T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İMALAT SANAYİNDE VERİ MADENCİLİĞİ DESTEKLİ TEDARİKÇİ SEÇİMİ UYGULAMASI

DOKTORA TEZİ

Endüstri Yük. Müh. Aslan ÇOBAN

Enstitü Anabilim Dalı : MAKİNA EĞİTİMİ
Tez Danışmanı : Prof. Dr. İsmet ÇEVİK

Mart 2006

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İMALAT SANAYİNDE VERİ MADENCİLİĞİ DESTEKLİ TEDARİKÇİ SEÇİMİ UYGULAMASI

DOKTORA TEZİ

Endüstri Yük. Müh. Aslan ÇOBAN

Enstitü Anabilim Dalı : MAKİNA EĞİTİMİ

Bu tez 20/03/2006 tarihinde aşağıdaki jüri tarafından Oybirliği/Oyçokluğu ile kabul edilmiştir.

Prof. Dr. İ.Mete DOĞRUER **Jüri Başkanı**	**Prof. Dr. İsmet ÇEVİK** **Jüri Üyesi**	**Prof. Dr. İbrahim ÖZSERT** **Jüri Üyesi**
Doç.Dr. İ. Hakkı BİÇER **Jüri Üyesi**		**Yrd. Doç. Dr. Bayram TOPAL** **Jüri Üyesi**

ÖNSÖZ

Veri madenciliği kavramı, bilgisayar teknolojilerinin işletme ve organizasyonlarda daha etkin bir biçimde kullanılmasının etkisiyle, iş dünyasında hızla gelişen olgulardan biri olarak karşımıza çıkmaya başlamıştır. Bankacılık, sigortacılık, marketçilik, perakendecilik, e-ticaret gibi birçok alanda veri madenciliği yaygın olarak kullanılmaktadır.

Bu çalışma, veri madenciliği tekniklerinin eğitim, sağlık, bankacılık, sigortacılık gibi sosyal içerikli alanların yanı sıra, imalat ve montaj sanayinin birçok alanında uygulanabilir olduğunu ortaya koymuştur. Çalışmanın, bu alanda bundan sonra yapılacak olan araştırmalar için ve yararlanmak isteyenlere bir kaynak olmasını ümit ediyorum.

Çalışmalarım sırasında, beni teşvik eden, engin bilgi ve deneyimlerini benden esirgemeyen değerli danışmanlarım sayın Prof.Dr. İsmet ÇEVİK ve sayın Yrd.Doç.Dr. Bayram TOPAL'a çok teşekkür ederim. Ayrıca uygulamaları yapmama fırsat tanıyan ve tedarikçi verilerini kullanmama izin veren TÜVASAŞ Genel Müdürlüğüne, özellikle satınalma biriminden sayın Eren BULDUR'a ve planlama daire başkanı sayın Metin BAYRAM'a teşekkürü bir borç bilirim. Analiz çalışmalarındaki katkılarından dolayı sayın Yrd.Doç.Dr. Ayhan DEMİRİZ'e ve SPSS Türkiye ofisinden sayın Selim DELİLOĞLU'na sonsuz teşekkür ederim.

Maddi ve manevi desteklerini her an hissettiğim aileme, çalışmalarım sırasında gösterdikleri sabır için sevgili eşim ve çocuklarıma da şükranlarımı sunuyorum.

20.03.2006 Aslan ÇOBAN

İÇİNDEKİLER

SİMGELER VE KISALTMALAR LİSTESİ

AID	: Automatic Interaction Detector
Ar-Ge	: Araştırma -Geliştirme
ERP	:Kurumsal Kaynak Planlama
DM	: Data Mining
ISO	: Uluslararası Kalite Standardı
KDD	: Knowledge Discovery in Databases
KK	: Kalite Kontrol
MRP	:Malzeme İhtiyaç Planlaması
OLAP	: On Line Analysis Processes
PC	: Personel Computer
SD	: Serbestlik Derecesi
TSE	: Türk Standartları Enstitüsü
TB	: Terra Byte
VM	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
$R	: C&R Tree Tahmini
$N	: Neural Network Tahmini
$C	: C5.0 Algoritması Tahmini

ŞEKİLLER LİSTESİ

TABLOLAR LİSTESİ

ÖZET

Anahtar Kelimeler: Veri, Veri Madenciliği, Tedarikçi İlişkileri Yönetimi

Bilginin temel yapısını oluşturan veri, son dönemde gelişen veri madenciliği kavramı ile daha bir önem kazanmıştır. Dünyada ve Türkiye'de veri madenciliğine olan ilgi ve yatırım büyük miktarlara ulaşmıştır. Dünyada perakendecilik – marketçilik, e-ticaret, bankacılık, sigortacılık, telekomünikasyon, sağlık ve eğitim alanlarında yaygın olarak kullanılan veri madenciliği, son dönemde Türkiye'de de özellikle marketçilik, banka ve sigortacılık ile e-devlet alanlarında kullanılmaya başlanmıştır.

Veri madenciliğinin üretim sektöründe kullanımı ise henüz yaygınlaşmamıştır. Buna gerekçe olarak bu alanda farklı tekniklerin kullanılması gösterilebilir. Ancak son zamanlarda veri madenciliği tekniklerinin, MRP ve ERP sistemleri ile birlikte kullanımı, olumlu sonuçlar vermeye başlamıştır. Hatta veri madenciliğini ERP sistemi içerisinde gösteren yaklaşımlar mevcuttur.

Bu çalışmada, veri madenciliğinin tanımı, kullanım alanları, model ve algoritmaları ayrıntılı olarak ele alınmıştır. Uygulama kısmında ise, üterim sektöründe faaliyet gösteren bir işletmenin gerçek verileri kullanılmıştır.

Birinci aşamada veriler düzenlenerek bir veri seti oluşturulmuş, daha sonra bu veri seti uygun model kurularak analiz edilmiştir. Analiz için SPSS Clementine 9.0 yazılımı kullanılmıştır. Elde edilen sonuçlar istatistik yöntemler kullanılarak test edilip, işletmenin tedarikçileri ile olan ilişkilerini etkileyecek anlamlı sonuçlar elde edilmiştir.

Son aşamada ise kurulan model; gerek verileri kullanılan işletmenin, gerekse benzer işletmelerin kullanabilecekleri dinamik bir yapıya dönüştürülmüştür.

Yaygın kullanım alanlarından farklı olarak, veri madenciliğinin üretim sektöründe de başarıyla kullanılabilir olduğunu göstermek, hem bu çalışmayı farklı kılmış, hem de bu alanda çalışmak isteyen araştırmacılara bir bakış açısı kazandırmıştır.

AN APPLICATION OF DATA MINING – AIDED SUPPLIER SELECTION MODEL IN MANUFACTURING INDUSTRY

SUMMARY

Key Words: Data, Data Mining, Supplier Relationship Management

Being the basic structure of knowledge, data has gained considerable importance with the emergence of the consept of data mining. Investment and interest in data mining has been growing and already reached big sums in the world as well as in Turkey. Data mining is used worldwide in various social and industrial areas such as retail marketing, e-commerce, banking, insurance, telecommunications, health and education. In Turkey, in recent years it is being utilesed especially in the areas of retail marketing, banking, insurance and e-state.

Using the data mining in manufacturing is not wide-spread for now. The reason for this is that so many different techniques for different areas are used in data mining. However nowadays, the usage of data mining techniques, with MRP and ERP systems is getting to give good results. In fact, there are some approacher which includes ERP systems in data mining.

In this resarch, the definition of data mining, the areas of its application, the models and the algoritms have been examined intensively. In the implemantation stage, real data taken from a manufacturing company have been used.

In the first stage, all data have been restored for creating a data-set then this set has been analyzed by using an appropriate model. For this purpose, SPSS Clementine 9.0 software has been used. The results obtained, have been tested using statistical methods and results making good sense and affecting the relations between the company and suppliers have been obtained.

Finally the model that is developed in this study has been given a dynamic structure that not only the company whose data were used benefits from it, but also similar companies can easily adapt for their applications.

Proving that data mining can be used in manufacturing area succesfuly has made this resarch, different from the others so, it has contributed a new point of view for the other researchers.

BÖLÜM 1. GİRİŞ

Bilişim teknolojilerinde yaşanan hızlı gelişmeler, günümüz işletmelerini rekabetin her an kıyasıya yaşandığı bir pazar yapısına götürmektedir. Bu yapı içinde "Bilgi" karşımıza yükselen bir değer olarak çıkmakta ve işletmelerin rekabet gücünü doğrudan etkilemekte; böylece, tıpkı diğer kaynaklar gibi bilginin de, işletmenin kalite politikaları içinde yönetilmesi gerekmektedir [1].

Günümüz işletmeleri için yaşam biçimi haline gelen bir olgu da rekabettir. Bu yeni yaşam biçimi, beraberinde pek çok yeni kavramı getirmiş, pek çok eski kavrama da yeni anlamlar yüklemiştir. Rekabetin işletmelere kazandırdığı en önemli kavramlardan birisi de 'bilgi'dir. Artık işletmeler, tozlu arşivlere kaldırılmış eski defterlerin içinde gerçek hazinelerin yattığını öğrenmişlerdir. Bilgi, işletmecilikte ilk öğretilen konular arasında yer alan üretim faktörleri içerisine stratejik bir değer olarak girmiş ve pek çok kurum başarının, bilginin paylaşımı ve verimli kullanılması ile eş anlamlı hale geldiğine tanık olmaya başlamıştır

Hızla gelişen bilgisayar, internet ve iletişim teknolojileri ekonomi başta olmak üzere, sosyal ve politik hayattan sağlık ve eğitim hayatına kadar kurumların ve insanların yaşamını etkilemektedir. İnternetin etkisi, elektronik ticaret olarak iş hayatına, ağlar üzerinde uzaktan eğitim olarak eğitime, uzaktan yapılabilen operasyonlar olarak tıbba yansımaktadır. Bu etki sanal değil gerçek bir etkidir. Bilgisayar, internet ve iletişim teknolojilerindeki baş döndürücü ilerlemeler sonucu piyasaya, her geçen gün birbirinden farklı yeni ürünler, hizmetler ve her kesime çok farklı dünyalar sunulmaktadır. Kim olursanız olun tüm bu yenilik ve ilerlemelerden bugün ya da yakın gelecekte etkilenmemeniz mümkün görünmemektedir[2].

1995 yılında birincisi düzenlenen "Knowledge Discovery in Databases" konferansı bildiri kitabı sunuşunda, enformasyon teknolojilerinin oluşturduğu veri dağları,

"Dünyadaki enformasyon miktarının her 20 ayda bir ikiye katlandığı tahmin edilmektedir. Bu ham veri seli ile ne yapmamız gerekmektedir? İnsan gözleri bunun ancak çok küçük bir kısmını görebilecektir. Bilgisayarlar bilgelik pınarı olmayı vaat etmekle, ancak veri sellerine neden olmaktadır" cümleleri ile vurgulanmaktadır[3].

Veri tabanı sistemlerinin artan kullanımı ve sakladıkları veri miktarlarındaki olağanüstü artış, organizasyonları elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır[4]. Geleneksel sorgu (Query) veya raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veri Tabanlarında Bilgi Keşfi -VTBK (Knowledge Discovery in Databases) adı altında, sürekli ve yeni arayışlara neden olmaktadır[5].

Desen tanıma ve sınıflama problemleri üzerinde yoğunlaşan yapay zekâ ve istatistik disiplinlerindeki gelişmeler veri madenciliğinin temellerini oluşturmaktadır. Ayrıca veri madenciliği, yapay zeka çalışmalarının uzantısı olan makine öğrenimi (Machine Learning) ve uzman sistemlerin (Expert Systems) yanı sıra, veri tabanları, optimizasyon, görselleştirme, yüksek performanslı paralel işlemciler gibi çeşitli disiplin ve teknolojilerdeki gelişmelerden de etkilenmektedir[1].

Makine öğreniminde birçok algoritma, bozuk yapılardan kalıbı ayırt etmeye çalışırlar. Her ne kadar bu algoritmaların seçici versiyonları aynı zamanda geliştirilse de, basit Bayesian sınıflandırıcıları ve sinirsel ağlar gibi bazı öğrenme algoritmaları, bütün mevcut özellikleri kullanırlar[6].

Her ne kadar veri madenciliği kullanımı hükümetlerin ve özel sektörün ham veriden bilgi çıkarma kapasitesini arttırsa da, veri madenciliği belirsiz bir terim olarak kalmaktadır[7].

Veri madenciliği, kamu ve özel sektörde yıllardan beri başarılı biçimde kullanılmaktadır. Özel sektörde bu uygulamalar, müşteri ilişki yönetimi, pazar araştırması, perakende satış ve tedarik zincir analizi, tıbbi analiz ve teşhis, mali analiz ve hile belirlemesini içerirken, kamuda veri madenciliği ilk olarak mali hile ve yolsuzlukları saptamada kullanılmıştır[8].

Veri madenciliği araçları aynı zamanda web madenciliği araçları olarak da kullanılmaktadır. Bu da web madenciliğinin veri madenciliğinin bir parçası olduğunu ve webdeki veri üzerinden madencilik yaptığını göstermektedir. Web ve veri madenciliği araştırmaları birlikte yürür hale gelmişler ve bu çalışmalarda bilgi yönetimi için iş modelleri ile bütünleşmiş web tabanlı veri madenciliği araçları kullanılmaya başlanmıştır[9].

Bu çalışmada, veri madenciliğinin tanımı, kullanım alanları, model ve algoritmaları tanıtılmış ve bir işletmenin tedarikçileri ile ilgili gerçek verileri, veri madenciliği çözümleri kullanılarak analiz edilmiştir. Elde edilen sonuçlar, işletmenin tedarikçileri ile olan ilişkilerine farklı bir bakış açısı getirmiş ve bundan sonra yapacağı mal alımları ile ilgili olarak yeni kriterler geliştirmeye yöneltmiştir.

BÖLÜM 2. VERİ, VERİ MODELLERİ, VERİ TABANLARI VE VERİ AMBARLARI

2.1. Veri Kavramı

Bilgi yönetiminin yapısının daha iyi anlaşılabilmesi için "veri-data", "bilgi-information" ve "kurumsal bilgi (çıkarım)-knowledge" kavramlarının açıklanması gerekmektedir.

Veri kelimesinin sözlük anlamı "gerçek"tir ve kökü latince bir kelime olan ‘datum’ dan gelmektedir. Veri bilginin doğası gereği sayısal veya alfa sayısal olmak üzere iki ana kategoriye ayrılabilir. Burada bahsi geçen bilgi ile veri, birbirleriyle alakalı kavramlardır. Bilgi, anlamlı bir şekilde derlenen ve birleştirilen veridir ve bu anlamda alınacak kararlar için gerçek bir değer niteliği taşır[10].

Diğer açıdan bilgi, belirsizliğin azaltılması için gereken özellikleri tanımlayan bir kavramdır. Bu açıdan bilginin kapsamına iletişim kanalının bir fonksiyonu olmak da girmektedir. Bilgi, anlamlı biçimde derlenen veridir. Veriye göre daha değerlidir ve gerçek bir değerdir.

2.2. Veri kaynakları:

- İçsel veri: Bu tip veriler insanlar, ürünler, servisler ve prosesler ile ilgilidir. Örneğin işçilere ait ödemeler muhasebe bölümünde, malzeme ve makineler ile ilgili veriler imalat bölümünde tutulmaktadır.

- Dışsal veri: Bu tip veriler uydular ve algılayıcılardan toplanan ticari verilerdir. Cd sürücülerden, internetten, film müzik veya seslerden, resimlerden,

televizyondan, grafik ve diyagramlardan elde edilen veriler bu kategoriye girer. Hükümet raporları, yerel bankalar, enstitüler, özel şirketler de önemli dışsal veri kaynaklarıdır.

- Personel Verisi: Nesnel satış tahminleri, rakiplerin neler yapabileceği ile ilgili fikirler, şirkete özgü haber portalları gibi işletmenin kendi uzmanlık bilgileriyle bir araya getirdikleri verilerdir.

Veri son olarak, bir kişinin formülleştirmeye veya kayıt etmeye değer bulduğu her şey olarak da tarif edilebilir. Veriyi tanımlamak için çok farklı kavram seçeneği mevcuttur. Bu kavramlar aşağıdaki gibi sıralanabilir:

- Veri (Data) : Herhangi bir özel anlam içermeyen, kayıt edilebilen, sınıflandırılabilen, depolanabilen, bir bilgi sistemine girilen, yapısal olmayan, işlenmemiş girdiler, nesneler, aktiviteler, işlemlerin tümüne denir. Veri sadece sayılar veya harfler değildir. Veri; sayılar, harfler ve onların anlamıdır. Veri hakkındaki bu veriye ‘meta data’ denir.

- Byte: En küçük adreslenebilir birim olan "bit"in 8 adedinin oluşturduğu bütündür.

- Veri Parçası: Alan veya veri elementi olarak da tanımlanabilecek veri parçası bir veya birden fazla byte'dan oluşan en küçük kimliklendirilmiş veridir.

- Veri Toplamı: Veri toplamı bir kayıt içerisindeki veri parçalarının birleşiminden oluşan bir bütündür. Örneğin tarih bir veri toplamı olarak düşünülecek olursa, bu veri toplamını oluşturan veri parçaları gün, ay ve yıl'dır.

- Kayıt: Kayıt, veri toplamlarının oluşturduğu bir bütündür.

- Kısım: Kısım terimi kayıt ve veri toplamı gibi veri bölümünü tarif eden iki tanımın gereksiz olduğuna inanan IBM gibi firmaların geliştirdiği bir kavramdır. Bu kavram kayıt ve veri toplamını kapsamaktadır.

- Dosya: Dosya, kayıtlar bütünüdür.

- Veri Tabanı: Veri parçaları, veri kayıtları ve bu kayıtlar arasındaki ilişkileri içeren bir bütündür.

- Bilgi (Information) : Herhangi birine söylendiğinde bireyin kafasında söylenen bu ifadeye ait bir anlam uyandıran, karar alma aşamalarında verilerin işlenip anlamlı hale getirilerek kullanıcıya sunulmuş halidir. Veri bilginin hammaddesidir. Veriyi bilgiye çevirmeye veri analizi denir. Verilerin belirli sonuçlara ulaşmak üzere işlenmesi ve anlam kazanması sonucu elde edilen bilgiler, yönetim süreçlerinin ve karar alma mekanizmalarının temel girdisi olmaktadır[11].

- Kurumsal Bilgi - Çıkarımı (Knowledge) : Belirli bir amaca yönelik olarak bilginin çeşitli analiz, sınıflama ve gruplama işlemlerinden geçirilerek, gerektiği zamanlarda potansiyel olarak kullanıma hazır hale getirilmesidir[12].

Türkçe'de günlük kullanımda bilgi sözcüğü ile hem 'Information', hem de 'Knowledge' ifade edilmekte olduğundan ve henüz kurumsal bilginin (çıkarımın) örgüt içinde kullanımı yaygınlaşmadığından, kavramların ifade edilmesi sırasında güçlükler yaşanmaktadır. Bu karmaşa ile 'Yönetim Bilgi Sistemleri' olarak adlandırılan 'Management Information System' kavramında da karşılaşılmaktadır. Buradaki kavram tek sözcük olarak değil, bir bütün olarak ele alınmalı ve 'Information' bilişim anlamında kullanılmalıdır

2.3. Veri Modelleri

Veri modeli, veriyi bir kurala göre yapılandırma şeklidir. Bu yapılandırma içerisinde iki unsur bulunur. Bu unsurlar; yapı ve işlemlerdir. Yapı; sistemin veriyi yapılandırma şeklidir. İşlemler ise kullanıcıların veri tabanındaki veriyi düzenleme imkânlarıdır.

Dünyanın tüm özellikleri bir model tarafından yansıtılamaz. Eğer bir model uygun olarak formülleştirilmişse kullanıcıların ihtiyaçlarını karşılayabilir. Modellerin

eksiklikleri iki grup altında toplanabilir. Birincisi, veri yapısının bir bölümünün temsil edilmemesi ve ikincisi çeşitli yollarla veri yapısı üzerinde değişiklik yapılamamasıdır[13].

Bir veri modeli, verinin hangi kurallara göre yapılandırılacağını belirler. Fakat yapılar verinin anlamı ve nasıl kullanılacağı hakkında tam bir açıklama vermezler. Veri modeli veri tabanında bulunan verilerin mantıksal organizasyonunu belirleyen kurallar kümesi olarak tanımlanabilir. Veri modelleri;

- Basit Veri Modelleri
- Geliştirilmiş Veri Modelleri, olarak ikiye ayrılır[14].

2.3.1. Basit veri modelleri

Basit veri modellerindeki amaç, verinin basit, anlaşılabilir bir yapıya sokulmasıdır. Bunlar genel yapılardır. Basit veri modelleri daha çok programlamaya dayalı bir veri modelidir. Dosyalama sistemleri oluşturmak amacıyla kullanılmaya başlanan veri modelidir. Aynı zamanda bilgisayarlarda veri işleme ihtiyacının ortaya çıkması ile dosyalama sistemleri oluşturmak amacı ile kullanılmaya başlanan veri modelleridir. Basit veri modelleri;

- Hiyerarşik veri modeli,
- Ağ veri modeli, olarak ikiye ayrılmaktadır.

2.3.1.1. Hiyerarşik veri modeli

Hiyerarşik veri modeli bir ağaç yapısı şeklindedir. Ayrıca hiyerarşi sıralamasında üstteki varlıklar ebeveyn, alttakiler ise çocuklar olarak isimlendirilir. Hiyerarşik modelleme tekniği varlıklar arasında l:n (bire çoklu) ilişki tiplerinin bulunduğu verilerin modellenmesi esnasında kullanılır. Bu teknikteki 1 kısmındaki kayıt tiplerine baba, n kısmındaki kayıt tiplerine oğul adı verilir. Oğullarında oğulları tanımlanabiliyorsa düğüm adını alır. Hiyerarşik bir şema kurulurken aşağıdaki şu özelliklere dikkat etmek gerekmektedir[15].

Hiyerarşik şemanın kökü olarak tanımlanan kayıt tipi, başka bir Baba - Oğul ilişkisinde oğul kayıt tipi olarak yer almaz. Kök dışındaki her oğul kayıt tipi sadece bir Baba - Oğul ilişkisinde yer alır. Bir kayıt tipi baba kayıt tipi olarak birden fazla Baba - Oğul ilişkisinde yer alabilir. Herhangi bir Baba - Oğul ilişkisinde baba olarak tanımlanmamış bir oğul kayıt tipi, hiyerarşik şemanın dalı olarak tanımlanır.

Hiyerarşik veri modelleme tekniğinde karşılaşılan en önemli sorunlardan biri m : n (çoklu) ilişkilerin veri tabanına aktarılmasıdır. Diğer bir problemde bir kayıt tipinin birden fazla Baba - Oğul ilişkisinde oğul kayıt tipi olarak bulunmasıdır. Hiyerarşik veri modellerinde çoklu ilişkileri temsil edebilmek için varlık tiplerinin her ilişkinin ayrı ayrı tanımlanması gereklidir. Bu ise gereksiz veri tekrarı demektir.[16].

Hiyerarşik veri modeli bir ağaç yapısını andırmaktadır. Modelde herhangi bir düğüm altındaki n sayıdaki düğüme bağlanabilirken, üzerinde sadece bir düğüme bağlı bulunabilir. En üst noktadaki düğüm ise kök adını almaktadır. Bu veri modelinin temeli veriler arasındaki bire çok ilişkilerin "aile - çocuk" şeklinde tanımlanmasına dayanır. Bu yapıda ailenin çok sayıda çocuğu olabilir ancak bir çocuğun birden fazla ailesi olamaz.

2.3.1.2. Ağ veri modeli

Hiyerarşik veri modelinin basit yapılı olmasına rağmen tek bir kökün olmadığı durumlarda modellemede sorunlar çıkmaktadır. Aynı zamanda ilişki tipleri ikili, yani varlık arasında kurulan birebir ilişki söz konusudur. Ağ veri modeli iki varlık arasında bire çoklu ilişkiden oluşan küme kavramını kullanır. "Bir" tarafında olan varlık kümenin sahibi, "Çok" tarafında olan varlık ise kümenin üyesidir. Bir üye başka bir kümenin sahibi olabilir. Fakat bir varlık aynı tipte iki kümeye birden üye olamaz. Buna karşılık bir üye aynı tipte olmayan veya daha fazla kümeye sahip olabilir[17].

1970'li yılların başında çoğu ticari veri tabanı uygulaması ağ modelleme tekniği yardımı ile oluşturulmuştu. Ağ veri modeline CODASYL (Conference on Data

System Languages) veri modeli adı verilir. Bu modelde varlıklar arasında bire çoklu ilişkiler vardır. Ağ modelinde işlemler oklar yardımı ile ifade edilir. Okun yönü her zaman sahipten üyeye doğrudur[18].

Ağ tekniğinin hiyerarşik teknikten ayrılan yanları şöyle sıralanabilir:

- İki düğüm arasında birden fazla bağlantı olması,
- Kök olarak tanımlanacak bir düğüm noktasının tanımlanamıyor olması,
- Bir düğümün birden fazla baba kaydına sahip olmasıdır.

Bu özelliklere sahip verilerin modellenmesi için ağ modelleme tekniği, hiyerarşik modelleme tekniğinden daha uygun imkânlar sunmaktadır. Ağ veri modelleri, tablo ve grafik temellidir. Grafikteki düğümler varlık tiplerine karşılık gelir ve tablolar şeklinde temsil edilir. Grafiğin okları, ilişkileri temsil eder ve tabloda bağlantılar olarak temsil edilir.

2.3.2. Geliştirilmiş veri modelleri

Varolan bir verinin üzerinde bilgisayar kullanarak işlem yapabilmek için o verinin bilgisayarda işlenmesi yeterli değildir. Burada aynı zamanda kullanıcıların ve veri üzerindeki işlem yapacak analist1erin bakış açıları da çok önemlidir. Tüm kullanıcıların farklı açılarını bütün1eşik bir model ile veri tabanına yansıtılması veri modeli oluşturmaktadır. Geliştirilmiş veri modelleri;

- Varlık - İlişki Veri Modelleri
- İlişkisel Veri Modelleri
- Nesne Yönelimli Veri Modelleri, olarak sıralanabilir.

2.3.2.1. Varlık - ilişki veri modeli

Varlık - İlişki işlemi, analizler ve şemalandırma için önemli bir tekniktir. Organizasyonun veri ve gereksinimlerinin yukarıdan aşağıya planlamasında kullanılır. Bu varlık - ilişki şeması, işletme açısından önemli olan iş varlıklarının

gösterildiği bir grafiktir. İş varlıkları olarak, bilgi saklanan birimlerden ya da basitçe, ilişkilerden söz edilebilir. Varlık gerçek veya soyut, kesin, görülebilir veya görülemez olabilir. Görülebilir varlıklara müşteri, çalışan, fatura ve bölüm örnek verilebilir. Görülemez varlıklara ise olay, iş adı, zaman periyodu ve kazanç merkezi örnek verilebilir. Kayıt etmek istenilen bir varlık, renk, boyut, maddi değer, yüzdelik değerlendirme, adres, maaş, tarih, kod veya cinsiyet gibi özniteliklere sahip olabilir. Varlıklar arasındaki ilişki, şema içindeki varlıkları çizelgeler ile bağlayarak gösterilebilir. Her çizgi bir çift mevcut arasındaki ilişkiyi gösterir. Varlıklar arasındaki ilişkilerin üç önemli çeşidi vardır. Bunlar;

- Bire bir ilişki
- Bire çoklu ilişki
- Çoklu ilişki

Bire bir ilişki: Bir varlıktan diğerine bire bir ilişkiler, birinci varlığın her bir değeri ikinci varlığın sadece bir değeri ile eşleşir.

Bire çoklu ilişki: A varlığından B varlığına bire çoklu ilişki, A varlığının bir değeri B varlığının sıfır bir veya birçok değerleriyle herhangi bir zamanda ilişkilendirilmiş olduğu anlamına gelir.

Çoklu ilişki: Bazı durumlarda, bir varlık - ilişki şemasında çoklu ilişkilere ihtiyaç duyulur. Örnek olarak, bir çalışanın sıfır, bir veya birçok projeyle ve bir projenin sıfır, bir veya birçok çalışanla nasıl birleştirilebileceğini göstermektedir[19].

Balon grafikler, verilerin bireysel kullanıcı veri görünümlerini belgelendirmek için kullanılan araçlardır. Balon grafikler ayrıca birçok kullanıcı görünümlerin birleşmesinden oluşan mantıksal veri modellerini belgelendirmekte de kullanılır. Veriden bahsederken kullanılan terimin, veri öğesini mi, genel bir veri kategorisini mi, belirli bir veri değerini mi, yoksa bir oluşumu mu ifade ettiği kesin olarak belirtilmelidir.

2.3.2.2. İlişkisel veri modeli

İlişkisel veri modeli tablolardan oluşur. Tablolar ilişki olarak isimlendirilir. Tablolar arasında ortak olan sütunlar ile ilişkiler sağlanmış olur. Tablolar iki boyutludur, satır ve sütunlardan oluşur. Tablolarla ilgili bir takım kurallar vardır; her sütunun kendine özgü bir ismi olmalıdır ve o sütundaki veriler sütun ismi ile uyumlu olmalıdır. Aynı şekilde her satırda bir diğerinden farklı olmalıdır. Avantaj ve dezavantajları da şu şekilde sıralanabilir: Diğer modellere göre uygulaması daha kolaydır. Kullanıcının alttaki veri tabanı yapısını bilmesine gerek yoktur. Ayrıca veri bağımsızlığı diğer veri modellerindekinden daha fazladır[20].

İlişkisel modelde ayrıştırılabilecek tek bir veri tipi vardır; o da ilişkidir. İlişkisel veri tabanı bir ilişkiler koleksiyonudur. Tüm sorgular bu ilişkiler üzerine kuruludur. Tek çeşit bileşik verinin olmasının nedeni, yeni veri eklemenin karmaşıklığını da beraberinde getirmesindendir. Bu durum sorgulama dilinin bir gereğidir. Bu tip bir dilde dört temel komut vardır; elde etme, yerleştirme, güncelleştirme ve silme. Eğer sayısız bileşik veri olsaydı her bir bileşik veri için bu dört temel komutun uygulanması gerekecekti.

İlişkisel modelde her şey özellikleri tanımlayan sütunlar ve nesneleri veya kişileri tanımlayan bilgilerin yer aldığı satırlardan oluşan basit bir tablodur. Çoklu ilişkiler bile tablolar yardımı ile basitçe gösterilebilir. Modelde ilişkileri tablolarla ifade etme görsel açıdan da tercih edilebilir olsa da kolaylık açısından kısa notasyonlar kullanmak tercih edilmektedir. Bu notasyonda ilişki, isim ile tanımlanır ve tanım kümelerinin isimleri parantez içerisinde verilir.

İlişkisel veri modelinin kendinden öncekilere göre önemli avantajı suni yapılara gerek duymamasıdır. Matematiksel ilişki temeline dayandığı için yapılandırılması daha kolaydır. İlişkisel veri modelinin bu kadar popüler olmasının sebebi veri üzerinde yapılan işlemler için kullanılan sorgu dilinin güçlü ve basit olmasıdır. İlişkisel model kullanıcı dostu olarak tanımlanabilecek bir modeldir. Sıradan bir kullanıcı bile tablo yapısını anlayabilir. Modelin güçlü matematiksel altyapısıyla

birlikte mantıksal arabirimi üzerinde durması onu çok kullanılan bir model haline getirmiştir.

2.3.2.3. Nesne yönelimli veri modeli

Nesne yönelimli veri modeli ilişkisel modelle karşılaştırıldığında yüksek seviyeli bir modeldir. Çünkü nesne yönelimli veri modeli ilişkisel modelde zor olan hiyerarşiler gibi yapılandırılmaları hızlandırmaktadır. Nesne yönelimli veri modelini önemli kılan bir başka özellik ise verilerin harmanlanması için özel bir yapı sunmasıdır.

Nesne yönelimli veri modelinde her şey bir nesnedir. Bir nesne olayı ise gerçek dünyadaki öğelerin bir yansıması olan spesifik bir nesnedir. Bir nesne sınıfı ise hepsi birbirine benzeyen nesne olayları için tanımlamaları içerir. Her şey nesne olarak tanımlandığından karmaşık nesne yapıları sistem tarafından mı yoksa kullanıcı tarafından mı dizayn edildiği tartışmasına gerek bırakmadan yapılabilir. Her nesne sistemce atanmış tek bir tanımlayıcıya sahiptir.

Nesne yönelimli sistemler farklı sistemler ve metodolojiler için kullanılmıştır. Genel olarak bu sistemler gerçek dünyadaki objeleri nesne denilen varlık şeklinde modellemeyi temel almaktadır. Nesneler ortak karakteristikler içeren nesnelerin bulunduğu sınıflar içerisinde gruplandırılırlar. Sınıf hiyerarşisi kavramı ise insanları nesneleri global anlamda kategorize etmelerini sağlamak için geliştirilmiştir.

Nesne yönelimli, başka bir deyişle semantik veri modelinin özellikleri şöyle sıralanabilir:

- Nesne kimliği,
- Karmaşık nesneler,
- Tip hiyerarşisi.

Nesne yönelimli yaklaşımın en önemli özelliği gerçek hayat durumlarının modellenmesinde daha doğal bir yol sağlamasıdır. Böylece karmaşık ilişkiler daha kolay anlaşılabilir, kolay düzenlenebilir bir çerçeve içinde ele alınabilir.

Nesne yönelimli sistemlerin çok farklı uygulama alanları olmasına rağmen belirgin özellikleri vardır. Bunlar:

- Nesne birimlerinin ve sınıflarının desteklenmesi
- Veri işlemlerinin kapsüllenmesi
- Nesnelerin özelleştirilmesi.

2.4. Veri Tabanları

Veri tabanı bir veya daha fazla uygulamaya hizmet vermek için bir araya toplanmış birbirleriyle ilişkili veriler toplamıdır. Veri tabanı sadece verinin alınması değil aynı zamanda o veri üzerinde değişiklik yapılmasına da imkân vermektedir. Veri tabanı bilgisayarda veri depolamak ve işlemek amacıyla kullanılmaktadır. Veri tabanı, çeşitli tiplerdeki varlıklara, bu varlıkların özniteliklerine ve bunlar arasındaki ilişkilere ev sahipliği yapan bir yapıdır. Bir veri tabanında soyutlama katmaları kullanılarak gerçek dünyanın kavramları bilgisayar ortamına adapte edilebilmektedir[21].

Fiziksel veri tabanı, disk üzerinde bulunan dosya ve indeks koleksiyonu ve bunlara ulaşmak için kullanılan depolama yapılarıdır. Kavramsal veri tabanı, gerçek hayatın bir soyutlamasıdır. Bu soyutlamayı gerçekleştirmek için veri tabanı yönetim sistemi, bir veri tabanı tanımlama dili kullanır. Veri tanımlama dili kavramsal veri tabanını veri modeli olarak tanımlayabilmemizi sağlar. Kavramsal veri tabanı, organizasyon tarafından kullanılan verinin bütünü temsil eder[22].

Görünüm katmanı veya alt şema, kavramsal veri tabanının bir parçasıdır. Bu alt şemalar veri düzenleme dili ile oluşturulabilir. Görünümler, bir veri tabanında güvenliğin sağlanması için de önemlidir. Çünkü bir görünüm, kullanıcının veri tabanının sadece bir bölümünü görmesine izin vermektedir.

Geleneksel dosya yönetim sisteminin dezavantajları şu şekilde sıralanabilir:

- Veri tekrarına yol açması

- Aynı verinin farklı kullanıcılar tarafından aynı anda güncellenmeye çalışılması. Veri bütünlüğü bu durumda tehlikeye girer.
- Veri depolama alanının gereksiz yere kullanılması.
- Veriye ulaşmak için kullanılan dilin kullanılan programa özgü olması ve kullanışlı olmaması.

Bir veri tabanı sistemi ise;

- Veri tekrarını önler.
- Veri bütünlüğünü sağlar.
- Veri depolama alanı ihtiyacım azaltır.
- SQL gibi bir standart sorgulama dili kullanılarak veriye ulaşabilmeyi sağlar[23].

Bununla birlikte veri tabanı sisteminin dezavantajları da bulunmaktadır. Bunlar şu şekilde sıralanabilir:

- Kurulum ve bakımının zor ve pahalı olması
- Bütünleşik sistemdeki bir bölüm veriye ulaşamamak tüm sistemin çalışmamasına sebep olur.

İşletmeleri veri tabanı yaklaşımına götüren pek çok problem mevcuttur. Bunlardan bazıları şunlardır:

- Basit ihtiyaçlara çabuk yanıtlar alınamaması.
- Düşük veri kalitesi ve doğruluğu
- Değişime hızlı ayak uyduramama
- Yüksek gelişim maliyetleri
- Gerçek dünya için geçersiz veri modeli kullanımı[24].

Veri tabanı sistemlerinin başlıca üç özelliği vardır:

Özerklik: Bir veri tabanı diğer veri tabanlarıyla etkileşimde olmak için kendi kontrol politikasını oluşturabilir.

Heterojenlik: Veri modelleri, sorgulama dilleri, veri tabanından veri tabanına farklılık gösterebilir.

Dağıtım: Fiziksel olarak farklı ortamlarda yerleşmiş bulunan veri tabanları.
Veri tabanlarını;

- Hiyerarşik veri tabanları,
- İlişkisel veri tabanları
- Nesne yönelimli veri tabanları, olarak başlıca üç kısımda incelemek mümkündür[25].

2.4.1. Hiyerarşik veri tabanları

IBM'in yönetim bilişim sistemleri ile popüler olan hiyerarşik veri tabanı yönetim sistemleri, kullanıcılara verileri ağaç yapısında modellemesine olanak tanımaktadır. Bu ağaç yapısı bir köke bağlı bir veya daha fazla daldan oluşmaktadır.

Hiyerarşik veri tabanları içindekileri bir ağaca benzeyen hiyerarşik model içinde organize eder. Hiyerarşik ağaç sadece bir veri tabanı içindeki veri elemanlarını tanımlamakla kalmaz, ayın zamanda bu veri elemanları arasındaki ilişkiyi de tanımlar. Çeşitli hiyerarşik veri tabanı modelleri mevcuttur. Bunlar; bire- bir, bire çok ve çoklu şeklindedir.[21].

2.4.2. İlişkisel veri tabanları

Verileri satır ve sütunlardan oluşan iki boyutlu bir tablo şeklinde organize eden veri tabanlarıdır. İlişkisel veri tabanlarının kullanılması hiyerarşik veri tabanlarına nispetle bir takım kolaylıkları da beraberinde getirmiştir. Bu kolaylıklar şu şekilde sıralanabilir;

- Yeni veri tabanı kayıtları veri tabanına sokulabilir ve bir öğenin herhangi bir parçası değiştirilebilir veya silinebilir.
- Öğenin kayıt numarası ve alan ismine bakmak suretiyle veri tabanı içerisindeki verinin yerinin bulunmasının kolaylaşması.
- Veri kayıtlarını sıralama ve yeniden düzenleyebilme ve veri elemanlarını birleştirebilme[26].

Bu kolaylıkların yanı sıra içinde bazı kısıtları da barındırmaktadır. Bu kısıtlar şu şekildedir:

- Veri türlerinin tanımlanmasındaki yetersizlikler nedeni ile uygunsuzluk. Çoğu veri tabanı sistemi tamsayı, karakter dizisi, tarih gibi standart türdeki veri tiplerini içerir. Kullanıcının kendisi yeni veri tipleri ve bu veri tipleri arasındaki yeni işlem tiplerini tanımlayamaz.
- Karmaşık bir veri yapısını, çok yapısal olan hiyerarşik veya ilişkisel olarak gösterememe.
- Örneğin, C gibi programlama dillerine göre verilerin idaresini yapan yordamların zayıflığı.
- Bir yapıyı modellemedeki zayıflık: İlişkisel sistem klasik satır sütun yapısındadır. Bu yaklaşımda karmaşık sorgulama ve işlemlerin yapılamayacağı açıktır.

İlişkisel modelin en önemli amacı son kullanıcının ve veri tabanı tasarımcısının işini kolaylaştırmaktır. Kolaylaştırıcı bir önlem olarak da veri basit bir tablo yapısında tutulmaktadır. Veri tabanı ise tablolar bütünü olarak gözükmektedir[27].

2.4.3. Nesne yönelimli veri tabanları

Nesne yönelimli veri tabanı, bir veri tabanının kavramsal modelinin karmaşık nesnelerini oluşturan ve kullanabilen bir veri tabanıdır[28]. Nesneye dayalı veri tabanı seçiminin ve kullanımının sebebi, veri tabanının yeterliliğinin arttırılmış olmasıdır. Bunlar; yüksek yeterliliğine sahip yüksek düzeyde sorgu dili, giriş ve

düzeltme işlemleri, aynı zamanda olabilirlik kontrolü ve geri alma, karmaşık nesnelerin saklanması, indeksler ve hızlı erişim imkânlarıdır.

Pek çok nesne yönelimli veri tabanı sistemi geliştirmeleri başlıca şu araştırma alanları üzerinde yoğunlaşmıştır:

Bilgi dağarcığı sunumu: Yapay zekâ kavramı içinde düğüm ve bağlantıların kullanılarak bilgi dağarcığı meydana getirilmesi ile ilgili anlam bilimsel ağlar ve çerçeve bazlı sistemler.

Anlam bilimsel veri modelleme: Şema tasarımında basit kalan ilişkisel modelin yapısal eksikliklerini tamamlamak için geliştirilen bir model olmuştur.

Nesne yönelimli programlama: Soyutlama ve dil desteği ihtiyacı, soyut veri tipleri ve nesne yönelimli programlama dilleri ile ilgili çalışmaları teşvik etmiştir.

2. 5. Veri Ambarları

Veri ambarcılığı çeşitli şekillerde tanımlanmıştır. Veri ambarcılığının babası sayılan Bill Inmon veri ambarını 1992' de şu şekilde tanımlamıştır: Veri ambarı, yönetimin karar sürecini desteklemede kullanılan, konuya yönelik, entegre, zamana bağlı, kalıcı veri topluluğudur[29].

Başka bir tanıma göre ise veri ambarı, basitleştirilmiş biçimde hareket sistemlerinden özetlenen ve kümelenen verinin saklandığı yerdir. Son kullanıcılar için yapılan veri erişimi ve raporlama araçları kullanıcıların karar desteği için veriyi elde etmelerini sağlamaktadır[30].

Veri ambarlarının üç çeşidi vardır:

- Tüm kuruma hizmet eden kurumsal (geleneksel) veri ambarı,
- İşletmedeki belirli bir iş birimini veya bölümü desteklemek üzere tasarlanmış minyatür bir veri ambarı olan veri pazarı (data mart),

- Veri ambarı tekniklerinin hareket sistemlerine uyarlandığı operasyonel veri deposu.[31].

2.5.1. Veri ambarının karakteristik özellikleri

1. Konuya Yönelik Olma: Operasyonel veri ihtiyacı, uygulamanın anlık ihtiyaçları ile ilgilidir ve o anda geçerli iş kurallarına dayanır. Veri ambarı dünyası ise müşteri, mal veren, ürün ve etkinlik gibi temel konular etrafında organize olur. Veri ambarındaki veri karar vermeye yöneliktir ve zaman derinliği çok daha fazla olduğundan daha karmaşık ilişkilere olanak tanır.

2. Veri Entegrasyonu: Sitemlerden veri ambarına veri aktarılırken veri entegre edilir ve hepsi aynı formata getirilir. Böylece değişik kaynaklardan gelen veri, veri ambarında tek ve genel olarak üzerinde anlaşmaya varılmış bir şekilde yer alır. Veri ambarındaki veri, temiz, geçerliliği onaylanmış ve uygun biçimde kümelenmiş olmalıdır. Benzer biçimde verinin doğru olması da gerekmektedir.

3. Kalıcı Ortam: Veri ambarında veri her zaman eklenir, üzerine yazılmaz. Veri tabanı sürekli olarak yeni veri alır ve eskisiyle entegre eder. Veri ambarlarında veri yükleme belirli zamanlarda yapılır ve kullanıcılar ancak bundan sonra veriye erişebilirler.

4. Zamana Bağlı Olma: Veri ambarındaki veri referans alınan zaman birimi ile birlikte kaydedilir ve veri bir kez doğru biçimde kaydedildikten sonra kullanıcılar tarafından güncellenemez. Veri ambarındaki veri tipik olarak 3-10 yıllık bir zaman dilimini kapsar[32].

2.5.2. Veri ambarının yapısı

Veri ambarında dört tip veri bulunur. Bunlar;

- Şu anda geçerli detay veri,

- Eski detay veri,
- Özetlenmiş veri,
- Meta Data, [33].

Bu veri tiplerinin hepsi aynı kayıt ortamında saklanmak zorunda değildir. Ancak kullanılan yazılım tümüne erişebilmelidir.

- Şu Anda Geçerli Detay Veri: Şu anda geçerli detay veri en son olan olayları içerir. Bu veri en alt düzey tanelikte saklanırsa çok yer tutabilir. Bu yüzden bu veri genellikle şu anda geçerli hareket verisinin temizlenip veri ambarına yüklenmiş biçimidir.

- Eski Detay Veri: Çoğu veri ambarında detay verinin belirli bir zaman sonra diskten bir büyük veri saklama ortamına aktarılmasını öngören kurallar vardır. Detay veriye hala erişilebilir ancak daha yavaş bir ortamda olduğu için erişim hızı yavaştır.

- Özetlenmiş Veri: Standart rakamları önceden tahmin edip veriyi buna göre özetlemek veri ambarının sorgulamalara daha hızlı cevap vermesini ve daha çok kullanılmasını sağlar. Ayrıca özetlemeler sayesinde hesaplamalar tekrar tekrar yapılmak zorunda kalmaz, ancak daha fazla veri saklama kapasitesi gerekir.

- Meta Data: Veri ambarının en önemli bileşenlerinden biri meta datadır. Veri ambarında verilerin tanımlandığı kısımdır. Meta data "veri hakkında veri" anlamındadır. Meta data her veri elementinin anlamını, hangi elementlerin hangileriyle nasıl ilişkili olduğunu ve kaynak verisi ile erişilecek veri gibi bilgileri içermektedir.

2.5.3. Veri ambarı projelerinde kaçınılması gereken hatalar

Veri ambarları işletmenin birçok bölümünü ilgilendiren en pahalı yatırımlardır. Bu projenin başarılı olabilmesi için sektörün tecrübeli veri ambarı proje yöneticilerinin ve bilgi sitemleri yöneticilerinin uyarılarını bilmek oldukça yararlı olabilecektir. Veri

ambarı enstitüsü bir yayınında, veri ambarı projelerinde kaçınılması gereken pek çok hatadan birkaçını şu şekilde sıralamaktadır:

- Yanlış sponsorluk zinciriyle işe başlamak,
- Ulaşılamayacak hedefler koymak ve gerçekler ortaya çıktığında üst düzey yöneticileri hüsrana uğratmak,
- Yalnızca elde olduğu için veri ambarına enformasyon yüklemek,
- Veri ambarı veri tabanı tasarımını, hareket veri tabanı tasarımıyla bir tutmak,
- Kullanıcıya değil, teknolojiye önem veren bir veri ambarı yöneticisi seçmek,
- Geleneksel olarak işletme içinde üretilen veriye odaklanıp, veri, metin, resim ve hatta ses ve video görüntülerinden oluşan harici verinin potansiyel değerini ihmal etmek,
- Üst üste çakışan veya kafa karıştırıcı şekilde tanımlanmış veri sağlamak,
- Performans, kapasite ve ölçeklenebilirlik sözlerine inanmak,
- Veri ambarı bir kez kurulduktan sonra tüm sorunlarımızın hallolduğuna inanmak[34].

2.5.4. Veri ambarı ihtiyacı

Bir işletmenin büyüklüğü veri ambarı ihtiyacının bir ölçüsü değildir. İşletmenin bir veri ambarına ihtiyacı olup olmadığına karar verirken işe bazı anahtar göstergelere bakarak başlanabilir[35]. Bu göstergelerden bazıları şunlardır:

- İşletme değişen, rekabetin çok yoğun olduğu bir pazarda faaliyet gösteriyorsa,
- Müşteriler hakkında sağlıklı enformasyon elde etme ihtiyacı varsa,
- Kazanç sağlayacak ve/veya verimliliği arttıracak enformasyona dayalı ürünler veya hizmetler oluşturma fırsatları varsa,
- Sık kullanılan ve birbiriyle ilişkili kurumsal veri birçok değişik yerde ve farklı sistemlerde bulunuyorsa,
- "Aynı veri ama farklı sonuç" şeklindeki sorun işletmede sürekli bir rahatsızlık haline gelmişse,

- Gerçek karar destek sistemlerine ihtiyaç varsa,
- Kullanıcılar daha etkili ve anlık sorgulama ve raporlama yapmak istiyorlarsa,
- Bir enformasyon dağıtımı alt yapısına ihtiyaç varsa.

Veri ambarı kurma kararı veren firmaların karakteristiklerine bakıldığında şu özellikler göze çarpmaktadır:

- Yönetimde enformasyona dayalı bir yaklaşım,
- Rekabetin yoğun olduğu, hızla değişen pazarlarda bulunmak,
- Çok sayıda ve farklı özelliklerde müşterilere sahip olmak,
- Verinin farklı sistemlerde bulunuyor olması,
- Aynı verinin farklı sistemlerde değişik biçimlerde gösteriliyor olması,
- Verinin çok teknik ve deşifresi zor biçimde kayıtlı olması[35].

2.5.5. Veri ambarının çeşitli sektörlerde kullanımı

Veri ambarı pek çok sektörde kullanılmaktadır. Aşağıda çeşitli sektörlerde veri ambarının nasıl kullanıldığı açıklanmaktadır.

2.5.5.1. Finans sektörü

Bankalar, sigorta şirketleri, leasing, factoring ve borsa şirketleri gibi kurum1arı içerisine alan bu sektörde kuruluşlar, değişik hesap kartlarında ve ayrı ayrı veri tabanlarında birçok değişik veriyi tutmak zorundadır. Tamamı ayrı formatlarda bilgilerin tutulduğu bu ortamlarda tek bir kişiye ait tüm verilere u1aşmak uzun saatler alan bir uğraştır. Veri ambarı verileri tek bir noktada toplamanın avantajı ile müşterilerin tüm bilgilerine doğrudan u1aşma imkânı verir.

Veri ambarı kuran pek çok banka, müşterilerinin neredeyse yarısından zarar ettiklerini ve karlarını oldukça orantısız oranlarda az müşteriden elde ettiklerini keşfetmişlerdir. İncelenen pek çok grup için %20 - %80 biçimindeki Pareto analizi prensibinin geçerli olduğu görülmüştür. Veri ambarı bankanın dikkatini karlı

müşterilerine yöneltmesini sağlayabilmektedir. Bankalar hızla tek kişiden oluşan pazar bölümlemesi kavramına doğru gitmeye başlamışlardır.

2.5.5.2. Üretim sektörü

Dayanıklı ve dayanıksız tüketim mamu1leri sektöründe veri ambarları, yedek parça envanterinin düşürülmesi, müşterileri ve kullandıkları ürünleri tanıma, parça ve kalite yönetimi, garantili ürünlerin yönetimi gibi birçok konuda kullanılabilmektedir.

2.5.5.3. Ulaşım sektörü

Ulaşım sektöründe veri ambarı genellikle müşteri eğilimlerini anlama ve onlara daha iyi hizmetler geliştirebilme konusunda kullanılmaktadır. Rezervasyon kayıtlarının incelenmesiyle, rezervasyon yaptırıp biletini a1mamayı alışkanlık haline getirmiş kişiler belirlenebilmekte ve kara listeler oluşturulmaktadır. Tanıtım broşürleri ve özel kampanya duyuruları gibi malzemelerin sadece ilgili kişilere gönderilmesi ile önemli ölçüde tasarruf sağlanabilmektedir.

2.5.5.4. Kamu sektörü

Bu alandaki en önemli uygulama vergi dairelerinde olmaktadır. Devletin ana gelir kaynağı olan vergiler ve bunları ödemekle yükümlü mükellefler iyi bir şekilde analiz edilerek devletin vergi alma etkinliği arttırılabilir.

Kamuda veri ambarı kullanımı çok çeşitlidir. Polis teşkilatlarında, sosyal hizmetler müdürlüklerinde, nüfus dairelerinde, yerel yönetimlerde ve planlama teşkilatlarında veri ambarı kullanılmaktadır. Ancak bu uygulamalar genellikle devletlerin gizlilik prensipleri nedeniyle pek açıklanmadığından detayları bilinmemektedir.

2.5.5.5. İletişim sektörü

Genellikle telefon şirketlerini bünyesinde bulunduran bu sektör içerisinde yer alan kuruluşlar, sahip oldukları teknolojik altyapı ve veri ambarları sayesinde

kullanıcıların taleplerini zamanında tespit edip onların ihtiyacı olan hizmeti götürebilme şansına sahip olabilmektedir.

Veri ambarı kuran bir çok telekomünikasyon şirketinin amacı, görüşme tarifelerini dengelemek ve telefon ağından geçen trafiği arttırmak için boş zamanlarda daha çok kullanımı teşvik ederek varlıklarının kullanımını geliştirmek olmuştur.

2.5.5.6. Perakendecilik sektörü

Veri ambarı konusunda başarılı olan kuruluşlar, en fazla geri dönüş sağlayan konular arasında ürün bazında satış analizi, mal veren fiyatlandırma ve performans analizi, tahmin ve yönetim sayesinde stok düzeylerini tam tutturabilme, bölgeye özel planlanan stratejik promosyonlar ve alışverişleri takip edilen müşterileri hedeflemeyi saymaktadırlar. Ayrıca işletmede hemen her bölümde verimliliğin ve yaratıcılığın artması ve yetenekli personelin özendirilip işletmede tutulabilmesi de önemli faydalar olarak belirtilmektedir.

Veri ambarı konusunda tecrübe kazanan perakendecilerin gelecekte yapmayı hedefledikleri uygulamalar arasında ise, mağazalar için detaylı ani modelleme; biten ya da azalan ürünün yerine otomatik olarak yeniden ürün konulması, detaylı yer planlaması, satış oranına göre detaylı fiyatlandırma ve müşteri / rakip davranış ilişkisi gibi uygulamalar sayılmaktadır.

Araştırmalar göstermiştir ki yeni bir müşteri kazanmak için yapılacak etkinliklerin perakendeciye maliyeti, mevcut bir müşterinin elden kaçırılmaması için yapılacak masrafın on katına ulaşmaktadır. Bunun yanına müşterinin ömür boyu değeri hesapları da eklendiğinde perakendeciler için müşterilerini daha iyi tanıyıp onlara daha iyi hizmet sunabilmek ve elde tutabilmek daha da önem kazanmaktadır. Perakendeciler bunun için bire bir pazarlama yöntemleri geliştirmeye başlamışlardır. Müşteri sadakati programları ve müşteri kartı dağıtımı uygulamaları bunlardan bazılarıdır,

Bu uygulamaların amacı perakendecinin müşterilerini tanıyıp her birine farklı bir şekilde davranabilmesini sağlamaktadır. Bu sayede müşterileriyle tek tek ilgilenen bir mahalle bakkalı gibi müşteriyi memnun edebilecek ve müşterinin demografik bilgileri, yaşam biçimi, alışkan1ıklan ve davranışları bilindiğinden ona uygun pazarlama yöntemleri daha etkili bir şekilde uygulanabilecektir. Bunu gerçekleştirmek için perakendeciler müşteri ilişkileri sürecinde şu kritik adımları atmalıdırlar:

- Müşterilerin tanınması,
- Müşterilerin ayrılması,
- Müşterilerle karşılıklı etkileşim içinde olunması,
- Müşterilerin bireysel özelliklerine dayanarak, kurumun müşterilerine davranış biçiminin özelleştirilmesi

Perakendecilikte veri ambarı teknolojisini kullanan firmaların elde ettikleri faydaları kabaca iki grupta topladıkları görülmektedir; satılan ürünleri daha iyi yönetimi (satış, kar veya her ikisini birden arttırmaya yönelik) ve daha iyi karar alabilme yeteneği (maliyetleri düşürmeye yönelik). Veri ambarı ile yapılabilecek pek çok analiz ve raporlama bu sonuçların elde edilmesini sağlamaktadır. Aşağıda bu analizlerin bazı örnekleri bulunmaktadır:

- Satış hızı yüksek ürün / rekabet edebilen fiyat analizi
- Satış hızı orta veya düşük ürün / fiyat elastisitesi analizi
- Fiyat arttırma / indirim fırsatı belirleme
- Promosyon fiyat analizi
- Başkasının markası / kendi markamız fiyat analizi
- Müşteri belirleme ve performans ölçümü
- Promosyon performansı analizi
- Benzerlik analizi
- Sepet analizi

Perakendecilikte hareket bilgilerinin miktarı çok fazladır. Türkiye' de ortalama bir perakendecide 40.000 civarında ürün kaydı olabilmekte ve günde on binlerce satış kaydı tutulmaktadır. Sezonluk ürünler, promosyonlar, ürünlerin alındığı pek çok tedarikçi ve depolar, çok sayıdaki personelin mesai ve komisyonları gibi diğer veriler de dikkate alındığında, kaydı tutulacak veri miktarının boyutları daha iyi anlaşılabilir. Ayrıca veri hazırlama konusu hafife alınmamalıdır. Stok, satış ve tedarikçi istatistikleri nispeten daha kolay elde edilebilmektedir. Doğru müşteri enformasyonuna ulaşmak ise zaman alan bir iştir, ancak bunu elde etmenin faydası da büyüktür.

Perakendecilikte ürünlerin dağıtımı, stok akışı, promosyon planlaması ve müşteri sadakati başarı için hayati konulardır. Perakendecilik işletmeleri için hareket sistemleri her zaman çekirdek teknoloji olmuştur. Veri ambarı ise, eğer uygun şekilde alınırsa, perakendecilerin daha güçlü rekabet avantajı sağlamalarına yol açacak bir teknolojidir.

2.5.6. Gelecek kuşak veri ambarları

Gelecek kuşak veri ambarı uygulamalarında ise her düzeyde müşteri ilişkisini düzenlemek için gerçek zamanlı analiz yöntemleri gerekecektir. Bugünün rekabetçi ortamındaki müşteri ilişkileri yönetimi veri ambarı uygulamalarını bire bir ilişkileri düzenlemek için yapılanma yönüne kaydıracaktır. Müşteriyle etkileşim analitik karar destek sistemleriyle birleşerek 'etkin veri ambarı' çözümlerine olgunluk kazandıracaktır. Bir bankadaki müşteri temsilcisinin durumunu düşünün. Bankaların müşteri temsilcisine sunduğu bilgiler genelde bir ürün için çapraz satış ve üst satış bilgileriyle sınırlı kalmaktadır. Bu bire bir hizmet değildir. Bire bir iletişimde, aynı ürünle ilgilenen benzer nitelikteki kişilere belli bir paket sunmaktan ziyade, her müşterinin bireysel gereksinimlerini belirleyerek ona uygun destek amaçlanır[36].

Ürüne dayalı çapraz ve üst satış teknikleri etkili olmakla birlikte müşteriyi bir birey olarak ele alamadığı için bire bir iletişimde yetersiz kalmaktadır. Gelecek kuşak müşteri ilişkilerini belirlerken, puanlama teknikleri, sorgulanan ürünle birlikte müşteriye özel satın alma modelleri ve demografı bilgilerini de kullanmak

durumundadır. Üstelik bu, müşteriyi bağlılık ve karlılık özellikleriyle birlikte değerlendirerek kişiye özel teklifler ve fiyatlandırma yoluna götürecektir. Burada gerçek zamanlı puanlamanın önemi ortaya çıkmaktadır, çünkü kişiye has özellikleri göz önüne alsa bile, önceden tanımlanmış puanlamanın en etkili özelliği, şu andaki ilişkiyi göz ardı etmesi durumudur[37].

Gerçek zamanlı müşteri iletişimini kullanmak ise teknolojiyi büyük ölçüde zorlayacaktır. Çevrim içi ve gerçek zamanlı puanlama, anlık yanıt zamanlarına gerek duyacaktır. Bu, çok büyük veri tabanlarıyla karmaşık sorgulamaların birlikteliğini getiren bir mimari gerektirmektedir ve geleneksel ilişkisel veri tabanlarıyla çözülmesi çok zor hatta olanaksızdır[38]. Ayrıca veri çekme işlemi gerçek zamanlı ya da ona yakın olmalıdır. Buna günümüz veri tabanlarının 365 gün 24 saatlik işletilebilme gereksinimini de eklemek gerekir. Unutulmaması gereken bir nokta da ölçeklenebilirlik sorunudur. Çünkü web siteleri ve e-ticaret uygulamalarıyla, artık aynı anda yüzlerce değil binlerce sorgulama yapılabilecektir. Gelecek kuşak veri ambarı uygulamaları birkaç yıl içinde yaygın1aşacaktır. Şu anda gerekli olan geleceği düşünerek yatırım ve tasarım yapmaktır[39].

BÖLÜM 3. VERİ MADENCİLİĞİ

3.1. Giriş

Bilgisayar sistemleri her geçen gün hem daha ucuzlamakta, hem de güçleri artmaktadır. İşlemciler gittikçe hızlanmakta, hafıza kapasiteleri artmaktadır. Artık bilgisayarlar daha büyük miktardaki veriyi saklayabilmekte ve daha kısa sürede işleyebilmektedir. Bunun yanında bilgisayar ağlarındaki ilerleme ile bu veriye başka bilgisayarlardan da hızla ulaşabilmek de mümkündür. Bilgisayarların ucuzlaması ile sayısal teknoloji daha yaygın olarak kullanılmaktadır. Veri doğrudan sayısal olarak toplanmakta ve saklanmaktadır. Bunun sonucu olarak da detaylı ve doğru bilgiye ulaşabilmek mümkün olmaktadır.

Örneğin eskiden süper marketteki 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 saklanabilmektedir. 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 mümkün olacaktır.

Süper market örneğinde, veri analizi yaparak her mal için bir sonraki ayın satış tahminleri çıkarılabilir; müşteriler satın aldıklara 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 gereği ortaya çıkacaktır. Veri madenciliği burada devreye girmektedir.

Peki, veri madenciliği nedir? Veri madenciliği, verilerden daha önceden bilinmeyen ve muhtemelen faydalı enformasyonun monoton olmayan bir süreçte çıkartılması işlemi olarak tanımlanmaktadır[40]. Bu süreç kümeleme, veri özetleme, sınıflama kurallarının öğrenilmesi, bağımlılık ağlarının bulunması, değişikliklerin analizi ve anomali tespiti gibi farklı bir çok teknik yaklaşımı kapsamaktadır[41]. Başka bir deyişle, veri madenciliği, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir.

Diğer bir tanımlama ise "Veri ambarlarında tutulan çok çeşitli ve çok miktarda veriye dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarmak, bunları karar verme ve eylem planını gerçekleştirmek için kullanma sürecidir"[19].

Ayrıca veri madenciliği, istatistik ve matematik tekniklerle birlikte örüntü tanıma (Pattern Recognition) teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi süreci olarak tanımlamaktadır[19].

Temel olarak veri madenciliği, veri setleri arasındaki desenlerin ya da düzenin, verinin analizi ve yazılım tekniklerinin kullanılması ile ilgilidir. Veriler arasındaki ilişkiyi, kuralları ve özellikleri belirlemekten bilgisayar yazılımları sorumludur. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir.

Veri madenciliğini istatistiksel bir yöntemler serisi olarak görmek mümkün olabilir. Ancak veri madenciliği, geleneksel istatistikten birkaç yönde farklılık gösterir. Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan – bilgisayar ara yüzünü birleştirir. Veri madenciliği sahası, istatistik, makine bilgisi, veri tabanları ve yüksek performanslı işlemler gibi temelleri de içerir[42].

Veri madenciliği konusunda bahsi geçen 'geniş veri'deki geniş ifadesi, tek bir iş istasyonunun belleğine sığamayacak kadar büyük veri kümelerini ifade etmektedir. Yüksek hacimli veri ise, tek bir iş istasyonundaki ya da bir grup iş istasyonundaki

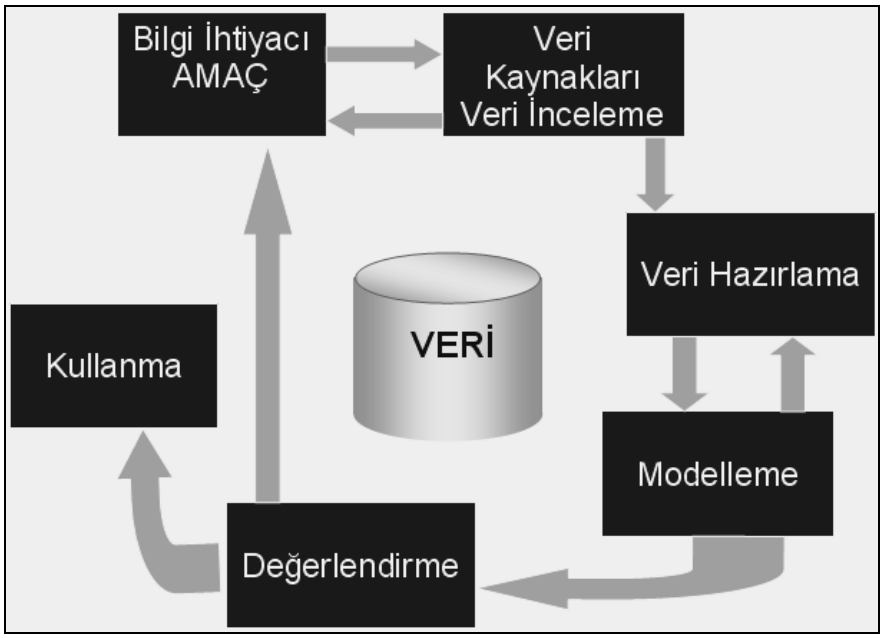
hafıza imkânlarına sığamayacak kadar fazla veri anlamındadır. Dağıtık veri ise, farklı coğrafi konumlarda bulunan verileri anlatmaktadır.

Veri madenciliğinde önemli bir role sahip olan makine öğrenimi, yapay zekâ araştırmalarında geliştirilen keşfedici (Heuristic) algoritmaların ileri istatistik tekniklerle bir harmanı olarak, son yıllarda bilim ve uygulama dünyasında önemini sürekli olarak arttırmaktadır. Makine öğrenimi teknikleri içerisinde yapay sinir ağları (artifical neural networks) ve genetik algoritmalar ön planda yer almaktadır[43]. Geniş veri tabanı ve makine öğrenimi olanaklarından yararlanılan veri madenciliğinde, özellikle sınıflama ve kümeleme konularında etkin çözümler elde edilmesi amaçlanmaktadır.

Veri madenciliği, veri tabanları, istatistik ve yapay öğrenme konularının kavramlarına dayanır ve onların tekniklerini kullanır[44]. Veri madenciliği aşağıda belirtilen özelliklere sahiptir:

- Amaç, büyük miktardaki ham veriden değerli bilginin çıkarılmasıdır.
- Çok miktarda, güvenilir veri ön şarttır. Çözümün kalitesi öncelikle verinin kalitesine bağlıdır.
- Veri madenciliği, uygulama alanındaki uzmanların ve bilgisayarın ortak çalışmasıdır.
- Uygulama ile ilgili ve yararlı olabilecek her tür bilginin öğrenmeye yardım için sisteme verilmesi gerekmektedir.
- Sonuçların tutarlılığının uzmanlar tarafından denetlenmesi gerekir.
- Veri madenciliği tek aşamalı bir çalışma değildir; tekrarlıdır. Sistem ayarlanana dek birçok deneme gerektirir.
- Veri madenciliği uzun soluklu bir çalışmadır, büyük beklentiler büyük hayal kırıklıklarına neden olabilir.

Şekil 3.1'de veri madenciliği standart süreci gösterilmiştir.



Şekil 3.1. Veri Madenciliği Standart Süreci [45]

Günümüzde oldukça yaygınlaşan elektronik ticaret ve on-line alışveriş mekanizmalarının da artmasıyla birlikte, bu alanda birbirlerine rakip olan firmaların çalışmaları, veri madenciliğinin önemini ön plana çıkarmaktadır.

Veri Madenciliği uzun süredir üzerinde çalışılan bir konu olmasına rağmen, ancak son zamanlarda iş dünyasında daha etkin bir maliyet kontrolü ve daha yüksek karlılık elde etme konusunda sağladığı katkılar ile ilgi görmeye başlamıştır. Gartner Group araştırma şirketi, gelecek on yıl içinde, hedef pazarlarda veri madenciliği kullanımının yüzde 80'lere ulaşacağı tahmininde bulunmaktadır[19].

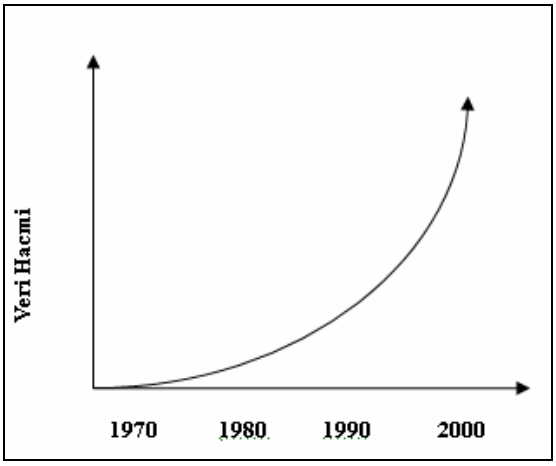
Veri Madenciliği, karar vericilerin kullanabileceği yeni bilgi oluşturabilmek için yapay zekâ (artificial intelligence) gibi yüksek teknoloji içeren yöntemler kullanmaktadır. Kurum işlemleri, müşteri geçmişi ve demografisi ve kredi büroları gibi değişik kaynaklardan toplanan veriler kullanılarak gerçek dünyanın bir modeli oluşturulmaktadır. Bu model, karar vermeyi ve yeni iş imkânlarını tahmin edebilmeyi destekleyen yöntemler üretir.

Şekil 3.2'de veri madenciliği piramidi, Şekil 3.3'te ise veri hacmindeki artış gösterilmiştir.

Şekil 3.2. Veri Madenciliği Piramidi [46]

Veri Madenciliği uygulamaları, pek çok sektör ve iş fonksiyonlarında kullanılmaktadır. Telekomünikasyon, hisse senedi işlemleri, kredi kartı ve sigorta şirketleri, hizmetlerinin istismar edilmesini önlemek için, tıp endüstrisi ameliyat prosedürlerinin, tıbbi testlerin ve ilaçla tedavinin etkinliğinin tahmini için; perakendecilik sektörü de özel uygulamalarının etkinliğini artırmada veri madenciliği tekniklerini kullanmaktadır.

Veri Madenciliği'nin tüm uygulama alanları içinde belki de en çok kullanılanı, veri tabanı pazarlamacılığı ve müşteri ilişkileri yönetimidir. Pazarlamacılar bu yolla hedefledikleri kampanyalar için uygun müşteri adaylarını belirlemekte ve müşterilerinin rakip firmaları tercih etmesinin nedenlerini saptamaktadırlar. Böylelikle maliyetler düşürülerek karlılık artırılmaktadır.



Şekil 3.3. Veri Hacmindeki Büyüme

3.2. Veri Madenciliğine Genel Bakış

Veri madenciliği ortaya çıkmadan önce, büyük veri tabanlarından faydalı örüntüler elde etmek için, çevrim-dışı veri üzerinde çalışan istatistiksel paketler kullanılmaktaydı[47]. İstatistiksel paketler kullanılırken öncelikle ne tür bilginin veri içinden çıkarılacağı belirlenmekte ve ilgili veriler toplanmaktaydı. Amaç belirlenip, kullanılacak veriler toplandıktan sonra istatistiksel paket çalıştırılıp, elde edilen sonuçlar bir alan uzmanı tarafından ayıklandıktan sonra yorumlanmaktaydı. İstatistiksel bir paket kullanmanın dezavantajı, bu işlemlerin her farklı ihtiyaç için tekrarlanmasıdır. Oysa veri tabanlarında bilgi keşfi (VTBK), önemli ve anlamlı örüntüleri çok büyük işletimsel veri içinden otomatik olarak belirleyerek kullanıcının amacına uygun biçimde bilgiye dönüştürmektedir. Bilgi, potansiyel olarak ilginç ve faydalı olan veriler arasındaki ilişkidir. Keşif ise, gizlenmiş veya daha önceden bilinmeyen nesnelerin bulunması anlamına gelir.

Günümüz ilişkisel veri tabanı teknolojisi bu biçimde bilgi elde etme olanakları açısından çok zayıftır. Aynı zamanda, akıllı veri analizinde kullanılan bilgi keşfi teknikleri gerçek dünya veri tabanları için henüz yeterince olgunlaşmamıştır. Örneğin bir kuruma ait bilgi sisteminde, kurumsal ihtiyaçlar çerçevesinde veriler farklı kaynaklardan toplanabilir. Bu, bilgi çıkarım tekniklerinin, farklı kaynaklardan ilgili verileri toplaması anlamına gelir. Bu durum veri toplama aşamasında gerçek bir güçlüğe sebep olabilir. Bu yüzden genel amaçlı VTBK sistemi henüz gerçeklikten oldukça uzaktır.

Veri madenciliği problemi, büyük veri tabanlarında bilgi keşfinin önemini vurgulamak ve araştırmacılar ve uygulama geliştiricilerin bu konuda çalışmalarını özendirmek amacıyla tanımlanmıştır. VTBK sistemleri; örüntü tanıma, makine öğrenimi, makine keşfi, veri tabanı yönetimi, istatistik, uzman sistemler, veri görselleştirme(data visua1ization) gibi farklı alanlarda kullanılan yöntemleri bir araya getirmiştir. Ancak VTBK, sözü edilen alanlardaki sonuçları kullansa bile amaçları bakımından bu alanlardan ayrılmaktadır. Sonraki bölümde VTBK ile, esinlendiği sahalar arasındaki temel farklılık ve/veya benzerlikleri incelenmiştir[48].

Veri modelleme ve verideki gürültüyü aza1tma açısından VTBK istatistik ile yakından ilgilidir. Son yıllarda istatistiğe dayalı veri madenciliği tekniklerinin kullanımında önemli bir artış görülmüştür. Bunlar özellik seçimi, veri bağımlılığı, tanıma dayalı nesnelerin sınıflandırılması, veri özeti, eksik değerlerin tahmini, sürekli değerlerin ayrımı vb. Veri analizinde kullanılan istatistiksel teknikler iyi tasarlanmıştır ve bazı durumlarda başka bir tekniğin uygulanmasına gerek yoktur. Ancak pek çok veri analizi problemi istatistiksel tekniklerin uygulanmasına elverişli değildir. İstatistiksel tekniklerdeki temel kısıtlama, ilişkileri tanıma ve genelleştirme yetersizliğidir[49].

VTBK sistemleri için veri modelleme ve örüntü çıkarmada gerekli olan teori ve algoritmalar makine öğrenimi ve örüntü tanıma alanlarında yapılan araştırmaların sonuçlarıdır (örneklerden öğrenme, olgular ile kavramı biçimleme, düzenli örüntülerin keşfi, gürültülü ve eksik veri, belirsizlik yönetimi vb.). Bununla birlikte VTBK araştırmaları bu teorilerin büyük veri kümelerinde uygulanmasını sağlamıştır. Veri madenciliği, veri tabanının öğrenme kümesi rolünü üstlendiği kuralların üretilme sürecidir. Bir veri madenciliği uygulaması, kontrol edilemeyen gerçek dünya verisini ele alabilecek biçimde makine öğrenimi tekniklerini genişletir.

Sahası gözlem ve deneye dayalı ampirik kuralların otomatik biçimde bulunması olan makine keşfi (machine discovery), VTBK sistemleri ile yakından ilgilidir. Ticari çevrelerde son yıllarda yaygınlaşan veri ambarlama ve çevrim-içi analitik işleme alanları VTBK gibi yeni kuşak stratejik bilgi çıkarım ve analiz araçlarıdır. Uzman sistemlerde bilgi edinme de veri madenciliği ile ilgilidir[50]. Ancak uzman sistemlerde bilgi, alan uzmanı ve bilgi mühendisinin gözetiminde gerçekleştirilir. Veri görselleştirmede kullanılan yöntemler, VTBK sistemi ile elde edilen örüntülerin kullanıcıya grafikler aracılığıyla sunumunu sağlar.

3.3. Veri Madenciliğinin Temelleri

Veri madenciliği işlemleri aşağıdaki gelişmeler doğrultusunda gelişmiştir. Böylece daha kolay ve daha hızlı yapılabilir hale gelmiştir[51];

- Veri madenciliği algoritmaları
- Güçlü, çoklu işlemcili bilgisayarlar
- Büyük hacimde veri toplama

Tablo 3.1'de veri işleme tekniklerinin veri tablolarından veri madenciliğine kadar olan gelişimi görülmektedir.

Etkin bir veri madenciliği uygulayabilmek için dikkat edilmesi gereken noktalar aşağıdaki gibi özetlenebilir;

- Farklı tipteki verileri ele alma: Gerçek hayattaki uygulamalar makine öğreniminde olduğu gibi yalnızca sembolik veya kategorik veri türleri değil, aynı zamanda tamsayı, kesirli sayılar, çoklu ortam verisi, coğrafik bilgi içeren veri gibi farklı tipteki veriler üzerinde işlem yapılmasını gerektirir. Kullanılan verinin saklandığı ortam, düz bir kütük veya ilişkisel veri tabanında yer alan tablolar olacağı gibi, nesneye yönelik veri taban1arı, çoklu ortam veri taban1arı, coğrafik veri taban1arı vb.de olabilir. Saklandığı ortama göre veri, basit tipte olabileceği gibi karmaşık veri tipleri de olabilir. Bununla birlikte veri tipi çeşitliliğinin fazla olması bir veri madenciliği algoritmasının tüm veri tiplerini ele alabilmesini zorlaştırmaktadır. Bu yüzden veri tipine özgü veri madenciliği algoritmaları geliştirilmektedir.

- Veri madenciliği algoritmasının etkinliği ve ölçeklenebilirliği: Çok büyük oylumlu veri içinden bilgi elde etmek için kullanılan veri madenciliği algoritması etkin ve ölçeklenebilir olmalıdır. Bu, veri madenciliği algoritmasının çalışma zamanının tahmin edilebilir ve kabul edilebilir bir süre olmasını gerektirir. Üssel veya çok terimli bir karmaşıklığa sahip bir veri madenciliği algoritmasının uygulanması kullanışlı değildir.

- Sonuçların yararlılık, kesinlik ve anlamlılık kıstaslarını sağlaması: Elde edilen sonuçlar analiz için kullanılan veri tabanını doğru biçimde yansıtmalıdır. Bunun yanı sıra gürültülü ve aykırı veriler ele alınmalıdır. Bu işlem elde edilen kuralların kalitesini belirlemede önemli bir rol oynar.

Tablo 3. 1. Veri İşleme Tekniklerinin Gelişimi[1]

Gelişme Basamakları	**İmkân Sağlayan Teknolojiler**	**Ürün Sağlayıcılar**	**Karakteristikler**
Veri Toplama (1960'lar)	Bilgisayarlar Kasetler Diskler	IBM, CDC	Statik veri ulaşımı
Veriye Ulaşım (1980'ler)	İlişkili veri tabanları (RDMS),SQL,ODBC	Oracle, Sybase, IBM, Informix, Microsoft	Kayıt esnasında veriye dinamik ulaşım
Veri Depolama ve Karar Destek (1990'lar)	OLAP, Çok boyutlu veri tabanları, Veri ambarları	Pilot, Cornshare, Arbor, Cognos, Microstrategy	Her seviyede veriye dinamik ulaşım
Veri Madenciliği (Günümüzde)	Gelişmiş algoritmalar, çoklu işlemcili bilgisayarlar, büyük hacimli veri tabanları	Pilot, Lockheed, IBM, SGI	Bilgiye önceden ulaşım

- Keşfedilen kuralların çeşitli biçimlerde gösterimi: Bu özellik keşfedilen bilginin gösterim biçiminin seçilebilmesini sağlayan yüksek düzeyli bir dil tanımının yapılmasını ve grafik ara yüzünü gerektirir.

- Farklı bir kaç soyutlama düzeyi ve etkileşimli veri madenciliği: Büyük veri tabanlarından elde edilecek bilginin tahmin edilmesi güçtür. Bu yüzden veri madenciliği sorgusu, elde edilen bilgilere göre kullanıcıya etkileşimli olarak sorgusunu değiştirebilmeyi, farklı açılardan ve farklı soyutlama düzeylerinden keşfedilen bilgiyi inceleyebilme esnekliğini sağlamalıdır.

- Farklı ortamlarda yer alan veri üzerinde işlem yapabilme: Kurumlar, yerel ağlar üzerinden pek çok dağınık ve heterojen veri tabanı üzerinde işlem yapmaktadırlar. Bu, veri madenciliğinin farklı kaynaklarda birikmiş formatlı ya da formatsız veriler üzerinde analiz yapabilmesini gerektirir. Verinin büyüklüğünün yanı sıra dağınık olması, farklı algoritmaların geliştirilmesi ve kullanılmasını gerektirmektedir.

- Gizlilik ve veri güvenliğinin sağlanması: Bir VTBK sisteminde keşfedilen bilgi pek çok farklı açıdan ve soyutlama düzeyinden izlenebildiği için, gizlilik ve veri güvenliği, veri madenciliği sistemini kullanan kullanıcının haklarına ve erişim yetkilerine göre sağlanmalıdır.

3.4 . Veri Tabanlarında Bilgi Keşfi Süreci ve Veri Madenciliği

Veri tabanı sistemlerinin artan kullanımı ve hacimlerindeki bu olağanüstü artış, organizasyonları elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır. Geleneksel sorgu (Query) veya raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veri Tabanlarında Bilgi Keşfi- VTBK (Knowledge Discovery in Databases) adı altında sürekli ve yeni arayışlara neden olmaktadır. VTBK süreci içerisinde, modelin kurulması ve değerlendirilmesi aşamalarından meydana gelen veri madenciliği en önemli kesimi oluşturmaktadır. Bu önem, birçok araştırmacı tarafından VTBK ile veri madenciliği terimlerinin eş anlamlı olarak da kullanılmasına neden olmaktadır.

Çeşitli veri kaynaklarından verilerin toplanması ile başlayan VTBK süreci, toplanan verilerin analiz için uygun hale getirilmesi aşaması ile devam etmektedir. Ancak veri ambarına (Data Warehouse) sahip olan kuruluşlarda, gerekli verilerin Data Mart olarak isimlendirilen işleve özel veri tabanlarına aktarılması ile doğrudan veri madenciliği işlemlerine başlanabilmesi de mümkündür[52].

Örüntü tanıma ve sınıflama problemleri üzerinde yoğunlaşan yapay zekâ ve istatistik disiplinlerindeki gelişmeler, veri madenciliğinin temellerini oluşturmaktadır. Ayrıca veri madenciliği, yapay zekâ çalışmalarının uzantısı olan makine öğrenimi (Machine Learning) ve uzman sistemlerin (Expert Systems) yanı sıra, veri tabanları optimizasyonu, görselleştirme (Visualization), yüksek performanslı paralel işlemciler (Massively Parallel Processing – MPP ve Symmetric Multiprocessing -SMP) gibi çeşitli disiplin ve teknolojilerdeki gelişmelerden de etkilenmektedir[53].

Diğer bir kaynağa göre; VTBK daha geniş bir disiplin olarak görülmektedir ve veri madenci1iği terimi, sadece bilgi keşfi metot1arıyla uğraşan VTBK sürecinde yer alan bir adımdır. VTBK sürecinde yer alan adımlar şöyledir[40];

Veri Seçimi (Data Selection): Bu adım birkaç veri kümesini birleştirerek, sorguya uygun örnekler kümesini elde etmeyi gerektirir.

Veri Temizleme ve Önişleme (Data Cleaning & Preprocessing)**:** Seçilen örneklemde yer alan hatalı tutanakların çıkarıldığı ve eksik nitelik değerlerinin değiştirildiği aşamadır. Bu aşama keşfedilen bilginin kalitesini arttırır.

Veri İndirgeme (Data Reduction): Seçilen örneklemden ilgisiz niteliklerin atıldığı ve tekrarlı tutanakların ayıklandığı adımdır. Bu aşama, seçilen veri madenciliği sorgusunun çalışma zamanını iyileştirir.

Veri Madenciliği (Data Mining): Verilen bir veri madenciliği sorgusunun işletilmesidir (sınıflama, kümeleme, eşleştirme, vb.).

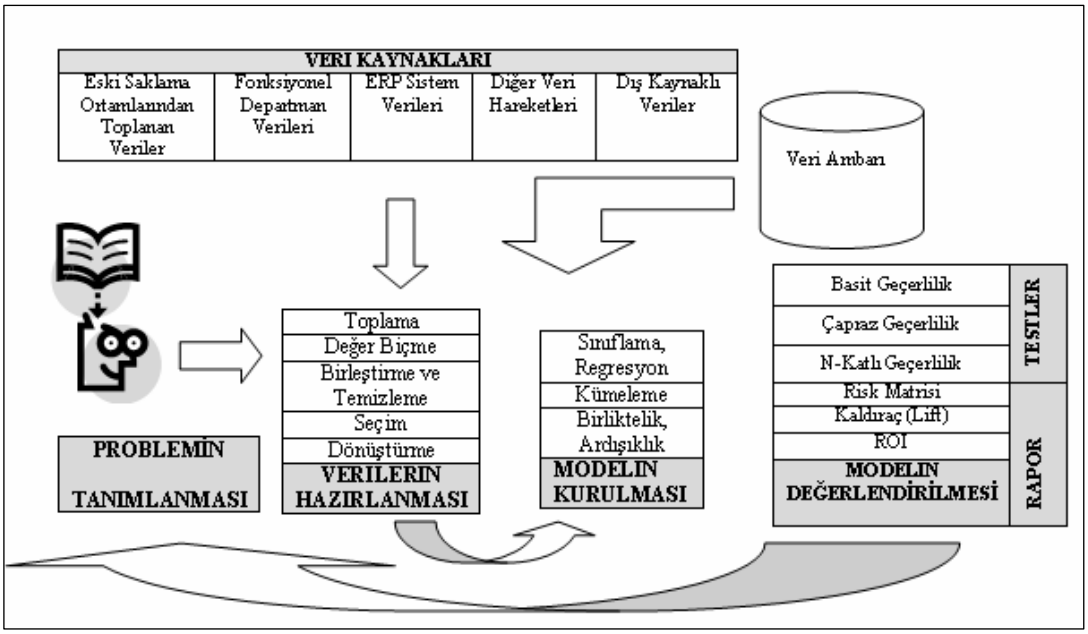
Değerlendirme (Evaluation): Keşfedilen bilginin geçerlilik, yenilik, yararlılık ve basitlik kıstaslarına göre değerlendirilmesi aşamasıdır[54].

3.5. Veri Tabanlarında Bilgi Keşfi Süreci

Ne kadar etkin olursa olsun, hiç bir veri madenciliği algoritması, 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[1]. Şekil 3.4' te ayrıntılı olarak görüldüğü gibi, veri tabanlarında bilgi keşfi süreci;

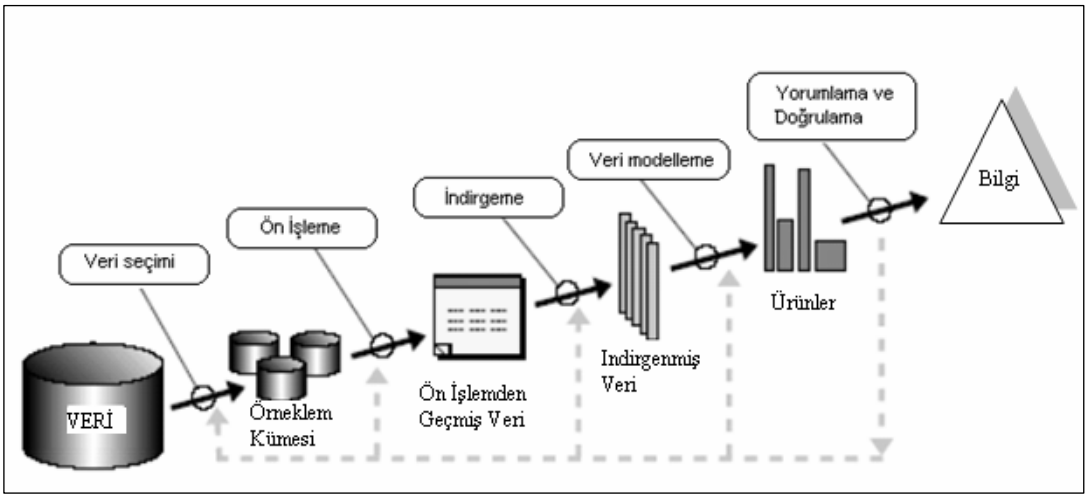
- Problemin Tanımlanması,
- Verilerin Hazırlanması,
- Modelin Kurulması ve Değerlendirilmesi,
- Modelin Kullanılması,

- Modelin İzlenmesi, veri tabanlarında bilgi keşfi sürecinde izlenmesi, gereken temel aşamalardır.



Şekil 3.4. Veri Tabanlarında Bilgi Keşfi Süreci ve Veri Madenciliği[1].

Veri tabanlarında bilgi keşfi sürecinde yer alan temel adımlar Şekil 5.3'te gösterilmiştir.



Şekil 3.5 VTBK Sürecinde Yer Alan Adımlar[12]

3.5.1. 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.

3.5.2. 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[12,40].

3.5.2.1. 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 faydalanmak mümkündür.

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

3.5.2.3. 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 gelişigüzel yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

3.5.2.4. 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 a1goritma1arı konu ile ilgisi olmayan bu tip değişkenleri otomatik olarak elese de, pratikte bu işlemin kullanılan yazılıma bırakılmaması daha akılcı olacaktır.

Verilerin görselleştirilmesine 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 olacaktır. 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.

3.5.2.5. Dönüştürme

Kredi riskinin tahmini için geliştirilen bir modelde, borç/gelir gibi önceden hesaplanmış bir oran yerine, ayrı ayrı borç ve gelir verilerinin kullanılması tercih edilmektedir. Ayrıca modelde kullanılan a1goritma, verilerin gösteriminde önemli rol oynayacaktır. Örneğin bir uygulamada, 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.

3.5.3. 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.

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 belirlenir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik 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 olunması durumunda, kullanılabilecek diğer bir yöntem çapraz geçerlilik testidir. Bu yöntemde veri kümesi tesadüfî 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 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.

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir. Model kuruluş çalışmalarına başlamadan önce, imkânsız 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.

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

Kaldıraç (Lift) oranı ve grafiği, bir modelin sağladığı faydanın değerlendirilmesinde kullanılan en önemli 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 tesadüfî 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 için katlanılacak maliyete bölünmesi ile edilecek olan yatırımın geri dönüş oranıdır.

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamıyla 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.

3.5.4. 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.

3.5.5. 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[12,54].

3.6. Veri Madenciliği İçin Gerekli Olan Altyapı

Veri madenciliği; PC, client / server ve main frame olan tüm sistemlerde kullanılabilir. İşletmenin tümüne yönelik uygulamalarda veri madenciliği uygulamalarının kapladığı alan 10 Gbyte ile 11 Tbyte arasındadır. Burada önemli olan iki nokta; sorgulamanın karmaşıklığı ve veri tabanının kapasitesidir. Bu iki değer ne kadar büyük olursa, ihtiyaç duyulan sistem büyüklüğü ve gücü de o kadar yüksek olur.

3.7. Veri Madenciliğine İhtiyaç Duyulmasının Sebepleri

İş süreçlerinde veri madenciliği işleminin karar verme mekanizmalarında ön plana çıkmasını sağlayan faktörler şunlardır:

- Geniş veri tabanlda işlenmemiş değerli verilerin varlığı,
- Belirli müşteri segmentine yönelik veri tabanı kayıtlarının bir araya getirilmesi,

- Veri tabanı kayıtlarının birleştirilmesinden sonra bilgi ve veri ambarları konseptlerine ulaşılması,
- Veri tabanı hacimlerinin veri madenciliği gerektirecek düzeye ulaşması, pazarlama, reklam ve imalatta küçük müşteri segmentlerine ve bireylere kadar ulaşılması gerekliliği.

3.8. Veri Madenciliğinin Amaçları

Veri Madenciliğinin amaçlarını aşağıdaki başlıklar altında toplamak mümkündür;

- Öngörü: Hangi ürünlerin, hangi dönemlerde, hangi şartlarda, hangi miktarlarda satılacağına ilişkin öngörülerde bulunmak

- Tanıma: Aldığı ürünlerden bir müşterinin tanınması, kullandığı programlar ve yaptığı işlemlerden bir kullanıcının tanınması

- Sınıflandırma: Birçok parametrenin birleşimi kullanılarak ürünlerin, müşterilerin vb. sınıflandırılması

- En İyileme: Belirli kısıtlamalar çerçevesinde zaman, yer, para ya da ham madde gibi sınırlı kaynakların kullanımını en iyileme ve üretim miktarı, satış miktarı ya da kazanç gibi değerleri büyütme de veri madenciliği amaçlarındandır.

3.9. Veri Madenciliğinin İşletmelerde Kullanımı

İstatistiğin amacı nasıl ana kütle hakkında anlamlı bilgiler elde etmek ve yorum yapmaksa, veri madenciliğinin amacı da anlamlı bilgiler elde etmek ve bunu eyleme dönüştürecek kararlar için kullanmaktır. İlgilendiği ana kütle mevcut veya potansiyel müşteriler olabilir. Müşterilerin profillerini, satın alma eğilimlerini, bir ürünü veya hizmeti kabul etme veya etmeme ihtimallerini tahmin etme, veri madenciliğinde hedeflenen amaçlar arasındadır. Bu tahminler, strateji belirlemede ve karar vermede kullanılır. Ürün ve hizmet sektöründe müşterilerle ilgili veri madenciliği uygulama

amaçlarına ilişkin çok çeşitli örnekler vermek mümkündür. En karlı pazar alanlarını, en karlı müşterileri, yeni bir promosyon kampanyasında müşteriye sunulan ürün veya hizmetin kabul edilme oranlarını saptamak, pazarlamada veri madenciliği uygulamasının önemli amaçlarındandır. Veri madenciliği uygulamalarından elde edilecek faydalara ilişkin bazı örnekler aşağıda sıralanmıştır[55].

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 müşterinin 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.

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.

3.10. Veri Madenciliğinde Karşılaşılan Problemler

Küçük veri kümelerinde hızlı ve doğru bir biçimde çalışan bir sistem, çok büyük veri tabanlarına uygulandığında tamamen farklı davranabilir. Bir veri madenciliği sistemi tutarlı veri üzerinde mükemmel çalışırken, aynı veriye gürültü eklendiğinde kayda değer bir biçimde kötü1eşebilir. Veri madenciliğinde karşılaşılan başlıca problemler aşağıda sıralanmıştır:

3.10.1. Veri tabanı boyutu

Veri tabanı boyutları inanılmaz bir hızla artmaktadır. Pek çok makine öğrenimi algoritması bir kaç yüz tutanaklık oldukça küçük örneklemeleri ele alabilecek biçimde geliştirilmiştir. Aynı algoritmaların yüz binlerce kat büyük örneklemelerde kullanılabilmesi için azami dikkat gerekmektedir. Örneklemenin büyük olması, örüntülerin gerçekten var olduğunu göstermesi açısından bir avantajdır ancak böyle bir örneklemeden elde edilebilecek olası örüntü sayısı çok büyüktür. Bu yüzden veri madenciliği sistemlerinin karşı karşıya olduğu en önemli sorunlardan biri veri tabanı boyutunun çok büyük olmasıdır. Dolayısıyla veri madenciliği yöntemleri ya sezgisel/buluşsal bir yaklaşımla arama uzayım taramalıdır ya da örneklemeyi yatay/dikey olarak indirgemelidir[41].

Yatay indirgeme, nitelik değerlerinin önceden belirlenmiş genelleme sıradüzenine göre, bir üst nitelik değeri ile değiştirilme işlemi yapıldıktan sonra aynı olan çoklukların çıkarılması işlemidir.

Dikey indirgeme artık niteliklerin indirgenmesi işlemidir. Özellik seçimi yöntemleri ya da nitelik bağımlılık çizelgesi uygulanarak yapılır.

3.10.2. Gürültülü veri

Büyük veri tabanlarında pek çok niteliğin değeri yanlış olabilir. Bu hata, veri girişi sırasında yapılan insan hataları veya girilen değerin yanlış ölçülmesinden kaynaklanır. Veri girişi ya da veri toplanması sırasında oluşan sistem dışı hatalara

gürültü adı verilir. Ancak günümüzde kullanılan ticari ilişkisel veri tabanları veri girişi sırasında oluşan hataları otomatik biçimde gidermek konusunda az bir destek sağlamaktadır. Hatalı veri gerçek dünya veri tabanlarında ciddi problem oluşturabilir. Bu durum, bir veri madenciliği yönteminin kullanılan veri kümesinde bulunan gürültülü verilere karşı daha az duyarlı olmasını gerektirir. Gürültülü verinin yol açtığı problemler tümevarımsal karar ağaçlarında uygulanan metotlar bağlamında kapsamlı bir biçimde araştırılmıştır. Eğer veri kümesi gürültülü ise sistem bozuk veriyi tanımalı ve ihmal etmelidir. Quinlan, gürültünün sınıflama üzerindeki etkisini araştırmak için bir dizi deney yapmıştır. Deneysel sonuçlar, etiketli öğrenmede etiket üzerindeki gürültü öğrenme algoritmasının performansını doğrudan etkileyerek düşmesine sebep olmuştur. Buna karşın eğitim kümesindeki nesnelerin özellikleri/nitelikleri üzerindeki en çok % 10'luk gürültü miktarı ayıklanabilmektedir. Chan ve Wong gürültünün etkisini analiz etmek için istatistiksel yöntemler kullanmışlardır [56].

3.10.3. Boş (null) değerler

Veri tabanlarında boş değeri birincil anahtarda yer almayan herhangi bir niteliğin değeri olabilir. Bir çokluda eğer bir nitelik değeri boş ise o nitelik bilinmeyen ve uygulanamaz bir değere sahiptir. Bu durum ilişkisel veri tabanlarında sıkça karşımıza çıkmaktadır. Bir ilişkide yer alan tüm çoklular aynı sayıda niteliğe, niteliğin değeri boş olsa bile, sahip olmalıdır. Örneğin kişisel bilgisayarların özelliklerini tutan bir ilişkide bazı model bilgisayarlar için ses kartı modeli niteliğinin değeri boş olabilir[57].

Lee, boş değerini, bilinmeyen, uygulanamaz ve bilinmeyen veya uygulanamaz olacak biçimde üçe ayıran bir yaklaşımı ilişkisel veri tabanlarını genişletmek için öne sürmüştür. Mevcut boş değer taşıyan veri için herhangi bir çözüm sunmayan bu yaklaşımın dışında bu konuda sadece bilinmeyen değer üzerinde çalışmalar yapılmıştır. Boş değerli nitelikler veri kümesinde bulunuyorsa, ya bu çoklular tamamıyla ihmal edilmeli ya da bu çoklularda niteliğe olası en yakın değer atanmalıdır[58].

3.10.4. Eksik veri

Evrendeki her nesnenin ayrıntılı bir biçimde tanımlandığı ve bu nesnelerin alabileceği değerler kümesinin belirli olduğu varsayılsın. Verilen bir bağlamda her bir nesnenin tanımı kesin ve yeterli olsa idi, sınıflama işlemi basitçe nesnelerin alt kümelerinden faydalanılarak yapılırdı. Bununla birlikte, veriler kurum ihtiyaçları göz önünde bulundurularak düzenlenip, toplandığından, mevcut veri gerçek hayatı yeterince yansıtmayabilir. Örneğin hastalığın tanısını koymak için kurallar sadece çok yaşlı insanların belirtilerinin bulunduğu bir veri kümesi kullanılarak üretilseydi, bu kurallara dayanarak bir çocuğa tanı koymak pek doğru olmazdı. Bu gibi koşullarda bilgi keşfi modeli belirli bir güvenlik derecesinde tahmini kararlar alabilmelidir [59].

3.10.5. Artık veri

Verilen veri kümesi, eldeki probleme uygun olmayan veya artık nitelikler içerebilir. Bu durum pek çok işlem sırasında karşımıza çıkabilir. Örneğin, eldeki problem ile ilgili veriyi elde etmek için iki ilişkiyi ortak nitelikler üzerinden birleştirirsek sonuç ilişkide kullanıcının farkında olmadığı artık nitelikler bulunur. Artık nitelikleri elemek için geliştirilmiş algoritmalar özellik seçimi olarak adlandırılır [60].

Özellik seçimi, tümevarıma dayalı öğrenmede budama öncesi yapılan bir işlemdir. Başka bir deyişle, özellik seçimi, verilen bir ilişkinin içsel tanımını, dışsal tanımın taşıdığı (veya içerdiği) bilgiyi bozmadan onu eldeki niteliklerden daha az sayıdaki niteliklerle (yeterli ve gerekli) ifadeleyebilmektir. Özellik seçimi yalnızca arama uzayını küçültmekle kalmayıp, sınıflama işleminin kalitesini de arttırır[61].

3.10.6. Dinamik veri

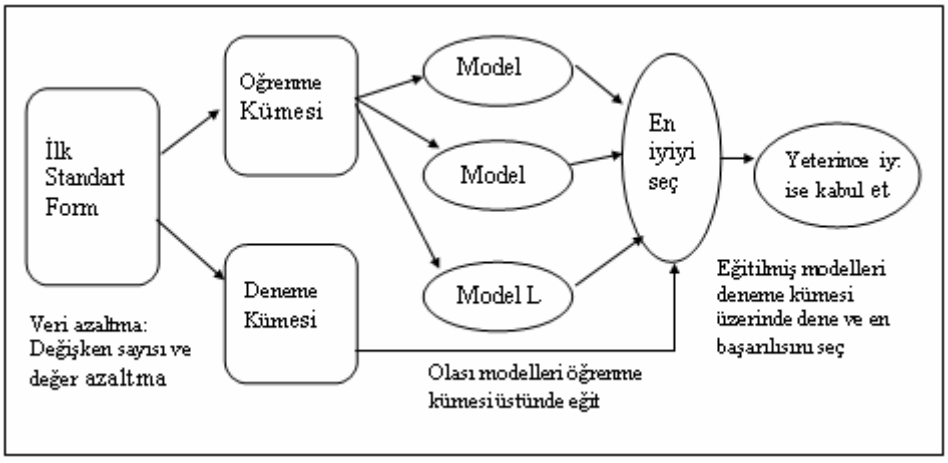
Kurumsal çevrim-içi veri tabanları dinamiktir, yani içeriği sürekli olarak değişir. Bu durum, bilgi keşfi metotları için önemli sakıncalar doğurmaktadır. İlk olarak sadece okuma yapan ve uzun süre çalışan bilgi keşfi metodu bir veri tabanı uygulaması olarak mevcut veri tabanı ile birlikte çalıştırıldığında mevcut uygulamanın da

performansı ciddi ölçüde düşer. Diğer bir sakınca ise, veri tabanında bulunan verilerin kalıcı olduğu varsayılıp, çevrimdışı veri üzerinde bilgi keşif metodu çalıştırıldığında, değişen verinin elde edilen örüntü1ere yansıması gerekmektedir. Bu işlem, bilgi keşfi metodunun ürettiği örüntü1eri zaman içinde değişen veriye göre sadece ilgili örüntü1eri yığmalı olarak günleme yeteneğine sahip olmasını gerektirir. Aktif veri tabanları tetikleme mekanizmalarına sahiptir ve bu özellik bilgi keşif metotları ile birlikte kullanılabilir[62].

BÖLÜM 4. VERİ MADENCİLİĞİNİN METODOLOJİSİ, KULLANIM ALANLARI, MODEL VE ALGORİTMALARI

4.1. Veri Madenciliği Metodolojisi

Bir veri madenciliği çalışmasında kullanılan metodoloji Şekil 4.1 '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 mümkün 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.



Şekil 4.1. Veri Madenciliğinde Kullanılan Metodoloji

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

4.2. Veri Madenciliğinin Kullanım ve Uygulama Alanları

Veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta tıp ve birçok başka alanda uygulanmaktadır. Son beş yıldır Amerika Birleşik Devletleri'nde çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılıklarının ortaya çıkarılmasına kadar çeşitli uygulamalarda kullanıldığı bilinmektedir[63].

Veri madenciliği, Türkiye'de ise gazetecilik, bankacılık, perakendecilik sektörlerinde ve üniversitelerde kullanılmaktadır.

Bununla birlikte günümüzde veri madenciliği teknikleri özellikle işletmelerde çeşitli alanlarda başarı ile kullanılmaktadır. Bu uygulamaların başlıcaları ilgili alanlara göre aşağıda özetlenmiştir.

4.2.1. Pazarlama

Bilgisayarlaştırılmış kredi kartlarının kullanımı sayesinde, pazarlamacılar her alışveriş işleminin detaylı kayıtlarını tutabilmektedirler. Bu, onlara farklı müşteri kesimlerini daha iyi tanımalarına imkân verir. Bu durumlar:

- Pazar sepeti analizi: Sepet analizi, müşterilerin hangi kalem malları birlikte satın alma eğiliminde olduğunu açıklar, gösterir.

- Zaman bazlı örüntülerin incelenmesi: Bu bilgi pazarlamacılara stok yapma kararlarında yardımcı olur. Eğer müşteri bugün bir kasetçalar satın aldıysa muhtemelen ne zaman ekstra bir batarya ve ayrıca kaset satın alacaktır, sorusuna cevap veren bir analizdir.

- Tahmini modeller geliştirmek: Pazarlamacılar, satın alınan mallara ya da satışlara bakarak belli davranışlarla müşteri profillerini geliştirebilirler. Daha sonra pazarlamacılar, bu bilgiyi sonuca odaklanmış maliyet-etkili promosyonları geliştirmek için kullanabilirler.

4.2.2. Bankacılık

Bankalar çok yönlü uygulamalar için bilgi keşfi sayesinde avantaj elde edebilirler. Bu uygulamalar arasında; sahtekârlık, dolandırıcılık tespiti, müşteri gruplaması ve benzeri operasyonları vardır.

- Dolandırıcılık tespiti: Bankacılık sektöründe kredi kartı işlemlerinde yapılan dolandırıcılık girişimleri fazlaca olmaktadır. Bankalar geçmiş işlemleri analiz ederek sahte olduğu daha sonra tespit edilen örüntü1eri tespit edebilirler. Örneğin, müşterisinin elektronik kayıtları üzerinde çok kısa bir sürede pek çok işlem olduğunu fark ettiği anda, işlemleri dolandırıcılık örüntü1erine uyan bu müşterisi ile konuşmadan işlemlerinin gerçekleşmesine onay vermeyebilir.

- Müşteri gruplaması: Müşteri grupları tespit edilerek bankalar, belli gruplara farklı çeşit hizmet sağlayarak kendi kendilerine ayrım yapabilirler. Örneğin; banka, sık sık seyahat eden müşterilerine benzer bir kart ve kredi kartlarını daima zamanında ödeyen müşterilerine de daha başka bir kart seçebilir. Bankalar ayrıca, belirli bir reklâmdan çoğunlukla benzer fayda getiren bölümleri tespit etmek için gruplamayı kullanır.

- Doğal süreç yönetim tahmini**:** Bilgi keşfi, bankanın her müşterisinin ömür değerini tahmin etmesinde ve her bir gruba uygun şekilde hizmet etmesinde yardımcı olur. Banka mevcut karlı müşterilerinin bir profilini çıkarır ve onların ortak özelliklerini birkaç yıl önceden belirlemek için bilgi çıkarım tekniklerini kullanır. Daha sonra yakın gelecekte muhtemelen karlı müşteri olacak özelliklere sahip müşterileri bugünden tespit etmeyi başarır. Banka bu müşterilerini kaybetmeme programları hedefleyebilir. Örneğin; özel mukaveleler sunarak ya da alacaklarından vazgeçmek gibi[64].

4.2.3. Haberleşme

Haberleşme şirketleri tüm dünyada kendilerini daha iyi tanıtmak ve mevcut müşterilerini kaybetmemek için ücretlendirme programlarına ve yeni müşteriler çekmeye zorlayan yükselen rekabet seviyeleri ile karşı karşıyadırlar. Haberleşme teknolojisinde bilgi keşfinin uygulamaları aşağıdaki şekilde sıralanabilir.

- Çağrı ayrıntı analizleri: Haberleşme şirketleri görüşmelerin ayrıntılı kayıtlarını tutarlar. Firmalar, bu verilerin analizi sonucunda benzer kullanıcı örüntü1eri ile müşteri gruplarım tespit ederek cazip fiyatlandırma programları geliştirebilirler.

- Müşteri bağlılığı: Bazı müşteriler sürekli hizmet sağlayıcılarını değiştirirler yada her bir şirketin özendirici promosyonlarından avantaj sağlamaya çalışırlar. Haberleşme şirketleri, kazancın en fazla olduğu pazara harcama yapmak amacıyla kendilerine bağlı olan müşterilerinin karakteristiklerini tespit etmek için bilgi keşfi metodunu kullanabilirler.

4.2.4. Sigortacılık

Sigorta şirketleri, daha etkin planlama için itme gücü olarak kullanabilecekleri, yıllarca birikmiş yüksek hacimli verilere sahiptirler. Bilgi keşfinin uygulamaları şöyledir:

- Dolandırıcı keşfi: Sigorta şirketleri iflas, kaza sigortası gibi yüksek meblağlı taleplerdeki örüntüleri inceleyerek hukukçular, doktorlar ve hak sahipleri arasındaki ilişkiyi belirleyerek sahtekârlıkları azaltabilirler.

- Ürün tasarımı: Sigortacılar en karlı poliçe miktarını ve poliçe seçeneklerini bilmek isterler. Sigortacılar bu bilgiyi gelecekte satışı en iyi olacak yeni ürün tasarımında kullanırlar.

- Risk analizi: Ödenmiş taleplerle ilişkin etmenler kombinasyonunu tespit etmeyle, sigortacılar maruz kaldıkları yükümlülüklerini azaltabilirler. ABD' de büyük bir sigorta şirketi son zamanlarda evli insanların talep ettiği dolar miktarını evli olamayan insanlara göre iki kat fazla olduğunu keşfetti. Şirket bu yeni bilgiyi değerlendirerek boş poliçelerinde evli insanları azaltma yönünde karar aldı.

4.2.5. Pazar analizleri ve yönetimi

Günümüz serbest rekabet ortamında zaman ve piyasa verilerinin karar destek amaçlı bilgi haline dönüşmüş şekli, satış ve pazarlama faaliyetleri açısından kritik öneme haiz iki unsurdur. Doğru kararlar alma ve bu kararların gecikmeden hayata geçirilmesi işletmelerin varlığım devam ettirebilmeleri açısından çok önemlidir.

Pazar analizlerinde kullanılacak verinin kaynağı kredi kartı işlemleri, indirim kuponları, müşteri şikâyet aramaları, yaşam stili çalışmaları, müşteri kayıtları olabilir. Bu kapsamda yapılan çalışmalar, ilgi alam, gelir düzeyi, harcama alışkanlıkları, vb. özellikler açısından benzer niteliklerdeki müşteriler için bir model belirlenmesi ve hedef pazarın saptanması; müşterinin fiyat artışı ile değişen satın alma alışkanlıklarının belirlenmesi; çapraz pazar analizleri ile ürün satışları arasındaki birlikteliklerin ve ilişkilerin belirlenmesi ve bu bilgilere dayanılarak ürün satış tahminleri yapılması; müşteri profili belirleme çalışmaları kapsamında hangi özelliklerdeki müşterilerin hangi ürünleri satın aldıklarının belirlenmesi (kümeleme veya sınıflama); müşteri ihtiyaçlarının belirlenmesi kapsamında farklı müşteri tipleri için en iyi ürünlerin neler olduğunun belirlenmesi ve yeni müşterileri çekmede hangi faktörlerin etkili olacağının tahmini; çok boyutlu özetleme raporları ve istatistiksel özetleme bilgileri şeklinde özetlenebilir.

4.2.6. Şirket analizleri ve yönetimi

Veri ayrıştırmanın şirkete analizlerinde ve risk yönetimi konusundaki kullanımı oldukça yaygındır. Bu yönde yapılan bazı çalışmalar finansal planlama ve aktif varlıkların değerlendirilmesi kapsamında nakit akışlarının analizi ve tahmini, aktif

varlıkların değerlendirilmesi için şüpheli alacakların analizi, zaman serileri, kaynakların planlanması, işletme performansı değerlendirme ve izleme, işletmenin performansı ile ilgili geleceğe yönelik tahminler, kaynakların ve harcamaların karşılaştırılması ve özetlenmesi, rekabetsel incelemeler kapsamında rakiplerin ve pazarlama yönelimlerinin incelenmesi, müşterilerin sınıflara ayrılması ve sınıflara göre fiyat politikası tayini, rekabetin yüksek olduğu bir pazarda fiyat politikalarının belirlenmesidir

4.2.7. Hilekârlıkların tespiti ve yönetimi

Clementine ve/veya SPSS ile geçmişe ait veriler kullanılarak, geçmişte hilekârlık yapmış kişilere ait veriler incelenebilir ve bunlara ait bir model kurulabilir. Geliştirilen bu model kullanılarak hilekârlığa meyilli olanlar tespit edilebilir. Hilekârlık belirlemenin en yaygın kullanım alanları sigortacılık sektörü, finans sektöründe kredi kartı servisleri, perakendecilik sektörü ve telekomünikasyon sektörüdür. Örneğin sigorta poliçelerinde yapılan hilekârlıkların tespiti amaçlı bir model ile hilekârlık yapmaya meyilli gruplar belirlenebilir, farklı müşteri gruplarına uygun poliçe türü tespit edilebilir, maliyeti yüksek poliçeler belirlenebilir ve mevcut poliçeler için poliçelerdeki risk tayini yapılabilir.

4.3. Scanner Data ve Veri Madenciliği

4.3.1. Tanımı

Scanner Data, birçok sektörde ürünler üzerindeki barkotlardan ürüne özel kodun okutulup, veri tabanından bu kodun eşleştiği gerekli bilgilerin alınmasını sağlayan sistemdir.

Perakendecilik sektöründe, öncelikle müşterinin faturasını hesaplamak için veri tabanından ürünle ilgili fiyatların alınmasıyla bu tekniğin kullanımı başlamıştır. Ardından, stok bilgilerinin tutularak stok denetiminin kolaylaşmasında, belirli bir seviyenin altına düşen ürünlerin siparişlerinin verilmesinde, en son olarak da

ürünlerin pazar paylan ve perakende müşterilerinin davranışlarının belirlenmesinde kullanılmaya başlanmıştır.

Alışveriş fişlerini hemen kaydeden barkod aletlerini ve barkoddan gelen bilgileri saklayan ve ileten bilgisayarın gelişmesine bağlı olarak özellikle scanner data kullanılmaya başlandı. Bu teknik, hangi müşterinin hangi ürünü, ne zaman, nereden ve hangi ürünlerle birlikte alındıklarını gösteriyor. Ayrıca scanner data sayesinde kredi kartı ödemeleri bilgileri de kaydedilebiliyor.

4.3.2. Scanner data'nın perakendecilik sektöründe işleyişi

Öncelikle mağazada satılmaya başlanacak ürünlerle ilgili olarak barkod numarası, ürünün türü (içecek, temel tüketim malzemesi, kuru gıda vb.), adı, fiyatı, stok miktarı vb. bilgiler veri tabanına aktarılır. Daha sonra, ürün kasaya geldiği zaman, ürün barkodu kasiyer tarafından barkod okuma aletine okutulur. Okunulan barkod bilgisiyle veri tabanındaki bilgiler eşleştirilir. Kasaya ürün adı ve ürün fiyatı bilgisi aktarılır, eş zamanlı olarak veri tabanından stok miktarından satılan ürün sayısı kadar düşülür. Ayrıca, sistemde bilgiler birikirken, ürünlerin müşteriyle bağlantısını kurmak amacıyla, müşterinin o sepette satın aldığı tüm ürünlerin ortak kaydı ve alışverişin yapıldığı tarih ve saat tutulur ve veri tabanına aktarılır. Eğer perakendecinin kart sistemi mevcutsa, bu anda müşteriye ait demografik bilgiler de elde edilmiş olur ve bu bilgiler veri tabanında müşteri alışveriş bilgilerinden farklı olarak statik bilgi adı altında tutulur.

Sistemin etkin ve doğru bir şekilde kullanılabilmesi için barkodda okutulan ürüne ait karakteristik bilgilerin veri tabanında sürekli olarak yenilenmesi gerekmektedir. Yeni ürün geldiğinde stokların düzenlenmesi, fiyat değişikliği oluştuğunda sisteme fiyatın yeniden kaydedilmesi bu durma örnektir.

Scanner data ile toplanan verilerin arkasındaki bilgilerin elde edilmesi de veri madenciliği sayesinde olur. Veri madenciliği ile bu veriler işlenerek;

- Hangi müşteriye hangi promosyon çalışması ile ulaşılmalıdır?

- Belirli bir müşteride planlanan promosyonun etkili olma olasılığı nedir?
- Bir sonraki dönemde en karlı hisseler tahmin edilebilir mi?
- Bu müşteri borcunu vadelerinde düzenli olarak ödeyebilir mi?
- Programlar için gelecekteki izlenme oranı tahmin edilebilir mi? sorularını cevaplandıracak bilgiler elde edilir.

4.3.3. Scanner data ve veri madenciliğinin beraber işleyişi

Perakendecilik sektöründe scanner data verileri, hangi malların, hangi müşterilerce, nasıl alındığı bilgilerini toplamaktadır. Bunun yanı sıra eldeki mal stoku ile ilgili, mevcut olan her türlü bilgi ve kayıt ayrıca demografik müşteri bilgilerini içeren veri depoları sağlanır.

Veri madenciliği işlemleri ile elde edilen scanner data verileri analiz edilir ve çıkan sonuçlar aşağıdaki konularda kullanılır:

- Stok Düzenleme
- Raf Düzenleme
- Fiyatlandırma, İndirimler
- Promosyon Faaliyetleri
- Tedarikçilerle İşbirliği
- Müşteri Profillerini İzleyerek Bire Bir Pazarlama
- Çapraz Satışlar

4.3.3.1. Stok düzenleme

Scanner data ile birlikte kullanılan barkod okuma sistemi ile satılan ürünler tür ve sayı bazında eş zamanlı olarak takip edilebilmektedir. Barkod sistemi, malların stoğa girişi sırasında da kullanıldığından eldeki stok miktarı düzenli olarak takip edilebilmektedir. Bu sayede, hem kategoriler bazında hem de ürünler bazında stoklarda ki değişiklikler izlenebilmekte, azalmalar anında görülerek gerekli siparişler ve işlemler geç kalınmadan yapılabilmektedir.

4.3.3.2. Raf düzenleme

Veri madenciliği ile elde edilen ürün satış bilgileri doğrultusunda raf düzenlemeleri yapılmaktadır. Hangi ürünlerin daha fazla, hangilerinin daha az satıldığı bilgisi doğrultusunda, çok satılan ürünler göz hizasında satılmayan ürünler ise daha alçak yada daha yüksek yerlere yerleştirilerek daha etkin satışlar yapılabilmektedir. Veri madenciliği çıktıları doğrultusunda promosyon kararı verilen ürünlerin raflarda etkin olarak sunulabilmesi için, müşterilerin bu ürünleri daha kolay fark etmelerini sağlamak amacıyla raf düzenlemesi de yapılmaktadır.

Veri madenciliği sonucunda müşterilerin çoğunlukla beraber aldıkları ürünler hakkında bilgiler elde edilmektedir. Bu bilgiler doğrultusunda, birlikte satın alma oranları yüksek olan ürünler raflarda birbirine yakın konumlandırılarak satışlar artırılmaktadır. Farklı bir uygulamada ise, söz konusu ürünler birbirlerinde daha uzak noktalarda konumlandırılarak, müşterilerin mağaza içinde daha çok zaman geçirmeleri, bu sayede satışların artması sağlanmaktadır.

Ayrıca, veri madenciliğinin ürün satış bilgilerini sunması ile ürünlerin raflarda kaplayacağı alanların büyüklükleri de belirlenebilmekte, raf düzenlemeleri buna göre yapılmaktadır.

4.3.3.3. Fiyatlandırma ve indirimler

Veri madenciliği çıktıları doğrultusunda, hangi ürünlerde yapılacak indirimlerin etkili olacağı belirlenebilmektedir. Bu indirimlerin etkileri de yine veri madenciliği aracılığı ile ölçülebilmektedir.

Örneğin, yapılan bir veri madenciliği araştırmasında ürünler üzerinde yapılan indirimlerin doğrudan satış miktarlarına olumlu yönde yansıdığı tespit edilmiş ve bu bilgi, promosyon çalışmalarında kullanılmaktadır. Aynı şekilde fiyatlandırma politikasının etkileri belirlenebilmekte ve çıktılar doğrultusunda fiyatlandırma politikaları yeniden düzenlenmektedir.

4.3.3.4. Promosyon

Veri madenciliği sonuçlarının en çok kullanıldığı alan promosyon çalışmalarıdır. Elde edilen veriler, promosyon çalışmaları için yol gösterici niteliktedir. Scanner data uygulamalarında, barkod sisteminin yanında müşterilerin demografik bilgilerini içeren kartlarında kullanılması ile promosyon çalışmaları yönlendirilmektedir. Müşterilerin ürün tercihleri, beklentileri, beğenileri, geçmişteki satın alma alışkanlıkları ve geçmiş promosyon çalışmalarına verdikleri tepkiler gibi, tutundurma faaliyetleri için büyük önem taşıyan veriler veri madenciliği ile elde edilir.

Perakendecilik uygulamalarında veri madenciliği çalışması olarak bölgesel bazlı değerlendirmelerde yapılır ve bu değerlendirmeler ile promosyon çalışmaları, bölgeye özgü olarak planlanır. Örneğin, Ege Bölgesi'nde zeytinyağı tüketimi, katı yağ tüketimine oranla fazla iken, Marmara Bölgesi'nde bu durum ters olarak gerçekleşmektedir. Elde edilen bu bilgiler ışığında Ege ve Marmara Bölgeleri'nde promosyonlar farklı ürünlerde yapılmaktadır.

Veri madenciliği ile yapılan bilimsel çalışmalar sonucunda uygulanan promosyon çalışmaları; müşteriler üzerinde olduğu kadar potansiyel müşteriler üzerinde de son derece etkili olmaktadır; çünkü çalışmaların bilimsel temelde yapılması ile genel eğilimler ve tercihler daha doğru olarak belirlenebilmektedir. Bu şekilde yapılan promosyonların potansiyel müşteriler üzerinde de etkili olduğu görülmüştür.

4.3.3.5. Tedarikçilerle işbirliği

Scanner data ile sağlanan geniş veri kümesi, sadece perakendecilere değil, aynı zaman da tedarikçilerine de önemli bilgiler sunmaktadır. Ancak bu bilgiler perakendeciler üzerinden alındığından, tedarikçilerin bu bilgileri almaları perakendeciler ile işbirliği içinde olmaları şarttır. ABD' de Wall-Mart 6 ülkedeki 2900 merkezinde 7,5 TB'lik veri deposuna sürekli olarak alışveriş verilerini kaydetmekte ve 3500 tedarikçisine bu verilere ulaşma ve onları analiz etme imkânı vermektedir. Bu bilgiler bölgesel merkez stoklarını kontrol etme, yeni iş fırsatlarını

bulma ve tedarikçilerin perakendeciye bağlı kalmadan promosyon yapmasını sağlamaktadır.

Aynı zamanda perakendecilerde, bu bilgileri tedarikçileri ile paylaşarak kendi promosyon çalışmaları sırasında oluşacak sorunları en aza indirgemekte, kendilerini güvence altına almaktadırlar. Tedarikçileri bazı promosyon çalışmalarına ikna etmekte (bazı promosyon faaliyetleri tedarikçi desteği olmadan yapılamamakta, bundan dolayı tedarikçi desteğinin sağlanması büyük önem taşımaktadır) ve sonuçlarım somut şekilde göstermekte veri madenciliği sonuçlarından yararlanılmaktadır.

4.3.3.6. Müşteri profillerini izleyerek bire bir pazarlama

Demografik bilgileri içeren müşteri kartları kullanılarak yapılan scanner data ve veri madenciliği faaliyetleri sonucunda daha öncede bahsedildiği gibi, bilgiler müşteri bazında da elde edilebilmektedir. Bu sayede, bire bir pazarlama faaliyetlerinin etkinliğinin artırılabilmesi için gerekli olan bilgiler elde edilebilmekte; yapılan promosyon ve fiyatlandırma çalışmaları müşterilere özel olarak da gerçekleştirilebilmektedir.

Örneğin, müşterilerin doğum günü veya evlilik yıl dönümü gibi bilgilerinin bilinmesi ve tercihleri, beklentileri, beğenileri, geçmişteki satın alma alışkanlıkları, geçmiş promosyon çalışmalarına verdikleri tepkilerin elde edilmesi ile bire bir pazarlama çalışmaları daha etkin olarak gerçekleşebilmektedir.

4.3.3.7. Çapraz pazarlama

Veri madenciliği uygulamaları, bir çok şirketten oluşan holding tipi büyük işletmeler için de önemli veriler sağlamaktadır. Perakendecilik sektörünün yüksek müşteri potansiyeli olması ve çok farklı müşterilere hizmet sunması ile, scanner data için elde edilen verilerin, veri madenciliği analizlerinin çok geniş bir yelpazede sonuçlar sunmasını sağlamaktadır. Bu genişlikte ve bilisel yöntemlerle yorumlanmış veriler ile çapraz pazarlama çalışmalarının etkinliği artmaktadır. Müşterilerin demografik

bilgileri yanında, satın alalarında da gerçekleşen değişikliklerin takip edilebilmesi, çapraz pazarlama faaliyetlerinin geliştirilmesi için önemli veri oluşturmaktadır.

4.4. Veri Madenciliği Modelleri

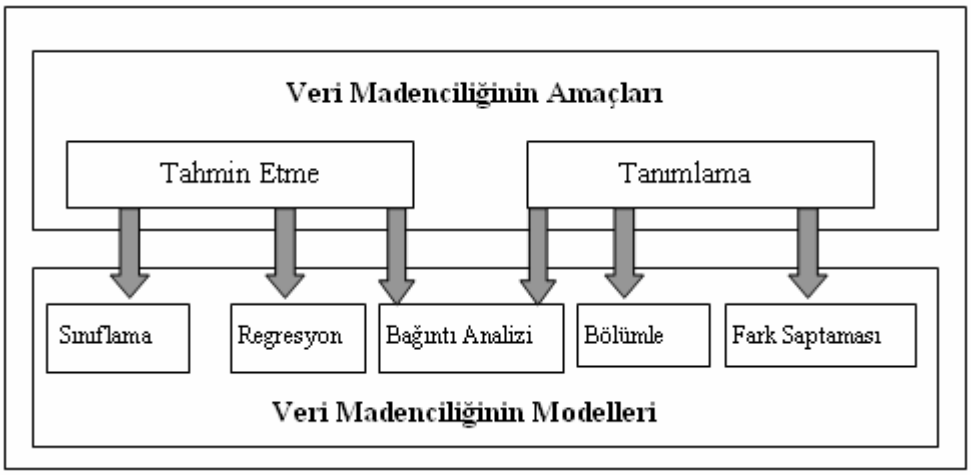
Veri madenciliğinde kullanılan modeller, tahmin edici (predictive) ve tanımlayıcı (descriptive) olmak üzere iki ana başlık altında incelenmektedir[65].

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.

Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntü1erin 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 örüntü1erinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir.

Veri madenciliği modellerini gördükleri işlevlere göre,

- Sınıflama ve regresyon,
- Kümeleme,
- Birliktelik kuralları ve ardışık zamanlı örüntüler, olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir. Veri madenciliği ile modelleri arasındaki ilişki Şekil 4.2'de gösterilmiştir.



Şekil 4.2. Veri madenciliği ve modelleri arasındaki bağıntı.

4.4.1. Sınıflama ve regresyon modelleri

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 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. Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler;

- Karar ağaçları (Decision Trees),
- Yapay sinir ağları (Artificial Neural Networks),
- Genetik algoritmalar (Genetic Algorithms),
- K-en yakın komşu (K-Nearest Neighbor),
- Bellek temelli nedenleme (Memory Based Reasoning),
- Lojistik regresyon (Logistic Regression)'dur.

4.4.2. Kümeleme modelleri

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.

4.4.3. Birliktelik kuralları ve ardışık zaman örüntüleri

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 örüntü1er, pazarlama amaçlı olarak pazar sepeti analizi 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[66].

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

- Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diet süt de satın alırlar.

- Ardışık zamanlı örüntü1er ise 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.

4.5. Veri Madenciliği Algoritmaları

Veri madenciliği süreci sonunda elde edilen örüntü1er kurallar biçiminde ifade edilir. Elde edilen kurallar iki değişkenin eşleştirme derecesini gösterir veya veriyi önceden tanımlanmış sınıflara paylaştırır ya da veriyi tanımlayan sonlu sayıda kümeye ayırır. Bu kurallar veri üzerinde belirli bir tekniğin (algoritmanın) yinelenmesiyle elde edilir. Elde edilen bilginin kalitesi veri analizi için kullanılan algoritmaya büyük ölçüde bağlıdır.

Veri madenciliği algoritmaları iki grupta toplanabilir. Bunlar doğrulamaya dayalı algoritmalar ve keşfe dayalı algoritmalardır. Doğrulamaya dayalı veri madenciliği algoritmasında kullanıcı bir hipotez öne sürer ve sistem bu hipotezi ispatlamaya çalışır. Doğrulamaya dayalı veri madenciliği algoritmalarının en yaygın olarak kullanıldığı yerler, istatistiksel analiz ve çok boyutlu analizdir[67]. Öte yandan keşfe dayalı algoritmalar otomatik olarak yeni bilgi çıkarırlar. Bu bölümün devamında veri madenciliği sistemlerinde kullanılan algoritmalardan önemli olanları incelenecektir.

4.5.1. Hipotez testi sorgusu

Hipotez testi sorgusu algoritması, doğrulamaya dayalı bir algoritmadır. Bir hipotez öne sürülür ve seçilen veri kümesinde hipotez doğruluğu test edilir. Öne sürülen hipotez genellikle belirli bir örüntünün veri tabanındaki varlığıyla ilgili bir tahmindir. Bu tip bir analiz özellikle keşfedilmiş bilginin genişletilmesi veya ayıklanması işlemleri sırasında yararlıdır[68].

Hipotez ya mantıksal bir kural, ya da mantıksal bir ifade ile gösterilir. Her iki biçimde de seçilen veri tabanındaki nitelik alanları kullanılır. X ve Y birer mantıksal ifade olmak üzere "IF X THEN Y" biçiminde bir hipotez öne sürülebilir.

Verilen hipotez seçilen veri tabanında doğruluk ve destek kıstasları baz alınarak sistem tarafından sınanır.

4.5.2. Sınıf1ama Sorgusu

Sınıflama sorgusu, yeni bir veri elemanını daha önceden belirlenmiş sınıflara atamayı amaçlar[69]. Veri tabanında yer alan çoklu1ar bir sınıflama fonksiyonu yardımıyla kullanıcı tarafından belirlenmiş ya da karar niteliğinin bazı değerlerine göre anlamlı ayrık alt sınıflara ayırır. Bu yüzden sınıflama, denetimli öğrenmeye girer. Sınıflama algoritması bir sınıfı diğerinden ayıran örüntüleri keşfeder. Sınıflama algoritmaları iki şekilde kullanılır[70].

- Karar Değişkeni ile Sınıf1ama: Seçilen bir niteliğin aldığı değerlere göre sınıflama işlemi yapılır. Seçilen nitelik karar değişkeni adını alır ve veri tabanındaki çoklular karar değişkeninin değerlerine göre sınıflara ayrılır. Bir sınıfla yer alan çoklular karar değişkeninin değeri açısından özdeştir.

- Örnek ile Sınıf1ama: Bu biçimdeki sınıflamada veri tabanındaki çoklu1ar iki kümeye ayrılır. Kümelerden biri pozitif, diğeri negatif çoklu1arı içerir.

Yaygın kullanım alanları, banka kredisi onaylama işlemi, kredi kartı sahteciliği tespiti ve sigorta risk analizidir.

4.6. Veri Madenciliği Teknikleri

Sadece dört temel veri madenciliği operasyonları varken, bu işlemleri mümkün kılan çok sayıda veri madenciliği teknikleri vardır. Normal olarak veri madenciliği sistemleri bu tekniklerin her birini desteklemez fakat veri madenciliği sistemleri çoğu zaman kullanıcının özek problemlerine bağlı olarak seçebileceği farklı teknikler arasından iki ya da daha fazlasını birleştirir[71]. Bundan dolayı, potansiyel bir kullanıcı, ihtiyaçlarına en iyi uyacak sisteme karar vermek amacıyla en yaygın teknikler üzerinde ön bilgiye sahip olması gerekir. İdeal olarak yukarıda tanımlanan işlemlerin her biri sadece bir tekniğe ayrılması gerekir. Bu işlem nispeten belirli bir

sistem için kararı basit1eştirecektir. Ne yazık ki bazı operasyonlar sadece bir tekniği kullanarak tamamen gerçekleştiremezken, bazı teknikler birden fazla operasyonu destekler.

4.6.1. İstatistiksel yöntemler

Veri madenciliği çalışması esas olarak bir istatistik uygulamasıdır. Verilen bir örnek kümesine bir kestirici oturtmayı amaçlar. İstatistik literatüründe son elli yılda bu amaç için değişik teknikler önerilmiştir. Bu teknikler istatistik literatüründe çok boyutlu analiz (multivariate analysis) başlığı altında toplanır ve genelde verinin parametrik bir modelden (çoğunlukla çok boyutlu bir Gauss dağılımından) geldiğini varsayar. Bu varsayım altında sınıflandırma (classification; discriminant analysis), regresyon, kümeieme (clustering), boyut azaltıma (dimensionality reduction), hipotez testi, varyans analizi, bağıntı kurma (association; dependency) teknikleri istatistikte uzun yıllardır kullanılmaktadır[72].

4.6.1.1. Binominal test

Binom modeli, istenilen sonucun olma olasılığı p iken, n bağımsız denemede tam x adet istenilen sonucun olması olasılığını veren modeldir.

Örnek: Demir bir para ile yazı tura atıldığında, yazı gelme olasılığı 1/2dir. Bu hipoteze dayanarak 40 defa yazı tura atılarak sonuçlar bir yere not edildiğinde, atılanların % 3/4'ünün yazı olması ve göz1emlenen anlamlılık derecesinin küçük (0,0027) olması durumunda, olasılığın 1/2 ihtimalinden uzak olması yani atılan paranın hileli olması söz konusudur.

4.6.1.2. Kümeleme analizi

Kümeleme analizi, bireylerin veya uyarıcıların benzerliklerine göre gruplarda veya kümelerde toplanmasını amaçlayan birçok değişkenli istatistik analizidir. Ayırma (diskriminant) analizinden farklı olarak kümeleme analizinde faktör analizindeki gibi veri matrisi analiz öncesi tahmin ve kriter alt setlerine bölüştürü1mez. Kümeleme

analizinde dikkatler, bireylerin araştırmada ölçülen tüm değişkenler üzerindeki değerlerini hesaba katarak ortaya çıkacak kümeler veya gruplar üzerinde toplanmıştır. Bireyler arasındaki benzerlikleri saptamak amacıyla uzaklık ölçüleri, korelasyon ölçüleri veya nitelik verilerinin benzerlik ölçüleri kullanılabilir[73].

Kümeleme analizinin pazarlama sorunlarının çözümüne uygulanması oldukça yaygın bir yöntemdir. Pazar bölümlenmesi, pazar testinin uygulanacağı bölgelerin saptanması bu konuda örnek verilebilecek birkaç konudur.

4.6.1.3. Ayırma (diskirminant) analizi

Ayırma analizi, iki veya daha fazla sayıdaki grubun ayırımı ile ilgilenen birçok değişkenli ilgi analizidir. Amaçları arasında analiz öncesi tanımlanmış iki veya daha fazla sayıda grubun ortalama nitelikleri arasında önemli farkların olup olmadığının test edilmesi, gruplar arasındaki farka her bir değişkenin katkısının saptanması ve grup içi değişime oranla gruplar arasındaki ayrımı maksimize eden tahmin değişkenleri kombinasyonunun belirlenmesi sayılabilir[74].

Ayırma analizi sonuçlarının test edilme olanağının bulunması sonuçların geçerliliğini ve güvenilirliğini ve dolayısıyla analizin gücünü artıran önemli bir etmendir.

4.6.1.4. Faktör analizi

Faktör analizi veriler arasındaki ilişkilere dayanarak verilerin daha anlamlı ve özet bir biçimde sunulmasını sağlayan birçok değişkenli istatistiksel analiz türüdür. Amaç esas olarak değişkenler arasındaki karşılıklı bağımlılığın kökenini araştırmaktadır[75].

Örnek: Pazarlama araştırmacısı tüketicilerin marka tercihleri, mağaza tercihleri, sosyo-ekonomik demografık ve psikolojik nitelikleriyle ilgili çeşitli verileri toplayabilir. Ancak, araştırmacının son amacı, tüketicilerin çeşitli markalara karşı tutumları veya eğilimleri gibi bazı temel değişkenlerin veya boyutların saptanmasıdır. Tüketicilerin markalara tutumları, aile büyüklüğü ve satın alma sıklığı

gibi çeşitli değişkenlerle ölçülebilir. Şayet bu tür değişkenler arasında önemli korelasyonlar var ise 'markalara karşı tutum' bir faktör olarak kabul edilir.

4.6.1.5. Ki - kare testi

Ki-kare ilgi analizi pazarlama araştırmalarında çok yaygın olarak kullanılan bir istatistiksel analiz türüdür. Bu yaygın kullanımın en önemli nedenleri, çok basit bir analiz türü olması, varsayımlarının azlığı ve çok güçsüz ölçeklerde ölçülmüş verilere uygulanabilmesidir. Amaçları ise;

- Örnek değerlerinin dağılımının belirli bir teorik dağılıma uyma derecesinin saptanması (uygunluk testi),

- İki veya daha fazla nitelik esas alınarak sınıflandırılan veriler değerlendirilerek bu nitelikler arasındaki ilginin derecesinin belirlenmesi (bağımsızlık testi)[76].

Araştırmacının amacı, örnek değerlerinde gözlenen ilgi hakkında bir yargıya varmaktır. Odak noktası bireylerin seçilen bazı nitelikleridir. İlginin fonksiyonel formunun doğrusal olması gerekmez. Analiz doğrusal olmayan ilişkilere de uygulanabilir.

Örnek: Belirli tip bir elektrik rezistansının dayanıklılığım test etmek amacıyla 360 rezistans tesadüfî olarak seçilmiş ve belli gözlem değerleri saptanmıştır. Dağılımın %5 önem derecesinde normal dağılımdan gelip gelmediğini anlamak için ki-kare uygunluk testi yapılabilir.

4.6.1.6. Korelâsyon analizi

Korelasyon analizi esas olarak tahmin ve kriter değişkenleri arasındaki ilginin yönü ve derecesi ile ilgilenir. Analizin en önemli varsayımı değişkenler arasındaki ilginin doğrusal olduğu yönündedir[77]. İlginin derecesini ölçmede korelasyon katsayısı "r"

kullanılır. Basit korelasyon analizinden söz edilebileceği gibi, çoklu korelasyon analizi yapmak da mümkündür.

4.6.1.7. Varyans analizi

İkiden fazla ana kütle aritmetik ortalamasının karşılaştırılması ile ilgili testte izlenecek süreç ANOVA tablosu ile özetlenebilir. Buna göre F test istatistiği varyans analizi yardımıyla kullanılır[78]. Farklı ana kütlelerden seçilen örnek aritmetik ortalamaları arasındaki farkların karelerinin ortalaması, her bir örneğin kendi içindeki farkların karelerinin ortalamasına bölünür. F test istatistiği belirlendikten sonra sonuca varılır.

Örnek: Bir firma yöneticileri yeni ambalaj makineleri satın almayı planlamaktadır. Buna göre piyasada en çok tutulan üç marka ambalaj makinesinden hangisini satın almaları gerektiğine karar verebilmek için her bir makine 5'er saat çalıştırılmış ve saat başına ambalaj miktarları saptanmıştır. Bu verilere dayanarak % 1 önem derecesinde firma yöneticilerinin üç makinenin üretim miktarları arasında önemli bir fark olup olmadığım test etmeleri gerekir ve verilere varyans analizi uygularlar.

4.6.2. Bellek tabanlı yöntemler

Bellek tabanlı veya örnek tabanlı bu yöntemler (memory-based, instance-based methods; case-based reasoning) istatistikte 1950'li yıllarda önerilmiş olmasına rağmen o yıllarda gerektirdiği hesaplama ve bellek yüzünden kullanılamamış ama günümüzde bilgisayarların ucuzlaması ve kapasitelerinin artmasıyla, özellikle de çok işlemcili sistemlerin yaygınlaşmasıyla, kullanılabilir olmuştur[79]. Bu yönteme en iyi örnek, k- en yakın komşu algoritmasıdır (k-nearest neighbor).

4.6.3. Karar ağaçları

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

- Kuruluşlarının ucuz olması,
- Y orum1anmalarının kolay olması,
- Veri tabanı sistemleri ile kolayca entegre edilebilmeleri,
- Güvenilirliklerinin daha iyi olması, nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahiptir[80].

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı sahalar;

- 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;

- 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[81].

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. AlD 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 coşku 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 birçok araştırmacı tarafından yayınlar yapılmıştır.

4.6.4. Yapay sinir ağları

Sinir ağları sınıflandırmada kullanılan başka bir tekniktir. Sinir ağları kayıtlardan sonuçlar çıkarmak amacıyla, bir kayıta ait değişkenlerin değerlerini aritmetik olarak birleştirir. Sinir ağları insan beyninin yapısını taklit eden veri modelleridir. Biyolojik sistemde olduğu gibi bağımsız sinir hücreleri birbirini aktifler. Yapay sinir ağları, "biyolojik sinir sisteminde olduğu gibi, sistem dışındaki nesnelerle etkileşebilen, birbirinden bağımsız ve paralel olarak çalışabilen işlem elemanları ve onların hiyerarşik bir şekilde organizasyonu" olarak tanımlanır. Beyin gibi sinir ağları da bir dizi veri girişleri sayesinde öğrenir ve veri tabanından örüntü1er bulmak amacıyla öğrenmiş olduğu bilgilere dayalı olarak modelin parametrelerini ayarlar. Modelin işlem yapısı, en doğru sonuçları üretmek için giriş değişkenleri ağırlıkları arasında bağıntı bulmayı amaçlar. Bu işlem sinir ağını veri ile eğiterek yapılır. Eğitim bazı öğrenme etkilerine sebep olur. Sinir ağlarını eğitmeye yönelik başka bir yöntemde, sinir ağına veri tipine özgü eğitilmemiş ağ örnekleri sunmak ve sonuç çıkış örneklerinin olması gerekenle arasındaki farka göre ağırlıkları ayarlama yaklaşımıdır. Bu ayar tekrar tekrar ve birçok giriş örnekleri, ağı tatmin edici bir şekilde çalışana kadar sürdürülür[82].

4.6.4.1. Yapay sinir ağlarının temel özellikleri

- Örneklerden öğrenme: Yapay sinir ağlarına, öğrenmesi istenen girdi/çıktı İlişkilerinin örnekleri verilir. Yapay sinir ağları bu örnekleri kullanarak genellemeler yapar.

- Biçim tanıma ve sınıflandırma: Yapay sinir ağlarına örnekler girdi olarak verilir ve yapay sinir ağları, oluşturulan girdi/çıktı eşleşmeleri ile bilgiyi depoladığı yerdeki yayılı belleğini kullanarak karşılık gelen çıktıyı üretir.

- Eksik bilgileri tamamlama: Ağın eksik bilgileri yeniden oluşturabilme özelliğidir. Eksik bilgiye sahip bir örnek verildiğinde ağ, eksik örnekteki kayıp olan bilgiyi belleğinde bulunan tam örnekteki bilgilerle bağdaştırarak, eksik örnekteki kayıp bilgiye karşılık gelen tam örnekteki bilgiyi bulabilir.

- Kendi kendine adapte olabilme: Bazı yapay sinir ağları, kendi kendine öğrenme yeteneğine sahiptir. Ortamda bazı değişiklikler olduğunda, bu tür sinir ağları bu yeni duruma kendilerini adapte edebilirler.

- Hatalara tolerans gösterme: Bazı işlem elemanlarının ağdan çıkarılması veya olmaması durumunda yapay sinir ağının başarısız olması gibi bir durum söz konusu değildir. Bilgi, bütün ağ boyunca yayılı olduğundan bazı bilgilerin kayıp oluşu veya yok edilişi, ağın performansım fazla etkilemeyecektir. Bu durum, kötü sonuçlar doğurmayacak şekilde performans azalmasına yol açmasına rağmen, sistem performansının tamamen başarısız olmasına sebep olmayacaktır. Bu özellik, hesaplamada ufak bir eksikliğin kötü sonuçlara yol açabileceği kritik ortamlarda çok faydalı olacaktır.

- Eksik bilgilerle çalışabilme: Bulanık veya eksik bilgiler ağa sunulduğu zaman yayılı bellek girdi için en uygun olan çıktıyı seçer. El yazısını tanıma bu özelliğe güzel bir örnektir.

4.6.4.2. Yapay sinir ağlarında öğrenme

İşlem elemanı ve ağın yapısı tasarlandıktan sonra, yapay sinir ağının eğitme işlemi başlatılabilir. Sinir ağlarının en önemli özelliği, eğitme yeteneğidir. Bir sinir ağında öğrenmenin anlamı, ağın belirli bir probleme ait olduğu çıktıları üretmesini sağlayacak optimum ağırlık değerinin bulunmasıdır. Bilgi ağ boyunca bağlantılarda ağırlıklar şeklinde dağıldığı için tek bir bağlantı herhangi bir anlamlı bilgiyi ifade etmez. Daha doğrusu, anlamlı bir bilgi oluşturmak için işlem elemanlarına sahip olan bir bağlantı grubu tasarlamak gerekmektedir. Bilgiler bağlantılar üzerinde dağıtıldığından yapılacak bir eğitme faaliyetinde ağ bir bütün olarak göz önüne alınır. Problemin çözümü için ağın bağlantılarına ait doğru ağırlık değerlerine sahip olması gerekmektedir. Bu eğitme olarak adlandırılan bir işlem vasıtası ile gerçekleştirilir[83]. Öğrenme, ağırlık değerlerinin nasıl değiştirileceğini ifade eden bir öğrenme kuralına dayanır. Geliştirilen birçok öğrenme kuralı vardır. Bu öğrenme kuralının temel ilkesi, benimsenen öğrenme stratejisi ile tanımlanır. Literatürde üç tip öğrenme stratejisinden söz edilir.

Denetimli öğrenme: Ağı eğitmek için bir öğretici gerektirir. Öğretici, çıktı katmanında ağ kararının ne olması gerektiğini söyler. Ayrıca problemin ağ tarafından iyi bir şekilde kavranabilmesi için öğrenmede kullanılacak olan örneklerin seçimi de öğretici tarafından yapılır. Bir girdi ve doğru çıktı örneği ağa verilir. Ağ, girdiyi işleyerek çıktıyı üretir ve üretilen çıktıyı doğru çıktı ile karşılaştırır. Bağlantılardaki ağırlıklar, daha iyi çıktıyı üretmek için yeniden ayarlanır ve bu işlem kabul edilebilir bir hata seviyesine ulaşıncaya kadar devam eder.

Destekli öğrenme: Destekli öğrenme de, bir öğretici gerektirir. Ancak çıktının ne olması gerektiği ağa söylenmez. Ağa bildirilen, üretilen çıktının doğru ve ya yanlış olup olmadığıdır.

Denetimsiz öğrenme: Denetimsiz öğrenme, ise bir öğreticiye gerek duymaz. Bu stratejide ağ, girdi/çıktı eşleştirmesini düzenlemek için kendi ölçütlerini geliştirir. Bu sebeple, denetimsiz öğrenme stratejisini kullanan ağlar, kendi kendini organize eden ağlar olarak adlandırılır.

Yapay sinir ağlarında bilgi temsili çok önemlidir. Ağ yapısı ne kadar güzel tasarlanırsa ya da öğrenme ne kadar iyi gerçekleşirse gerçekleşsin, eğer bilgi tutarlı olmayan bir şekilde temsil edilirse, çözüm sonuçları da tutarlı ve isabetli olmayacaktır.

Öğrenme seti, ağın öğretilmesinde kullanılan girdi ve çıktılardan oluşan bir settir. Denetimli öğrenmede, olması gerek çıktılar sette bulunurken diğer iki stratejide bulunmaz.

Yapay sinir ağlatının iki ana problemi vardır[84];

Birincisi**;** sonuç ağının kapalı bir kutu olması ve sonuçların yeterince açık olmamasıdır. Böylece açıklamaların eksikliğinden dolayı, güven, kabullenme ve elde edilen sonuçların uygulanması zor olmaktadır.

İkincisi ise; yapay sinir ağlarının olabildiğinden daha doğru gibi görülen sonuçlar üretmesidir. Giriş verisi olarak ondalık basamağı olmayan veriler kullanılmasına rağmen, tüm müşterilerin %12,235'i belli bir sınıfa aittir sonucu üretmesi buna verilebilecek tipik bir örnektir.

Sinir ağları, test verisinin sonuçlarına bağlı olarak yeni veri üzerinden tahminde bulunmada oldukça iyi sonuçlar verir. Ayrıca, yüksek hacimli verileri işleyebilirler fakat bunu yapmak için uzun öğrenme zamanına ihtiyaç duyarlar. Sinir ağları bundan dolayı, verinin lineer olmayan pek çok etkileşimleri olduğunda hedef değişkeni tahmininde faydalıdırlar ve bu açıdan diğer istatistikî tekniklerden güven aralığı daha yüksek sonuçlar verirler. Fakat bu ilişkilerin açıklanmasına ihtiyaç duyulduğunda istenilen neticeyi vermezler.

4.6.5. Görselleştirme

Görselleştirme, klasik manada bir teknik değildir. Görselleştirme, veri tabanından verinin grafiksel özetlerini sağlar. Bundan dolayı, analizi yapan kişiye diğer veri

madenciliği teknikleri kullanarak çıkarılan bilginin iç yüzünü daha derinlemesine kavramaya yardımcı olur. Görselleştirme, gözlemcinin dikkatini önemli örüntü1er ve eğilimler üzerine odaklamasına ve bunları derinliğine incelemesine imkân verir. Veri tabanındaki satır ve sütunlar taranarak özelliklerin bulunması zordur. Grafiksel olarak bakıldığında çoğu zaman özellikler daha aşikâr olmaktadır. Görselleştirme araçları, analiz için metot olarak İnsan algılamasından faydalanır. Görselleştirme sayesinde kullanıcıyı veri analiz işleminin içine entegre etmek mümkün olmaktadır. Kullanıcı verinin bütün hacmi altında ezilmez, fakat ilgili, yararlı bilgiyi iyi araştırmasına imkân sağlar. Görselleştirme insanların sezgisi ve yaratıcılığı ile bilgisayarların hesaplama güçlerinin birleşmesine izin verir[85].

OLAP (Online Analytical Processing Technique) çok bilinen görselleştirme tekniğidir. OLAP, özel bilgi teknolojisi bilgisine ihtiyaç olmaksızın, karmaşık, çok boyutlu veriye hızlı bir bakışa, kavrayışa izin veren bir dizi yazılım araçlarından oluşur. Fakat diğer veri madenciliği tekniklerinin aksine OLAP, bir analizcinin bir sorguyu formüle etmesine gereksinim duyar. Çok sayıda sordu yaparak, ilginç bilgi parçası ya da eğilimi bulunabilir. OLAP yaklaşımı bundan dolayı, analizcinin sezgisine güvenir[86].

Görselleştirme, belirli bir veri madenciliği işlemine tahsis edilemez. Görselleştirme her veri madenciliği işlemini destekleyen yardımcı bir araçtır. Bundan dolayı, veri madenciliği ve veri görselleştirme, özellikle birbirlerini çok iyi tamamlar. Veri madenciliği, hem belirli görselleştirme metodu hem de görüntülenen bilginin tipini değiştirmek için kullanıcıya kolay bir şekilde ve hızlıca müdahale izni veren etkileşimli görselleştirme araçlarının kullanımım gerekli kılar.

4.6.6. Sepet analizi

Sepet analizinde amaç alanlar arasındaki ilişkileri bulmaktır. Bu ilişkilerin bilinmesi şirketin karını arttırmak için kullanılabilir. Eğer X malını alanların Y malını da çok yüksek olasılıkla aldıklarım biliyorsanız ve eğer bir müşteri X malını alıyor ama Y malını almıyorsa o potansiyel bir Y müşterisidir.

Örneğin internet üzerinden kitap satan Amazon şirketi, Book Matcher adlı programıyla müşterilerine okudukları ve sevdikleri kitaplara göre satın almaları için kitap tavsiye etmektedir.

Eğer elimizdeki veride mallar için sadece satın alındı/alınmadı bilgisi varsa, sepet analizinde mallar arasındaki bağıntı, destek ve güven kıstasları aracılığıyla hesaplanır. İki mal, X ve Y, için destek ve güven tanımları şöyledir:

Destek: P(X ve X) = X ve Y mallarını satın almış müşteri sayısı / Toplam müşteri sayısı

Güven: P(X/Y)=P(X ve Y)/P(Y) = X ve Y mallarını satın almış müşteri sayısı / Y malını satın almış müşteri sayısı

Destek veride bu bağıntının ne kadar sık olduğunu, güven de Y malını almış bir kişinin hangi olasılıkla X malını alacağını söyler. Bağıntının önemli olması için her iki değerin de olabildiğince büyük olması gerekir.

Eğer elimizde malların müşteri tarafından ne kadar tüketildiği, ne kadar beğenildiği ile ilgili bilgi varsa o zaman bağıntı daha iyi hesaplanabilir. Örneğin süper markette müşterinin aylık toplam X malı kullanma miktarı hesaplanabilir. Amazon'un Book Matcher programı, okuyuculara okudukları her kitap için 1 ile 5 arasında bir beğeni notu vermelerini ister. Bu durumda X ve Y nümerik veriler olduğundan X ile Y'nin korelasyonu hesaplanabilir:

Corr(X,Y)= Cov(X,Y) / (Std(X)*Std(y))
X ile Y'nin kovaryansı birbirlerine göre doğrusal olarak nasıl değer aldıklarını belirtir:

$Cov(X,Y)=E[(X-m_x)\ (Y-m_y)]$

m_x X'lerin ortalaması, std(X)'de standart sapmasıdır. Örneğimizde m_x, X malının ortalama olarak ne kadar beğenildiğini, std(X) de beğenilerin bu ortalama etrafında ne kadar değişken olduğunu gösterir.

Eğer X'i sevenler genelde Y'yi de sevdiyse hem X, hem de Y değeri ortalamadan daha yüksek olacak ve Cov(X,Y)>O olacaktır. Aynı şekilde X ve Y beraber beğenilmiyorsa her iki değer de ortalamadan küçük olacak ve yine Cov(X, Y)>O olacaktır. Eğer X'i beğenenler Y'yi beğenmediyse (veya aksi takdirde) değerlerden biri ortalamadan yüksek, diğeri ortalamadan düşük olacak ve Cov(X, Y)<O olacaktır. Corr(X,Y)'de Cov(x,y)'nin -1 ile +1 arasında standart sapmalara göre normalize edilmiş halidir. Corr(X,Y) değerinin 0 olması X ile Y arasında (doğrusal) bağlantı olmadığını, negatif değer ters, pozitif değer de doğrudan bağıntı olduğunu gösterir.

Bu şekilde olası bütün mallar arasında korelasyon bilgileri varsa X'i kullanan ve seven kişiye tavsiye edilecektir. Y, müşterinin kullanmadığı diğer bütün mallar arasında X ile korelasyonu en fazla ve olabildiğince l' e yakın olan mal olmaktadır.

BÖLÜM 5. TEDARİK SİSTEMİNİN İNCELENMESİ VE MODELİN KURULMASI

5.1. Giriş

Tedarik etme veya satın alma kavramının önemi uzun yıllardan beri bilinmektedir. Geçmiş yıllarda firmaların nihai ürünün bütün parçalarını üretmeleri bir saygınlık konusu olarak görülmüştür. Ancak günümüzde rekabetin de zorlaması ile ürün detaylarının artması, kullanılan parça adedinin çoğalması ve bu yeni parçalar için yeni ve pahalı yatırımların gerekmesi firmaları bazı parçaların tedarik edilmesi fikrine yöneltmiştir. Bu durum; tedarik, tedarikçi, tedarik yönetimi, tedarikçi ilişkileri yönetimi gibi kavramları öne çıkarmıştır.

5.2. Tedarikçi Seçimi ve Tedarikçi İlişkileri Yönetimi

5.2.1. Tedarikçi yönetimi

Tedarikçi yönetimi; toplam maliyetin minimizasyonu için tedarikçilerin yönetimi çalışmalarının bütününe verilen addır. Tedarikçiler, alımın bir kereye mahsus ya da sürekli yapılmasının söz konusu olmasına göre ve tedarikçi ile kurulması düşünülen stratejik ilişkiden mesafeli ilişki biçimlerine kadar genişleyen bir yelpazede ayrıma tabi tutulmalıdır [87]. Tedarikçi yönetimi aynı zamanda tedarik merkezi sayısında indirimin sağlanmasını da içermektedir. Çünkü birçok işletme gereğinden fazla sayıda tedarikçi firma ile ilgilenmek durumunda kalmaktadır.

Bir işletme, tedarik merkezi sayısını azaltarak, daha az sayıda tedarikçi ile harcamalarında düzenlemeye, böylece daha düşük toplam maliyete ulaşabilir. Daha az tedarikçi, aynı zamanda, kilit tedarikçiler ile daha iyi ilişkilerin geliştirilebilmesi

anlamına da gelmektedir.

5.2.2. Tedarikçi ilişkileri yönetimi

Tedarikçi ilişkileri yönetimi (Supplier Relationship Management), işletmelerin; tedarikçiden neyi ne kadara aldıkları, tedarikçiden kaynaklanan risklerin boyutlarının ne olduğu, alınan ürünlerin kalitesinin firma kalite hedeflerine uygunluğu, satın alma uygulamalarında zaman içerisinde yaşanan değişiklikler, satın alma etkinliklerinin firma genel hedeflerine uygunluğu gibi cevabını aradıkları soruların cevaplanmasına yardımcı olan yönetim sistemidir [88].

Tedarik zinciri yönetimi kullanımının gündeme gelmesi ile birlikte, tedarikçi ilişkileri yönetimi kavramı da ortaya çıkmaktadır. Tedarikçi ilişkileri yönetimi, tedarikçilerin değerlendirilmelerinin dışında, var olan tedarikçilerle kurulacak olan iletişimin organizasyonunu ve yönetim sorumluluklarını içermektedir. Bu amaçla günümüzde kullanılan yazılımlar tedarikçi - üretici arasında ihtiyaç duyulan bilgi akışının son derece hızlı, koordineli ve amaca hizmet edebilir yapıda olmasını sağlamaktadır. Bu şekilde paylaşılan bilgi, gerek üreticilerin gerekse bunlara ait tedarikçilerin stok ve üretim maliyetlerinin azalmasını mümkün kılar.

Tedarikçi ilişkileri yönetimi, kilit tedarikçilerin belirlenmesi süreci ile başlayıp en uçtaki tedarikçiye kadar genişleyen bir yelpazede geliştirilecek stratejileri, yaklaşımları ve organizasyonu içerisinde barındırır. Tedarikçi ilişkileri yönetimi, uzun vadede, tedarikçi değerlendirme sürecinin, özellikle niteliksel kriterlerin oluşmasında önemli bir rol oynamaktadır[89].

5.2.3. Tedarikçi seçimi karar süreci

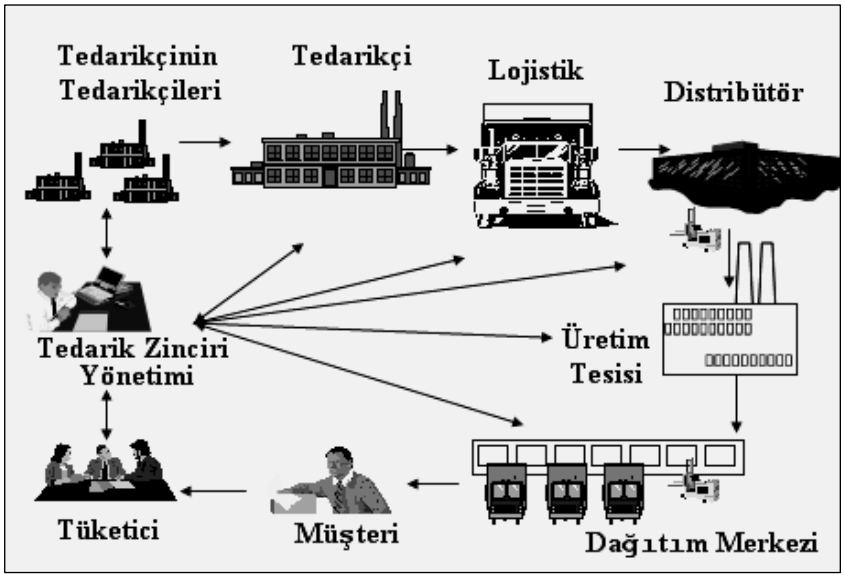
Tedarikçi seçimi kararını verirken göz önünde bulundurulması gereken en önemli noktalar şunlardır:

- Birçok ürünün esasını satın alınan materyaller (hammadde ve malzemeler) oluşturur.

- Tedarikçilerden kaliteli materyaller alınması önemlidir.
- Tedarikçi seçimi kritiktir.
- İşletmeler, çoğu kez tedarikçilerine büyük miktarda yatırım yaparlar.
- Rekabetçi indirimlerden yararlanmaya çalışmak yerine, makul tedarikçi seçimi tercih edilmelidir[87].

Tedarikçilerin seçiminde, değerlendirme yaparken, tek bir mükemmel yol olduğu önyargısı kesinlikle yanlıştır. Seçim metodu, bir çok türde faktöre dayanmaktadır. Bunlar:

- Sözleşme tek bir kaynağı mı yoksa birden fazla tedarikçiyi mi içermektedir?
- Fiyat ve kalitenin bağıl önemi nedir?
- Tedarikçi ile uzun vadeli bir ilişki istenmekte midir?
- İşletmenin ve tedarikçilerin birlikte olmalarından oluşacak bağıl güç nedir?
- Tedarikçi tasarıma destek verecek midir, yoksa sadece tedarik mi edecektir?
- Hepsinin üstünde, işletme tedarikçilerin riskini minimize etmek ve değerlerini ise maksimize etmek amacındadır. [89].



Şekil 5.1 Tedarik Zinciri[90]

5.2.4. Tedarikçi seçiminde izlenen değerlendirme prosedürü

Bu aşamada, işletme; yeni bir ürün veya ürün bileşeni için yeni bir tedarikçiye ihtiyaç duyabilir ya da mevcut bir tedarikçiyi değiştirmek istiyor olabilir. İşletme öncelikle, bir tedarikçiyi seçerken kendisi için nelerin önemli olduğunu belirlemelidir. Bu bilgi değerlendirme sürecini sonlandırmaya yarayacaktır[87].

- Kaynak temini stratejisinin belirlenmesi
- Potansiyel tedarik kaynaklarının belirlenmesi
- İlk belirleme; havuzdaki tedarikçiler
- Tedarikçi değerlendirme ve seçme metodunun belirlenmesi
- Tedarikçinin seçimi için bir başlangıç tedarikçi değerlendirme ve seçme şablonu oluşturulması

5.2.5. Tedarikçi seçiminde kullanılan tedarikçi değerlendirme kriterleri

Tedarikçi değerlendirme ve seçme aşamasında, tüm bileşenler için geçerli olan üç ana kriter söz konusudur. Bunlar:

- Fiyat
- Kalite
- Teslim

Bunların yanı sıra, kritik bileşenler için daha derin bir araştırma yapılması, yani daha farklı kriterler gereklidir. Bu kriterler ana başlıklar halinde Tablo 5.1'deki gibi sıralanabilir:

Tablo 5.1. Tedarikçi seçim kriterleri[87]

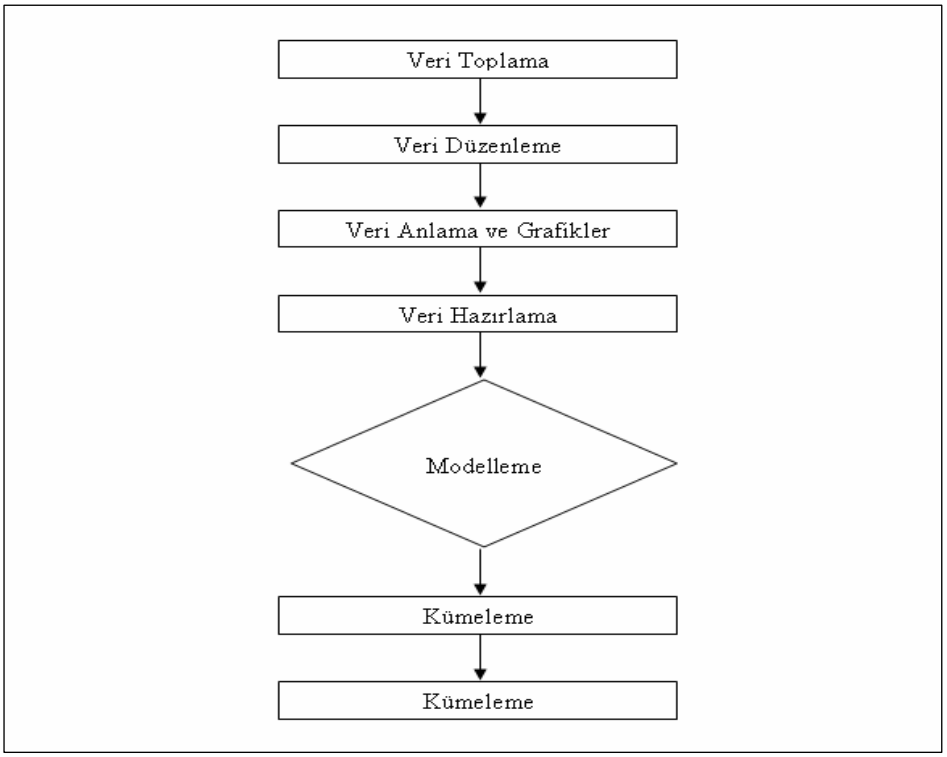
Tedarikçi Seçim Kriterleri	
Fiyat	Maliyet Hesaplama Prosedürleri
Finansal Uygunluk	İşletme Geçmişi
Tavırlar	Garantiler
Eğitim Kaynakları	Bilgi Paylaşımı
Tesislerin Konumu	Şirket Ünü
Bilgi Teknolojileri Kaynakları	Kalite Sistemi
Kapasite	İşgücü İle İlişkiler
Hız	Paketleme İmkânları
Teslimat Performansı	Nakliye Yetenekleri
Tazminat	Çevrim Süresi
Zamanında Teslimler	Esneklik
Ürün Çıkış Doğruluğu	Bağımlılık Oluşturabilirlilik
Stok Dışı Kalma Sıklığı	Sipariş Çevrim Zamanı
Sipariş Süreç Uyumluluğu	Gecikme Zamanı
Ürün Bulunabilirliği	Elverişlilik
Güvenilirlik	Faturalandırma Hataları
Hak Talebi Uyumsuzluk Sayısı	Kalite Kontrol

5.3. Mevcut Yapının İncelenmesi

Bu çalışmada, Adapazarı'nda kurulu bulunan ve özerk bir kamu kuruluşu olan, Türkiye'nin vagon üretimi yapan tek işletmesi Türkiye Vagon Sanayi A.Ş (TÜVASAŞ)' ın tedarikçi verileri kullanılmıştır. Adı geçen işletme, yıllık 150 adet çeşitli tiplerde vagon imalatı ile 700 adet vagon tamiri kapasitesine sahiptir. Ancak kapasite kullanımının artırılabilmesi için işletmede yeni yatırımlara ve mevcut yapının modernizasyonuna ihtiyaç vardır. Fabrikada 2005 yılı itibarı ile 55 adet çeşitli tipte vagon üretimi gerçekleştirilmiştir. Fabrikanın kendi bünyesinde ürettiği parçaların oranı %10'un altında, dışarıdan tedarik ettiği hammadde, yarı mamul ve mamul parçaların oranı ise %90'ın üzerindedir[91]. Bu kadar yüksek oranda tedarik ihtiyacı olan bir işletmenin tedarikçileri ile olan ilişkileri ve uyguladığı tedarik yöntemi önem arz etmektedir.

Bu çalışmada, işletmenin 2004 ve 2005 yılında yapmış olduğu ihalelerde, ihale kazanan 400 firma içerisinde örnekleme yöntemi ile 94 firma seçilmiştir. Bu firmalara ait ihale bilgileri ve firmaların demografik bilgileri birleştirilerek bir veri seti oluşturulmuştur. Bu veri seti, bir veri madenciliği çözüm paketi olan Clementine 9.0 yazılımı kullanılarak analiz edilmiş ve elde edilen sonuçlar işletmenin yararına kullanılacak şekilde düzenlenmiştir.

Veri toplama işleminden analizin sonuçlanmasına kadar yapılan işlerin akış diyagramı Şekil 5.2'de verilmiştir



Şekil 5.2. Veri Analizi Süreci Akış Diyagramı

5.4. Veri Toplama Aşaması

İşletmenin bilgisayar sistemine geçişindeki gecikme nedeniyle, tüm verilerin elektronik ortamdan alınması mümkün olmamıştır. Bu nedenle, ihtiyaç duyulan ve farklı birimlerde dağınık olarak bulunan veriler toplanarak birleştirilmiştir. Ancak kullanılması düşünülen tedarikçi verilerinin güncellenmesi ihtiyacı doğmuştur. Bu amaçla işletmenin satınalma ve planlama dairelerinin, geçmiş yıllarda tedarikçilerine gönderdikleri "Tedarikçi bilgi Formu" geliştirilerek, "Tedarikçi Bilgi Güncelleme

Formu" adı altında yeniden tedarikçilere gönderilmiş ve alınan cevaplar bir veritabanında kaydedilerek tedarikçi verileri güncellenmiştir. Güncellenen tedarikçi bilgi formu Şekil 5.3'de verilmiştir.

TÜVASAŞ GENEL MÜDÜRLÜĞÜ	**TARİH:**	**DOKÜMAN NO:**
TEDARİKÇİ BİLGİ GÜNCELLEME FORMU		
Kuruluş Şekli	☐ Kamu ☐ A.Ş ☐ Ltd. Şti. ☐ Adi Ort. ☐ Diğer	
Firma Adı		
Firma Yetkilisi		
Firma Adresi		
Tel. Ve Fax No:	e-Mail Adresi:	Web Adresi:
Firma Tipi (Üretici/Satıcı/İthalatçı..vb)		
Firmanızın Faaliyet Alanı (Sektörü)		
Ürettiğiniz/Sattığınız Ürünler		
Yıllık Üretim Kapasiteniz (Adet, Ton, Kg…/Yıl)		
Kalite Güvence Belgesi (ISO,TSE,TSEK ..vs)		
Ürettiğiniz/Sattığınız Ürünlerin Garanti Belgesi Var mı?		
Firmanızın İş Tecrübeleri (Referansları)		
Firmanızdaki Teknik ve İdari Personel Sayısı, Unvanları		
Makine/Teçhizat/Araç Parkınız		
Firmanız Bünyesinde AR-GE ve/veya Kalite Kontrol Departmanı Var mı?		
Firma Yetkilisinin Adı Soyadı, Unvanı:	Tarih: İmza:	
Araştırma ve Pazarlama Şube Müdürü	Satınalma ve Ticaret Daire Başkanı	

Şekil 5.3. Tedarikçi bilgi güncelleme formu

Oluşturulan veri setinde değişkenler, firma bilgileri ile ihale bilgilerinden oluşmaktadır. Model, bu değişkenler dikkate alınarak kurulmuştur. Veri setini oluşturan değişkenler Tablo 5.2'de verilmiştir.

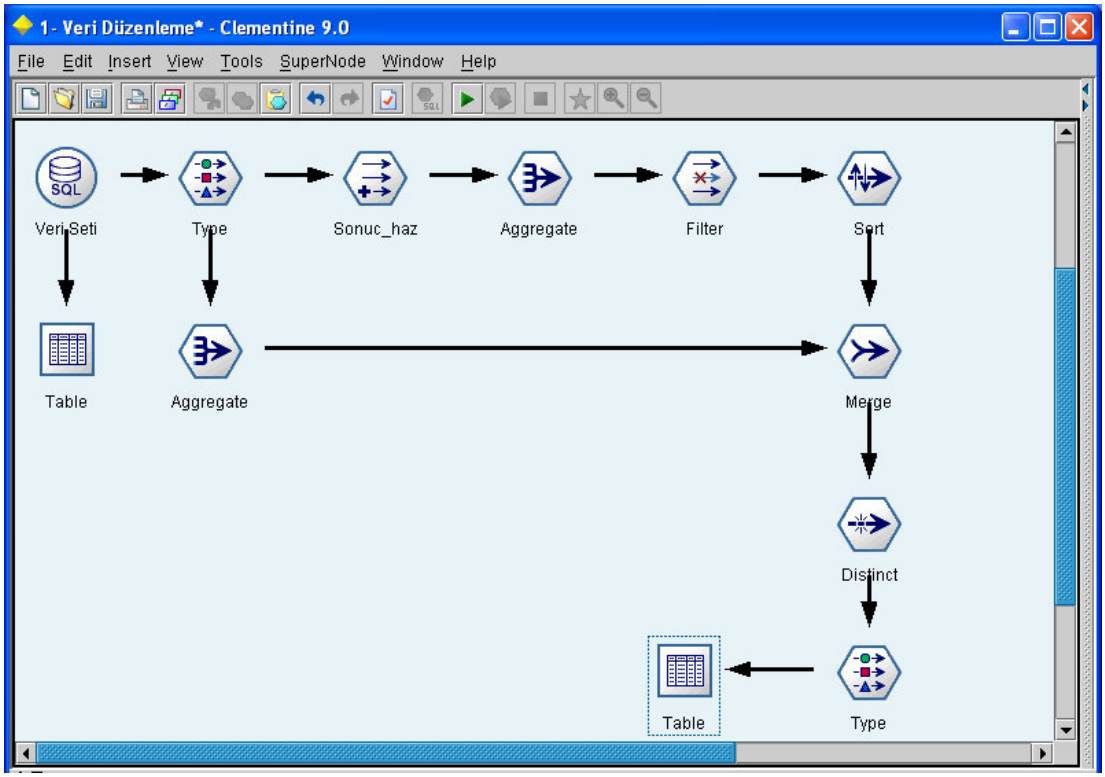
Tablo 5.2. Veri setini oluşturan değişkenler

Firma Bilgileri	**İhale Bilgileri**
Firma Kodu	Dosya No
Firma Adı	Sipariş No
İli	Sipariş Tarihi
Kazandığı İhale Sayısı	İhale Tarihi
Kuruluş Şekli	Malzeme No
Firma Tipi	Malzeme Adı
Sektörü	Miktarı
Kalite Belgesi	Fiyatı
Garanti Belgesi	Sözleşme Tarihi
Teknik Personel Sayısı	Teslim Tarihi
İdari Personel Sayısı	Teslim Miktarı
Ar-Ge/Kalite Kontrol Departmanı	Sonuç

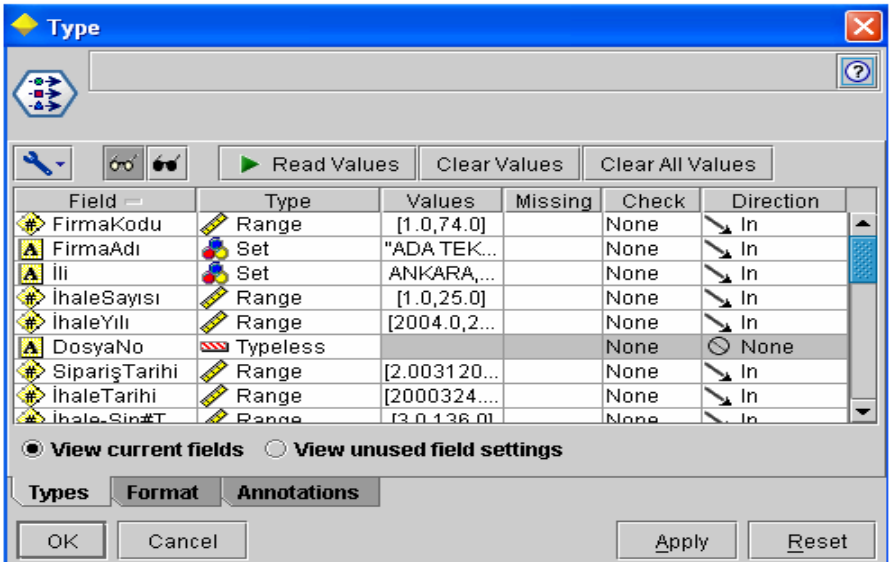
5.5. Veri Düzenleme

Excel tablosu olarak oluşturulan veri seti önce Clementine programına tanıtılmıştır. Veriler, Clementine yazılımı içerisinde bulunan nodlardan yararlanılarak programın istediği formata uygun olarak düzenlenmiştir. Veri düzenleme ekranı Şekil 5.4'te gösterilmiştir

Veri seti programa tanıtıldıktan sonra, Şekil 5.5'te fonksiyonları gösterilen "Type" nodu ile değişkenler tanımlanmıştır. Kurulacak modelde ihtiyaç duyulan değişkenler seçilmiş, modelde yer almayacak olan değişkenler pasif konuma getirilmiştir



Şekil 5.4. Veri Düzenleme Clementine Ekranı



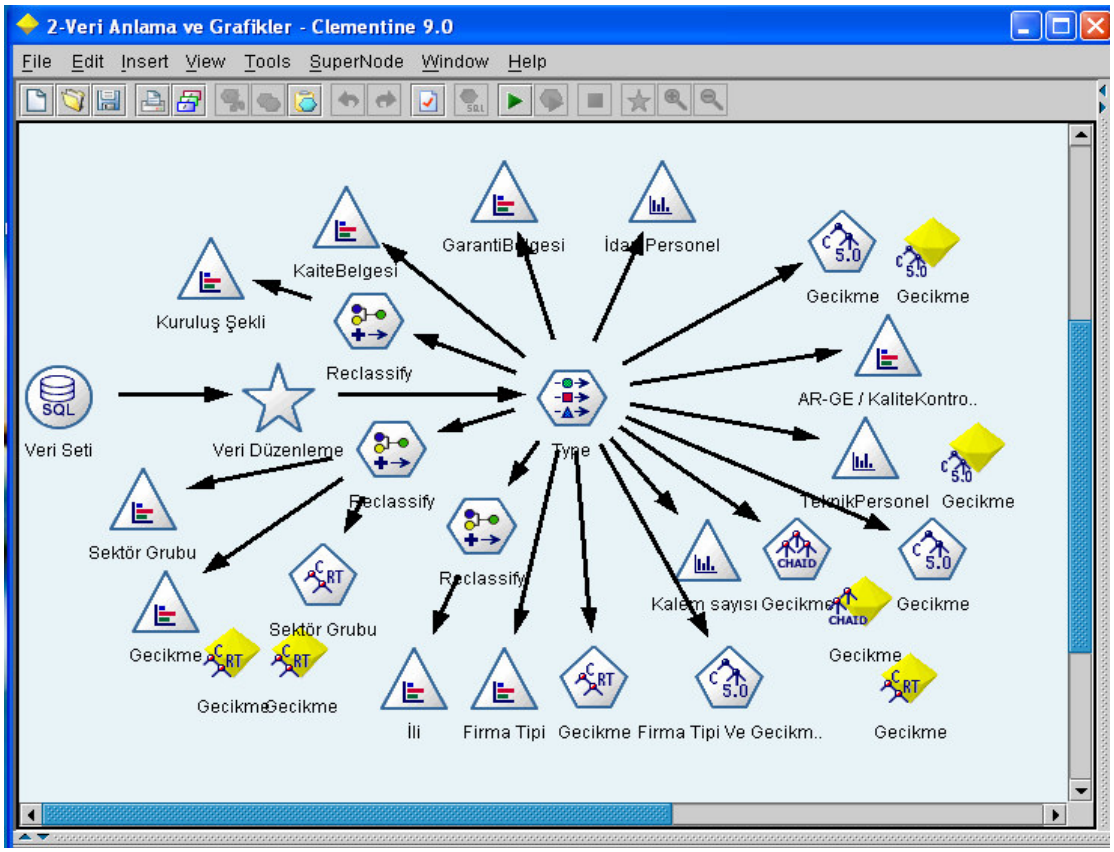
Şekil 5.5. Type Nodu Fonksiyonları

Derive nodunda kriter tanımı yapılmıştır. Buna göre, zamanında teslim = 1, geç teslim, eksik teslim ve red olunan durumlar = 0 olarak tanımlanmıştır.

Aggregate nodunda yukarıdaki kritere göre gecikmeler ve zamanında teslim sayıları hesaplattırılmış, filter nodu ile tabloda yer alması istenmeyen değişkenler devre dışı bırakılmış ve yeni tabloda veriler düzenli hale getirilmiştir.

5.6. Veri Anlama ve Grafikler

Veri anlama ve grafik elde etme aşamasında, veri düzenleme aşamasında yapılan işlemler bir süper nod icerisinde toplanarak veri setine bağlanmıştır. Mevcut yapıya bir type nodu ilave edilerek değişkenler yeniden tanımlanmış ve karar değişkeni olarak "gecikme" seçilmiştir. Şekil 5.6'da veri anlama ve grafikler aşaması gösterilmiştir.



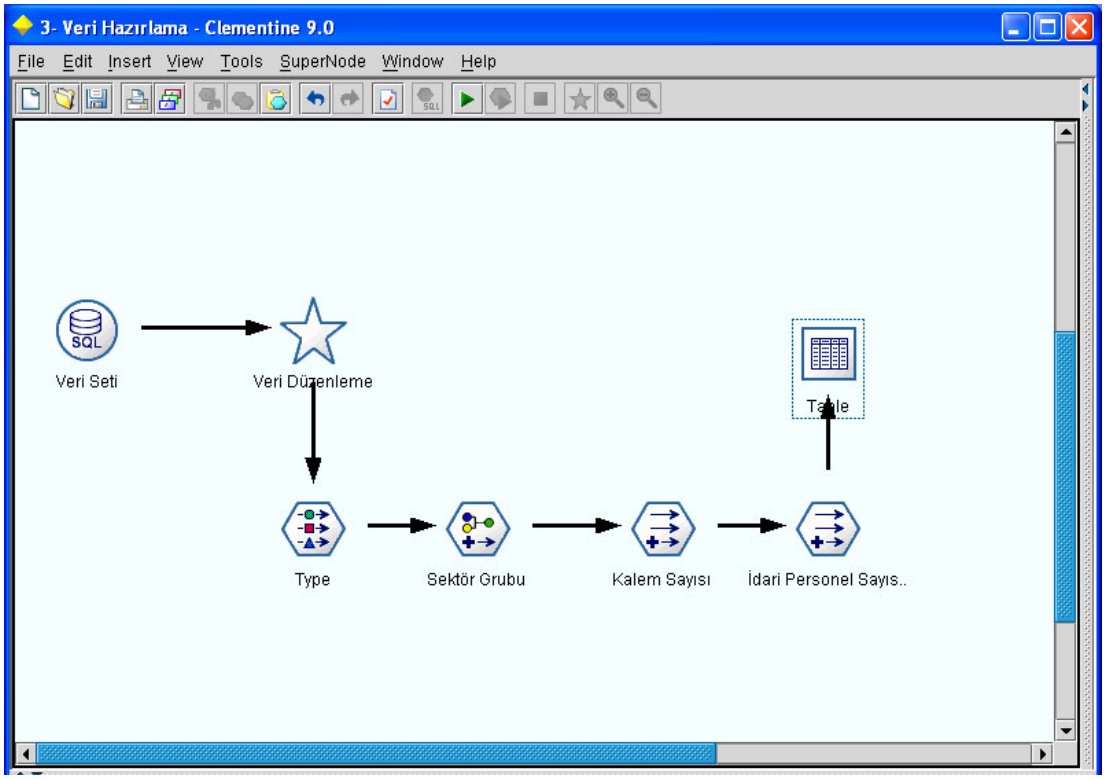
Şekil 5.6. Grafikler ve Veri Anlama Clementine Ekranı

Veri anlama aşamasında her değişkenin gecikme ile ilişkisi grafik ya da uygun algoritmanın çözümü olarak elde edilmiştir. Elde edilen bu sonuçlar, sonuç bölümünde ayrıntılı olarak verilmiş, her bir sonucun anlamlılığı istatistik olarak test edilmiş ve yorumlanmıştır.

5.7. Veri Hazırlama

Veri düzenleme aşaması bir süper nod içerisinde toplanarak, şeması Şekil 5.7'de gösterilen veri hazırlama aşamasına bağlanmıştır. Bir önceki aşamada elde edilen sonuçlar, bazı değişkenlerin yeniden tanımlanması gereğini ortaya çıkarmıştır.

Firmaların bağlı oldukları sektörler ile gecikme arasında anlamlı bir ilişki bulunmakla beraber, veri setinde tanımlanan sektör sayısının fazla olması ve sektör başına düşen firma sayısının azlığı, karar verme sürecinde dikkate değer olmadığından, benzer sektörler, sektör grupları şekline dönüştürülmüştür. Böylece, sektör grubu başına düşen firma sayısı artırılmış ve sektör grubu ile gecikme arasındaki ilişkinin karar sürecinde yorumlanması kolaylaşmıştır



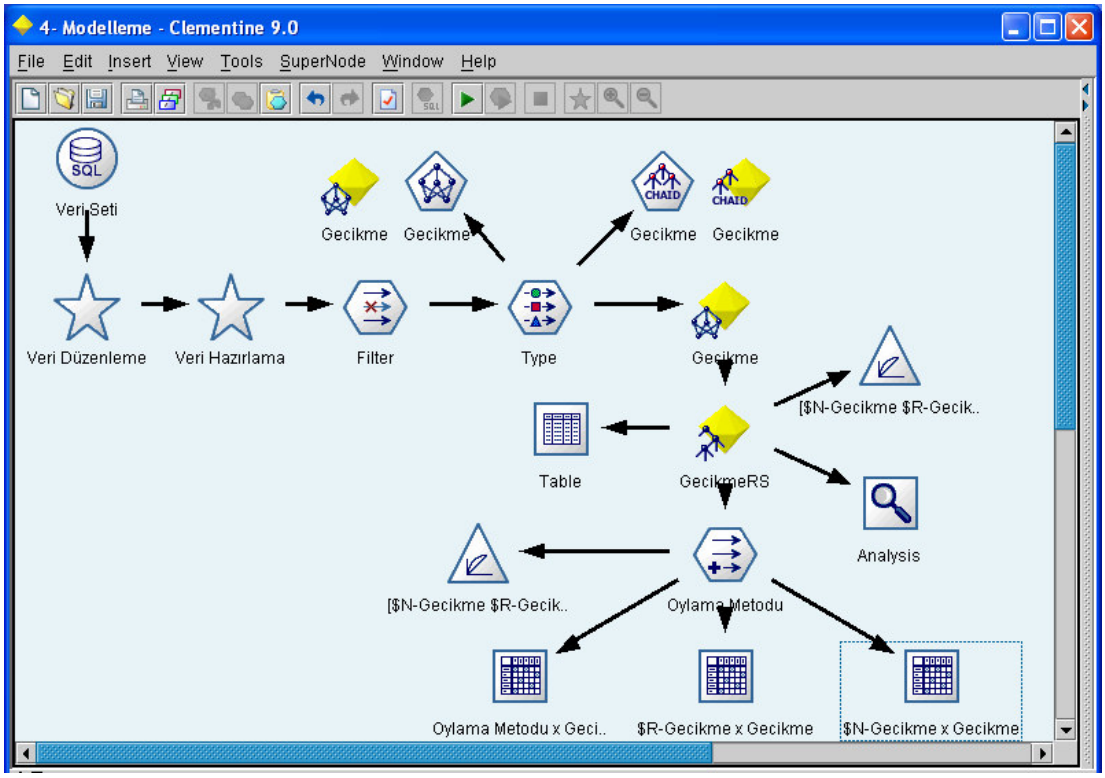
Şekil 5.7. Veri Hazırlama Clementine Ekranı

Veri anlama aşamasında, bir ihalede satın alınan malzemelerin sayısı (kalem sayısı) ile gecikme arasında bir ilişki bulunmuş ve kritik nokta 5 olarak tespit edilmiştir. Bir ihalede satın alınacak malzeme kalemlerinin sayısı 5 ya da daha az ise zamanında teslim oranı yüksek, 5'ten fazla ise gecikme oranı yüksek çıkmıştır. Bu durum

dikkate alınarak ihalelerdeki kalem sayıları 5'ten küçük olanlar ve büyük olanlar şeklinde yeniden düzenlenmiştir.

Gecikme ile ilişkisi bakımından anlamlı bulunan değişkenlerden biri de firmaların idari personel sayılarıdır. İdari personel sayısında kritik nokta 2 olarak tespit edildiğinden, idari personel sayısı 2'den az olanlar ve 2'den fazla olanlar hesaplanarak bu değişken yeniden düzenlenmiştir.

5.8. Modelleme



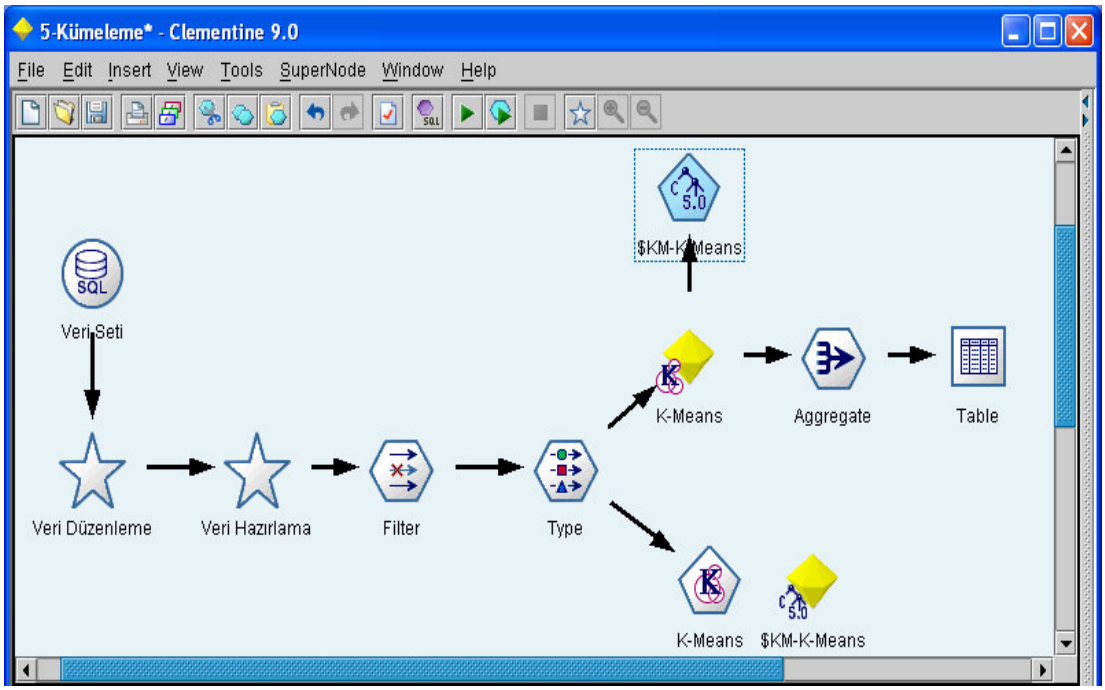
Şekil5.8. Modelleme Clementine Ekranı

Veri düzenleme aşamasında tanımlanan, veri anlama aşamasında gecikme ile ilişkisi tespit edilip anlamlı bulunan, veri düzenleme aşamasında yeniden düzenlenen değişkenler Şekil 5.8'de görüldüğü gibi bu aşamada modellenmiştir. Karar değişkeni olarak belirlenen gecikme ile ilişkili olan tüm değişkenler modelde yer almıştır. Bu aşamada, oluşturulan yeni veri seti, karar ağacı ve yapay sinir ağı algoritmaları kullanılarak analiz edilmiştir. Her iki algoritmanın gecikme tahminleri elde edilmiş ve bu tahminlerin etkinlikleri grafik olarak ifade edilmiştir. Bu algoritmaların,

tahminlerinde uyuştukları ve ayrıştıkları durumlar belirlenerek, her iki tahminin mukayese edildiği tablolar oluşturulmuştur.

5.9. Kümeleme

Veri setinde yer alan firma sayısının azlığı nedeniyle, firmalar 2 kümeye ayrılmıştır. Kümeleme işlemi için, K-Means kümeleme algoritması kullanılmıştır. Kümeleme ekranı Şekil 5.9'de gösterilmiştir.



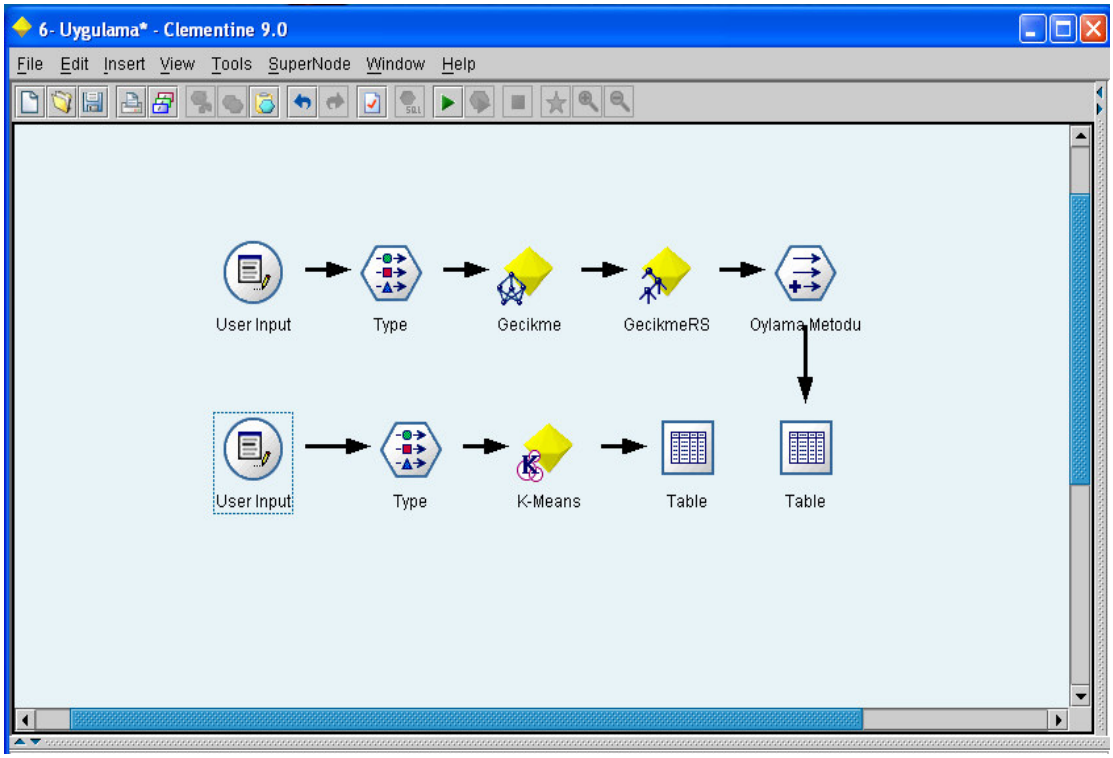
Şekil 5.9. Kümeleme Clementine Ekranı

Hiç gecikmesi olmayan firmalar Cluser-2'de, en az bir gecikmesi olan firmalar Cluster-1'de yer almıştır. Bu iki kümenin elemanları, C5.0 karar ağacı algoritmasıyla yeniden analiz edilmiş ve hangi değişkenin bu kümelerde ne oranda temsil edildiğine ilişkin karar ağaçları oluşturulmuştur.

5.10. Uygulama

Veri madenciliği standart sürecinde deployment olarak adlandırılan bu aşamada, kurulan model dinamik bir yapıya dönüştürülmüştür. Buna göre sisteme giren yeni tedarikçi bir firmanın, ihale süreci başlamadan ve mal teslimi gerçekleşmeden,

sadece firmanın demografik bilgileri içerisinde model tarafından anlamlı bulunan değişkenler girilerek, firma hakkında genel bir kanaat sahibi olmak ve bu firmanın hangi kümenin bir elemanı olacağını tahmin etmek mümkün olacaktır. Bu durum, modelin amacına ulaştığının bir göstergesi olarak kabul edilebilir. Şekil 5.10'da modeli dinamik hale dönüştüren sistemin yapısı gösterilmiştir.



Şekil 5.10. Uygulama Clementine Ekranı

BÖLÜM 6. SONUÇLAR

Veri madenciliğinin temel amacı, işletmelerin bünyesinde bulunan ve tek başına bir anlam ifade etmeyen verilerin, bir program çerçevesinde derlenip, uygun teknikler kullanılarak analiz edilmesi ve bu verilerden bilgi çıkarılmasıdır. Bu çalışmada bahsedilen amaca uygun olarak veri seti oluşturulmuş ve profesyonel bir yazılım kullanılarak analizler yapılmıştır. Teknik altyapısı son derece zengin olan yazılımın içerisinden, belirlenen hedefe uygun olan algoritmalar model içerisinde denenmiş ve en sağlıklı çözümü veren algoritmaların sonuçları dikkate alınmıştır. Clementine ile elde edilen sonuçlar Ki-Kare bağımsızlık testine de tabi tutularak, anlamlı ilişkiler perçinlenmiştir.

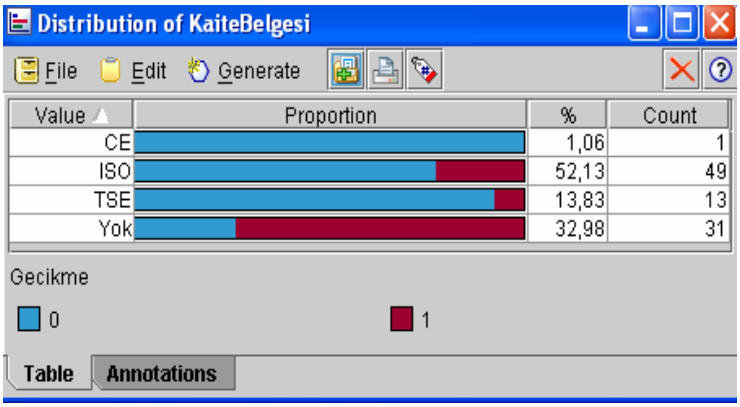
Analizlerde, yapay sinir ağı, karar ağacı algoritmalarından C&R Tree, C5.0, Chad ve Quest, kümeleme algoritmalarından K-Means ve Kohonen kullanılmıştır.

Veri setinde bulunan değişkenlerin içerisinde, karar değişkeni olarak "gecikme" seçilmiştir. Her bir değişkenin gecikme ile olan ilişkisi kurulan modelde incelenmiş ve gecikmeyi etkileyen değişkenler tespit edilmiştir

Bu çalışmanın sonucunda, uygulama yapılan işletmenin tedarikçilerinin analizi, veri madenciliği standart sürecinde yer alan bütün aşamalar dikkate alınarak yapılmış ve aşağıdaki sonuçlar elde edilmiştir. Başarı düzeyi ölçülürken "Gecikme = 1" , "Zamanında teslim = 0" şeklinde ifade edilmiştir. Bütün tablo, şekil ve grafiklerde yer alan 0, başarıyı; 1 ise başarısızlığı temsil etmektedir.

Her bir değişken ile ilgili olarak, önce Clementine ile elde edilen grafik, tablo ya da karar ağacı diyagramı verilmiş, daha sonra ki-kare testi ile elde edilen sonuç tabloları ve bu sonuçların yansıtıldığı şekiller çizilmiş, elde edilen tüm sonuçlar kendi başlıkları altında yorumlanmıştır.

1. Kalite Belgesi: Kalite Belgesi, Şekil 6.1'de görüldüğü gibi model tarafından anlamlı bulunmuş ve belgesi olan firmaların olmayanlara oranla daha başarılı olduğu sonucuna ulaşılmıştır.



Şekil 6.1. Kalite Belgesi-Gecikme İlişkisi

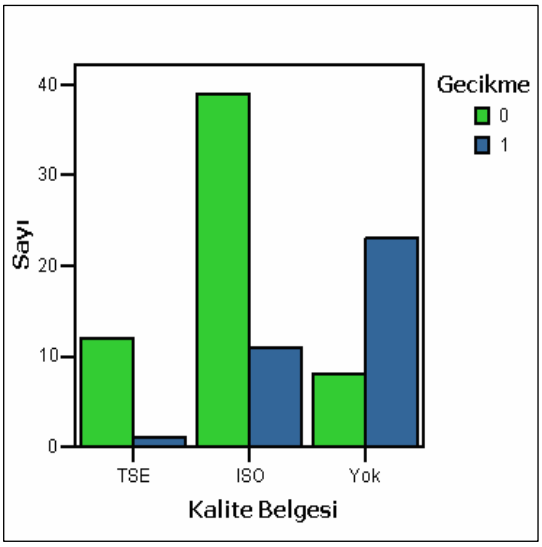
Kalite Belgesi olan 63 firmada gecikme oranı daha düşük, buna karşılık kalite belgesi olmayan 31 firmada bu oran oldukça yüksek çıkmıştır. CE belgesine sahip firma sayısı 1 olduğundan, sonraki aşamalarda bu firma ISO içerisinde değerlendirilmiştir. Elde edilen sonuçlar %5 anlam düzeyinde Ki-Kare bağımsızlık testi ile istatistik açıdan test edilmiş ve sonuçları Tablo 6.1'de gösterilmiştir.

Tablo 6.1. Kalite Belgesi-Gecikme İlişkisi

			Gecikme		**Toplam**
			0	**1**	
Kalite Belgesi	ISO	Sayı	39	11	50
		Beklenen	31,4	18,6	50,0
	TSE	Sayı	12	1	13
		Beklenen	8,2	4,8	13,0
	Yok	Sayı	8	23	31
		Beklenen	19,5	11,5	31,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		27,939		2	0,000

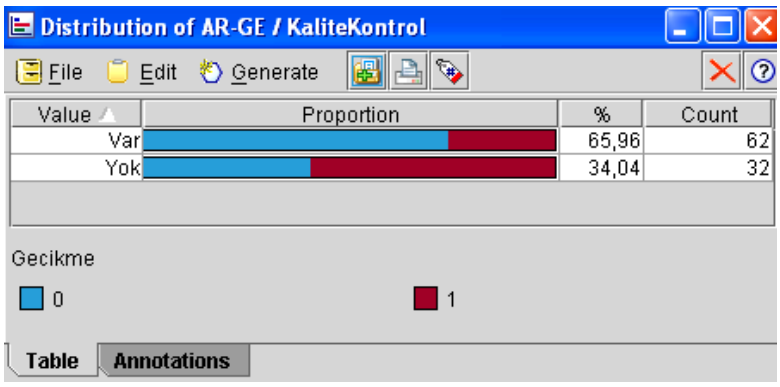
Tablo 6.1'de verilen sonuçlara göre kalite belgesi ile gecikme arasındaki ilişki çok önemli düzeyde anlamlıdır. Yukarıdaki tabloda, ISO kalite belgesine sahip olan 50 firmadan, zamanında teslim etmesi beklenen firma sayısı 31,4 olmasına karşın, 39

firma zamanında teslimat yapmış, 18,6 firmanın gecikmesi beklenirken sadece 11 firma gecikmiştir. TSE belgesine sahip olan 13 firmada da zamanında teslim oranı yüksek, gecikme oranı düşük çıkmıştır. Oysa herhangi bir kalite belgesi olmayan 31 firmada, zamanında teslim beklenen 19,5 firmaya karşın sadece 8 firma zamanında teslim gerçekleştirmiş, gecikmesi beklenen 11,5 firmaya karşın 23 firma gecikmiştir. Elde edilen bu sonuçlar, kalite belgesi ile gecikme arasındaki ilişkinin son derece anlamlı olduğunu ortaya koymuştur. Kalite belgesi ile gecikme arasındaki ilişkinin grafiği Şekil 6.2'de gösterilmiştir.



Şekil 6.2. Kalite Belgesi-Gecikme İlişkisi Grafiği

2. AR-GE/Kalite Kontrol Departmanı: AR-GE/Kalite Kontrol Departmanı ile gecikme arasındaki ilişki Şekil 6.3'te gösterilmiştir.

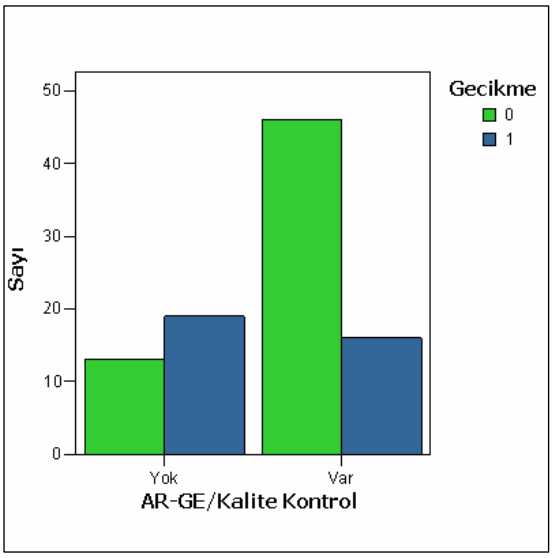


Şekil 6.3. AR-GE/Kalite Kontrol Departmanı - Gecikme Arasındaki İlişki

Model, veri setindeki AR-GE/Kalite Kontrol Departmanı değişkenini anlamlı bulmuş ve bu departmana sahip olan 62 firmanın gecikme oranının düşük, bu departmana sahip olmayan 32 firmanın gecikme oranlarının ise yüksek olduğunu tespit etmiştir. Elde edilen sonuç, istatistik açıdan test edilmiş ve test sonucu Tablo 6.2'de, grafiği de Şekil 6.4'te verilmiştir.

Tablo 6.2. Ar-Ge/Kalite Kontrol Deparmanı -Gecikme İlişkisi

			Gecikme		Toplam
			0	1	
AR-GE/ Kalite Kontrol Departmanı	Var	Sayı	46	16	62
		Beklenen	38,9	23,1	62,0
	Yok	Sayı	13	19	32
		Beklenen	20,1	11,9	32,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		10,177		1	0,001

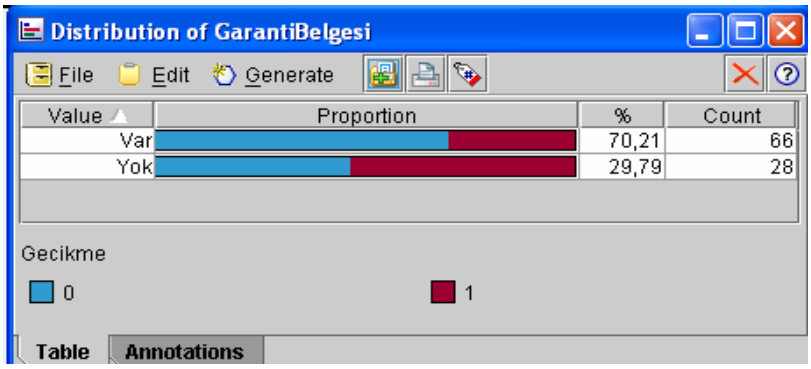


Şekil 6.4. Ar-Ge/KK Departmanı-Gecikme İlişkisi Grafiği

Tablo 6.2'de verilen sonuçlara göre Ar-Ge/Kalite kontrol departmanı ile gecikme arasındaki ilişki çok önemli düzeyde anlamlıdır. Sonuç tablosunda da görüldüğü gibi Ar-Ge/KK departmanına sahip olan 62 firmada, zamanında teslim beklenen 38,9 firmaya karşılık 46 firma (%74,19) başarılı olmuş, bu departmana sahip olmayan 32

firmadan sadece 13 firma (%40,6) zamanında teslim gerçekleştirmiştir. Elde edilen sonuçlar firmaların, Ar-Ge ve/veya kalite kontrol departmanına sahip olmalarının başarılarına olumlu katkı sağladığını ortaya koymuştur.

3. Garanti Belgesi: Firmaların ürettikleri ya da sattıkları malzemeler için garanti belgesi olup olmaması incelenmiş ve sonuçları Şekil 6.5'te verilmiştir. Bu sonuçlara göre garanti belgesi model tarafından anlamlı bulunmuş ve karar verme sürecinde etkili olan unsurlar içerisinde yer almıştır. Garanti belgesine sahip olan 66 firmada gecikme oranı düşük, belgesi olmayan 28 firmada bu oran daha yüksek çıkmıştır.



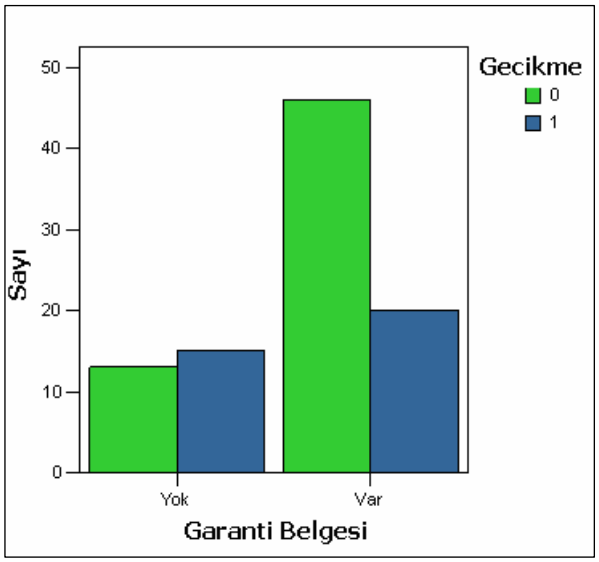
Şekil 6.5. Garanti Belgesi - Gecikme İlişkisi

Clementine ile elde edilen sonuca Ki-Kare testi uygulandığında Tablo 6.3'teki sonuçlar elde edilmiştir.

Tablo 6.3. Garanti Belgesi - Gecikme İlişkisi

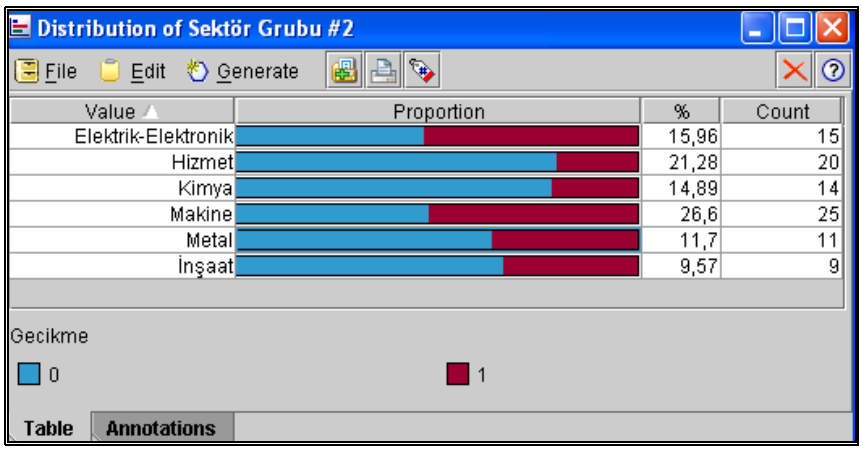
			Gecikme		Toplam
			0	1	
Garanti Belgesi	Var	Sayı	46	20	66
		Beklenen	41,4	21,6	66,0
	Yok	Sayı	13	15	28
		Beklenen	17,6	10,4	28,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		4,555		1	0,033

Tablo 6.3'te elde edilen sonuçlara göre garanti belgesi ile gecikme arasındaki ilişki anlamlıdır. Garanti belgesi olan 66 firmadan 20 firma (%30,3) gecikmiş, buna karşılık garanti belgesi olmayan 28 firmadan 15 firma (%53,5) gecikmiştir.



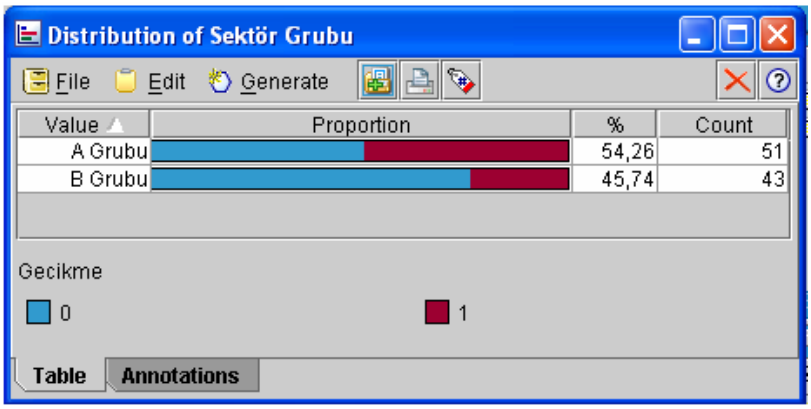
Şekil 6.6. Garanti Belgesi-Gecikme İlişkisi Grafiği

4. Sektör Grubu: Veri setinde, firmaların ait oldukları sektör sayısı fazla olduğundan, benzer sektörler birleştirilerek 6 grup oluşturulmuştur. Şekil 6.7.'de sektör grubu ile gecikme arasındaki ilişki görülmektedir. Sektör gruplarının başarı düzeyleri birbirine yakın çıkmış, ancak elektrik-elektronik ve makine sektöründe yer alan firmaların gecikme oranları diğer sektörlere göre biraz daha yüksek çıkmıştır.



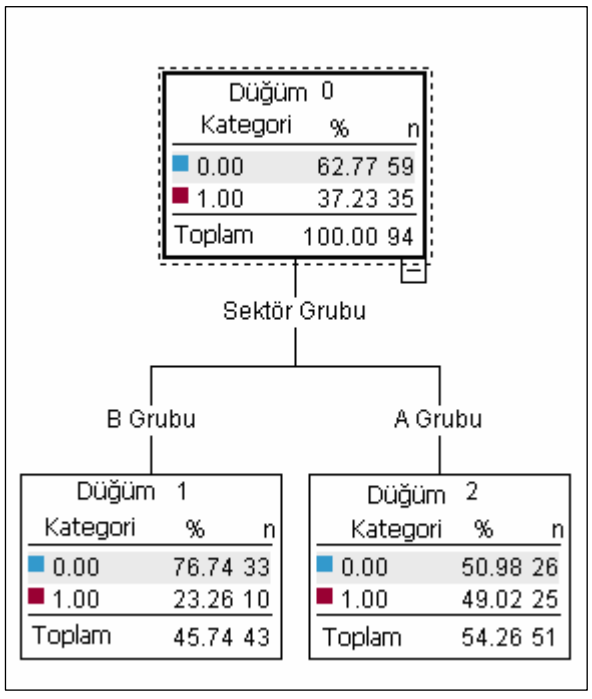
Şekil 6.7. Sektör Grubu -Gecikme İlişkisi

Sektör grubu ile gecikme arasındaki ilişkinin gerçek düzeyinin daha net ölçülebilmesi için, makine, elektrik-elektronik ve metal sektörleri üretimi temsil etmek üzere A Grubu, diğer 3 sektör grubu da B Grubu olmak üzere düzenlenmiş ve yeniden analize tabi tutulmuştur. Şekil 6.8'de görüldüğü gibi iki grup arasındaki fark daha belirginleşmiştir.



Şekil 6.8. Düzenlenmiş Sektör Grubu-Gecikme İlişkisi

Sektörler iki gruba ayrıldıktan sonra karar ağcı grafiği de Şekil 6.9'da gösterildiği gibi oluşmuştur.



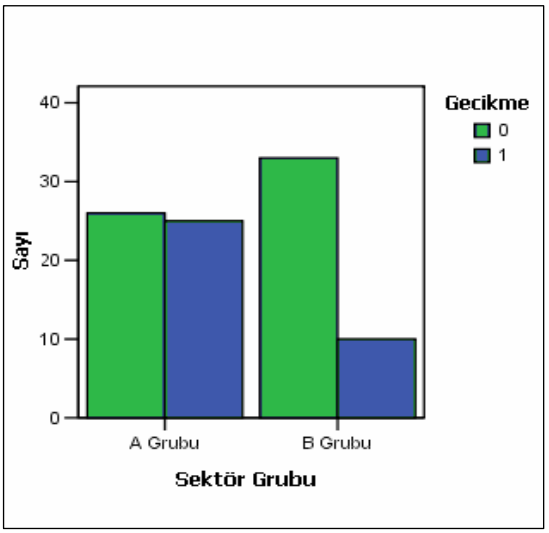
Şekil 6.9. Düzenlenmiş Sektör Grubu-Gecikme İlişkisi Karar Ağacı Grafiği

Veri setinde yer alan 94 firmadan, 59 firma (%62,77) zamanında teslimat yapmış (başarılı), 35 firma (%37,23) gecikmiştir (başarısız). Bu sonuçlar karar ağacının birinci seviyesinde görülmektedir. İkinci seviyede ise A ve B grubunun başarı/başarısızlık oranları görülmektedir. Buna göre, B grubunda yer alan firmaların %76,74'ü başarılı iken %23,26'sı başarısız, A grubunda yer alan firmaların ise %50.98'i başarılı, %49.02'si başarısız olmuştur. Her grupta yer alan firma sayısı ile başarı düzeyleri Ki-kare bağımsızlık testine tabi tutulduğunda Tablo 6.4'teki sonuçlar elde edilmiştir.

Tablo 6.4. Sektör Grubu-Gecikme İlişkisi

			Gecikme		Toplam
			0	1	
Sektör Grubu	A Grubu	Sayı	26	25	51
		Beklenen	32.0	19.0	51,0
	B Grubu	Sayı	33	10	43
		Beklenen	27.0	16.0	43,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		6,626		1	0,009

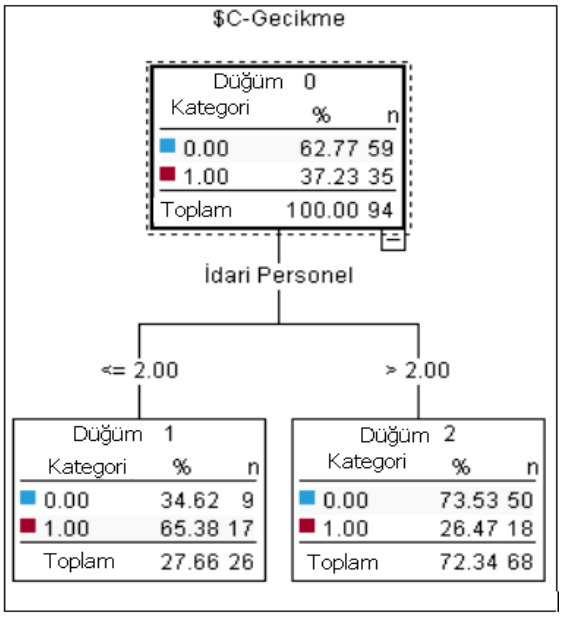
Sektör grubu ile gecikme arasındaki ilişki çok önemli düzeyde anlamlıdır. Çünkü B grubunda gerçekleşen başarı, beklenen başarıdan yüksek; A grubunda ise gerçekleşen başarı beklenenin altında çıkmıştır.



Şekil 6.10. Sektör Grubu- Gecikme İlişkisi Grafiği

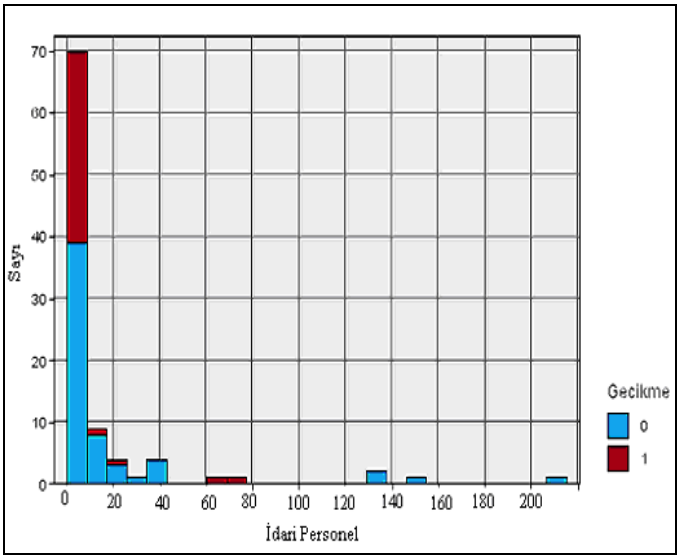
Sektör grubu ile ilgili elde edilen sonuçlar, analiz edilen veri seti için anlamlıdır. Verileri kullanılan işletmenin tedarikçileri ile, bağlı oldukları sektör arasında başarı/başarısızlık bakımından anlamlı bir ilişki bulunmuştur. Ancak, gerçek hayatta firmaları değerlendirirken, bağlı oldukları sektörü bir ön yargı aracı olarak kullanmak doğru bir yaklaşım olamaz.

5. İdari Personel Sayısı: Model, idari personel sayında belirleyici değeri 2 olarak tespit etmiş ve idari personel sayısını karar verme sürecinde anlamlı bulmuştur. Buna göre, idari personel sayısı iki ve daha az olan firmaların gecikme oranları yüksek çıkmıştır. İdari personel sayısı iki ve daha az olan 35 firmanın gecikme oranı %65.38 iken, idari personel sayısı ikiden fazla olan 59 firmanın başarı oranı %73.53'tür. Şekil 6.11'de idari personel ile gecikme arasındaki ilişkiyi gösteren karar ağacı diyagramı gösterilmiştir.



Şekil 6.11. İdari Personel Sayısı-Gecikme İlişkisi Karar Ağacı Diyagramı

İdari personel sayısı ile gecikme arasındaki ilişki Şekil 6.12'de histogram olarak verilmiştir. Bu tabloda, idari personel sayısı 60-80 arasında olan bir firma dışında, genel olarak idari personel sayısı artışına bağlı olarak gecikme oranının önemli düzeyde azaldığı görülmektedir.



Şekil 6.12. İdari Personel Sayısı-Gecikme İlişkisi

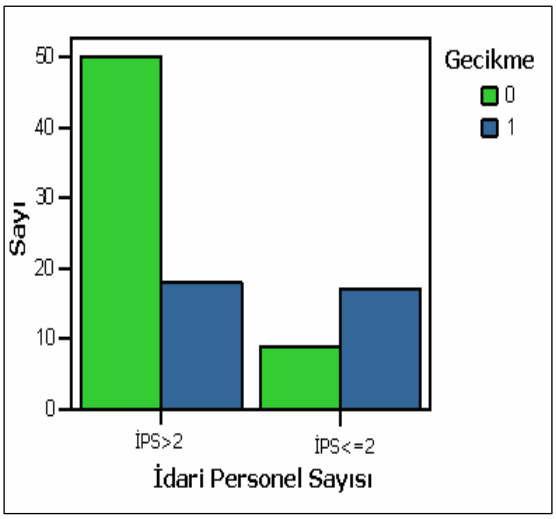
İdari personel sayısı ile gecikme arasındaki ilişki, ki-kare bağımsızlık testi ile test edildiğinde Tablo 6.5'teki sonuçlar elde edilmiştir.

Tablo 6.5. İdari Personel Sayısı-Gecikme İlişkisi

			Gecikme		Toplam
			0	1	
İdari Personel Sayısı	< 2	Sayı	50	18	68
		Beklenen	42,7	25,3	68,0
	<= 2	Sayı	9	17	26
		Beklenen	16,3	9,7	26,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		12,187		1	0,000

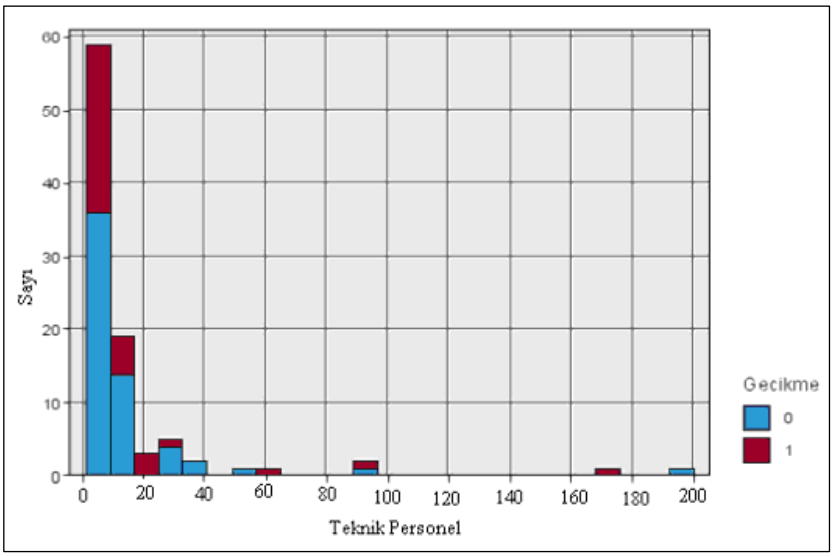
İdari personel sayısı ile gecikme arasındaki ilişki çok önemli düzeyde anlamlıdır. Çünkü idari personel sayısı 2'den fazla olan firmaların diğer gruba göre zamanında teslim oranı beklenenden yüksek, gecikme oranları beklenenden düşüktür. Buna paralel olarak idari personel sayısı 2'den az olan firmalarda ise, gecikme oranı yüksek, zamanında teslim oranı düşüktür.

İdari personel sayısı ile gecikme arasındaki ilişki grafiği Şekil 6.13'te verilmiştir.



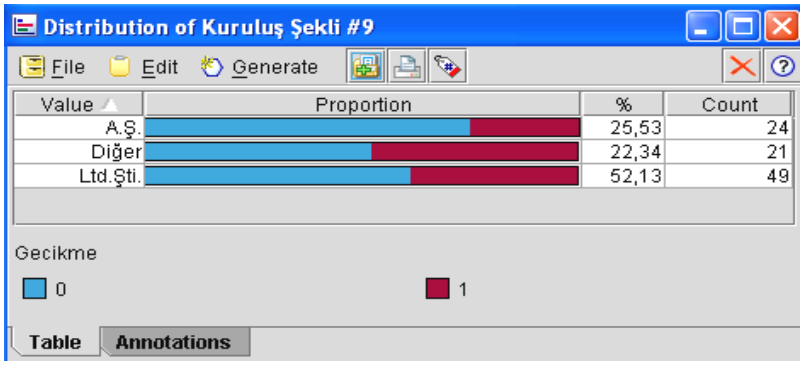
Şekil 6.13.İdari Personel Sayısı -Gecikme İlişkisi Grafiği

6. Teknik Personel Sayısı: Model, teknik personel sayısının az ya da çok olması ile gecikme arasında, karar sürecini etkileyecek anlamlı bir ilişki bulamamıştır. Buna gerekçe olarak, teknik personelin daha çok üretimle ilgili olması, firmaların ihale, satış ve sevkıyat işleri ile idari personelin ilgili olması gösterilebilir. Şekil 6.14'te teknik personel ile gecikme arasındaki ilişki gösterilmiştir.



Şekil 6.14. Teknik Personel Sayısı-Gecikme İlişkisi

7. Kuruluş Şekli: Firmaların kuruluş şekli ile gecikme arasındaki ilişki Şekil 6.15'te gösterilmiştir. Şekilde de görüldüğü gibi gecikme oranları birbirine yakın çıkmıştır.

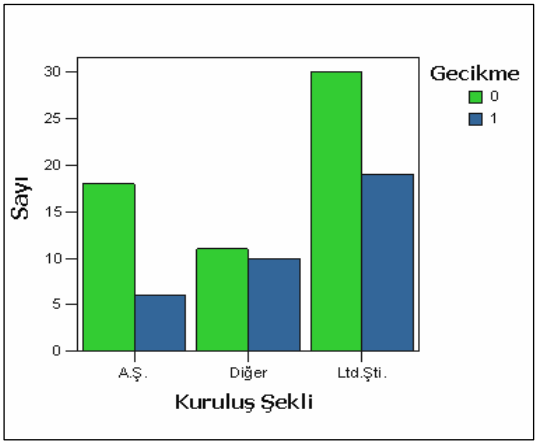


Şekil 6.15. Kuruluş Şekli – Gecikme İlişkisi

Kuruluş şekli ile gecikme arasındaki ilişkinin anlamlı olup olmadığı ki-kare testi ile ölçüldüğünde Tablo 6.6'daki sonuçlar elde edilmiştir. Şekil 6.16'da kuruluş şekli ile gecikme arasındaki ilişki grafiği verilmiştir.

Tablo 6.6. Kuruluş Şekli-Gecikme İlişkisi

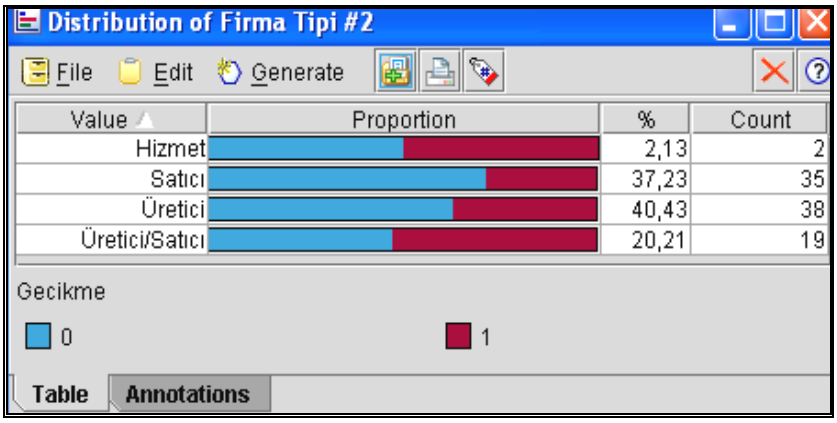
			Gecikme		Toplam
			0	1	
Kuruluş Şekli	A.Ş.	Sayı	18	6	24
		Beklenen	15,1	8,9	24,0
	Ltd.Şti.	Sayı	30	19	49
		Beklenen	30,8	19,2	49,0
	Diğer	Sayı	11	10	21
		Beklenen	13,2	7,8	21,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		2,556		2	0,279



Şekil 6.16. Kuruluş Şekli -Gecikme İlişkisi Grafiği

Kuruluş şekli ile gecikme arasındaki ilişki anlamlı değildir. Çünkü %5 anlam düzeyinde yapılan ki-kare testinin sonucu %27,9 olarak çıkmıştır.

8. Firma Tipi: Veri setinde tanımlanan 4 çeşit firma tipi ile gecikme arasındaki ilişki Şekil 6.17'de verilmiştir.



Şekil 6.17. Firma Tipi – Gecikme İlişkisi

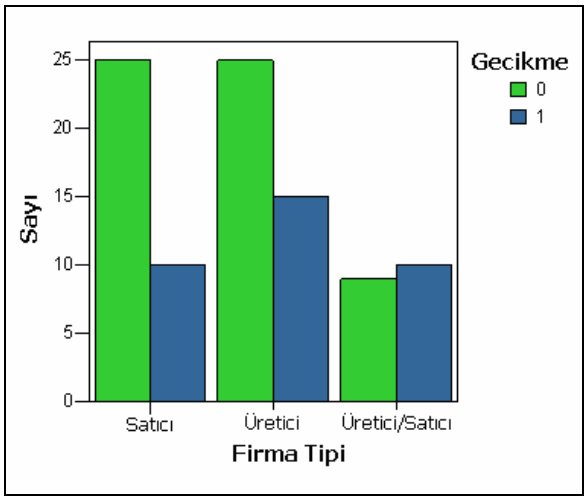
Firma tipine göre tedarikçilerin gecikme oranları birbirine yakın çıkmıştır. Ancak başarı düzeyleri diğerlerinden biraz daha yüksek olan "üretici" ve "satıcı" firmaların, toplam firmaların %78'i gibi yüksek bir orana sahip olmasından dolayı, bu değişkenin gecikme ile olan ilişkisinin anlamlı olup olmadığı ki-kare testi ile test edilmiştir. Elde edilen sonuçlar Tablo 6.7'de gösterilmiştir.

Tablo 6.7. Firma Tipi-Gecikme İlişkisi

			Gecikme		**Toplam**
			0	**1**	
Firma Tipi	Satıcı	Sayı	25	10	35
		Beklenen	22,0	13,0	35,0
	Üretici	Sayı	25	15	40
		Beklenen	25,1	14,9	40,0
	Üretici/Satıcı	Sayı	9	10	19
		Beklenen	11,9	7,1	19,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare Değeri		Değeri		S.Derecesi	Olasılığı
		3,053		2	0,217

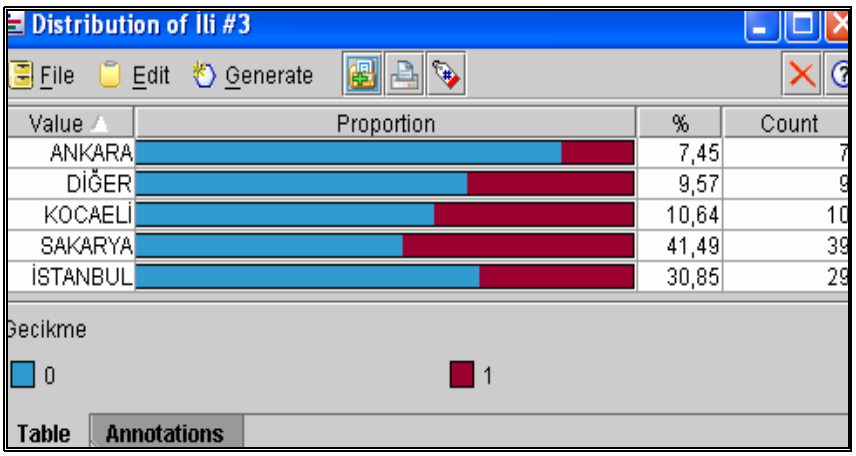
Tablodan elde edilen sonuçlara göre firma tipi ile gecikme arasında anlamlı bir ilişki bulunamamıştır. %5 anlam düzeyinde yapılan testin sonucu %21,7 çıkmıştır.

Firma tipi ile gecikme arasındaki ilişki grafiği Şekil 6.18'de gösterilmiştir.



Şekil 6.18. Firma Tipi-Gecikme İlişkisi Diyagramı

9. İli: Veri setinde, tedarikçilerin bağlı oldukları 9 il mevcuttur. Ancak firma sayıları dikkate alındığında bu firmaların %90'ı Sakarya, İstanbul, Kocaeli ve Ankara illerine bağlı olduğundan bu iller kendi isimleri ile belirtilmiş, kalan 5 il ise "Diğer" adlı grubun içerisinde yer almıştır. Şekil 6.19'da firmaların bağlı olduğu iller ile gecikme arasındaki ilişki verilmiştir.



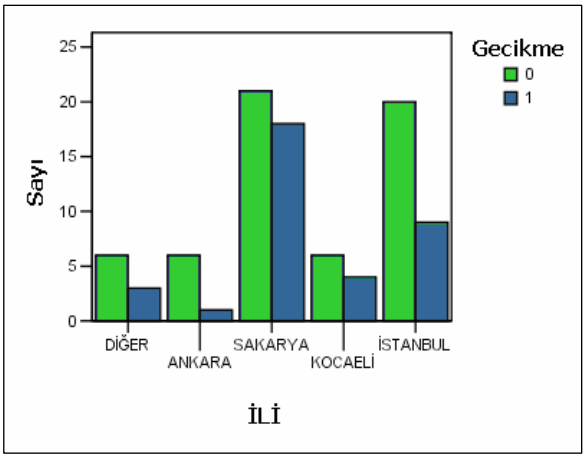
Şekil 6.19. Firmaların Bağlı Oldukları İl–Gecikme İlişkisi

İllere göre firmaların gecikme ve zamanında teslim oranları birbirine yakın çıkmıştır. Ancak tedarikçi verileri kullanılan işletmenin Sakarya'da bulunması, toplam firma sayısı içinde Sakarya iline bağlı firmaların oranının %41 olması ve gecikme oranı en yüksek ilin Sakarya olarak gözükmesi göz önüne alındığında, işletmenin bu durumu dikkate alması önem arz etmektedir.

Firmaların bağlı oldukları il ile gecikme arasındaki ilişkinin anlamlı olup olmadığı ki-kare testi ile de test edilmiş ve sonuçlar Tablo 6.8'de, il-gecikme ilişkisi Şekil 6.20'de gösterilmiştir.

Tablo 6.8. Firmaların Bağlı Olduğu İl-Gecikme İlişkisi

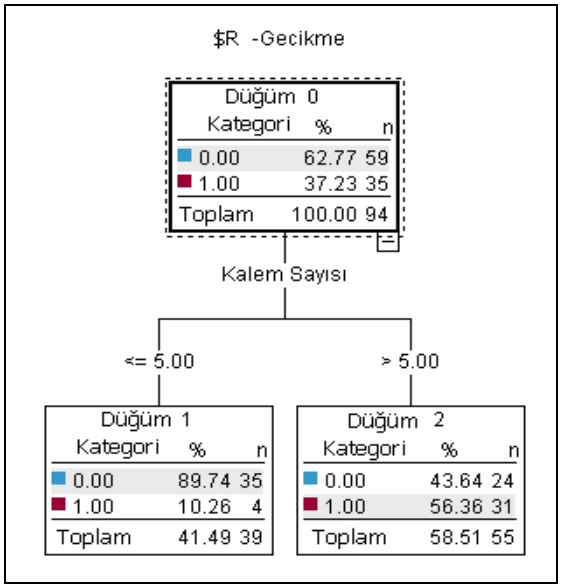
			Gecikme		Toplam
			0	1	
İli	Sakarya	Sayı	21	18	39
		Beklenen	24,5	14,5	39,0
	İstanbul	Sayı	20	9	29
		Beklenen	18,2	10,8	29,0
	Kocaeli	Sayı	6	4	10
		Beklenen	6,3	3,7	10,0
	Ankara	Sayı	6	1	7
		Beklenen	4,4	2,6	9,0
	Diğer	Sayı	6	3	9
		Beklenen	5,6	3,4	9,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		3,473		4	0,482



Şekil 6.20. Firmaların Bağlı Oldukları İl–Gecikme İlişkisi Diyagramı

Tablo 6.8'de elde edilen sonuçlara göre firmaların bağlı olduğu il ile gecikme arasında anlamlı bir ilişki bulunmamıştır. %5 anlam düzeyinde yapılan testin sonucu %48,7 çıkmıştır.

10. Kalem Sayısı: Kurulan model, bir ihalede satın alınan malzeme sayısının az ya da çok olması ile gecikme arasında anlamlı bir ilişki tespit etmiştir. Bu ilişki, C&R Tree karar ağacı algoritması kullanılarak elde edilmiştir. Şekil 6.21'de kalem sayısı ile gecikme arasındaki ilişki gösterilmiştir.

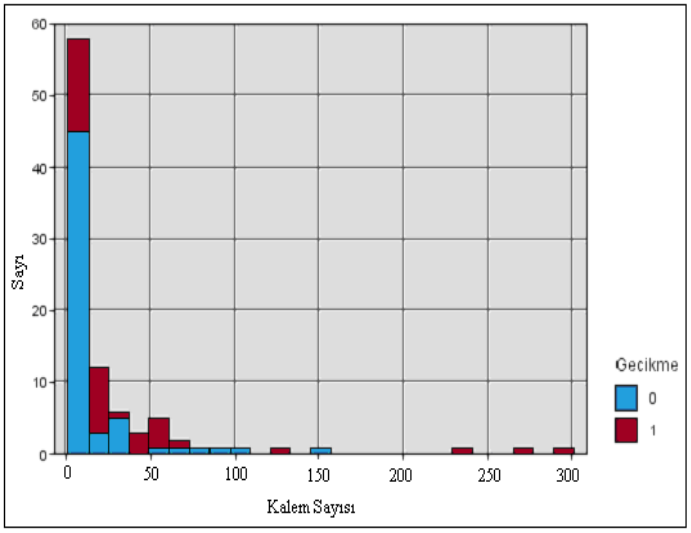


Şekil 6.21. Kalem Sayısı-Gecikme İlişkisi Karar Ağacı Diyagramı

Buna göre bir ihalede satın alınan malzeme kalemlerinin sayısı 5'ten az ise firmaların zamanında teslim oranı %89,74 gecikme oranı %10,26 iken; kalem sayısı 5'ten fazla ise zamanında teslim oranı %43,64 gecikme oranı %56,36 çıkmıştır.

Elde edilen oranlar, kalem sayısı değişkeninin karar verme sürecinde dikkate alınması gereken faktörlerden birisi olduğunu ortaya koymuştur.

Kalem sayısı ile gecikme arasındaki ilişki histogram olarak da Şekil 6.22'de verilmiştir.



Şekil 6.22. Kalem Sayısı – Gecikme İlişkisi

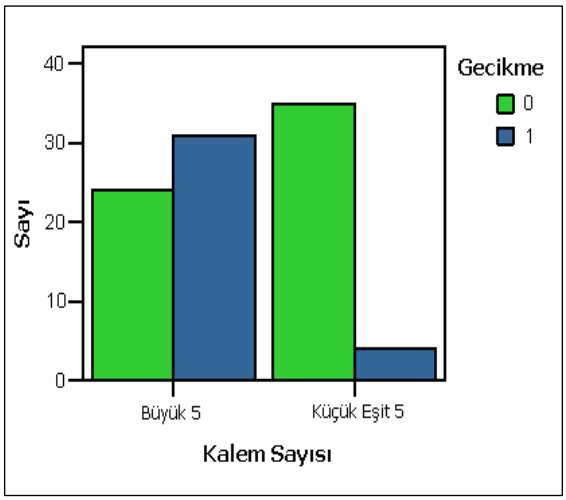
Kalem sayısı ile gecikme arasında anlamlı olduğu düşünülen ilişki ki- kare testi ile test edilerek, Tablo 6.9'de verilen sonuçlar elde edilmiştir.

Tablo 6.9. Kalem Sayısı-Gecikme İlişkisi

			Gecikme		Toplam
			0	1	
Kalem Sayısı	< 5	Sayı	24	31	55
		Beklenen	34,5	20,5	55,0
	=> 5	Sayı	35	4	39
		Beklenen	24,5	14,5	39,0
Toplam		Sayı	59	35	94
		Beklenen	59,0	35,0	94,0
Ki-Kare		Değeri		S.Derecesi	Olasılığı
		20,757		1	0,000

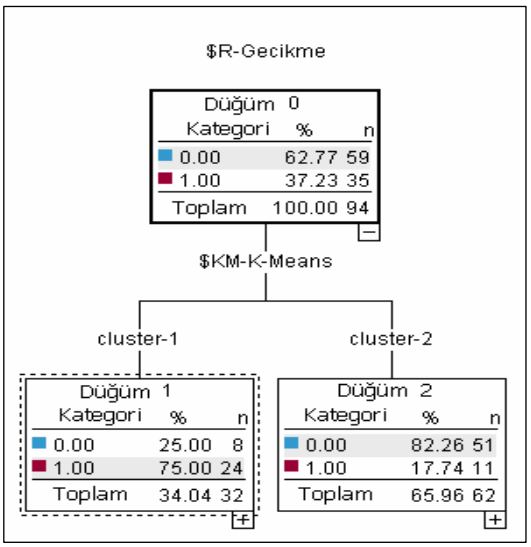
Tablo 6.9'da elde edilen sonuçlara göre kalem sayısı ile gecikme arasındaki ilişki çok önemli düzeyde anlamlıdır. Çünkü bir ihalede satın alınan malzeme kalemi sayısı 5'ten fazla ise gecikme oranları oldukça yüksek, 5'ten az ise başarı oranları beklenenin çok üzerinde yüksek çıkmıştır.

Kalem sayısı ile gecikme arasındaki ilişki Şekil 6.23'te gösterilmiştir.



Şekil 6.23. Kalem Sayısı-Gecikme İlişkisi Diyagramı

11. Firmaların başarı düzeyini ölçmek için yapılan analizlerden biri de kümeleme (cluster) analizidir. Bu çalışmada kümeleme algoritmalarından K-Means, Kohonen ve Two Step algoritmaları kullanılmış, ancak bu algoritmalardan K-Means daha sağlıklı sonuç vermiştir.



Şekil 6.24. Cluster-Gecikme İlişkisini Gösteren Karar Ağacı Diyagramı

Veri setinin küçük olması nedeniyle K-Means firmaları 2 kümeye ayırmıştır. Bu algoritma ile yapılan kümelemede, hiç gecikmesi olmayan 59 firma Cluster-2'de, en az bir gecikmesi olan 35 firma Cluster-1'de yer almıştır. Şekil 6.24'te görüldüğü gibi Cluster-2'de yer alan firmaların zamanında teslim oranları %82,26, gecikme oranları

%17,74 iken, cluster–1' de yer alan firmaların zamanında teslim oranları %25, gecikme oranları %75 çıkmıştır.

Model tarafından anlamlı bulunan bazı değişkenlerin hangi kümede yer aldıkları Tablo 6.10'da gösterilmiştir.

Tablo 6.10. Anlamlı Değişkenlerin Kümelere Dağılımı

Değişken Adı		Cluster–2 (59 Firma)	Cluster–1 (35 Firma)
AR-GE/Kalite Kontrol	Var	46	16
	Yok	13	20
Garanti Belgesi	Var	46	20
	Yok	13	15
Sektör Grubu	A	19	21
	B	40	14
Kalite Belgesi	ISO	39	11
	TSE	12	1
	Yok	8	23

Yukarıdaki tabloda da açıkça görüldüğü gibi, Ar-Ge/Kalite kontrol departmanına sahip olan, garanti belgesi olan, B sektör grubuna mensup, ISO ya da TSE kalite belgesi olan firmalar çok büyük bir oranda, başarılı küme olan Cluster–2' de yer almışlardır.

12**.** Kurulan modelde, yapay sinir ağı ve karar ağacı algoritmaları ile tahmin yapılmıştır. Her iki algoritmanın doğru ve yanlış tahminleri elde edilmiş, her iki tahmin kıyaslanmış ve etkinlikleri ölçülmüştür.

Tablo 6.11'de yapay sinir ağı algoritması ile elde edilen sonuçlar verilmiştir. Önceki aşamalarda olduğu gibi, zamanında teslim '0', gecikme '1' ile gösterilmiştir. Tabloda, 0 satırı ile 0 sütununun kesiştiği nokta zamanında teslim eden firmaların doğru tahmini, 1 satırı ile 1 sütununun kesiştiği nokta ise geciken firmaların doğru tahminini göstermektedir.

Tablo 6.11. Yapay Sinir Ağı Algoritmasının Tahmini

Gecikme				
Yapay Sinir Ağı		**0**	**1**	**Toplam**
0	Sayı	**42**	4	46
	Satır %	**91.304**	8.696	100
	Sütun %	71.186	11.429	48.936
	Toplam %	44.681	4.255	48.936
1	Sayı	17	**31**	48
	Satır %	35.417	**64.583**	100
	Sütun %	28.814	88.571	51.064
	Toplam %	18.085	32.979	51.064
Toplam	Sayı	59	35	94
	Satır %	62.766	37.234	100
	Sütun %	100	100	100
	Toplam %	62.766	32.234	100

Yapay inir ağı algoritması, veri setinde yer alan 94 firmadan 46 firmanın gecikmeyeceğini tahmin etmiş ve bu 46 firmadan 42 firma gerçekten gecikmemiştir. Yapay sinir ağı algoritmasının 'gecikmeme' tahminindeki başarısı %91,304 olarak gerçekleşmiştir. Aynı algoritma 48 firmanın gecikeceğini öngörmüş, bu firmaların 31 tanesi gerçekten gecikmiştir. Algoritmanın 'gecikme' tahminindeki başarısı %64,583 olarak gerçekleşmiştir.

Aynı veri seti için karar ağacı algoritması ile de tahmin yapılmıştır. Tablo 6.12'de karar ağacı algoritması ile yapılan tahminin sonuçları gösterilmiştir.

Tablo 6.12. Karar Ağacı Algoritmasının Tahmini

Gecikme				
Karar Ağacı		**0**	**1**	**Toplam**
0	Sayı	**51**	12	63
	Satır %	**80.952**	19.048	100
	Sütun %	86.441	34.286	67.021
	Toplam %	54.255	12.766	67.021
1	Sayı	8	**23**	31
	Satır %	25.806	**74.194**	100
	Sütun %	13.559	65.714	32.979
	Toplam %	8.511	24.468	32.979
Toplam	Sayı	59	35	94
	Satır %	62.766	37.234	100
	Sütun %	100	100	100
	Toplam %	62.766	37.234	100

Yapay sinir ağı ve karar ağacı algoritmalarının tahmin sonuçlarının karşılaştırılması Tablo 6.13'de gösterilmiştir.

Tablo 6.13. Tahminlerin Karşılaştırılması

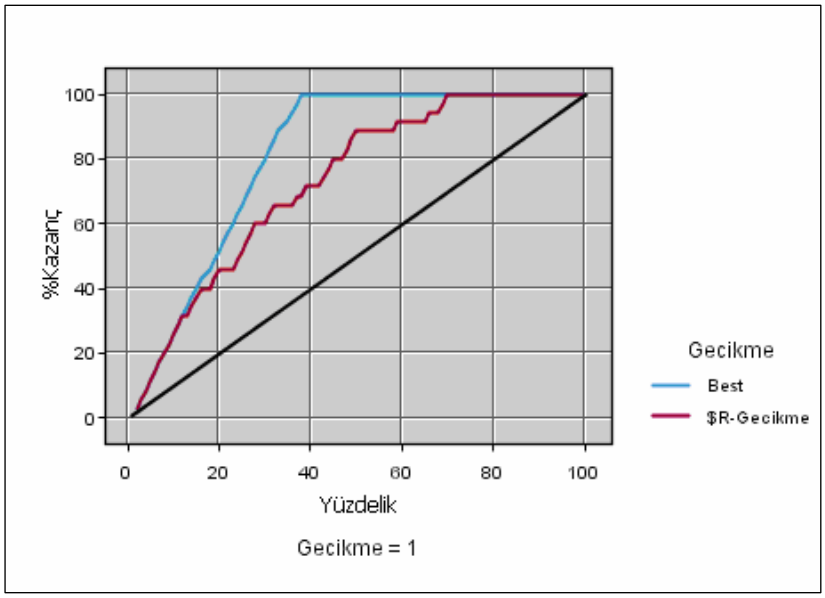
Gerçekleşen Sonuçlar ile Yapay Sinir Ağı Tahminin Karşılaştırılması		
Doğru	73	77.66%
Yanlış	21	22.34%
Toplam	94	100%
Gerçekleşen Sonuçlar ile Karar Ağacı Tahminin Karşılaştırılması		
Doğru	74	78.22%
Yanlış	20	21.285%
Toplam	94	100%
Karar Ağacı ile Yapay Sinir Ağının Tahminlerinin Uyuştuğu Noktalar		
Doğru	77	81.91%
Yanlış	17	18.09%
Toplam	94	100%
Gerçekleşen Sonuçlar İki Algoritmanın Uyuşmasının Kıyaslanması		
Doğru	65	84.42%
Yanlış	12	15.52%
Toplam	77	100%

Karar ağacı algoritması, veri setinde yer alan 94 firmadan 63 firmanın gecikmeyeceğini tahmin etmiş ve bu 63 firmadan 51 firma gerçekten gecikmemiştir. Karar ağacı algoritmasının 'gecikmeme' tahminindeki başarısı %80,952 olarak gerçekleşmiştir. Aynı algoritma 31 firmanın gecikeceğini öngörmüş, bu firmaların 23 tanesi gerçekten gecikmiştir. Algoritmanın 'gecikme' tahminindeki başarısı %74,194 olarak gerçekleşmiştir.

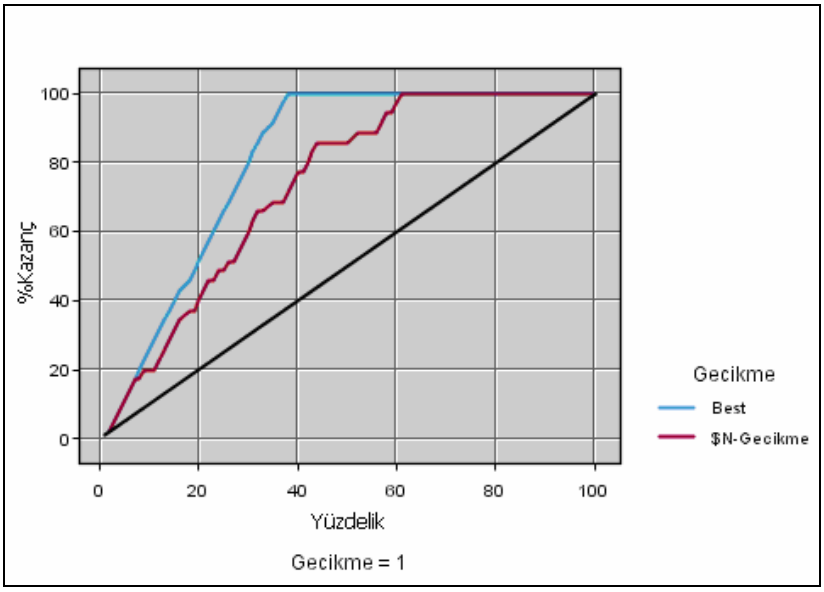
Sonuçlar karşılaştırıldığında, yapay sinir ağı algoritması 'gecikmeme' tahmininde daha başarılı, karar ağacı algoritması ise 'gecikme' tahmininde daha başarılı olmuştur.

Karar ağacı algoritmasının etkinlik grafiği Şekil 6.25'te, yapay sinir ağı algoritması

etkinlik grafiği de Şekil 6.26'da gösterilmiştir.

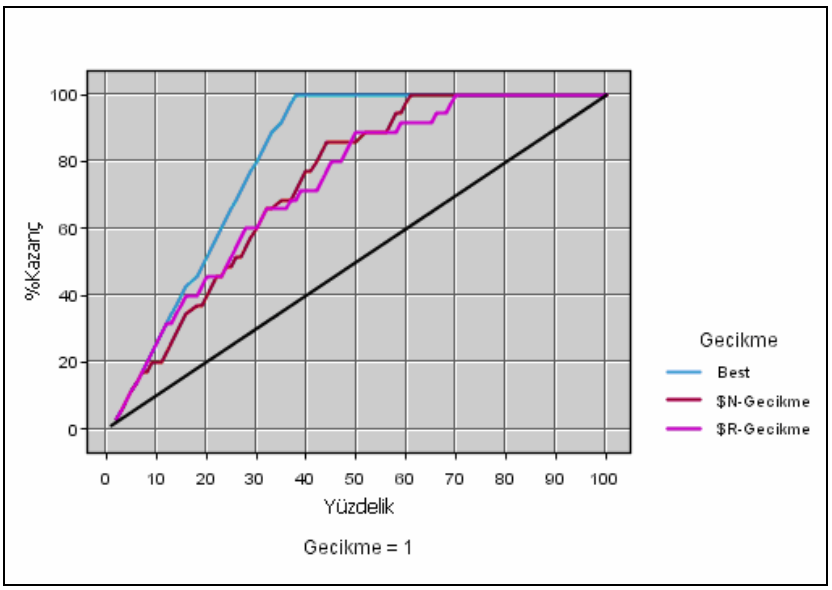


Şekil 6.25. Karar Ağacı Algoritmasının Etkinlik Grafiği



Şekil 6.26. Yapay Sinir Ağı Algoritmasının Etkinlik Grafiği

Yapay sinir ağı ve karar ağacı algoritmalarının tahminlerdeki etkinliklerinin karşılaştırılması Şekil 6.27'de grafik olarak gösterilmiştir.



Şekil 6.27. Yapay Sinir Ağı ve Karar Ağacı Algoritmalarının Etkinliklerinin Karşılaştırması

Bir tahminin başarısı, merkez çizgiden uzak ve ideal (best) çizgiye yakın olması ile ölçülmektedir. Bu ölçüye göre, her iki algoritmanın tahminleri başarılı sayılabilir. Etkinlik grafikleri incelendiğinde, yapay sinir ağı algoritmasının başlangıçta, karar ağacı algoritmasının ise süreç içerisinde daha etkin olduğu gözlenmektedir.

BÖLÜM 7. TARTIŞMA VE ÖNERİLER

Günümüz piyasa koşullarında yan sanayi, ana sanayinin vazgeçilmez unsurlarından biridir. Yan sanayi, ana sanayini gereksiz yatırımdan, atıl kapasite oluşturmaktan ve daha fazla personel ihtiyacından koruyacaktır. Elbette ana sanayi olmadan yan sanayi ve tedarikçiler de olmayacaktır. Bu yüzden bu ilişki sağlıklı bir zemine oturtulmalıdır.

Veri madenciliği, ana sanayi ile yan sanayi arasında tedarik zinciri oluşturulmasında yardımcı olarak kullanılabilir. Bu güne kadar kullanıldığı alanlardan farklı olarak, imalat ve montaj sanayinde etkin bir tedarik zinciri kurma ve tedarikçi seçiminde, stok kontrolü, imalat planlama ve kontrolü, personel yönetimi, kalite geliştirme, müşteri ilişkileri yönetimi, tedarikçi ilişkileri yönetimi gibi birçok alanda veri madenciliği tekniklerinden yararlanabilmek için aşağıdaki çalışmaların yapılması gerekli görülmüştür:

1. Veri toplama aşamasında yaşanılan sorunlar göz önüne alındığında, hem ana sanayinin, hem de yan sanayinin faaliyetlerini sağlıklı bir şekilde sürdürebilmesi için, her türlü bilginin derlenmesi ve saklanması, birimler ve kurumlar arasında iletişim kurulması ve fonksiyonel bir hafızanın oluşturulması gerekmektedir. Bu bağlamda; ihalelere iştirak eden firmaların demografik bilgileri, ihale bilgileri, ürün ve fiyat bilgileri, ürün tesliminde geciken firmaların gecikme süreleri ve nedenleri, iptal edilen ihalelerin gerekçeleri, reddedilen malzemelerin ret nedenleri gibi bilgiler, daha sonra yapılacak ihalelere ışık tutması amacıyla mutlaka kayıt altına alınmalıdır.

2. Analiz sonuçları göstermiştir ki, ISO kalite belgesine sahip firmalar ile kalite belgeli ürün satan firmalar daha başarılı olmuşlardır. Analizi yapılan 94 firmadan bu belgelere sahip 63 firma, ürün tesliminde %91 oranında başarılı olurken, kalite belgesine sahip olmayan 31 firmanın %75'i gecikmiştir. Benzer şekilde Ar-Ge ve/

veya kalite kontrol departmanına sahip olan firmaların %75'i başarılı olurken, bu departmana sahip olmayan firmaların %60'ı zamanında teslimat yapmayarak başarısız olmuşlardır. Bu nedenle tedarikçi firmalardan, uluslar arası geçerliliği olan kalite standartları mutlaka istenmeli ve Ar-Ge çalışmaları teşvik edilmelidir. İhalelerde, tedarik edilecek malzemeler için mutlaka kalite belgesi olma şartı konulmalıdır.

3. Garanti belgesi, tedarik edilecek ürünler için son derece önemlidir. Kusurlu malzemeler için garanti şartlarının uygulanması işletmelere maddi avantajlar sağlayacaktır.

4. Şüphesiz ki ana sanayilerde ürünü oluşturan tüm parçaların üretilmesi mümkün değildir. Mümkün olsa bile, bunun için yapılması gereken teknoloji ve kalifiye personel yatırımları oldukça yüksek olmaktadır. Bu nedenle bazı hammadde, yarı mamul ve mamullerin dışarıdan tedarik edilmesi daha ekonomiktir. Ancak tedarik edilmesinde sürekli sıkıntı yaşanan malzemeler için, işletmenin kendi bünyesinde üretim alternatifi her zaman hesaba katılmalıdır.

5. Stratejik öneme sahip olan malzemeler için mutlaka alternatif tedarikçiler belirlenmelidir.

6. Tedarik zincirinin başarılı olabilmesi için kurum elemanlarının eğitilmesi ve eğitimlerinin güncellenmesi gereklidir. Yan sanayiler için de eğitim faaliyetleri düzenlenmeli ve firmalar teşvik edilmelidir.

7. Mevcut tedarikçiler performanslarına göre mutlaka puanlanmalı ve bu puanlama sistemi, ihalelerde aktif belirleyici bir unsur olarak kullanılmalıdır.

8. ERP sistemi artık birçok orta ve büyük ölçekli işletmelerde kullanılmaya başlamıştır. Tedarik zinciri yönetimi veri madenciliği ile ERP sistemlerine entegre edilmelidir.

9. Bürokrasi, her alanda çalışmaları yavaşlatan bir engeldir. Ana sanayi ile tedarikçiler arasındaki ilişkilerin daha hızlı yürüyebilmesi için bürokrasi en az seviyeye indirilmelidir.

10. Günümüzde gerek işletmelerin tanıtımı, gerekse e-ticaret için bilgisayar ve internet teknolojisindeki gelişmeler son derece önem arz etmektedir. Bu nedenle işletmelerin, gerek ürünlerinin tanıtımı, gerekse ihtiyaç duydukları malzemelerin daha çok tedarikçi tarafından izlenebilmesi için web ortamını iyi değerlendirmeleri gerekmektedir.

11. Veri madenciliğinin işletmelere katkı sağlayabilmesinin en önemli şartı kaliteli veridir. Bu nedenle işletmeler, ürün, tedarikçi, kullandıkları teknoloji, personel, stok, ihale sistemi, finans yönetimi gibi birçok alan ile ilgili verileri, uygun veritabanı sistemi kullanarak, işlenmeye hazır durumda elektronik ortamda tutmaları gerekmektedir.

KAYNAKLAR

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

[2] ÖZMEN, Ş., "Ağ Ekonomisinde Yeni Ticaret Yolu E-TİCARET", İstanbul Bilgi Üniveristesi Yayınları, 1/2003, ISBN: 975-6857-44-7,İstanbul, 2001

[3] SPSS INC. – Data Mining and Introduction, Clementine™ – Working With Health Care, WPDMINTRO-0699, SPSS Inc., White Paper, USA, 1999.

[4] AHMED, I., " Data Warehousingin Construction Organizations", Construction Congress VI 2000 Proceeding, ASCE, 2000, pp: 194–203

[5] AKPINAR, H., "Kendini Düzenleyen Haritalar, Avrupa Birliği'ne Üye ve Aday Ülkelerin Karşılaştırılması, İ.Ü.İşletme Fakültesi, www.isletme.istanbul.edu.tr/ akpinar , 2001

[6] JENSEN, D., NEVILLE J., ve RATTIGAN M., "Randomization Tests for Relational Learning", University of Massachusetts, , 2003, pp: 03-05

[7] AGOSTA, L., "Data Mining is Dead-Long Live Predictive Analytics", ForresterResearch, http://www.forrester.com/research/legacyıt/0,7208,33030.html., oct. 30, 2003.

[8] U.S. General Accounting Office, "For more Information On The Uses of Data Mining in Gao audits, Data Mining: Results and Challenges for Government Programs, Audits, and İnvestigations", Washington, 2003, pp: 3-59

[9] KIM, K-J., CHO, S-B., "Fuzzy Integration of Structure Adaptive SOMs for Web Content Mining", Elsevier B.V., Fuzzy Sets and Systems, 2004, pp: 93-101

[10] WANGA, X., ABRAHAM B, A., SMITHA, K.A., "Intelligent Web Traffic Mining and Analysis", Elsevier Ltd., Journal of Network and Computer Applications, 2004, pp: 13-21

[11] BERSON, A., SMITH, S., THEARLING, K., "Building Data Mining Application for CRM", McGraw Hill, 1999, pp: 2–13

[12] FAYYAD, U., PIATETSKY-SAPHIRO, G., SMITH, P., UTHURUSAMY, R., "Advances in Knowledge Discovery and Data Mining", Cambridge, MA, London 1996

[13] HEINRICHS, J., LIM, J-S., "Integrating web-based data mining tools with business models for knowledge management", Elsevier Science B.V., Decision Support Systems, 2002, pp:103– 112

[14] GORLA, N., " Features to Consider in a Data Warehousing System", Communications of the ACM, 46(11), 2003, pp: 111–115

[15] PHILLIPS-WREN, G. E, HAHN, E. D., FORGIONNE, G. A., "A Multiplecriteria Framework for Evaluation of Decision Support Systems", OMEGA, 32(4), 2004, pp: 323–332.

[16] COCHRAN, J. K., CHEN, H., "Fuzzy Multi-Criteria Selection of Object-Oriented Simulation Software for Production System Analysis", Computers and Operations Research, 32(1), 2005, pp: 153–168

[17] NGAI, E. W. T., CHAN, E. W. C., "Evaluation of Knowledge Management Tools Using AHP", Expert Systems with Applications, Vol. 29, 2005, pp: 889–899

[18] MANNILA, H., "Methods and Problems in Data Mining", Working Paper, University of Helsinki, Helsinki 1997

[19] ZHAO, J., SCHEWE, K.-D., "Using Abstract State Machines for Distributed Data Warehouse Design", In: Hartmann, S., Roddick, J. (Eds.), Conceptual Modelling 2004—First Asia-Pacific Conference on Conceptual Modelling, Vol. 31 of CRPIT. Australian Computer Society, Dunedin, New Zealand, 2004, pp: 49–58.

[20] ABIDI, S. S. R., "Knowledge Management in Healthcare: Towards Knowledge-Driven Decision-Support Services", International Journal of Medical Informatics 63**,** 2001, pp: 5–18

[21] COREY, M., ABBEY, M., ABRAMSOM, I., "Oracle 8 Data Warehousing—A practical Guide to Successful Data Warehouse Analysis", ORACLE Press, 1998

[22] ALLSOPP, D. J., HARRISON, A., SHEPPARD, C., "A Database Architecture for Reusable CommonKADS Agent Specification Components", Knowledge-Based Systems 15, 2002, pp: 275–283

[23] ANAND, S. S., BELL, D. A., HUGHES, J. G., "EDM: A General Framework for Data Mining Based on Evidence Theory", Data and Knowledge Engineering 18**,** 1996, pp. 189–223

[24] WESTPHAL, C., BLAXTON, T., "Data Mining Solutions", Methods and Tools for Solving Real-World Problems, New York, 1998, pp: 21-30

[25] HEIJST, G., SPEK, R., KRUIZINGA, E., "Corporate Memories as a Tool for Knowledge Management", Expert Systems With Applications , 1997, pp: 41–54.

[26] HENDRIKS, P. H. J., VRIENS, D. J., "Knowledge-Based Systems and Knowledge Management: Friends or Foes?", Information and Management 35, 1999, pp: 113–125

[27] JOHANNESSEN, J. A., OLSEN, B., OLAISEN, T., "Aspects of Innovation Theory Based on Knowledge-Management", International Journal of Information Management 19, 1999, pp: 121–139

[28] KNIGHT, B., MA, J., "Temporal Management Using Relative Time in Knowledge-Based Process Control", Engineering Applications Artificial Intelligence 10, 1997, pp: 11-19

[29] WANG, X., ABRAHAM, A., SMITH, KA., "Web traffic mining using a oncurrent neuro-fuzzy approach" In Proceedings of the 2nd International Conference on Hybrid Intelligent Systems, Computing Systems: Design, Management and Applications, Santiago, Chile , 2002, pp: 853–862

[30] LAUDON, K.C., LAUDON, J. P., "Essential of Management Information Systems", Prentice Hall, New Jersey, 2002

[31] INMON, W. H., "Building the Data Warehouse", John Wiley and Sons, New York, 1993, pp:102-125

[32] BONIFATI, A., CASATI, F., DAYAL, U., SHAN M.-C., "Warehousing Workflow Data: Challenges and Opportunities", Proceedings of VLDB, Rome, Italy, 2001, pp: 649–652

[33] CHEUNG, D., KAO, B., LEE, J., "Discovering User Access Patterns on the World Wide Web", In Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD(97)), Vol. 10, 1997, pp: 463–70

[34] CHANG, G., HEALEY, M., J., MCHUGH, J.A.M., WANG, J.T.L., "Web Mining. In Mining the World Wide Web—An Information Search Approach", Dordetch: Kluwer; 2001,pp:43-47

[35] HAN, J., PEI, J., YIN, Y., "Mining Frequent Patterns Without Candidate Generation", In Proceedings of the 2000 ACMSIGMOD International Conference on Management of Data (SIGMOD(00), Dallas, TX, USA 2000, pp:1–12

[36] AGRAWAL, R., SRIKANT, R., "Mining Sequential Patterns", 11th International Conference on Data Engineering, Taipei, Taiwan, 1995, pp:73-79

[37] CHRYSANTHIS, P. K., RAMAMRITHAM, K., "ACTA: A Framework for Specifying and Reasoning About Transaction Structure and Behavior", Proceedings ACM SIGMOD, May 1990, pp: 194–203

[38] THOMSON, E., "OLAP Solutions: Building Multidimensional Information Systems", John Wiley & Sons, New York, 2002, pp: 71-79

[39] MENCZER, F., "Complementing Search Engines With Online Web Mining Agents", Elsevier Science B.V., Decision Support Systems 35, PII: S0167-9236(02)00106-9, 2002, pp: 195– 212

[40] FAYYAD, U., PIATETSKY-SHAPIRO, G., Smyth, P., "The KDD Process for Extracting Useful Knowledge From Volumes of Data", Communications of ACM, 39(11), 1996, pp: 27-34

[41] DILLY, R., "Data Mining: An Introduction", www-pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book1.html, 1995

[42] ARSLANTEKİN, S., "Ankara Üniversitesi Bilgi Hizmetlerinde Veri Madenciliği Çalışmaları", Veri Tabanlarının Performans Ölçümü Workshop'u, AB-04 Akademik Bilişim Programı, Karadeniz Teknik Üniversitesi, Trabzon, 2001

[43] SPSS, "Data Mining and Statistics Gain a Competitive Advantage", SPSS Inc.,White Paper, DATAMINP-1296M, USA, 1999, pp: 7-9

[44] INOUE, A., KILLIAN, L., "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?", Econometric Reviews 23, 2004, pp: 371–402

[45] AKPINAR, H., "Yapay Sinir Ağları ve İşletmecilik Uygulamaları", İ.Ü. İşletme Fakültesi Dergisi, 1994, s: 41-78

[46] HELBERG, C., "Data Mining with Confidence", SPSS, Integral Solution Ltd.,White Paper, USA., 2002, pp:19-26

[47] HUANG,S. J., LIN, J. M., "Enhancement of Power System Data Debugging Using GSA-Based Data-Mining Technique", IEEE Trans. Power Syst 17(4), 2002, pp: 1022–1029

[48] XU T., RENMU, H., PENG, W., XU, D., "Applications of Data Mining Technique for Power System Transient Stability Prediction, Electric Utility Deregulation, Restructuring and Power Technologies", 2004 (DRPT 2004, Hongkong). Proceedings of the 2004 IEEE International Conference, Vol. 5; 5–8 April, 2004. pp. 392–398

[49] WU H. C., LU C. N., "A Data Mining Approach for Spatial Modeling in Small area Load Forecast", IEEE Trans. Power Syst. 17(2), 2002, pp: 516–21

[50] Ogilvie, T., Swidenbang, E., Hogg, B.W., "Use of Data Mining Techniques in the Performance Monitoring and Optimisation of a Thermal Power Plant",

Proceedings of IEE Colloquium on Knowledge Discovery and Data Mining, May 1998, London, pp: 7/1–7/4.

[51] HAN, J., KAMBER, M., " Data Mining: Concepts and Techniques", Morgan Kaufmann, Los Altos, CA, 2001, pp:1-22

[52] THURAISINGHAM, B., "Web Data Mining and Applications in Business Intelligence and Counter-Terrorism", CRC PRESS, Boca Raton London New York Washington, D.C., 2003, pp: 45-47

[53] STEEL, J. A., MCDONALD, J. R., ARCY, C. D., " Knowledge Discovery in Databases: Applications in the Electrical Power Engineering Domain", Proceedings of IEE Colloquium on IT Strategies for Information Overload, December 1997, pp: 8/1–8/4

[54] FAYYAD, U., "Advances in Knowledge Discovery and Data Mining", Cambirdge MA: MIT. press, California, 1996

[55] TADESSE, T., WILHITE, D. A., HARMS, S. K., HAYES, M. J., GODDARD, S., "Drought Monitoring Using Data Mining Techniques", A Case Study from Nebraska USA, Nat. Hazards 33, 2004, pp: 137–159

[56] CHAN, A.N., WONG., J.R., "A query-driven approach to the design and management of flexible database systems," Journal of Management Information Systems 19(3), Winter 2002-2003, pp.121-154

[57] QUINLAN, J. R., "The Effect of Noise on Concept Learning. Machine Learning: An Artificial Intelligence Approach", San Mateo, CA, Morgan Kauffmann Inc.,1986, pp:149-166

[58] LEE, S. K., "An Extended Relational Database Model for Uncertain and Imprecise Information", 18th International Conference on Very Large Databases, VLDB'92, Vancouver, British Columbia, Canada, 1992, pp: 211-218

[59] LUBA, T., ve LASOCKI, R., "On Unknown Attribute Values in Functional Dependencies", T.Y. Lin ed., The International Workshop on Rough Sets and Soft Computing, San Jose, California, 1994, pp: 490-497

[60] TOLUN, M. R., SEVER, H., ULUDAĞ, M., "Improved Rule Discovery Performance on Uncertainty", Research and Development in Knowledge Discovery and Data Mining, Second Pacific-Asia Conference, PAKDD-98, Melbourne, Australia, Lecture Notes in Computer Science, Vol. 1394, Springer 1998, ISBN 3-540-64383-4, pp: 310-321

[61] CHOUBEY, S. K., DEOGUN, J. S., "A Comparison of Feature Selection Algorithms in the Context of Rough Classifiers", The 5th IEEE International Conference on Fuzzy Systems, New Orleans, 1996, pp: 1122-1128

[62] DEOGUN, J. S., RAGHAVAN, V. V., SEVER, H., "Exploiting Upper Approximations in the Rough Set Methodology", The First International Conference on Knowledge Discovery and Data Mining, Montreal, Quebec, Canada, 1995, pp: 69-74

[63] HULTEN, G., SPENCER, L., DOMINGOS, F., "Mining Time-Changing Data Streams", 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Fransisco, CA, ACM Press., 2001

[64] TOPAL, B., "A General Framework for Intelligent Data Mining", 3. International Symposium on Intelligent Manufacturing Systems, Sakarya, August 2001, pp: 30-31

[65] ZAHN, E., "Informationstechnologie und Informationsmanagement", München**,** 1997, pp: 300-357

[66] ARAYA, S., SILVA, M., WEBER, R., "A Methodology for Web Usage Mining and Its Application to Target Group Identification", Elsevier B.V., Fuzzy Sets and System, 2004, pp:103-107

[67] SIMOUDIS, E., "Reality Check for Data Mining", IEEE Expert: Intelligent Systems and Their Applications, 11(5), 1996, pp: 26-33

[68] RAGHAVAN, V. V., DEOGUN, J.S., SEVER, H., "Data Mining: Trends and Issues. Journal of American Society for Information Science and Technology, 49(5), 1998, pp: 397-402

[69] WEISS, S. M., KULIKOWSKI, C.A., "Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems", Morgan Kaufman, 1991, pp:21-32

[70] PAWLAK, Z., "Rough classification", International Journal of Man-Machine Studies, Vol. 20, 1984, pp: 469–483

[71] SARWAR, B., "Sparsity, Scalability and Distribution in Recommender Systems", PhD Thesis, University of Minnesota, 2001

[72] BERKAN, R. C., TRUBATCH, S L., "Fuzzy Logic and Hybrid Approaches to Web Intelligence Gathering and Information Management", In Proceedings of 2002 World Congress on Computational Intelligence, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE(02) Special Session on Computational Web Intelligence (CWI), Honolulu, Hawaii, USA, 2002, pp:1033–1038,

[73] CHEN, P. M., KUO, F. C., "An Information Retrieval System Based on an User Profile", J Syst Software; 54, 2000, pp: 3–8

[74] DE, S. K., KRISHNA, P. R., "Clustering Web Transactions Using Rough Approximation", Elsevier B.V., Fuzzy Sets and Systems, 2004, pp:201-210

[75] ZAIANE, O. R., "Building Virtual Web Views", J Data Knowledge Engineering; 39(2), 2001, pp: 143-163

[76] SPILIOPOULOU, M., FAULSTICH, L. C., "WUM: A Web Utilization Miner", In Proceedings of Workshop on the Web and Data Bases (WebDB(98), Valencia, Spain, 1998, pp:109–115

[77] COOLEY, R., TAN, P. N., SRIVASTAVA, J., "WebSIFT: The Web Site Information Filter System", In Proceedings of the Web Usage Analysis and User Profiling (WebKDD'99) Workshop on Web Mining, San Diego, CA, USA, 1999, pp:163–182

[78] PERKOWITZ, M., ETZIONI, O., "Adaptive Web Sites: Automatically Synthesizing Web Pages", In Proceedings of the 15th National Conference on Artificial Intelligence and 20th Innovative Applications of Artificial Intelligence Conference (AAAI (98), IAAI (98)), Madison, Wisconsin, USA, 1998, pp: 727–732

[79] CHEN, M.-C., CHIU, A.-L., CHANG, H.-H., "Mining Changes in Customer Behavior in Retail Marketing", Expert Systems with Applications, 28(4), 2005, pp: 773–781

[80] JOACHIMS, T., FREITAG, D., MITCHELL, T., "Web Watcher: A Tour Guide for the World Wide Web" In Proceedings of the 15th International Joint Conference on Artificial, 2002, pp:104-112

[81] MASSEGLIA, F., PONCELE, T. M., TEISSEIRE, M., "Using Data Mining Techniques on Web Access Logs to Dynamically Improve Hypertext Structure", ACM SigWeb Lett, 8(3),1999, pp: 1-19

[82] PAL, S. K., TALWAR, V., MITRA, P., "Web Mining in Soft Computing Framework: Relevance. State of the Art and Future Directions", IEEE Transaction on Neural Networks, 13(5), 1999, pp: 1163–1177

[83] ZHANG, Y. Q., LIN, T. Y., "Computational Web Intelligence (CWI): Synergy of Computational Intelligence and Web Technology", In Proceedings of 2002 World Congress on Computational Intelligence, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'02) Special Session on Computational Web Interlligence (CWI), Honolulu, Hawaii, USA, May 2002, pp: 12–17

[84] WUA, H., GORDON, M,. DEMAAGD, K., FAN, W., "Mining Web Navigations for Intelligence", Elsevier B. V., Decision Support System, 2004, pp: 36-44

[85] BAUTISTA, M. J., ANCHEZ, D. S. J., INEZ, C. M., SERRANO, J. M., VILA, M. A., "Mining Web Documents to _nd Additional Query Terms Using Fuzzy Association Rules", Elsevier Ltd., Fuzzy Sets and Systems, 2004, pp: 55-63

[86] PETERSON, T., PINKELMAN, J., "Microsoft OLAP Unleashed", SAMS, Indianapolis, 1999

[87] RICHARDSON, R., "Purchasing and Supply Chain Management, School of Engineering Technology& Management", Southern Polytechnic State University, 2002, pp:7-15

[88] CURTIS, C., "Supplier Development - SupplierRelationship Management", Institute of Supply Management, 2001, pp: 16-19

[89] ÖZ, E., BAYKOÇ, Ö. F. , "Tedarikçi Seçimi Problemine Karar Teorisi Destekli Uzman Sistem Yaklaşımı", Gazi Üniv. Müh. Mim. Fak. Dergisi, Cilt 19, No: 3, 2004, s: 275-286

[90] TANYAŞ, M., "Tedarik Zinciri Yönetimi ve Kalder Grup Kıyaslama Projesi", Lojistik Derneği , İstanbul, 2005

[91] TUVASAŞ 2005 Yılı İmalat Fabrikası İş Programı, Adapazarı, 2005

ÖZGEÇMİŞ

1971 yılında Malatya'nın Hekimhan ilçesine bağlı Dumlu köyünde doğan Aslan ÇOBAN, ilk, orta ve lise öğrenimini Hekimhan'da tamamladı. 1988 yılında girdiği İ.T.Ü. Sakarya Mühendislik Fakültesi Endüstri Mühendisliği Bölümünü 1992 yılında bitirdi. 1994 yılında Sakarya Üniversitesi Teknik Eğitim Fakültesi Makine Eğitimi Bölümü'ne araştırma görevlisi olarak girdi ve halen bu görevine devam etmektedir. 1994 yılında girdiği Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği A.B.D'da 1996 yılında yüksek mühendis olarak mezun oldu. Aynı enstitünün Makine Eğitimi Anabilim Dalında 1997 yılında doktora çalışmalarına başladı. Aslan ÇOBAN evli ve 2 çocuk babasıdır.