birlikte alma olasılığını belirtir. A ürün grubunu alanların B ürün grubunu da alma durumu, yani birliktelik kuralı $A \rightarrow B$ biçiminde gösterilir. Bu durumda destek değeri Denklem 1’deki gibi ifade edilir (Özkan, 2008).

$$\text{Destek}(A \rightarrow B) = \text{sayı}(A,B) / N \quad (3.1.)$$

Denklem 3.1.'de $\text{sayı}(A,B)$ destek sayısı A ve B ürün gruplarını birlikte içeren alışveriş sayısını göstermektedir. N ise tüm alışverişlerin sayısını göstermektedir. A ve B ürün gruplarının birlikte satın alınması olasılığını ifade eden güven değeri ise Denklem 3.2.'deki gibi hesaplanır (Özkan, 2008).

$$\text{Güven}(A \rightarrow B) = \text{sayı}(A,B) / \text{sayı}(A) \quad (3.2.)$$

Veriler arasındaki bağıntının önemli olması destek ve güven kriterlerinin yüksekliği ile ölçülmektedir.

3.1.8.2. Birliktelik kuralları uygulaması

Tablo 3.1.'de Birliktelik Kuralları oluşturmak için örnek alışveriş hareketi listesi verilmiştir.

ID	Ürünler
1	Etek, Kazak, Hırka, Pantolon
2	Etek, Hırka, Kazak, Balık
3	Ayakkabı, Kazak, Hırka, Pantolon
4	Kazak, Hırka, Pantolon
5	Etek, Ayakkabı, Hırka, Pantolon

Tablo 3.1.'deki hareketlere bakılarak Hırka ve Pantolonun Etek ile olan ilişki analizi Denklem 3.3.'deki gibi yapılabilir.

$$\% \text{Destek} = \frac{(\text{Hırka,Pantolon,Etek})}{\text{Toplam Alışveriş Hareketi}} = 2/5 = 0,4 \quad (3.3)$$

Denklem 3.3'te toplam alışveriş hareketi tüm alışveriş listesi için işlem sayısını ifade ederken, (Hırka, Pantolon, Etek) ise Hırka, Pantolon ve Etek üçlüsünün aynı anda yer aldığı işlem sayısını gösterir.

$$\%Güven = \frac{(Hırka,Pantolon,Etek)}{(Hırka,Pantolon)} = 2/4 = 0,5 \quad (3.4)$$

Denklem 3.4'te (Hırka, Pantolon, Etek) üçlüsü Hırka, Pantolon ve Etek ürünlerinin aynı anda bulunduğu işlem sayısını, (Hırka,Pantolon) ise Hırka ve Pantolonun birlikte bulundu işlem sayısını ifade etmektedir.

Denklem 3.3. ve 3.4.'e bakıldığında {Hırka,Pantolon) alanların Etek alma kuralı için %40 destek ve %50 güven değeri ortaya çıkmıştır. Yapılan hesaplamalara göre Hırka ve Pantolonu birlikte alanların Etek alma durumu %50 iken, bu durumların veri setinde birlikte bulunma oranı %40'tır.

Birbiriyle ilişkili farklı olayların birlikte belirlenmesinde kullanılan çok büyük hacimli veriler üzerine veri madenciliği teknikleri uygulandığında, birliktelik kuralları ortaya çıkarmak için uygulanan algoritmalar hızlı şekilde çalışmalıdır (Agrawal ve ark., 1993).

Birliktelik kuralı çıkarmak için geliştirilen bazı algoritmalar ise şunlardır;

- AIS Algoritması,
- SETM Algoritması,
- Apriori Algoritması ,
- Partition Algoritması,
- RARM Algoritması ,
- CHARM Algoritması (Özçakır ve Çamurcu, 2007)

Bu algoritmalarından ilk kullanılan AIS olmakla birlikte, en bilinen algoritma ise apriori algoritmasıdır (Agrawal ve Srikant, 1995).

Bu çalışmada birliktelik analizi için apriori algoritmasının Rapidminer programında kullanılan özel hali Fp-growth algoritması kullanılmıştır.

3.1.9. Apriori algoritması

Agrawal ve arkadaşları birliktelik kuralının matematiksel modelini 1993 yılında ortaya çıkarmışlardır (Agrawal ve ark., 1993). Apriori algoritması birliktelik kurallarının uygulanmasında en yaygın yöntem olmakla birlikte, algoritma bir önceki adımın bilgilerini kullanarak iteratif olarak sık geçen öge kümesinin tekrarını tespit eden niteliğe sahiptir (Han ve ark., 2006). Birliktelik analizi sık geçen verilerin bulunması ve bu öğelerden güçlü birlikteliklerin tespit edilmesi olarak 2 aşamada gerçekleşir.

- Sık geçen nesne kümesinin bulunması; her nesnenin veri setinde bulunma sıklığının minimum destek değerinden büyük olması,
- Bulunan bu kümelerden güçlü ilişkilerin tespit edilmesi; minimum destek ve güven değerlerinin sağlanmasıdır.

Apriori algoritması sık geçen verilerin tespitinde kullanılarak ilk aşamayı gerçekleştirir. Algoritma ilk taramada her nesnenin tek tek destek değerini yani veri setinde tekrarlama sayısını hesaplayarak kullanıcı tarafından belirlenen destek değeri ile karşılaştırır. Sonraki aşamada ise bir önceki taramadan elde edilen nesnelere başlar ve nesne kümeleri oluşturur. Bulunan bu nesne kümelerinin destek değerleri hesaplanır. Tüm işlem adımları yeni nesne kümesi bulunamayınca kadar devam eder (Agrawal ve Srikant, 1994; Han ve ark., 2006).

Fp-growth algoritması ise 2000 yılında geliştirilmiş olup Apriori ve diğer birliktelik kuralı algoritmalarına kıyasla böl-yönet stratejisi uygulayarak daha performanslı çalışmaktadır. Diğer algoritmalar veri setini defalarca tararken, Fp-growth, tarama işlemini yalnızca iki kez gerçekleştirmektedir. Kurallar ilk taramada destek değerlerine göre sıralanır, ikinci taramada sıkıştırılır ve ağaç veri yapısına yerleştirilir. Belirlenen güven destek değerine göre kurallar seçilir.

3.1.10. Kullanılan araç-gereçler

SABİS altyapı ve bilgiye erişim itibariyle Türkiye'deki üniversitelere referans olabilecek niteliktedir. OIS yeni gelen öğrencilerin çevrimiçi kaydedildiği, öğrencilere

ait kişisel bilgilerin, ders/not bilgilerinin tutulduğu, sınav bilgi ve dokümanlarına çevrimiçi erişebilen, sanal ders/sınav yapılabilen, mezun öğrencilerin arşivlenip, e devlet üzerinde anlık bildirimlerin yapıldığı, harç bilgilerinin bankalar ile çevrimiçi çalıştığı, teknolojinin tüm imkânlarını en iyi şekilde kullanan bir sistemdir. Ayrıca tüm bu işlemlerin mobil olarak da yapılabilmesine imkân sağlayan ara yüzlere sahiptir.

Çalışmada mezunların çalıştığı sektör bilgisine erişilebilmek için sosyal ağlardan LinkedIn kullanılmış, çeşitli arama moturu taramaları yapılmış, mezunların not bilgilerine erişebilmek için SABİS ve Microsoft SQL Server Management Studio 2017 kullanılmış, birliktelik kurallarının çıkarılması için ise RapidMiner Studio Free (7.6.003) programında çalışılmıştır.

SABİS'e erişebilmek ve sosyal medya aracılığı ile mezun öğrencilere ulaşım, bu öğrencilerin sektörel tercihleri ile çalışma yapabilmek adına Sakarya Üniversitesi Etik Kuruluna 23.11.2017 tarihinde başvuru yapılmıştır. Çalışma Etik Kurulu Başkanlığının 06.12.2017 tarihinde yapılan toplantı ve 61923333/044/ sayılı alınan kararı ile etik açıdan uygun bulunmuştur. İlgili karar çalışmanın EK kısmında görülebilmektedir.

3.2. Materyal

3.2.1. Mezun öğrencilerin sosyal ağlardan tespiti

Çalışmada kullanılacak olan mezun öğrenci bilgilerine LinkedIn sosyal iş ağından arama yapılarak erişilmiştir. Sosyal ağlarda lisans eğitimi olarak Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi olarak bildirimde bulunan kullanıcılar tespit edilmiş ve listelenmiştir. Bazı mezunların bilgilerine ise kişisel çabalar sonucu ulaşılmıştır.

3.2.2. Mezun öğrenci bilgileri için veri ön işleme

Veriler sosyal ağlardan elde edildikten sonra incelenerek, Sakarya Üniversitesi Öğrenci Bilgi Sistemindeki karşılıklarına erişilmeye çalışılmıştır. Kişilerin sosyal

ağlarda kullandıkları isim, ikinci isim, soy isim ve önceki soy isimleri ile mail adresleri sistemdeki veriler ile karşılaştırılarak veri önışleme yapılmıştır. Veri önışleme aşamasında tek bir verinin bile kullanılabilirliğini yitirmemesi amacıyla aşğıdaki maddeler halinde kontrol yapılarak isim soy isim eşleřtirmesi yapılmıştır.

- Kız öğrencilerden evlilik sonrası soyadı deęişenler için sosyal ağlarda bulunan mail adreslerine bakılmış,
- Sosyal ağlarda Türkçe karakter içeren isim soy isim bilgilerini İngilizce karakter ile yazmış olanlar için karakter deęişimi gerçekleştirilerek kontrol edilmiş,
- İki isme sahip olup bir ismini beyan eden veya isim kısaltması kullananlar için yine mail adresine bakılmış,
- Yabancı uyruklu öğrenciler için isimlerinin Türkçe/kendi dilinde yazılıřları ve mail adresleri ile kontrol sağlanmış.

Benzerlik bulunan kayıtlar doęru kabul edilerek sistemde kayıtlı olan 834 adet mezun öğrencinin benzersiz öğrenci numarasına erişilmiştir. Doęru eşleşme olması ile SABİS'deki öğrenci numarası öğrenci için kullanılan tek bilgi olup, çalışmanın bu aşamasından sonra öğrencilerin kişisel hiçbir bilgisiyle uygulama yapılmayarak, öğrenci numarası ile tüm işlemler tamamlanmıştır. Öğrenci numaraları ve çalıştıkları kurum, kuruluş, işletme adı bulunan öğrenciler için bundan sonraki aşamada sadece bu veriler kullanılmıştır.

3.2.3. Mezun öğrencilerin not bilgilerine ulaşılması

Yapılan çalışmada belirlenen dersler Sakarya Üniversitesi Bilgisayar ve Biliřim Bilimleri Fakültesinde okutulan temel dersler olarak seçilmiştir. Bu dersler Tablo 3.2.'de görölmektedir.

Tablo 3.2. Mezun öğrencilerin aldıkları dersler ve kısa adları

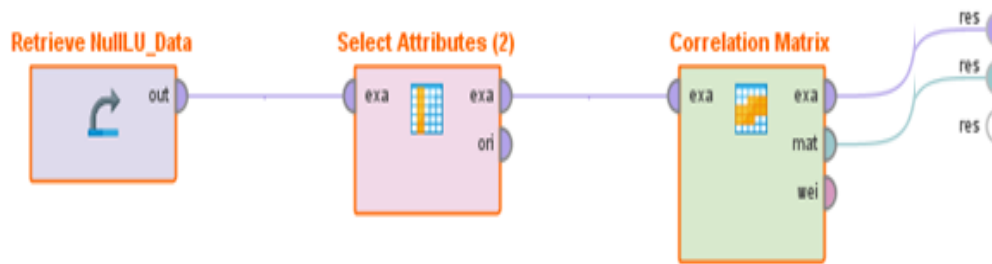
Ders Adı	Kısa Adı
Ayrık Matematik/ Ayrık İşlemsel Yapılar	AYRIK
Veri İletişimi Ve Bilgisayar Ağlarına Giriş Bilgisayar Ağları	AG
Yazılım Mühendisliği	YAZILIM
Elektrik Devre Temelleri Elektrik Ve Elektronik Devreler	ELKTRK
İşletim Sistemleri	ISLTM
Web Programlama	WEB
Veri tabanı Tasarımı Ve Uygulamaları Veri tabanı Yönetim Sistemleri	VERI
Olasılık Teorisi Olasılık Ve İstatistik	OLASLK
Nesneye Dayalı Programlama Nesne Yönelimli Programlama	NESNE
Algoritma Ve Programlamaya Giriş Veri Yapıları Ve Algoritmalar	ALGRM

Fakülte bazında temel ders olarak seçilen dersler farklı zamanlarda farklı isimlerle açılmış olabilmektedir. Bu sebeple genel olarak derslerin içerik benzeşimi dikkate alınmıştır.

3.2.4. Veri indirgeme, dönüştürme ve eksik veri doldurma

Çalışmada Bilgisayar ve Bilişim Mühendisliği bölümünde açıldığı yıl itibariyle ders olarak okutulan temel 10 ders seçilmiştir. Bu derslerden bazıları Fakültenin kuruluşundan itibaren verilirken bazıları ise belirli bir yıldan sonra zorunlu olarak okutulmuştur. Bu sebeple bazı öğrencilerin giriş yılı sebebiyle aldıkları dersler farklı olduğu için notlarına erişilememiştir. Çalışmada 10 ders bilgisinden tüm notlarına erişilebilen mezun öğrenci sayısı 589'dur.

Belirli notlarına erişim sağlanamamış öğrenciler için ise korelasyon ve regresyon ilişkileri RapidMiner programı ile tespit edilerek eksik not bilgileri tamamlanmış olup 834 adet veri işleme alınmıştır. Şekil 3.4.'te Rapidminer programında bu amaçla kurgulanan korelasyon matrisinin modeli görülmektedir. Şekil 3.4.'te görülen ilk süreç ile ham veri sisteme tanıtılmış olup, ikinci süreç ile korelasyon matrisinin uygulanacağı nitelik verisi seçilmiş, üçüncü süreç ile matris oluşturulmuştur.



Şekil 3.4. Eksik bilgi için kullanılan korelasyon matrisinin Rapidminer modeli

Bir başka veri dönüştürme işlemi ise öğrencilerin not durumları değerlendirilirken Tablo 3.3.'te belirtildiği gibi yapılmıştır. Öğrencinin dersten başarısız olma durumunda mezun olması mümkün olmadığından;

- COKIYI
- IYI
- ORTA

olmak üzere 3 farklı kategoride not etiketlenmesi yapılmıştır.

Tablo 3.3. Not durumlarının kategorik olarak ayrılması

Öğrenci Harf Notu	Değerlendirme
AA/BA	Çok İyi
BB/CB	İyi
YETERLİ / MUAF	İyi
CC/DC/DD	Orta

Tablo 3.3.'teki kabullere göre her ders notu da 3 ana başlıkta ifade edilmiştir. Tablo 3.4.'te görüleceği üzere derslerden alınan her notun 3 ayrı kategoride karşılığı bulunmuştur. Çalışmanın bu aşamasından sonra ders notları harf/puan olarak değil, çok iyi, iyi, orta derecesine sahip olarak nitelendirilecektir.

Tablo 3.4. Mezun öğrencilerin ders notlarının kategorize edilmesi

Ders Adı	Çok İyi	İyi	Orta
Ayrık	AYRIK_COKIYI	AYRIK_IYI	AYRIK_ORTA
Ag	AG_COKIYI	AG_IYI	AG_ORTA
Yazılım	YAZILIM_COKIYI	YAZILIM_IYI	YAZILIM_ORTA
Elktrk	ELKTRK_COKIYI	ELKTRK_IYI	ELKTRK_ORTA
Isltm	ISLTM_COKIYI	ISLTM_IYI	ISLTM_ORTA
Web	WEB_COKIYI	WEB_IYI	WEB_ORTA
Veri	VERI_COKIYI	VERI_IYI	VERI_ORTA
Olaslk	OLASLK_COKIYI	OLASLK_IYI	OLASLK_ORTA
Nesne	NESNE_COKIYI	NESNE_IYI	NESNE_ORTA
Algrtm	ALGRTM_COKIYI	ALGRTM_IYI	ALGRTM_ORTA

Çalışmada örneklenen mezun öğrencilerin çalıştıkları firmalar için Tablo 3.5.'teki gibi çeşitli sınıflandırma ve kabuller yapılmıştır. Kabuller için kullanılan anahtar kelimeler belirlenirken firmanın ismi, faaliyet gösterdiği sektör ve hizmet alanı, mezun öğrencinin firmadaki pozisyon bilgisi gibi detaylar göz önünde bulundurulmuştur. Bu kabuller neticesinde mezun öğrencilerin sektör bilgileri etiketlenerek, ilgili çalışma için veri indirilmesi yapılmıştır.

Tablo 3.5. Mezun öğrencilerin çalıştıkları sektör tespiti

Çalışılan Firma veya Pozisyon Bilgisinde Belirlenen Anahtar Kelimeler	Etiketlenen Sektör Bilgisi
Banka, Katılım, Finans, Investment, Yatırım, Payment, Card, Risk	Finans
Netaş, Botaş, Sepaş, Enerji, Elektrik, Elektronik, Dağıtım	Enerji Sektörü
Borusan, OTOKAR, Toyota, Ford, Motor, Hyundai, TOFAŞ, Goodyear, Bridgestone	Otomotiv
Türkiye, Turkish, Bakanlık, Belediye, Genel Müdürlük, TSK, ISKI, IETT, Hastane	Hizmet Sektörü(Kamu)
Üniversite, University, MEB	Eğitim
Danışman, Proje, ,Eğitim, Education, Consult, Intelligence	Danışmanlık/Arge
ERP, Dynamics Ax, SAP, ABAP, Kurumsal,	ERP
Endüstriyel, Industrial	
“.com”, “.net”, Hepsiburada, Tatil, ETS, GittiGidiyor	E-Ticaret
İletişim, Turkcell, Vodafone, Ericsson, Turk Telekom, Wireless, Network, Telekomünikasyon, Güvenlik	İletişim/Güvenlik
Freelance, Costum, Self-Employed	Freelance (Serbest Çalışan)
Medya, Media, TV, Reklam, Ajans	Medya
Yazılım, Bilişim, Bilgi, Soft, Information, Teknoloji, Technology, Solution, Web	Bilişim

Fakülte bazında temel ders olarak seçilen dersler aynı zamanda ana bilim dalı olarak kategorize edilmiş, böylelikle öğrencilerin derslerdeki genel not durumunun bilimsel olarak da ortaya çıkarılması amaçlanmıştır. Ders seçimi sonrası belirlenen dersler için SABİS vasıtasıyla öğrencilerin notlarına erişilmeye çalışılmıştır. Notlarına erişilmeye çalışılan bu derslerin kısa adı ve bilimsel kategorileri Tablo 3.6.'da görülmektedir.

Tablo 3.6. Mezun öğrencilerin aldıkları dersler ve kategorileri

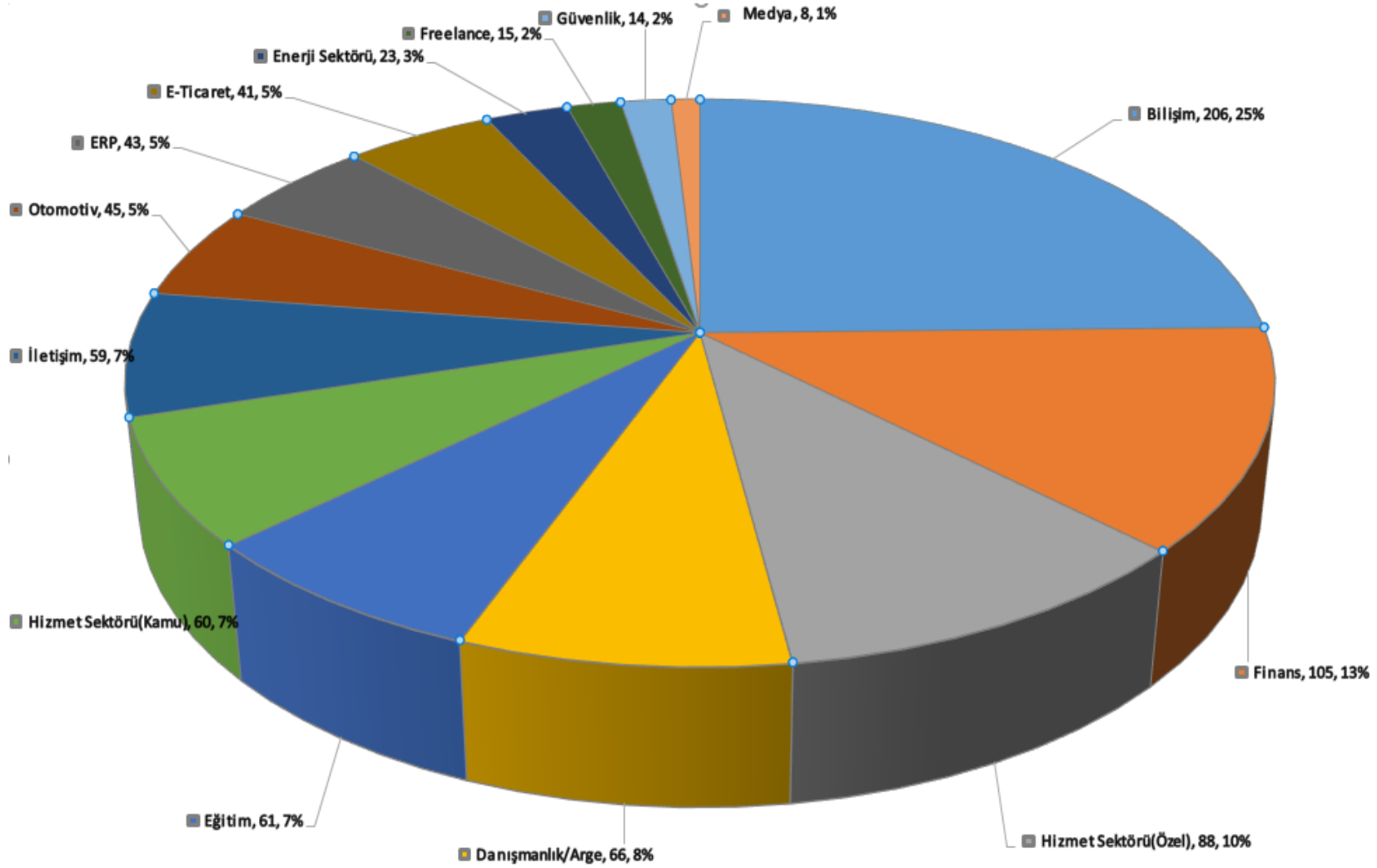
Kısa Ders Adı	Kategori
AYRIK	Bilgisayar Bilimleri
AG	Donanım
YAZILIM	Yazılım
ELKTRK	Donanım
ISLTM	Donanım
WEB	Yazılım
VERI	Yazılım
OLASLK	Bilgisayar Bilimleri
NESNE	Yazılım
ALGRM	Yazılım

3.2.5. Veriler ile ilgili istatistiki bilgiler

Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi'nden 01.09.2017 tarihi itibarıyla toplam 2225 öğrenci mezun olmuştur. Çalışmada LinkedIn sosyal iş ağından lisans mezuniyet bilgisi Sakarya Üniversitesi Bilgisayar ve Bilişim Sistemleri Mühendisliği olan ve çeşitli şekilde toplanan 834 adet mezun öğrenciye ulaşılmıştır. Bu bağlamda örnekleme uzayı toplam mezun öğrenci sayısının %35' inden fazlası olarak görülmüştür.

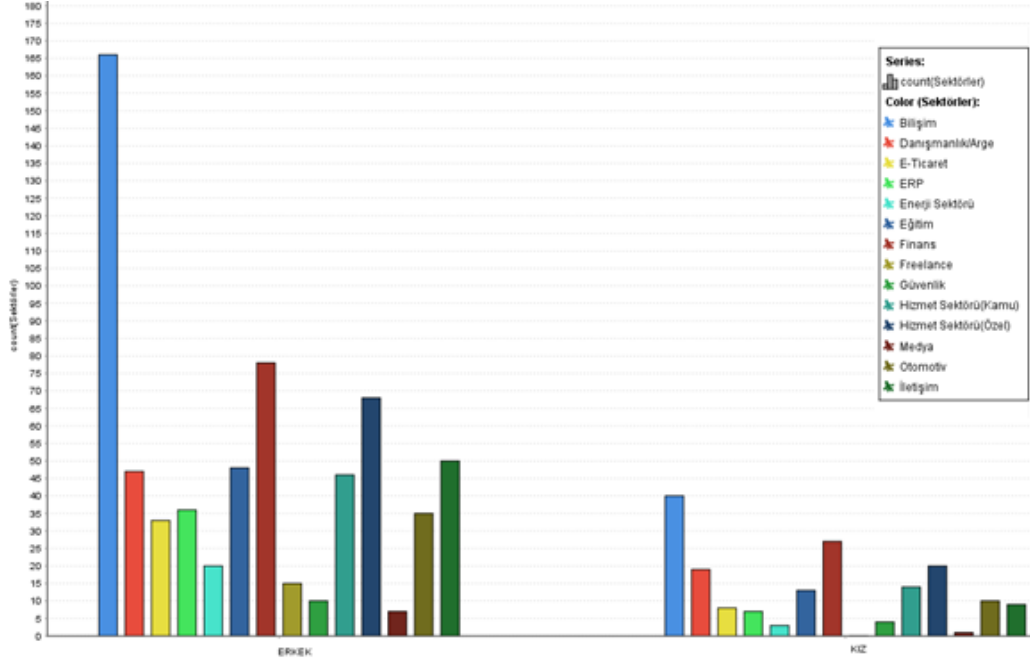
Düzenlenen veriler ile belirlenen sektörlerde çalışan mezun öğrenci sayısı ve tüm öğrenci sayısına yüzde oranı Şekil 3.5.'te gösterilmiştir.

Çalışma neticesinde elde edilen 834 mezun öğrenciden 206'sının Bilişim sektörünü 8'inin ise Medya sektörünü seçtiği belirlenmiştir. Bu bakış açısıyla mezunların en çok Bilişim, Finans ve Hizmet Sektörü(Özel)'nde faaliyet gösterdiği, en az Güvenlik ve Medya sektörünü tercih ettiği gözlenmiştir.



Şekil 3.5. Örnekleme de yer alan mezun öğrencilerin sektörel dağılımı

Yapılan çalışma sonucunda 175 adet kız 659 adet erkek mezun öğrenciye ulaşılmıştır. Mezunların cinsiyetine göre sektörel dağılımı ise Şekil 3.6.'daki grafikte görülebilmektedir.



Şekil 3.6. Mezun öğrenci ve çalıştıkları sektör dağılımı

Şekil 3.7.'de her ders için yapılan kabul neticesinde oluşan kategorideki toplam sayı parantez içinde belirtilmiştir. Örnelemek gerekirse 834 adet mezundan Ayrık Matematik dersini COKIYI derecesiyle geçen öğrenci sayısı 193, IYI derecesiyle geçen öğrenci sayısı 288, ORTA derecesiyle geçen öğrenci sayısı 353'tür.



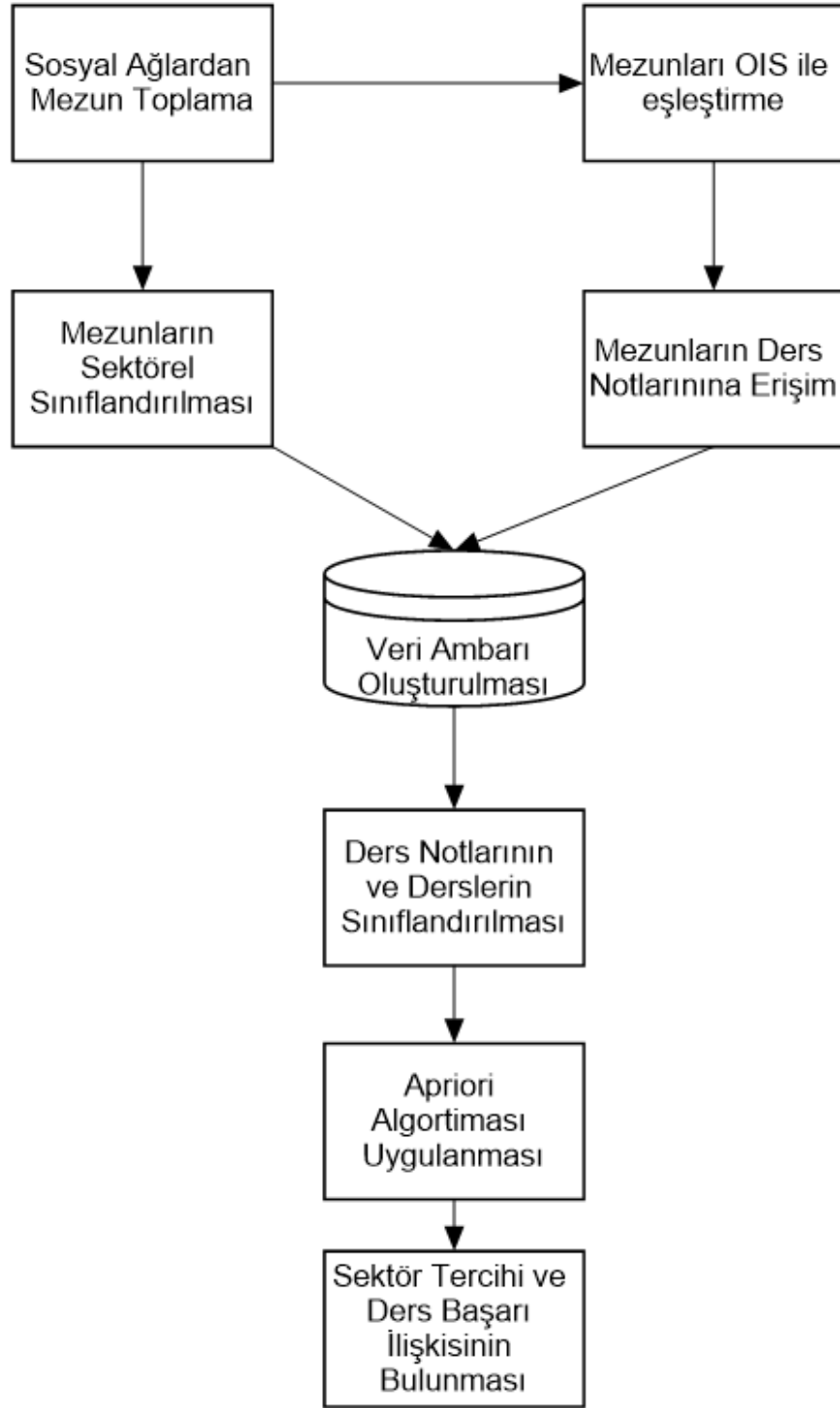
Şekil 3.7. Sınıflandırılmış not durumlarının dağılımı

BÖLÜM 4. ARAŞTIRMA BULGULARI

4.1. Araştırma Akış Diyagramı

Bu tez kapsamında sosyal ağlardan çeşitli şekilde toplanan Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi mezunlarının çalıştıkları sektör ile eğitim öğretim hayatları boyunca aldıkları bazı derslerin not bilgilerine erişilerek, öğrencilerin başarı durumları ile sektör eğilimleri arasındaki ilişkiler birliktelik kuralları ile çıkarılmıştır.

Tez çalışmasında yapılan işlem adımları akış diyagramı şeklinde Şekil 4.1.'de gösterilmiştir.

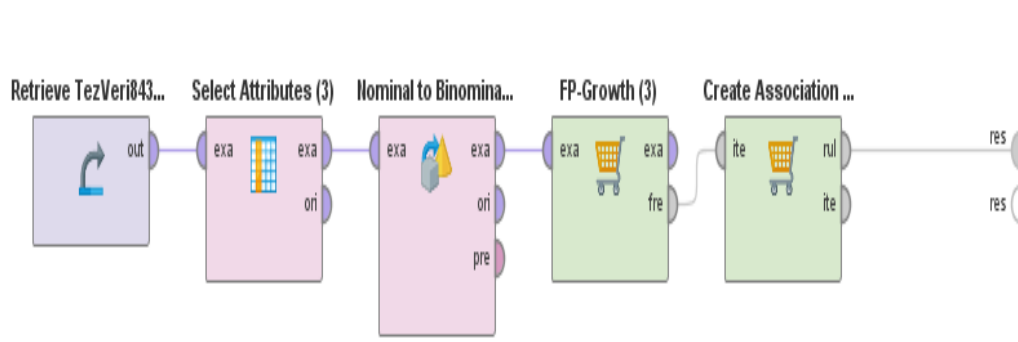


Şekil 4.1. Yapılan çalışmaya ait akış diyagramı

4.2. Modelleme

Bu çalışmada birliktelik kurallarını ortaya çıkarmak için RapidMiner programı kullanılmıştır. RapidMiner programında hazırlanan modelin ekran görüntüsü Şekil 4.2.'de gösterildiği gibidir. Şekil 4.2.'de bulunan modeli kısaca açıklarsak;

- İlk süreçte verinin RapidMiner programına tanıtılması gerçekleşmiş,
- İkinci süreçte veri içerisinde çalışılacak nitelikler seçilmiş,
- Üçüncü süreçte tek terimli veriler birden çok terimli hale getirilip, modelin işlemesi kolaylaştırılmış,
- Dördüncü süreçte apriori algoritmasının özel bir hali olan FP-Growth ile veriler taranmış ve son olarak da birliktelik kuralları oluşturulmuştur.



Şekil 4.2. Rapidminer programında hazırlanan modelin görüntüsü

4.3. Regresyon ve Korelasyon Katsayılarının Bulunması

Belirlenen 10 ders için tüm öğrencilerin ders notlarına Bölüm 3.2.4.'te belirtilen sebepler ile erişim sağlanamamıştır. Bu öğrencilere ait verileri kaybetmemek adına uygulanan Regresyon ve Korelasyon işlemi RapidMiner programı yardımı ile yapılarak, tüm derslerin birbiri ile olan ilişkisi ortaya çıkarılmış ve eksik veriler tamamlanmıştır.

Şekil 4.3.'te Rapidminer programı ile kurulan model ve regresyon katsayılarının bulunduğu tablo görülmektedir. Şekil 4.3.'ü örneklemek gerekirse, Ayrık dersi için en

yakın ilişki Olasılık dersiyle görülmektedir. Bu durumda Ayrık dersi için notu olmayan mezun öğrencilerin notunun Olasılık dersi notu ile güncellenerek, süreç dahil edilme işlemi gerçekleştirilmiştir. Başka bir örnek olarak Web Programlama dersi içinse yine en yakın ilişki tespiti yapılmış, bu ders için notu olmayan öğrencilerin Yazılım dersinde aldıkları not ile güncelleme yapılmıştır.

Attribut...	Ayrık	Ağ	yazılım	elektrik	işletim	web prog	veritaba...	olasılık	nesne	algoritma
Ayrık	1	-0.081	0.012	-0.152	0.119	0.044	-0.054	0.131	0.011	-0.122
Ağ	-0.081	1	-0.170	0.003	0.002	-0.220	0.150	-0.198	-0.133	-0.019
yazılım	0.012	-0.170	1	0.226	0.043	0.270	-0.084	0.138	0.224	0.152
elektrik	-0.152	0.003	0.226	1	-0.070	0.209	0.022	0.023	0.094	0.207
işletim	0.119	0.002	0.043	-0.070	1	0.028	0.111	0.029	-0.051	-0.085
web prog	0.044	-0.220	0.270	0.209	0.028	1	-0.129	0.172	0.195	0.141
veritabanı	-0.054	0.150	-0.084	0.022	0.111	-0.129	1	-0.057	-0.152	-0.004
olasılık	0.131	-0.198	0.138	0.023	0.029	0.172	-0.057	1	0.066	0.022
nesne	0.011	-0.133	0.224	0.094	-0.051	0.195	-0.152	0.066	1	0.101
algoritma	-0.122	-0.019	0.152	0.207	-0.085	0.141	-0.004	0.022	0.101	1

Şekil 4.3. Örnekleme belirlenen dersler arasındaki regresyon katsayıları

4.4. Apriori Algoritması ile Elde Edilen Sonuçlar

Sektör belirleme işlemi yapıldıktan sonra tüm verilerdeki sıklık oranı Tablo 4.1.'deki gibidir. En çok çalışılan sektör Bilişim sektörü olarak etiketlenmiş olup, tüm sektörler içinde yüzde 24,7'lik bir orana sahiptir. İkinci sırada Finans sektörü yüzde 12,6'lık bir paya sahipken, Hizmet Sektörü(Özel) ise yüzde 10,6'lık bir oranla üçüncü sırada bulunmaktadır.

Tablo 4.1. Belirlenen sektörlere göre sıklık analizi

Sektör Adı	Sıklık	Yüzde(%)
Bilişim	206	24,7
Finans	105	12,6
Hizmet Sektörü(Özel)	88	10,6
Danışmanlık/Arge	66	7,9

Tablo 4.1. Belirlenen sektörlere göre sıklık analizi (devamı)

Eğitim	61	7,3
Hizmet Sektörü(Kamu)	60	7,2
İletişim	59	7,1
Otomotiv	45	5,4
ERP	43	5,2
E-Ticaret	41	4,9
Enerji Sektörü	23	2,8
Freelance	15	1,8
Güvenlik	14	1,7
Medya	8	1,0

Veri ambarı olarak oluşturulan mezun öğrenci bilgilerine gerekli dönüşüm işlemi yapıldıktan sonra dersleri Tablo 4.2., 4.3., 4.4.'teki gibi kategorilere ayırarak Apriori Algoritması uygulanmıştır. Böylece sektörel tercihlerin derslerdeki ilişkisi daha net şekilde ortaya çıkacaktır. Tablo 4.5.'te ise birliktelik kurallarının tespiti amacıyla %2 destek değeri ve %20 güven değeri olarak belirlenmiş ve sektörlerin tüm dersler ile olan ilişkileri tespit edilmeye çalışılmıştır.

Tablo 4.2.'de tespit edilen sektör ve bilgisayar bilimleri derslerinin birliktelik analizi sonucu elde edilen bazı kurallar bulunmaktadır. İşlem sonucunda oluşan bazı birliktelik kurallarını yorumlamak gerekirse;

- 1 numaralı kural ile Ayrık dersi orta olan ve Sektör olarak Danışmanlık/Arge de çalışan öğrencilerin %68 olasılıkla Olasılık dersinin orta derecesine sahip olduğu öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %2'dir.
- 3 numaralı kural ile Ayrık ve Olasılık dersi orta derecesine sahip öğrencilerin %30 olasılıkla Bilişim sektöründe çalıştığı öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %6'dir.
- 5 numaralı kurala bakıldığında ise Ayrık dersi çok iyi olan öğrencilerden sektör olarak Finans seçenlerin %50 olasılıkla Olasılık derslerinin çok iyi olduğu ve bu durumun veri setinde birlikte bulunma olasılığının %2 olduğu ortaya çıkmıştır.

Tablo 4.2. Sektör ve Bilgisayar Bilimleri Dersleri İlişkisi

Önerme	Sonuç	Destek	Güven
1 Ayırık = AYRIK_ORTA, Sektörler = Danışmanlık/Arge	Olasılık = OLASLK_ORTA	0,02	0,68
2 Olasılık = OLASLK_IYI, Sektörler = Hizmet Sektörü(Özel)	Ayrık = AYRIK_ORTA	0,03	0,60
3 Ayırık = AYRIK_ORTA, olasılık = OLASLK_ORTA	Sektörler = Bilişim	0,06	0,30
4 Olasılık = OLASLK_COKIYI, Sektörler = Finans	Ayrık = AYRIK_COKIYI	0,02	0,47
5 Ayırık = AYRIK_COKIYI, Sektörler = Finans	Olasılık=OLASLK_COKIYI	0,02	0,50

Tablo 4.3.'te tespit edilen sektör ve donanım derslerinin birliktelik analizi sonucu elde edilen bazı kurallar bulunmaktadır. İşlem sonucunda oluşan bazı birliktelik kurallarını yorumlamak gerekirse;

- 1 numaralı kural ile Elektrik ve İşletim Sistemleri dersi orta derecesine sahip öğrencilerden Sektör olarak İletişim seçenlerin %97 olasılıkla Ağ dersinin orta derecesine sahip olduğu öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %3'tür.
- 4 numaralı kural ile Ağ dersi orta ve Elektrik dersi çok iyi derecesine sahip öğrencilerin %31 olasılıkla Bilişim sektöründe çalıştığı öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %3'tür.
- 7 numaralı kural ile İşletim Sistemleri dersi çok iyi derecesine sahip öğrencilerden Sektör olarak Bilişim seçenlerin %67 olasılıkla Ağ dersinin orta derecesine sahip olduğu öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %3'tür.

Tablo 4.3. Sektör ve Donanım Dersleri İlişkisi

Önerme	Sonuç	Destek	Güven
1 Elektrik = ELKTRK_ORTA, işletim = ISLTM_ORTA, Sektörler = İletişim	Ağ = AG_ORTA	0,03	0,97
2 Elektrik = ELKTRK_ORTA, işletim = ISLTM_ORTA, Sektörler = Hizmet Sektörü(Kamu)	Ağ = AG_ORTA	0,03	0,96
3 Elektrik = ELKTRK_ORTA, işletim = ISLTM_ORTA, Sektörler = Finans	Ağ = AG_ORTA	0,03	0,88
4 Ağ = AG_ORTA, elektrik = ELKTRK_COKIYI	Sektörler = Bilişim	0,03	0,31
5 Sektörler = Bilişim, işletim = ISLTM_COKIYI	Elektrik=ELKTRK_ORTA	0,02	0,48
6 Sektörler = Bilişim, elektrik = ELKTRK_COKIYI	Ağ = AG_ORTA	0,03	0,71
7 Sektörler = Bilişim, işletim = ISLTM_COKIYI	Ağ = AG_ORTA	0,03	0,67

Tablo 4.4.'te tespit edilen sektör ve yazılım derslerinin birliktelik analizi sonucu elde edilen bazı kurallar bulunmaktadır. İşlem sonucunda oluşan bazı birliktelik kurallarını yorumlamak gerekirse;

- 1 numaralı kural ile Web Programlama, Veritabanı ve Nesneye Dayalı Programlama ve Algoritma dersleri orta derecesine sahip öğrencilerin %27 olasılıkla Bilişim sektöründe çalıştığı öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %3'tür.
- 2 numaralı kural ile Web Programlama, Yazılım dersleri orta ve Algoritma dersi iyi derecesine sahip öğrencilerin %32 olasılıkla Bilişim sektöründe çalıştığı öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %2'dir.
- 4 numaralı kural ile Web Programlama dersi çok iyi ve sektör olarak iletişim seçen öğrencilerin %74 olasılıkla Veritabanı derslerinin çok iyi derecesine

sahip olduğu öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %2'dir.

Tablo 4.4. Sektör ve Yazılım Dersleri İlişkisi

Önerme	Sonuç	Destek	Güven
1 Algoritma = ALGRM_ORTA, nesne = NESNE_ORTA, web prog = WEB_ORTA, veritabanı = VERITBN_ORTA	Sektörler = Bilişim	0,03	0,27
2 Yazılım = YAZILIM_ORTA, web prog = WEB_ORTA, algoritma = ALGRM_IYI	Sektörler = Bilişim	0,02	0,32
3 Veritabanı = VERITBN_COKIYI, nesne = NESNE_COKIYI, algoritma = ALGRM_COKIYI, yazılım = YAZILIM_COKIYI	web prog = WEB_COKIYI	0,03	0,88
4 Web prog = WEB_COKIYI, Sektörler = İletişim	Veritabanı=VERITBN_CO KIYI	0,02	0,74
5 Yazılım = YAZILIM_ORTA, algoritma = ALGRM_ORTA, nesne = NESNE_ORTA, veritabanı = VERITBN_ORTA, Sektörler = Bilişim	Web prog = WEB_ORTA	0,02	0,91
6 Algoritma = ALGRM_ORTA, nesne = NESNE_ORTA, web prog = WEB_ORTA, Sektörler = Hizmet Sektörü(Özel)	Yazılım=YAZILIM_ORTA	0,02	0,86
7 Nesne = NESNE_ORTA, Sektörler = Hizmet Sektörü(Kamu)	Web prog = WEB_ORTA	0,03	0,78
8 Web prog = WEB_ORTA, veritabanı = VERITBN_ORTA, Sektörler = Finans	Nesne = NESNE_ORTA	0,02	0,86
9 Web prog = WEB_ORTA, veritabanı = VERITBN_ORTA, Sektörler = Danışmanlık/Arge	Algoritma=ALGRM_OR TA	0,02	0,78

Tablo 4.5.'te tespit edilen sektör ve tüm derslerinin birliktelik analizi sonucu elde edilen bazı kurallar bulunmaktadır. İşlem sonucunda oluşan bazı birliktelik kurallarını yorumlamak gerekirse;

- 1 numaralı kural ile Web Programlama, Veri tabanı ve Nesneye Dayalı Programlama, Yazılım ve Algoritma dersleri orta derecesine sahip öğrencilerin %27 olasılıkla Bilişim sektöründe çalıştığı öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %2'dir.
- 3 numaralı kurala başka bir bakış açısıyla bakıldığında Nesneye Dayalı ve Programlama dersi orta, Web Programlama dersi çok iyi derecesine sahip öğrencilerin %55 olasılıkla Ağ ve Elektrik derslerinin orta olduğu ortaya çıkmış, bu durumların veri setinde birlikte bulunma olasılığı %5 olarak tespit edilmiştir.
- 8 numaralı kural ile Nesneye Dayalı Programlama ve Algoritma dersleri Çok iyi derecesine sahip öğrencilerin %19 olasılıkla Finans sektöründe çalıştığı öngörülmektedir. Ayrıca bu durumların veri setinde birlikte bulunma olasılığı %2'dir.
- 10 numaralı kurala bakıldığında ise Ayrık dersi çok iyi olan öğrencilerden sektör olarak Finans seçenlerin %50 olasılıkla Olasılık derslerinin çok iyi olduğu ve bu durumun veri setinde birlikte bulunma olasılığının %2 olduğu ortaya çıkmıştır.

Tablo 4.5. Sektör ve Tüm Dersler İlişkisi

	Önerme	Sonuç	Destek	Güven
1	Yazılım = YAZILIM_ORTA, algoritma = ALGRM_ORTA, nesne = NESNE_ORTA, web prog = WEB_ORTA, veritabanı = VERITBN_ORTA	Sektörler = Bilişim	0,02	0,27
2	Ağ = AG_ORTA, yazılım = YAZILIM_ORTA, algoritma = ALGRM_ORTA, nesne = NESNE_ORTA, ayrık = AYRIK_ORTA, veritabanı = VERITBN_ORTA	Sektörler = Bilişim	0,02	0,39
3	nesne = NESNE_ORTA, web prog = WEB_COKIYI	Ağ = AG_ORTA, elektrik=ELKTRK_ORTA	0,05	0,55
4	Ağ = AG_ORTA, elektrik = ELKTRK_ORTA, işletim = ISLTM_IYI	Sektörler = Finans	0,03	0,22
5	Ağ = AG_ORTA, elektrik = ELKTRK_ORTA, işletim = ISLTM_IYI	Sektörler = Bilişim	0,03	0,28
6	elektrik = ELKTRK_ORTA, işletim = ISLTM_ORTA, Sektörler = Hizmet Sektörü(Kamu)	Ağ = AG_ORTA	0,03	0,96
7	algoritma = ALGRM_ORTA, nesne = NESNE_ORTA, web prog = WEB_ORTA, sektörler = Hizmet Sektörü(Özel)	Yazılım=YAZILIM_ORT A	0,02	0,85
8	nesne = NESNE_COKIYI, algoritma = ALGRM_COKIYI	Sektörler = Finans	0,02	0,19
9	web prog = WEB_COKIYI, veritabanı = VERITBN_COKIYI, nesne = NESNE_COKIYI	Sektörler = Bilişim	0,02	0,22
10	ayrık = AYRIK_COKIYI, sektörler = Finans	Olasılık=OLASLK_COKI YI	0,02	0,5

BÖLÜM 5. TARTIŞMA VE SONUÇ

5.1. Tartışma

Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi 13.10.2010 tarihinden itibaren eğitim öğretim hayatına devam etmektedir.

Sakarya Üniversitesi Bilgisayar Mühendisliği bölümü ise akademik faaliyetlerine ilk olarak Mühendislik Fakültesi bünyesinde 1994 yılında başlamıştır. 2017-2018 eğitim öğretim yılı itibari ile Bilgisayar ve Bilişim Bilimleri Fakültesi çatısı altında Bilgisayar Mühendisliği lisans (1.öğretim, 2.öğretim ve uzaktan eğitim), yüksek lisans (tezli ve uzaktan eğitim) ve doktora programları yürütülmektedir.

Sakarya Üniversitesi Bilişim Sistemleri Bölümü ise 2011 yılında Bilişim Sistemleri Mühendisliği programı ile akademik faaliyetlerine başlamıştır.

Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesinde 2015, 2016 ve 2017 akademik yıllarında seçmeli olarak açılan ve açılmayan bazı derslerin örnekleme yer alan öğrencilerin tercih ettiği sektörel analizdeki karşılıkları Tablo 5.1.'de görülmektedir. Örnekleme yer alan sektörlerin seçmeli derslerdeki karşılıkları incelendiğinde, sektörle ilişkili seçmeli dersi açılmayan bazı sektörleri tercih eden mezunların olduğu ortaya çıkmaktadır. Örnek vermek gerekirse,

- “ERP Sistemleri” dersi →ERP, Finans, Danışmanlık/Arge
Sektörüyle ilişkili olmasına ve bu sektörde çalışan örnekleme yer alan toplam mezun oranı yaklaşık olarak %25 olmasına rağmen son 3 yıldır dersin açılmadığı görülmektedir.
- “CRM İş Zekâsı Uygulamaları”dersi →Bilişim, E-Ticaret, Finans

Sektörüyle ilişkili olmasına ve bu sektörde çalışan örnekleme yer alan toplam mezun oranı yaklaşık olarak %40 olmasına rağmen son 3 yıldır dersin açılmadığı görülmektedir.

Bu bağlamda mezunların sektörel tercihleri göz önünde bulundurularak, sonraki senelerde açılacak/açılmayacak derslerin seçiminde Tablo 5.1.'deki ilişkiler fayda sağlayabilir. Benzer şekilde seçmeli derslerdeki öğrenci kotası hesaplanırken yine Tablo 5.1.'den faydalanarak karar verilebilir. Ayrıca sektör bilgileri ve ders ilişkileri incelenerek müfredata yönelik işlevsel bir model oluşturulabilir.

Tablo 5.1. Fakülte seçmeli derslerin yıllara göre açık/kapalı olma durumu

Seçmeli Ders Ad	Etkileşimli Olduğu Sektör	2015	2016	2017
3D Modelleme Ve Animasyon	Medya, İletişim	Kapalı	Kapalı	Kapalı
Bilgisayar Mimarileri	İletişim, Güvenlik	Kapalı	Kapalı	Kapalı
Bilişim Hukuğu	Bilişim, Danışmanlık/Arge, Medya	Açık	Kapalı	Kapalı
Biyobilişime Giriş	Bilişim	Kapalı	Açık	Kapalı
Bulanık Sistem Ve Yapay Sinir Ağları	Bilişim, Otomotiv, Danışmanlık/Arge	Kapalı	Kapalı	Kapalı
Bulut Bilişim	Bilişim, İletişim	Açık	Kapalı	Kapalı
Büyük Veri Analizi	Bilişim, İletişim	Açık	Kapalı	Açık
Crm İş Zekâsı Uygulamaları	Bilişim, E-Ticaret, Finans	Kapalı	Kapalı	Kapalı
Derin Öğrenme Ve Evrişimli Sinir Ağları	Bilişim, Otomotiv, Danışmanlık/Arge	Açık	Kapalı	Kapalı
E-İş Ve E-Ticaret Teknolojileri	E-Ticaret, Danışmanlık/Arge	Kapalı	Kapalı	Kapalı
Erp Sistemleri	Erp, Finans, Danışmanlık/Arge	Kapalı	Kapalı	Kapalı
Finansal Bt Yönetimi	Erp, Finans, Danışmanlık/Arge	Kapalı	Açık	Açık
Gerçek Zamanlı Ağ Sistemleri	Bilişim, İletişim	Kapalı	Kapalı	Kapalı
Gömülü Sistemler	Bilişim, İletişim, Otomotiv	Açık	Açık	Kapalı
Karar Destek Sistemleri	Erp, Finans, Danışmanlık/Arge	Kapalı	Kapalı	Kapalı
Robotik	Bilişim, Otomotiv, Danışmanlık/Arge	Kapalı	Kapalı	Kapalı
Semantic Web Teknolojileri	E-Ticaret, Danışmanlık/Arge	Kapalı	Kapalı	Kapalı
Siber Güvenliğe Giriş	Bilişim, İletişim, Güvenlik, Medya	Kapalı	Kapalı	Kapalı
Yönetim Muhasebesi	Erp, Finans, Danışmanlık/Arge	Açık	Kapalı	Kapalı

5.2. Sonuç

Sosyal ağların sıklıkla kullanılması ve kullanım alanlarının artması, araştırmacılar ve bilim insanlarının dikkatini çekmekte ve çözülmesi gereken yeni problemlerin varlığını ortaya koymaktadır.

Gelişen teknoloji ile birlikte sosyal ağlardaki verilerin takip edilerek anlamlandırılması ve işlenmesi ile keşfedilmemiş bilgiye erişilebileceği gün yüzüne çıkmıştır. Sosyal ağlar ile tespit edilen verilerin farklı ortamlardan elde edilen ilişkili verilerle harmanlanıp, veri madenciliği çalışmaları sonucu çok farklı amaçlar için kullanılabilmesi de mümkündür. Veri madenciliği teknikleri ilk bakışta istatistiksel çalışmalar ve uygulama alanı olarak bilgisayar bilimlerini kapsıyor olsa da gelecekte sosyal ağlar üzerinden elde edilen bilgiler ile yapılan uygulamalar sayesinde davranış, toplum bilimleri veya işletme gibi çok farklı disiplinlerle ortak çalışma alanları oluşturacağı düşünülmektedir. Örneklemek gerekirse, sosyal ağ bildirimleri üzerinde yapılan veri madenciliği çalışmaları ile kişilerin ruh halleri, alışveriş eğilimleri veya sosyal alışkanlıklarından çalışma/beslenme rutinlerine kadar olan geniş bir yelpaze de keşfedilmemiş bilgiye erişim sağlanabilmektedir.

Veri madenciliğine verilen önem, gösterilen ilgi tüm dünyaya benzer şekilde ülkemizde de her geçen yıl artmakta ve uygulama, kullanım ve tercih alanları ile genişleyerek yayılmaktadır. Bu çalışmada daha önceden yapılan veri madenciliği uygulamaları incelenmiş, özellikle olarak eğitim alanında gerçekleştirilen uygulamaların çoğunun öğrenci başarısı üzerine analizler gerçekleştirmek için yapıldığı görülmüştür. Eğitim alanında gerçekleştirilen bu çalışmalar ileride yapılacak olan çalışmalara temel niteliğinde olup, eğitim faaliyetlerindeki gelişmeler için işlevsel ve faydalı modeller çıkarılabileceği düşünülmektedir.

Tüm bu gelişmeler göz önüne alındığında gelişen teknoloji ile kurum ve kuruluşların mevcut sistem verileri üzerine veri madenciliği tekniklerini kullanarak, faydalı ve bilinmeyen bilgiye erişmesi kaçınılmazdır. Benzer şekilde ilgili kuruluşların

teknolojik ve diđer geliřmelerini elde edilen bilgiye dayalı gerekleřtirmeleri ise faydalarına olacaktır.

Ayrıca veri madenciliđinin kullanıldıđı alanların eřitlendirilmesi ile lkemiz kurum ve kuruluřları ile bu kuruluřlardan hizmet alan lkemiz insanlarına da byk faydalar sađlayacaktır.

KAYNAKLAR

- Adriaans, P.Z., Zantinge, D., n.d. D.(1996). Data Mining. London: Addison Wesley Longman Ltd.
- Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: *Acm Sigmod Record*. ACM, pp. 207–216.
- Agrawal, R., Srikant, R., 1995. Mining sequential patterns, in: *Data Engineering, 1995. Proceedings of the Eleventh International Conference On*. IEEE, pp. 3–14.
- Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules, in: *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*. pp. 487–499.
- Akçapınar, G., 2014. Çevrimiçi öğrenme ortamındaki etkileşim verilerine göre öğrencilerin akademik performanslarının veri madenciliği yaklaşımı ile modellenmesi.
- Akgöbek Ömer, K.S., 2011. Veri Madenciliği Teknikleri İle Veri Kümelerinden Bilgi Keşfi: Medikal Veri Madenciliği Uygulaması. *Genetics* 20, 18–21.
- Akpınar, H., 2014. Data: veri madenciliği veri analizi. Papatya Yayıncılık Eğitim.
- Aksaraylı, M., 2017. Uzaktan Eğitimi Tercih Etme Nedenleri ve Başarı Arasındaki İlişkinin Kümeleme Analizi İle İncelenmesi *Analysis of Relations Between Reasons of Preference and Success of Distance Education With Clustering Analysis* 8, 37–48.
- Argüden, Y., Erşahin, B., 2008. Veri Madenciliği: Veriden Bilgiye, Masraftan Değere.
- Asif, R., Merceron, A., Ali, S.A., Haider, N.G., 2017. Analyzing undergraduate students' performance using educational data mining. *Computers and Education* 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Bozkır, A.S., Mazman, S.G., Sezer, E.A., 2010. Identification of user patterns in social networks by data mining techniques: Facebook case, in: *International Symposium on Information Management in a Changing World*. Springer, pp. 145–153.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A., 1998. *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc.
- Çeşmeli, M.Ş., Bozkurt, Ö.Ç., Kalkan, A., Pençe, İ., 2015. Yönetim Bilişim Sistemleri Bölümü Öğrencilerinin Yönetim Ve Bilişim Derslerindeki Başarılarının Veri Madenciliği Yöntemleri İle İncelenmesi. *Yönetim Bilişim Sistemleri Dergisi* 1, 36–47.
- Cihan, P., Kalıpsız, O., Cingiz, M.Ö., Doksöz, M., 2013. Yazılım Geliştirme Dersleri Öğrenci Projelerinin Birliktelik Kuralı ile Değerlendirilmesi.

- Dalkilic, F., Aydın, O., 2017. Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Öğrencilerinin Devamsızlık Davranışlarını Etkileyen Faktörler. *Journal of Higher Education and Science* 7, 546. <https://doi.org/10.5961/jhes.2017.231>
- Doğan, A., Özcan, U., Söylemez, İ., 2016. TRAFİK Kazalarında Birliktelik Kural Analizi: Ankara İli Örneği. *Ege Akademik Bakis (Ege Academic Review)* 16. <https://doi.org/10.21121/eab.2016OZEL24423>
- Dunham, M.H., Ming, D., 2003. *Introductory and advanced topics*. Prentice Hall.
- Ekim, U., 2011. Veri madenciliği algoritmalarını kullanarak öğrenci verilerinden birliktelik kurallarının çıkarılması. Selçuk Üniversitesi Fen Bilimleri Enstitüsü.
- Ertam, F., 2017. Sosyal Medya Verileri için Etkili Bir Sınıflandırma Yaklaşımı 29, 67–73.
- Farboudi, S., 2009. Tıp Bilişiminde İstatistiksel Veri Madenciliği. Yayımlanmamış Yüksek Lisans Tezi, Hacette Üniversitesi, Ankara.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., 1996. *Advances in knowledge discovery and data mining*. AAAI press Menlo Park.
- Güngör, E., Yalçın, N., Yurtay, N., 2013. Apriori algoritması ile teknik seçmeli ders seçim analizi. UZEM 2013 Ulusal Uzaktan Eğitim ve Teknolojileri Sempozyumu 01–03.
- Gürsakal, N., 2014. *Büyük Veri*. Baskı, Bursa: Dora.
- Han, J., Kamber, M., Pei, J., 2006. *Data preprocessing*. *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann 47–97.
- Han, J., Pei, J., Kamber, M., 2011. *Data mining: concepts and techniques*. Elsevier.
- Hand, D.J., 1998. Data mining: Statistics and more? *The American Statistician* 52, 112–118.
- Holsheimer, M., Siebes, A., 1994. *Data mining: The search for knowledge in databases*. Centrum voor Wiskunde en Informatica.
- Jacobs, P., 1999. Data mining: What general managers need to know. *Harvard Management Update* 4, 8.
- Kittler, R., Wang, W., 2000. Data mining for yield improvements. *Proceedings from MASM*.
- Kılınç, Ç., 2015. Üniversite Öğrenci Başarısı Üzerine Etki Eden Faktörlerin Veri Madenciliği Yöntemleri İle İncelenmesi.
- Kurt, Ç., Erdem, O.A., 2012. Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi. *Polytechnic* 111–116.
- Maimon, O., Rokach, L., 2007. *Soft computing for knowledge discovery and data mining*. Springer Science & Business Media.
- Onat, A., 2008. Veri madenciliğinin web tabanlı uygulamalarda insan uyumluluklarının tesbiti üzerine bir çalışma. Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

- Özçakir, F.C., Çamurcu, A.Y., 2007. Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi 6, 21–37.
- Özkan, Y., 2008. Veri madenciliği yöntemleri. Papatya Yayıncılık Eğitim.
- Piatetski, G., Frawley, W., 1991. Knowledge discovery in databases. MIT press.
- Piatetsky-Shapiro, G., 1990. Knowledge discovery in real databases: A report on the IJCAI-89 Workshop. AI magazine 11, 68.
- Roiger, R., Geatz, M., n.d. Data mining: A tutorial-based primer. 2003. Boston MA: Addison Wesley.
- Savaş, S., Arıcı, N., 2009. Web tabanlı uzaktan eğitimde iki farklı öğretim modelinin öğrenci başarısı üzerindeki etkilerinin incelenmesi. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09) 13–15.
- Silahtaroglu, G., 2008. Veri madenciliği. Papatya Yayınları, İstanbul.
- Velickov, S., Solomatine, D., 2000. Predictive data mining: practical examples, in: 2nd Joint Workshop on Applied AI in Civil Engineering.
- Yavuz, Ü., Ekim, U., KÖKLÜ, M., 2011. Üniversite Öğrencilerin Ortak Zorunlu Derslerdeki Başarılarının K-Means Algoritması ile İncelenmesi. Engineering Sciences 6, 342–347.

EKLER

EK 1: Etik Kurul Kararı

Evrak Tarih ve Sayısı: 08/12/2017-E.53017



T.C.
SAKARYA ÜNİVERSİTESİ REKTÖRLÜĞÜ
Etik Kurulu

Sayı :61923333/044/
Konu :77/03 Esin Ayşe ZAIMOĞLU

Sayın Ayşe Esin ZAIMOĞLU

İlgi : Esin Ayşe ZAIMOĞLU 23/11/2017 tarihli ve 0 sayılı yazı

Üniversitemiz Etik Kurulu Başkanlığının 06.12.2017 tarihli ve 77 sayılı toplantısında alınan "3" nolu karar örneği ekte sunulmuştur.
Bilgilerinizi rica ederim.

Prof.Dr. Haluk SELVİ
Etik Kurulu Başkanı

3- Esin Ayşe ZAIMOĞLU'nun " Veri madenciliği teknikleri kullanılarak sosyal ağlar aracılığı ile Bilgisayar ve Bilişim Mühendisliği mezun öğrenci profillerinin belirlenmesi" başlıklı çalışması görüşmeye açıldı.

Yapılan görüşmeler sonunda; Esin Ayşe ZAIMOĞLU'nun " Veri madenciliği teknikleri kullanılarak sosyal ağlar aracılığı ile Bilgisayar ve Bilişim Mühendisliği mezun öğrenci profillerinin belirlenmesi" başlıklı çalışmasında katılımcı sayılarının belirlenmesi ve nasıl ulaşılabileceğinin belirlenmesi şartıyla Etik açıdan uygun olduğuna oy birliği ile karar verildi.

ÖZGEÇMİŞ

Esin Ayşe Zaimođlu, 1985 yılında Gaziantep'te doğdu. İlk, orta ve lise eğitimini burada tamamladı. 2003 yılında Gaziantep FNT Anadolu Lisesi'nden mezun oldu. 2003 yılında başladığı Fırat Üniversitesi Bilgisayar Mühendisliği Bölümü'nü 2007 yılında bitirdi. 2010 yılına kadar Microsoft Dynamics-AX Yazılım Uzmanı olarak çeşitli kurumsal firmalarda yazılım danışmanlığı yaptı. 2010 yılında Sakarya Valiliđi'ne kamu personeli olarak atandı. Çeşitli kurumlarda çalıştıktan sonra 2015 yılında Sakarya Üniversitesine naklen tayin oldu. 2016 yılında Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü'nde yüksek lisans eğitimine başladı. Halen Sakarya Üniversitesinde görev yapmaktadır.