T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİR ÜRETİM İŞLETMESİNDE VERİ MADENCİLİĞİ UYGULAMASI

YÜKSEK LİSANS TEZİ

Endüstri Müh. Muhammet ÇETİN

Enstitü Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ

Enstitü Bilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ

Tez Danışmanı : Yrd. Doç. Dr. Bayram TOPAL

Haziran 2009

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİR ÜRETİM İŞLETMESİNDE VERİ MADENCİLİĞİ UYGULAMASI

YÜKSEK LİSANS TEZİ

Endüstri Müh. Muhammet ÇETİN

Enstitü Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ

Enstitü Bilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ

Bu tez 17 / 06 / 2009 tarihinde aşağıdaki jüri tarafından Oybirliği ile kabul edilmiştir.

Yrd. Doç. Dr. Bayram TOPAL	Doç. Dr. Cemalettin KUBAT	Yrd. Doç. Dr. İsmail GÜMÜŞ
Jüri Başkanı	Üye	Üye

TEŞEKKÜR

Çalışamalarım süresince, bilgi ve deneyimini esirgemeyen danışman Hocam Sayın Yrd. Doç.Dr. Bayram TOPAL'a çok teşekkür ederim.

Maddi ve manevi desteklerini her an hissettiğim sevgili eşim, annem, babam ve kardeşime çalışmalarım sırasında gösterdikleri sabır için şükranlarımı sunuyorum.

İÇİNDEKİLER

SİMGELER VE KISALTMALAR LİSTESİ

AID	: Automatic Interaction Detector
C&RT	: Classification and Regression Trees
CHAID	: Chi-Squared Automatic Interaction Detector
GA	: Genetik Algoritma
MARS	: Multivariate Adaptive Regression Splines
OLTP	: Online Transaction Processing
QUEST	: Quick, Unbiased, Efficient Statistical Tree
SLIQ	: Supervised Learning in Quest
SPRINT	: Scalable Parallelizable Induction of Decision Trees

ŞEKİLLER LİSTESİ

TABLO LİSTESİ

ÖZET

Anahtar kelimeler: Veri Madenciliği, Üretim Sektörü, Uygunsuz Ürün

Veri madenciliği, büyük veri yığınlarından anlamlı bilgiyi ortaya çıkarma sürecidir. Veri depolama ve bilgisayar sistemlerindeki hızlı gelişim ve düşük maliyetler nedeni ile veri madenciliği iş dünyasında hızla gelişen bir olgu olarak karşımıza çıkmaktadır. Günümüzde veri madenciliği pazarlama, finans, bankacılık, sigortacılık, parakendecilik, telekomünikasyon, imalat gibi pek çok alanda kullanılmaktadır.

Bu çalışmada, veri madenciliği ayrıntılı bir şekilde ele alınmıştır. Uygulama kısmında, bir üretim işletmesinde, üretilen ürünlerin uygunsuz olarak ayrılmasının nedenleri belirlenerek, bu nedenlerin analizi ile uygunsuz ürünlerin sayısını azaltıcı stratejiler geliştirilmesi hedeflenmektedir. Çalışmada analiz için SPSS Clementine 11.1 yazılımı kullanılmıştır. Neden analizi için karar ağaçları ve yapay sinir ağları ile bir model geliştirilmiştir.

Veri madenciliği üretim sektöründe pek fazla kullanılmamaktadır. Bu çalışma ile veri madenciliğinin üretim sektöründe başarıyla kullanılabilir olduğunu göstermek amaçlanmıştır.

AN APPLICATION OF DATA MINING IN A MANUFACTURING INDUSTRY

SUMMARY

Key Words: Data Mining, Production Sector, Incorrect Product

Data minig is the process of finding hidden and unknown patterns in huge amounts of data. Data mining seems in business world as fastly developing fact owing to fast development and low cost on data storage and computer systems. Data mining is used in various areas such as marketing, e-commerce, banking, insurance, telecommunications etc.

In this work, data mining have been examined intensively. In the implemantation stage, it is determined causes of selection of incorrect products from products which is produced in a manufacturing company. After determination, with resaults of this analysis it is aimed at developing strategies which is used to reduce count of incorrect products. In work, SPSS Clementine 11.1 software was used. A model was developed with desicion trees and artificial neural netwoks for analysis.

Data mining isn`t used widely in manufacturing areas. With this work, it is aimed at showing that data mining can be used in manufacturing area succesfuly.

BÖLÜM 1. GİRİŞ

Günümüzde şirketler bilgisayar ve veri depolama sistemlerine düşük maliyetlerde sahip olabilmektedirler. Bilgisayar sistemlerinin kullanımının hızla yaygınlaşmasına paralel olarak sayısal veri üretiminin artmış ve veri depolama teknolojilerinin gittikçe güçlenmesi nedeni ile de veri tabanlarında daha fazla veri depolanmaya başlanmıştır. İşte veri tabanlarında ki bu teknolojik gelişme ve hacimlerindeki bu olağanüstü artış, veri yığınının yönetilmesi, bu verilerin anlamlı hale getirilmesi ve işe yarar bilgilerin çıkarılması konusunda ciddi boyutta sorun oluşturmaya başlamıştır.

Bilgisayar sistemleri ile üretilen bu veriler tek başlarına değersizdirler (Özellikle veri tabanlarının bilgiyi sadece saklamak için dizayn edildiği düşünüldüğünde). Çünkü çıplak gözle bakıldığında verilerin bir anlam ifade etmediğini söyleyebiliriz. Bu veriler belli bir amaç doğrultusunda işlendiği zaman anlamlı hale gelmektedir. İşte ham veriyi bilgiye veya anlamlı hale dönüştürme işini veri madenciliği ile yapabiliriz.

Örneğin eskiden süpermarketteki kasa basit bir toplama makinesinden ibaretti. Müşterinin o anda satın almış olduğu malların toplamını hesaplamak için kullanılırdı. Günümüzde ise kasa yerine kullanılan satış noktası terminalleri sayesinde bu hareketin bütün detayları saklanabiliyor. Saklanan bu binlerce malın ve binlerce müşterinin hareket bilgileri sayesinde her malın zaman içindeki hareketleri ve eğer müşteriler bir müşteri numarası ile kodlanmışsa bir müşterinin zaman içindeki verilerine ulaşmak ve analiz etmek olasıdır.

Veri tek başına değersizdir. Veriler genellikle tanımlanmamış kullanım ve başvuruları içeren ham gerçekleri göz önünde tutarlar. Bilgi seçeneklere etkiyen işlenmiş veri olmak üzere göz önünde tutulur. Veri bazen formatlanır, filtrelenir ve özetlenir. Veriyi bilgiye çevirmeye veri analizi denir. Araştırmacılar veriyi

hipotezleri test etmek için toplarlar, böylece veri, işlenmemiş ve analiz edilmemiş sayılara bağlıdır.

Veri analizi yaparak her mal için bir sonraki ayın satış tahminleri çıkarılabilir; müşteriler satın aldıkları mallara bağlı olarak gruplanabilir; yeni bir ürün için potansiyel müşteriler belirlenebilir; müşterilerin zaman içindeki hareketleri incelenerek onların davranışları ile ilgili tahminler yapılabilir. Binlerce malın ve müşterinin olabileceği düşünülürse bu analizin gözle ve elle yapılamayacağı, otomatik olarak yapılmasının gerektiği ortaya çıkar. Veri madenciliği burada devreye girer.

Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır.

BÖLÜM 2. VERİ MADENCİLİĞİ

2.1. Veri Madenciliğinin Tanımı

Basit bir tanım yapmak gerekir ise veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, bilgiyi madenleme işidir. Ya da bir anlamda büyük veri yığınları içerisinden gelecek ile ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranmasıdır. Veri madenciliği, eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır. Başka bir deyişle, veri madenciliği, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir.

Veri madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreçtir [38]. Büyük veritabanlarından gizli kalmış örüntüleri çıkarma sürecine veri madenciliği adı verilmektedir. Geleneksel yöntemler kullanılarak çözülmesi çok zaman alan problemlere veri madenciliği süreci kullanılarak daha hızlı bir şekilde çözüm bulunabilir [18].

Veri madenciliği; önceden bilinmeyen, geçerli ve uygulanabilir bilginin veri yığınlarından dinamik bir süreç ile elde edilmesi olarak tanımlanabilir. Bu süreçte kümeleme, veri özetleme, sınıflama kurallarının öğrenilmesi, bağımlılık ağlarının bulunması, değişkenlik analizi ve anomali tespiti gibi farklı birçok teknik kullanılmaktadır.

Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan - bilgisayar arayüzü birleştirilir.

Veri madenciliği kendi başına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır. Veri madenciliği; analiste, iş yapma aşamasında oluşan veriler arasındaki şablonları ve ilişkileri bulması konusunda yardım etmektedir.

Veri madenciliği, verilerden, belirli ancak bilinmeyen bir sınıfta yer alan nesne veya olayları ifade eden örüntülerin çıkarılması amacıyla belirli algoritmaların uygulanmasıdır [14]. Bir başka yerde veri madenciliği, örgütün sahip olduğu veri, enformasyon kaynaklarında, yönetici veya analistin sormayı düşünmediği sorulara, örgüt hakkındaki cevapların aranması olarak tanımlanmıştır [31]. Amerika Birleşik Devletleri Kongresinde, yönetimin uyguladığı veri madenciliği faaliyetlerini kongreye raporlamasına yönelik olarak verilen Veri Madenciliği Raporlama Kanunu 2003 önerisinde veri madenciliği; bir veya daha fazla elektronik veritabanının, sorgulanması, araştırılması veya diğer bir şekilde analizi olarak belirtilmiştir [21].

Veri madenciliği, ham verinin tek başına sunamadığı bilgiyi çıkaran veri analizi sürecidir [20]. Veri madenciliği insanın asla bulmayı hayal bile edemeyeceği trendlerin keşfedilmesini sağlamaktadır [8]. Veri madenciliği büyük hacimli verilerdeki örüntüleri araştıran matematiksel algoritmaları kullanmaktadır. Veri madenciliği hipotezleri keşfeder, sonuçları birleştirmek için insan yeteneğini kullanır. Veri madenciliğinin sadece bir bilim olmadığı, aynı zamanda bir sanat olduğu da söylenebilir [10]. Başka bir tanımda, veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır [16]. Diğer bir tanımda ise veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır [25].

Sonuç olarak veri madenciliği, büyük veri yığınlarından önceden bilinmeyen ilişki ve kuralların bulunması ile anlamlı bilgilerin çıkarılması yoludur. Veri madenciliği ile büyük veri yığınlarından oluşan veritabanı sistemleri içerisinde gizli kalmış bilgilerin çekilmesi sağlanır. Bu işlem, istatistik, matematik disiplinleri, modelleme teknikleri, veritabanı teknolojisi ve çeşitli bilgisayar programları

kullanılarak yapılır. İşletmelerin düşük maliyetler ile yüksek depolama kapasitesine sahip teknolojilere sahip olmaları ile daha da önem kazanan bir süreçtir.

2.2. Veri Madenciliğinin Gelişim Süreci

Bilgisayarların etkin kullanımı verilerin depolanması ile başlamaktadır. İlk haliyle karmaşık hesaplamaları yapmaya yönelik geliştirilen bilgisayarlar, kullanıcı ihtiyaçları doğrultusunda veri depolama işlemleri için de kullanılmaya başlandı. Bu sayede veri tabanları ortaya çıktı. Veri tabanlarının genişleme trendi içinde olması donanımsal olarak bu verilerin tutulacakları ortamların da genişlemesini gerektirdi. Veri ambarı kavramının ortaya çıkışı bu dönemlere rastlamaktadır. Kaybedilmek istemeyen veriler, bir ambar misali fiziksel sürücülerde tekrar kullanılmak üzere saklanmaktaydı.

Gittikçe büyüyen veri tabanlarının organizasyonu, düzenlenmesi ve yönetimi de buna paralel olarak güç bir hal almaya başladı. Bu safhada veri modelleme kavramı ortaya çıktı. İlk olarak basit veri modelleri olan Hiyerarşik ve Şebeke veri modelleri geliştirildi. Hiyerarşik veri modelleri, ağaç yapısına sahip, temelinde bir kök olan ve bu kök vasıtasıyla üstünde her daim bir, altında ise n sayıda düğüm bulunan veri modelleriydi. Şebeke veri modelleri ise kayıt tipi ve bağlantıların olduğu, kayıt tiplerinin varlık, bağlantılarınsa ilişki tiplerini belirlediği bir veri modeliydi. Şebeke veri modelinde herhangi bir eleman bir diğeri ile ilişki içerisine girebiliyordu. Ancak çoklu ilişki kurmak söz konusu değildi. Hiyerarşik veri modellerinde ise bu daha da kısıtlıydı. Dolayısıyla kullanıcıların ihtiyaçlarını tam olarak karşılayamadılar. Bu ihtiyaçlar doğrultusunda Geliştirilmiş Veri Modelleri geliştirildi. Bunlar Varlık–İlişki, İlişkisel ve Nesne–Yönelimli veri modelleri olarak bilinmektedirler. Günümüzde en sık kullanılanı İlişkisel veri modelidir. Nesne–Yönelimli veri modelleri ise hala gelişim süreci içerisindedir. İhtiyaçlar doğrultusunda şekillenen veri tabanları ve veri modelleme çeşitleri hızla yaygınlaşırken, donanımlar da bu sürece ayak uydurdular. Günümüzde milyarlarca bit veriyi ufacık belleklerde tutmak mümkün hale gelmiştir. İhtiyaçlar her ne kadar teknolojiyi ciddi anlamda şekillendirse de yanında sorunları daim olarak getirmektedir. Verileri saklanması,

düzenlenmesi, organize edilmesi her ne kadar bir sorun gibi görünmese de bu kadar çok veri ile istenilen sonuca ulaşmak başlı başına bir sorun halini almıştır.

Veri madenciliği, kavramsal olarak 1960`lı yıllarda, bilgisayarların veri analiz problemlerini çözmek için kullanılmaya başlamasıyla ortaya çıktı. O dönemlerde, bilgisayar yardımıyla, yeterince uzun bir tarama yapıldığında, istenilen verilere ulaşmanın mümkün olacağı gerçeği kabullenildi. Bu işleme veri madenciliği yerine önceleri veri taraması (data dredging), veri yakalanması (data fishing) gibi isimler verildi. 1990`lı yıllara gelindiğinde veri madenciliği ismi, bilgisayar mühendisleri tarafından ortaya atıldı. Bu camianın amacı, geleneksel istatistiksel yöntemler yerine, veri analizinin algoritmik bilgisayar modülleri tarafından değerlendirmesini vurgulamaktı. Bu noktadan sonra bilim adamları veri madenciliğine çeşitli yaklaşımlar getirmeye başladılar. Bu yaklaşımların kökeninde istatistik, makine öğrenimi (machine learning), veritabanları, otomasyon, pazarlama, araştırma gibi disiplinler ve kavramlar yatmaktaydı. İstatistik, süre gelen zaman içerisinde verilerin değerlendirilmesi ve analizleri konusunda hizmet veren bir yöntemler topluluğuydu.

Bilgisayarların veri analizi için kullanılmaya başlamasıyla istatistiksel çalışmalar hız kazandı. Hatta bilgisayarın varlığı daha önce yapılması mümkün olmayan istatistiksel araştırmaları mümkün kıldı. 1990`lardan sonra istatistik, veri madenciliği ile ortak bir platforma taşındı. Verinin, yığınlar içerisinden çekip çıkarılması ve analizinin yapılarak kullanıma hazırlanması sürecinde veri madenciliği ve istatistik sıkı bir çalışma birlikteliği içine girmiş bulundular. Bunun yanı sıra veri madenciliği, veri tabanları ve makine öğrenimi disipliniyle birlikte yol aldı. Günümüzdeki Yapay Zeka çalışmalarının temelini oluşturan makine öğrenimi kavramı, bilgisayarların bazı işlemlerden çıkarsamalar yaparak yeni işlemler üretmesidir. Önceleri makineler, insan öğrenimine benzer bir yapıda inşa edilmeye çalışıldı. Ancak 1980`lerden sonra bu konuda yaklaşım değişti ve makineler daha spesifik konularda kestirim algoritmaları üretmeye yönelik inşa edildi. Bu durum ister istemez uygulamalı istatistik ile makine öğrenim kavramlarını, veri madenciliği altında bir araya getirdi.

2.3. Veri Madencisi Kimdir?

Cevap aranılan soru veya çözülecek problem için kurulan bir modelin başarılı olabilmesi sadece metodolojilerin derinlemesine biliniyor olmasına bağlı değildir. Veriyi ve pazarı tanımak, kurumun iş hedeflerini biliyor olmak, modelin altyapısını oluşturan metodolojilerden çok daha önemlidir [6].

Her alanda olduğu gibi veri madenciliğinde de teknoloji ile deneyimin birleşimi en doğru sonuca ulaştırmaktadır. Deneyimin elde edilen sonuçlar üzerindeki etkisi oldukça yüksektir. Veri madenciliği bilincinin artması ile birlikte, bu tür çalışmalara ağırlık vermek isteyen şirketlerin büyük bölümü iki önemli hata yapmaktadırlar [6].

- Çalışmaları gerçekleştirmek için teknik konulara hakim istatistik uzmanları veya teknik analistleri işe alarak, modelleri kurgulamalarını istemek: Bu kategorideki uzmanlar teknik konularda çok yetkin olmalarına rağmen, gerekli iş kavrayışına yeterince sahip olmamaları nedeniyle arzu edilen sonuçlara çoğunlukla ulaşılamamaktadır.

- Sofistike veri madenciliği yazılımları satın almak: Konu ile ilgili çok detaylı, tüm metodolojileri içeren yazılımlar mevcuttur ancak yazılımlardan faydalı sonuçlar alabilmek için doğru modeli kurgulamak ve doğru girdileri sunmak gereklidir. Bu düşünce sürecinden geçmeden yazılımdan faydalı sonuçlar elde etmek mümkün değildir. Her iki yaklaşımda da; hedefi oluşturma, veriyi elde etme, veriyi hazırlama, modeli uygulama, sonuçları değerlendirme gibi önemli alanlarda bilgi eksikliği söz konusu olabilir. Bu alanların herhangi birinde yapılacak hata çok maliyetli olabileceği gibi tamamen yanlış sonuçlara da götürebilir. İstatistiksel araçları çok iyi bilen en iyi teknik analistlere sahip olmak kadar bunu gerçek dünyanın problemlerine nasıl uyarlayacaklarını bilmek de önemlidir. Bu aşamada veri madenciliğinin 3 farklı boyutuna bakmakta, ilişkileri ve gereksinimleri anlamak açısından fayda vardır [6].

- Yanıtlanacak soru nedir? / Neye cevap aranmaktadır?
- Cevap aranan konuyu hangi veri madenciliği fonksiyonu ile çözümlemek gerekir?

- İlgili veri madenciliği fonksiyonu için hangi algoritma ile model oluşturmak uygun olur?

Cevap aranılan sorunun tanımlanması ve uygun fonksiyonun seçilmesi aşamasında faaliyeen sorumlu olan profesyonellerin daha etkin rol alması, seçilen fonksiyona uygun algoritmanın belirlenmesi ve işletilmesi aşamasında istatistik uzmanlarının daha etkin rol alması gerekir.

2.4. Veri Madenciliğinin Uygulama Alanları

Azalan bilgi işleme maliyeti, verinin toplanması ve saklanmasındaki kolaylık, veritabanı yönetim sistemi teknolojilerindeki ilerlemeler, kullanılabilecek analitik araçların oldukça fazlalaşmasıyla birlikte veri madenciliği uygulamalarına olan ilgi artmaktadır [30].

Bir çok alanda uygulanma imkanı bulan veri madenciliği sektörlerde aşağıda belirtilen konularda uygulanabilmektedir [19].

Pazarlama alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Müşteri segmentasyonunda,
- Müşterilerin demografik özellikleri arasındaki bağlantıların kurulmasında,
- Çeşitli pazarlama kampanyalarında,
- Mevcut müşterilerin elde tutulması için geliştirilecek pazarlama stratejilerinin oluşturulmasında,
- Pazar sepeti analizinde,
- Çapraz satış analizleri,
- Müşteri değerleme,
- Müşteri ilişkileri yönetiminde,
- Çeşitli müşteri analizlerinde,
- Satış tahminlerinde,

Bankacılık alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Farklı finansal göstergeler arasındaki gizli korelasyonların bulunmasında,
- Kredi kartı dolandırıcılıklarının tespitinde,
- Müşteri segmentasyonunda,
- Kredi taleplerinin değerlendirilmesinde,
- Usulsüzlük tespiti,
- Risk analizleri,
- Risk yönetimi,

Sigortacılık alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Yeni poliçe talep edecek müşterilerin tahmin edilmesinde,
- Sigorta dolandırıcılıklarının tespitinde,
- Riskli müşteri tipinin belirlenmesinde.

Perakendecilik alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Satış noktası veri analizleri,
- Alış-veriş sepeti analizleri,
- Tedarik ve mağaza yerleşim optimizasyonu,

Borsa alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Hisse senedi fiyat tahmini,
- Genel piyasa analizleri,
- Alım-satım stratejilerinin optimizasyonu.

Telekomünikasyon alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Kalite ve iyileştirme analizlerinde,
- Hisse tespitlerinde,
- Hatların yoğunluk tahminlerinde.

Sağlık ve İlaç alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Test sonuçlarının tahmini,
- Ürün geliştirme,
- Tıbbi teşhis,
- Tedavi sürecinin belirlenmesinde.

Endüstri alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Kalite kontrol analizlerinde,
- Lojistik,
- Üretim süreçlerinin optimizasyonunda.

Bilim ve Mühendislik alanında aşağıdaki konuların analizinde kullanılmaktadır:

- Ampirik veriler üzerinde modeller kurarak bilimsel ve teknik problemlerin çözümlenmesi.

Veri madenciliğinin asıl amacı veri yığınlarından anlamlı bilgiler elde etmek ve bunu eyleme dönüştürecek kararlar için kullanmak olduğuna göre birkaç analiz örneği olarak aşağıdaki konular verilebilir [17]:

- Bir işletme kendi müşterisiyken rakibine giden müşterilerle ilgili analizler yaparak rakiplerini tercih eden müşterilerinin özelliklerini elde edebilir ve bundan yola çıkarak gelecek dönemlerde kaybetme olasılığı olan müşterilerin kimler olabileceği yolunda tahminlerde bulunarak onları kaybetmemek, kaybettiklerini geri kazanmak için strateji geliştirebilir.

- Ürün veya hizmette hangi özelliklerin ne derecede müşteri memnuniyetini etkilediği, hangi özelliklerinden dolayı müşterini bunları tercih ettiği ortaya çıkarılabilir.

- Müşterilerin kredi riskleri hesaplanarak hangi müşterilerin kredi riskinin yüksek olduğu, hangi müşterilerin geri ödemesini zamanında yapamayabileceği kestirilebilir. Kredi kartı ödemelerini aksatan, gecikmeli olarak yapan veya hiç yapmayanların

özelliklerinden yola çıkılarak bundan sonra aynı duruma düşebilecek muhtemel kişiler saptanabilir.

- Ürün talebi bazında müşteri profillerini belirleyerek, müşteri segmentasyonuna gitmek ve çapraz satış olanakları yaratmakta kullanılabilir.

- Piyasada oluşabilecek değişikliklere mevcut müşteri portföyünün vereceği tepkinin firma üzerinde yaratabileceği etkinin tespitinde kullanılabilir.

- En karlı mevcut müşteriler saptanarak, potansiyel müşteriler arasından en karlı olabilecekler belirlenebilir. Karlı müşteriler tespit edilerek onlara özel kampanyalar uygulanabilir. En masraflı müşteriler daha masrafsız müşteri haline dönüştürülebilir. Örneğin en çok bankacılık işlemi yapanlar ortaya çıkarılıp bunlar şube bankacılığı yerine daha masrafsız Internet bankacılığına yönlendirilebilir.

- Bir ürün veya hizmetle ilgili bir kampanya programı oluşturmak için hedef kitlenin seçiminden başlayarak bunun hedef kitleye hangi kanallardan sunulacağı kararına kadar olan süreçte veri madenciliği kullanılabilir.

- Operasyonel süreçte oluşabilecek olası kayıpların veya suistimallerin tespitinde kullanılabilir.

- Kurum teknik kaynaklarının en optimal şekilde kullanılmasını sağlamakta kullanılabilir.

- Firmanın finansal yapısının, makro ekonomik değişmeler karşısındaki duyarlılığı ve oluşabilecek risklerin tespitinde kullanabilir.

- Geçmiş ve mevcut yapı analiz edilerek geleceğe yönelik tahminlerde bulunulabilir. Özellikle ciro, karlılık, pazar payı, gibi analizlerde veri madenciliği çok rahat kullanılabilir.

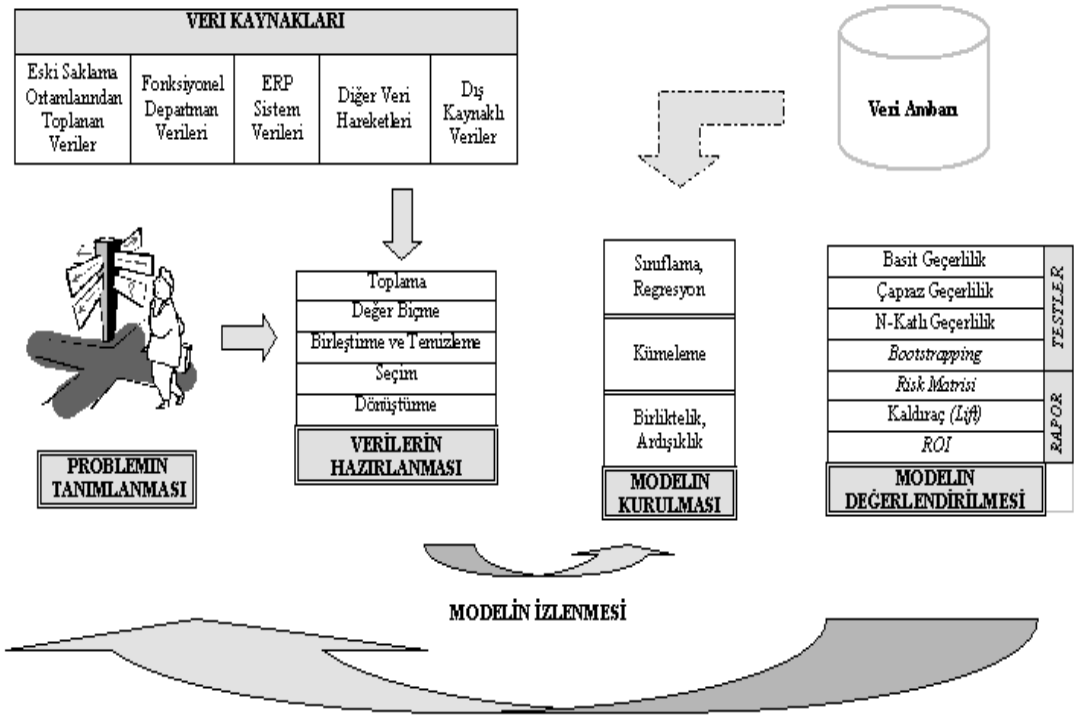
2.5. Veri Tabanlarında Bilgi Keşfi Süreci

Ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlaması mümkün değildir. Bu nedenle aşağıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi / anlaşılması başarının ilk şartı olacaktır [3].

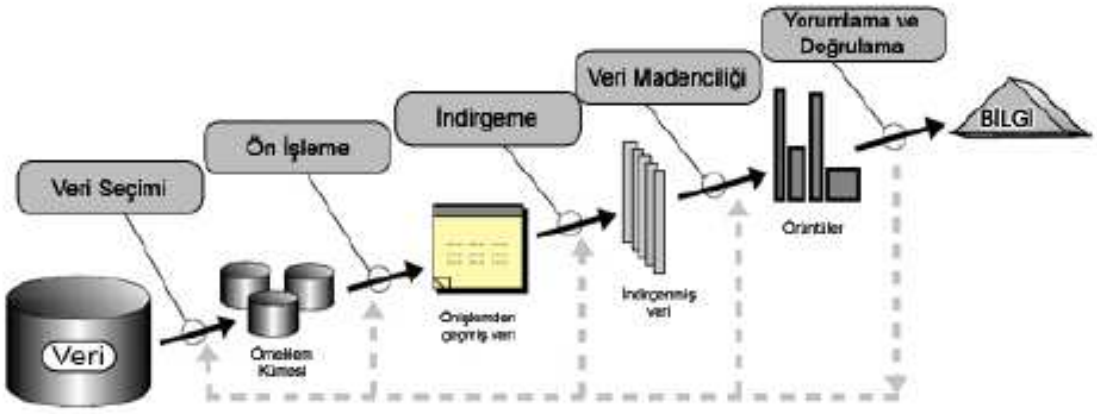
Şekil 2.1`de ayrıntılı olarak görüldüğü gibi,

- Problemin Tanımlanması,
- Verilerin Hazırlanması,
- Modelin Kurulması ve Değerlendirilmesi,
- Modelin Kullanılması,
- Modelin İzlenmesi

veri tabanlarında bilgi keşfi sürecinde izlenmesi gereken temel aşamalardır [3].



Şekil 2.1. Veri tabanlarında bilgi keşfi süreci ve veri madenciliği (Akpınar 2000)



Şekil 2.2. Bilgi keşfi sürecinde veri madenciliğinin yeri

2.5.1. Problemin tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın hangi işletme amacı için yapılacağının açık bir şekilde tanımlanmasıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir.

2.5.2. Verilerin hazırlanması

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının % 50 - % 85'ini harcamasına neden olmaktadır.

Verilerin hazırlanması aşaması kendi içerisinde toplama, değer biçme, birleştirme ve temizleme, seçme ve dönüştürme adımlarından meydana gelmektedir.

2.5.2.1. Toplama

Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımıdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir.

İş dünyasında veriler birçok farklı ortamda depolanmaktadır. Örneğin; Microsoft'da veriler yüzlerce OLTP veritabanında ve 70'in üzerinde veri ambarında saklanmaktadır. Burada ilk adım veri tabanlarından veya veri ambarlarından yapılacak uygulama için uygun verileri çekmektir [35].

Veri toplama işleminde, veriler test ve analiz veri seti olarak iki gruba ayrılmalıdır. Genellikle yapılan uygulamalarda verilerin %80'i analiz %20'si ise test verisi olarak ayrılır [35].

2.5.2.2. Değer biçme

Veri madenciliğinde kullanılacak verilerin farklı kaynaklardan toplanması, doğal olarak veri uyumsuzluklarına neden olacaktır. Bu uyumsuzlukların başlıcaları farklı zamanlara ait olmaları, kodlama farklılıkları (örneğin bir veri tabanında cinsiyet özelliğinin e/k, diğer bir veri tabanında 0/1 olarak kodlanması), farklı ölçü birimleridir. Ayrıca verilerin nasıl, nerede ve hangi koşullar altında toplandığı da önem taşımaktadır.

Bu nedenlerle, iyi sonuç alınacak modeller ancak iyi verilerin üzerine kurulabileceği için, toplanan verilerin ne ölçüde uyumlu oldukları bu adımda incelenerek değerlendirilmelidir.

2.5.2.3. Birleştirme ve temizleme

Bu adımda farklı kaynaklardan toplanan verilerde bulunan ve bir önceki adımda belirlenen sorunlar mümkün olduğu ölçüde giderilerek veriler tek bir veri tabanında toplanır. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

Veri temizleme işleminin amacı, veriler içindeki uygun olmayan veya hatalı girilmiş verileri ayıklamaktır [35].

2.5.2.4. Seçim

Bu adımda kurulacak modele bağlı olarak veri seçimi yapılır. Örneğin tahmin edici bir model için, bu adım bağımlı ve bağımsız değişkenlerin ve modelin eğitiminde kullanılacak veri kümesinin seçilmesi anlamını taşımaktadır.

Sıra numarası, kimlik numarası gibi anlamlı olmayan ve diğer değişkenlerin modeldeki ağırlığının azalmasına da neden olabilecek değişkenlerin modele girmemesi gerekmektedir. Bazı veri madenciliği algoritmaları konu ile ilgisi olmayan bu tip değişkenleri otomatik olarak elese de, pratikte bu işlemin kullanılan yazılıma bırakılmaması daha akılcı olacaktır.

Verilerin göreselleştirilmesine olanak sağlayan grafik araçlar ve bunların sunduğu ilişkiler, bağımsız değişkenlerin seçilmesinde önemli yararlar sağlayabilir.

Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir.

Modelde kullanılan veri tabanının çok büyük olması durumunda tesadüfîliği bozmayacak şekilde örnekleme yapılması uygun olabilir. Günümüzde hesaplama olanakları ne kadar gelişmiş olursa olsun, çok büyük veri tabanları üzerinde çok

sayıda modelin denenmesi zaman kısıtı nedeni ile mümkün olamamaktadır. Bu nedenle tüm veri tabanını kullanarak bir kaç model denemek yerine, tesadüfî olarak örneklenmiş bir veri tabanı parçası üzerinde birçok modelin denenmesi ve bunlar arasından en güvenilir ve güçlü modelin seçilmesi daha uygun olacaktır.

2.5.2.5. Dönüştürme

Veri dönüşümünün amacı, elimizdeki kaynak veriyi farklı formatlara veya değerlere dönüştürmektir [35]. Mesela kredi riskinin tahmini için geliştirilen bir modelde, borç/gelir gibi önceden hesaplanmış bir oran yerine, ayrı ayrı borç ve gelir verilerinin kullanılması tercih edilebilir. Ayrıca modelde kullanılan algoritma, verilerin gösteriminde önemli rol oynayacaktır. Örneğin bir uygulamada bir yapay sinir ağı algoritmasının kullanılması durumunda kategorik değişken değerlerinin evet/hayır olması; bir karar ağacı algoritmasının kullanılması durumunda ise örneğin gelir değişken değerlerinin yüksek/orta/düşük olarak gruplanmış olması modelin etkinliğini artıracaktır.

2.5.3. Modelin kurulması ve değerlendirilmesi

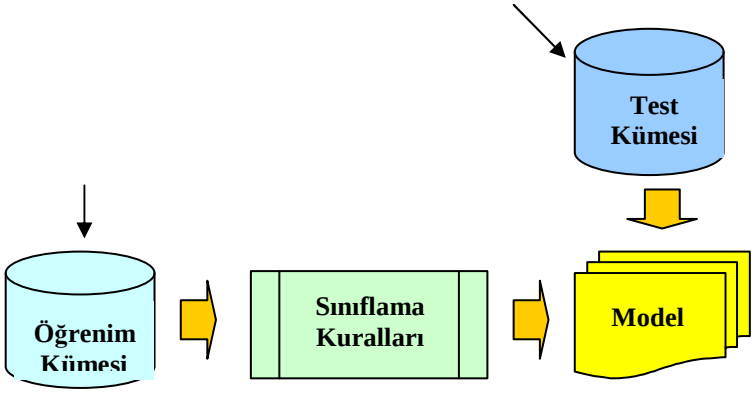
Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir.

Model kuruluş süreci denetimli (Supervised) ve denetimsiz (Unsupervised) öğrenimin kullanıldığı modellere göre farklılık göstermektedir.

Örnekten öğrenme olarak da isimlendirilen denetimli öğrenimde, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir.

Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklere uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir.

Denetimsiz öğrenmede, kümeleme analizinde olduğu gibi ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır.



Şekil 2.3. Denetimli öğrenme

Denetimli öğrenimde seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenimi, diğer kısmı ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin öğrenimi öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi (Accuracy) belirlenir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (Simple Validation) testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. (Doğruluk Oranı = 1 - Hata Oranı)

Sınırlı miktarda veriye sahip olunulması durumunda, kullanılabilecek diğer bir yöntem çapraz geçerlilik (Cross Validation) testidir. Bu yöntemde veri kümesi

tesadüfi olarak iki eşit parçaya ayrılır. İlk aşamada a parçası üzerinde model eğitimi ve b parçası üzerinde test işlemi; ikinci aşamada ise b parçası üzerinde model eğitimi ve a parçası üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır.

Bir kaç bin veya daha az satırdan meydana gelen küçük veri tabanlarında, verilerin n gruba ayrıldığı n katlı çapraz geçerlilik (N-Fold Cross Validation) testi tercih edilebilir. Verilerin örneğin 10 gruba ayrıldığı bu yöntemde, ilk aşamada birinci grup test, diğer gruplar öğrenim için kullanılır. Bu süreç her defasında bir grubun test, diğer grupların öğrenim amaçlı kullanılması ile sürdürülür. Sonuçta elde edilen on hata oranının ortalaması, kurulan modelin tahmini hata oranı olacaktır.

Bootstrapping küçük veri kümeleri için modelin hata düzeyinin tahmininde kullanılan bir başka tekniktir. Çapraz geçerlilikte olduğu gibi model bütün veri kümesi üzerine kurulur. Daha sonra en az 200, bazen binin üzerinde olmak üzere çok fazla sayıda öğrenim kümesi tekrarlı örneklemelerle veri kümesinden oluşturularak hata oranı hesaplanır.

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir. Model kuruluş çalışmalarına başlamazdan önce, imkansız olmasa da hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar bulunmaktadır.

Özellikle sınıflama problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde basit ancak faydalı bir araç olan risk matrisi kullanılmaktadır. Aşağıda bir örneği görülen bu matriste sütunlarda fiili, satırlarda ise tahmini sınıflama değerleri yer almaktadır. Örneğin fiilen B sınıfına ait olması gereken 46 elemanın, kurulan model tarafından 2’sinin A, 38’inin B, 6’sının ise C olarak sınıflandırıldığı matrisde kolayca görülebilmektedir.

Tablo 2.1. Örnek matris

	Fiili		
Tahmini	**A Sınıfı**	**B Sınıfı**	**C Sınıfı**
A Sınıfı	45	2	3
B Sınıfı	10	38	2
C Sınıfı	4	6	40

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, bir çok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Çok ender olarak yorumlanamayacak kadar karmaşıklaşsalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

Kaldıraç (Lift) oranı ve grafiği, bir modelin sağladığı faydanın değerlendirilmesinde kullanılan önemli bir yardımcıdır. Örneğin kredi kartını muhtemelen iade edecek müşterilerin belirlenmesi amacını taşıyan bir uygulamada, kullanılan modelin belirlediği 100 kişinin 35'i gerçekten bir süre sonra kredi kartını iade ediyorsa ve tesadüfi olarak seçilen 100 müşterinin aynı zaman diliminde sadece 5'i kredi kartını iade ediyorsa kaldıraç oranı 7 olarak bulunacaktır.

Kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçü, model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın gerçekleştirilmesi çin katlanılacak maliyete bölünmesi ile edilecek olan yatırımın geri dönüş (Return On Investment) oranıdır.

Kurulan modelin doğruluk derecesi nedenli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca nedenler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. Örneğin modelin kurulması sırasında varsayılan enflasyon oranının zaman içerisinde değişmesi, bireyin satın alma davranışını belirgin olarak etkileyecektir.

2.5.4. Modelin kullanılması

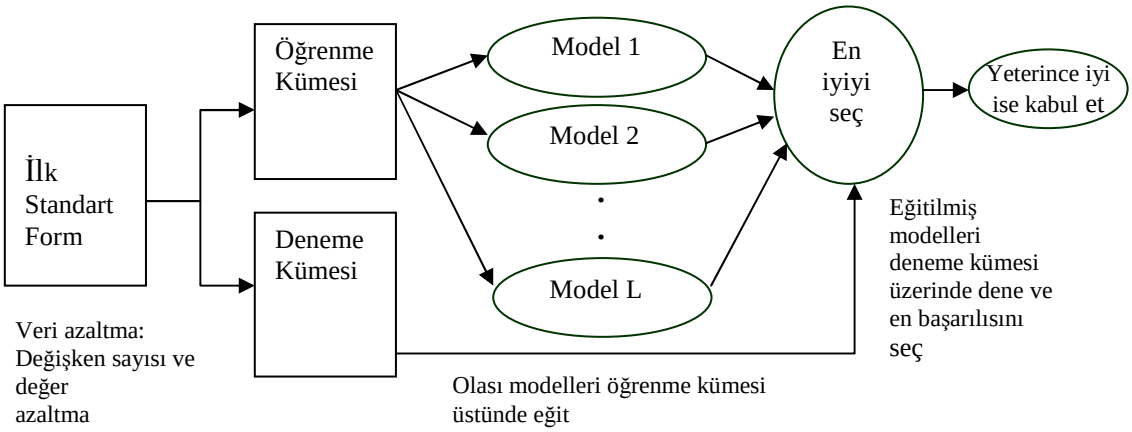
Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabileceği gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir.

2.5.5. Modelin izlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.

2.6. Veri Madenciliğinin Metodolojisi

Bir veri madenciliği çalışmasında kullanılan metodoloji Şekil 2.4’de verilmiştir. Standart form içinde verilen veri, öğrenme ve deneme olmak üzere ikiye ayrılır. Her uygulamada kullanılabilecek birden çok teknik vardır ve önceden hangisinin en başarılı olacağını kestirmek olası değildir. Bu yüzden öğrenme kümesi üzerinde L değişik teknik kullanılarak L tane model oluşturulur. Sonra bu L model deneme kümesi üzerinde denenerek en başarılı olanı, yani deneme kümesi üzerindeki tahmin başarısı en yüksek olanı seçilir [4].



Şekil 2.4. Veri madenciliği çalışmasında kullanılan metodoloji (Alpaydın 2000)

Eğer bu en iyi model yeterince başarılıysa kullanılır, aksi takdirde başa dönerek çalışma tekrarlanır. Tekrar sırasında başarısız olan örnekler incelenerek bunlar üzerindeki başarının nasıl arttırılabileceği araştırılır. Örneğin standart forma yeni alanlar ekleyerek programa verilen bilgi arttırılabilir veya olan bilgi değişik bir şekilde kodlanabilir veya amaç daha değişik bir şekilde tanımlanabilir [4].

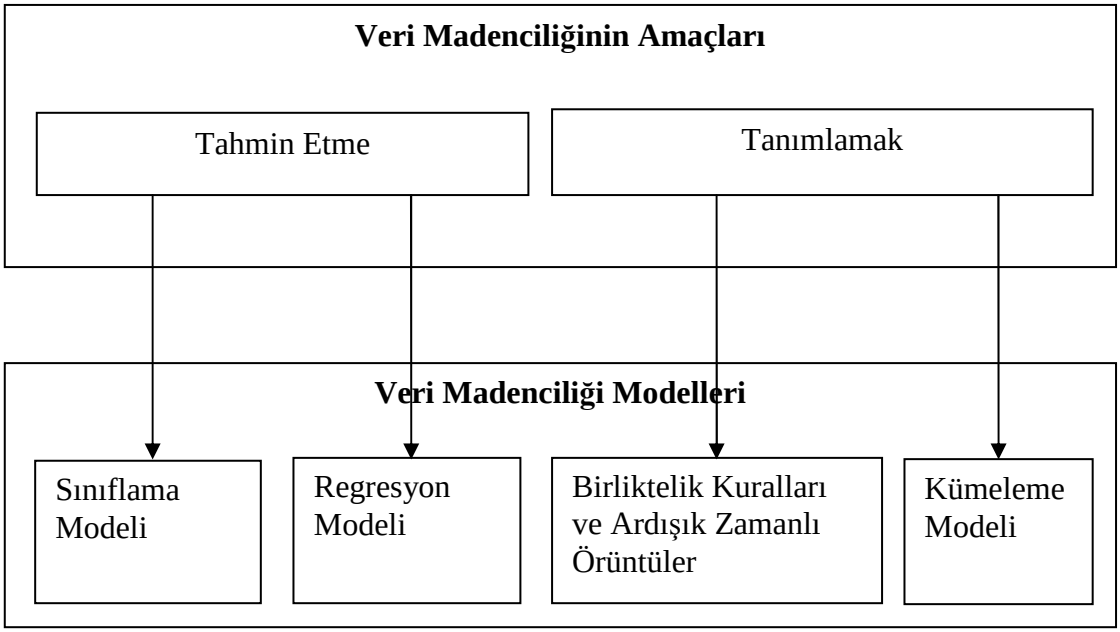
2.7. Veri Madenciliğinin Fonksiyonları

Veri madenciliği, yapılan analizde bir sonucu tahmin etmek ya da belirli bir sonucu tanımlamak amacı ile kullanılmaktadır. Bu nedenle veri madenciliği fonksiyonları, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir.

Veri madenciliği modellerini gördükleri işlevlere göre,

- Sınıflama (Classification),
- Regresyon (Regression),
- Kümeleme (Clustering),
- Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns)

olmak üzere dört ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir.



Şekil 2.5. Tahmin edici ve tanımlayıcı modeller

2.7.1. Tahmin / öngörü (Supervised) fonksiyonları

Geçmiş verilerden yararlanarak, gelecek ile ilgili bir sonucu tahmin etmek için kullanılan fonksiyonlardır. Yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Modellemelerinde olası sonucu öngörmeye yarayan faktörler ve sonuç yer alır. Model kurulurken geçmiş deneyimlerde, faktörlerin aldığı değerlere göre elde edilen sonuçlar girdi olarak kullanılır. Beklenen sonuç; "Katılır-Katılmaz" şeklinde kategorik değer veya rakamsal değerdir. Tahmin edilen sonuçların kalitesi (ne kadar iyi tahmin edildiği) tahmin edilen sonuç kadar önemlidir. Çoğunlukla tahmin edilen sonuç ile birlikte, bu sonucun kalitesine yönelik; güvenlik aralığı, olasılığı, vb. değerleri belirlenir [6].

Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Örneğin bir banka

önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken değeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır.

Tahmin edici modellerde amaç veritabanındaki bazı alanların diğer alanlara bağlı olarak tahmin edilmesidir. Tahmin edilecek alan eğer sayısal (sürekli) bir değişken ise tahmin problemi bir regresyon problemidir. Eğer tahmin edilecek alan kategorik bir değişken ise sınıflama problemidir. Sınıflama ve regresyon için kullanılan çok fazla sayıda değişken bulunmaktadır. Tahmin edici modellerde problem; diğer alanlardaki (girdiler), her gözlem için hedef değişken değerinin verilmiş olduğu eğitim veri seti ve problem hakkında önceden sahip olunan bilgileri yansıtan varsayımların kümesinin verilmesi durumunda tahmin edilecek değişkenin alabileceği muhtemel değerin belirlenmesi şeklinde özetlenebilir [24].

2.7.1.1. Sınıflandırma (Classification)

En temel veri madenciliği fonksiyonlarından biriside kategorik sonuçları tahmin etmek için kullanılan modellerdir. Modeli kurabilmek için, sonuçları önceden bilinen durumlar ve bu durumlarda ilgili faktörlerin aldığı değerler gereklidir. Bu değerler “eğitim verisi” olarak adlandırılır. Elde edilmesi beklenen sonuç “müşteri %80 ihtimal ile bu kampanyaya olumlu yanıt verecek” şeklinde belirli bir olasılık ile birlikte sunulur. Sonuçlar “Hizmeti Bırakır-Hizmeti Bırakmaz” şeklinde iki alternatifli olabileceği gibi “Kesin Tercih Eder-Tercih Eder-Yanıt Vermez-Tercih Etmez-Kesinlikle Tercih Etmez” şeklinde çoklu alternatifli de olabilir. Bir deneme kümesi modelin doğruluğunu belirlemek için kullanılır. Genellikle verilen veri kümesi öğrenme ve deneme kümesi olarak ikiye ayrılır. Öğrenme kümesi modeli oluşturulmasında, deneme kümesi modelin doğrulanmasında kullanılır. Örneğin bir otomobil satıcısı şirket geçmiş müşteri hareketlerinin analizi ile yukarıdaki gibi iki kural bulursa genç kadınların okuduğu bir dergiye reklam verirken küçük modelinin reklamını verir [6].

Sınıflama belki de veri madenciliği uygulamalarında en çok kullanılan yöntemdir [30]. Sınıflama, daha önceden belirlenmiş kriterlere göre, örneğin yaşa, cinsiyete, gelir durumuna, eğitim düzeyine ve müşterinin kredi borcunu zamanında ödeyip ödememesine, bir kampanyaya olumlu cevap verip vermemesine, hedeflenen değerlerin üzerinde bulunup bulunmamasına yani ilgilenilen herhangi bir özelliğe veya birkaç kritere göre yapılır.

Uygulama Alanları : Potansiyel müşteriler için düzenlenen kampanyalara dönüşler, mevcut müşterilerin belirli bir hizmeti almaktan vazgeçme olasılıkları, kredi başvurularının risk seviyeleri, çeşitli belirtilere göre hastalık ihtimalleri, vb. [6].

Örnek Model **:** Satışlarını artırmak için kampanya düzenlemek isteyen bir otomobil firması, kampanyasına katılma ihtimali olan potansiyel alıcıları belirlemek için daha önceden satış yapmış olduğu müşterilerinin verilerini (sonuçlarını) kullanarak, hangi özelliklere sahip adayların kampanyaya katılabileceğini belirli bir olasılık aralığında tahmin edebilir. Bu şekilde; ihtiyacı kadar veri satın alarak (eğer adayların verisini dışarıdan alıyorsa) ve sadece alma potansiyeli yüksek olan adaylara ulaşmaya çalışarak tasarruf sağlamaktadır [6].

Sınıflama modellerinde kullanılan başlıca yöntemler / algoritmalar şunlardır [6]:

- Yapay Sinir Ağları (Neural Networks),
- Bayes Sınıflandırması (Bayesian Classification),
- En Yakın Komşu (Nearest Neighbour),
- Karar Destek Makineleri (Support Vector Machines),
- Zaman Serisi Analizi (Time Series Analysis),
- Karar Ağaçları (Decision Trees),
- Lojistik Regresyon (Logistic Regression)

2.7.1.2. Regresyon / eğri uydurma (Regression)

Süreklilik gösteren değerleri tahmin etmek için kullanılan fonksiyonlardır. Regresyon ile amaç girdiler ile çıktıyı ilişkilendirecek modeli oluşturup, en iyi

tahmine ulaşmaktır. Sonuç “bağımlı değişken”, girdiler “bağımsız değişken” olarak adlandırılır. Sonucun alacağı değer genellikle bir güvenlik aralığı içinde belirtilir. Girdiler, çözülecek probleme göre bir veya birden fazla olabilir. Örneğin; bir inşaat firması konut satışlarının, faaliyet gösterdiği bölgede elde edilen toplam gelir ile ilişkili olduğunu düşünüyorsa, sadece bölgesel gelire dayalı bir model oluşturarak, bölgesel gelirdeki değişime göre satacağı ev sayısını tahmin etme yoluna gidebilir. Ancak gerçek hayatta çözülecek problemlerin hemen hepsinde doğru tahmine ulaşmak için birden fazla girdiden faydalanmak gereklidir. Bu noktada önemli olan konu girdilerin sonucun doğru tahmin edilmesine yaptıkları katkıdır. Bazı durumlarda sonuca katkısı limitli olan girdileri modelden çıkarmak, daha etkin bir model oluşturmak için önemli bir gerekliliktir [6].

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok terimli lojistik regresyon (multinomial logistic regression) gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır.

Uygulama Alanları : Finansal tahminler, zaman serisi tahminleri, biomedikal ve ilaç reaksiyonları, konut fiyatı değerlendirmeleri, müşterinin yaşam çevrimi boyunca yarattığı değer, vb. [6].

Örnek Model : Bir dergiye ilk kez reklam vermeye başlayacak olan bir şirket daha önce reklam vermiş olduğu dergilerin sayfa maliyetlerini kullanarak, çalışılmaya başlanılacak olan derginin vermiş olduğu fiyatın uygunluk seviyesini belirli bir güven aralığı içinde değerlendirebilir. Ya da daha sonra yapacağı kampanyalarda çalışmakta olduğu dergilerin verecekleri fiyatların ne kadar makul olduğunu önceden öngörebilir [6].

Regresyon modellerinde kullanılan başlıca yöntemler / algoritmalar şunlardır [6]:

- Yapay Sinir Ağları (Neural Networks),
- Karar Destek Makineleri (Support Vector Machines),
- Karar Ağaçları (Decision Trees),
- Doğrusal Regresyon (Linear Regression)

2.7.2. Tanımlama (Unsupervised) fonksiyonları

Fonksiyonların amacı belirli bir hedefi tahmin etmek değildir. Amaç veri setinde yer alan veriler arasındaki ilişkileri, bağlantıları ve davranışları bulmaktır. Var olan verileri yorumlayarak davranış biçimleri ile ilgili tespitler yapmayı ve bu davranış biçimini gösteren alt veri setlerinin özelliklerini tanımlamayı hedefler. Tanımı bilmek; tekrarlanan bir faaliyete veya tanımı bilinen yeni bir verinin yapıya katılmasında ne şekilde hareket edileceği konusunda karar almaya destek olur [6].

Tanımlayıcı modeller karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir. 25 yaş altı bekar kişiler ile 25 yaş üstü evli kişiler üzerinde yapılan ve ödeme performanslarını gösteren bir analiz yine tanımlayıcı modellere örnek olarak verilebilir.

2.7.2.1. Kümeleme/gruplama/demetleme/öbekleme (Clustering)

Bölümleme olarak da bilinen kümeleme, öngörülecek alanların belirlenmesini ve birbirine benzeyen verilerin altkümelere ayrılmasını hedefler. Kümeleme analizinin hedefi, veri setinde doğal olarak meydana gelen altsınıfları bulmaktır [14]. Denetimsiz öğrenme olarak da görülen kümeleme, veri setinin, kümeler olarak adlandırılan sınıflar seti haline getirmek amacıyla bölümlenmesi sürecidir [22]. Her kümenin üyeleri bazı ortak ilginç özellikleri paylaşmaktadır.

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir. Başlangıç aşamasında veri tabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı bilinmemekte, konunun uzmanı olan bir kişi tarafından kümelerin neler olacağı tahmin edilmektedir.

Sınıflamada olduğu gibi ayrılması istenen küme sayısı önceden bilinmediğinden, kümeleme algoritmaları tipik olarak iki aşamalı bir arama gerçekleştirirler. Mümkün küme sayıları üzerinde dıştan bir döngü ve belirli sayıdaki küme için mümkün olan en iyi kümelemeye ulaşmak için içsel bir döngü gerçekleştirilir [30]. Kümeleme, müşterilere ait bir veri deposunda yapılırken müşteriler, birçok özellikleriyle birlikte analiz edilir ve sonuçta müşteri kimlikleriyle, müşteri adlarına, posta kodlarına veya tanımlanan müşteri numarasına göre kendiliğinden gruplanırlar. Tüm müşteriler kendisiyle benzer özelliklere, niteliklere sahip olan müşterilerle aynı gruba atanır. Kümeleme analizinin sonuçlarını kullanacak kişilerin, ayrışan bu grupları daha sonradan tanımlaması ve pazar bölümü olarak hedeflemesi mümkündür. Çünkü kendi içinde çok çeşitli açılardan benzer özellikler, benzer tutum ve davranışlar gösteren bu grupların pazarlama faaliyetlerinde de benzer tepkiler oluşturacağı varsayılmaktadır [28].

Sınıflandırma fonksiyonunda tanımlı girdiler ve bunların geçmişte aldıkları değerler temel modeli oluştururken, kümeleme fonksiyonunda önceden tanımlanmış girdiler ve örnekler yoktur. Veriler kendi içlerindeki benzerliklere göre gruplanırlar. Hangi promosyon kampanyasına müşteriler en iyi tepkiyi verirler diye değerlendirmek yerine öncelikli olarak müşterilerin belirli kümelere ayrılmasının ardından her küme için en iyi promosyon kampanyasının ne olacağı belirlenebilir [6].

Uygulama Alanları : Benzer hücreleri tanımlamak, benzer davranışlar gösteren perakende müşterilerini tanımlamak, gen ve protein analizleri, ürün gruplaması, hastalık belirtileri, metin madenciliği [6].

Örnek Model : İki boyutlu bir örnekte kümeleme fonksiyonunu algılamak oldukça kolaydır. Yaş ve gelir düzeyleri belirtilmiş 40 kişiden oluşan bir grubu, grafik yardımı ile kümelerine ayırmak mümkündür. Yaş ve gelir düzeyi değerlerinin histograma yerleştirilmesi ve en yoğun durumların merkez olarak belirlenmesi en basit anlamda bir kümeleme işlemidir. Bu örnekte veri madenciliği yöntemleri kullanılmadan kümeler oluşturulmuştur. Ancak onlarca değişken olduğunda verileri kolayca kümelemek mümkün değildir, bu aşamada kümeleme fonksiyonuna özgü algoritmaları kullanmak gereklidir [6].

Kümeleme modellerinde kullanılan başlıca yöntemler / algoritmalar şunlardır [6]:

- Bölme yöntemleri (Partitioning methods),
- Hiyerarşik yöntemler (Hierarchical methods),
- Yoğunluk tabanlı yöntemler (Density-based methods),
- Grid tabanlı yöntemler (Grid-based methods),
- Model tabanlı yöntemler (Model-based methods)

2.7.2.2. Birliktelik analizi / bağıntı / eşleme / ilişki kuralları (Association Rules)

Büyük veri kümeleri içinde farklı veriler arasındaki birliktelik ilişkilerini bulma işlemidir. Birliktelik analizi, belirli bir veri kümesinde yüksek sıklıkta birlikte görülen özellik değerlerine ait ilişkisel kuralların keşfidir. Sonuçta elde edilen birliktelik kuralları (A B) şeklinde sunulur. Birliktelik analizi şirketlerin karar alma işlemlerini daha verimli hale getirmektedir. En klasik örneği sepet analizidir (basket analysis). Bu analizde müşterilerin beraber satın aldığı ürünlerin analizi yapılır. Amaç ürünler arasındaki pozitif veya negatif korelasyonları bularak müşterilerin satınalma alışkanlıklarını ortaya çıkarmaktır. Çocuk bezi alan müşterilerin mama da satın alacağını veya deterjan satın alanların yumuşatıcıda alacağını tahmin edebiliriz ancak manuel olmayan bir analiz bütün olasılıkları göz önüne alır ve kolay düşünülemeyecek, “mama” ve “yumuşatıcı” gibi bağıntıları da bulur. Bu verilere sahip olan marketler, birlikte satılan ürünleri yakın raflara koyarak, katalogda birlikte satılan ürünlerin birlikte görülmesini sağlayarak veya müşteriler için cazip ürün paketleri oluşturarak satışları artırabilirler [6].

Birliktelik kuralları, birbiriyle ilişkili olan değişkenlerin ortaya çıkarılması ve aralarındaki bağlantının büyüklüğünün tespit edilmesine yöneliktir. Birliktelik kuralları belirli türlerdeki veri yapıları arasındaki ilişkileri tanımlamaya çalışan bir yöntemdir [12]. Birliktelik kuralları ile veriler arasındaki olasılıksal korelasyon bulunmaya çalışılır. Olaylar arasında görülen korelasyon ise bu olayların sık sık beraber gözlendiklerini ifade etmektedir [32].

Bir alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi, müşteriye daha fazla ürünün satılmasını sağlama yollarından biridir. Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak pazar sepeti analizi (Market Basket Analysis) adı altında veri madenciliğinde yaygın olarak kullanılmaktadır. Bununla birlikte bu teknikler, tıp, finans ve farklı olayların birbirleri ile ilişkili olduğunun belirlenmesi sonucunda değerli bilgi kazanımının söz konusu olduğu ortamlarda da önem taşımaktadır.

Birliktelik kuralları aşağıda sunulan örneklerde görüldüğü gibi eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır.

- Müşteriler kola satın aldığında, % 75 ihtimalle patates cipsi de alırlar,
- Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diet süt de satın alırlar.

Uygulama Alanları : Birlikte hareket eden verilerin bulunması ile verimlik sağlanacak her alanda kullanılabilir. Süpermarkette birlikte satılan ürünler, otomobilde sunulacak ekstra özellikler, depolarda birbirine yakın konumlandırılması gereken ürünler, alışveriş merkezinde olması gereken mağazalar, vb. [6].

Örnek Model : Bir A ürününü satın alan müşteriler aynı zamanda B ürününü da satın alıyorlarsa, bu durum A B [destek = %2, güven = %60] şeklinde ifade edilir. Buradaki destek ve güven değerleri, birliktelik kuralının ilginçlik ölçüleridir. “Destek” tanımlanan kuralın sıklığını ve “güven” tanımlanan kuralın kabul edilebilirliğini gösterir. %2 oranındaki bir destek değeri, analiz edilen tüm

alışverişlerden %2'sinde A ile B ürünlerinin birlikte satıldığını belirtir. %60 oranındaki güven değeri ise A ürününü satın alan müşterilerinin %60'ının aynı alışverişte B ürününü de satın aldığını ortaya koyar. Kullanıcı tarafından minimum destek eşik değeri ve minimum güven değeri belirlenir ve bu değerleri aşan birliktelik kuralları dikkate alınır. Büyük veri tabanlarında birliktelik kuralları bulunurken, iki aşamalı bir süreç işletilir. İlk aşamada sık tekrarlanan öğeler bulunur: Bu öğelerin her biri en az, önceden belirlenen minimum destek sayısı kadar sık tekrarlanırlar. İkinci aşamada sık tekrarlanan öğeler arasından güçlü birliktelik kuralları oluşturulur [6].

Birliktelik analizi modellerinde kullanılan başlıca yöntemler / algoritmalar şunlardır [6]:

- Apriori

2.7.2.3. Sıralı dizi analizi (Sequence Analysis / Sequential Paerns)

Gözlem sonuçlarının zaman ve mekan özelliklerine göre sıralanmış olarak gösteren sayı dizileridir. Sayısal sıralı verilerdeki trendleri ve döngüleri anlamak için kullanılır. Bu fonksiyonda ilişkili kayıtlar incelenir ve zaman içinde sıkça rastlanan trendler ve benzer trendler bulunur. Bu trendler daha sonra veri içindeki ilişkileri tanımlamak için kullanılır. Bir beyaz eşya perakendecisinin veritabanından buzdolabı alımını takip eden beyaz eşya alımının bulaşık makinesi olduğunun belirlenmesi, doğal afetler veritabanından 6 büyüklüğünde bir deprem olduktan 3 gün sonra Klimanjaro dağının püskürmesi, banka veritabanından ilk üç taksitinden iki veya daha fazlasını geç ödemiş olan müşterilerin % 60 olasılıkla kanuni takibe gidiyor olduklarının belirlenmesi gibi örnekleri vardır. Kredi kartı örneğinde belirlenen davranış skoru (behavioral score), başvuru skorundan farklı olarak kredi almış ve taksitleri ödeyen bir kişinin sonraki taksitlerini ödeme/geciktirme davranışını notlamayı amaçlar. Seriler özelliklerine göre “zaman serileri”, “mekan serileri”, “bölünme serileri” ve “bileşik seriler” olmak üzere dört başlık altında incelenebilirler [6].

Ardışık zamanlı örüntüler aşağıda sunulan örneklerde görüldüğü gibi birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır.

- X ameliyatı yapıldığında, 15 gün içinde % 45 ihtimalle Y enfeksiyonu oluşacaktır,
- İMKB endeksi düşerken A hisse senedinin değeri % 15'den daha fazla artacak olursa, üç iş günü içerisinde B hisse senedinin değeri % 60 ihtimalle artacaktır,
- Çekiç satın alan bir müşteri, ilk üç ay içerisinde % 15, bu dönemi izleyen üç ay içerisinde % 10 ihtimalle çivi satın alacaktır.

Zaman Serisi Analizi / Benzer Zaman Sıraları/ Zaman İçinde Sıralı Örüntüler (Similar Time Sequences / Time Series) : Gözlem sonuçlarının zamana göre sıralanmış şeklidir. Borsada yer alan hisselerin davranışları sık rastlanan bir örneğidir. Günlere göre hisse değeri, yıllara göre faiz oranları, aylara göre üretim fire oranı, vb. gibi örnekleri vardır. Tek bir seri dışında, birden fazla hareket serisi arasında da bağıntı kurmak mümkündür. Bunlar örneğin iki malın zaman içindeki satış miktarları olabilir. Örneğin dondurma satışları ile kola satışları arasında pozitif, dondurma satışları ile salep satışları arasında negatif bir bağıntı beklenebilir. Zaman serisinde yer alan verilerin davranışları trend ve döngüler (cycle) ile tanımlanır. "Trend" serideki verilerin ortalama değerinde yaşanan değişimi tanımlamak için kullanılır. "Döngü"veride tekrar eden herhangi bir davranışı tanımlamak için kullanılır. Sezonsal veya dönemsel olabilir. Sezonsal olanlar tahmin edilebilir zamanlarda gerçekleşir, (her pazartesi, her yılbaşı, vb.) dönemsel olanlar "n" zaman aralıkları ile kendini tekrarlar. Zaman serisi analizlerinde veri serisindeki davranışları belirlemek kadar gelecek değerleri tahmin etme çalışmaları da gerçekleştirilir. (Hisse değerlerini, ekonomik değerleri, ürün talebini hava durumunu tahmin etmek, vb.)

Mekan Serisi : Gözlem sonuçlarının mekana göre sıralanmış şeklidir. Bölgelere göre satış rakamları, ülkelere göre yaşam süresi, vb.

Bölünme Serisi (Frekans) : Gözlem sonuçlarının belirlenen kriterlere göre sıralanmış şeklidir.

Bileşik Seri : Gözlem sonuçlarının iki ya da daha fazla özelliğe göre bir arada gösterilmiş şeklidir.

2.8. Veri Madenciliğinin Algoritmaları (Metotları/Teknikleri)

Veri madenciliği, sahip olunan verilerden yola çıkarak daha önce keşfedilmemiş bilgileri ortaya çıkarma ve bunları karar alma sürecinde kullanma yöntemidir. Veri madenciliği, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların analiz ve yazılım tekniklerinin kullanılarak ortaya çıkarılmasıdır. Bu açıdan bakıldığında veri madenciliği istatistiksel bir yöntemler serisi olarak görülebilir. Benzer şekilde veri madenciliğiyle ilgili yazılım ürünleri ve uygulamalara bakıldığında da veri madenciliğinin esasen istatistiğin kullanıldığı bir teknik olduğu görülmektedir. Ancak önemli olan kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği sadece istatistik değildir, insan merkezli bir uygulamadır [6].

Veri madenciliği fonksiyonlarının kullandığı bazı kritik teknikler ve tanımlamaları şu şekildedir;

- Karar Ağaçları
- Regresyon
- Lojistik Regresyon
- Bayes
- Apriori
- Kümeleme Teknikleri
- Yapay Sinir Ağları
- Genetik Algoritmalar

2.8.1. Karar ağaçları

İstatistiksel yöntemlerde veya yapay sinir ağlarında veriden bir fonksiyon öğrenildikten sonra bu fonksiyonun insanlar tarafından anlaşılabilecek bir kural olarak yorumlanması zordur. Karar ağaçları ise veriden oluşturulduktan sonra ağaç kökten yaprağa doğru inilerek kurallar (IF-THEN rules) yazılabilir [26]. Bu şekilde kural çıkarma (rule extraction), veri madenciliği çalışmasının sonucunun doğrulanmasını sağlar. Bu kurallar uygulama konusunda uzman bir kişiye gösterilerek sonucun anlamlı olup olmadığı denetlenebilir. Sonradan başka bir teknik kullanılacak bile olsa karar ağacı ile önce bir kısa çalışma yapmak, önemli değişkenler ve yaklaşık kurallar konusunda analiste bilgi verir ve daha sonraki analizler için yol gösterici olabilir [19].

Karar ağaçları genellikle sınıflama amacıyla kullanılan bir veri madenciliği tekniğidir. Karar ağacı, akış diyagramına benzer bir ağaç yapısında olup, her bir dal bir testin sonucunu, yaprak düğümleri ise sınıfları temsil eder. Bilinmeyen bir örneklemi sınıflamak için örneklemin nitelik değerleri karar ağacı karşısında test edilir. Kökten, o örneklem için sınıf tahminini içeren yaprak düğümüne kadar bir yol izlenir. Karar ağaçları kolaylıkla sınıflama kurallarına dönüştürülebilir [37]. Diğer tekniklerle karşılaştırıldığında karar ağaçlarının yorumlanması, anlaşılması ve yapılandırılması daha kolaydır [1]. Bu teknikte sınıflama yapılırken ilk önce veri setinden bir ağaç meydana getirilir. Bu ağaç meydana getirildikten sonra veri setindeki her bir kayıt bu ağaca uygulanarak bu kayıt sınıflandırılır.

Veri madenciliğinde karar ağacı modelleri veriyi incelemek ve tahmin yapmak için sıklıkla kullanılmaktadır [9]. Karar ağaçları, yinelenen bölünmelerle verileri farklı gruplara ayırarak büyür ve bu ayırmanın amacı her bölünmede veri grupları arasındaki uzaklığı arttırmaktır [38]. Kategorisel değişkenleri tahmin etmekte kullanılan karar ağaçları, olayları kategori ya da sınıflara ayırdığı için aynı zamanda sınıflama ağaçları (classification trees) olarak da adlandırılır. Sürekli değişkenleri tahmin etmekte kullanılan karar ağaçları ise regresyon ağaçları olarak adlandırılırlar [5].

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde

- Kuruluşlarının ucuz olması,
- Yorumlanmalarının kolay olması,
- Veri tabanı sistemleri ile kolayca entegre edilebilmeleri,
- Güvenilirliklerinin daha iyi olması

nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahiptir.

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı sahalar, [3]

- Belirli bir sınıfın muhtemel üyesi olacak elemanların belirlenmesi (Segmentation),
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması (Stratification),
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması,
- Parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden faydalı olacakların seçilmesi,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması**,**
- Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikliye dönüştürülmesidir.

Karar ağacı temelli tipik uygulamalar ise, [3]

- Hangi demografik grupların mektupla yapılan pazarlama uygulamalarında yüksek cevaplama oranına sahip olduğunun belirlenmesi (Direct Mail),
- Bireylerin kredi geçmişlerini kullanarak kredi kararlarının verilmesi (Credit Scoring),
- Geçmişte işletmeye en faydalı olan bireylerin özelliklerini kullanarak işe alma süreçlerinin belirlenmesi,
- Tıbbi gözlem verilerinden yararlanarak en etkin kararların verilmesi,
- Hangi değişkenlerin satışları etkilediğinin belirlenmesi,
- Üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin belirlenmesidir.

Gerçek dünyanın sosyal ve ekonomik olaylarını daha güvenilir bir şekilde gösterebilmek için standart istatistik tekniklerin dışında yeni analiz tekniklerinin geliştirilmesi ile ilgilenen Morgan ve Sonquist tarafından University of Michigan'da 1970'li yılların başlarında kullanıma alınan Automatic Interaction Detector – AID karar ağacı temelli ilk algoritma ve yazılımdır. AID tekniği en kuvvetli ve en iyi tahmini gerçekleştirebilmek için bağımlı ve bağımsız değişkenler arasındaki mümkün bütün ilişkilerin incelenmesine dayanmaktadır. Karar ağacı tekniğinin sağladığı kuruluş ve yorumlama kolaylıkları, AID yazılımının başlangıçta istatistikçi ve veri analistleri tarafından büyük çoşku ile karşılanmasına neden olmuştur. Ancak AID'in bağımlı ve bağımsız değişkenler arasındaki ilişkilerin tanımlanmasında aşırı saldırgan davrandığı ve bunun sonucunda anlamlı ve anlamsız ilişkileri ayırt edemediği yönünde Einhorn başta olmak üzere bir çok araştırmacı tarafından yayınlar yapılmıştır.

İlk temelleri AID yöntemi ile atılan karar ağacı modelleri çeşitli algoritmalar ile sürdürülmüştür.

Geliştirilen bu algoritmalar içerisinde

- CHAID (Chi-Squared Automatic Interaction Detector; G.V. Kass; 1980),
- C&RT (Classification and Regression Trees; Breiman, Friedman, Olshen ve Stone; 1984),
- ID3 (Quinlan; 1986),
- Exhaustive CHAID (Biggs, de Ville ve Suen; 1991),
- C4.5 (Quinlan; 1993),
- MARS (Multivariate Adaptive Regression Splines; Friedman),
- QUEST (Quick, Unbiased, Efficient Statistical Tree; Loh ve Shih, 1997),
- C5.0 (Quinlan),
- SLIQ (Supervised Learning in Quest; Mehta, Agarwal ve Rissanen),
- SPRINT (Scalable Parallelizable Induction of Decision Trees; Shafer, Agrawal ve Mehta)

başlıcalarıdır.

Karar ağacı, adında belirtildiği şekilde ağaç görünümünde bir tekniktir. Karar düğümleri, dallar ve yapraklardan oluşur [6].

Karar düğümü : Veriye uygulanacak test tanımlanır. Her düğüm bir özellikteki testi gösterir. Test sonucunda ağacın dalları oluşur. Dalları oluştururken veri kaybı yaşanmaması için verilerin tümünü kapsayacak sayıda farklı dal oluşturulmalıdır.

Dal : Testin sonucunu gösterir. Elde edilen her dal ile tanımlanacak sınıfın belirlenmesi amaçlanır. Ancak dalın sonucunda sınıflandırma tamamlanamıyorsa tekrar bir karar düğümü oluşur. Karar düğümünden elde edilen dalların sonucunda sınıflandırmanın tamamlanıp tamamlanmadığı tekrar kontrol edilerek devam edilir.

Yaprak : Dalın sonucunda bir sınıflandırma elde edilebiliyorsa yaprak elde edilmiş olur. Yaprak, verileri kullanarak elde edilmek istenen sınıflandırmanın sınıflarından birini tanımlar.

Başlangıçta bütün öğrenme örnekleri kök düğümdedir, örnekler seçilmiş özelliklere tekrarlamalı olarak göre bölündükten sonra ağacı temizlemek için (Tree pruning) gürültü ve istisna kararları içeren dallar belirlenir ve kaldırılır. Karar ağacı tekniğini kullanarak verinin sınıflanması üç aşamadan oluşur.

Öğrenme : Önceden sonuçları bilinen verilerden (eğitim verisi) model oluşturulur.

Sınıflama : Yeni bir veri seti (test verisi) modele uygulanır, bu şekilde karar ağacının doğruluğu belirlenir. Test verisine uygulanan bir modelin doğruluğu, yaptığı doğru sınıflamanın test verisindeki tüm sınıflara oranıdır. Her test örneğinde bilinen sınıf, model tarafından tahmin edilen sınıf ile karsılaştırılır.

Uygulama : Eğer doğruluk kabul edilebilir oranda ise, karar ağacı yeni verilerin sınıflanması amacıyla kullanılır.

2.8.2. Regresyon analizi (Regression Analysis)

Bir ya da daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini sağlayacak ilişkiler bulmak ve bunları tanımlamaktır. Regresyon analizinin temelinde gözlenen bir olayın değerlendirilirken, hangi olaylardan etkilendiğini belirlemek yatmaktadır. Bu olaylar bir veya birden çok olacağı gibi etki düzeyleri farklı seviyelerde de olabilir [6].

Regresyonda, verilerin matematiksel gösterimle, bir fonksiyon olarak tanımlanması gerekmektedir. Regresyon analizi yapılırken kurulan matematiksel modelde yer alan değişkenler bir bağımlı değişken ve bir veya birden çok bağımsız değişkenden oluşmaktadır. Değişkenler sayılabilir veya ölçülebilir niteliktedir. Örneğin bir hissenin fiyatını ile ona dolaylı veya direkt etkili olan faiz oranları, enflasyon, vb. gibi bir veya birden çok değişken ile ilişkilendirmek mümkündür. Sadece faiz oranlarının etkisi ile ilgileniyorsak, tek değişkenli bir matematiksel model, faiz oranları ile birlikte enflasyon oranı ile de ilgileniyorsak, iki değişkenli bir matematiksel model kurulmalıdır. Tek değişkenli modeller basit doğrusal regresyon (doğrusal ilişkiyi temsil eden bir doğrunun denklemi formüle edilir), birden fazla bağımsız değişkenli modeller çoklu regresyon modeli konusunu oluşturmaktadır [6].

Tek Değişkenli Regresyon - Lineer Regresyon : Basit lineer regresyon iki sürekli değişken (tahmin edilmeye çalışılan bağımlı değişken ve bağımsız değişken) arasındaki ilişkiyi tanımlamayı amaçlayan bir tekniktir. Teknik verileri kullanarak bir doğru denklemi oluşturmayı hedefler. Bu doğru oluşturulurken tüm veri noktalarından tahmin edilen eğriye olan uzaklığın karelerinin minimize edilmesi ile doğrunun optimize edilmesi sağlanır. Doğru elde edildikten sonra iki değişken arasındaki ilişkinin gücü R-kare (R-Square) değeri ile tanımlanır. R-kare verinin değişiminin ne ölçüde oluşturulan model (çizilen doğru) ile açıklanabildiğini gösterir.

Tek Değişkenli Regresyon – Lineer Olmayan Regresyon : Bazı durumlarda bağımlı vebağımsız değişkenler arasındaki ilişki doğrusal olmayabilir. Bu gibi durumlarda daha iyi bir uyum için bağımsız değişkeni modifiye etmek gerekebilir.

Çoklu Regresyon : Pazarlama, risk yönetimi, müşteri ilişkileri yönetimi konularında model oluşturulurken birden fazla değişkenin bağımlı değişken üzerinde etki ediyor olması çok doğal ve genellikle rastlanan bir durumdur. Bazı durumlarda değişkenler yüzler ile ifade edilecek seviyelere çıkabilir.

2.8.3. Lojistik regresyon (Logistic Regression)

Lojistik regresyon lineer regresyona çok benzer olmakla birlikte, lojistik regresyonda bağımlı değişkenin kesikli veya kategorik olması (sürekli olmaması) en önemli farklılıktır. Bu fark özellikle bir teklife yanıt veya bir seçim yapmak gibi kesikli aksiyonları belirlemeye yönelik sınıflandırma modellerinde önem kazanmaktadır. (Sınıflandırma analizlerinde doğrusal regresyonun kullanılması mümkün olmamaktadır.) Lojistik regresyon, çok değişkenli normal dağılım varsayımına ihtiyaç göstermediğinden bu tür uygulamalarda avantaj sağlamaktadır. Lojistik regresyon ile bağımsız değişkenleri kullanarak ikili çıktısı olan bağımlı değişkenin istenilen durumunun gerçekleşme olasılığını hesaplanır. Regresyon yapabilmek için bağımlı değişken sürekli değere dönüştürülür. Bu değer beklenen olayın olma olasılığıdır [6].

2.8.4. Bayes

İstatistiksel bir sınıflandırıcıdır. Belirsiz durumlarda tahmin yapmak, sınıflandırma yapmak için kullanılır. Eldeki verilerin belirlenmiş olan sınıflara ait olma olasılıklarını öngörür. İstatistikteki Bayes teoremine dayanır. Bu teorem; belirsizlik taşıyan herhangi bir durumun modelinin oluşturularak, bu durumla ilgili evrensel doğrular ve gerçekçi gözlemler doğrultusunda belli sonuçlar elde edilmesine olanak sağlar. Belirsizlik taşıyan durumlarda karar verme konusunda çok kullanışlıdır. En önemli zafiyeti değişkenler arası ilişkinin modellenmiyor olması ve değişkenlerin birbirinden tamamen bağımsız olduğu varsayımıdır [6].

Bayes yöntemi koşullu olasılık durumları ile ilgilidir. Her hangi bir koşullu olasılık durumu P(X=x | Y=y) = R şeklinde tanımlanır. Bu ifade; “Eğer Y = y doğru ise, X = x olma olasılığı R’dir” anlamına gelmektedir. X ve Y ‘nin alabileceği değerlerin her

kombinasyonu için koşullu olasılıkları belirleyen tabloya koşullu olasılık dağılımı adı verilir ve P(X|Y) ile ifade edilir [6].

2.8.5. Apriori algoritması

Apriori algoritmasını, sık geçen örüntülerin yakalanmasında kullanılan temel bir algoritmadır. Bu algoritma özünde yinelemeli bir yaklaşım sunar [34]. Örüntüleri ya da diğer bir deyişle sık geçen öge kümelerini bulmak için birçok kez veritabanını taramak gerekir. İlk taramada bir elemanlı minimum destek ölçüsünü sağlayan sık geçen öge kümeleri bulunur. Destek ölçüsü, öge kümesinin veritabanındaki kaç satır tarafından ya da veritabanındaki toplam satır sayısının yüzde kaçı tarafından doğrulandığını belirtir. İzleyen taramalarda bir önceki taramada bulunan sık geçen öge kümeleri aday kümeler adı verilen yeni potansiyel sık geçen öge kümelerini üretmek için kullanılır. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek metriğini saglayan kümeler o geçişte üretilen sık geçen öge kümeleri olur. Sık geçen öge kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen öge kümesi bulunmayana kadar devam eder. Bu algoritmadaki temel yaklaşım eğer *k*-öge kümesi minimum destek metriğini sağlıyorsa bu kümenin alt kümelerinin de minimum destek metriğini sağladığıdır [15].

2.8.6. Kümeleme yöntemleri

Bölünmeli yöntemler: Veriyi bölerek, her grubu belirlenmiş bir kritere göre değerlendirir. En yaygın olarak kullanılan iki algoritma vardır.

K-ortalaması (K-means) : K-ortalaması, en çok kullanılan merkezi sınıflama tekniklerinden biridir. Bu algoritmada, veri grubu tekrarlanarak –gözlem ve sınıf merkezi arasındaki öklit uzaklığının karesinin ortalamasını küçülterek– K adet sınıfa bölünür [2]. Başlangıç olarak verinin kaç kümeye ayrılacağını belirlemek gereklidir. Küme sayısı “k”değeri olarak adlandırılır. k-means algoritmasının 4 aşaması vardır [6]:

- Veri kümesinin rastsal olarak k altkümeye ayrılması (her küme bir altküme)

- Her kümenin ortalaması olan merkez noktanın (kümedeki nesnelerin niteliklerinin ortalaması) hesaplanması

- Nesnelerin küme merkezine olan uzaklıklarının değerlendirilmesi ve dahil olduğu kümenin merkezinden başka bir küme merkezine daha yakın olan nesnelerin yakın oldukları kümeye dahil edilmesi

- Yeni nesnelerle artan veya dışarıya nesne vererek azalan kümelerin ortalaması olan merkez noktalarıyeniden hesaplanır ve nesnelerin kümelenmesinde değişiklik olmayana kadar aynı şekilde devam edilir.

K-means yöntemi kurgulaması kolay vekarmaşıklığı az olan bir tekniktir. Ancak zayıf olduğu bazı önemli noktalar vardır. Sonuçları ilk başta merkez noktaların seçimine bağımlıdır. Merkez noktaların seçimine göre farklı sonuçlar ortaya çıkabilir. Bununla birlikte veri grupları farklı boyutlarda ise, veri gruplarının şekli küresel değilse ve veri içinde ortalamayı önemli ölçüde etkileyecek büyük bileşenler varsa çok iyi sonuçlar alınamayabilir [6].

K-medoids : K-means yönteminde; sadece kümenin ortalamasının tanımlanabildiği durumlarda kullanılma ve değeri çok büyük bir nesnenin kümede olması durumunda (kümenin ortalaması ve merkez noktası büyük ölçüde değişebileceğinden) kümenin hassasiyetinin bozulabilmesi gibi iki önemli zafiyet vardır. Bu sorunu gidermek için kümedeki nesnelerin ortalamasını almak yerine, kümede ortaya en yakın noktada konumlanmış olan nesne (medoid) kullanılabilmekte ve bu işlem k-medoids yöntemi ile tanımlanmaktadır. k-medoids yöntemi şu aşamalardan oluşur [6];

- Veri kümesi merkezi bir medoid olan k adet kümeye ayrılır.

- Veri kümesindeki nesneler, kendilerine en yakın olan medoide göre k adet kümeye yerleşirler.

- Bu bölünmelerin ardından kümenin ortasına en yakın olan nesneyi bulmak için medoid, medoid olmayan her nesne ile yer değiştirir. Bu işlem en verimli medoid bulunana kadar devam eder.

2.8.7. Yapay sinir ağları (Artificial Neural Networks)

Yapay sinir ağları, veri madenciliğinde belki de en çok bilinen fakat en az anlaşılan tekniktir [7]. Yapay sinir ağları değişik şekillerde tanımlanmaktadır. Bu tanımlardan bazılarında Yapay sinir ağları, insan beyninden esinlenerek tasarlanan, ağırlıklı bağlantılar yardımıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından meydana gelen paralel ve dağıtılmış bilgi işleme yapıları olarak belirtilmiştir.

Sinir ağları, veri madenciliğinde tahmin ve sınıflandırma analizlerinde kullanılan bir yöntemdir. Yapay sinir ağlarında beynin sinir sisteminin çalışma prensipleri model olarak alınmıştır. İnsan beyninin fiziksel düşünme işlemini taklit eden makinelere, yani yapay zekaya artan bir ilgi söz konusudur. Sinir hücresi, diğer sinir hücrelerini yani nöronları uyaran, sonrasında kendini de uyaran bir anahtar (switch) gibi çalışır. Nörondan çıkan akson diğer nöronların dentritlerine elektriksel olarak aktif bir link sağlar. Aksonlar ve dentritler, nöronları birbirine elektriksel olarak bağlayan tellerdir. Akson ve dentritin kesişim noktasına sinaps denir. Bu basit biyolojik model sinir ağlarının geliştirilmesinde kullanılan bir metafordur.

Yapay sinir ağları, insan beyninin fonksiyonel özelliklerine benzer bir şekilde [29],

- Öğrenme
- İlişki Kurma
- Sınıflama
- Genelleme
- Özellik belirleme
- Optimizasyon

gibi konularda da etkin olarak kullanılabilmektedir. Yapay Sinir Ağları, kendilerine verilen örneklerden elde ettikleri bilgiler ile kendi deneyimlerini meydana getirir ve daha sonra bu konulara benzer durumlarda benzer kararları verirler.

Yapay sinir ağları temelde aşağıdaki özelliklere sahiptir:

Yapay sinir ağlarının hesaplama ve bilgi işleme gücünü, paralel dağılmış yapısından, öğrenebilme ve genelleme yeteneğinden aldığı söylenebilir. Genelleme, eğitim ya da öğrenme sürecinde karşılaşılmayan girişler için de yapay sinir ağlarının uygun tepkileri üretmesi olarak tanımlanır. Bu üstün özellikleri, yapay sinir ağlarının karmaşık problemleri çözebilme yeteneğini gösterir. Günümüzde birçok bilim alanında yapay sinir ağları, aşağıdaki özellikleri nedeniyle etkin olmuş ve uygulama yeri bulmuştur.

Doğrusal Olmama : Yapay sinir ağlarının temel işlem elemanı olan hücre, doğrusal değildir. Dolayısıyla hücrelerin birleşmesinden meydana gelen yapay sinir ağlarında doğrusal değildir ve bu özellik bütün ağa yayılmış durumdadır. Bu özelliği ile yapay sinir ağları, doğrusal olmayan karmaşık problemlerin çözümünde en önemli araç olmuştur.

Öğrenme : Yapay sinir ağlarının arzu edilen davranışı gösterebilmesi için amaca uygun olarak ayarlanması gerekir. Bu, hücreler arasında doğru bağlantıların yapılması ve bağlantıların uygun ağırlıklara sahip olması gerektiğini ifade eder. Yapay sinir ağlarının karmaşık yapısı nedeniyle bağlantılar ve ağırlıklar önceden ayarlı olarak verilemez ya da tasarlanamaz. Bu nedenle yapay sinir ağları, istenen davranışı gösterecek şekilde ilgilendiği problemden aldığı eğitim örneklerini kullanarak problemi öğrenmelidir.

Genelleme : Yapay sinir ağları, ilgilendiği problemi öğrendikten sonra eğitim sırasında karşılaşmadığı test örnekleri için de arzu edilen tepkiyi üretebilir. Örneğin, karakter tanıma amacıyla eğitilmiş bir yapay sinir ağları, bozuk karakter girişlerinde de doğru karakterleri verebilir ya da bir sistemin eğitilmiş yapay sinir ağları modeli,

eğitim sürecinde verilmeyen giriş sinyalleri için de sistemle aynı davranışı gösterebilir.

Uyarlanabilirlik : Yapay sinir ağları, ilgilendiği problemdeki değişikliklere göre ağırlıklarını ayarlar. Yani, belirli bir problemi çözmek amacıyla eğitilen yapay sinir ağları, problemdeki değişimlere göre tekrar eğitilebilir ve değişimler devamlı ise gerçek zamanda da eğitime devam edilebilir. Bu özelliği ile yapay sinir ağları, uyarlamalı örnek tanıma, sinyal işleme, sistem tanılama ve denetim gibi alanlarda etkin olarak kullanılır.

Hata Toleransı : Yapay sinir ağları, çok sayıda hücrenin çeşitli şekillerde bağlanmasından oluştuğu için paralel dağılmış bir yapıya sahiptir ve ağın sahip olduğu bilgi, ağdaki bütün bağlantılar üzerine dağılmış durumdadır. Bu nedenle, eğitilmiş bir yapay sinir ağlarının bazı bağlantılarının hatta bazı hücrelerinin etkisiz hale gelmesi, ağın doğru bilgi üretmesini önemli ölçüde etkilemez. Bu nedenle, geleneksel yöntemlere göre hatayı tolere etme yetenekleri son derece yüksektir.

2.8.8. Genetik algoritmalar

Genetik Algoritmalar (GA), karmaşık optimizasyon problemlerinin çözülmesinde kullanılan bir teknolojidir. Genetik algoritmalar sezgisel bir yöntem olup geniş uygulama alanına sahiptir [26]. Genetik algoritmalar, finans, ekonomi, muhasebe, yöneylem araştırması, ulaştırma, dağıtım ve bunun gibi birçok alanda diğer arama algoritmalarına alternatif olarak oldukça popüler bir araç haline gelmiştir. Bunun temel nedeni, diğer sezgisel arama metotları bazen yerel optimum sonuçları verirken, genetik algoritmaların farklı sonuçları eşzamanlı olarak değerlendirerek ve arama eğilimini daha iyi çözüm alanlarına yönlendirerek bundan çoğunlukla kaçınmış olmasıdır [23]. Genetik algoritmalar, veri madenciliği amaçlı olarak başarılı bir şekilde kullanılmaktadır [39].

Genetik algoritmalar, veri yıgınlarındaki örüntüleri bulmak için tek başlarına kullanılmazlar, bunun yerine örnegin sinir agları gibi ögrenme temelli veri madenciligi algoritmalarına rehberlik etme amaçlı kullanılırlar. Aslında genetik

algoritmalar, çözüm uzayında bulunan en iyi modelin tespit edilmesinde kullanılan metotlardan biridir [38].

Bir çok alanda uygulama imkanı ve uygulamaları olan genetik algoritmaların işleme adımları şöyle açıklanabilir [13]:

- Arama uzayındaki tüm mümkün çözümler dizi olarak kodlanır.

- Genellikle rastsal bir çözüm kümesi seçilir ve başlangıç populasyonu olarak kabul edilir.

- Her bir dizi için bir uygunluk değeri hesaplanır, bulunan uygunluk değerleri dizilerin çözüm kalitesini gösterir.

- Bir grup dizi belirli bir olasılık değerine göre rastsal olarak seçilip çoğalma işlemi gerçekleştirilir.

- Yeni bireylerin uygunluk değerleri hesaplanarak, çaprazlama ve mutasyon işlemlerine tabi tutulur.

- Önceden belirlenen kuşak sayısı boyunca yukarıdaki işlemler devam ettirilir.

- İterasyon, belirlenen kuşak sayısına ulaşınca işlem sona erdirilir.

- Amaç fonksiyonuna göre en uygun olan dizi seçilir.

Genetik algoritmalar ancak;

- Arama uzayının büyük ve karmaşık olduğu,
- Mevcut bilgiyle sınırlı arama uzayında çözümün zor olduğu,
- Problemin belirli bir matematiksel modelle ifade edilemediği,
- Geleneksel eniyileme yöntemlerinden istenen sonucun alınmadığı alanlarda etkili ve kullanışlıdır.

Genetik algoritmaların özellikleri şu şekilde sıralanabilir:

- Genetik algoritmalar problemlerin çözümünü parametrelerin değerleriyle değil, kodlarıyla arar. Parametreler kodlanabildiği sürece çözüm üretilebilir. Bu sebeple genetik algoritmalar ne yaptığı konusunda bilgi içermez, nasıl yaptığını bilir.

- Genetik algoritmalar aramaya tek bir noktadan değil, noktalar kümesinden başlar. Bu nedenle çoğunlukla yerel en iyi çözümde sıkışıp kalmazlar.

- Genetik algoritmalar türev yerine uygunluk fonksiyonunun değerini kullanır. Bu değerin kullanılması ayrıca yardımcı bir bilginin kullanılmasını gerektirmez.

- Genetik algoritmalar gerekirci kuralları değil olasılıksal kuralları kullanır.

Bir genetik algoritmanın temel elemanları kısaca şöyle tanımlanmaktadır [29].

Kromozom ve Gen : Kromozon, GA'nın çözmesi istenen problemin her bir mümkün çözümünü göstermektedir. Belirlenen problem için N tane çözüm olabilir. Genetik algoritmanın bunların içinden en iyisini arayıp bulması istenmektedir. Kromozomlar ise bu çözümleri gösterirler. Başlangıçta rasgele olarak atanan çözümler daha sonra genetik algoritmanın çalışma prensiplerine göre iyileştirilmektedir. Bir kromozomun elemanlarından her birisi çözümün bir özelliğini göstermektedir. Bunlara da "Gen" adı verilmektedir.

Çözüm Havuzu : Problemin en iyi çözümünü aramak için kullanılan ve rasgele olarak belirlenmiş çözüm setidir. Problemin yapısına ve içeriğine göre değişen sayıda kromozom belirlenebilir. Çözüm havuzu içerisindeki her nokta mümkün bir çözümü ifade eder.

Çaprazlama : Çaprazlama, biyolojik terim olarak ele alınırsa üreme kromozomlarının birbirleriyle yapmış oldukları gen alışverişidir. Çaprazlama, problem çözüm havuzunda bulunan çözümleri (kromozomları) ikişer ikişer birleştirerek yeni çözümleri meydana getirmektir. GA' da seçim sonrası belli bir olasılıkla

kromozomların ikili olarak seçilmesi ve rasgele belirlenen bir noktanın sağ tarafındaki genlerinin karşılıklı olarak değiştirilmesi sonucu çaprazlama işlemi gerçekleştirilmiş olur. İki kromozomdan iki adet yeni kromozom üretilmektedir. Bir problem çözüm uzayından kaç adet kromozomum çaprazlanacağı önceden belirlenen çaprazlama oranına göre belirlenmektedir.

Mutasyon : Mutasyon, organizmalarda radyasyon veya başka sebeplerle değişiklik meydana gelmesinden esinlenerek geliştirilmiş bir operatördür. Çaprazlama neticesinde farklı çözümlere ulaşmak bazen zor olmaktadır.

Mutasyon, yeni çözümleri aramanın kolaylaştırılması ve aramanın yönünü değiştirmek amacı ile bir kromozomun bir elemanının (gen) değiştirilmesi işlemidir. Bir problemin havuzu içinden kaç kromozomun mutasyona uğratılacağına mutasyon oranına göre karar verilmektedir. Kromozomlarda tesadüfî olarak belirlenen bir noktada yapılan değişikliktir [33].

Uygunluk Fonksiyonu : GA tarafından bulunan çözümlerin uygunluk derecelerinin ölçülmesini sağlayan bir fonksiyondur [11]. Uygunluk fonksiyonu, kromozomların ne kadar iyi bir çözüm verdiğini hesaplayarak gösteren, genetik algoritmanın en temel yapısıdır. Her problem için bir uygunluk fonksiyonun belirlenmesi gerekmektedir.

Genetik algoritmada probleme özgü çalışan tek kısım uygunluk fonksiyonudur. Bu fonksiyon, probleme yapısına ve içeriğine göre değişmektedir. Mesela, bir iş çizelgelemesi (sıralanması) probleminde belirlenen her sıraya göre, işlerin tamamının bitirilmesinin maliyeti uygunluk ölçütü olarak belirlenebilir. En düşük maliyetle işleri bitiren iş sırasının belirlenmesi istenmektedir. İşlerin toplam maliyeti azaltıldığı sürece farklı iş sıralarının aranmasına devam eder. Bu maliyetin daha fazla azaltılamayacağı bir noktada en iyi çözümün GA tarafından bulunmuş olduğu kabul edilmektedir.

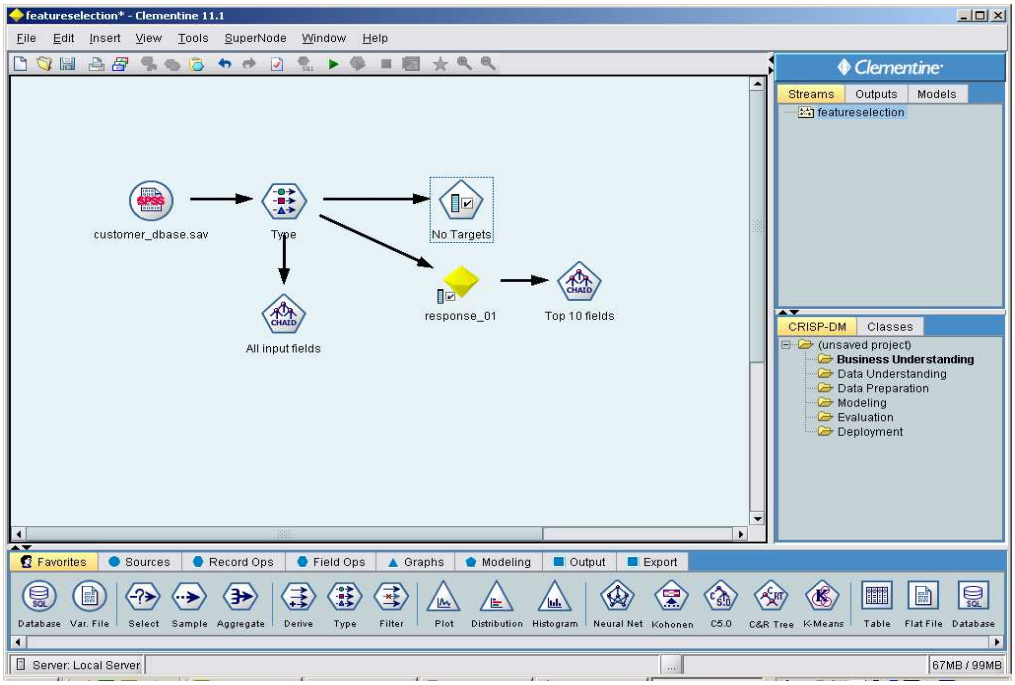
Yeniden Üretim : Çözüm havuzundaki kromozomlar çaprazlama ve mutasyon neticesinde meydana getirilen yeni kromozomlar sebebi ile artacaktır. Bunların

arasından problem havuz büyüklüğüne göre kromozomlar bazıları seçilerek diğerleri atılır. Seçilen kromozomlar ise bir sonraki nesil çözümü olarak yeniden çaprazlanıp gelecek çözümleri üretirler.

BÖLÜM 3. UYGULAMA

3.1. Uygulamada Kullanılan Clementine Programı

Clementine, SPSS Inc. Şirketi tarafından veri madenciliği uygulamaları için tasarlanmış ve veri madenciliği dünyasının yazılımları arasında tercih sıralamasında üç kez birincilik ödülünü almış bir yazılımdır. Görsellik önemli kılınarak tasarlanmış çalışma ekranında sürükle bırak ile nesne yerleştirme ve nesneleri birbirine bağlama işlemi kolaylıkla yapılabilmektedir. Clementine ile veri madenciliği adımları olan verinin hazırlanması, veri temizleme, veri birleştirme, seçme, dönüştürme, veri kalitesini belirleme, hata ayıklama, model kurma, modelin değerlendirilmesi ve modelin izlenmesi konularını gelişmiş bir teknoloji ile gerçekleştirme imkânı sunmaktadır. Aşağıdaki şekilde Clementine uygulama ekranından bir görüntü gösterilmektedir.



Şekil 3.1. Clementine Uygulama Ekranı

Clementine de veri modelleme aşamasında zengin bir içerik sunmaktadır. Clementine içerisinde yer alan modelleme yöntemleri 3 ana grup altında toplanmaktadır.

- Prediktif Modeller: Neural Networks, iki farklı rule induction tekniği C5.0 ve C&Rtree, regresyon, lojistik regresyon ve sequence detection olmak üzere 6 ayrı teknik çermektedir. Prediktif modellerde bir dizi input değeri baz alınarak bir "sonuç" değerinin tahmin edilmesi amaçlı modeller söz konusudur.

- Sınıflama Amaçlı Modeller: Benzer nitelik gösteren segmentlerin belirlenmesi amaçlıdır. Kohonen ağları, K-ortalama, İki adımlı sınıflama olmak üzere üç ayrı sınıflama yöntemi bulunmaktadır.

- Birliktelik Teknikleri: Genelleştirilmiş prediktif yöntemler olarak da tanımlanmakta olup, belirli bir sonucu bir dizi kural ile ilişkilendirmeye çalışırlar. Clementine içerisinde APRIORI ve GRI olmak üzere iki ayrı ilişkisel kural belirleme yöntemi vardır.

İş problemlerinin irdelenmesi aşamasında iş deneyimi önemlidir. Bu ilk adımda projenin amaç ve gerekliliklerinin iş persfektifi ile anlaşımlası, bu bilginin veri madenciliği problem tanımı olarak netleştirilmesi ve hedeflere ulaşma amaçlı ilk planların oluşturulması söz konusundur. Clementine ile birlikte opsiyonel olarak lisanslanan uygulama şablonları SPSS in farklı veri madenciliği projelerine dair ciddi bir iş deneyimini kullanıcılarına aktarmayı amaçlayarak hazırlanmış bir programdır.

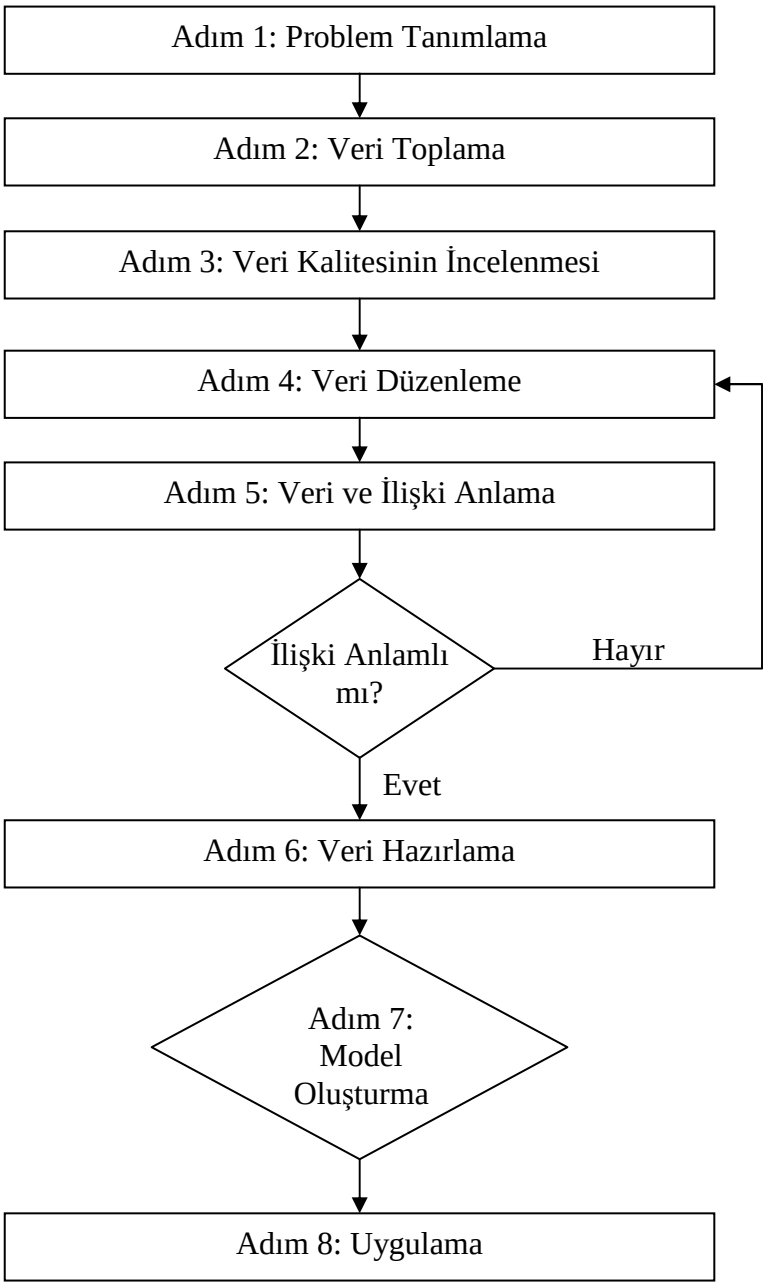
Verinin anlaşılması aşamasında veri kaynaklarına bağlanma, veriyi tanıma, verinin kalitesini anlama ve verinin grafiksel olarak incelenmesi, hipotezleri oluşturma amaçlı veri gruplarını değerlendirme çalışmalarında Clementine grafikler ve tablolar üzerinde belli bölgelerin seçimini yapma seçeneği sunmaktadır. Clementine içerisinde yer alan histogram, line plot, point plot, web associaion graphs, statistics, distribution graphs, data audit işlemcileri verinin ön incelemesinde sıkça kullanılan işlemcilerden bazılarıdır. Aşağıdaki tablo Clementine programında herbir aşamada amaçlanmış görevler listelenmiştir [36].

Tablo 3.1. Clementine programında her bir aşamada amaçlanmış görevler

İş Anlama	**Veri Anlama**	**Veri Hazırlık**	**Modelleme**	**Değerlendirme**	**Dağıtım**
İş hedefleri belirleme	İlk veri toplama	Veri seti	Modelleme tekniğini seçme	Veri keşfinin değerlendirme sonuçları	Plan dağıtımı
Arka plan hedefleri belirleme	İlk veri raporlama	Veri kümesinini tanımlama	Modelleme varsayımını belirleme	Madencilik sonuçları	Plan izleme ve bakımı
İş başarı hedefleri belirleme				İş başarı ölçütleri	Sonuç raporunu oluşturmak
	Veriyi Tanımlama	Veri seçimi			
Kaynakların stokları, gereksinimler, varsayımlar ve kısıtlar, Riskler ve yükümlülükler, terminoloji, maliyetler ve faydalar hakkında durum değerlendirme	Veri açıklama raporu	Ekleme çıkarma gerekçesi	Test tasarımı oluşturma	Onaylanmış modeller	Projeyi gözden geçirmek
					Deneyim dökümantasyonu oluşturmak
Veri madenciliği amaçlarını belirleme	Veri keşfi	Veri temizleme	Yapı modeli için parametre ayarı	Gözden geçirme süreci	
Veri madenciliği başarı kriterlerini belirleme	Veri arama raporu	Veri temizleme raporu	Model açıklamaları		
				Sonraki adımları belirlemek	
Proje planı oluşturmak	Veri kalitesini doğrulama	Veri yapısı	Değerlendirme modeli	Olası adımlar listesi	
Proje planı için araç ve tekniklerin ilk keşfi	Veri kalitesi raporu	Türetilmiş öznitelikler, oluşturulmuş kayıtlar	Düzenlenmiş parametre ayarı	Karar	

3.2. Uygulama Süreci

Bu çalışmada, bir üretim işletmesinde uygunsuz olarak görülen ve redlenen ürünler üzerinde bir analiz yapılmıştır. Bu analizde uygunsuz olarak görülen ve ayrılan malzemeler üzerine nelerin etkili olduğu ortaya konulmaya çalışılmıştır. Şekil 3.2 uygulama sürecinin adımlarını göstermektedir.



Şekil 3.2. Uygulama adımları

Uygulama sürecinde ilk olarak analiz yapılacak olan problemin girdi ve çıktıları tanımlanmıştır. Daha sonra problem için berirlenen girdi ve çıktıları destekleyecek veriler toplanmıştır. Toplanan veriler düzenlenip bir ön incelemeye tabi tutulmuştur. Ön inceleme sonucunda anlamlı görülmeyen veri setleri modelden ayrıştırılmış ve anlamlı ilişkiye sahip olan veri setleri ise modele dahil edilmek üzere hazırlanmış ve model oluşturulmuştur. Uygulama sürecinin son adımında ise model çıktıları yorumlanmıştır.

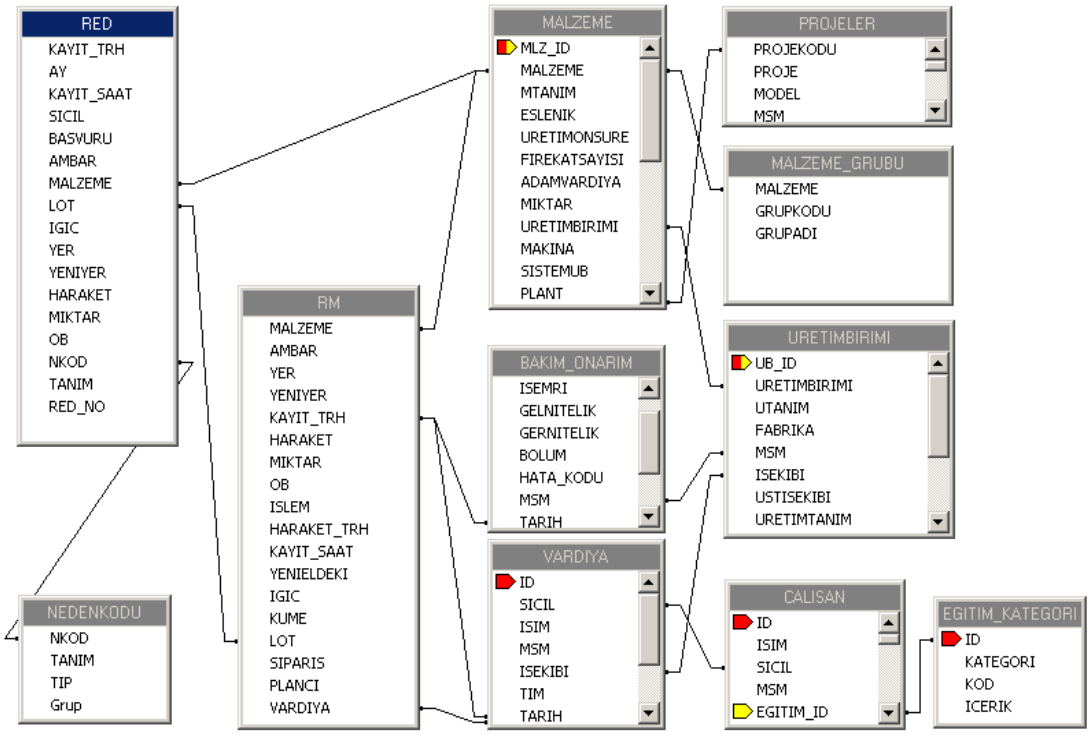
3.3. Uygulama Adımları

3.3.1. Problemin tanımlanması

Problemin tanımlanması aşamasında, ilk olarak analizde kullanılacak olan girdi ve çıktılar belirlenmiştir. Problem, uygunsuz olarak görülen ve redlenen malzemeler üzerinde nelerin ne kadar etkiye sahip olduğunun belirlenmesidir. Bu amaçla redlenme ile redlenme tarihi, redlenen ürünü üreten ekibin durumu, eğitimi, kadrolu ve taşeron çalışanlar, redlenen ürünün üretildiği tarih ve üretildiği vardiya, redlenme nedeni, redlenmenin insan ya da makineden mi kaynaklandığı, redlenen ürünün üretim sıklığı, redlenen ürünün grubu, redlenen ürünün seri üretim olup olmaması, redlenen ürünün firması arasındaki ilişkiler ortaya konulmaya çalışılmıştır.

3.3.2. Veri toplama

Tanımlanan problem için gerekli olan veriler işletmenin bilgisayar sistemleri ile elektronik ortamda toplanmıştır. Veriler red bilgileri, üretim bilgileri, çalışan bilgileri, malzeme bilgileri, proje ve firma bilgileri olmak üzere 6 kategoriye ayrılmış ve toplanan tüm veriler bir Microsoft Access Database (MDB) dosyası içerisinde tablolanmıştır. Oluşturulan veri setindeki bu veriler dikkate alınarak model kurulmuştur.



Şekil 3.3. Veri tabanındaki tablolar ve ilişkiler

MDB dosya formatında oluşturulan veri seti ile Clementine programına “Sources” paleti üzerindeki “Database” nodu vasıtası ile bağlantı sağlanmıştır. Veriler 11 ayrı tabloda bulunduğu için yazılan SQL cümleleri ile tek bir tabloda birleştirilmiştir. Kullanılan SQL cümleleri:

Tablo 3.2. Veri tabanı 1`e bağlantı SQL cümlesi

```
SELECT DISTINCT
RM.MALZEME,RM.LOT,RM.MIKTAR AS RMMIKTAR,RM.VARDIYA,RM.KAYIT_TRH AS TARIH,RM.HAFTA,
RED.KAYIT_TRH AS REDTRH,RED.MIKTAR AS REDMIKTAR,
NEDENKODU.NKOD,TRIM(NEDENKODU.TANIM) AS REV_TANIM,TRIM(NEDENKODU.GRUP) AS REV_GRUP,
MALZEME.URETIMONSURE,
PROJELER.PROJE,
URETIMBIRIMI.FABRIKA,URETIMBIRIMI.MSM,URETIMBIRIMI.ISEKIBI,URETIMBIRIMI.URETIMBIRIMI,
TRIM(MALZEME_GRUBU.GRUPADI) AS REV_GRUPADI,
MUSTERI_REF.MUSTERI
FROM
((((((RM RM INNER JOIN MALZEME MALZEME ON RM.MALZEME=MALZEME.MALZEME)
INNER JOIN MALZEME_GRUBU MALZEME_GRUBU ON MALZEME_GRUBU.MALZEME=MALZEME.MALZEME)
INNER JOIN PROJELER PROJELER ON MALZEME.PROJEKODU=PROJELER.PROJEKODU)
INNER JOIN MUSTERI_REF MUSTERI_REF ON MUSTERI_REF.MALZEME=MALZEME.MALZEME)
INNER JOIN URETIMBIRIMI URETIMBIRIMI ON MALZEME.URETIMBIRIMI=URETIMBIRIMI.URETIMBIRIMI)
LEFT OUTER JOIN RED RED ON RM.LOT=RED.LOT AND RM.MALZEME=RED.MALZEME)
LEFT OUTER JOIN NEDENKODU NEDENKODU ON RED.NKOD=NEDENKODU.NKOD
WHERE RM.KAYIT_TRH > #2008-01-01# AND RM.MALZEME LIKE '0%' AND RM.ISLEM<>'9900'
ORDER BY RM.KAYIT_TRH,RM.MALZEME
```

Tablo 3.3. Veri tabanı 2`ye bağlantı SQL cümlesi

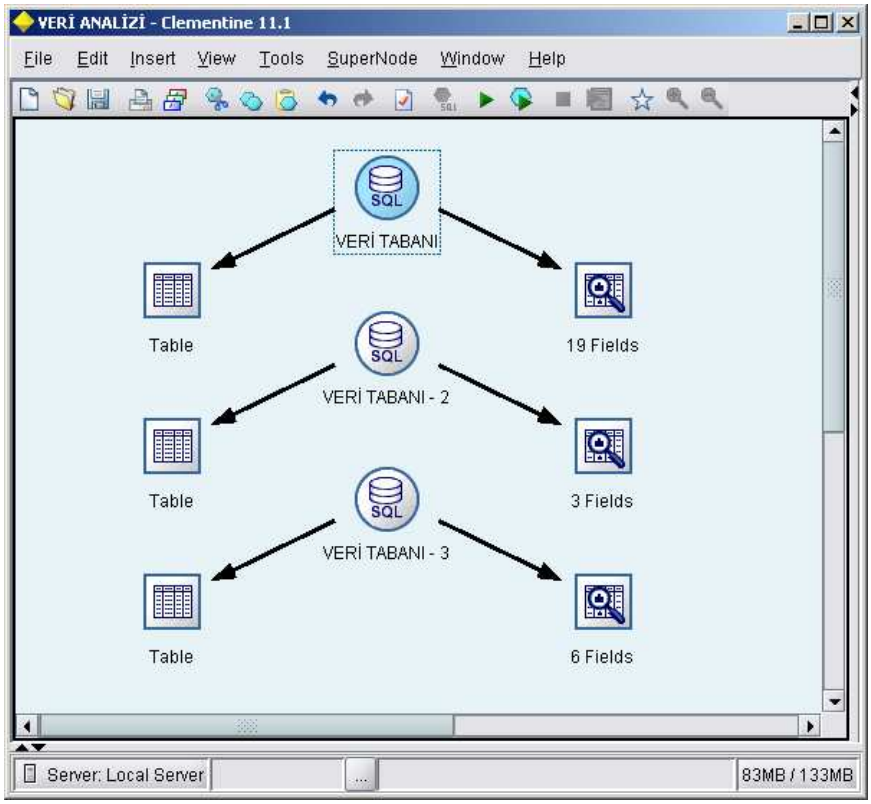
```
SELECT DISTINCT
ISEMRI,BAKIM_ONARIM.MSM,BAKIM_ONARIM.TARIH
FROM
BAKIM_ONARIM
```

Tablo 3.4. Veri tabanı 3`e bağlantı SQL cümlesi

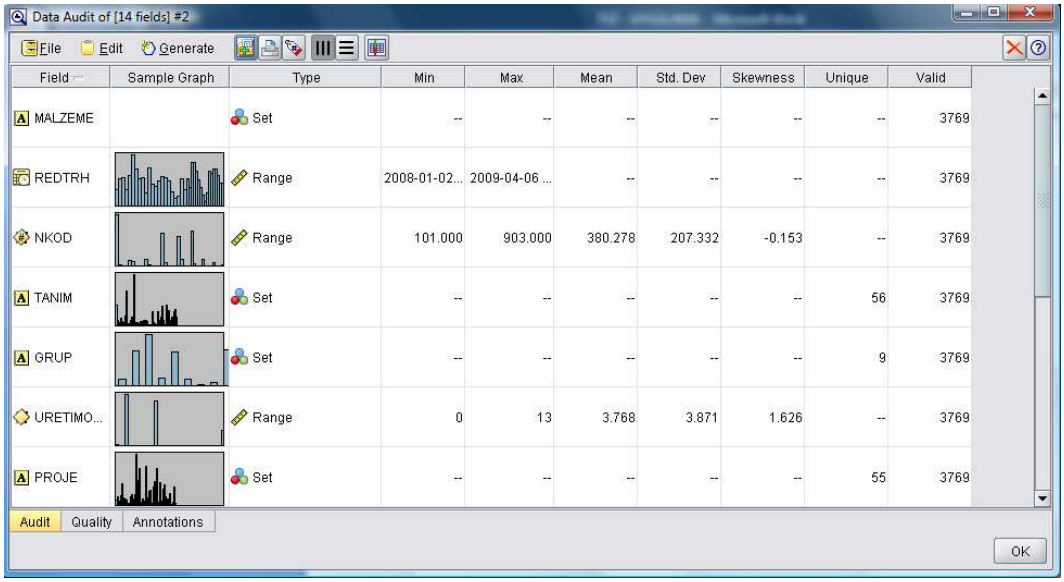
```
SELECT
ISEKIBI,VARDIYA.HAFTA,VARDIYA.VARDIYA,
CALISAND,EGITIM, VARDIYA.SICIL
FROM
VARDIYA VARDIYA INNER JOIN CALISAN CALISAN ON VARDIYA.SICIL=CALISAN.SICIL
```

3.3.3. Veri kalitesinin incelenmesi

Veri seti programa tanıtıldıktan sonra, verilerde bir sapma, anormal bir değer olup olmadığının tespiti için öncelikle veri kalitesi incelenmiştir. Şekil 3.4`de Clementinede veri kalitesinin incelenmesi için yapılmış node bağlantıları görülmektedir. Şekil 3.5`de oluşturulan akımın çalıştırılması sonucu oluşan sonuç tablosu görülmektedir.



Şekil 3.4. Veri kalitesinin incelenmesi

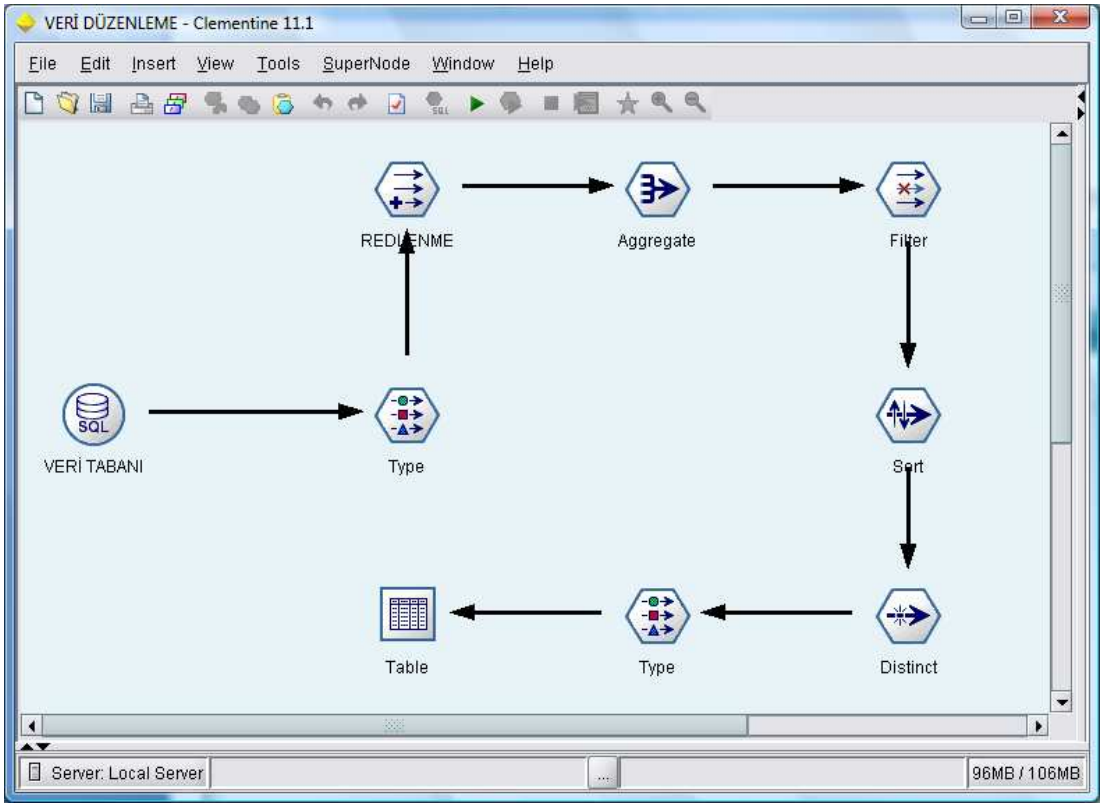


Şekil 3.5. Veri kalitesi incelenme sonuçları

İnceleme sonrasında herhangi bir sapma ve kayıp değer rastlanılmamıştır.

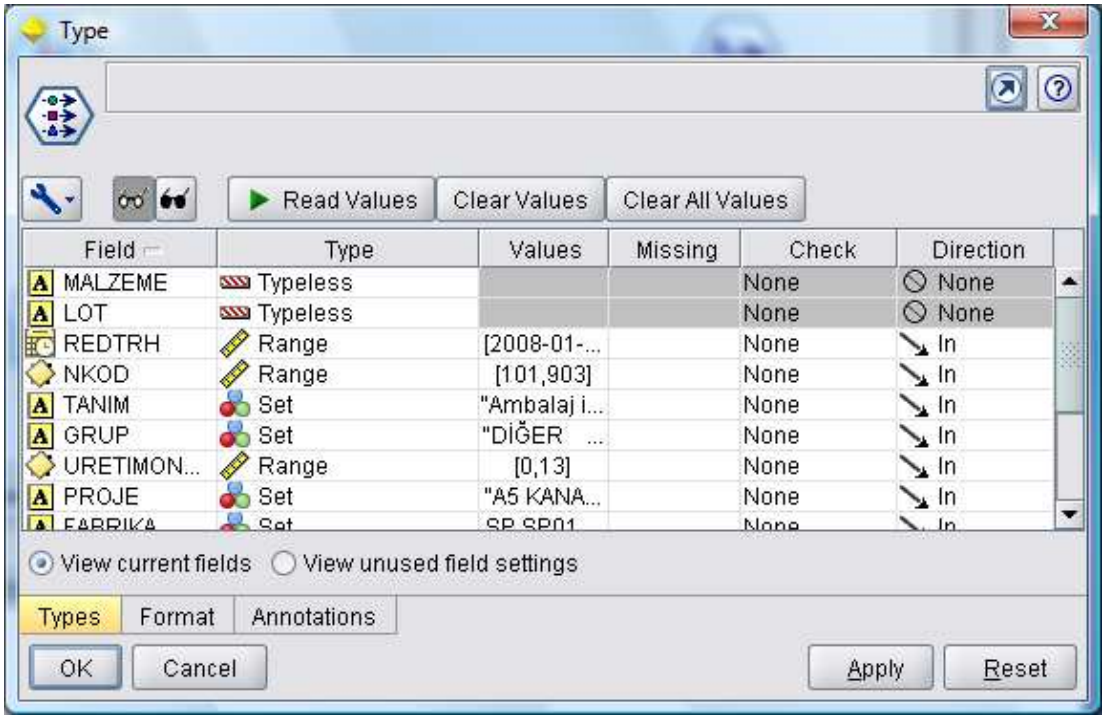
3.3.4. Veri düzenleme

Veri kalitesi incelendikten sonra modelde kullanılmak üzere veriler düzenlenmiştir. Şekil 3.6 oluşturulan veri düzenleme akımını göstermektedir.



Şekil 3.6. Clementinede oluşturulan veri düzenleme ekranı

Veri düzenleme akımında Şekil 3.7'de fonksiyonları gösterilen "Type" nodu ile değişkenler tanımlanmıştır. Kurulacak modelde ihtiyaç duyulan değişkenler seçilmiş, modelde yer almayacak olan değişkenler pasif konuma getirilmiştir.

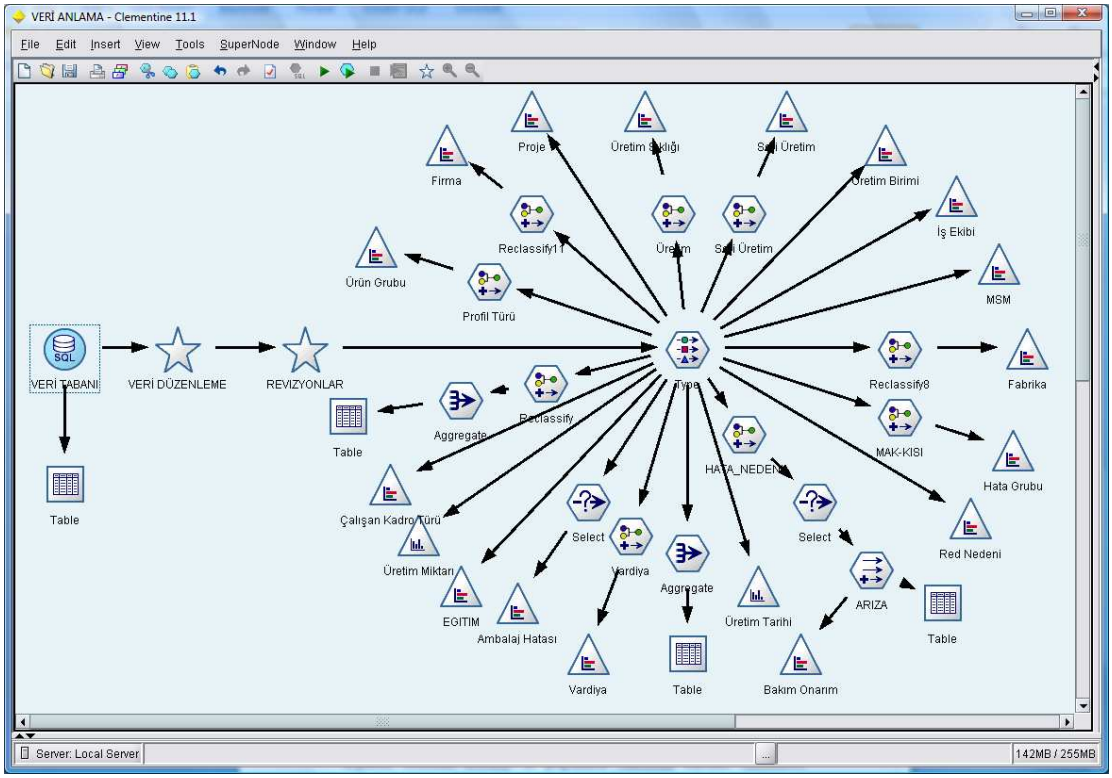


Şekil 3.7. Type nodu ile veri düzenleme ekranı

"Field Options" paletindeki "Derive" nodu ile redlenme işlemi = 1, uygun = 0 olarak tanımlanmış ve yeni bir alan olarak veri setimize eklenmiştir. "Aggregate" nodu ile de tanımlamış olduğumuz redlenme ve uygun görülme işlemi için toplam ve ortalama değerleri hesaplanmıştır. "Filter" nodu ile ihtiyaç duyulmayan değişkenler temizlenmiştir. "Sort" nodu ile veri seti sıralanmıştır. "Distinct" nodu ile veri tekrarları engelenmiş ve son olarak veriler son kez düzenlemiştir. Bu işlemlerin ardından yeni bir düzenlenmiş tablo elde edilmiştir.

3.3.5. Veri ve ilişki anlama

Veri ve ilişki anlama aşamasında, veri hazırlama için yapılan işlemler bir "Supernode" içerisinde toplanmış ve veri tabanına bağlanmıştır. "Type" nodu ile değişkenler yeniden tanımlanıp, girdi ve çıktı değişkenleri belirlenmiştir. Karar değişkeni olarak redlenmenin seçildiği uygulamada herbir değişken ile karar değişkeni arasındaki ilişki, grafiksel ve istatistiksel olarak ortaya konulmuştur.



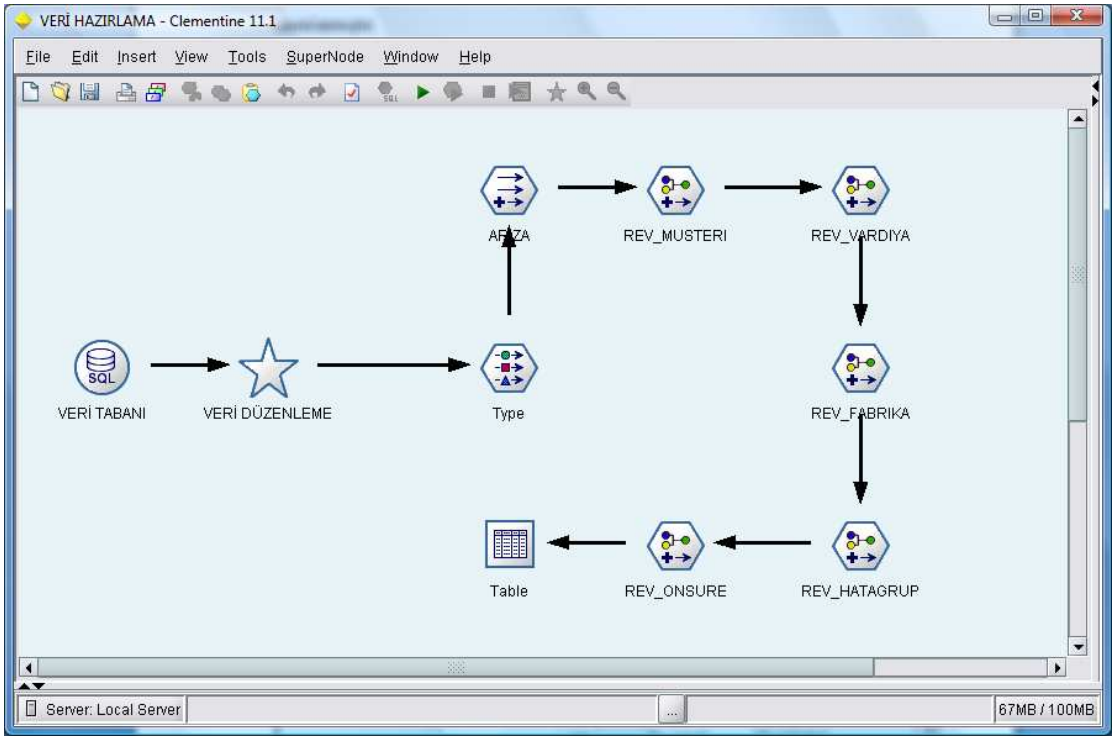
Şekil 3.8. Veri ve ilişki anlama aşamasında karar değişkeni ile ilişkiler

Bu aşamada elde edilen sonuç ve grafikler uygulamanın sonuç bölümünde ayrıntılı bir şekilde gösterilmiş ve yorumlanmıştır.

3.3.6. Veri hazırlama

Veri ve ilişki anlama aşamasından sonra verilerde tekar bir düzenlemeye gidilmesi gerektiği anlaşılmıştır.

Üretim önsürelerine göre üretilen malzemeler aylık üretilen, haftalık üretilen ve günlük üretilen malzemeler olmak üzere 3 ayrı sınıfda toplanmıştır. Üretilen ürünün uygunsuz olmasının nedenleri de gruplanarak makinadan kaynaklı redlenmeler, çalışandan kaynaklı redlenmeler ve makina-çalışandan kaynaklı redlenmeler olmak üzere 3 grupta toplanmıştır. Makinadan kaynaklı redlenmelerin incelenmesi aşamasında oluşabilecek makina arızaları için 1, arıza olmama durumu için ise 0 tanımlanmıştır.

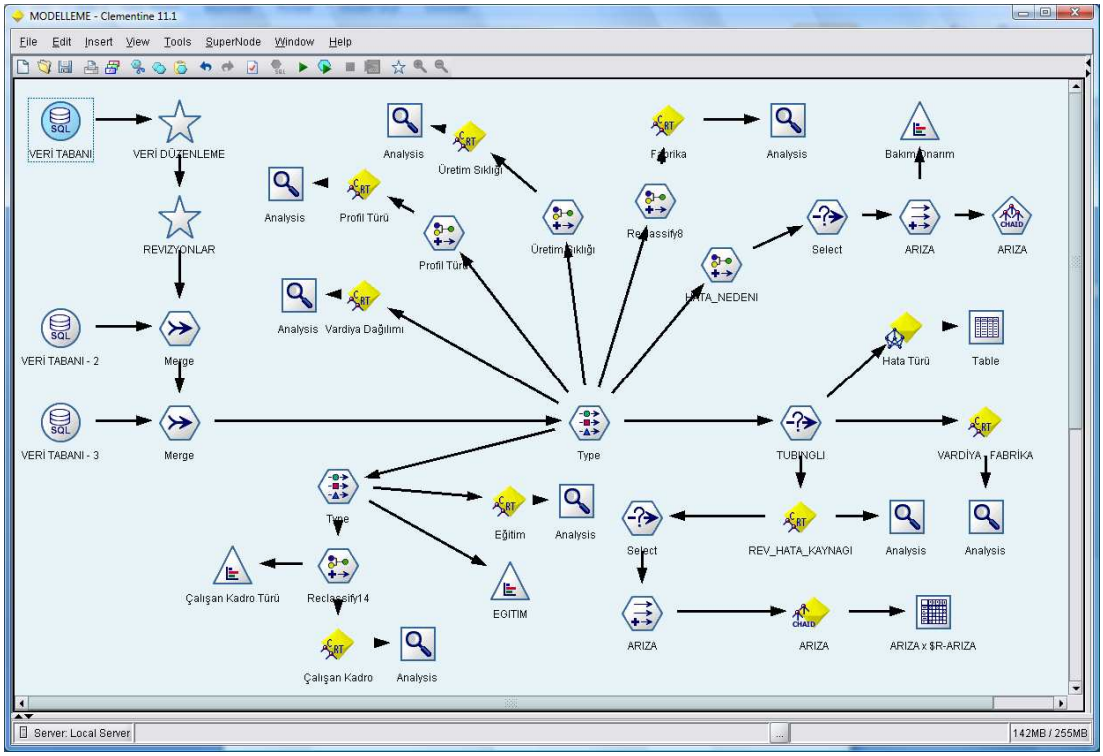


Şekil 3.9. Veri hazırlama ekranı

Müşteriler için, aynı firmaya ait fakat farklı plant olan malzemeler aynı müşteri olarak tanımlanmıştır. Üretim saatlerine göre vardiya tanımlamaları yapılmıştır. Saat 08:00 ile 16:00 arasındaki üretimler 1. vardiya (08:00-16:00), saat 16:00 ile 24:00 arasındaki üretimler 2. vardiya (16:00-24:00), saat 00:00 ile 08:00 arasındaki üretimler 3. vardiya (00:00-08:00) ve fazla mesaili çalışma olan 08:00 ile 20:00 saatleri arasında 4. vardiya (08:00-20:00) olarak tanımlanmıştır.

3.3.7. Model oluşturma

Analiz için toplanan verilerin düzenlenmesinin ardından, redlenme ile karar değişkenleri arasındaki tüm ilişkileri kapsayan model bu adımda oluşturulmuştur. Oluşturulan modelde karar ağaçları, yapay sinir ağları, dağılım grafikleri ve ilgili tablolar ile analiz gerçekleştirilmiştir.

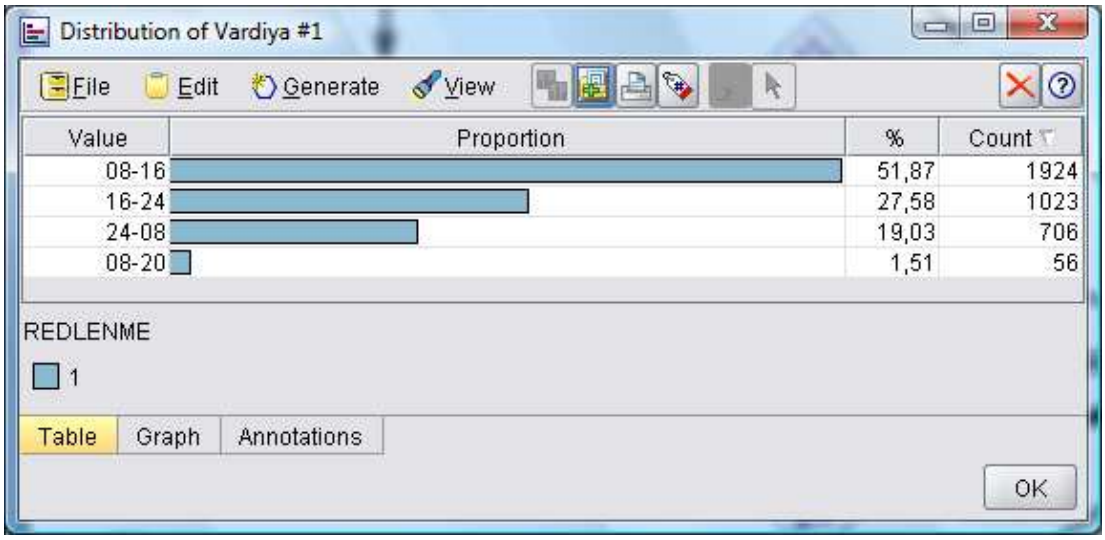


Şekil 3.10. Oluşturulan model

BÖLÜM 4. UYGULAMA SONUÇLARI

4.1. Karar Değişkenlerinin Modele Etkisi

1. Vardiya Düzeni : Vardiya düzeni ile üretilen ürünün redlenmesi arasındaki ilişki Şekil 4.1`de gösterilmektedir.



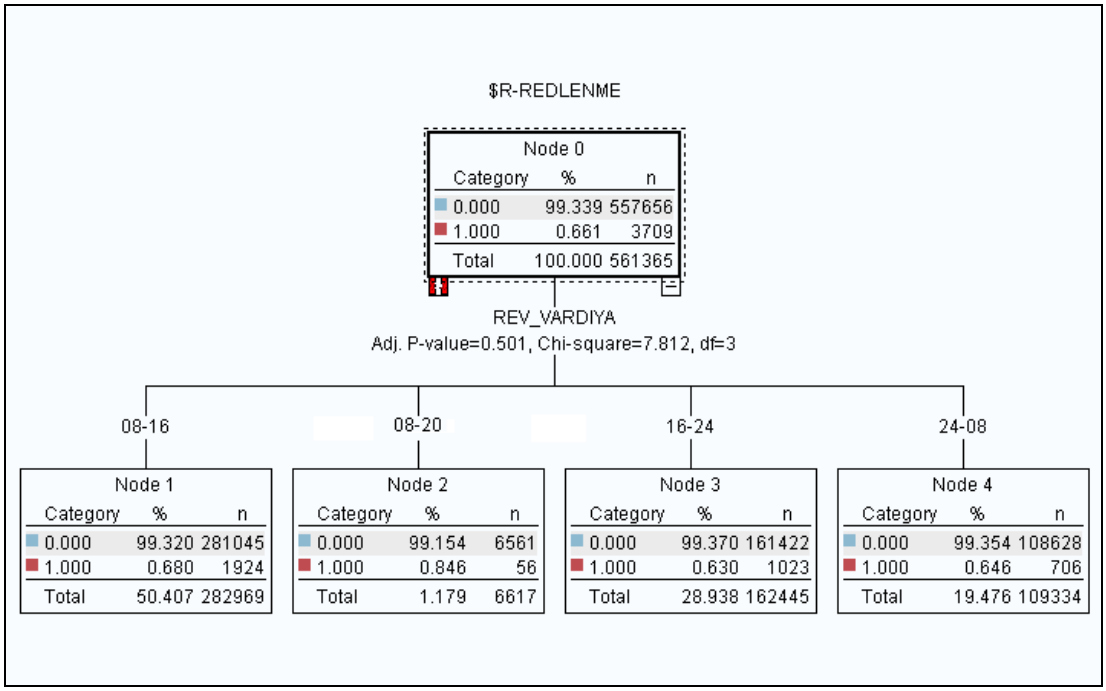
Şekil 4.1. Vardiya düzeni - redlenme ilişkisi

Model veri setindeki vardiya değişkenini anlamlı bulmuştur. 08-16 vardiyasındaki redlenme oranı toplam redlenmelerin % 51.87 `idir. 16-24 vardiyasında meydana gelen redlenme oranı % 27.58, 24-08 vardiyasında % 19.03 ve fazla mesaili çalışma olan 08-20 vardiyasındaki redlenme oranı ise % 1,51 olmuştur. Tablo 4.1`de vardiyalardaki toplam redlenme ve toplam üretim miktarlarına karşılık redlenme oranları verilmiştir.

Tablo 4.1. Vardiyaların redlenme - üretim oranı

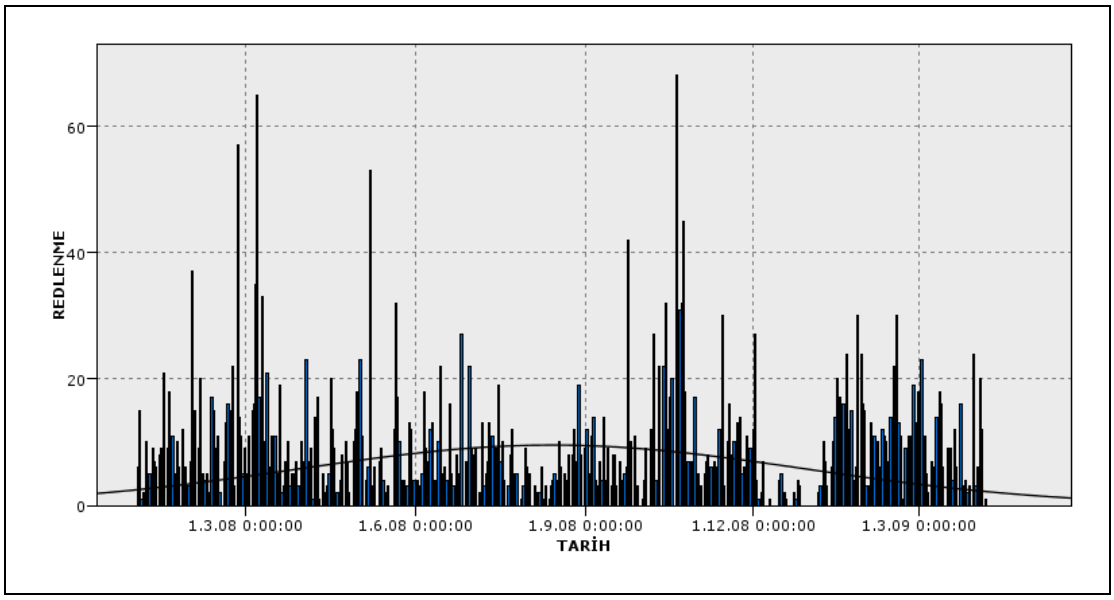
Vardiya	Toplam Redlenme	Üretim Sayısı	Redlenme/Üretim
20-08	0	40	0.0
08-20	56	6577	0.0085
24-08	706	109334	0.0064
08-16	1924	282969	0.0067
16-24	1023	162445	0.0063

Şekil 4.2`de vardiya ile redlenme arasındaki ilişkiyi gösteren karar ağacı verilmiştir. Redlenme sayısı en fazla 08-16 vardiyasında olmasına rağmen en düşük üretim yapılan vardiya olan fazla mesaili 08-20 vardiyasında redlenme oranı % 0.85 ile en yüksek değerdir. 08-16 vardiyası 08-20 vardiyasından sonra % 0.68 ile en yüksek redlenme oranına sahiptir.



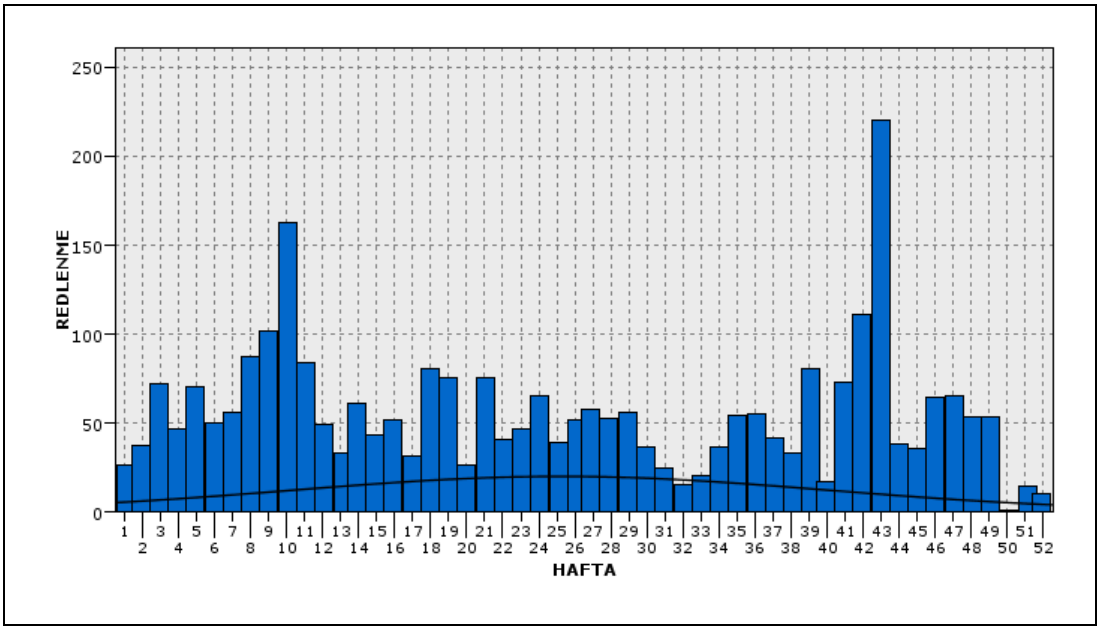
Şekil 4.2. Vardiya - redlenme ilişkisi karar ağacı

2. Üretim Periyodu : Üretimin yapıldığı periyodun redlenme üzerindeki etkisi incelendiğinde Şekil 4.3`deki histogram elde edilmiştir. Grafikten de görüldüğü üzere yılın ilk çeyreğinde ve son çeyreğinde redlenme oranlarında yükselme olmaktadır.



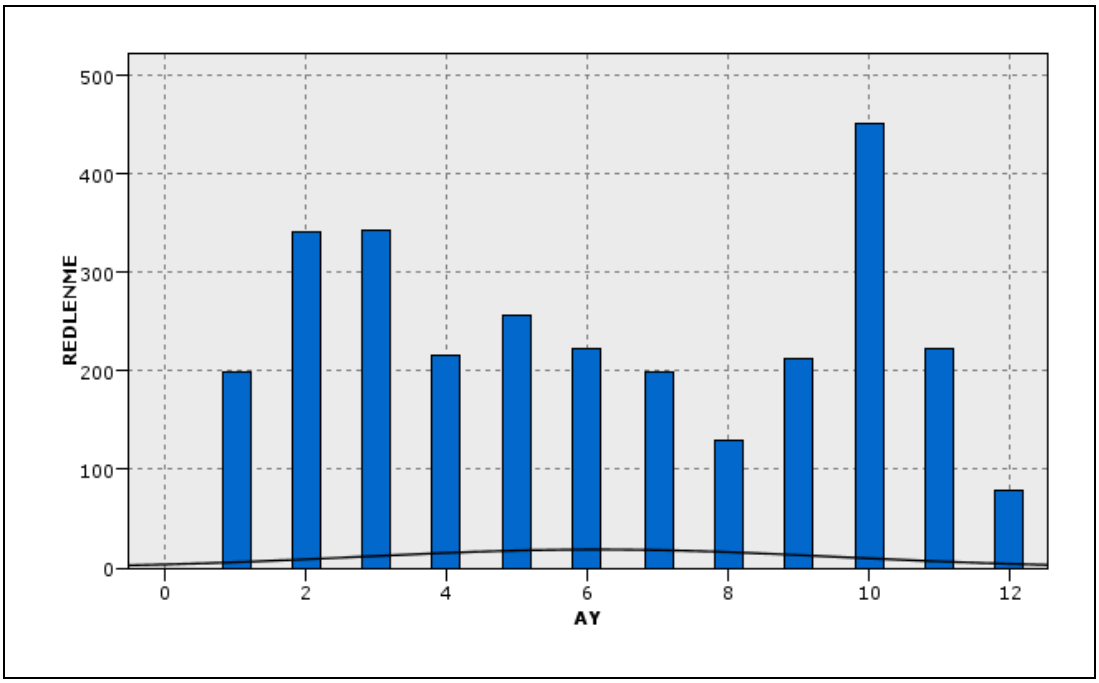
Şekil 4.3. Üretim periyodu - redlenme ilişkisi grafiği

Redlenen ürünlerin üretiminin yapıldığı haftaların dağılımı Şekil 4.4`de görülmektedir. Grafiktende görüldüğü üzere yılın ilk çeyreğinde 8. hafta ile 11. hafta arasında üretilen ürünlerde redlenme miktarı artmıştır. Aynı şekilde yılın son çeyreğinde 42. hafta ve 43. haftada redlenme oranları oldukça yüksektir.



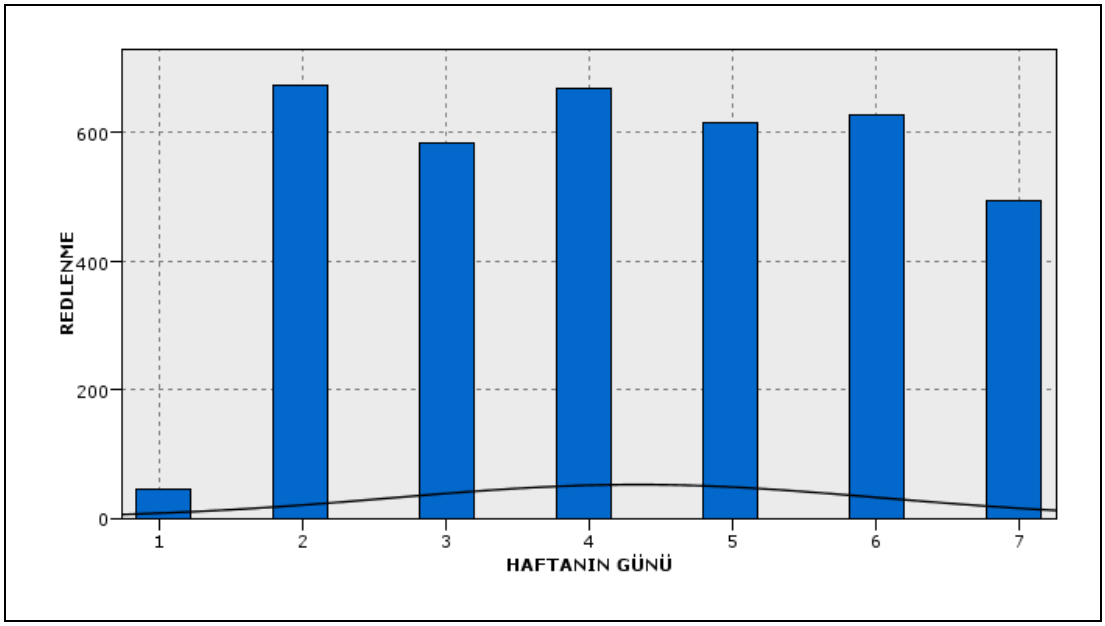
Şekil 4.4. Üretim haftası - redlenme ilişkisi grafiği

Şekil 4.5 redlenen ürünlerin üretimlerinin aylara dağılımını göstermektedir. 2. ve 3. aylarda yapılan üretimler ile 10. ayda yapılan üretimlerde redlenme oranları diğer aylara göre daha yüksektir. 10. aydaki redlenmeler yıl içerisindeki en yüksek redlenme oranıdır.



Şekil 4.5. Üretim ayı - redlenme ilişkisi grafiği

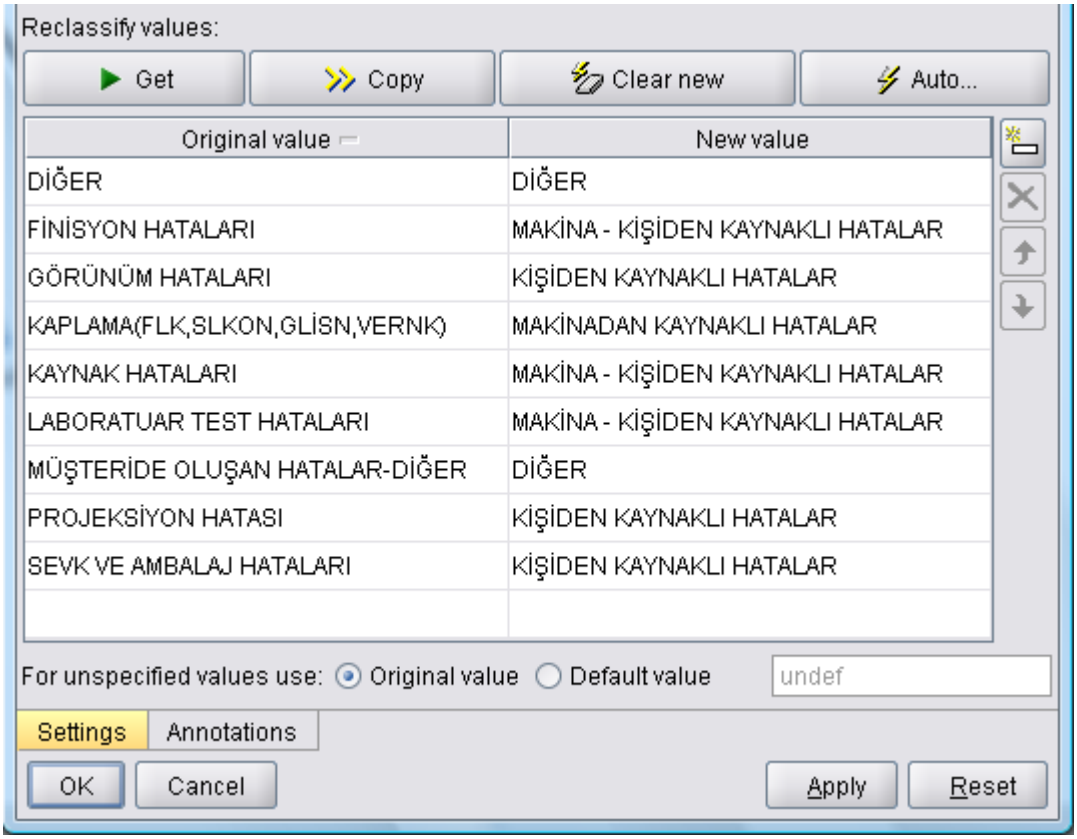
Şekil 4.6 redlenen ürünlerin üretildiği haftanın gününün redlenme miktarları ile ilişkisini göstermektedir. Grafikte 1 pazarı, 7 cumartesiyi ifade etmektedir.



Şekil 4.6. Üretim günü - redlenme ilişkisi grafiği

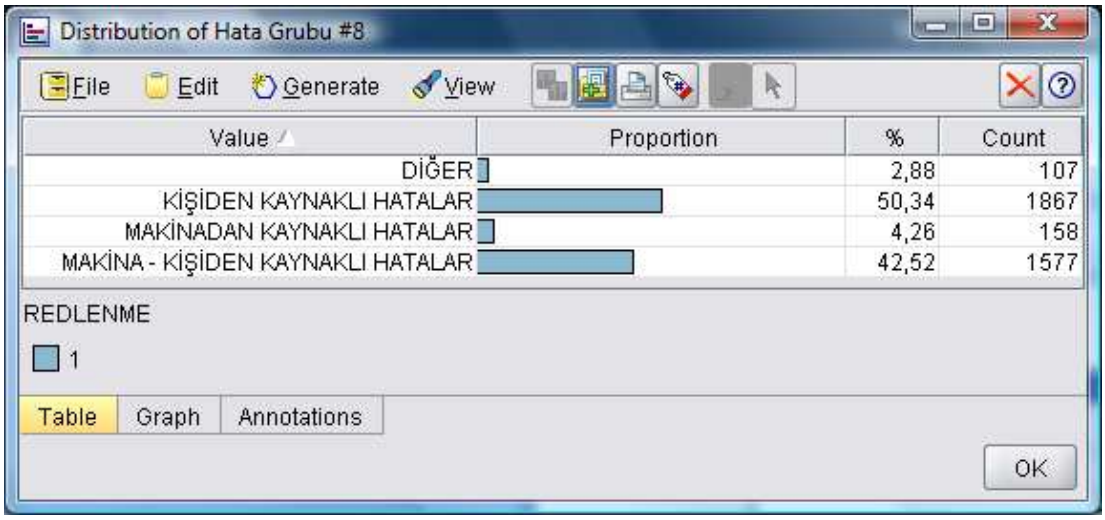
Grafiktende görüldüğü gibi redlenme en çok pazartesi günleri üretilen ürünlerde görülmektedir. Normalde çalışma olmayan pazar günü çok az miktarda fazla mesai yapıldığından çok düşük oranda redlenme olmuştur. Cumartesi ise haftanın normal mesaili çalışmasında en düşük redlenme oranına sahip günüdür.

3. Çalışanlar : Çalışanların redlenen ürün üzerindeki etkisini incelemek için hata nedenleri alt bir grupta toplanmış ve bu alt grupta Şekil 4.7`de görüldüğü üzere kişiden kaynaklı hatalar, makineden kaynaklı hatalar, kişi ve makineden kaynaklı hatalar ve diğer hatalar olmak üzere 4 ana grupta toplanmıştır.



Şekil 4.7. Hataların gruplanması

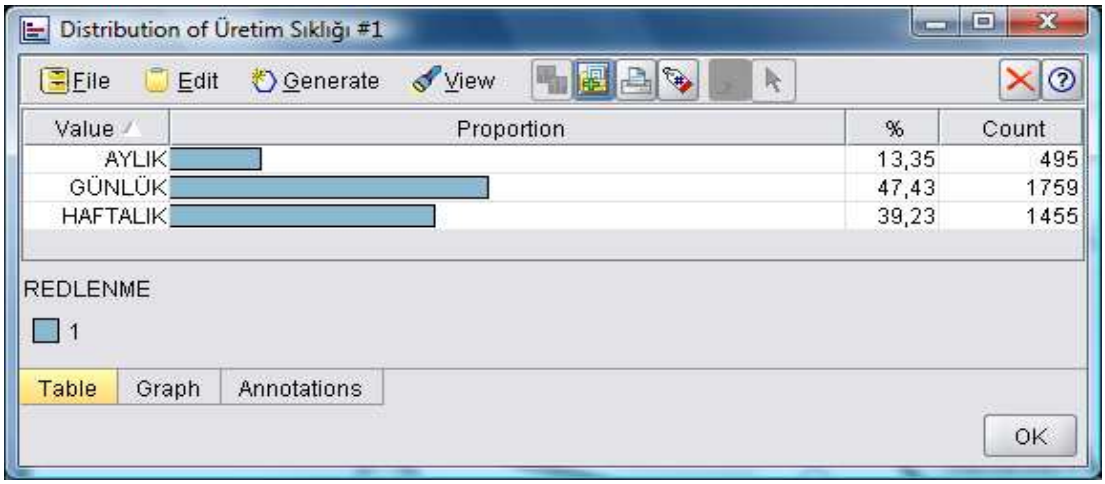
Model çalıştırıldığında bu değişkenler anlamlı bulunmuştur. Elde edilen sonuçlar Şekil 4.8`de gösterilmektedir.



Şekil 4.8 Hata grupları - redlenme ilişkisi

Üretilen ürünlerin uygunsuz olmasında sadece çalışan hataları % 50.34, makineden kaynaklı hatalar % 4.26 ve makine-çalışandan kaynaklı hatalar % 42.52 düzeyindedir. Makine-çalışandan kaynaklı hatalar üzerindeki çalışanın etkisi ikinci bir analiz ile ortaya konacaktır.

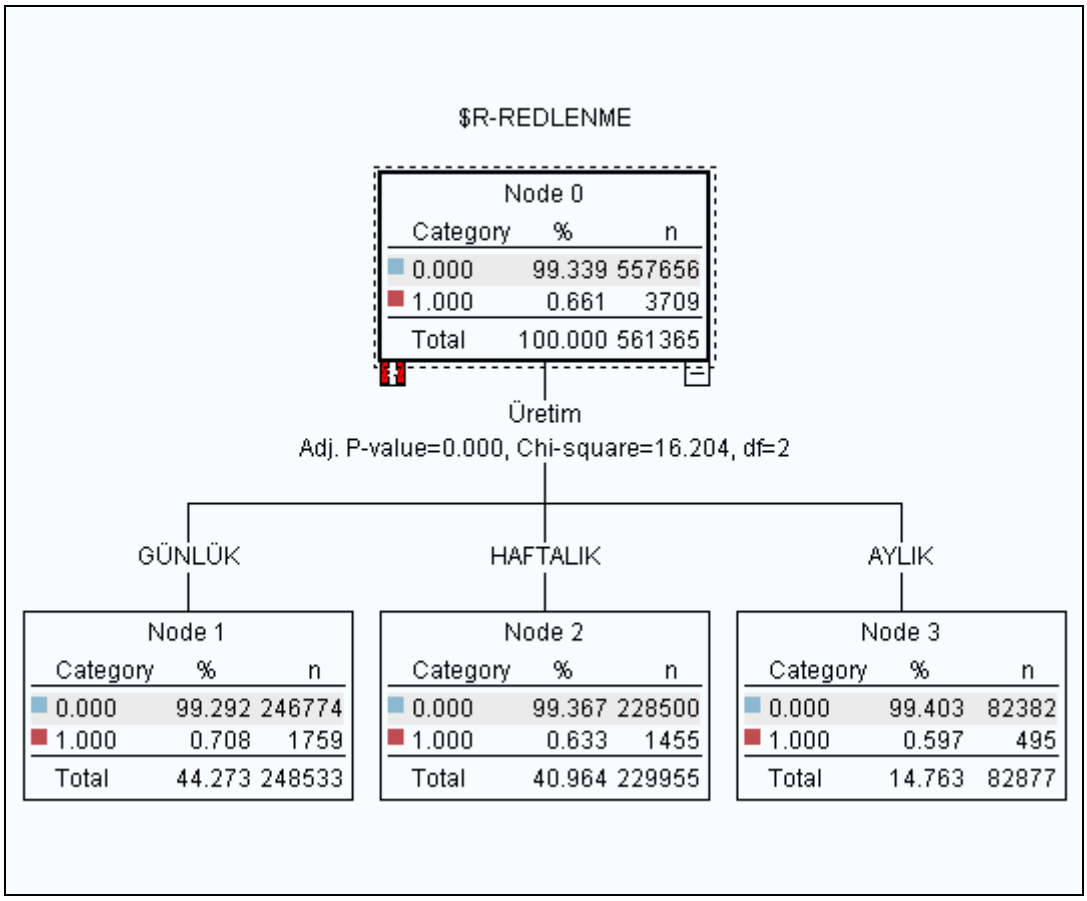
4. Üretim Sıklığı : İşletmede ürünler, üretim sıklığı olarak 3 gruba dağılmaktadır. Günlük olarak üretimi yapılan malzemeler, haftalık üretimi yapılan malzemeler ve aylık üretimi yapılan malzemeler olarak sınıflandırılmaktadır. Bu sınıflama, gelecek siparişlere bakılarak malzemelerin sipariş sıklığına göre bir formülasyon yapılarak belirlenen üretim ön sürelerinden ortaya çıkmaktadır. Şekil 4.9`da üretim sıklığının redlenme ile ilişkisi gösterilmektedir.



Şekil 4.9. Üretim sıklığı - redlenme ilişkisi

Günlük üretimi yapılan malzemelerin redlenme oranı, toplam redlenen malzemelerin % 47,49`u, haftalık üretimi yapılan malzemelerin redlenme oranı % 39,2`si ve aylık üretimi yapılan malzemelerin redlenme oranı ise toplam redlenen malzemelerin % 13,32 sidir.

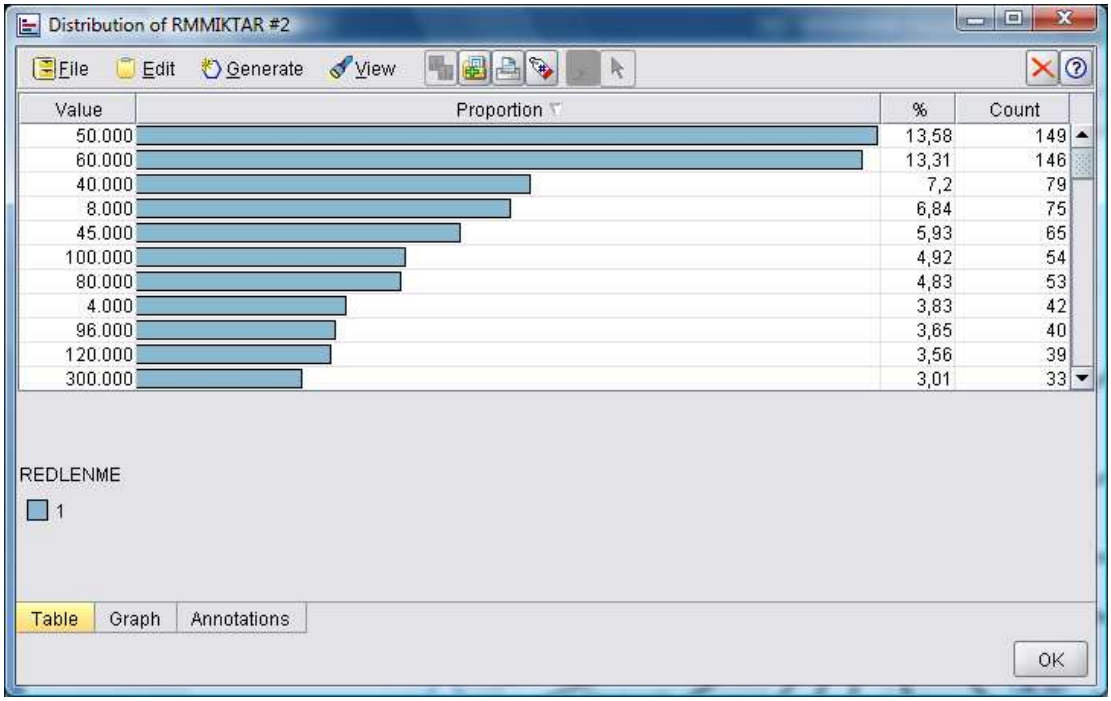
Üretim sıklığının redlenme üzerindeki etkisini gösteren karar ağacı Şekil 4.10`da verilmiştir.



Şekil 4.10. Üretim sıklığı - redlenme ilişkisi karar ağacı

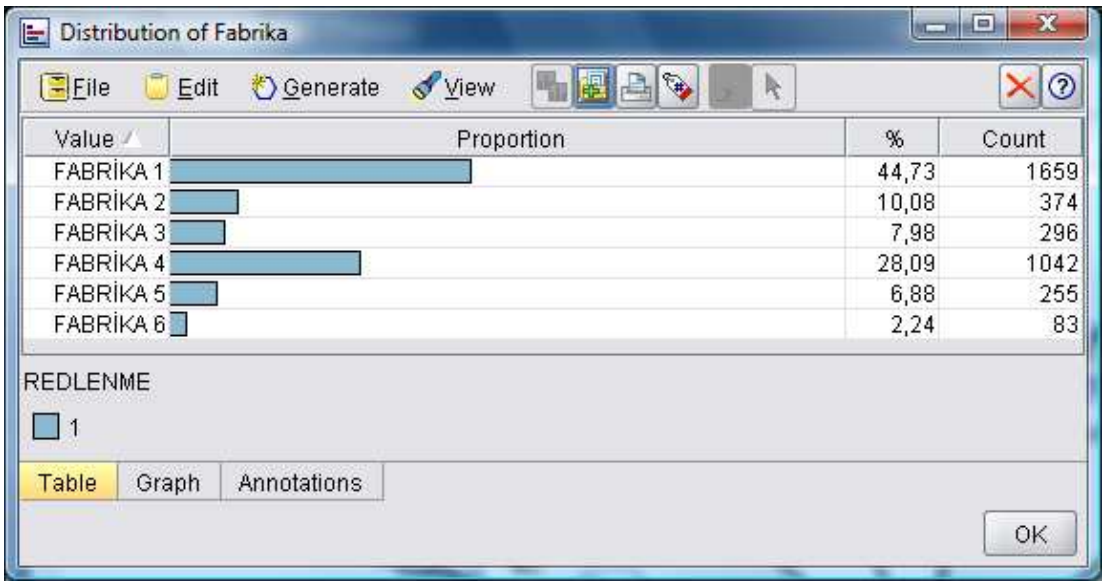
En fazla redlenme % 0.71 ile günlük olarak üretilen malzemelerde meydana gelmiştir. Haftalık olarak üretimi yapılan malzemelerde % 0.63 ve aylık olarak üretimi yapılan malzemelerde ise % 0.60 oranında redlenme meydana gelmiştir.

5. Ambalaj İçine Konulan Miktar : Görünüm hataları nedeni ile redlenen malzemeler üzerinde ambalaja konulan malzeme sayısının etkisi incelendiğinde karar sürecini etkileyen bir değişken olmadığı ve anlamlı bir ilişki oluşturmadığı tespit edilmiştir.



Şekil 4.11. Ambalaj içindeki miktar - redlenme ilişkisi

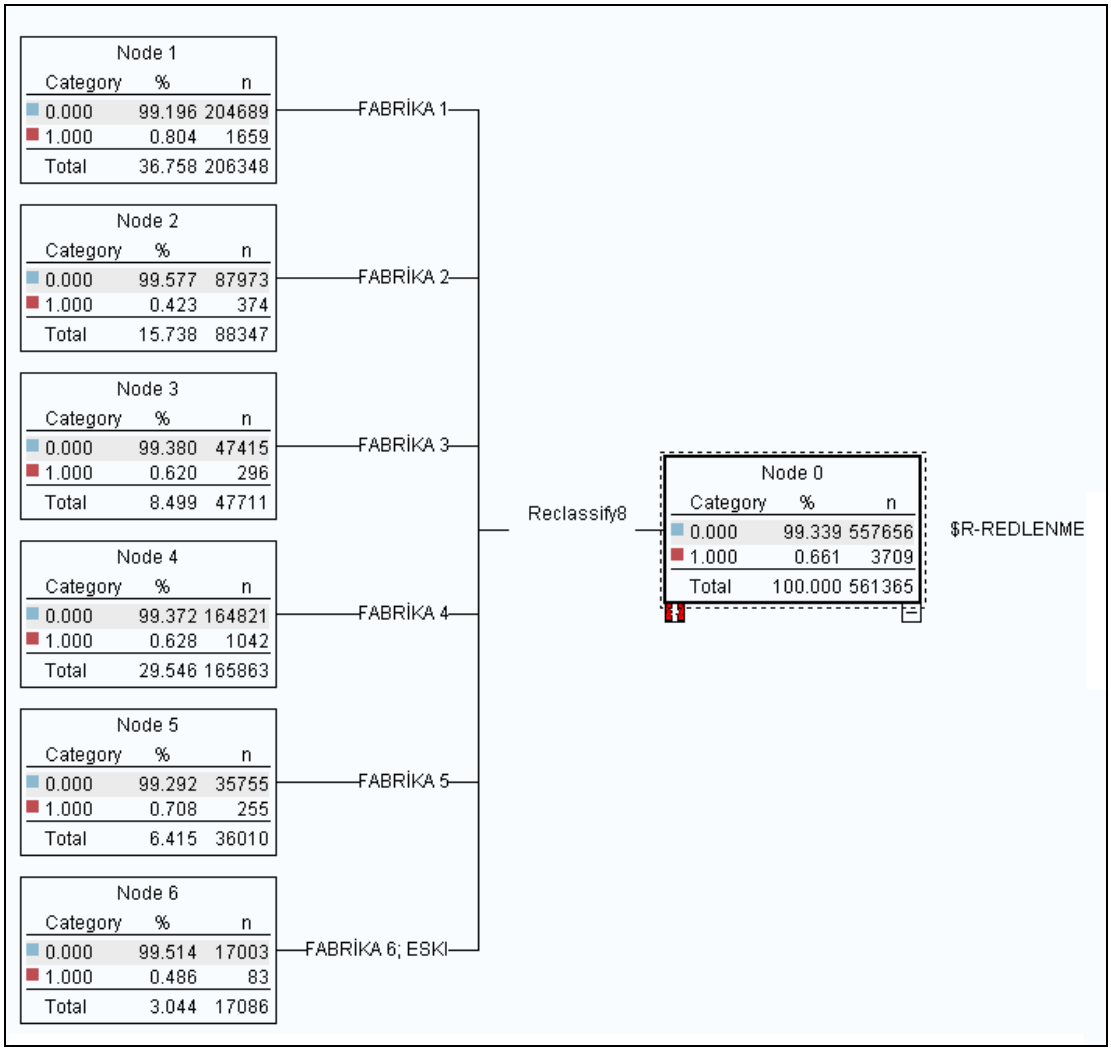
6. Fabrika : İşletme kendi içerisinde 6 adet fabrikaya bölünmüştür. Bu fabrikalarda meydana gelen redlenme miktarları Şekil 4.12`de verilmiştir.



Şekil 4.12. Üretim fabrikası - redlenme ilişkisi

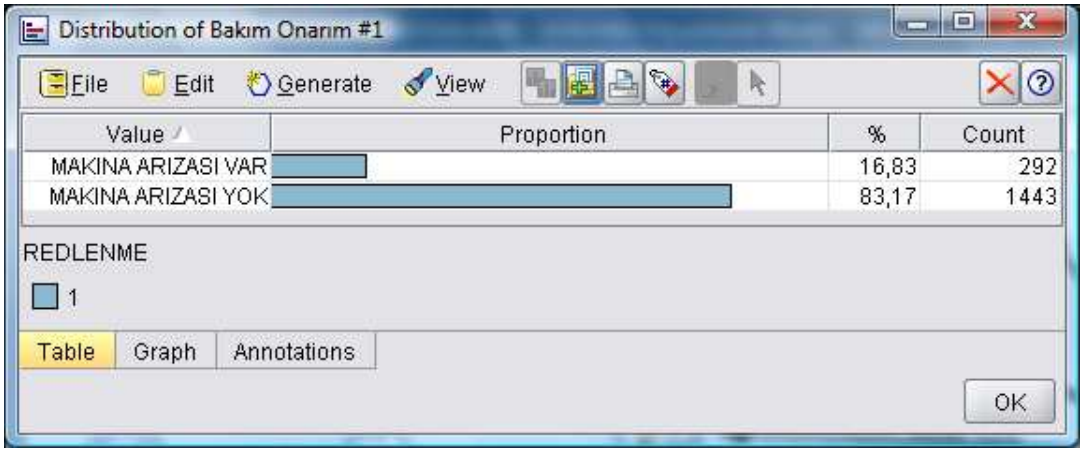
Şekil 4.13`de fabrikaların redlenme üzerine etkisini gösteren karar ağacı verilmiştir. Redlenme oranının en yüksek olduğu fabrika % 0.80 redlenme oranı ile Fabrika 1

olmuştur. Fabrika 5`de ise % 0.71 oranında redlenme meydana gelmiştir. Fabrika 2, % 0.42 ile en düşük redlenme oranına sahip fabrika olmuştur.



Şekil 4.13. Üretim fabrikası - redlenme ilişkisi karar ağacı

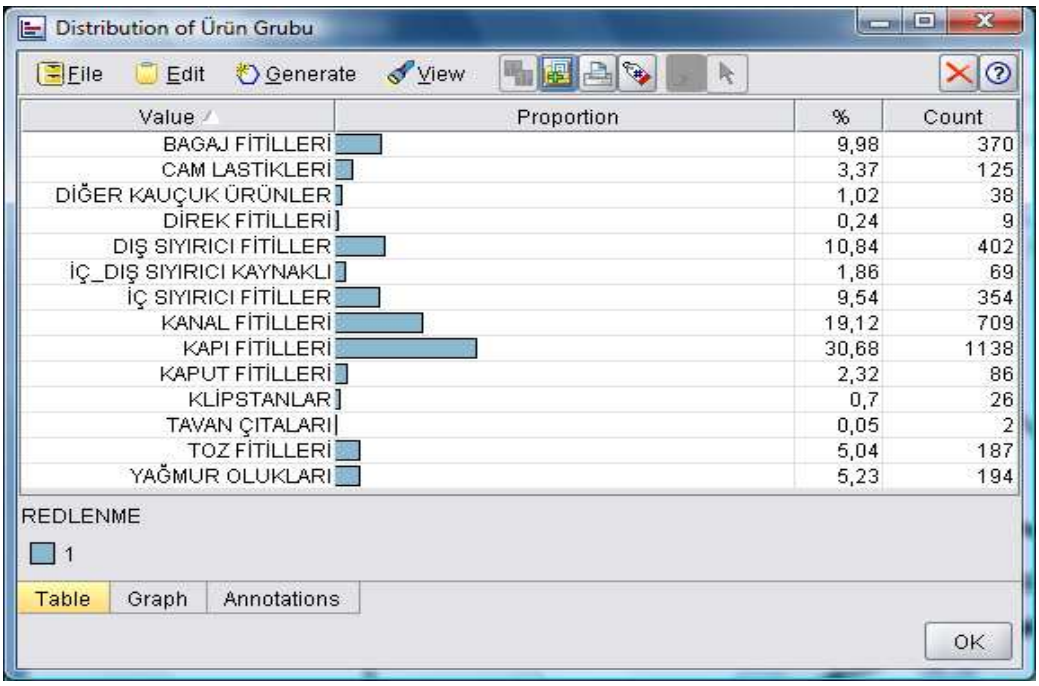
7. Makine Arızası : Makine hatası veya makine-insan hatası neden ile redlenen ürünler ile makine arızası arasındaki ilişki Şekil 4.14`de görülmektedir. Çalışanın redlenme üzerindeki etkisi incelenirken makine-insan hatası oranı, toplam redlenmelerin % 43`ü olduğunu tespit edilmişti. Sadece makine hatası ise toplam redlenmelerin % 4`ü dür.



Şekil 4.14. Makina Arızası - Redlenme İlişkisi Grafiği

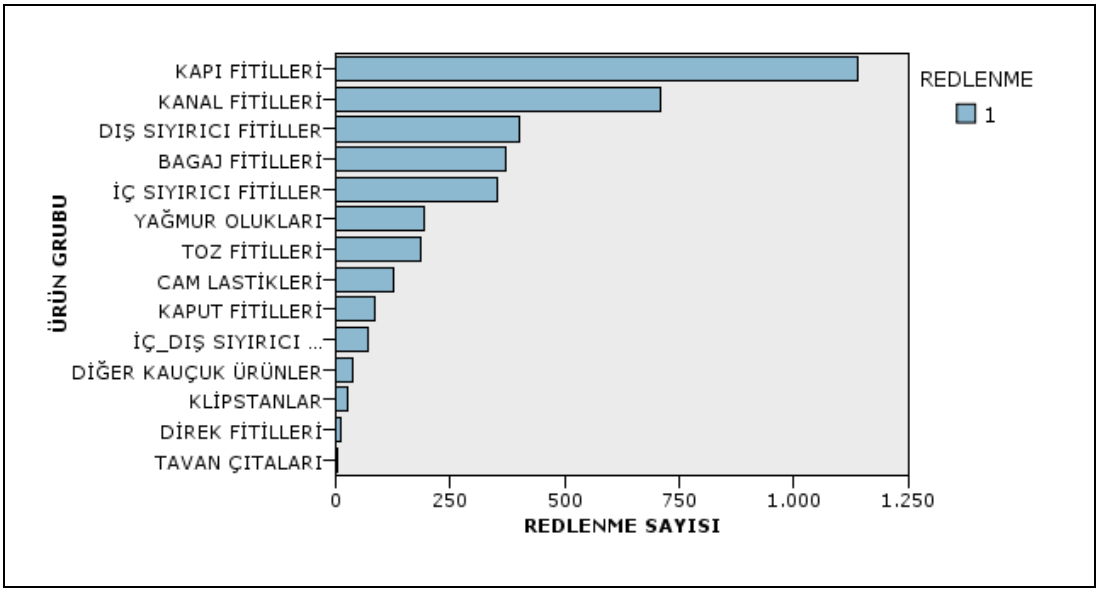
Makine hatası ve makine-insan hatası olarak grupladığımız nedenler yüzünden redlenen ürünün üretildiği üretim biriminde, üretim esnasında operasyonların gerçekleştirildiği makinalardaki arızanın redlenme üzerine etkisi incelendiğinde, makine hatası ve makine-insan hatası nedeni ile redlenmelerin % 16.83`ünde makine arızası mevcut olduğu görülmektedir.

8. Ürün Grubu : İşletmede üretilen ürünler 13 gruba ayrılmaktadır. Şekil 4.15 ürün grupları ile redlenme arasındaki ilişkiyi göstermektedir.



Şekil 4.15. Ürün Grubu - Redlenme İlişkisi

İşletmede üretilen ürünler içerisinde redlenen ürünlerin % 30,68`i kapı fitilleri, % 19,12`si kapı camı kanal fitilleri, % 10,84`ü ise dış sıyırıcı fitilleridir.

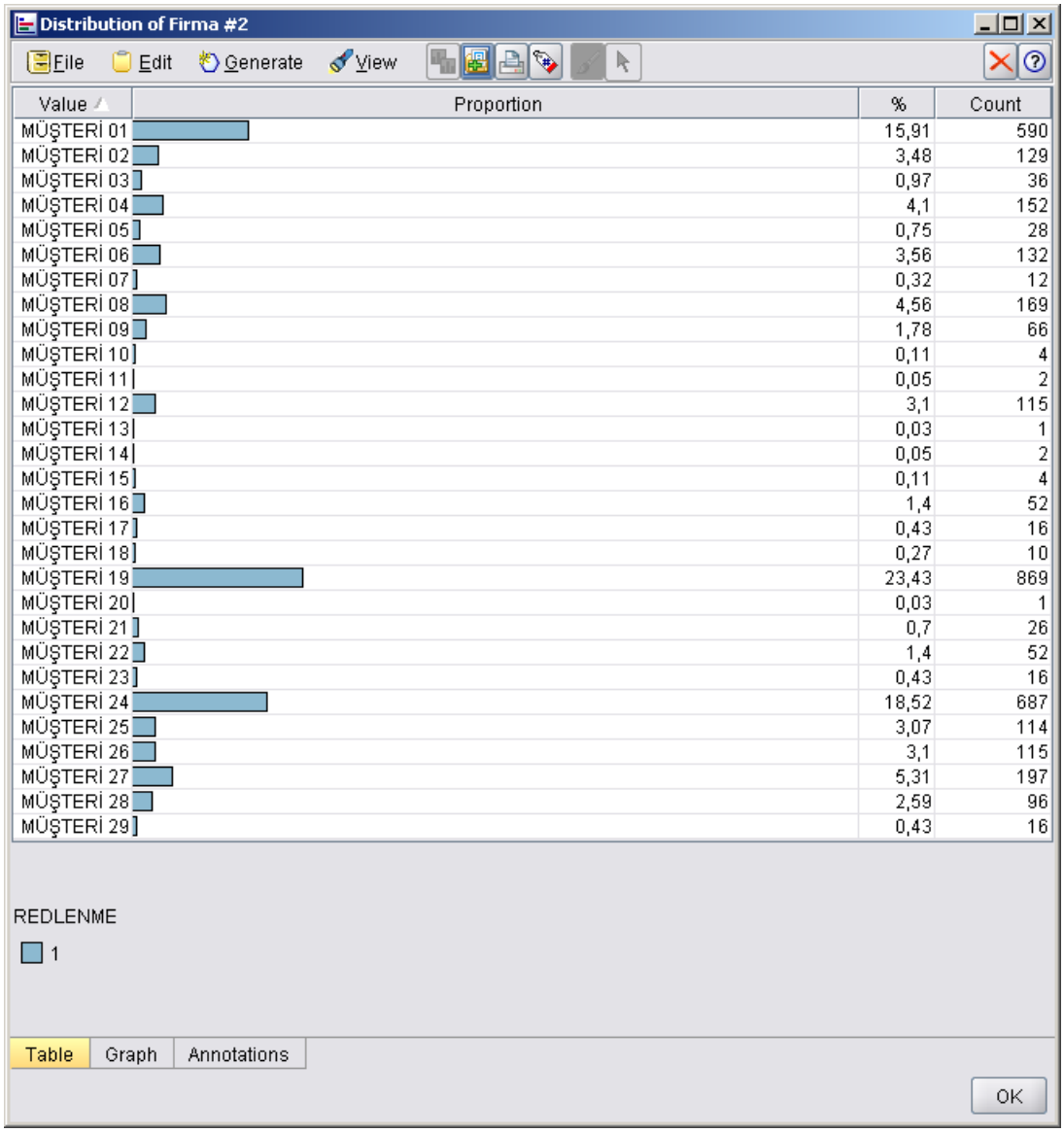


Şekil 4.16. Ürün grubu - redlenme ilişkisi grafiği

Tablo 4.2. Her bir ürün grubu için redlenme - üretim Oranı

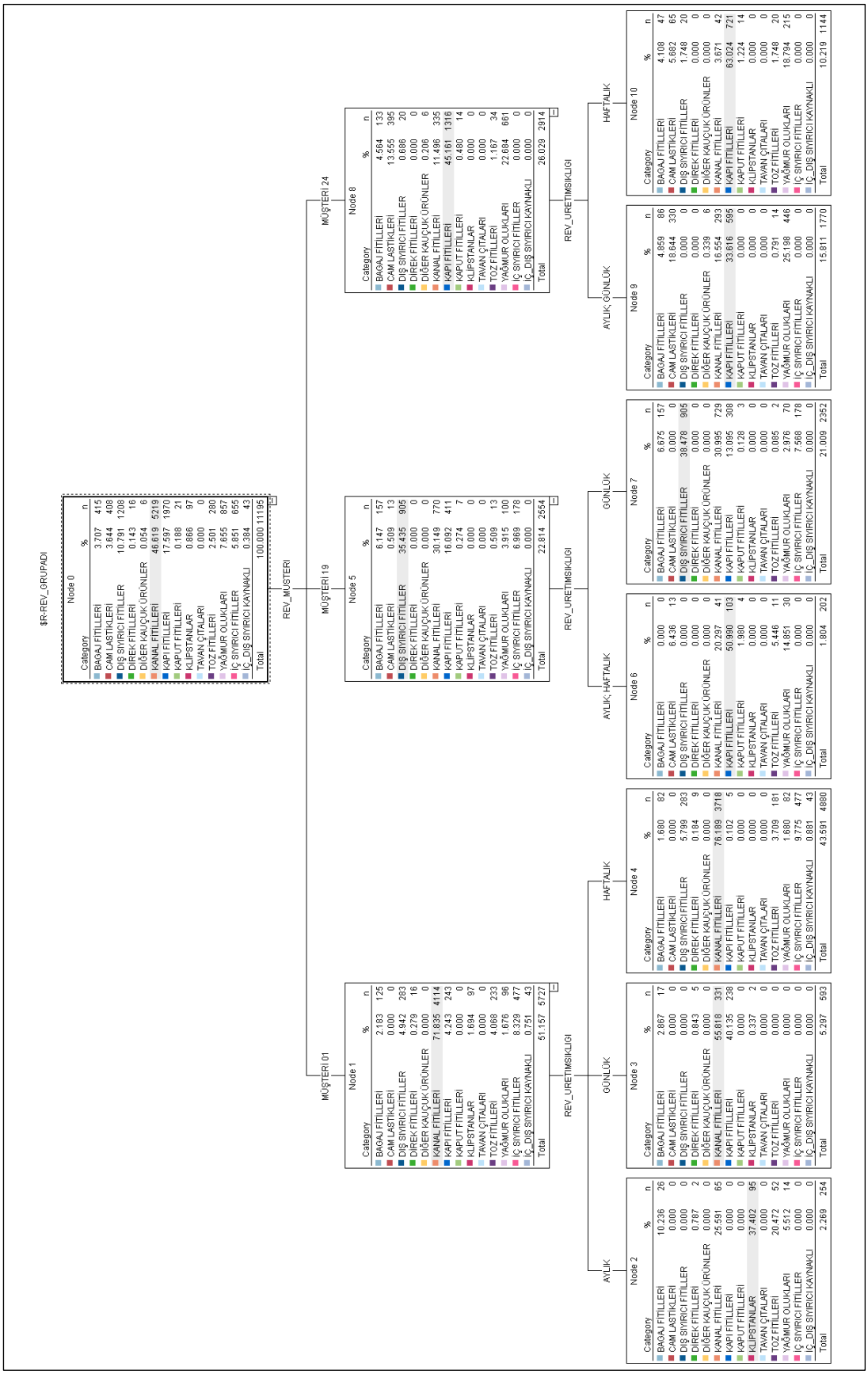
Ürün Çeşidi	Red Sayısı	Üretim Sayısı	Red Sayısı/Üretim
İÇ_DIŞ SIYIRICI KAYNAKLI	69	9974	0.0069
TAVAN ÇITALARI	2	1175	0.0017
KLİPSTANLAR	26	7541	0.0034
DİĞER KAUÇUK ÜRÜNLER	38	6960	0.0054
TOZ FİTİLLERİ	187	26932	0.0069
YAĞMUR OLUKLARI	194	32960	0.0058
DIŞ SIYIRICI FİTİLLER	402	88120	0.0045
İÇ SIYIRICI FİTİLLER	354	66539	0.0053
KANAL FİTİLLERİ	709	144225	0.0049
CAM LASTİKLERİ	125	15202	0.0082
DİREK FİTİLLERİ	9	2853	0.0031
BAGAJ FİTİLLERİ	370	27182	0.0136
KAPUT FİTİLLERİ	86	12386	0.0069
KAPI FİTİLLERİ	1138	119316	0.0095

9. Müşteri : Redlenen ürünlerin müşteri ürünlerine dağılımı incelenmiştir. İnceleme sonucunda 3 firmanın ürünlerinde yüksek oranlar tespit edilmiştir. Şekil 4.17 redlenmelerin müşteri ürünlerine dağılımını göstermektedir.



Şekil 4.17. Müşteri - redlenme ilişkisi

Redlenmeler en fazla Müşteri 01, Müşteri 19 ve Müşteri 24`ün ürünlerinde olmuştur. Yaklaşık olarak bu üç firmanın ürünlerinin redlenme oranı toplam redlenmelerin % 60`ı kadardır. Şekil 4.18`de bu üç firmanın redlenen ürünlerinin, üretim sıklığı ve ürün grubu ile ilişkisini gösteren karar ağacı verilmiştir.



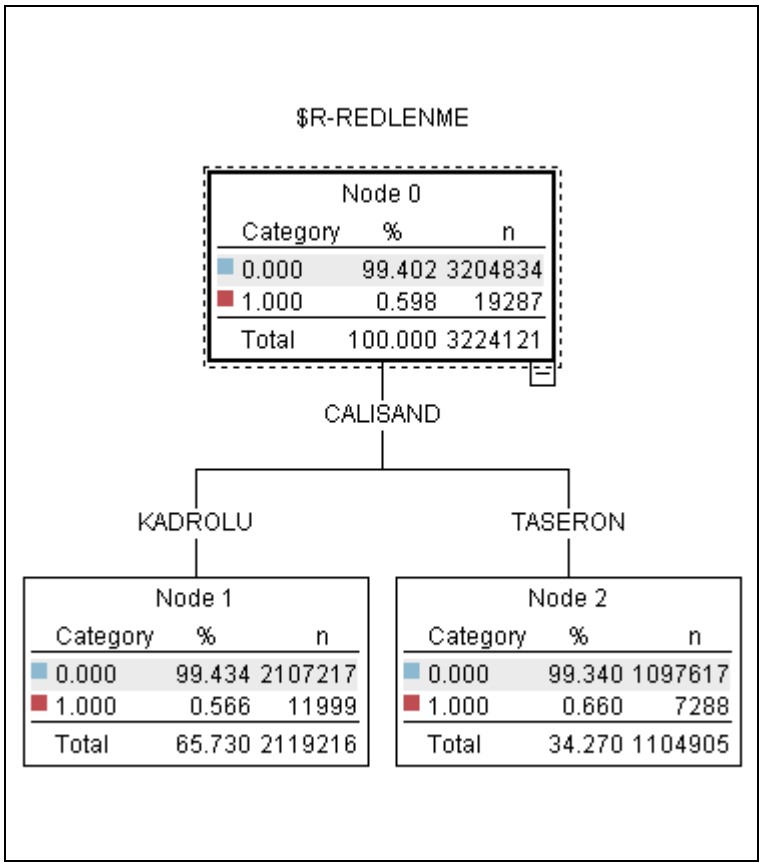
Şekil 4.18. 3 büyük müşterinin ürünlerinin üretim sıklığı – ürün grubu ilişkisi karar ağacı

Müşteri 01`in aylık üretilen ürünlerinden redlenenlerin % 37`si klipstanlar, % 26`sı kanal fitilleri, % 20`si toz fitilleridir. Günlük üretilen ürünlerinden redlenenlerin % 56`i kanal fitilleri, % 40`ı kapı fitilleridir. Haftalık üretilen ürünlerin ise, redlenmelerin % 76`sı kanal fitilidir.

Müşteri 19`un üretilen ürününün üretim sıklığı aylık veya haftalık ise redlenen ürünlerinin % 51`i kapı fitilleridir. Günlük üretilen ürünlerinde ise redlenenlerin % 31`i kanal fitilleri, % 38`i dış sıyırıcı fitilleridir.

Müşteri 24`ün üretilen ürününün üretim sıklığı aylık veya günlük ise redlenen ürünlerinin % 33`ü kapı fitili, % 25`i yağmur oluklarıdır. Haftalık üretilen ürünlerinde ise redlenenlerin % 63`ü kapı fitilleridir.

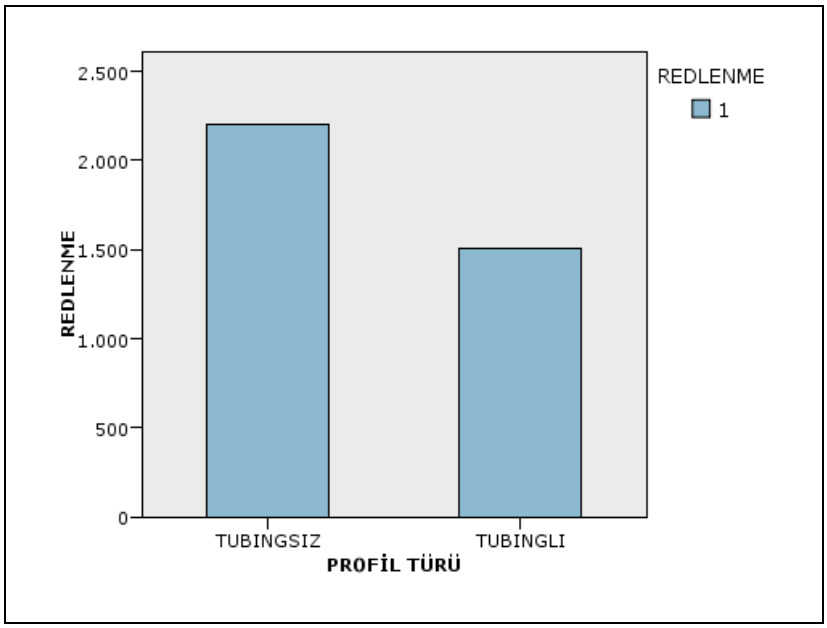
10. Çalışan Kadro : Çalışanların kadrolu ya da taşeron olmasının redlenme üzerindeki etkisini incelemek için redlenmenin incelendiği tüm süreç içerindeki üretimler de dikkate alınmıştır. Şekil 4.19`daki karar ağacı çalışanların kadro türü ile redlenme arasındaki ilişkiyi göstermektedir.



Şekil 4.19. Çalışan kadro durumu - redlenme ilişkisi karar ağacı

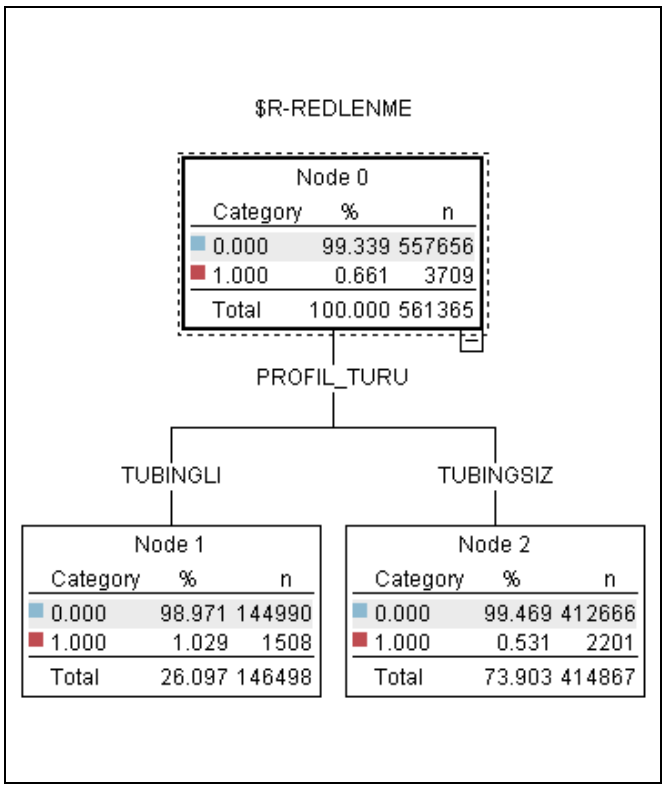
Üretiminde kadrolu çalışanların bulunduğu malzemelerin redlenme oranı % 0.56 ve başarı oranı % 99.44'dür. Üretiminde taşeron çalışanların bulunduğu malzemelerde ise başarı oranı % 99.34 iken redlenme oranı % 0.66 olarak tespit edilmiştir.

11. Profil Türü : Üretilen mamulde kullanılan profil türünün redlenme üzerine etkisi incelenmiş ve sonuçları Şekil 4.20`de verilmiştir. Model profil türünü redlenmeye etkisi konusunda anlamlı bulmuştur. Redlenmeler tubingli profiller de tubingsiz profillere göre daha fazla olmuştur.



Şekil 4.20. Profil türü - redlenme ilişkisi

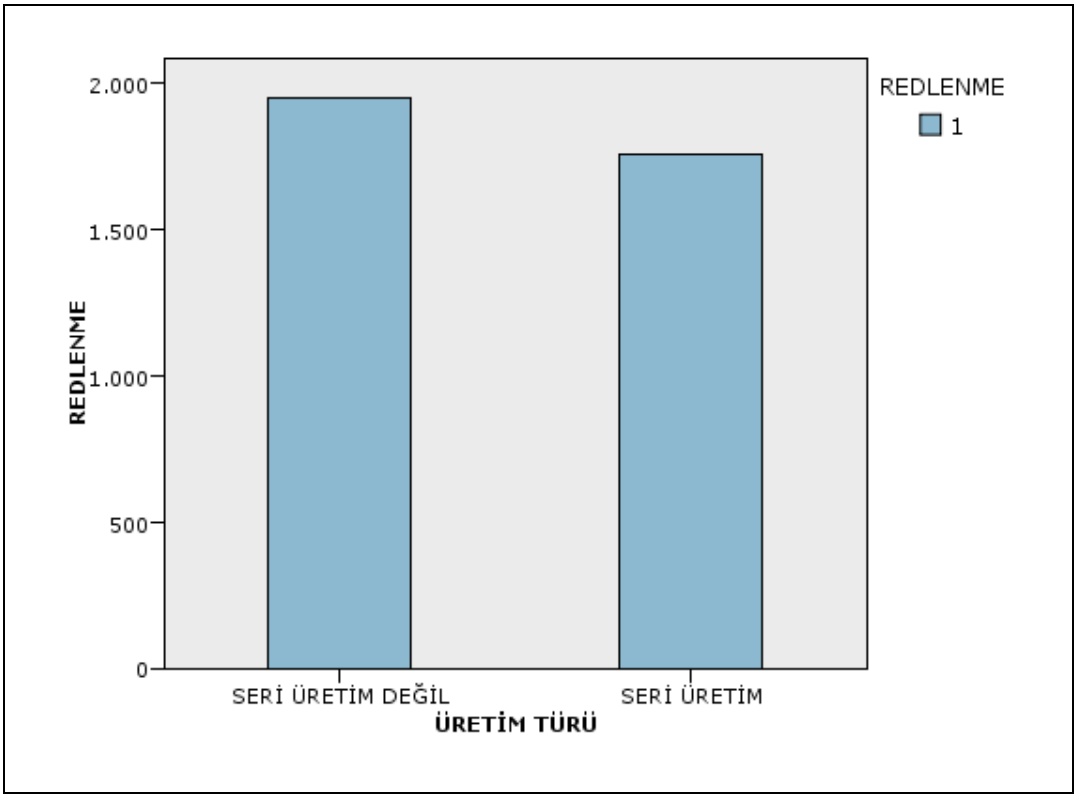
Profiller tubingli ve tubingsiz olmak üzere iki gruba ayrıldıktan sonra karar ağacı grafiği Şekil 4.21`de gösterildiği gibi olmuştur.



Şekil 4.21. Profil türü - redlenme ilişkisi karar ağacı

Veri setinde yer alan üretilmiş malzemelerde redlenme oranı % 0.66, başarı oranı % 99.34 olarak tespit edilmiştir. Tubingli profil ile üretim yapılan malzemelerin redlenme oranı % 1.03 ve başarı oranı % 98.97 dir. Tubingsiz olan profil ile üretim yapılmış malzemelerde ise başarı oranı % 99.47 iken redlenme oranı % 0.53 olarak bulunmuştur.

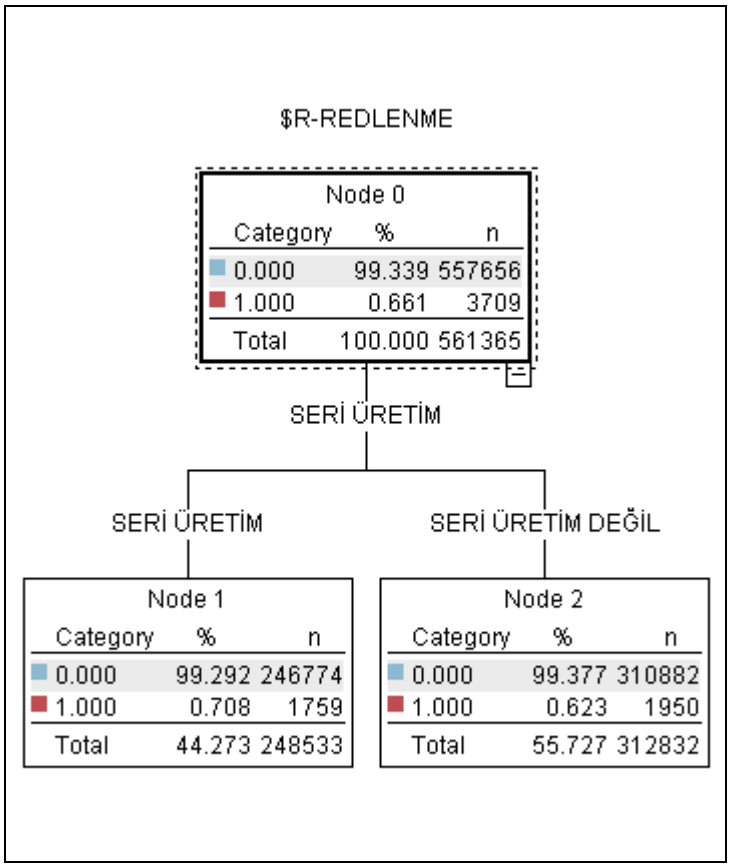
12. Üretim Türü : İşletmede üretilen ürünler, seri üretim ve siparişe göre üretim olmak üzere ikiye ayrılmaktadır. Veri setindeki malzemeler seri üretim ve seri üretim değil olarak iki grubda düzenlenerek analiz edilmiştir. Şekil 4.22 üretim şekli ile redlenme arasındaki ilişkiyi göstermektedir.



Şekil 4.22. Üretim türü - redlenme ilişkisi

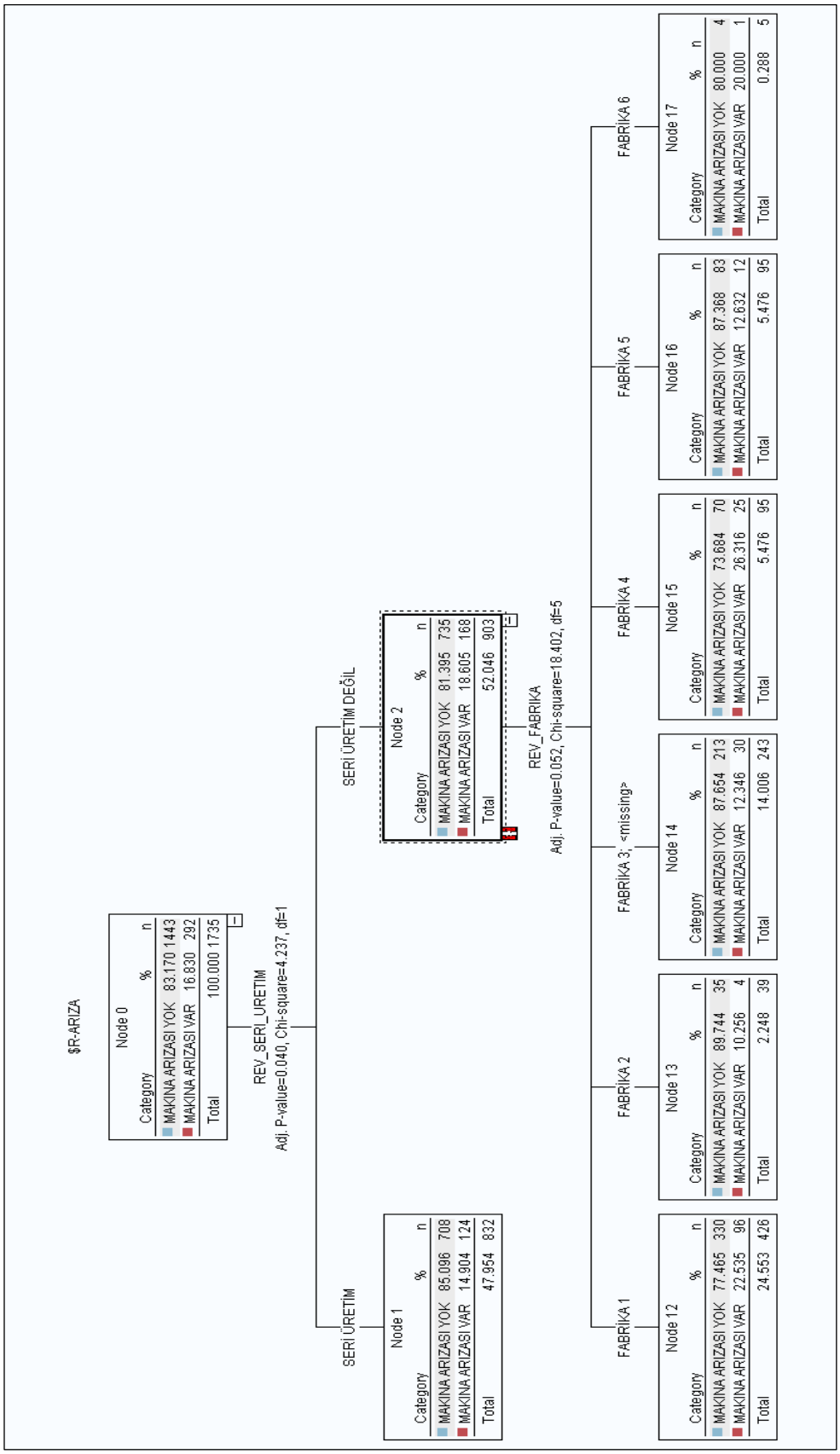
Seri olarak üretilmeyen malzemlerdeki redlenme oranı seri üretilen malzemelerin redlenme oranından daha falza olduğu ortaya çıkmıştır. Seri üretimde redlenme sayısı 1759 iken sipariş geldikçe üretim yapılan malzemelerde ise redlenme sayısı 1950`dir. Üretimler de dikkate alındığı zaman seri olarak üretilen malzemelerde redlenme oranı % 0.71 ve başarı oranı % 99.29 olarak tespit edilmiştir. Seri olarak

üretilmeyen malzemelerde ise redlenme oranı % 0.62 ve başarı oranı % 99.38`dır. Şekil 4.23`de seri üretimin redlenme üzerine etkisini gösteren karar ağacı verilmiştir.



Şekil 4.23. Üretim türü - redlenme ilişkisi karar ağacı

Şekil 4.24 redlenme nedenlerinden, makineden kaynaklı redlenmeler ve makine-insandan kaynaklı redlenmeler dikkate alındığında makine arızası üzerinde seri üretim ve fabrikanın etkisini göstermektedir.



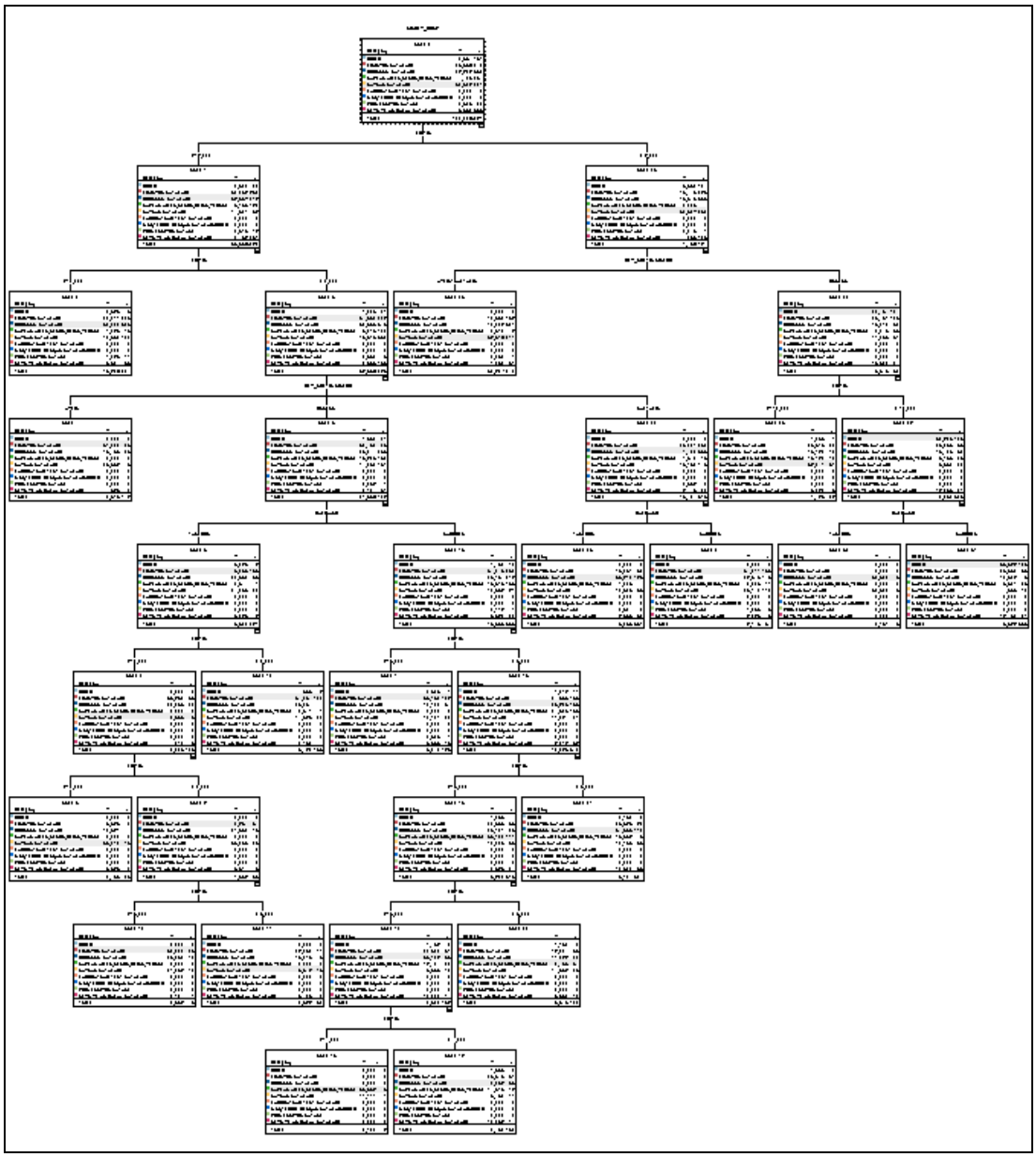
Şekil 4.24. Makine ve makine - insandan kaynaklı redlenmelerde makine arızasının seri üretim ve fabrika ile ilişki karar ağacı

Makineden kaynaklı redlenmeler ve makine-insandan kaynaklı redlenmelerde seri üretimde redlenen ürünlerin % 14.91`unda makine arızası olmuştur, % 85.19`da makina arızası görülmemiştir. Seri üretim olmayan redlenen ürünlerin % 18.60`ından makine arızası olurken, % 81,40`ında makine arızası görülmemiştir. Seri üretim olmayan redlenen ürünlerin fabrikalara dağılımına bakıldığında, fabrika 4 % 26.32 ile en yüksek redlenme oranına sahiptir. Fabrika 1 % 22.54 ile redlenme oranı en yüksek 2. fabrika olarak gözükmektedir. % 10.26 redlenme oranı ile fabrika 2 en düşük redlenme oranına sahiptir.

Şekil 4.25`deki karar ağacı, redlenme oranının daha yüksek olduğu tubingli profil ile üretimi yapılan ürünlerin üretiminde, çalışan kadro türü, çalışanların operasyonel eğitim düzeyi ve üretim sıklığının hata türü üzerindeki etkisini göstermektedir. Hataların % 58.53`ü eğitim düzeyi 9 ve 9`dan küçük olan çalışanların çalıştırığı birimlerde olmaktadır. Bu dallanmada % 39 oranından görünüm hataları, % 31 oranında finisyon hataları olmaktadır. Eğer eğitim düzeyi 0 ise, çalıştığı üretim birimindeki redlenmelerin % 57`si görünüm hatası sebeplidir. Eğitim düzeyi 1 ile 9 arasında olan çalışanların bulunduğu üretim birimlerinde aylık olarak üretimi yapılan ürünlerde meydana gelen hataların % 51`i finisyon hataları ve % 28`i kaynak hatalarından meydana gelmektedir.

Günlük olarak üretimi yapılan ürünlerin üretiminde eğitim düzeyi 2`nin altında olan taşeron çalışanların etkisi ile meydana gelen redlenmelerin % 65`i kaynak hatalarından oluşmaktadır. Eğitim düzeyi 2 ile 9 arasında olan taşeron çalışanların bulunduğu üretim birimlerinde ise bu kategorideki redlenmeler içerisinde kaynak hatası oranı % 33`e düşmektedir. Ama aynı zamanda eğitim düzeyi arttığı için verilen sorumluluklarında artması ile finisyon makinası kullanma yetkileride arttmaktadır. Bu nedenle eğitim düzeyi arttıkça kaynak hatası oranı düşmesine rağmen finisyon hatası oranı % 8`den % 40`a çıkmıştır.

Günlük olarak üretimi yapılan ürünlerde kadrolu çalışanların eğitim düzeyi 3`ün altında ise hataların % 65`ini finisyon hataları kapsamaktadır. Eğitim düzeyi 3 ile 9 arasında ise hata oranının finisyon, görünüm ve kaplama hatalarında düzgün bir şekilde dağıldığı görülmektedir.



Şekil 4.25. Tubingli profil ile üretimlerde kadro-eğitim-üretim sıklığı-hata türü ilişkisi karar ağacı

Haftalık olarak üretimi yapılan malzemelerde taşeron çalışanların etkisi ile meydana gelen redlenmelerin % 59`u görünüm hatası olurken bu hata oranı kadrolu çalışanlarda % 29`a düşmektedir.

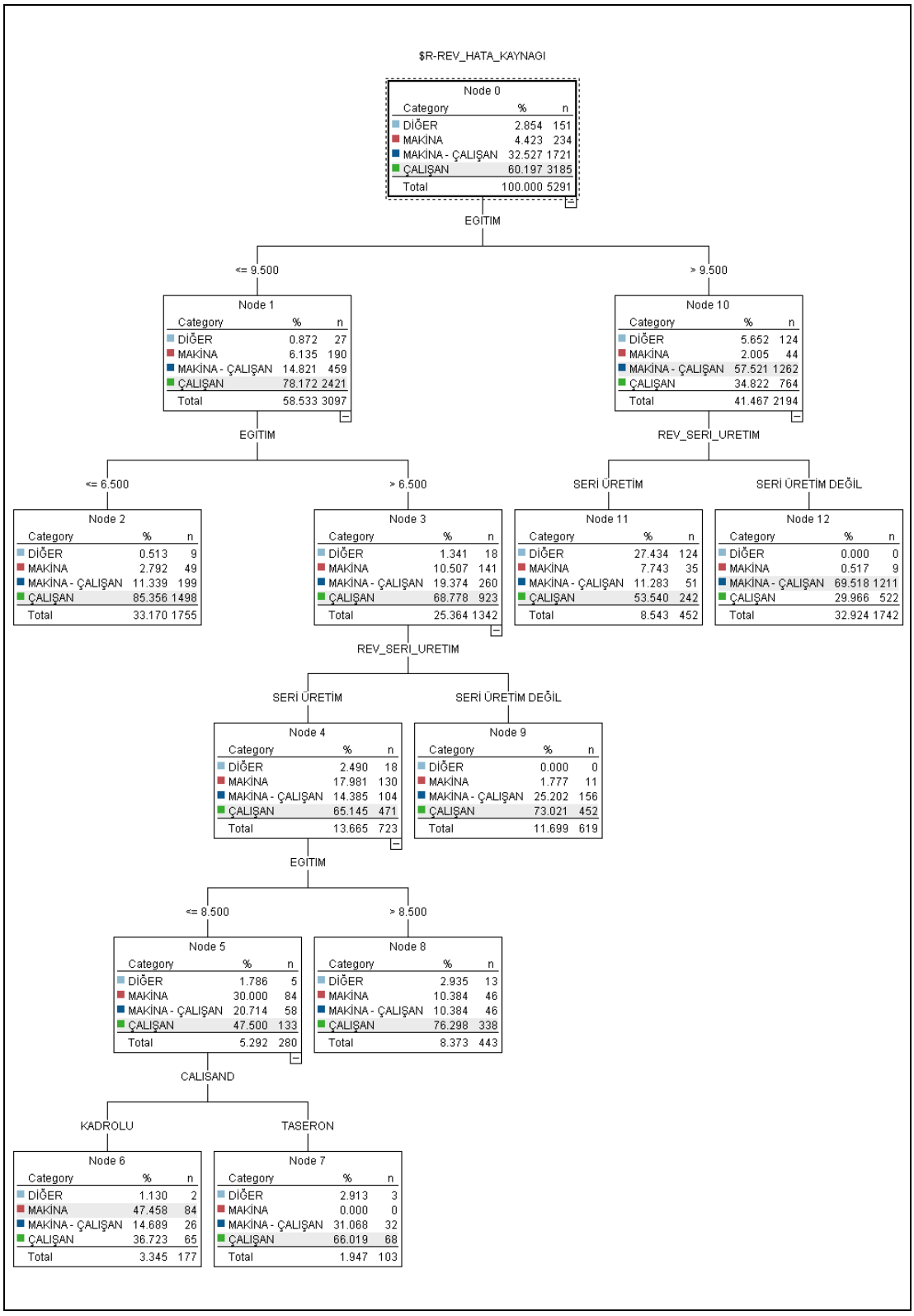
Eğitim düzeyi 9`un üzerinde olan çalışanların üretim yaptıkları üretim biriminde görünüm hataları % 16`ya düşerken, Kaynak hatalarının oranı % 14`den % 57`ye yükselmiştir.

Üretim sıklığı günlük olan ürünlerin üretiminde eğitim düzeyi 10`nun üzerinde olan çalışanlardan taşeron olanların neden oldukları redlenmelerin % 62`si görünüm hataları, kadrolu çalışanların neden oldukları redlenmelerin ise 12`si görünüm hatasıdır.

Şekil 4.26`daki karar ağacında tubingli profil ile üretilen ürünlerde, hata kaynağı üzerine üretimin seri üretim olup olmamasının, çalışan kadro türünün ve çalışan eğitim seviyesinin etkisini göstermektedir.

Tubingli profil ile üretilen ürünlerden redlenenlerin üretiminde çalışanların eğitim düzeyi 9.5`un altında düştükçe, çalışandan kaynaklı hataların oranı % 70 seviyesinde olmaktadır. Eğitim seviyesi 9.5`in üzerine çıkınca ise çalışandan kaynaklı hata oranı % 34`e düşmektedir. Eğer seri üretim yapılan bir ürünü üreten ekipde çalışanların seviyesi 9.5`in üzerinde ise redlenme üzerine çalışanın direk etkisi % 53`dür. Aynı şartlar altında seri üretim olmayan bir üretim gerçekleştiği takdirde redlenme üzerine çalışanın etkisi % 29`dur.

Eğitim düzeyi 6.5 ile 9.5 arasında bir değer ise seri üretim yapılması durumunda çalışanın etkisi % 65, seri üretim olmayan bir üretim yapılması durumunda ise % 73`dir. Eğer eğitim düzeyi 6.5 ile 8.5 arasında ise kadrolu çalışanın redlenme üzerine etkisi % 36, taşeron çalışanın redlenme üzerine etkisi de % 66`dır.

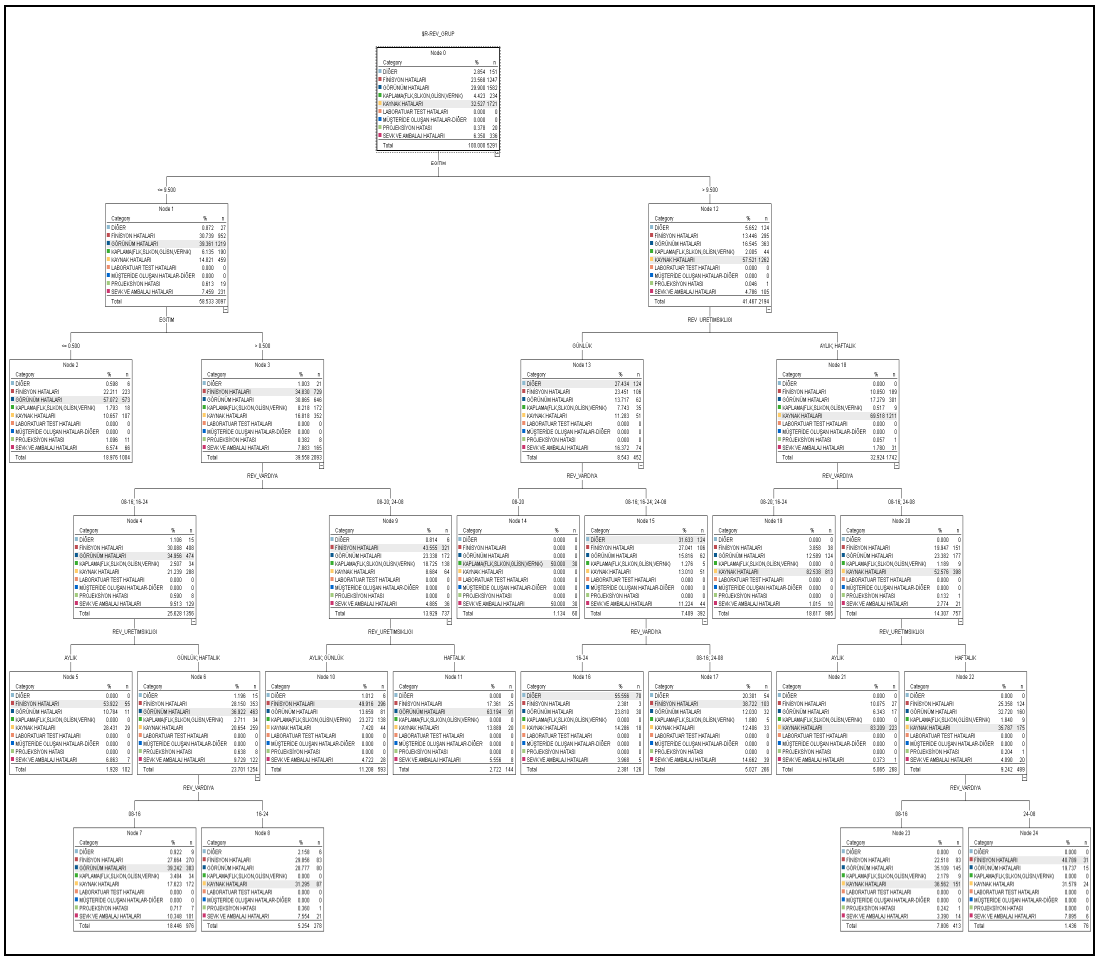


Şekil 4.26. Tubingli profil ile üretimlerde hata kaynağı–kadro–eğitim-seri üretim ilişkisi karar ağacı

Şekil 4.27 hata türü üzerine üretim sıklığının, fabrika ve vardiyanın etkisini gösteren karar ağacını vermektedir. Eğitim düzeyi 0.5`den küçük ise % 57 oranında görünüm hataları oluşmaktadır. Eğitim düzeyi 0.5 ile 9.5 arasında, vardiya 08-16 veya 16-24 vardiyası ve üretim sıklığı aylık ise redlenmelerin % 53`ü finisyon hatasıdır. Eğitim düzeyi 0.5 ile 9.5 arasında ve üretim sıklığı günlük veya haftalık ise 08-16 vardiyasında % 39`u görünüm hataları nedeni ile 16-24 vardiyasında ise % 31 oranında kaynak hatası yüzünden redlenmeler olmaktadır.

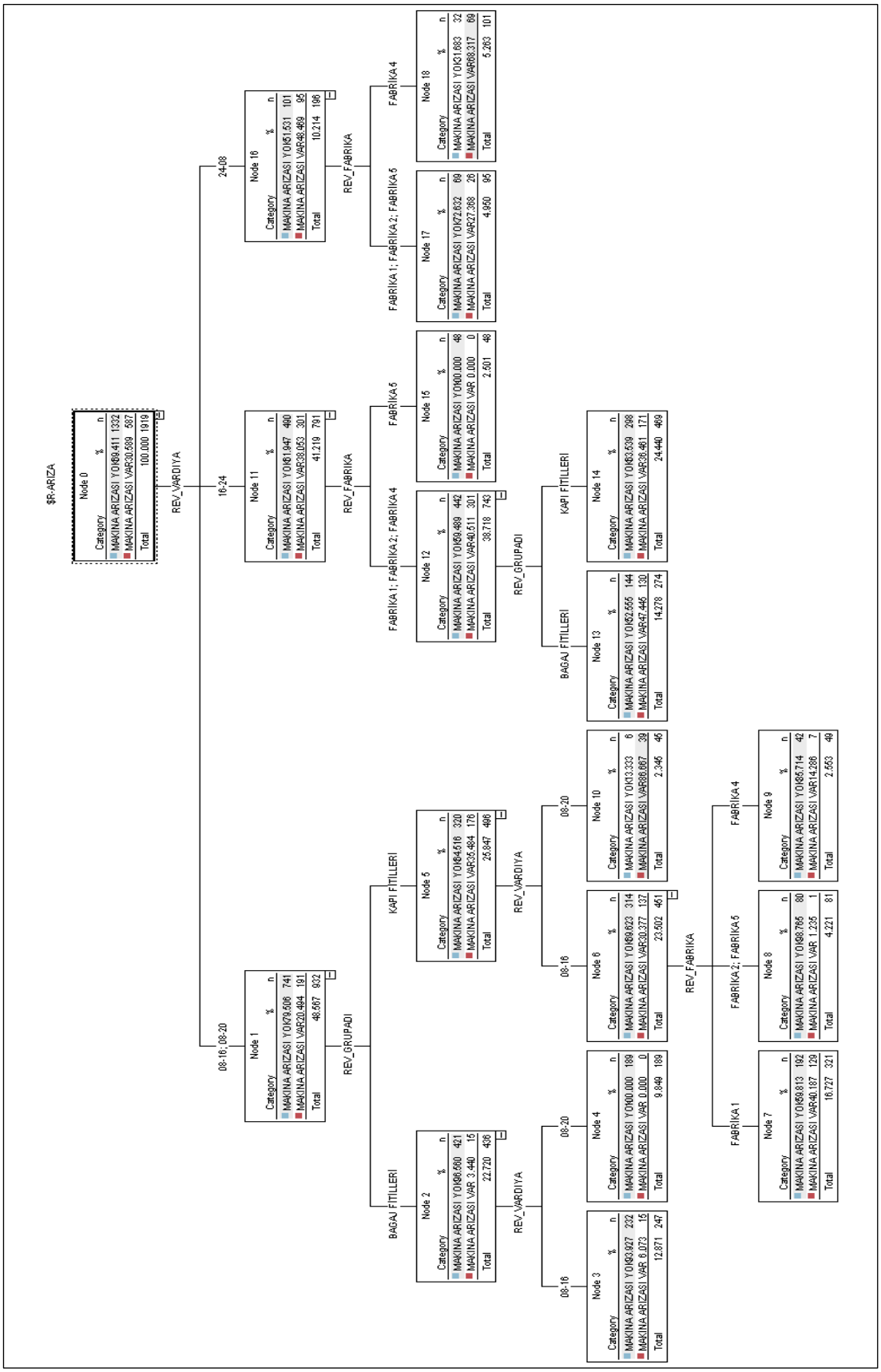
Eğitim düzeyi 0.5 ile 9.5 arasında ve vardiya 08-20 veya 24-08 ise üretim sıklığı aylık veya günlük olduğu üretimlerde redlenmelerin % 50`si finisyon hatası nedeni ile olurken üretim sıklığı haftalık olduğu üretimlerde ise % 63 görünüm hataları meydana gelmiştir.

Eğitim düzeyi 9.5`in üzerinde olduğu zaman günlük üretimlerde 08-20 vardiyasında redlenmelerin % 50`si kaplama hatası nedeni ile geriye kalan % 50`si sevk ve ambalaj hatası nedeni ile olmaktadır. Vardiya 08-16 veya 24-08 ise görünüm hataları % 39 oranında meydana gelmektedir. Eğer üretim sıklığı aylık ise redlenmelerin % 83`ü kaynak hatası nedeni iledir. Üretim sıklığı haftalık ve vardiya 08-16 ise % 37 oranında kaynak hatası meydana gelmektedir. Vardiya 24-08 ise redlenmelerin % 40`ı finisyon hatası nedeni ile meydana gelmektedir.



Şekil 4.27. Üretim sıklığı – fabrika – vardiya - red nedeni ilişkisi

Şekil 4.28`de makine arızası üzerine vardiya düzenini, ürünün üretildiği fabrikanın ve ürün gurubunun etkisini gösteren karar ağacı verilmiştir.



Şekil 4.28. Makine arızası – vardiya – fabrika – ürün grubu ilişkisi karar ağacı

Makineden ve çalışan-makine kaynaklı hatalarda vardiya 08-16 ve 08-20 ise % 79 oranında makine arızası meydana gelmektedir. Bagaj fitilleri üretiminde bu vardiyalarda arıza olma olasılığı % 3.5`dur. Kapı fitillerinde ise % 35.5`dur. Eğer kapı fitilleri fabrika 1`de üretilen ürünler ise makine arızası olasılığı % 40`dır. Fabrika 4`de üretilen kapı fitillerinde aynı vardiyada % 14 oranında makine arızası meydana gelmektedir.

16-24 vardiyasında fabrika 1, fabrika 2 ve fabrika 4`de üretilen ürün bagaj fitili ise makine arızası olasılığı % 47, kapı fitilleri üretiliyor ise makine arızası olasılığı % 36`dır. 24-08 vardiyasında fabrika 4`de makine arızası meydana gelme olasılığı % 68`dir.

4.2. Modelde Kullanılan Algoritmaların Karşılaştırılması

Hata kaynağı tahmininde karar ağacı aloritması diğer olarak gruplanmış 149 redlenmeye çalışandan kaynaklı demiştir. Makineden kaynaklı hataları % 35 oranında doğru tespit ederken, % 60`ını çalışandan kaynaklı olarak yorumlamıştır. Makine ve çalışandan kaynaklı olabilecek hataları % 70 oranında doğru tespit etmiştir. % 28`ini çalışan olarak yorumlamıştır. Çalışandan kaynaklı hataları ise % 81 oranında doğru bulmuştur.

Tablo 4.3. Karar ağacı algoritmasının hata kaynağı tahmini

HATA KAYNAĞI		**Makine**	**Makine - Çalışan**	**Çalışan**
Diğer	Sayı	2	0	149
	Satır	1.32	0.0	98.68
Makine	Sayı	84	9	141
	Satır	35.90	3.84	60.26
Makine - Çalışan	Sayı	26	1211	484
	Satır	1.51	70.36	28.13
Çalışan	Sayı	65	522	2598
	Satır	2.04	16.39	81.57

Hata kaynağı tahmininde yapay sinir ağı algoritması, diğer hata kaynaklarına % 100 çalışandan kaynaklı tespitinde bulunmuştur. Makine hatalarına ise % 96 çalışandan kaynaklı yorumu yapmıştır. Makine ve çalışandan kaynaklı hataları % 70 oranında doğru tespit etmiştir. Çalışandan kaynaklı hataları ise % 80 oranında doğru yorumlamıştır.

Tablo 4.4. Yapay sinir ağı algoritmasının hata kaynağı tahmini

HATA KAYNAĞI		**Makine - Çalışan**	**Çalışan**
Diğer	Sayı	0	151
	Satır	0.0	100.0
Makine	Sayı	9	225
	Satır	3.84	96.16
Makine - Çalışan	Sayı	1211	510
	Satır	70.36	29.64
Çalışan	Sayı	522	2663
	Satır	16.39	83.61

Hata kaynağı tespitinde, makine ve çalışandan kaynaklı hatalarda yapay sinir ağı ve karar ağacı % 70 oranında başarılı yorumda bulunmuşlardır. Çalışandan kaynaklı hatalarda ise yapay sinir ağı, karar ağacının % 81`lik başarılı sonucuna karşın, % 84 gibi bir başarı oranı ile daha doğru sonuçları vermiştir.

Makine arızası tahminin de karar ağacı algoritması, makine arızası olan durumlara % 59 oranında makine arızası var tespitinde bulunurken % 41 oranında makine arızası yok yorumu yapmıştır. Makine arızası olmayan durumlarda ise karar ağacı % 8 oranında makina arızası olduğu yorumu yaparken, % 91 oranında makine arızası yok tespitinde bulunmuştur.

Tablo 4.5. Karar ağacı algoritmasının makine arızası tahmini

ARIZA		**Makine Arızası Var**	**Makine Arızası Yok**
Makine Arızası Var	Sayı	349	238
	Satır %	59.45	40.56
Makine Arızası Yok	Sayı	108	1224
	Satır %	8.11	91.89

Yapay sinir ağı algoritması ise makine arızası tahmininde makine arızası olan durumları % 22 oranında tespit ederken, % 78 oranında makine arızası yok yorumu yapmıştır. Makine arızasının olmadığı durumlarda yapay sinir ağı % 3 oranında makine arızası var yorumu yaparken, % 97 lik bir başarı oranı ile makine arızası yok tespitinde bulunmuştur.

Tablo 4.6. Yapay sinir ağı algoritmasının makine arızası tahmini

ARIZA		**Makine Arızası Var**	**Makine Arızası Yok**
Makine Arızası Var	Sayı	127	460
	Satır %	21.63	78.37
Makine Arızası Yok	Sayı	34	1298
	Satır %	2.55	97.45

Makine arızası tahminin de, makine arızası olan durumları tespit etmekde karar ağacı daha başarılı olmuştur. Karar ağacı % 59 oranında, yapay sinir ağı ise % 22 oranında makine arızalarını tespit edebilmektedir. Makine arızası olmayan durumların tespitinde ise yapay sinir ağı daha başarı olmuştur. Yapay sinir ağı % 97 oranında, karar ağacı ise % 91 oranında makine arızası yok yorumunda bulunmuştur.

Hata türü tahmininde karar ağacı finisyon hatalarını % 81 oranında, görünüm hatalarını % 72 oranında, kaplama hatalarını % 74 oranında, laboratuar hatalarını % 88 oranında, projeksiyon hatalarını % 69 oranında ve sevk hatalarını % 66 oranında doğru tahmin etmiştir.

Tablo 4.7. Karar ağacı algoritmasının hata türü tahmini

HATA TÜRÜ		DİĞER	FİNİSYON	GÖRÜNÜM	KAPLAMA	KAYNAK	LAB.	PROJEK.	SEVK
DİĞER	Sayı	507	32	66	0	5	0	0	61
	Satır %	75.55	4.76	9.83	0.0	0.74	0.0	0.0	9.09
FİNİSYON	Sayı	40	4411	335	180	262	29	0	182
	Satır %	0.73	81.09	6.15	3.30	4.81	0.53	0.0	3.34
GÖRÜNÜM	Sayı	32	532	3172	73	272	0	12	263
	Satır %	0.73	12.21	72.81	1.67	6.24	0.0	0.27	6.03
KAPLAMA	Sayı	22	102	118	913	71	0	0	0
	Satır %	1.79	8.31	9.62	74.46	5.79	0.0	0.0	0.0
KAYNAK	Sayı	40	395	370	120	2434	3	0	200
	Satır %	1.12	11.08	10.38	3.36	68.33	0.08	0.0	5.61
LAB.	Sayı	0	12	0	0	0	91	0	0
	Satır %	0.0	11.65	0.0	0.0	0.0	88.34	0.0	0.0
PROJEK	Sayı	0	43	12	0	4	0	132	0
	Satır %	0.0	22.51	6.28	0.0	2.09	0.0	69.10	0.0
SEVK	Sayı	17	398	385	116	264	0	2	2331
	Satır %	0.48	11.32	10.95	3.30	7.51	0.0	0.05	66.35

Hata türü tahmininde yapay sinir ağı algoritması finisyon hatalarını % 64 oranında görünüm hatalarını % 56 oranında kaplama hatalarını da % 40 oranında doğru tespitte bulunmuştur.

Tablo 4.8. Yapay sinir ağı algoritmasının hata türü tahmini

HATA TÜRÜ		DİĞER	FİNİSYON	GÖRÜNÜM	KAPLAMA
DİĞER	Sayı	241	152	139	139
	Satır %	35.91	22.65	20.71	20.71
FİNİSYON	Sayı	149	3491	961	838
	Satır %	2.73	64.18	17.66	15.40
GÖRÜNÜM	Sayı	84	857	2451	964
	Satır %	1.92	19.67	56.26	22.13
KAPLAMA	Sayı	0	480	257	489
	Satır %	0.0	39.15	20.96	39.88
KAYNAK	Sayı	42	548	2252	720
	Satır %	1.17	15.38	63.22	20.21
LABORATUAR	Sayı	0	30	64	9
	Satır %	0.0	29.12	62.13	8.73
PROJEKSİYON	Sayı	15	80	87	9
	Satır %	7.85	41.88	45.54	4.71
SEVK	Sayı	131	559	869	1954
	Satır %	3.72	15.91	24.73	55.62

Hata türünün tahmin edilmesinde karar ağacı algoritması, yapay sinir ağı algoritmasına göre daha başarılı olmuştur. Yapay sinir ağı algoritması kaynak hataları, laboratuar hataları, projeksiyon hataları ve sevk hataları konusunda doğru tespitlerde bulunamamıştır.

Tablo 4.9. Karar ağacı algoritmasının redlenme tahmini

REDLENME		**0**	**1**
0	Sayı	3204693	367
	Satır %	99.98	0.01
1	Sayı	5147	13914
	Satır %	26.72	73.28

Redlenme tahmininde karar ağacı ile yapay sinir ağı bir birine yakın değerler elde etmişlerdir. Karar ağacı algoritmasının redlenmeyi doğru tespit etme oranı % 73`dür. Yapay sinir ağı algoritmasında ise bu oran % 70`dir.

Tablo 4.10. Tahminlerin karşılaştırılması

%	**Hata Kaynağı Tahmini**		**Makine Arızası Tahmini**		**Hata Türü Tahmini**		**Redlenme Tahmini**	
Algoritma	Doğru	Yanlış	Doğru	Yanlış	Doğru	Yanlış	Doğru	Yanlış
Karar Ağacı	73.58	26.42	81.97	18.03	73.40	26.60	72.79	27.21
Yapay Sinir Ağı	73.22	26.78	74.26	25.74	42.69	57.31	70.48	29.42

Tablo 4.10`da karar ağacı algoritması ile yapay sinir ağı algoritmasının tahmin sonuçlarının karşılaştırılması verilmiştir. Tablodan da görüldüğü üzere yapılan tahminlerde karar ağacı daha başarılı sonuçlar elde etmiştir. Hata türü tespitinde yapay sinir ağı % 43 gibi düşük bir oran ile doğru sonucu tahmin edebilmiştir.

BÖLÜM 5. DEĞERLENDİRME VE ÖNERİLER

Günümüzde firmalar pazar paylarını artırmak için müşteri memnuniyetini artırmak ve üretim maliyetlerinin düşürülmesi konularına çok önem vermektedir. Bir ürünün müşteriye ulaşmadan önce, tespit edilen hata, kusur ve eksikliklere neden olan etkenlerin ve bu hataların oratadan kaldırılması için gerekli olan tüm faaliyetler çok ciddi maliyetlere yol açar ve bu maliyetler iç başarısızlık maliyetlerini artırır. Firmalar hataları minimize etmek ve hata nedenlerini analiz etmek için çok çaba harcamaktadırlar.

Veri madenciliği, hataların ortadan kaldırılmasında, hata nedenlerinin analizi ile nelerin üzerine gidilmesi gerektiği, hataya sebebiyet veren değişkenlerinin tespitinde yardımcı olarak kullanılabilir. Ayrıca veri madeciliği ile gelecek tahminleri yapılarak yapılan iyileştirmeler izlenebilir ve değişkenlerin değişmesi durumunda tahmini veriler elde edilebilir.

İşletmelerde sağlıklı bir bilgi akışı sağlanması ve herşeyin sistematik hale getirilmesi için gerekli olan verilerin uygun bir şekilde depolanması ve gerektiğinde kullanılabilir olması sağlanmalıdır. Ama işletmeler bu amaca ulaşmak için bilgisayar ve depolama sistemlerinin hızla gelişmesi ve kapasitelerinde artması nedeni ile gerekli gereksiz tüm verileri saklamaya ve kirli veri yığınları oluşturmaya başlamışlardır. Uygulama aşamasında yaşanan zorluklar göz önüne alındığında veri yığınlarından değerli verilerin çıkarılması ve gerekli verilere ulaşma zorluğu, veri madenciliğinde veri toplama aşamasında en ciddi sorunu oluşturmaktadır.

Uygulama yapılan işletmede veri saklama işleminin sistematik olması ve geniş çaplı olarak yapılması nedeni ile üretilen malzemelerin hatalı olmasını etkileyen bir çok değişken tespit edilmiş ve analiz için veri setine dahil edilmiştir. Gerekli olan tüm

verilerin sistemli bir şekilde saklanması işletmelerde karar verme aşamasında birçok girdi olmasına olanak sağlayacaktır.

Bir ürünün redlenmesi analiz edildiğinde, hatalı ürün olma nedeni, redlenme tarihi ve saati, hatalı ürün miktarı, malzeme bilgileri, ürünlerin üretim tarihi ve saati, üretim miktarı, üretilen ürünü üreten ekip, ekibin vardiya düzeni, çalışanların aldığı eğitimler, çalışanların özellikleri, üretim birimi bilgileri, üretimde kullanılan makinaların arıza ve bakım bilgileri gibi bilgiler daha sonraki üretimlere ışık tutması hata nedenlerinin belirlenmesi ve olası uygunsuzlukların tespiti için kayıt altına alınması gerekmektedir. Üretim üzerine etkisi olan her veri, analize dahil edilebilir. Ama verilerin redlenme ile anlamlı bir ilişkiye sahip olup olmadığı analiz sonucunda ortaya çıkacaktır.

Analiz sonuçları göstermiştir ki;

- Yılın ilk çeyreğinde işletmede zam oranlarının açıklanması sonucu performans düşüklüğü meydana gelmektedir. Buna bağlı olarak da yılın ilk çeyreğinde redlenme oranı yükselmiştir. Yılın son çeyreğinde ise yıllık izinler kullanılmakta olduğundan, izine çıkış ve izinden döndükten sonraki bir dönemde redlenmeler artmıştır. Bunun en büyük nedeni olarak çalışanların konsantrasyon düşüklüğü gösterilebilir. Redlenmeler en fazla haftanın pazartesi günü meydana gelmektedir. Hafta içerisinde en düşük redlenme oranı cumartesi gününde olmaktadır. Çalışanların tatil dönüşlerinde performansını artıracak, ilgisi toplamaya yardımcı olacak yöntemler geliştirilmelidir.

- Redlenme oranı, fazla mesaili çalışma olan 08-20 vardiyasında en yüksek oranda olmaktadır. Diğer vardiyalar içerisinde üretimi en düşük olan 24-08 gece vardiyası redlenme oranı sırasında gündüz vardiyasından sonra 2. sıradadır. 08-20 vardiyası fazla mesaili çalışma olduğundan dikkat dağılımı ve konsantrasyon eksikliği daha fazla olmaktadır. Bu nedenle en fazla bu vardiya redlenme olmuştur. Mümkün oldukça bu vardiya tercih edilmemelidir. Eğer bu vardiya seri üretim olan bir üretim için kurulacak ise kadrolu çalışanlardan eğitim düzeyi 9.5`in üzerinde olan çalışanlar

tercih edilmelidir. Eğer proje seri üretim değil ise kadrolu ya da taşeron olması çok önemli değildir

- Çalışanların eğitim düzeyi 0 ise redlenme nedenleri en fazla yüzey, görünüm hatası olmaktadır. Eğitim düzeyi arttıkça çalışanların yetkinliklerinin artması nedeni ile hata oranı diğer hata gruplarına dağılmaktadır. Günlük olarak üretimi yapılan ürünlerde eğitim düzeyi 2`nin altında olan taşeron çalışanların çalışması durumunda hata nedeni % 65 oranında kaynak hatası nedeni ile olmaktadır. Eğer taşeron çalışanların eğitim düzeyi 2 ile 9 arasında olur ise kaynak hatasının redlenmeye etkisi % 33`e düşmektedir. Günlük olarak üretimi yapılan ürünlerde eğitim düzeyi 3`ün altında olan kadrolu çalışanlarda redlenme nedeni % 65 oranında finisyon hatası nedeni ile olmaktadır. Eğitim düzeyi 3 ile 9 arasında olması durumunda kadrolu çalışanların ürettiği ürünlerin redlenmesinde ciddi bir düşüş olduğu ve hataların diğer hata gruplarına düzgün bir şekilde dağıldığı gözlenmektedir.

- Kadrolu çalışanların eğitim düzeyi 9`un üzerine çıkınca kritik operasyon olan kaynak işleminde redlenmeler artmaktadır. Kritik operasyonda yetkinlikler arttığı için çalışanların kendine aşırı güveni hataya neden olmaktadır. Taşeron çalışanların eğitim düzeyi 10`un üzerinde ise belli bir süre azalış gösteren görünüm hataları oranı % 62`ye çıkmaktadır.

- Eğer seri üretim yapılan bir ürünü üreten ekipte çalışanların eğitim seviyesi 9.5`in üzerinde ise redlenme üzerine çalışanın direk etkisi % 53`dür. Aynı şartlar altında seri üretim olmayan bir üretim gerçekleştiği takdirde redlenme üzerine çalışanın etkisi % 29`dur.

- Eğitim düzeyi 0.5 ile 9.5 arasında ve vardiya 08-20 veya 24-08 ise üretim sıklığı aylık veya günlük olduğu üretimlerde redlenmelerin % 50`si finisyon hatası nedeni ile olurken üretim sıklığı haftalık olduğu üretimlerde ise % 63 görünüm hataları meydana gelmiştir. 08-20 ve 24-08 vardiyalarında üretim sıklığı haftalık olan ürünlerin üretiminde görünüm hatalarının oranını düşürmek için taşeron çalışanlar tercih edilmemeli, kadrolu çalışanlar tercih edilmelidir.

- Çalışanların aldıkları operasyonel eğitimler yetersizdir. Eğitim seviyeleri artırılmalı fakat yine de kontrol altında olmalıdır. Taşeron çalışanlar daha çok görünüm hatalarına neden olmaktadır. Eğitim seviyeleri artırılana kadar taşeron çalışanlar yüzey kontrolü ve ambalajlama işleminde çalıştırılmamalıdır. Kadrolu işçiler daha çok kaynak hatalarına neden olmaktadır. Kaynak operasyonu için tekrar bir eğitim verilmeli ve taşeron çalışanlardan bazıları bu operasyon için kadrolu çalışanların yanına verilmelidir. Böylelikle kadrolu çalışanların daha dikkatli ve özenli çalışması aynı zamanda taşeron çalışanların tecrübe kazanması beklenir.

- Hata oranı fabrika 1`de daha fazla çıkmıştır. Fabrika 1`in genel eğitim oranı diğer fabrikalara göre daha düşüktür. Ayrıca taşeron çalışan oranı bu fabrikada daha fazladır. Bu nedenle fabrika 1`de eğitim seviyesinde ve kadrolu çalışan sayısında bir dengeleme yapılmalıdır.

- Redlenmelerle en fazla günlük olarak üretimi yapılan seri ürünlerde karşılaşılmaktadır. Tubingli profil ile üretimi yapılan ürünlerdeki redlenme oranı tubingsiz profil ile üretim yapılan ürünlerin redlenme oranının yaklaşık 2 katıdır.

- Redlenmerler daha çok bagaj fitili ve kapı fitilinde olmaktadır. Bu grubu en çok etkileyen karar değişkenlerinden biri makine arızasıdır. 08-16 ve 08-20 vardiyalarında fabrika 1`de üretilen kapı fitillerinde yüksek oranda makine arızası olmaktadır. Farika 4`de ise kapı fitilleri üretiminde dikkate değer bir makine arızası meydana gelme ihtimali vardır. 16-24 vardiyasında fabrika 1, fabrika 2 ve fabrika 4`de bagaj fitili veya kapı fitili üretimi yapılıyor ise makine arızası olasılığı oldukça fazladır. Fabrika 4`de makineden kaynaklı ve makine çalışandan kaynaklı hataların % 68`i 24-08 vardiyasında makine arızası nedeni ile oluşmaktadır.

- Müşteri 01`in ürünlerinden üretim sıklığı günlük olan kanal fitilleri ve kapı fitilleri üretimlerinde redlenme riski yüksektir. Haftalık üretilen ürünlerde kanal fitillerinde üretim kontrol altında tutulmalıdır. Müşteri 19`un üretilen ürününün üretim sıklığı aylık veya haftalık ise kapı fitilleri üretiminde redlenme oranı yüksek olacaktır. Günlük üretilen ürünlerinde ise dış sıyırıcı fitillerine dikkat edilmelidir. Müşteri 24`ün haftalık üretilen ürünlerinde ise kapı fitilleri ve yağmur oluklarında redlenme

oranı yüksektir. Kısacası Müşteri 01`in günlük üretilen kanal ve kapı fitilleri, haftalık üretilen kalan fitilleri, Müşteri 19`un aylık veya haftalık üretimi yapılan kapı fitilleri ve günlük üretilen dış sıyırıcı fitilleri ve Müşteri 24`ün haftalık üretimi yapılan kapı fitilleri ile yağmur oluklarında üretim biriminde kontrol elemanı sayısı artırılmalı, daha iyi bir yönetim kontrolü sağlanabilecek 08-16 vardiyasında üretimler geçekleştirilmedir.

KAYNAKLAR

[1] AGRAWAL, R., IMIELINSKI, T., "Database Mining: A Performance Perspective," IEEE Transactions on Knowledge and Data Engineering, pp. 914-925, 1993.

[2] AHOLA, J., RINTA-RUNSALA, E., "Data Mining Case Studies in Customer Profiling," VTT Information Technology, Espoo., pp. 25, 2001, http://www.vtt.fi/inf/julkaisut/muut/2001/dm_case_studies.pdf, Erişim: 13.04.2009

[3] AKPINAR, H., "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", Ü. İşletme Fakültesi Dergisi, Cilt:29, Sayı: 1, sf. 1–22, 2000

[4] ALPAYDIN E., Bilişim 2000 Eğitim Semineri, "Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri", Bilgisayar Mühendisliği Bölümü, Boğaziçi Üniversitesi, 2000

[5] ANDERSON, DAVID, R.., SWEENY, D.J., WILLIAMS T.A., "Statistics for Business and Economics", West Publishing Company, USA, 1996

[6] ARGÜDEN, Y., ERŞAHİN, B., ARGE Danışmanlık, "Veri Madenciliği: Veriden Bilgiye, Masraftan Değere", 2008

[7] BERRY, MICHAEL J.A., LINOFF, G.S., "Mastering Data Mining:The Art and Science of Customer Relationship Management", Wiley Computer Publishing, pp. 122, USA, 2000

[8] BRANSTEN, L., “Technology – power tools – looking for patterns: data mining enables companies to better manage the ream of statistics they collect; the goal: spot the unexpected”, Wall Street Journal, 27 (12), pp. 16-20, 1999

[9] CERTO, SAMUEL, C., “Modern Management”, New Jersey Prentice Hall International Inc., 1997

[10] DAVIS, B., “Data mining transformed”, Information Week, 751: 86, 1999

[11] DEBORAH, R., CARVALHO, A., FREITAS, A., “A Hybrid Decision Tree/Genetic Algorithm Method for Data Mining,” Information Sciences,, No: 163, pp. 18, 2004

[12] DUNHAM, M., “Data Mining Introductory and Advanced Topics”, Prentice Hall, pp. 8, USA, 2003

[13] ENGİN O. , “Akış Tipi Çizelgeleme Problemlerinin Genetik Algoritma ile Çözüm Performansının Arttırılmasında Parametre Optimizasyonu”, ITÜ, Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, 2001

[14] FAYYAD, USAMA, PIATETSKY, G.S., PADHIRAIC S., “Knowledge Discovery and Data Mining: Towards a Unifying Framework,” Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), pp. 98-152, Portland, 1999

[15] HAN, J., KAMBER M., “Data Mining: Concepts and Techniques”, San Francisco: Morgan Kaufmann Publishers Inc., 2000

[16] HAND, D.J., “Data mining: statistics and more?”, The American Statistician, pp. 112-118, 1998

[17] HUDAIRY, H., “Data mining and decision making support in the governmental sector”, Master Thesis, Faculty of Graduate School of The University of Louisville, Kentucky, pp. 1-5, 2004

[18] HUNG, S., YEN, D., C., WANG, H., ‘Applying data mining to telecom churn management’, Expert Systems with Applications, pp. 1-10, 2005

[19] INTERNET, “Veri Madenciliği Veya Bilgi Keşfi”, http://www.bilgiyonetimi.org/ cm pages/mkl_gos.php?nt=538, Erişim: 11.04.2009

[20] JACOBS, P., ”Data Mining: What general managers need to know”, Harvard Management Update, 4 (10): 8, 1999

[21] JEFFERY, W., SEIFERT, “Data Mining and the Search for Security: Challenges for Connecting the Dots and Databases,” Government Information Quarterly, No: 21, pp. 462, 2004

[22] JIAWEI, H., V.D., “DBMiner: A System for Data Mining in Relational Databases and Data Warehouses”, pp. 9, 1997

[23] YONGSEOG, K., STREET, W.N., “An Intelligent System for Customer Targeting: A Data Mining Approach,” Decision Support Systems, No: 37, pp. 215-228, 2004

[24] KİREMİTÇİ B., “Veri Ambarlarında Veri Madenciliği ve Ulaştırma-Lojistik Sektöründe Bir Uygulama”, sf. 38, İstanbul, 2005

[25] KITLER, R., WANG, W., “The emerging role of data mining”, Solid State Technology, 42 (11): 45, 1998

[26] METHA, K., BHATTACHARYYA, S., “Adequacy of Training Data for Evolutionary Mining of Trading Rules,” Decision Support Systems, No:

37, pp. 462, 2004

[27] MITCHEL, T., McGraw-Hill “Machine Learning”, 1997

[28] ÖZMEN, Ş., “Ağ-Ekonomisinde Yeni Ticaret Yolu: e-Ticaret”, İstanbul Bilgi Üniversitesi Yayınları, İstanbul, 2003

[29] ÖZTEMEL, E., “Yapay Sinir Ağları”, Papatya Yayıncılık, sf. 29, İstanbul, 2003

[30] PARK, S.C., SELWYN, P., SHAW, M.J., “Dynamic Rule Refinement in Knowledge-based Data Mining Systems,” Decision Support Systems, No: 31, pp. 205-222, 2001

[31] GRAY, P., WATSON, H.J., “Decision Support in The Data Warehouse”, pp. 144, U.S.A., 1998

[32] WESLEY, S.C., LEE, C.F., HSU, Y.J., “On-line Personalized Sales Promotion in Electronic Commerce,” Expert Systems with Applications, No: 27, pp. 37, 2004

[33] SCHMITT, L.M., “Theory of genetic algorithms”, Theoretical Computer Science, No: 259, pp. 1-61, 2001

[34] SEVER, H., BUKET, O., “Veritabanlarında Bilgi Keşfine Formal Bir Yaklaşım”, 2002

[35] TANG, Z., MACLENNAN, J., “Data Mining with Sql Server 2005”, Wiley, 2005

[36] THULASI, C.R.,“SPSS Clementine for Data Mining Institutional research, University of Northern”, 2004, http://www.ir.uni.edu/dbWeb/pdf/present/dm_spss.pdf, Erişim: 13.04.2009

[37] TOKTAŞ, P., DEMİRHAN, M.B., "Risk Analizinde Veri Madenciliği Uygulamaları," Yöneylem Araştırması/Endüstri Mühendisliği – XXIV Ulusal Kongresi (Bildiri), 2004

[38] TWO CROWS CORPORATION, "Introduction to Data Mining and Knowledge Discovery," U.S.A., 1999, http://www.twocrows.com/intro-dm.pdf, Erişim: 13.04.2009

[39] WESLEY, R., FREITAS, A.A., GIMENES, I.M., "Discovering Interesting Knowledge from a Science and Technology Database with a Genetic Algorithm," "Applied Soft Computing, University of Kent Canterbury, pp. 2, UK, 2006

EKLER

EK A Tubingli Profil İle Üretimlerde Kadro-Eğitim-Üretim Sıklığı-Hata Türü İlişkisi Karar Ağacı Kural Seti

```
Rules for DİĞER - contains 1 rule(s)
        Rule 1 for  DİĞER
                if EGITIM > 9
                and REV_URETIMSIKLIGI in [ "GÜNLÜK" ]
                and EGITIM > 10
                and CALISAND = KADROLU
                then DİĞER
Rules for FİNİSYON HATALARI - contains 5 rule(s)
        Rule 1 for  FİNİSYON HATALARI
                if EGITIM <= 9
                and EGITIM > 0
                and REV_URETIMSIKLIGI = AYLIK
                then FİNİSYON HATALARI
        Rule 2 for  FİNİSYON HATALARI
                if EGITIM <= 9
                and EGITIM > 0
                and REV_URETIMSIKLIGI = GÜNLÜK
                and CALISAND = TASERON
                and EGITIM <= 7
                and EGITIM > 2
                and EGITIM <= 6
                then FİNİSYON HATALARI
        Rule 3 for  FİNİSYON HATALARI
                if EGITIM <= 9
                and EGITIM > 0
                and REV_URETIMSIKLIGI = GÜNLÜK
                and CALISAND = TASERON
                and EGITIM > 7
                then FİNİSYON HATALARI
        Rule 4 for  FİNİSYON HATALARI
                if EGITIM <= 9
                and EGITIM > 0
                and REV_URETIMSIKLIGI = GÜNLÜK
                and CALISAND = KADROLU
                and EGITIM <= 3
                then FİNİSYON HATALARI
        Rule 5 for  FİNİSYON HATALARI
                if EGITIM <= 9
                and EGITIM > 0
                and REV_URETIMSIKLIGI = HAFTALIK
                and CALISAND = KADROLU
                then FİNİSYON HATALARI
Rules for GÖRÜNÜM HATALARI - contains 5 rule(s)
        Rule 1 for  GÖRÜNÜM HATALARI
                if EGITIM <= 9
                and EGITIM <= 0
                then GÖRÜNÜM HATALARI
        Rule 2 for  GÖRÜNÜM HATALARI
```

EK A (Devam) Tubingli Profil İle Üretimlerde Kadro-Eğitim-Üretim Sıklığı-Hata Türü İlişkisi Karar Ağacı Kural Seti

```
                    if EGITIM <= 9
                    and EGITIM > 0
                    and REV_URETIMSIKLIGI = GÜNLÜK
                    and CALISAND = KADROLU
                    and EGITIM > 3
                    and EGITIM <= 8
                    and EGITIM <= 6
                    and EGITIM > 4
                    then GÖRÜNÜM HATALARI
          Rule 3 for  GÖRÜNÜM HATALARI
                    if EGITIM <= 9
                    and EGITIM > 0
                    and REV_URETIMSIKLIGI = GÜNLÜK
                    and CALISAND = KADROLU
                    and EGITIM > 3
                    and EGITIM > 8
                    then GÖRÜNÜM HATALARI
          Rule 4 for  GÖRÜNÜM HATALARI
                    if EGITIM <= 9
                    and EGITIM > 0
                    and REV_URETIMSIKLIGI = HAFTALIK
                    and CALISAND = TASERON
                    then GÖRÜNÜM HATALARI
          Rule 5 for  GÖRÜNÜM HATALARI
                    if EGITIM > 9
                    and REV_URETIMSIKLIGI in [ "GÜNLÜK" ]
                    and EGITIM > 10
                    and CALISAND = TASERON
                    then GÖRÜNÜM HATALARI
Rules for KAPLAMA(FLK,SLKON,GLİSN,VERNK) - contains 2 rule(s)
          Rule 1 for  KAPLAMA(FLK,SLKON,GLİSN,VERNK)
                    if EGITIM <= 9
                    and EGITIM > 0
                    and REV_URETIMSIKLIGI = GÜNLÜK
                    and CALISAND = KADROLU
                    and EGITIM > 3
                    and EGITIM <= 8
                    and EGITIM <= 6
                    and EGITIM <= 4
                    then KAPLAMA(FLK,SLKON,GLİSN,VERNK)
          Rule 2 for  KAPLAMA(FLK,SLKON,GLİSN,VERNK)
                    if EGITIM <= 9
                    and EGITIM > 0
                    and REV_URETIMSIKLIGI = GÜNLÜK
                    and CALISAND = KADROLU
                    and EGITIM > 3
                    and EGITIM <= 8
                    and EGITIM > 6
                    then KAPLAMA(FLK,SLKON,GLİSN,VERNK)
Rules for KAYNAK HATALARI - contains 4 rule(s)
          Rule 1 for  KAYNAK HATALARI
                    if EGITIM <= 9
                    and EGITIM > 0
                    and REV_URETIMSIKLIGI = GÜNLÜK
                    and CALISAND = TASERON
                    and EGITIM <= 7
                    and EGITIM <= 2
                    then KAYNAK HATALARI
          Rule 2 for  KAYNAK HATALARI
```

EK A (Devam) Tubingli Profil İle Üretimlerde Kadro-Eğitim-Üretim Sıklığı-Hata Türü İlişkisi Karar Ağacı Kural Seti

```
if EGITIM <= 9
        and EGITIM > 0
        and REV_URETIMSIKLIGI = GÜNLÜK
        and CALISAND = TASERON
        and EGITIM <= 7
        and EGITIM > 2
        and EGITIM > 6
        then KAYNAK HATALARI
Rule 3 for  KAYNAK HATALARI
        if EGITIM > 9
        and REV_URETIMSIKLIGI in [ "AYLIK" "HAFTALIK" ]
        then KAYNAK HATALARI
Rule 4 for  KAYNAK HATALARI
       if EGITIM > 9
       and REV_URETIMSIKLIGI in [ "GÜNLÜK" ]
       and EGITIM <= 10
       then KAYNAK HATALARI
      Default: KAYNAK HATALARI
```

EK B Tubingli Profil İle Üretimlerde Hata Kaynağı – Kadro – Eğitim - Seri Üretim İlişkisi Karar Ağacı Kural Seti

```
Rules for ÇALIŞAN - contains 5 rule(s)
        Rule 1 for  ÇALIŞAN
                if EGITIM <= 9,500
                and EGITIM <= 6,500
                then ÇALIŞAN
        Rule 2 for  ÇALIŞAN
                if EGITIM <= 9,500
                and EGITIM > 6,500
                and REV_SERI_URETIM in [ "SERİ ÜRETİM" ]
                and EGITIM <= 8,500
                and CALISAND in [ "TASERON" ]
                then ÇALIŞAN
        Rule 3 for  ÇALIŞAN
                if EGITIM <= 9,500
                and EGITIM > 6,500
                and REV_SERI_URETIM in [ "SERİ ÜRETİM" ]
                and EGITIM > 8,500
                then ÇALIŞAN
        Rule 4 for  ÇALIŞAN
                if EGITIM <= 9,500
                and EGITIM > 6,500
                and REV_SERI_URETIM in [ "SERİ ÜRETİM DEĞİL" ]
                then ÇALIŞAN
        Rule 5 for  ÇALIŞAN
                if EGITIM > 9,500
                and REV_SERI_URETIM in [ "SERİ ÜRETİM" ]
                then ÇALIŞAN
Rules for MAKİNA - contains 1 rule(s)
        Rule 1 for  MAKİNA
                if EGITIM <= 9,500
                and EGITIM > 6,500
                and REV_SERI_URETIM in [ "SERİ ÜRETİM" ]
                and EGITIM <= 8,500
                and CALISAND in [ "KADROLU" ]
                then MAKİNA
Rules for MAKİNA - ÇALIŞAN - contains 1 rule(s)
        Rule 1 for  MAKİNA - ÇALIŞAN
                if EGITIM > 9,500
                and REV_SERI_URETIM in [ "SERİ ÜRETİM DEĞİL" ]
                then MAKİNA - ÇALIŞAN
Default: ÇALIŞAN
```

EK C Üretim Sıklığı-Fabrika-Vardiya-Red Nedeni İlişkisi Karar Ağacı Kural Seti

```
Rules for DİĞER - contains 1 rule(s)
        Rule 1 for  DİĞER
                if EGITIM > 9,500
                and REV_URETIMSIKLIGI in [ "GÜNLÜK" ]
                and REV_VARDIYA in [ "08-16" "16-24" "24-08" ]
                and REV_VARDIYA in [ "16-24" ]
                then DİĞER
Rules for FİNİSYON HATALARI - contains 4 rule(s)
        Rule 1 for  FİNİSYON HATALARI
                if EGITIM <= 9,500
                and EGITIM > 0,500
                and REV_VARDIYA in [ "08-16" "16-24" ]
                and REV_URETIMSIKLIGI in [ "AYLIK" ]
                then FİNİSYON HATALARI
        Rule 2 for  FİNİSYON HATALARI
                if EGITIM <= 9,500
                and EGITIM > 0,500
                and REV_VARDIYA in [ "08-20" "24-08" ]
                and REV_URETIMSIKLIGI in [ "AYLIK" "GÜNLÜK" ]
                then FİNİSYON HATALARI
        Rule 3 for  FİNİSYON HATALARI
                if EGITIM > 9,500
                and REV_URETIMSIKLIGI in [ "GÜNLÜK" ]
                and REV_VARDIYA in [ "08-16" "16-24" "24-08" ]
                and REV_VARDIYA in [ "08-16" "24-08" ]
                then FİNİSYON HATALARI
        Rule 4 for  FİNİSYON HATALARI
                if EGITIM > 9,500
                and REV_URETIMSIKLIGI in [ "AYLIK" "HAFTALIK" ]
                and REV_VARDIYA in [ "08-16" "24-08" ]
                and REV_URETIMSIKLIGI in [ "HAFTALIK" ]
                and REV_VARDIYA in [ "24-08" ]
                then FİNİSYON HATALARI
Rules for GÖRÜNÜM HATALARI - contains 3 rule(s)
        Rule 1 for  GÖRÜNÜM HATALARI
                if EGITIM <= 9,500
                and EGITIM <= 0,500
                then GÖRÜNÜM HATALARI
        Rule 2 for  GÖRÜNÜM HATALARI
                if EGITIM <= 9,500
                and EGITIM > 0,500
                and REV_VARDIYA in [ "08-16" "16-24" ]
                and REV_URETIMSIKLIGI in [ "GÜNLÜK" "HAFTALIK" ]
                and REV_VARDIYA in [ "08-16" ]
                then GÖRÜNÜM HATALARI
        Rule 3 for  GÖRÜNÜM HATALARI
                if EGITIM <= 9,500
                and EGITIM > 0,500
                and REV_VARDIYA in [ "08-20" "24-08" ]
                and REV_URETIMSIKLIGI in [ "HAFTALIK" ]
                then GÖRÜNÜM HATALARI
Rules for KAPLAMA(FLK,SLKON,GLİSN,VERNK) - contains 1 rule(s)
        Rule 1 for  KAPLAMA(FLK,SLKON,GLİSN,VERNK)
                if EGITIM > 9,500
                and REV_URETIMSIKLIGI in [ "GÜNLÜK" ]
                and REV_VARDIYA in [ "08-20" ]
                then KAPLAMA(FLK,SLKON,GLİSN,VERNK)
Rules for KAYNAK HATALARI - contains 4 rule(s)
        Rule 1 for  KAYNAK HATALARI
                if EGITIM <= 9,500
                and EGITIM > 0,500
                and REV_VARDIYA in [ "08-16" "16-24" ]
```

EK C (Devam) Üretim Sıklığı-Fabrika-Vardiya-Red Nedeni İlişkisi Karar Ağacı Kural Seti

```
                    and REV_URETIMSIKLIGI in [ "GÜNLÜK" "HAFTALIK" ]
                    and REV_VARDIYA in [ "16-24" ]
                    then KAYNAK HATALARI
          Rule 2 for  KAYNAK HATALARI
                    if EGITIM > 9,500
                    and REV_URETIMSIKLIGI in [ "AYLIK" "HAFTALIK" ]
                    and REV_VARDIYA in [ "08-20" "16-24" ]
                    then KAYNAK HATALARI
          Rule 3 for  KAYNAK HATALARI
                    if EGITIM > 9,500
                    and REV_URETIMSIKLIGI in [ "AYLIK" "HAFTALIK" ]
                    and REV_VARDIYA in [ "08-16" "24-08" ]
                    and REV_URETIMSIKLIGI in [ "AYLIK" ]
                    then KAYNAK HATALARI
          Rule 4 for  KAYNAK HATALARI
                    if EGITIM > 9,500
                    and REV_URETIMSIKLIGI in [ "AYLIK" "HAFTALIK" ]
                    and REV_VARDIYA in [ "08-16" "24-08" ]
                    and REV_URETIMSIKLIGI in [ "HAFTALIK" ]
                    and REV_VARDIYA in [ "08-16" ]
                    then KAYNAK HATALARI
Default: KAYNAK HATALARI
```

EK D Makine Arızası – Vardiya – Fabrika – Ürün Gurbu İlişkisi Karar Ağacı Kural Seti

```
Rules for MAKINA ARIZASI VAR - contains 2 rule(s)
        Rule 1 for  MAKINA ARIZASI VAR
                if REV_VARDIYA in [ "08-16" "08-20" ]
                and REV_GRUPADI in [ "KAPI FİTİLLERİ" ]
                and REV_VARDIYA in [ "08-20" ]
                then MAKINA ARIZASI VAR
        Rule 2 for  MAKINA ARIZASI VAR
                if REV_VARDIYA in [ "24-08" ]
                and REV_FABRIKA in [ "FABRİKA 4" ]
                then MAKINA ARIZASI VAR
Rules for MAKINA ARIZASI YOK - contains 9 rule(s)
        Rule 1 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "08-16" "08-20" ]
                and REV_GRUPADI in [ "BAGAJ FİTİLLERİ" ]
                and REV_VARDIYA in [ "08-16" ]
                then MAKINA ARIZASI YOK
        Rule 2 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "08-16" "08-20" ]
                and REV_GRUPADI in [ "BAGAJ FİTİLLERİ" ]
                and REV_VARDIYA in [ "08-20" ]
                then MAKINA ARIZASI YOK
        Rule 3 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "08-16" "08-20" ]
                and REV_GRUPADI in [ "KAPI FİTİLLERİ" ]
                and REV_VARDIYA in [ "08-16" ]
                and REV_FABRIKA in [ "FABRİKA 1" ]
                then MAKINA ARIZASI YOK
        Rule 4 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "08-16" "08-20" ]
                and REV_GRUPADI in [ "KAPI FİTİLLERİ" ]
                and REV_VARDIYA in [ "08-16" ]
                and REV_FABRIKA in [ "FABRİKA 2" "FABRİKA 5" ]
                then MAKINA ARIZASI YOK
        Rule 5 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "08-16" "08-20" ]
                and REV_GRUPADI in [ "KAPI FİTİLLERİ" ]
                and REV_VARDIYA in [ "08-16" ]
                and REV_FABRIKA in [ "FABRİKA 4" ]
                then MAKINA ARIZASI YOK
        Rule 6 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "16-24" ]
                and REV_FABRIKA in [ "FABRİKA 1" "FABRİKA 2" "FABRİKA 4" ]
                and REV_GRUPADI in [ "BAGAJ FİTİLLERİ" ]
                then MAKINA ARIZASI YOK
        Rule 7 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "16-24" ]
                and REV_FABRIKA in [ "FABRİKA 1" "FABRİKA 2" "FABRİKA 4" ]
                and REV_GRUPADI in [ "KAPI FİTİLLERİ" ]
                then MAKINA ARIZASI YOK
        Rule 8 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "16-24" ]
                and REV_FABRIKA in [ "FABRİKA 5" ]
                then MAKINA ARIZASI YOK
        Rule 9 for  MAKINA ARIZASI YOK
                if REV_VARDIYA in [ "24-08" ]
                and REV_FABRIKA in [ "FABRİKA 1" "FABRİKA 2" "FABRİKA 5" ]
                then MAKINA ARIZASI YOK
Default: MAKINA ARIZASI YOK
```

ÖZGEÇMİŞ

Muhammet Çetin, 24.04.1983 de Zonguldak`ın Kdz.Ereğli ilçesinde doğdu. İlkokulu ve orta 2`ye kadar ki eğitimini Kdz Ereğli`de tamamladı. Orta 3 ve lise eğitimini Düzce'de tamamladı. 2001 yılında Düzce Süper Lisesinde, Fen Bölümünden mezun oldu. 2002 yılında başladığı Selçuk Üniversitesi Endüstri Mühendisliği bölümünü 2006 yılında bitirdi. 2007 yılından Standard Profil Tic. ve San A.Ş`ye üretim sorumlusu olarak girdi. 2008 yılında Planlama ve Lojistik Müdürlüğünde Planlama Mühendisi olarak görev aldı. 2009 yılında Bilgi Teknolojileri Müdürlüğünde Yazılım Uzmanı olarak çalışmaya başladı. Şu anda Standard Profil Tic. ve San. A.Ş.`de Yazılım Uzmanı olarak görev yapmaktadır. Muhammet Çetin evli ve mutlu bir aileye sahiptir.