

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**SİGORTACILIK SEKTÖRÜNDE RİSK ANALİZİ: VERİ MADENCİLİĞİ
UYGULAMASI**

YÜKSEK LİSANS TEZİ

Duygu MUSLU

Anabilim Dalı : Endüstri Mühendisliği

Programı : Mühendislik Yönetimi

OCAK 2009

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**SİGORTACILIK SEKTÖRÜNDE RİSK ANALİZİ: VERİ MADENCİLİĞİ
UYGULAMASI**

**YÜKSEK LİSANS TEZİ
Duygu MUSLU
(507051207)**

Tezin Enstitüye Verildiği Tarih : 26 Aralık 2008

Tezin Savunulduğu Tarih : 19 Ocak 2009

**Tez Danışmanı : Doç. Dr. Mehmet Mutlu YENİSEY (İTÜ)
Diğer Jüri Üyeleri : Doç. Dr. Raziye SELİM (İTÜ)
Öğr. Gör. Dr. Bülent CERİT (İTÜ)**

ŞUBAT 2009

ÖNSÖZ

Yararlı ve önemli bilgilere ulaşmakta günümüzde yaygın olarak kullanılan başarılı yöntemler bütünü olan veri madenciliği, gelişen teknoloji ve artan veriler karşısında vazgeçilmez konumundadır. Bu tez çalışmasında sigortacılık sektöründe sahte hasar ihbarlarını tespit ederek risk faktörlerinin belirlenmesi hedeflenmiştir. Risk faktörlerinin belirlenmesinde veri madenciliği yöntemlerinden faydalanılmıştır.

Tez çalışmam süresince benden yardımlarını, desteğini, sabrını ve bilgisini esirgemeyen danışmanım Sayın Doç. Dr. Mehmet Mutlu Yenisey'e teşekkür etmeyi bir borç bilirim

Ayrıca çalışmalarım sırasında bana gerekli moral ve motivasyonu sağlayan kıymetli arkadaşım Pürnur Fırat'a ve desteklerini eksik etmeyen yöneticim Bahadır Kaan Özütam'a ve çalışma arkadaşlarıma teşekkür ederim.

Büyük fedakarlıklar göstererek bugünlere gelmemi sağlayan, sevgi ve desteğini hiçbir zaman eksik etmeyen annem, babam ve kardeşlerime tez çalışmalarım sırasında da sıkıntılarımı paylaşıp, anlayış göstererek bana yardımcı oldukları için minnettarım.

Aralık 2008

Duygu Muslu

Matematik Mühendisi

İÇİNDEKİLER

Sayfa

ÖZET	xi
SUMMARY	xiii
1. GİRİŞ	1
2. VERİ TABANLARINDA BİLGİ KEŞFİ	3
2.1 Problemin Tanımlanması	5
2.2 Verilerin Hazırlanması	6
2.2.1 Toplama	6
2.2.2 Değer biçme	6
2.2.3 Birleştirme ve temizleme	7
2.2.4 Seçme	8
2.2.5 Dönüştürme	8
2.2.5.1 Max-min normalizasyonu	8
2.2.5.2 Z-score standartlaştırma	10
2.3 Modelin Kurulması ve Değerlendirilmesi	11
2.3.1 Çapraz Doğrulama	11
2.3.2 ROC Eğrisi	12
2.4 Modelin Kullanılması	12
2.5 Modelin İzlenmesi	13
3. VERİ MADENCİLİĞİ	15
3.1 Veri Madenciliği Modelleri	19
3.1.1 Tahmin edici modeller	19
3.1.1.1 Sınıflandırma	20
3.1.1.2 Regresyon	21
3.1.2 Tanımlayıcı modeller	22
3.1.2.1 Kümeleme	22
3.1.2.2 Ardışık zamanlı örüntüler	23
3.1.2.3 Birliktelik kuralları	24
3.2 K En Yakın Komşu Yöntemi ile Sınıflandırma	25
3.3 Yapay Sinir Ağları Yöntemi ile Sınıflandırma	26
3.4 Karar Ağaçları Yöntemi ile Sınıflandırma	27
3.4.1 Karar ağaçlarında dallanma kriterleri	29
4. RİSK ANALİZİ	39
4.1 Risk Tanımı	39
4.2 Risk Analizi	40
4.2.1 Nitel risk analizi	42
4.2.2 Nicel risk analizi	45
4.2.2.1 Duyarlılık analizi	45
4.2.2.2 Simülasyon analizi	46
4.2.2.3 Karar ağacı analizi	47
5. SİGORTACILIK SEKTÖRÜNDE VERİ MADENCİLİĞİ UYGULAMASI 49	

5.1 Sigortacılık Sektörü Hakkında Bilgi.....	49
5.1.1 Sigortacılık sektörünün tarihi gelişimi	50
5.1.2 Sigortacılığın genel işleyişi	52
5.2 Sigortacılık Sektöründe Veri Madenciliği Uygulaması.....	54
5.2.1 Problemin tanımlanması	54
5.2.2 Verilerin hazırlanması	55
5.2.3 Modelin kurulması ve değerlendirilmesi.....	60
5.2.3.1 QT Orange yazılımı hakkında genel bilgi	60
5.2.3.2 QT Orange Canvas sınıflandırma ağacı uygulaması	64
5.2.4 Modelin kullanılması	71
6. SONUÇ VE ÖNERİLER	93
KAYNAKLAR	95
EKLER	99

KISALTMALAR

VTBK	: Veri Tabanından Bilgi Keşfi
CHAID	: Chi-Squared Automatic Interaction Detector
CART	: Classification and Regression Trees
ID3	: Induction of Decision Trees
TPR	: True Positive Rate
TNR	: True Negative Rate
ROC	: Receiver-Operating Curve
AUC	: Area Under ROC Curve

ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 2.1 : Eğitim verileri.	9
Çizelge 2.2 : Eğitim verileri.	10
Çizelge 3.1 : Müşteri eğitim verileri.	30
Çizelge 3.2 : Müşteri eğitim verileri.	31
Çizelge 3.3 : Birikim ve Kredi Riski nitelik değerleri.	31
Çizelge 3.4 : Mal Varlığı orta olan müşteri verileri.	33
Çizelge 3.5 : Gelir ve Kredi Riski nitelik değerleri.	34
Çizelge 4.1 : Olasılık ölçeği.	43
Çizelge 4.2 : Etki ölçeği.	43
Çizelge 4.3 : Risk Etki-Olasılık değerlendirme grafiği.	43
Çizelge 4.4 : Risk verileri.	44
Çizelge 4.5 : Örnek risk olasılık-etki değerlendirme grafiği.	44
Çizelge 5.1 : Uygulama verileri NULL değeri.	56
Çizelge 5.2 : Sabit değer atanmış uygulama verileri.	57
Çizelge 5.3 : Teminat bedeli 0 olan uygulama verileri.	58
Çizelge 5.4 : Ortalama değeri hesaplanan uygulama verileri.	58
Çizelge 5.5 : Nitelik değerleri.	60
Çizelge A.1 : Uygulama testi için ihbar verileri.	100

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Veri tabanında bilgi keşfi süreci [21].	5
Şekil 2.2 : İdeal ve kötü performans göstergesi olan ROC eğrileri [47].	12
Şekil 3.1 : Veri madencilğinin kronolojik gelişimi [52].	16
Şekil 3.2 : Veri madenciliği uygulama alanları [41].	19
Şekil 3.3 : Zaman serileri arasındaki bağımlılıkları gösteren örnek grafik.	23
Şekil 3.4 : Verilen bir ■ noktasına en yakın k=3 komşusunun belirlenmesi.	26
Şekil 3.5 : Biyolojik sinir hücresi.	27
Şekil 3.6 : Yapay sinir hücresi.	27
Şekil 3.7 : Mal Varlığı ile ilgili işlemler.	33
Şekil 3.8 : Sonuç karar ağacı.	35
Şekil 4.1 : Kıymet, tehdit ve açıklık fonksiyonu olarak risk.	41
Şekil 4.2 : Duyarlılık analizi akış diyagramı.	46
Şekil 4.3 : Karar Ağacı.	47
Şekil 5.1 : Orange için Python il yazılmış kod örneği.	62
Şekil 5.2 : Orange için Python il yazılmış kod örneği.	62
Şekil 5.3 : QT Orange Canvas bileşen iletişim örneği.	63
Şekil 5.4 : Bileşen iletişim kanalı örneği.	63
Şekil 5.5 : Uygulama bileşen seçimi.	64
Şekil 5.6 : Uygulama nitelik seçimi.	65
Şekil 5.7 : Ağaç kriterlerinin seçimi.	66
Şekil 5.8 : Ağacın hiyerarşik metinsel çıktısı.	67
Şekil 5.9 : Ağacın grafik gösterimi.	68
Şekil 5.10 : Ağacın önizlemesi.	69
Şekil 5.11 : Çapraz doğrulama testi için bileşen seçimi.	70
Şekil 5.12 : Çapraz doğrulama testi sonucu.	71
Şekil A.1 : Uygulama karar ağacının grafik gösterimi	101

SİGORTACILIK SEKTÖRÜNDE RİSK ANALİZİ: VERİ MADENİLİĞİ UYGULAMASI

ÖZET

Türkiye’de sigortacılık sektöründe Avrupa birliği uyum süreci ile hızla gelişmeye başlamıştır. Bunun yanında her sektörde olduğu gibi sigortacılık sektöründe de teknolojik gelişmeler takip edilmeye başlanmıştır. Poliçe bilgileri, müşteri bilgileri, poliçelerin hasar ihbarları gibi bilgiler büyük veri tabanlarında saklanmaya başlanmıştır. Yıllar geçtikçe sigorta sektörüne ait tonlarca veri toplanmaktadır.

Büyük veri tabanlarından istenilen bilgilere sorgulama yöntemleri ile kolayca ulaşılabilir. Fakat bu büyük veri tabanları belirli raporlar için gerekli bilgilerin yanı sıra şirket için yararlı birçok gizli bilgiyi de barındırmaktadır.

Veri madenciliği büyük miktarlardaki verinin içinden anlamlı ve yararlı, ilişki ve kuralların bilgisayar programları aracılığıyla aranması ve analizidir. Birinci bölümde veri tabanlarından bilgi keşfedilmesi süreçleri, bilgi keşfi sürecinin en önemli adımı olan veri madenciliği ve veri madenciliğine ait yaklaşımlar incelenecektir.

İkinci bölümde risk analizi yapabilmemiz için riskin açıklaması, risk analizi ve yöntemlerine değinilecektir.

Bu çalışmada sigortacılık sektörünün önemli adımlarından biri olan hasar ihbarlarının olumsuz sonuçlanması kuralları belirlenerek ve yeni ihbarların sonucu tahmin edilmeye çalışılacaktır. Olumsuz sonuçlanmasına sebep olacak risklerin belirlenmesinde veri madenciliği yöntemlerinden biri olan karar ağaçlarından faydalanılacaktır. Bu amaçla çeşitli veri madenciliği yazılımlarından biri olan Orange yazılımı kullanılacaktır.

Son bölümde uygulama geliştirilecek ve uygulama sonucunda ortaya çıkan karar ağacının kuralları değerlendirilecektir. Çıkan kurallardan risk maddesi olarak anlamlı olacağı belirlenecektir. Bu çalışma sonuçları ile şirketin yeni gelen ihbarların olumlu ya da olumsuz sonuçlanacağını tahmin etmesinde yardımcı olacaktır.

RISK ANALYSIS ON INSURANCE INDUSTRY WITH DATA MINING APPLICATION

SUMMARY

Insurance industry is growing rapidly with Europe Community adaptation process in Turkey. Besides this, technological developments are followed by insurance industry, as other industries. All information about the industry like insurance policy, notice of claim, customer information is stored in huge databases by these technologies. Tons of data are accumulated in insurance industry over years.

By query methods, companies can easily reach required information from the huge databases. These databases do not only provide reports information but also hold lots of hidden information that are useful for the companies.

Data mining is an analyzing tool and a process of finding meaningful and useful correlations or patterns among dozens data form with computer software. The first section is to identify process of knowledge discovering in databases. Data mining is the most important step of knowledge discovering in databases. In this section, data mining and its approaches are studied.

Second section is to identify meaning of risk. In addition, it touches on risk analysis and methods of risk analysis.

Notice of claim is the major step of insurance sector. In this study, the rules that will help to predict results of notice of claims like positive or negative, will be determined. Set of negative result rules will help us to identify risk factors. Decision trees will be used to identify rules and risk factors. There is lots of commercial data mining software today. We will use Orange software to apply decision tree.

As the last section, decision tree will be developed by claims data. Rules of decision tree's leaves will be determined whether they are risk factors or not. Risk factors will be tested by test data to decide that they are meaningful or not. This study will help to insurance company to predict the notice of claim's results as positive or negative.

1. GİRİŞ

Sigorta kavramı belirli bir prim karşılığında, öngörölmüş bir rizikonun gerçekleşmesine bağılı zarar ya da hasarın karşılanması sağılayan sözleşme olarak karşımıza çıkmakta ve günümüz dünyasında, vazgeçilmez bir mali piyasa oyuncusu olarak ekonomi içerisindeki yerini almış bulunmaktadır.

Kentleşme olgusunun giderek yoğunluk kazanması, sanayi hareketleri sürecine paralel olarak üretime ve üretim araçlarına yatırılan sermayenin artması ile birlikte sigortacılık, dünya genelinde büyüme ivmesini giderek yükselten bir sektör olmuştur. Bunun yanında çoğı sektör gibi sigortacılık sektörü de gelişen bilgi teknolojilerine uyum sağlamıştır.

Bilgi sistemleri ve iletişim teknolojilerindeki gelişmeler sayesinde tıp ve sağılık alanındaki birçok veri sayısal ortamda saklanabilir ve kolaylıkla erişilebilir hale gelmiştir. Kayıt altına alınan veriler sayesinde organizasyonlar veri yığınlarını anlamlandırarak ve içlerindeki gizli bilgileri tespit ederek organizasyon yararına bilgiler elde edebilmektedirler. Bu bilgileri elde etmek klasik sorgulama yöntemleriyle veya basit gözlemlerle elde edilemeyeceğı için anlamlı ve yararlı örüntüleri tespit edebilmek için geliştirilmiş olan veri madenciliğinden faydalanmaktadırlar.

Veri madenciliğı büyük verilerden yararlı örüntüleri çıkarma bir sürecidir. Birçok veri madenciliğı yöntemi bulunmaktadır. Ulaşılmak istenen bilgi ve veri şekline en uygun veri madenciliğı yöntemi seçilerek uygulanmaktadır.

Organizasyon için risk oluşturabilecek veri yığınları içinde gizlenmiş şekilde keşfedilmeyi bekliyen durumlar olabilir. Bu çalışmada sigortacılık verilerine veri madenciliğı tekniklerinden biri olan karar ağağını uygulayarak, risk teşkil edecek durumlar önceden tespit edilmeye çalışılacaktır.

Birinci bölümde veri tabanlarında bilgi keşfi, ikinci bölümde ise veri madenciliğı ile ilgili genel bilgiler verilecek ve veri madenciliğı yöntemleri anlatılacaktır. Bu yanında bu bölümde tezin uygulama aşamasında kullanılacak olan karar ağacı

yöntemi detaylandırılacaktır. Oluşturulacak karar ağacından yola çıkılarak risk yorumlaması yapılacağı için üçüncü bölümde risk analizine değinilecektir. Dördüncü bölümde ise incelenen sektör olan sigortacılık sektörü hakkında bilgi verilecektir. Uygulamada kullanılacak prgoram hakkında genel bilgi verildikten sonra son olarak uygulama geliştirilip, yorumlanacaktır.

2. VERİ TABANLARINDA BİLGİ KEŞFİ

Bilgi toplumlarının temel hammaddesi bilgidir. Bilgisayar ve iletişim teknolojileri geliştikçe bilginin üretilmesi, taşınması ve depolanması kolaylaşmıştır. Miniwatts'a (2006) göre elektronik ortamların gün geçtikçe yaygınlaşması ve kullanımının artmasıyla birlikte bilgi miktarında da hızlı bir artış gözlenmektedir [40]. Günümüz dünyasında gelişen bilgi teknolojilerinin bir sonucu olarak bilgisayar sistemleri her geçen gün hem daha ucuzlamakta, hem de güçleri artmaktadır [12]. İşlemciler gittikçe hızlanmakta ve disklerin 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 mümkün olabilmektedir.

Bilgisayarların ucuzlaması ile bilgisayar ve otomatik veri toplama araçları yaygın bir şekilde kullanılmaya başlamıştır. Bu yaygın kullanım sonucunda insanların yaptığı işlemler kayıt altına alınmakta ve veriler doğrudan sayısal olarak toplanmakta ve saklanmaktadır [11]. Bunun sonucu olarak da büyük ölçekli işletimsel veriler oluşmaktadır.

Bu verileri saklamak ve işletilmek üzere, çok boyutlu verilerin yönetilmesi için uygun bir araç olan veri tabanlarında tutulur. Fakat veri tabanı teknolojileri büyük hacimli verilerdeki anlamlı örüntüler çıkarabilmek için yeterli gelmemektedir. Porter (1998) veri saklama kapasiteleri yaklaşık olarak her dokuz ayda bir iki katına çıktığını belirtmiştir. Bunun yanında Braynt ve O'Hallaran (2003) veri hesaplama gücünün Moore kanununa göre her on sekiz ayda bir ikiye katlandığını belirtmiştir. Buna göre, verinin saklama oranı ile veriyi işleme oranı çok daha büyük olduğu için birçok veri ulaşılamamak ve işlenememek üzere saklanmaktadır.

Veri tek başına değersizdir. Bilgi ise bir amaca hizmet etmek doğrultusunda işlenen verilerdir. Diğer bir deyişle bilgi sorulan sorulara cevap vermek için veriden çıkardığımız anlamdır [27].

Çok çeşitli alanlarda organizasyonlarda biriken veri yığınları içinden anlamlı, değerli, önceden bilinmeyen bilgiyi keşfetmek için geleneksel sorgulama (query)

veya raporlama araçları yetersiz hale gelmesi, verilerin toplanması ve etkileyici şekilde birikmesi karşısında, hızla büyüyen dijital verileri kullanışlı bilgilere dönüştürmede insanlara yardımcı olacak, yeni nesil hesaplama teorilerine ve araçlara duyulan acil ihtiyaç, Veri Tabanından Bilgi Keşfi (VTBK) adı altında yeni arayışlara neden olmaktadır. Aktif araştırma alanlarından biri olan veri tabanlarında bilgi keşfi (VTBK) disiplini, çok büyük oylumlu verileri tam veya yarı otomatik bir biçimde analiz eden yeni kuşak araç ve tekniklerin üretilmesi ile ilgilenen son yılların gözde araştırma konularından biridir [34]. VTBK uygulamaları, faaliyet alanına yönelik karar destek mekanizmaları için gerekli ön bilgileri temin etmek için veri madenciliğini kullanılır [13].

VTBK süreci verilerden kullanışlı tüm bilgi keşfi süreçlerini kapsamaktadır. VTBK süreçlerinde ek olarak veri hazırlama, veri seçimi, veri temizleme, daha önce elde olan uygun bilgilerle birleştirme ve madencilik sonuçlarının uygun yorumu, verilerden faydalı bilgi çıkarımını sağlamak için gereklidir [35]. Veri madenciliği en önemli kesimini oluşturur.

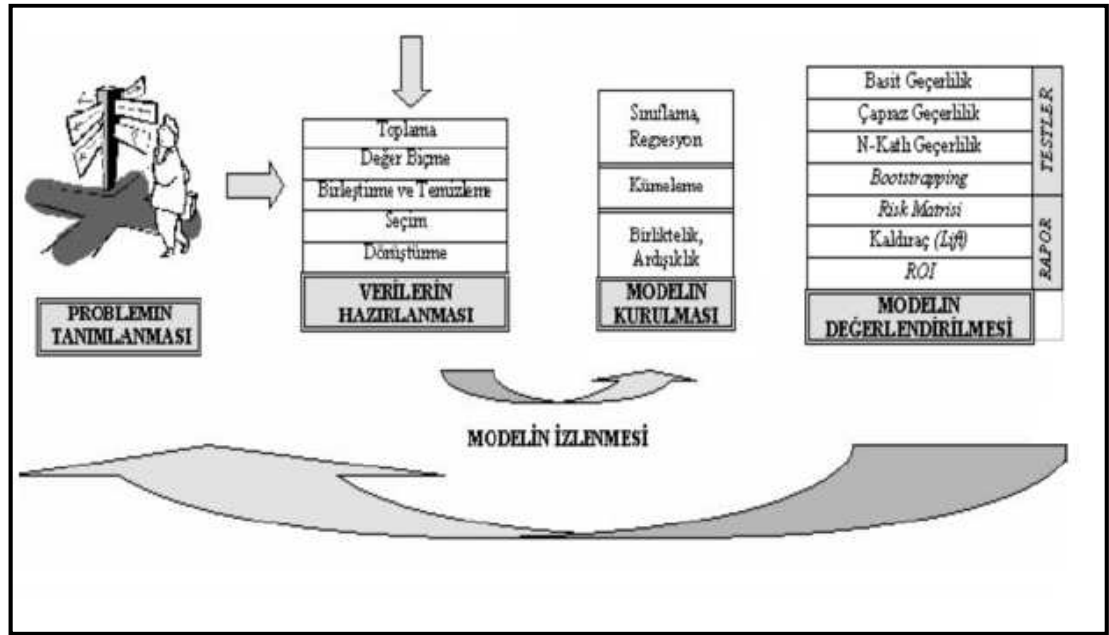
Süper market kasa sistemini örnek verecek olursak, günümüz sistemleri ile gün içindeki bütün hareketlerin detayları saklayabilmektedirler. Tutulan bu verilerden ürün barkodları ile ürünlerin dönemsel hareketlerini, eğer müşteri kodlama sistemleri mevcut ise müşteri ürün tercihlerini tespit edebilmektedirler. Acaba A ürününü alan müşteri yanında başka bir B veya C gibi ürün alıyor mu? B ürününü alan kişinin aynı gün C ürününü alma olasılığı nedir? Bütün bu soruların cevaplarına uygun veri madenciliği yöntem ve teknikleri ile ulaşabilirler. Bu bilgiler dahilinde promosyonlar veya kampanyalarla kendilerine satış stratejisi belirleyebilir ya da tedarik aşamasında hangi ürünlerden ne kadar alınması gerektiğini belirleyebilirler.

Örneğin en basit fatura kayıtları incelendiğinde, tıraş bıçağı alan müşterilerin %60 nın kalem pil de aldığı ortaya çıkmıştır. Buna dayanarak firma, tıraş bıçağı ve kalem pil reyonlarını bir araya getirmek suretiyle kalem pil satışlarını %14 arttırmıştır. Ürünler ve satışları arasındaki bu ilişkilerin belirlenmesiyle, satış stratejileri değiştirilip kazancın artırılması mümkündür[13].

Uygulanacak veri madenciliği algoritması ne kadar etkin olursa olsun, inceleme yapılacak işin ve verilerin özelliklerinin bilinmemesi durumunda sağlıklı ve faydalı bilgi elde edilmesi mümkün değildir. Bu nedenle yapılacak inceleme süreçlerinin

tüm aşamalarının öğrenilmesi veya anlaşılması için aşağıdaki adımlar takip edilmelidir.

1. Problemin tanımlanması
2. Verilerin hazırlanması
3. Modelin kurulması ve değerlendirilmesi
4. Modelin kullanılması
5. Modelin izlenmesi



Şekil 2.1 : Veri tabanında bilgi keşfi süreci [21].

2.1 Problemin Tanımlanması

Veri madenciliği çalışmalarının birinci adımı problemin tanımlanmasıdır. Veri madencisi çalışmanın kesin amacı üzerinde çalışmalı ve çalışma için gerekenler üzerine objektif olarak yoğunlaşmalıdır [7]. Diğer bir deyişle çalışmanın başarılı olmasının ilk şartı uygulamanın hangi işletme amacı için yapılacağını açık bir şekilde tanımlanmalıdır. İşletme amacının işletme problemi üzerine yoğunlaşacak şekilde açık bir dille ifade edilmiş olmalı ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği belirlenmelidir. Sorun ile örtüşmeyen bir veri madenciliği çalışmaları sorunu çözmeyeceği gibi kayıp maliyetlerine de sebep olur. Bunun yanında uygulanacak veri madenciliği algoritması sonucunda elde edilen

tahminlerin yanlış veya doğru olması durumunda karşılaşılabilecek kayıp veya kazançların belirlenmesine ilişkin tahminlere de bu aşama yer verilmelidir.

2.2 Verilerin Hazırlanması

Modelin kurulması aşamasında karşılaşılan sorunlar sık sık geri dönüp verilerin yeniden düzenlenmesine sebep olacaktır. Bu nedenle verilerin hazırlanması, veri madenciliği işlemi için gereken enerji ve zamanının büyük kısmını harcanmasına neden olmaktadır.

Verilerin hazırlanması verilerin toplanması, değer biçilmesi, birleştirmesi ve temizlenmesi, seçilmesi ve dönüştürme adımlarından oluşmaktadır.

2.2.1 Toplama

İşletme problemi için kullanılacak veriler ve bu verilerin hangi veri tabanlarından temin edileceğinin belirlendiği aşamadır. Genellikle büyük veritabanlarından rastgele örnekler alınarak analizde kullanılır. Farklı veritabanlarından veriler de içerebilir. Bu veritabanları içsel(müşterilerin geçmiş satın almaları) veya dışsal (kredi oranları) olabilir [1]. Seçilen ve kullanılan veri kaynakları sürecin sonucunu ve güvenilirliğini etkileyeceği için bu kaynakların güvenilirliği dikkat edilmesi gereken bir husustur. Veri madenciliği uygulaması genellikle büyük veritabanlarının sadece binlerce veya yüz binlerce kaydı üzerine uygulanır [1].

2.2.2 Değer biçme

Veri madenciliğinde kullanılmak üzere bir araya getirilen farklı kaynaklı veriler arasında uyumsuzluklar meydana gelmektedir. Farklı veritabanlarından ya da veri kaynaklarından elde edilen verilerin birlikte değerlendirmeye alınabilmesi için farklı türdeki verilerin tek türe dönüştürülmesi söz konusu olacaktır [2]. Bu uyumsuzlukların başlıcaları farklı zamanlara ait olmaları, güncelleme hataları, format farklılıklarının olması veya kodlamanın farklı olması (Örneğin cinsiyet ayrımı bir kaynakta 0 /1 olarak kodlanırken bir diğerinde E/K olarak kodlanması), farklı ölçü birimlerinin kullanılması ve varsayım farklılıklarıdır.

2.2.3 Birleřtirme ve temizleme

Deęer bięme iřlemi sırasında uyumlulařtırılan verilerin tek bir veri tabanında toplanması ařamasıdır. Bu ařamada unutulmaması gereken en önemli nokta, deęer bięme ve birleřtirme iřlemleri dikkatli yapılmazsa ilerleyen ařamalarda büyük sorunların kaynaęı olacaęıdır [22].

Veri tabanında yer alan tutarsız ve hatalı veriler gürültü olarak deęerlendirilmektedir. Bu gibi durumlarda verinin söz konusu sorunlardan temizlenmesi gerekecektir [2]. Ayrıca yanlış ve ařırı uęta bulunan verilerin ortadan kaldırılması da verilerin temizlenmesi konusuna girer. Bu kayıp verilerin yaratacaęı sorunları ortadan kaldırmak için kullanılan teknikler ařaęıdaki gibi özetlenebilir [3]:

1. *Eksik deęer ięeren kayıtlar veri kümesinden atılabilir:* Eęer kayıp verilerin kayıt sayısı, toplam kayıt sayısına oranlandığında sonuçların hassasiyetini etkilemeyecek kadar küçükse bu yöntem kullanılabilir.
2. *Kayıp deęerlerin yerine genel bir sabit kullanılabilir:* Örneęin eęitim seviyesi eksik olan bilgiler yerine “X” gibi bir sabit atanabilir. Bu ęalıřma sonucunda eęitim durumunun X olması anlamlı bir sonuçmuř gibi ıkabilir yada eęitim durumunun X olması belirleyici ve ayırt edici bir özellikmiř gibi ıkabilir. Yani bu bilgiler veri madencilięini yanıltabilir. Bununla beraber bu yöntemin kullanılması, bazı durumlarda veri madencilięinin gerçek amacına hizmet ederek bilinmeyen ortaya ıkabilir.
3. *Deęiřkenin tüm verilerinin ortalaması veya sadece bir sınıfa alınarak eksik deęerin yerine bu deęer kullanılabilir:* Örneęin elimizde sigorta hasar ihbar verileri olsun. Teminat tutarı eksik olan kısma teminat tutarlarının ortalaması alınıp atanabilir.
4. *Verilere uygun bir tahmin yapılarak eksik deęerin yerine konabilir:* Eldeki eksik olmayan veriler kullanılarak bir regresyon denklemi ve regresyon katsayıları elde edilerek kayıp veriler tahmin edilebilir. Regresyon dıřında zaman serileri analizi, Bayesyen sınıflandırma, karar aęaęları, maksimum beklenti gibi veri madencilięinde kullanılan dięer yöntem ve teknikler de kayıp verilerin tahmin edilmesinde kullanılabilir [2,3].

Bunların yanında eksik olan veri kimlik numarası veya vergi numarası gibi sabit bilgileri içeriyorsa, yukarıda bahsedilen yöntemler kullanılamaz. Bu durumlarda eksik verilerin manüel ya da bir program aracılığı ile gerçek verilere atanabilir.

2.2.4 Seçme

Modelin belirlenmesinden sonra modele kullanılması uygun olan verilerin seçilmesi aşamasıdır. Örneğin bir tahmin modeli için modelde kullanılacak bağımlı ve bağımsız değişkenlerin seçilmesi işlemidir. Seçme işlemi yaparken sıra numarası, id gibi alanların modele dahil edilmez. Seçilecek olan veri kümesinin bütünü temsil edecek şekilde seçilmesi büyük önem taşımaktadır [22].

2.2.5 Dönüştürme

Verilere aldıkları değerlere göre dönüştürme işlemi yapılır. Örneğin yapılan harcamalar verisi, $\text{harcama} > 1000$ ve $\text{harcama} \leq 1000$ olarak gruplanarak model içerisinde kullanılabilir veya uç değerler varsa bunların modele aynen katılması uygun olmayabilir. Bu durumda veriler üzerinde standartlaştırma veya normalizasyon işlemleri uygulamak uygun olacaktır [1,2].

Ayrıca modelde kullanılan algoritma, verilerin gösteriminde önemli rol oynayacaktır. Örneğin bir uygulamada bir yapay sinir ağı algoritmasının kullanılması durumunda kategorik değişken değerlerinin evet/hayır olması; bir karar ağacı algoritmasının kullanılması durumunda ise örneğin gelir değişken değerlerinin yüksek/orta/düşük olarak gruplanmış olması modelin etkinliğini artıracaktır [52].

2.2.5.1 Max-min normalizasyonu

Normalizasyon mevcut verilerin 0 ile 1 aralığı gibi aralık değerlerine dönüştürülmesi işlemine denir. Normalizasyon yöntemlerinden biri olan max-min normalizasyon işleminde minimum en düşük değer alan veriyi, maksimum ise en yüksek değer alan veriyi ifade eder. Dönüştürme işleminde yeni değerler aşağıdaki formül kullanılarak hesaplanır.

$$s' = \frac{s - s_{\min}}{s_{\max} - s_{\min}} \quad (2.1)$$

Çizelge 2.1 : Eğitim verileri.

Veri s	Normalizasyon değeri s'	Aralık belirtilmiş norm. değeri s'
36	0.95	2.9
23	0.36	1.72
15	0	1
37	1	3

Formüldeki s değişkeni dönüştürülecek veriyi, s' ise verinin normalize edilmiş değerini ifade eder. Bir örnekle açıklayacak olursak, çizelgede verilen eğitim verilerinin normalizasyon değerleri aşağıdaki gibi hesaplanır:

$$s'_1 = \frac{36-15}{37-15} = 0.95$$

$$s'_2 = \frac{23-15}{37-15} = 0.36$$

$$s'_3 = \frac{15-15}{37-15} = 0$$

$$s'_4 = \frac{37-15}{37-15} = 1$$

Dönüştürme işlemleri sonrasında normalizasyon değerlerinin 0-1 aralığından farklı bir aralıkta olması istenirse, istenilen aralığın üst değeri $a_{\max}$ ve alt değeri $a_{\min}$ olacak şekilde aşağıdaki formül kullanılır.

$$s' = \frac{s - s_{\min}}{s_{\max} - s_{\min}} (a_{\max} - a_{\min}) + a_{\min} \quad (2.2)$$

$$s'_1 = \frac{36-15}{37-15} (3-1) + 1 = 2.90$$

2.2.5.2 Z-score standartlaştırma

İstatistik çözümlerde sıkça kullanılan bir diğer dönüşüm biçimi Z-score adıyla anılmaktadır. Bu yöntem, verilerin ortalaması ve standart hatası göz önüne alınarak yeni değerlere dönüştürülmesi esasına dayanmaktadır [2]. Bu yöntemin bir diğer adı da sıfır-ortalama normalizasyonudur. Dönüştürme işleminde yeni değerler aşağıdaki formül kullanılarak hesaplanır.

$$s' = \frac{s - \bar{s}}{\sigma_s} \quad (2.3)$$

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^n (s_i - \bar{s})^2}{n-1}} \quad (2.4)$$

Max-min normalizasyonu örneğindeki veriler için z-score dönüşüm değeri aşağıdaki gibi hesaplanır:

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i = \frac{36 + 23 + 15 + 27}{4} = 25.25$$

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^n (s_i - \bar{s})^2}{n-1}} = 9.15$$

$$s'_1 = \frac{36 - 25.25}{9.15} = 1.175$$

Çizelge 2.2 : Eğitim verileri.

Veri s	Z-score değeri s'
36	1.175
23	-0.245
15	-1.120
37	1.284

2.3 Modelin Kurulması ve Değerlendirilmesi

Birçok veri madenciliği modeli vardır. Seçilen model hem geleneksel hem de veriye uyarlanabilir olmalı fakat hepsinden önemlisi verilerin ve yöntemin birbirlerine uygunluğudur [4]. Mevcut veri madenciliği modellerine ilerleyen konularda değinilecektir.

Modelin değerlendirilmesi için birçok yöntem bulunmaktadır. Uygulamada model değerlendirmesi aşamasında modelin doğruluk oranını test etmek için çapraz doğrulama yöntemi kullanılacağı için bu yöntem detaylandırılacaktır.

2.3.1 Çapraz Doğrulama

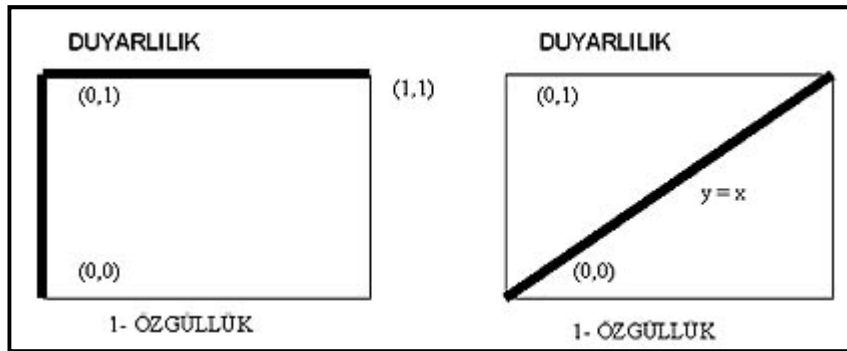
- *Basit Doğrulama (Simple Validation)*: Büyük veri setleri için kullanılır. Verilerin %5 ile %33 lük kısmı test kümesi olarak ayrılır, geri kalanıyla model kurulur. Daha sonra kurulan model, test kümesi ile test edilir ve modelin doğruluk oranı hesaplanır. Eğer farklı eğitim kümesi ve test verileri kullanılmazsa modelin geçerliliği tahmin üstü olur.
- *Çapraz Doğrulama (Cross Validation)*: Model kurmak için az sayıda veri varsa, basit doğrulamada ki kadar veri alınmaz. Çapraz doğrulama tüm veriyi kullanmayı sağlayan bir metottur. Veri kümesini rastgele 2 eşit kısma ayrılır. İlk veri kümesiyle model kurulur, kurulan model diğer veri kümeleriyle test edilir ve doğruluk oranı hesaplanır. Daha sonra ikinci veri kümesiyle model kurulur ve kurulan model birinci veri kümesiyle tahmin edilip, doğruluk oranı hesaplanır. En sonunda tüm veriler kullanılarak model oluşturulur. Daha önceden hesaplanmış olan 2 doğruluk oranının ortalamasıyla, son model karşılaştırılır.
- *N Kere Çapraz Doğrulama (N Fold Cross Validation)*: Veri kümesi rastgele N gruba ayrılır. 1.grup test için ayrılırken geriye gruplarla model kurulur. Kurulan model test için ayrılan veriler üzerinde tahminler ve doğruluk oranı hesaplanır. Süreç N defa tekrar eder ve modelin doğruluk oranı, N tane doğruluk oranının ortalaması kadar olur.

Çapraz doğrulamanın avantajı hata hesaplama teknikleri genel performansı göstermede yardımcı olmaktadır. N-kat çapraz onaylama ile elimizdeki veriyi daha iyi kullanmasını sağlamaktadır. [45].

2.3.2 ROC Eğrisi

ROC (Receiver Operating Curve) eğrisi, modellerin doğruluk değerlendirilmesi ve kıyaslanması için yaygın olarak kullanılan bir yöntemdir. ROC eğrisi, duyarlılık ve özgüllük oranlarını kullanarak birimleri sınıflarına ayıran en uygun kesim noktasını belirler. Sınıflamanın doğruluğu, ROC eğrisi altında kalan alanın büyüklüğüne bağlıdır.

ROC eğrisinin oluşturulacağı koordinat sisteminin ordinatında modelin gerçek pozitif değeri (duyarlılık), apsisinde ise yanlış pozitif değeri (1-özgüllük) yer alır. Model ne kadar iyi ise eğri o kadar yukarıya (yüksek duyarlılık bölgesi) ve sola (düşük yanlış pozitif oranı bölgesi) doğru kayar. Yanlış değerlere sahip olmayan ideal bir testte ROC eğrisi (0,0)-(0,1)-(1,1) noktalarını birleştirmektedir. Buna karşın ROC çizimi $y=x$ fonksiyonuna yaklaştıkça başarısız bir test ortaya çıkar. Çünkü bu testte yanlış değerlerin oranı yükselmektedir. Bu fonksiyonun altındaki ROC eğrisine sahip test başarısızdır.



Şekil 2.2 : İdeal ve kötü performans göstergesi olan ROC eğrileri [47].

Şekil 2.2’de görüldüğü gibi ROC eğrisi altında kalan alan 1 ile 0.5 aralığında değer almaktadır. Alan değeri ne kadar büyük ise modelin okadar iyi ayırım yeteneğine sahip olduğunu gösterir.

2.4 Modelin Kullanılması

Uygulanan modelin geçerliliği kabul edildikten sonra doğrudan bir uygulama olarak veya başka uygulamalara destek olacak alt uygulama olarak kullanılabilir. Örneğin kurulan model ile doğrudan risk analizinde kullanılabilir.

2.5 Modelin İzlenmesi

Zaman içerisinde sistem özelliklerinin değişmesi ve dolayısıyla ürettikleri verilerin değişmesi sebebiyle, kurulan modelin sürekli izlenmesi ve ihtiyaç duyulursa yeniden düzenlemesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.

3. VERİ MADENCİLİĞİ

Literatürde faydalı ve gizli örüntülerin (patterns) ortaya çıkarılmasında karşımıza çıkan terimlerden biride Veri Madenciliğidir. Veri madenciliği için yapılan tanımlamalardan birkaçı şöyledir: Fayyad'a (1996) göre veri mandeciliği önceden bilinmeyen, veri içinde gizli, anlamlı ve yararlı örüntülerin büyük ölçekli veri tabanlarından otomatik biçimde elde edilmesini sağlayan VTBK süreci içinde bir adımdır. Gartner Group'a göre ambarlarda depolanmış verileri, istatistiksel ve matematiksek teknikler yardımıyla inceleyerek anlamlı yeni korelasyonların, örüntülerin ve eğilimleri ortaya çıkarılması sürecidir [5]. Raghavan ve Sever'e (1994) göre ise, veri madenciliği büyük veri kümesi içinde saklı olan genel örüntülerin ve ilişkilerin bulunmasıdır. Deogun, Raghavan, Sarkar ve Sever'e (1997) göre veri madenciliği, makine öğrenimi, istatistik, veri tabanı yönetim sistemleri, veri ambarlama, koşut programlama gibi farklı disiplinlerde kullanılan yaklaşımları birleştirmektedir. Swift (2001), veri madenciliğini veri kaynağında tutulan çok çeşitli verilere dayanarak daha önce keşfedilmemiş verileri ortaya çıkarmak, bunları karar vermek ve gerçekleştirmek için kullanma süreci olarak tanımlamaktadır [24]. Hand (2001) ise büyük veri kümelerinden yararlı bilgilerin çıkarılması olarak ifade etmiştir [1].

Günümüzde bankacılık, sigortacılık ve borsa gibi birçok alandaki bütün bilgiler bilgisayar sistemleri sayesinde çeşitli veri tabanlarına kaydedilmektedir. Bütün bunlara bir de uydulardan elde edilen veriler katıldığından, ulaşılan veri hacminin inanılmaz boyutlara vardığı açık bir gerçektir. Veri tabanlarında klasik sorgulama programları veya yöntemleri yardımı ile kullanıcı sorgulamak istediği sorularına yanıt bulabilmektedir. Bu sorgulama araçları kullanıcının sormayı düşündüğü sorgularla sınırlıdır fakat veri tabanı kullanıcının aklına gelmeyecek veya gözlemleyemeyeceği bilgileri de barındırabilir. Veri madenciliğinin asıl amacı bu bilgilere ulaşmaktır.

Veri madenciliği, eldeki yapısız veriden, anlamlı ve kullanışlı bilgiyi çıkarmaya yönelik çalışmalarının bütünü olmuştur. Sonuç olarak yıllar ilerledikçe ortaya çıkan

veri yığınlarına bir düzen verme, başka bir anlamda potansiyel olarak kullanışlı bilgi haline getirme amacıyla veri madenciliği ortaya çıkmıştır.

Veri madenciliği, büyük miktardaki veriler arasından gelecek ile ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranması sağlayan bilgiye ulaşma işidir. Yani veri işlenerek amaçlanan bilgi elde edilir [10].

Veri madenciliğinin kronolojik gelişimi aşağıdaki çizelgedeki gibidir:

Gelişim Adımları	Cevaplanan Karar Problemi	Kullanılabilen Teknolojiler	Ürün Sağlayıcıları	Karakteristikler
Veri Toplama (1960'lar)	"Benim toplam karım geçen 5 yılda ne kadardı?"	Bilgisayarlar, Teypler, Diskler	IBM, CDC	Geriye döntük , statik veri dağıtımı
Veri Erişimi (1980'ler)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	İlişkisel Veritabanları, SQL, ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Kayıt düzeyinde geriye döntük, dinamik veri dağıtımı
Veri Ambarlama ve Karar Destek Sistemleri (1990'lar)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	OLAP, Çok Boyutlu Veritabanı Sistemleri, Veri ambarları	Pilot, Comshare, Arbor, Cognos, Microstrategy	Çoklu düzeylerde, geriye döntük dinamik veri dağıtımı
Veri Madenciliği (Bugün)	"Gelecek ay Boston'daki birim satışlar muhtemelen ne olabilir, niçin?"	İleri düzeyde algoritmalar, çok işlemcili bilgisayarlar, büyük veritabanları	Pilot, Lockheed, IBM, SGI, SPSS, SAS, Microsoft vs.	Geleceğe döntük ,proaktif enformasyon dağıtımı

Şekil 3.1 : Veri madenciliğinin kronolojik gelişimi [52].

Veri madenciliği, bilgisayar bilimi için makine öğrenimi ve istatistik tekniklerine dayanmaktadır. Makine öğrenimi ve istatistik gibi veri içerisindeki örüntüleri bulmayı amaçlar. Özel durumlarda veri madencisi bu teknik ve araçları, daha yararlı veri madenciliği yapmak amacıyla kusursuz bir şekilde eşleştirebilmelidir [6].

Veri madenciliği algoritmalarının çekirdeğini makine öğreniminde kullanılan algoritmalar oluşturur. Makine öğreniminde kullanılan sınıflama, kümeleme ve ilişkilendirme algoritmaları gibi birçok algoritma veri madenciliğinde kullanılmaktadır. Makine öğrenimi ile veri madenciliği arasında bu söylenen benzerliklerin bulunmasının yanı sıra aralarında çok büyük farklar da vardır. Örneğin, Makine öğrenimi küçük hacimli ve genelde deneysel verilerle uğraşırken,

Veri madenciliği büyük hacimli gerçek dünya verileriyle uğraşır. Makine öğreniminin örnekleme kümesi genelde 100-1000 arasındayken, Veri madenciliği uygulandığında milyonlarca veriden söz edilmektedir. Veri madenciliği ve makine öğrenimi arasındaki diğer bir fark da, makine öğreniminin aksine veri madenciliğinin gürültülü, eksik, artık ve bos (NULL) değerleri işleyebilmesidir [45].

Veri madenciliği, veri çıkarımlarında ki-kare testi gibi istatistiksel metotlardan yararlanır. Veri tabanları veri madenciliğine sorgulama yapmak için girdi oluşturmaktadır. Veri tabanından sorgularla veri madenciliği için örnekleme kümesi oluşturulur. Veri madenciliğinin veri tabanından farkı veri tabanında bilinen örüntüler için sorgulama yapılırken, veri madenciliği sorguları belirsiz örüntüleri keşfetmeye dayanır.

Veri madenciliği pazarlama, bankacılık, sigortacılık gibi sektörlerde yaygın olarak kullanılmaktadır. Karar verme süreçlerinde başarılı sonuçlar elde edildiği için taşımacılık-ulaşım-konaklama, eğitim öğretim, perakendecilik gibi konularda da kullanılabilir. Kullanım alanlarını genel olarak aşağıdaki gibi gruplayabiliriz:

İş ve Elektronik Ticaret Verileri

Her türlü ofis uygulamaları ve iş süreçleri sonucunda büyük çaplarda veriler üretilmektedir. Bunların karar verme süreçlerinde etkili olarak kullanılabilmesi ticari kuruluşların temel yapı taşlarından biri olmalıdır.

Bilimsel, Mühendislik Verileri

Günümüzde bilimsel veriler daha karmaşık hale gelmişlerdir. Bilim adamları ve mühendisler bu verileri kullanarak simülasyon ve sistem kullanımının arttırılması hedefindedirler.

Web Verileri

İnternet ve web üzerindeki veriler hem hacim hem de karmaşıklık olarak hızla artmaktadır. Web verilerinde düz metin ve resimden başka akan (streaming) ve nümerik veriler gibi farklı yapılarda veriler de yer alabilmektedir [25].

Pazarlama

- Müşterilerin satın alma örüntülerinin belirlenmesi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,

- Posta kampanyalarında cevap verme oranının artırılması,
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Pazar sepeti analizi,
- Müşteri ilişkileri yönetimi,
- Müşteri değerlendirmesi,
- Satış tahmini,

Bankacılık

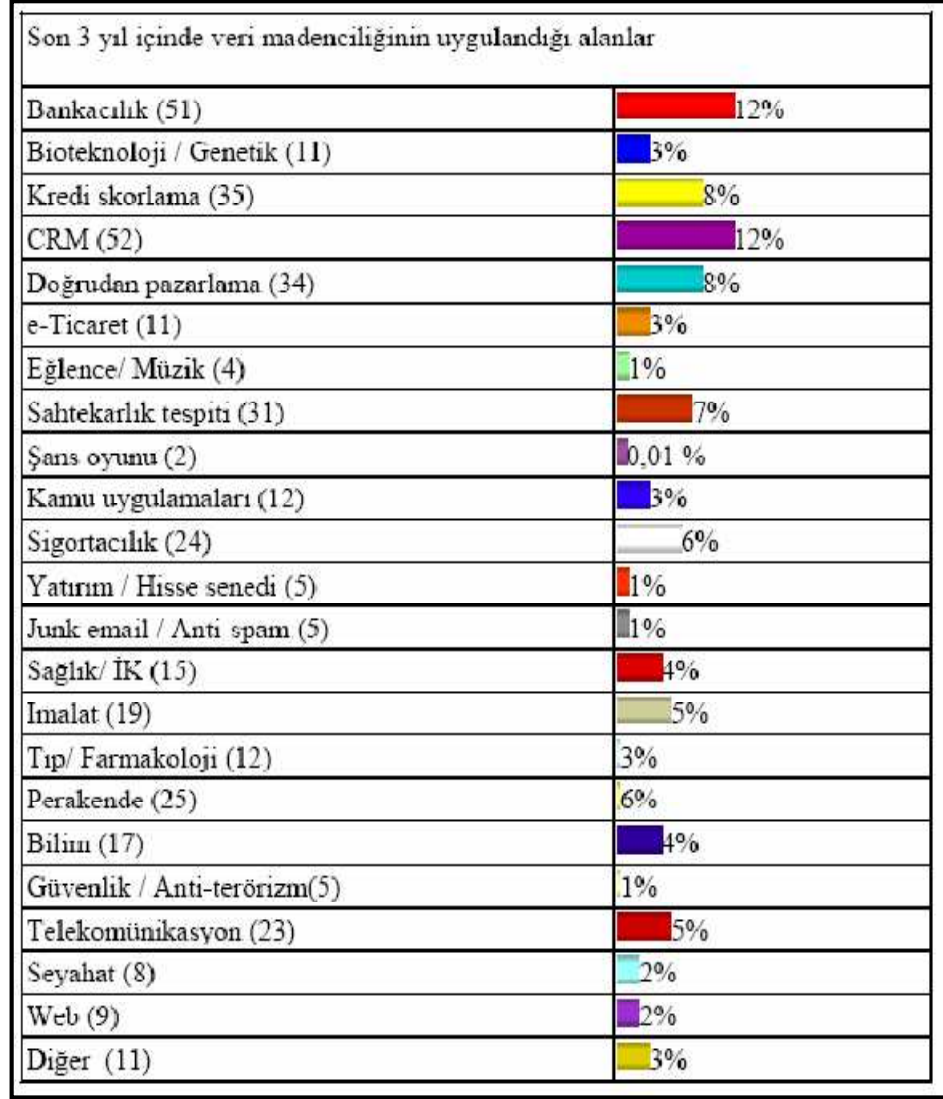
- Farklı finansal göstergeler arasında gizli korelasyonların bulunması,
- Kredi kartı dolandırıcılıklarının tespiti
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi
- Kredi taleplerinin değerlendirilmesi

Sigortacılık

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi
- Sigorta dolandırıcılıklarının tespiti
- Riskli müşteri örüntülerinin belirlenmesi [21]

Sağlık

- Belirli bir hastalığa sahip kişilerin ortak özelliklerinin tahmin edilmesi
- Tıbbi tedaviden sonra hastaların durumlarının tahmin edilmesi
- Hastane maliyetlerinin tahmin edilmesi
- Ölüm oranları ve salgın hastalıkların tahmin edilmesi [44].



Şekil 3.2 : Veri madenciliği uygulama alanları [41].

3.1 Veri Madenciliği Modelleri

Veri Madenciliği modelleri tahmin edici (prediction) ve tanımlayıcı (description) modeller olmak üzere ikiye ayrılır.

3.1.1 Tahmin edici modeller

Sonuçlanmış veri kümelerinden yola çıkarak model geliştirilmesinin ve bu modeli sonucu bilinmeyen veriler üzerinde uygulayarak bir tahmin sonucunun elde edilmesinin amaçlandığı modeldir. Örneğin bir banka önceki dönem kredi alan müşterilerinin verilerine sahiptir. Bu verilerde müşteri bilgileri bağımsız değişkenler olurken kredinin geri ödenip ödenmediği bağımlı değişkendir. Bu veriler incelenerek oluşturulan model yardımıyla gelecek dönem kredi talebinde bulunan müşterilerin

krediyi geri ödeyip ödemeyeceğinin tahmini yapılabilir. Tahmin edici model tabanlı geliştirilen modeller, Sınıflandırma (Classification) ve Regresyon (Regression) Modelleridir.

3.1.1.1 Sınıflandırma

Sınıflandırma en çok bilinen veri madenciliği tekniklerinden birisidir. Verileri belli nitelik değerlerine göre sınıflandırmaya yarar. Bunun için girdi olarak verilerin belirlenen belli nitelikleri, çıktı olarak da verilerin belli bir niteliği verilir ve algoritma bu çıktı niteliğindeki değerlere ulaşmak için hangi girdi nitelik değerlerinin kullanılması gerektiğini keşfeder [11]. Resim, örüntü, tanıma, hastalık tanıları, dolandırıcılık tespiti, kalite kontrol çalışmaları ve pazarlama konuları sınıflandırma tekniklerinin bolca kullanıldığı alanlardır [2]. Sınıflandırmada kategorik hedef değişkeni söz konusudur. Bu model, her kaydın tahmin edilen kategorik değişken için bilgi içerdiği büyük veri kümelerini inceler [5].

Sınıflandırma verileri önceden tanımlanmış sınıflara atayan bir öğrenme fonksiyonudur [8]. Matematiksel olarak ifadesi şöyle tanımlanabilir [2]:

$D = \{t_1, t_2, \dots, t_n\}$ bir veri tabanı olsun ve herbir t_i bir kaydı temsil etsin.

$C = \{C_1, C_2, \dots, C_m\}$ ise m adet sınıftan oluşan sınıflar kümesini temsil etsin.

$f : D \rightarrow C$ e her bir t_i bir sınıfa dahil olmalıdır.

Her bir C_j ayrı bir sınıftır ve her bir sınıf kendisine ait kayıtları içerir. Yani

$$C_j = \{t_i \mid f(t_i) = C_j, 1 \leq i \leq n, \text{ ve } t_i \in D\}. \quad (3.1)$$

Sınıflandırmada amaç, yeni bir veri elemanını daha önceden belirlenmiş sınıflara atamak için bir model oluşturmaktır. Veritabanında yer alan örnekler, bir sınıflama fonksiyonu yardımıyla kullanıcı tarafından belirlenmiş ya da karar niteliğinin bazı değerlerine göre anlamlı ayrık alt sınıflara ayrılır. Sınıflama algoritması, bir sınıfı diğerinden ayıran örüntüleri keşfeder. Bu bağlamda kullanılan belki de en önemli değerlendirme kriterleri tahmini doğruluk ve anlaşılabilirliktir. Tahmini doğruluk, genelleme olarak ta bilinir ve oluşturulan modelin daha önce görülmemiş örnekleri

sınıflandırmada ne kadar performanslı olduğunun bir ölçüsüdür. Anlaşılabilirlik ise, oluşturulan modelin kullanıcılar tarafından anlaşılabilir olmasını sağlar [26].

Sınıflandırmada kullanılan en genel veri madenciliği yöntemleri k-en yakın komşu, karar ağacı ve yapay sinir ağlarıdır [6]. İlerleyen bölümlerde bu sınıflandırma yöntemleri anlatılacaktır.

Sınıflandırma modeli örneklerinin birkaçı şöyledir [5]:

- Hileli kredi kartı işlemlerinin tespit edilmesi
- Üniversite sınavında belirli bir puanın üzerinde, devlet lisesi mezunu öğrencilerin tespit edilmesi

3.1.1.2 Regresyon

Regresyon analizi herhangi bir değişkenin bir veya daha başka değişkenler arasındaki ilişkinin matematiksel bir denklem şeklinde ifade yazılmasıdır. Yazılan bu denkleme regresyon veya eğri uydurma denklemi denir [2]. Regresyon modelini sınıflandırma modelinden ayıran özellik tahmin edilen hedef değişkenin kategorik değil sayısal (süreklilik gösteren) değişken olmasıdır.

Araştırmacılar mevcut veriler üzerinde çalışırlar ve hedef değişken ile tahminler arasındaki ilişkiyi yeni gözlemlerin üzerinde kullanırlar. Örneğin cinsiyet, yaş, boy ve ağırlıklara göre yetişkin insanların normal tansiyon değerleri modellenenbilir. Bu model bize yeni bir hastanın cinsiyetine, yaşına, ağırlığına ve boyuna göre normal tansiyon değerini hesaplamamızı mümkün kılar [9].

Regresyon analizi, araştırma, matematik, finans, ekonomi, tıp gibi bilim alanlarında yoğun olarak kullanılmaktadır. Eğri uydurma modeli örneklerinin birkaçı şöyledir:

- Şehir nüfusu ile suç işleme oranı arasındaki ilişkinin belirlenmesi
- Bir albüme harcanan reklam bütçesiyle o albüm piyasaya çıkmadan önceki hafta albümün radyoda çalınma sayısının ve albümü yapan grubun çekiciliğinin albüm satışları üzerindeki etkisini belirlenmesi

3.1.2 Tanımlayıcı modeller

Karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanmasını sağlayan modellerdir. Belirli bir gelir düzeyindeki çocuklu ailelerin ve bu gelir düzeyinden düşük geliri olan çocuksuz ailelerin aynı satın alma eğilimlerinde olmalarının belirlenmesi tanımlayıcı modele bir örnektir. Tanımlayıcı modeli temel alan modeller, Kümeleme (Clustering), Ardışık Zamanlı Örüntüler (Sequential Pattern) ve Birliktelik Kuralları (Association Rules) Modelleridir.

3.1.2.1 Kümeleme

Bölümlenme olarak da bilinen kümeleme, birbirinden farklı kümelerin belirlenmesi ve benzer özellik taşıyan verilerin ortak kümelere toplanması sonucunda veri tabanındaki verileri kümelere ayıran modeldir. Kümeleme modelinde, sınıflama modelinde olan veri sınıfları yoktur [30]. Sınıflandırma modelinde veri sınıfları önceden bilinmektedir ve yeni gelen verinin hangi sınıfa dahil olduğunun tahmin edilmektedir. Kümeleme modelinde ise herhangi bir sınıfa dahil edilmeyen verilerin gruplar halinde kümelere ayrılırlar [30]. Yani önceden saptanmış bir değişkene bağımlı olarak sınıflandırma yapmak söz konusu değildir. Örneğin, cinsiyete veya her hangi bir durumdaki başarıya göre bir sınıflandırma söz konusu değildir. Eğer veriler çeşitli özelliklere göre bazı gruplara kendiliğinden ayrışıyorsa, bu kümeler onların bazı özellikleri açısından birbirlerine benzer ve yakın olduklarını gösterir. Kendi içlerinde homojen olan ancak birbirlerinden farklı özellikler taşıyan gruplar oluşur. Bu yöntem daha çok doğal bir kümelene olduğu hissedilirse kullanılabilir. Bazı uygulamalarda kümeleme modeli sınıflandırma modelinin bir ön işlemi olarak görev alabilir [30].

Birbirine benzeyen nesnelerin aynı grupta toplanması mantığına dayanan bu modelde en büyük etken hangi kriterlere göre kümeleme yapılacağıdır. Bu işlem konu ile ilgili uzman tarafından tahmin edilir [12]. Veriler kümeleme işleminde aynı sınıfta yer almalarına rağmen farklı gruplarda da yer alabilir [22].

Nüfus bilimi, astronomi, arkeoloji, kimya, eğitim, psikoloji, dilbilimi ve sosyoloji alanında kullanımları yaygındır [1]. Elektronik ticaret uygulamalarında müşterilere özel hizmet sunabilmek için gerekli olan pazar bölümlenmesi kümeleme sayesinde yerine getirilebilmektedir [28]. Müşterilere ait bir veri deposunda yapılırken müşteriler, birçok özellikleriyle birlikte analiz edilir ve sonuçta müşteri kimlikleriyle,

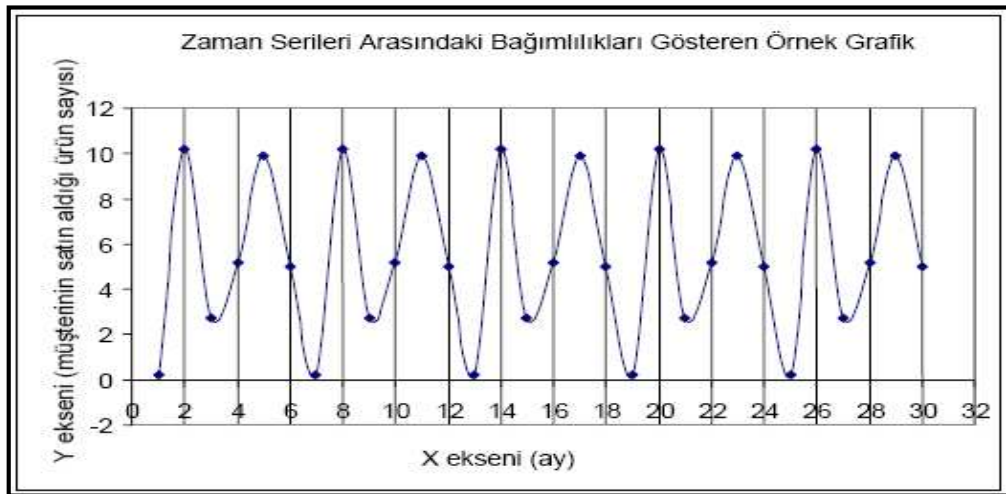
müşteri adlarına, posta kodlarına veya tanımlanan müşteri numarasına göre kendiliğinden gruplanırlar. Tüm müşteriler, kendisiyle benzer özelliklere, niteliklere sahip olan müşterilerle aynı gruba atanır. Kümeleme analizinin sonuçlarını kullanacak kişilerin, ayırtan bu grupları daha sonradan tanımlaması ve pazar bölümü olarak hedeflemesi mümkündür. Çünkü kendi içinde çok çeşitli açılardan benzer özellikler, benzer tutum ve davranışlar gösteren bu grupların pazarlama faaliyetlerinde de benzer tepkiler oluşturacağı varsayılmaktadır.

Örnekleme modeli örneklerinin birkaçı şöyledir:

- Biyolojide bitki ve hayvanların kümelenmesi
- Şehir planlama için şehirlerdeki yapı tiplerine, değerlerine ve coğrafik konumlarına göre kümelenmesi
- Elektronik ticaret uygulamalarında müşterilere özel hizmet sunabilmek için gerekli olan pazar bölümlenmesi

3.1.2.2 Ardışık zamanlı örüntüler

Zaman serilerindeki örüntü belli bir periyotta, belli bir sıklıkta gerçekleşen olaylardır. Bu yöntem belirli bir olay veya eylemin bir başkasını izlemesindeki örüntüleri yakalamak için kullanılır [45]. Şekil 3.3'te X eksenini sıfır zamanını referans alan ve geçen süreyi ay cinsinden ifade ediyorsa, Y eksenini müşterinin satın aldığı ürün sayısını gösteriyorsa bu grafikteki 6 ay süreyle tekrarlanan örüntü keşfedilebilmektedir [11].



Şekil 3.3 : Zaman serileri arasındaki bağımlılıkları gösteren örnek grafik.

Örneğin, müşteriler e-ticarete yazın yazlık ürünlere, kışın da kışlık ürünlere rağbet gösteriyorsa bu 6 ay periyotla tekrarlanan bir örüntüdür. Ardışık zamanlı örüntü modeli bu örüntüleri bulmayı amaçlar.

Belli frekansla tekrarlanan bu olaylar zaman serileriyle yapılan veri madenciliği algoritmaları sayesinde keşfedilir. Sıralı örüntü bulma işleminde, belirli zaman aralıklarında oturumlar incelenir ve karşılaştırmalar yapılır [28]. Sıralı örüntülerin bulunması gelecekteki eğilimi tahmin edecek web pazarlamacıları için oldukça anlamlıdır.

Ardışık zamanlı örüntüler modeli örneklerinin birkaçı şöyledir:

- İlk üç taksitini ödemeyen veya geç ödemiş müşteriler %60 olasılıkla kanuni takibe gidiyor olmasının belirlenmesi
- Basketbol oyununda hangi oyuncu tarafından hangi yolları izleyerek hareket ettiği belirlenerek, belirli bir oyuncunu 3 veya 5 saniye sonra ne yapacağının önceden belirlenmesi
- X ameliyatı yapıldığında, 15 gün içinde % 45 ihtimalle Y enfeksiyonu oluşacaktır [21].

3.1.2.3 Birliktelik kuralları

Birliktelik kuralı belirli türlerdeki veri ilişkilerini tanımlayan bir modeldir. Bu sebeple tanımlayıcı modeller grubuna dahil olur.

Çok sayıda verinin depolandığı bir veri tabanı içinde çeşitli nitelikler arasında hemen fark edilemeyen bir takım ilişkiler mevcut olabilir. Bu tip ilişkilerin ortaya çıkarılması stratejik kararların alınmasına yardımcı olabilir. Ancak, bu ilişkilerin çok sayıda verinin içinden elde edilmesi basit bir süreç değildir.

Birliktelik kuralları verilerin if-then kurallarına göre şekillenmelerini sağlar. If-then koşuluna göre şekillendirme mantığından dolayı birliktelik kuralları modeli olasılıksal bir modeldir [1]. If-then kuralları aşağıdaki gibi gösterilirler:

If <bazı şartlar sağlanırsa> Then <bazı niteliklerin değerlerini tahmin et>

Bu gerek birbirini izleyen gerekse eş zamanlı durumlarda araştırma yapar. Ürünler ve bu ürünlerin birlikte alınmaları söz konusu olunca, hemen anlaşılacağı gibi birliktelik kuralları daha çok perakendecilik sektöründe faaliyet gösteren işletmelerde

kullanılmaktadır [3]. Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak pazar sepeti analizi (Market Basket Analysis) adı altında veri madenciliğinde yaygın olarak kullanılmaktadır [10]. Bunun yanında finans ve tıp alanında da kullanılmaktadır.

Birliktelik kuralı modeli örneklerinin birkaçı şöyledir:

- Bir mağazadan gömlek alan müşterilerin %55'inin aynı alışverişte kravat satın aldıklarının tespit edilmesi [3]
- Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diyet süt satın alması [10].

3.2 K En Yakın Komşu Yöntemi ile Sınıflandırma

En yaygın kullanılan algoritmalarından birisidir. K en yakın komşu (K-Nearest Neighbour - KNN) algoritması, eğitilmiş öğrenme algoritmasıdır ve amacı, yeni bir örnek geldiğinde var olan öğrenme verisi üzerinde sınıflandırma yapmaktır. Algoritma, yeni bir örnek geldiğinde, onun en yakın k komşusuna bakarak örneğin sınıfına karar verir [39]. Algoritmanın adından da anlaşılacağı gibi k adet kayıt, başka bir deyişle veritabanındaki nokta, mesafesi hesaplanan noktaya diğer kayıtlara nazaran en yakın olan kayıtlardır [3].

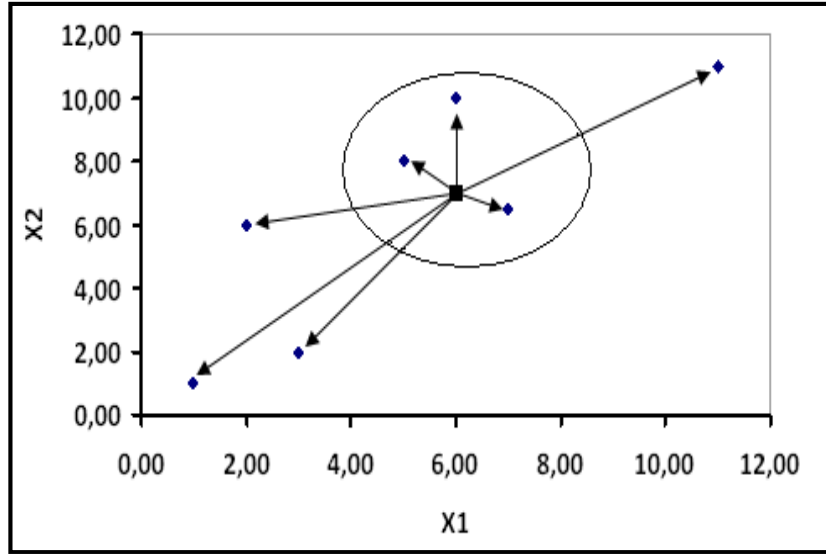
Algoritmanın uygulanması şu adımlardan ibarettir:

- Bir k parametresi belirlenir. Bu parametre verilen bir noktaya en yakın komşularının sayısıdır.
- Bu algoritma verilen bir noktaya en yakın komşuları belirleyeceği için, söz konusu nokta ile diğer tüm noktalar arasındaki uzaklıklar tek tek hesaplanır. Uzaklıkların hesaplanmasında i ve j noktaları için aşağıdaki Öklit uzaklık formülü kullanılabilir.

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3.2)$$

- Hesaplanan uzaklıklara göre satırlar sıralanır ve bunlar arasından en küçük olan k tanesi seçilir.
- Seçilen satırların hangi kategoriye ait oldukları belirlenir ve en çok tekrarlanan kategori değeri seçilir.

- Seçilen kategori, tahmin edilmesi beklenen gözlem değerinin kategorisi olarak kabul edilir [2].



Şekil 3.4 : Verilen bir ■ noktasına en yakın $k=3$ komşusunun belirlenmesi.

Bu yöntem coğrafi bilgi sistemlerinde çok kullanılır, örneğin belirlenen bir noktaya en yakın şehir, istasyon vs belirlenmesi aslında KNN algoritmasının temelini oluşturur [Beyer, 1999] [3].

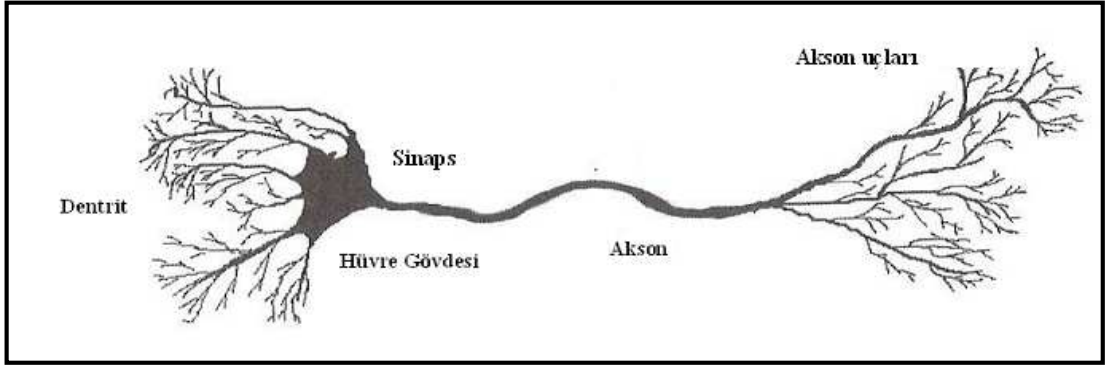
3.3 Yapay Sinir Ağları Yöntemi ile Sınıflandırma

Yapay sinir ağları, biyolojik sinir ağlarından esinlenerek geliştirilmiş bir bilgi işleme sistemidir [3]. İnsan beyninin fizyolojisini taklit ederler. Komplike ve belirsiz veriden bilgi üretirler. Keşfettikleri örüntü ve eğilimler, insanlar ya da bilgisayarlarca kolay keşfedilemez [41]. Bu tür karmaşık problemlerde birbirleriyle etkileşimli yüzlerce değişken bulunur [49]. En önemli özelliği, deneyimlerden (tecrübe) yararlanarak öğrenebilmesidir. Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilmişlerdir [42].

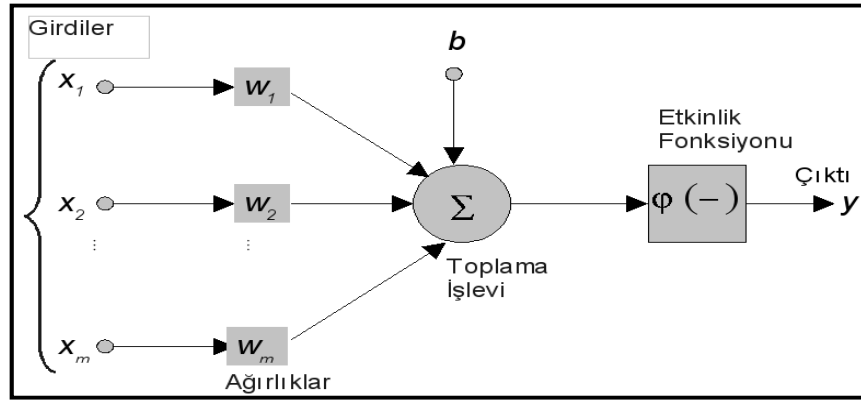
Bu teknik, veritabanındaki örüntüleri, sınıflandırma ve tahminde kullanılmak üzere genelleştirir. Sinir ağları algoritmaları sadece sayısal veriler üzerinde çalışırlar [41].

Yapay sinir ağlarının endüstriyel uygulamalar, finans uygulamaları, askeri ve savunma uygulamaları, tıp ve sağlık uygulamaları, mühendislik uygulamaları,

robotbilim, görüntü işleme, örüntü tanıma dışında iletişim sanayi, eğlence amaçlı tahmin gibi özel uygulama alanları da bulunmaktadır [42].



Şekil 3.5 : Biyolojik sinir hücresi.



Şekil 3.6 : Yapay sinir hücresi.

Yapay sinir ağları esas olarak biyolojik nöronların basitleştirilmiş bir şekilde matematiksel olarak modellenmesine dayanmaktadır [13,14]. Biyolojik ve yapay sinir hücreleri arasındaki benzeşim, hücreler arasındaki bağlantılar (hücreye gelen ve hücreden çıkan uyarımlar) akson ve dendritlere; ağırlık faktörleri (w_m) sinapslara ve sınır değerler (b) hücre içi kararlılık haline karşılık gelecek şekilde oluşturulmaktadır. Şekil 3.6'da gösterilen yapay sinir hücresinde, ağırlık faktörünün etkisine bağlı olarak ($w_1, w_2, \dots, w_m$) hücreye gelen uyarımlar ($x_1, x_2, \dots, x_m$) hücre içi denge durumu veya sınır değeri (d) de dikkate alınarak doğrusal olmayan bir aktivasyon fonksiyonu $\phi(-)$ yardımıyla çıktı şeklinde uyarımlara (y) dönüştürülür [43].

3.4 Karar Ağaçları Yöntemi ile Sınıflandırma

Karar Ağaçları, veri madenciliğinde kullanılan birçok sınıflandırma yönteminden bir tanesidir. 'Böl-ve-yönet' stratejisini uygular. Karar ağaçları verileri saf alt sınıflara ayırmaya çalışır. Kurulumlarının ucuz olması, yorumlanabilmelerinin kolay olması,

veri tabanlarına kolay entegre edilebilmeleri ve güvenilir olmaları sebebiyle karar ağaçlarının kullanımı oldukça yaygındır.

Quinlan (1993), karar ağaçlarını akış şemalarına benzeyen yapılar olarak tanımlamaktadır. Her bir nitelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının elemanlarıdır. En son yapı yaprak, en üst yapı kök ve bunların arasında kalan yapılar ise dal olarak isimlendirilir [3]. Karar ağacının tasarlanması, karar düğümlerinin toplanması, dallarla birbirine bağlanması, ana düğümden yaprak düğümlerde son bulana kadar aşağıya doğru uzanması adımlarından oluşmaktadır. Karar ağacını diyagramının başında ana düğüm, kök bulunur, niteliklerin olası sonuçları bir dal oluşturacak şekilde karar düğümlerinde test edilir. Her bir dal diğer bir karar düğüme veya sonlandırılmış yaprak düğüme ulaşır. Bu teknikte her kayıt ağaca uygulanır ve çıkan sonuca göre de bu kayıt sınıflandırılır. Temel olarak iki adımdan oluşur: Birincisi ağacın kurulması, ikincisi de verilerin teker teker ağaca uygulanarak sınıflandırmanın gerçekleştirilmesi şeklindedir [2].

Karar ağaçlarını dallanmanın hangi kritere göre yapılacağına göre çeşitli algoritmalar geliştirilmiştir. Bu algoritmaları entropiye dayalı algoritmalar, sınıflandırma ve regresyon ağaçları (CART), bellek tabanlı sınıflandırma algoritmaları olarak üç grupta toplayabiliriz.

En yaygın kullanılan karar ağacı algoritmaları;

- CHAID (Chi-Squared Automatic Interaction Detector , Kass 1980),
- CART (Classification and Regression Trees, Breiman ve Friedman, 1984),
- ID3 (Induction of Decision Trees, Quinlan, 1986),
- C4.5 (Quinlan, 1993) [41].

Entropiye dayalı bölümlmeyi kullanan algoritmalara örnek ID3 ve onun gelişmiş biçimi olan C4.5 algoritmaları verilebilir. Sınıflandırma ve regresyon ağaçları konusunda ise Twoing ve Gini algoritmalarından söz edilebilir. Bellek tabanlı sınıflandırma yöntemleri arasında k-en yakın komşu algoritması sayılabilir [2].

Karar ağaçları belirli bir sınıfın muhtemel üyelerinin belirlenmesi (Segmentation), vakaların risk gruplarına göre kategorize edilmesi (Stratification), gelecekteki olayları tahmin edebilmek için kuralların oluşturulması, parametrik modellerin kurulması için çok çeşitli verilere içinden faydalı olacakların seçilmesi, kategorilerin birleştirilerek sürekli değişkenlerin kesikli hale dönüştürülmesi gibi alanlarda kullanımı oldukça yaygındır [21].

3.4.1 Karar ağaçlarında dallanma kriterleri

Karar ağaçlarının inşa edilmesinde en önemli husus dallanmanın ve bölünmenin hangi nitelikten başlayacağını belirlenmesidir. Bu işlem sırasında entropi kavramından yararlanılır. Entropi rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Entropi kavramı, Dunham'a (2003) göre eldeki bilginin sayısallaştırılmasıdır, Fiske'ye (1998) göre ise entropi beklentisizliğin maksimumlaştırılmasıdır [3].

$D = \{t_1, t_2, \dots, t_n\}$ bir veri tabanı olsun. t_i verilerinin oluşma olasılıkları p_i 'dir. $P = \{p_1, p_2, \dots, p_n\}$ olasılık dağılımına sahip D veritabanının entropisi $H(D)$ olarak ifade edilir aşağıdaki formül ile hesaplanır:

$$H(D) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (3.3)$$

Entropi 0-1 arasında değer alır. Bütün olasılıklar eşit olduğunda entropi maksimum değerine ulaşacaktır [3]. Verilerin tümü aynı sınıfa dahil olması yani olasılığın 1 veya 0 olması durumunda entropi minimum değerini alır. Örneğin veri kümesindeki kişilerin hepsinin yaşının 26 olması durumunda entropi değeri sıfırdır.

$$P = \{0,1\} \quad H(D) = 0$$

$$P = \{0.67, 0.33\} \quad H(D) = 0.92$$

$$P = \{0.5, 0.5\} \quad H(D) = 1$$

Örnek: Çizelge 3.1'deki örnek verilerin kredi riski bağımsız değişkeni için entropi değerini bulalım.

Çizelge 3.1 : Müşteri eğitim verileri.

Müşteri	Birikim	Mal Varlığı	Gelir (1000\$)	Kredi Riski
1	Orta	Yüksek	75	İyi
2	Düşük	Düşük	50	Kötü
3	Yüksek	Orta	25	Kötü
4	Orta	Orta	50	İyi
5	Düşük	Orta	100	İyi
6	Yüksek	Yüksek	25	İyi
7	Düşük	Düşük	25	Kötü
8	Orta	Orta	75	İyi

Toplam 8 adet kredi riski verilerinden 3'ü kötü, 5'i iyi kredi sınıfında bulunmaktadır. Bu durumda,

$$H(D) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (3.4)$$

$$H(D) = -\frac{3}{8} \log_2\left(\frac{3}{8}\right) - \frac{5}{8} \log_2\left(\frac{5}{8}\right) = 0.9544$$

Dallanmanın hangi nitelikten başlayacağını belirlemek için her niteliğin kazanç değeri (*Gain*) hesaplanır. Kazanç değeri hesaplanırken yukarıda bahsettiğimiz entropi değerinden yararlanılır ve kazanç değeri en büyük olan nitelik karar ağacının kökünü oluşturur.

Hedef değişkeni T ile, hedef niteliği olmayan bir X niteliğine bağlı olarak ayrıldığı alt kümelerini $T_1, T_2, \dots, T_n$ ile ifade edelim. T 'nin bir elemanının sınıfını belirlemek için gerekli bilgi şu şekilde hesaplanır:

$$H(X, T) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i) \quad (3.5)$$

T veri tabanının X testine göre bölmekle elde edilen bilgileri ölçmek için:

$$Kazanç(X, T) = H(T) - H(X, T) \quad (3.6)$$

Kazanç değeri en yüksek olan yani en yüksek bilgi kazancı sağlayan X değeri seçilir.

Örneğin aşağıdaki müşteri eğitim verilerin kredi riski hedef değişkenine göre birikim niteliğinin kazanç ölçütünü hesaplayalım.

Çizelge 3.2 : Müşteri eğitim verileri.

Müşteri	Birikim	Mal Varlığı	Gelir (1000\$)	Kredi Riski
1	Orta	Yüksek	75	İyi
2	Düşük	Düşük	50	Kötü
3	Yüksek	Orta	25	Kötü
4	Orta	Orta	50	İyi
5	Düşük	Orta	100	İyi
6	Yüksek	Yüksek	25	İyi
7	Düşük	Düşük	25	Kötü
8	Orta	Orta	75	İyi

Adım 1:

Öncelikle sırasıyla her niteliğin bağımsız değişkene göre kazanç ölçütleri hesaplanır. Birikim niteliğinin kredi riski hedef değişkenine göre kazanç ölçütünü hesaplayalım.

Çizelge 3.3 : Birikim ve Kredi Riski nitelik değerleri.

Birikim	Kredi Riski
Orta	İyi
Düşük	Kötü
Yüksek	Kötü
Orta	İyi
Düşük	İyi
Yüksek	İyi
Düşük	Kötü
Orta	İyi

$$H(Brikim, Risk) = \frac{2}{8} H(Birikim_{yüksek}) + \frac{3}{8} H(Birikim_{orta}) + \frac{3}{8} H(Birikim_{düşük})$$

Yukarıdaki ifade içinde yer alan $H(Birikim_{yüksek})$, $H(Birikim_{orta})$, $H(Birikim_{düşük})$ entropileri şu şekilde hesaplanır:

Birkimin Yüksek olması durumunda 1 kötü, 1 iyi kredi riski değeri,

Birkimin Orta olması durumunda 3 iyi kredi riski değeri,

Birkimin Düşük olması durumunda 2 kötü, 1 iyi kredi riski değeri karşılık gelmektedir.

$$H(Birikim_{yüksek}) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$H(Birikim_{orta}) = -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} = 0$$

$$H(Birikim_{düşük}) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.918$$

$$H(Brikim, Risk) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i) \quad H(Brikim, Risk) = \frac{2}{8}(1) + \frac{3}{8}(0) + \frac{3}{8}(0.918) = 0.5942$$

Kredi riski bağımsız değişkeni için entropi değerini bir önceki örnekte hesaplamıştık:

$$H(D) = 0.9544$$

Bu durumda birikim niteliğinin kazanç ölçütü:

$$Kazanç (Birikim, Risk) = 0.9544 - 0.5942 = 0.3602$$

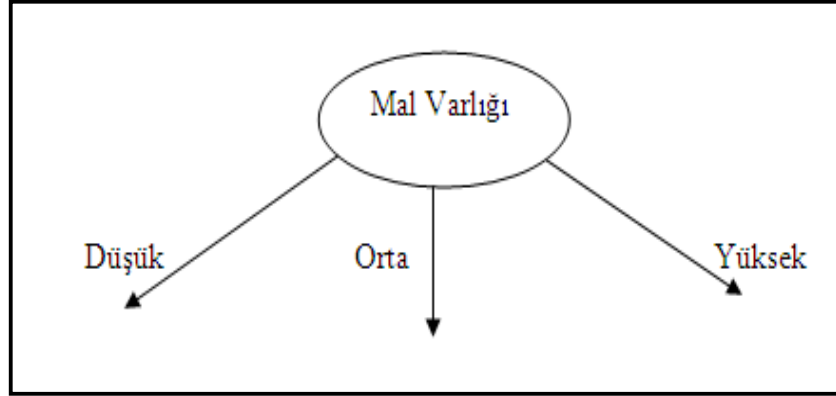
Benzer şekilde malvarlığı ve gelir nitelik değerlerine göre kazanç ölçütlerini aşağıda hesaplanmıştır. Gelir nitelik değerinin kazanç değerini hesaplarken verileri gelir ≤ 50 ve gelir > 50 kriterlerine göre sınıflandırdık.

$$Kazanç (MalVarligi, Risk) = 0.5488$$

$$Kazanç (Gelir, Risk) = 0.004$$

$$Kazanç (Birikim, Risk) = 0.3602$$

Karar ağacı en yüksek kazanç değerini elde ettiğimiz nitelik değerinden yani mal varlığı niteliğinden dallanmaya başlayacaktır (Şekil-3.7). İkinci seçenek olarak yakın değerler olduğu için birikim niteliği de tercih edilebilir.



Şekil 3.7 : Mal Varlığı ile ilgili işlemler.

Adım 2:

Kök değişkeninin değerlerine göre alt karar ağacı hesaplamalarına gidilir ve süreç dallanma yapraklarla son bulana kadar devam eder.

Mal varlığının niteliğinin orta değeri içi alt karar ağacının düzenleyelim.

Çizelge 3.4 : Mal Varlığı orta olan müşteri verileri.

Müşteri	Birikim	Mal Varlığı	Gelir (1000\$)	Kredi Riski
3	Yüksek	Orta	25	Kötü
4	Orta	Orta	50	İyi
5	Düşük	Orta	100	İyi
8	Orta	Orta	75	İyi

Yeni ağaç oluşturacak verilerin entropi değeri:

$$H(D) = -\frac{1}{4} \log_2 \left(\frac{1}{4} \right) - \frac{3}{4} \log_2 \left(\frac{3}{4} \right) = 0.8113$$

Gelir niteliğinin kredi riski hedef değişkenine göre kazanç ölçütünü hesaplayalım. Daha önce gelir niteliğinin değerlerini ≤ 50 olması düşük, > 50 olması yüksek olarak sınıflandığımızı belirtmiştik.

Çizelge 3.5 : Gelir ve Kredi Riski nitelik değerleri.

Gelir (1000\$)	Kredi Riski
25	Kötü
50	İyi
100	İyi
75	İyi

$$H(\text{Gelir}, \text{Risk}) = \frac{2}{4} H(\text{Gelir}_{\text{yüksek}}) + \frac{2}{4} H(\text{Gelir}_{\text{düşük}})$$

Yukarıdaki ifade içinde yer alan $H(\text{Gelir}_{\text{yüksek}})$, $H(\text{Gelir}_{\text{düşük}})$ entropileri şu şekilde hesaplanır:

Gelirin Düşük olması durumunda 1 iyi, 1 kötü kredi riski değeri,

Gelirin Yüksek olması durumunda 2 iyi kredi riski değeri karşılık gelmektedir.

$$H(\text{Gelir}_{\text{yüksek}}) = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} = 0$$

$$H(\text{Gelir}_{\text{düşük}}) = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$H(\text{Gelir}, \text{Risk}) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i) \quad H(\text{Gelir}, \text{Risk}) = \frac{2}{4}(1) + \frac{2}{4}(0) = 0.5$$

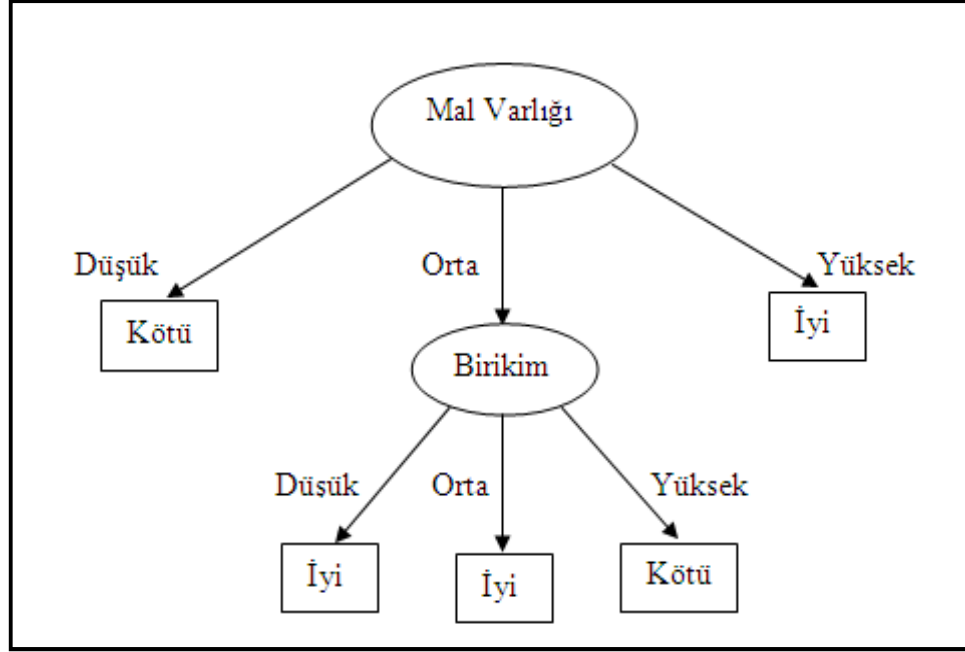
Bu durumda birikim niteliğinin kazanç ölçütü:

$$\text{Kazanç}(\text{Gelir}, \text{Risk}) = 0.8113 - 0.5 = 0.3113$$

Benzer şekilde birikim nitelik değerine göre kazanç ölçütü değeri hesaplanmıştır.

$$\text{Kazanç}(\text{Birikim}, \text{Risk}) = 0.8113$$

Orta düzey mal varlığına göre oluşacak alt ağacın dallanma kriteri kazanç ölçütü en yüksek olan birikim niteliği olacaktır. Bu şekilde her dal için hesaplamalar yapıldığında karar ağacının sonhali aşağıdaki gibi olacaktır:



Şekil 3.8 : Sonuç karar ağacı.

Entropi ve kazanç dışında farklı karar ağacı algoritmalarında farklı algoritmalarından yararlanıldığından bahsetmiştik. Bunlar kazanç oranı, towing ve gini algoritmalarıdır.

Kazanç oranı algoritması:

$$Kazanç\ oranı(X,T) = \frac{Kazanç(X,T)}{H(P_{X,T})} \quad (3.7)$$

Towing algoritması adımları:

- Niteliklerin içerdiği değerler gözönüne alınarak eğitim kümesi iki ayrı dala ayrılır. Bir t düğümünde “sağ” ve “sol” olmak üzere iki ayrı dal bulunur. Bu bölünlenen kümeler $t_{sağ}$ ve t_{sol} biçimindedir.
- Bölünmelerin her biri için P_{sol} , $P(j|t_{sol})$ ve $P_{sağ}$, $P(j|t_{sağ})$ olasılıkları aşağıdaki formüller yardımı ile hesaplanır.

x : t_{sol} 'daki herbir nitelik değerin ilgili nitelik sütunundaki tekrar sayısı

y : Eğitim kümesindeki kayıtların sayısı

$$P_{sol} = \frac{x}{y} \quad (3.8)$$

z : t_{sol} 'daki kayıtların j sınıfları sayısı

w : t_{sol} 'daki herbir nitelik değerin ilgili nitelik sütunundaki tekrar sayısı

$$P(j | t_{sol}) = \frac{z}{w} \quad (3.9)$$

- $\Phi(s | t)$, t düğümündeki s bölünmelerin uygunluk ölçüyü olsun. Uygunluk ölçütü en büyük olanı seçilir. Bu değerin ilgili olduğu bölünme satırı bize dallanmanın yapılacağı satırı bildirir.

$$\Phi(s | t) = 2P_{sol}P_{sağ} \sum_{j=1}^n |P(j | t_{sol}) - P(j | t_{sağ})| \quad (3.10)$$

- Birinci adıma dönülerek ağacın alt kümesine aynı işlemler uygulanır.

Gini algoritması adımları:

- Her nitelik değerleri ikili olacak biçimde gruplanır. Bu şekilde elde edilen sol ve sağ bölünmelere karşılık gelen sınıf değerleri gruplandırılır.

- Herbir nitelik ile ilgili sol ve sağ taraftaki bölünmeler için $Gini_{sol}$ ve $Gini_{sağ}$ değerleri hesaplanır.

k : Sınıfların sayısı

T : Bir düğümdeki örnekler

$|T_{sol}|$: Sol taraftaki örneklerin sayısı

$|T_{sağ}|$: Sağ taraftaki örneklerin sayısı

L_i : Sol taraftaki kategorisindeki örneklerin sayısı

R_i : Sağ taraftaki kategorisindeki örneklerin sayısı

$$Gini_{sol} = 1 - \sum_{i=1}^k \left(\frac{L_i}{|T_{sol}|} \right)^2 \quad (3.11)$$

$$Gini_{sağ} = 1 - \sum_{i=1}^k \left(\frac{R_i}{|T_{sağ}|} \right)^2 \quad (3.12)$$

- Her j niteliği için, n eğitim kümesindeki satır sayısı olmak üzere aşağıdaki bağıntının değeri hesaplanır. En küçük $Gini_j$ değeri seçilir ve bölünme bu nitelik üzerinden gerçekleştirilir.

$$Gini_j = \frac{1}{n} \left(|T_{sol}| Gini_{sol} + |T_{sağ}| Gini_{sağ} \right) \quad (3.13)$$

- Baştaki adıma dönülerek ağacın alt kümesi için işleme devam edilir.

Tez kapsamında sigortacılık sektörü verileri karar ağacı yöntemi ile incelenerek hangi koşulların risk teşkil edeceğinin belirlenmesi için çalışılmıştır.

4. RİSK ANALİZİ

4.1 Risk Tanımı

Risk kelimesi günlük hayatta sıkça kullandığımız bir kelimedir. Risk sözcüğünün kökeni Arapça rızık/rısk (risq) yada Latince riziko (risicum) sözcüklerinden çıkmıştır. Rızık, kişiye Tanrı tarafından verilen ve üzerinden kar elde edilen herhangi bir şey olarak tanımlanabilir. Burada rızık rassal ve istenen iyi sonuç anlamı kazanır. Riziko ise, bir denizcinin karşılaştığı kayalık alan gibi bir engel olarak tanımlanır. Burada riziko, rassal istenmeyen kötü sonuç anlamına gelir. Arapça'dan ödünç alınan bu sözcük olumlu ya da olumsuz etkilere sahip rassal olayları açıklamak için kullanılmıştır. Çağdaş Fransızca'da risk (risque) sözcüğü yine rassal olaylara ilişkin olup, çoğunlukla olumsuz ama bazen olumlu anlam içerir [18]. Gerek İngilizce'de gerekse Türkçe'de risk (ya da riziko), kayıp, hasar tehlikesi ya da kayıp, hasar tehlikesi olasılığı, sigorta edilen şey ya da kimse olarak tanımlanır; ayrıca fiil olarak risk tehlikeye girmek ya da göze almak anlamlarında kullanılır ve sıfat olarak risk tehlikeli anlamına gelir [15]. Risk kelimesinin farklı şekillerde birçok tanımı vardır. Bunlardan bazıları:

Risk, bir olayın yada olaylar setinin ortaya çıkma olasılığıdır [17]. Risk, gerek belirsizlik gerekse belirsizliğin sonuçları olarak tanımlanabilir [15]. Risk, gerçekleştiğinde organizasyonun hedefleri üzerinde olumsuz etki yapan ve gerçekleşme ihtimali olan rast gele olaylardır [14]. Risk, zarar verici bir aktiviteye açık olma potansiyeli olarak tanımlanabilir. Tehditler, zayıflıklar, etkiler ve olabilirlik riskin bileşenleridir. Teknik anlamda risk, getirilere ilişkin olasılık değerlerinin ortalama değer etrafındaki dağılımı ile ifade edilebilir [19]. Risk belirsizliği, şüpheyi, kayıp olasılığını, zarar ihtimalini ifade eder. Risk belirli bir tehdidin sistemin belirli bir zayıflığından faydalanarak sisteme zarar verme olasılığıdır.

4.2 Risk Analizi

İnsanoğlunun rahat iletişim kurabilmeleri ve işlerini daha kolay sonuçlandırabilmeleri için çok çeşitli çalışmalar yapılmaktadır. Gelişen teknoloji ile süreçler günümüzdeki halini almışlardır.

Teknolojinin gelişimi ve bilgisayarlı sistemlerin kullanılmaya başlaması ile birçok faydanın yanı sıra güvenlik açıkları da meydana gelmiştir. Sistemlerin işleyişini, firmalarının karını veya süreçlerini olumsuz etkileyecek bu faktörlerin (arıza, saldırı vs.) belirlenmesi için risk oluşturan olayların tespit edilmesi ve bu tespitler sonucunda engelleyici faaliyetlerin oluşturulması gerekmektedir. Bu olayların belirlenmesi uygulanacak yöntemlerden biri Risk Analizidir. Risk Analizinin birçok tanımı vardır. Bunlardan bazıları:

Risk analizi, stratejik kararlarda ele alınan değişkenle ilgili olan riskin kapsamlı olarak anlaşılmasını sağlayan yöntemlerin bütünüdür [15]. Proje başarısını veya hedeflerine ulaşmasını tehlikeye sokacak faktörlerin tanımlanması ve değerlendirilmesi tekniğidir [16]. Risklerin niceliksel veya niteliksel olarak değerlendirildiği süreçtir [20]. Bir başka ifadeyle risk analizi risk faktörlerinin belirlenip, ölçeklendirilip, önlem alınması gereken alanların belirlenmesi işlemlerinin bütünüdür. Risk analizinin yapılabilmesi için öncelikle kullanılacak olan genel kavramların iyi algılanması gerekmektedir.

Kıymet (Asset): Korunması gereken her şeydir. Veri, kaynak, zaman veya saygınlık kıymetlere örnektir. Risk analizinde, iki farklı türde kıymet vardır. Somut (**tangible**) ve soyut (**intangible**) kıymetler. Donanım ve insan somut, yazılım, veri ve politikalar ise soyut kıymetlerdir.

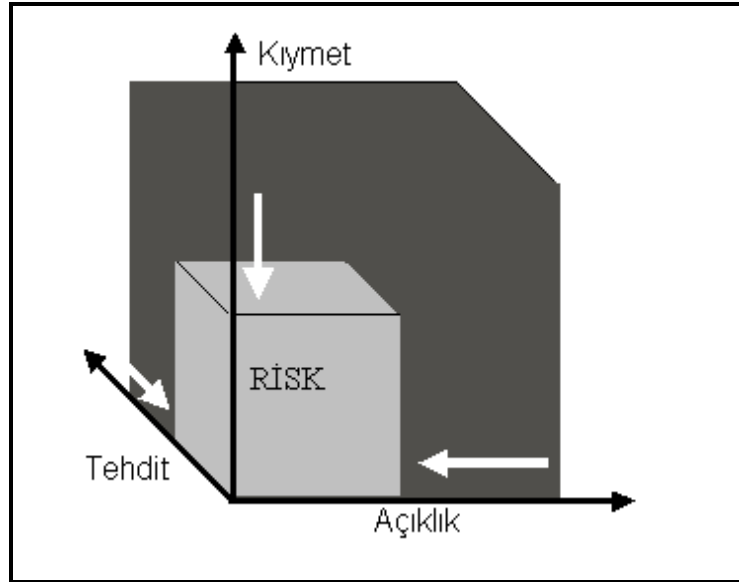
Tehdit (Threat): Bir kıymetteki açıklıkları kullanarak kıymete kısmen yada tamamen zarar veren etkenlerdir. Tehdit, zarar oluşması olasılığı olarak değerlendirilebilir. İki türlü tehdit vardır. Bunlar insanlar ve doğal kaynaklardır. Örneğin bir bilgisayar sistemindeki tehdit unsurları fırtına, yangın gibi doğal tehditlerle olabileceği gibi, kasıtsız veya bir amaca yönelik sistemi çökertmeye çalışan kasıtlı saldırılar da olabilir.

Açıklık (Vulnerability): Kıymeti tehditlere karşı korunmasız hale getiren unsurlardır. Bir başka ifadeyle bir sistemde yetkilendirilmemiş eylemlere izin veren zafiyetler veya tehdidin gerçekleşmesine sebep olabilecek zayıflıklardır. Örneğin bir

bilgi sistemindeki açıklık, bilgisayar programında kodsız hatalardan veya virüs tehdidine karşı antivirüs yazılımının bulunmamasından kaynaklanabileceği gibi doğal etkenlere (toz, nem... vb) karşı korunmasız olmasından da kaynaklanabilir. Tehditler bu açıklıkları kullanarak kıymetlere zarar verdikleri için açıklıklar riskin en önemli nedenidir.

Karşı Önlem (Safeguard): Kıymette bulunan açıklıkları kullanan tehditlerin verdiği zararları sıfırlamak yada en aza indirmek için alınan tedbirlere karşı önlem denir. Art niyetli eylemlerden ve etkilerden korunmak üzere alınan ve sürdürülen koruyucu önlemler sonucunda güvenlik durumu oluşur.

Arta Kalan Risk: Karşı önlemler uygulandıktan sonra geriye kalan risklere denir.



Şekil 4.1 : Kıymet, tehdit ve açıklık fonksiyonu olarak risk.

Risk analizinin uygulanmasında ilk olarak şirketin tüm kıymetleri ve her bir kıymetin içinde bulunduğu tehditleri belirlenir. Belirlenen kıymet, zayıflık ve tehditlerin girdileri matematiksel ve mantıksal olarak hesaplanıp, kıymet risk ağırlıklandırması yapılır. Risk analizi, kıymetleri, kıymetlerdeki açıklıkları, tehditleri çıkarır fakat nasıl karşı önlemler alınacağı üzerinde durmaz çünkü belirlenen risk maddelerine göre alınacak önlemleri belirleme işlemi risk yönetimi yönteminin uygulanması ile gerçekleştirilir. Risk analizi risk yönetimi için girdi oluşturur [33].

Örneğin bir sigorta şirketi nüfus bilimine yararlanarak belirli periyotlardaki sigortalıların ölümünün tahmin edebilir. Bu tahminleri polişe bilgileriyle eşleştirerek belirlenen periyotta sigortalının ne kadar ödeyeceğini belirleyebilir [20].

Nicel analiz ve nitel analiz olmak üzere iki temel risk analizi yöntemi vardır.

4.2.1 Nitel risk analizi

Nitel risk analizi daha çok riski değerlendirmede sayısal veriler elde etmek mümkün olmadığı durumlarda kullanılır. Ayrıca nitel risk analizine katılacak kişilerin matematik bilgisine sahip olmaları da gerekmez. Bu değerlendirmenin başarılı bir şekilde yapılabilmesi daha çok, katılımcıların kurumsal süreçlere hâkim olmalarına ve benzer projelerde deneyim sahibi olmalarına bağlıdır [50].

Nitel risk analizinde risk, sayısal değerlerden çok tanımlar ile ifade edilir. Riski hesaplarken ve ifade ederken nümerik değerler yerine yüksek, çok yüksek gibi tanımlayıcı değerler kullanılır.

Nitel risk analizi için risk olasılık / etki değerlendirmesi tekniği kullanılır.

Riskin gerçekleşme olasılığı 0 ile 1 arasında bir değerdir. Nitel risk analizinde riskin olma olasılığı; farklı şekillerde sınıflandırılabilir. Bazı durumlarda düşük, orta yüksek şeklinde belirlenebildiği gibi bazı risk analizlerinde, riskin olma olasılığı; her zaman, bazen, nadiren, neredeyse hiç şeklinde sınıflandırılabilmektedir. Ancak yapılan bütün sınıflandırmalar sonuçta 0 ile 1 aralığını bölmelere ayırmaktan ibarettir. Riskin olması durumunda oluşturacağı etki de aynı şekilde düşük, orta, yüksek şeklinde olabileceği gibi çok düşük, düşük, orta, önemli etki şeklinde de sınıflandırılabilir [53].

Risklerin gerçekleşme olasılıkları ve gerçekleşme etkileri belirlendikten sonra risk etki grafiğine yerleştirilerek değerlendirilir.

Örneğin bir fabrikada meydana gelebilecek iş kazaları ve etkileri için risk analizini uygulayalım. Kazaların meydana gelme olasılık ve etki ölçekleri sırasıyla Çizelge-4.1 ve Çizelge-4.2'teki gibi tanımlansın.

Çizelge 4.1 : Olasılık ölçeği.

Olasılık	Ortaya çıkma sıklığı / frekans için derecelendirme basamakları
Çok Küçük	Yılda Bir
Küçük	Üç Ayda Bir
Orta	Ayda Bir
Yüksek	Haftada Bir
Çok Yüksek	Her Gün

Çizelge 4.2 : Etki ölçeği.

Etki	Derecelendirme
Çok Hafif	İş saati kaybı yok, ilk yardım gerektiren
Hafif	İş günü kaybı yok, ilk yardım gerektiren
Orta	Hafif yaralanma tedavi gerektirir
Ciddi	Ölüm, ciddi yaralanma, meslek hastalığı
Çok Ciddi	Birden çok ölüm, sürekli iş görmezlik

Çizelge 4.3 : Risk Etki-Olasılık değerlendirme grafiği.

5-Çok Yüksek	5	10	15	20	25
4-Yüksek	4	8	12	16	20
3-Orta	3	6	9	12	15
2-Düşük	2	4	6	8	10
1-Çok Düşük	1	2	3	4	5
Olasılık / Etki	1-Çok Hafif	2-Hafif	3-Orta	4-Ciddi	5-Çok Ciddi

Risk etki-olasılık değerlendirme grafiğinde kırmızı ile işaretlenen risk kabul edilemez, sarı renk ile işaretlenen risk dikkate değer, yeşil ile işaretlenen risk ise kabul edilebilir riskleri ifade eder. Kabul edilemez riskler hemen müdahale

edilmesi gereken risklerdir. Dikkate değer riskler mümkün olduğu kadar çabuk müdahale edilmesi gereken risklerdir. Kabul edilebilir riskler ise acil tedbir alınmasını gerektirmeyen risklerdir [30].

Çizelge 4.4 : Risk verileri.

Risk	Olasılık	Etki
Risk1	Yüksek	Hafif
Risk2	Orta	Ciddi
Risk3	Çok Yüksek	Orta
Risk4	Düşük	Çok Hafif
Risk5	Yüksek	Çok Ciddi

Çizelge 4.5 : Örnek risk olasılık-etki değerlendirme grafiği.

5-Çok Yüksek			Risk3		
4-Yüksek	Risk1				Risk5
3-Orta				Risk2	
2-Düşük	Risk4				
1-Çok Düşük					
Olasılık / Etki	1-Çok Hafif	2-Hafif	3-Orta	4-Ciddi	5-Çok Ciddi

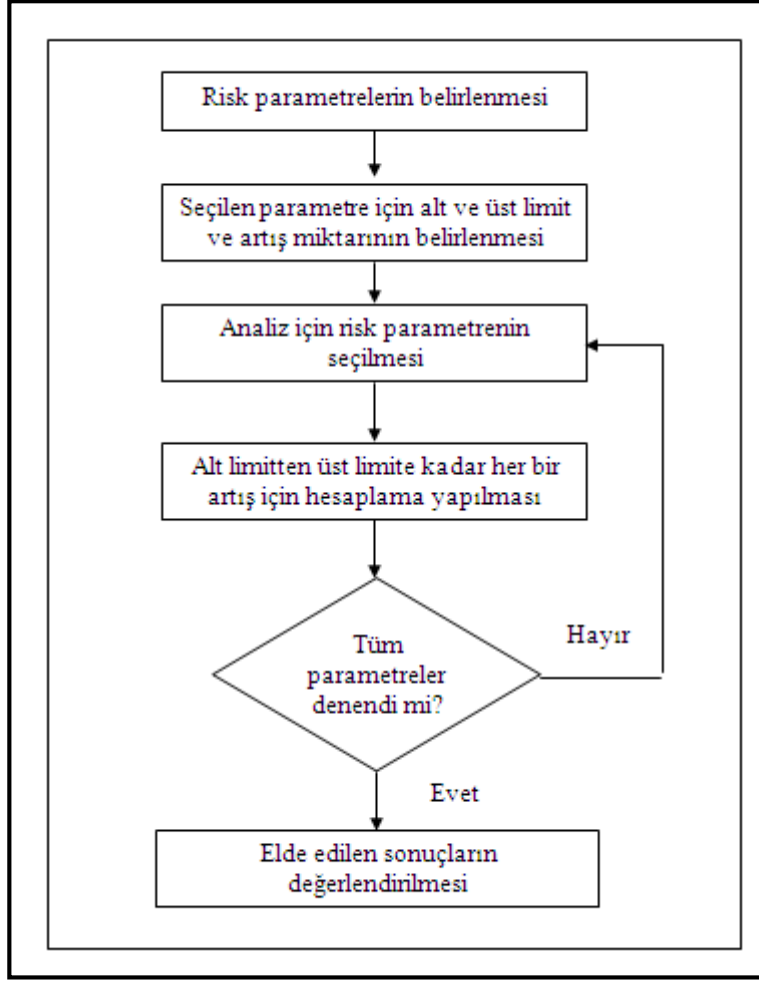
Bu grafiğe göre örneğin risklerini değerlendirecek olursak, Risk3 ve Risk5 en üst düzey risk yönetimi uygulanması gereken risklerken, Risk 2 izlenmesi ve yönetilmesi gereken riskler, Risk1 ve Risk4 ise acil müdahale edilmesi gerekmeyen, kabul edilebilir risklerdir.

4.2.2 Nicel risk analizi

Geçmişten dersler alınarak, daha sayısal değerlerle risklerin tehdidinin belirlenmesi çalışmasıdır. Nicel risk analizi, risklerin sonuçlarını ve olasılıklarını ölçmek, risklerin kıymetler üzerinde taşıdığı anlamı (etkiyi) tahmin etmektir [48]. Nicel risk analizi riski hesaplarken sayısal yöntemlerden yararlanır. Nicel risk analizinde kıymet, açıklık, tehdidin olma ihtimali, tehdidin etkisi gibi değerlere sayısal değerler verilir ve bu değerler matematiksel ve mantıksal metotlar ile proses edilip risk değeri belirlenir. Bu aşamada aynen nitel risk analizinde yer alan girdilere ihtiyaç duyulur. Kullanılacak araç ve teknikler ise şunlardır: Duyarlılık Analizleri, Karar Ağacı Analizleri, Simülasyon (Bu teknikler istatistik, yöneylem araştırması gibi matematiksel disiplinlerin türevleridir).

4.2.2.1 Duyarlılık analizi

Duyarlılık analizi, hangi risk parametresinin organizasyon hedeflerini veya projenin başarısını daha çok etkilediği tespit edilmeye yönelik bir tekniktir. Her parametrenin değişiminin % olarak hesaplanması işlemi, her defasında sadece bir değişkeni değiştirip diğerlerini sabit tutarak gerçekleştirilmektedir. Duyarlılık analizi yapabilmek için, duyarlılığı saptanacak değişken için bir alt limit, üst limit ve bu aralıkta artış miktarı belirlenmekte ve modele girdi olarak verilmektedir [31]. Duyarlılık analizinin akış şeması Şekil-4.2'deki gibidir:



Şekil 4.2 : Duyarlılık analizi akış diyagramı.

Bu yöntem riski büyüklük olarak ölçmeyi hedef etmez. Duyarlılık analizi riske duyarlı olan faktörleri tespit etmeyi amaçlar. Duyarlılık analizi karar vericini tüm “Böyle olursa sonuç ne olur?” tipindeki sorulara cevap arar. Örneğin, “Satış fiyatı % 10 düşerse normal şartlarda durumu ne olur? Projenin ömrü 5 değil de 3 yıl olursa iç verim oranı ne olur?” sorulara cevap verir. Duyarlılık analizi sonucu duyarlı olduğu belirlenen parametreler üzerinde işlemler yapılır.

4.2.2.2 Simülasyon analizi

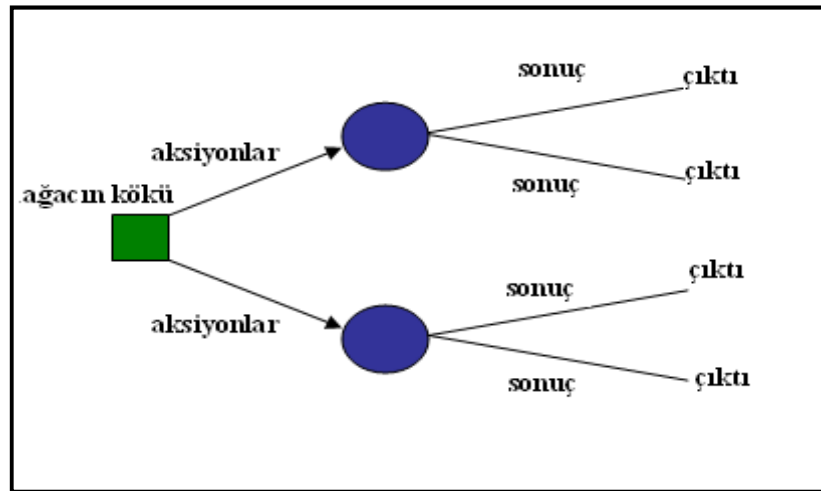
Gelişimi ve sonuçları öğrenmek istenen bir olayın laboratuvar ya da bilgisayar ortamında temsili olarak gerçekleştirilmesine simülasyon adı verilir. Benzetim, olayın ya da deneyin gerçekleştirilmesinin pahalı, tehlikeli ya da zor olması; olayın doğal dengelere bağlı bulunması nedeniyle insan eliyle kontrol edilememesi; analitik yöntemlerle hesaplanamaması; olayın değişik hızlarda gözlenmek istenmesi, bazı değer ve koşulların belirsizliği vb. durumlarda uygulanır [32].

En çok kullanılan simülasyon yöntemi Monte Carlo simülasyon yöntemidir. Monte Carlo simülasyonu birçok alanda kullanılan bir yöneylem araştırması tekniğidir. Monte Carlo yöntemi, deneysel ve istatistiksel problemlerin çözümüne rastgele sayılarla yaklaşımlara verilen genel bir isimdir. Metodun bir probleme uygulanması, problemin tesadüfi sayıları kullanarak simülasyon edilip hesap edilmek istenen parametrenin bu simülasyonlarının sonuçlarına bakılarak yaklaşık hesaplanması fikrine dayanır. Bu yaklaşıma göre yatırımın ana özelliklerini içeren matematiksel bir modelin kurulmasını öngörüyor.

4.2.2.3 Karar ağacı analizi

Karar ağacı analizi yöntemleri belirsiz bir ortamda karar verme olanağı sağlarlar. Karar ağacı, karar vericinin büyük bir projenin amaçlarını gerçekleştiren en iyi alternatifleri tanımlamasında, belli sayıda bilinen alternatifler ve onların olası sonuçlarını grafiksel olarak ağaç formunda gösterir. Karar ağacı yöntemi inşaat yapım yöntemlerine karar vermede, alternatif projeleri seçmede ve bir fikir ile devam etmek veya başarılı bir fikrin benzeri yaklaşımında bulunmak gibi sözleşme problemlerinde karar vermede kullanılır [29].

Karar ağacı yöntemi temel olarak beklenen maddi değer analizinin grafiğe dökülmüş halidir [54]. Bu yöntemde en soldan, ağacın kökünden, başlanarak sağa doğru, ağacın dallarına doğru, grafik çizilir. Her bir dalın beklenen maddi değeri hesaplanır. En soldaki noktadan en sağdaki noktaya varabilmek için en uygun(avantajlı) yol seçilir. Şekil 4.3 örnek bir karar ağacını göstermektedir.



Şekil 4.3 : Karar Ağacı.

5. SİGORTACILIK SEKTÖRÜNDE VERİ MADENCİLİĞİ UYGULAMASI

5.1 Sigortacılık Sektörü Hakkında Bilgi

Sigorta öncelikle çeşitli riskler karşısında kişi, kurum ve daha geniş olarak ülke için bir güvence aracıdır. Sigortanın ülke ve dünya ekonomisinde, ekonomik ve risk yönetimi ve girişim açısından işlevleri bulunmaktadır. Sigortanın ekonomik işlevleri arasında; önemli bir tasarruf kaynağı olması, kişi ve kurumların sosyo-ekonomik çöküntü ve kayıplarını önlemesi, uluslar arası ekonomik ilişkileri ve ticareti geliştirmesi ve vergi kaynağı olması yer almaktadır. Risk yönetimi ve girişimler yönünden işlevleri ise; bir dayanışma organizasyonu olması bakımından güven sağlaması, yatırım kararlarının alınması sırasında olumlu etkisinin bulunması, risk karşılığında atıl tutulan sermaye miktarını en aza indirmesi, fiyatların daha gerçekçi düzeyde oluşmasını sağlaması şeklinde sıralanabilmektedir.

Sigorta kavramının tek bir tanımı bulunmamakta, hukuk, ekonomi, tarih, aktüeryal bilimler, risk teorisi ve sosyoloji gibi birçok disiplinler tarafından tanımlanmaya çalışılmaktadır. “American Risk and Insurance Association”a ait risk ve sigorta terminolojisine göre sigorta; “tesadüfî zararları bir havuzda toplayıp, riskleri, gelecekte karşılaşılabileceği zararlar için kendisini teminat altına alan sigortalılar arasında dağıtımını gerçekleştirerek, maddi kazanç sağlanmakta veya risk ile ilgili servisleri birleştirmek” biçiminde tanımlanabilmektedir.

Sigorta, aynı tür tehlikeye maruz kişilerin, sadece o tehlikeye bilfiil maruz kalanların zararının telafisinde kullanılması için belirli bir miktar para toplamak amacıyla oluşturdukları gruplaşma ile zarar veya insan yaşamındaki tehlikeli bir olayın gerçekleşmesi durumunda önceden kararlaştırılan parayı ödemeyi öngördükleri bir akittir. Ancak, para prim denen bir ücret karşılığında tehlikenin meydana gelmesinden önce toplanmaktadır.

Türk Ticaret Kanununda sigorta; “Sigorta bir akittir ki, bununla sigortacı bir prim karşılığında diğer bir kimsenin para ile ölçülebilir bir menfaatini halele uğratan bir tehlikenin meydana gelmesi halinde tazminat vermeyi yahut bir veya birkaç kimsenin hayat müddetleri sebebiyle veya hayatlarında meydana gelen belli birtakım

hadiseler dolayısıyla bir para ödemeyi veya sair edalarda bulunmayı üzerine alır” şeklinde tanımlanmaktadır (6762 sayılı Türk Ticaret Kanununun 1263. maddesi) [36].

Sigorta işletmeleri; riskleri azaltmak ve risklerin gerçekleşmesi halinde ortaya çıkan zararları belirginleştirmek amacıyla oluşturulmuş iktisadi ve sosyal kurumlardır. Sigorta işletmeleri belli bir prim karşılığında riskleri dağıtarak, etkileri azaltmaya çalışan kuruluşlar şeklinde tanımlanabilir. Bu amaca yönelik olarak sigorta işletmeleri, riskin belirli ve parayla ölçülebilir olma koşulunu aramakta ve sigortalı adına riskin yönetimi sorumluluğunu üstlenmektedir.

Sigorta tanımlamalarından yola çıkarak sigorta işletmeleri, bir amacı gerçekleştirmek üzere oluşturulan organizasyonlar olarak tanımlanır. Öncelikle sigorta hizmeti sunulması görevi, sigorta işletmeni tarafından yerine getirilmektedir. Bu işletmeler ekonomide aynı riskin tehdidi altında bulunan kişi ve kurumları bir araya getirme görevini yerine getirmekte, güven pazarlamakta ve bir organizasyon olarak risk ve belirsizliklerle dolu geleceğin güvence altına alınması gereksinimine yönelik hizmetleri gerçekleştirmektedirler.

Başka bir görüşe göre ise; "Sigorta işletmesi; belirli bir sahada başkalarının gereksinmesini karşılamak üzere sürekli olarak hizmet üretiminde bulunan ortaya koyduğu bu hizmetleri bir fiyat karşılığında (sigorta primi) başkalarına satmak suretiyle, kar sağlamak amacını güden bir kuruluş olduğu” şeklinde tanımlanmaktadır. Amacı olarak da, "çeşitli risklerin tehdidi altında bulunan kişileri ve sermaye unsurlarını sigorta teminatı altına almak suretiyle topluma hizmet etmek" olarak belirtilmektedir [37]. Farklı bir kaynak ise, sigorta işletmeciliğini; "Sigorta ile ilgi faaliyetlerin yürütülmesi amacıyla emek, sermaye, doğal kaynaklar ve girişimci gibi üretim unsurlarının bir araya getirilmesiyle oluşan birim olarak" tarif etmektedir [38].

5.1.1 Sigortacılık sektörünün tarihi gelişimi

Türkiye'de sigorta sektörünün 150 yıldan fazla bir geçmişi vardır. Önceleri yabancılar tarafından yapılan sigortacılık faaliyetlerinin Türk sigortacıları eliyle yapılması aslında 1945 yıllarından itibaren başlamıştır denilebilir.

Osmanlı İmparatorluğu döneminde sigortacılık faaliyetleri gelişme olanağı bulamamıştır. İlk sigorta şirketleri, İstanbul'daki büyük yangınların etkisiyle 1872

yılında İngiliz, 1878 yılında Fransız sigorta şirketleri açtıkları temsilciliklerle Türkiye'de ilk sigortacılık faaliyetlerini başlatmışlardır. Osmanlı Devleti'nde kurulan ilk sigorta şirketi ise, sermayesi yabancılara ait olan Osmanlı Bankası ile Tütün Rejisi ve Duyun-i Umumiye İdaresi'nin ortaklığıyla kurulan Osmanlı Umum Sigorta Şirketi'dir.

Cumhuriyeti izleyen 1924' de, 1920 yılında Fransız Urbain Şirketi'nin iştiraki ile kurulmuş olan "Milli Sigorta Şirketi" daha sonra da Avrupa Anonim Türk Sigorta Şirketi ile İtimad-i Milli Sigorta Şirketleri kurulmuştur. Sonraları da 1925' de Güneş Sigorta A.Ş., 1929 da Milli Reasürans T.A.Ş kurulmuştur. 1980' lerin ortalarından itibaren hazinenin yeni şirket oluşumuna yeşil ışık yakmasıyla birlikte sektörde şirket sayıları 2-3 misline çıkmıştır. 1988 yılı sonunda 35 olan şirket sayısı, 1997 yılı sonunda 59'a çıkmıştır.

2001 yılında kabul edilen "Bireysel Emeklilik Tasarruf ve Yatırım Sistemi Kanunu" ile kurulan Bireysel Emeklilik Sistemi'nin 27.10.2003 tarihinde faaliyete geçmesiyle sigorta işletmelerindeki tarihi gelişim süreci gelişimini sürdürmeye devam etmiştir [37].

Tüm gelişmiş finansal piyasalarda sigorta sektörü etkin bir rol almaktadır. Ancak Türkiye'de sigorta sektörünün finansal piyasalar içindeki payının 2006 yılında %3,3 olması, bu sektörün henüz istenilen seviyeye ulaşamadığının göstergesiydi. Fakat son yıllarda ekonomimizde yaşanan olumlu gelişmelere paralel olarak sigortacılık sektöründe de önemli ilerlemeler kaydedilmiştir [36]. 13 yıldır beklenen Sigortacılık Kanunu yürürlüğe girmiştir. Türk Sigorta sektörü için 2007 yılında yaşanan en önemli gelişme sektörün 13 yıldır beklediği, Haziran ayında yürürlüğe giren “Sigortacılık Kanunu” olmuştur. Bu gelişmeyle birlikte Türk sigorta sektörünün uluslararası standartlara uygunluğunun da sağlanması amaçlanmıştır. Yasa ile sigorta ve reasürans şirketlerinin faaliyetlerinin, varlıklarının, iştiraklerinin, alacaklarının, öz kaynaklarının, borçları ile mali bünyeyi ve idari yapıyı etkileyen tüm unsurlarının daha etkin şekilde denetimi sağlanacaktır. Kanun kapsamında yeni olarak bir tahkim müessesesi ve sigorta şirketlerini, acenteleri ve eksperleri temsilen birer icra komisyonu kurulmuştur.

Sigortacılık kanunu sonrasında yürürlüğe düzenlemeler konmuştur ve konmaya da devam etmektedir. Bu düzenlemeler sigorta şirketlerini sermaye yapılarını

güçlendirmeye yönlendirmektedir. Bunun sonucu olarak uzun vadede sektöre güven ve sektör karlılığı artacaktır.

5.1.2 Sigortacılığın genel işleyişi

Sigorta, temelde benzer riske maruz kalan kişilerin (sigortalılar) maddi güçlerini birleştirerek yardımlaşmasıdır. Sigorta şirketlerinin işlevi, bu kişilerin birbirlerini bulmasını ve gerçekleşen zararların giderilmesini organize etmektir.

Aynı riske maruz kalan kişiler bir araya gelerek gerçekleşen riskleri hep birlikte karşıladıklarında, kişi başına düşen ödeme miktarı düştüğünden büyük bir risk bile herkes için karşılanabilir hale gelebilir. Bir araya gelen kişi sayısı ne kadar fazla olursa, kişi başına düşen zarar miktarı o kadar tahmin edilebilir hale gelir. Olasılık hesaplarında büyük adetler kanunu (veya büyük sayılar kanunu) adı verilen kanuna göre, aynı riske konu olan kıymetlerin sayısı ne kadar fazla olursa, gerçekleşecek hasar, riskin oluşma ihtimaline o denli yakın olacaktır.

Bir sigorta şirketinin organizasyonu altında bir araya gelen sigortalılar gelecekteki hasarları karşılamak için prim adı verilen ücreti öderler. Prim tutarına, hasarlar için yapılacak ödemelerin yanı sıra, sigorta şirketinin işletme masrafları, kârı, vergiler, acente komisyonları gibi kalemler de dâhildir.

Sigorta, öngörülemez ve maddi zarar doğuracak olaylara (yani risklere) karşı yapılır. Ne zaman olacağı önceden bilinen olaylara karşı sigorta yapılmaz. Kumar, hisse senedi alım-satımı gibi spekülatif riskler ve kanuna aykırı olarak yapılan işler sigortalanamaz. Keza, kanunlardan kaynaklanan cezalara karşı sigorta olmaz.

Sigorta teminatının işlemeye başlaması için sigortalı ile sigorta şirketi arasında sigorta sözleşmesi yapılır. Poliçe, sigorta sözleşmesinin koşullarını içeren ve ispatlayan belgedir. Poliçe üzerinde sigorta şirketinin unvanı, adresi, sigortalının adı, riskle ilgili bilgiler, prim, sigorta başlangıç ve bitiş tarihleri, sigorta koşulları vs. yer alır.

Sigorta satışları bazen doğrudan sigorta şirketleri tarafından yapılmakla birlikte, genellikle sigorta aracıları tarafından yapılır. Sigorta aracıları belli bir sigorta şirketine bağlı çalışan acenteler ve sigorta şirketlerinden bağımsız çalışan sigorta brokerleridir. Aracıların başlıca görevleri, sigorta müşterisini riskler ve buna uygun çözümler hakkında bilinçlendirmek, prim ödemelerini, hasarların tazmin edilmesini

ve poliçenin yenilenmesini takip etmek, gerektiğinde hatırlatmada bulunmaktır. Prim tahsil etme yetkisi bulunan acenteler, müşterilerden aldıkları primleri sigorta şirketine öderler. Bu hizmetlerine karşılık aracılar komisyon alırlar. Komisyon genellikle prim tutarına dâhil edilir.

Sigorta edilmiş bir riskin gördüğü zarara hasar denir. Bir hasar gerçekleştiğinde, sigortalı, sigorta şirketinde ihbarda bulunur. Sigorta şirketi, hasarı kendisinden bağımsız bir eksperin incelemesini sağlar. Daha sonra eksper raporunu inceleyerek, sigorta teminatı içindeki zararı tazmin eder, sigortalıya tazminat öder.

Can sigortalarında ve bedeni hasarlarda ise tazminat, aktüer denilen uzmanların yaptığı aktüerya hesaplarının sonucuna göre ödenir.

Eğer bir hasar üçüncü bir şahsın kusuru sonucu oluşmuşsa ve sigortalının o şahıstan tazminat talep etme hakkı veya daha başka hakları doğduysa, halefiyet denilen hukuki ilke uyarınca bu haklar ödediği hasar oranında sigorta şirketine geçer. Sigorta şirketinin bu hakkını kullanarak kusurlu taraftan tazminat talep etmesine rücu adı verilir. Örneğin, A ve B araçlarının karıştığı bir kazada, B aracı kusurlu ise, A'nın sigortacısı önce A'ya tazminat öder, daha sonra, ödediği tazminatı B'den talep edebilir.

Sigorta şirketleri önce bu haklarını kusurlu kişilerle anlaşarak (sulhen) kullanmayı denerler. Sulhen giderilemeyen sorunların çözümü için mahkeme yoluna gidilir. Çağdaş sigorta şirketlerinde rücu ve hukuk işlerinde uzmanlaşmış bölümler bulunur.

Deprem gibi büyük felaketler sonucu ödenen hasarların sigorta şirketlerinin mali güçlerini aşan tutarlara ulaşması nedeniyle, sigortacılar da kendi üstlendikleri riskleri sigortalılarla. Bu yeniden sigortalama işlemine reasürans denir. Reasürans konusunda çalışan şirketlere reasürör adı verilir. Reasürörler, verdikleri teminat karşılığında sigorta şirketlerinden prim alırlar. Reasürans, uluslararası bir iştir. Bir sigorta şirketi, risklerini pek çok reasüröre devredebilir. Böylece, deprem gibi felaket boyutlarındaki riskler bile dünya çapında pek çok şirkete dağılarak ödenebilir hale gelir. Bu durumun bir yansıması olarak, dünyanın herhangi bir yerinde meydana gelen bir felaket, tüm sigorta piyasasını etkiler. Örneğin, ABD'de meydana gelen 11 Eylül Saldırıları sonucunda, Türkiye'deki sigorta primleri de (Dünya çapında olduğu gibi) arttı.

Artan rekabet sonucu, çağdaş sigorta şirketleri müşterilerine tazminat ödeyerek yardımcı olmanın dışında sayısı giderek artan bazı yardım hizmetleri, örneğin yolda kalan araçlar için çekici, bozulan musluklar için tesisatçı temini gibi ek hizmetler vermektedir.

Sigorta şirketleri çoğu ülkede (ve Türkiye'de) hayat ve hayat dışı branşlarda çalışan şirketler olarak ikiye ayrılır.

Türkiye'de sigortacılık Hazine Müsteşarlığı Sigortacılık Genel Müdürlüğü tarafından düzenlenmekte ve denetlenmektedir.

Türkiye'deki sigorta müşterileri, sigorta şirketi ile anlaşmazlığa düştüklerinde, Türkiye Sigorta ve Reasürans Şirketleri Birliği'ne [1] veya Hazine Müsteşarlığı'na başvurabilir [37].

5.2 Sigortacılık Sektöründe Veri Madenciliği Uygulaması

Veri madenciliği uygulamalarının problemin tanımlanması, verilerin hazırlanması, modelin kurulması ve değerlendirilmesi, modelin kullanılması ve izlenmesi aşamalarından oluşmakta olduğunu söylemiştik. Uygulamada bu adımlardan sırasıyla geçeceğiz.

5.2.1 Problemin tanımlanması

Sigortacılık sektörünü anlatırken bahsettiğimiz gibi sigortacılık yapılan sigorta işleminin poliçelerle takip edildiği ve poliçede belirlenen olası risklerden herhangi birinin gerçekleşmesi durumunda bu işlemlerin hasar ihbarları ile takip edilmektedir. Açılan hasar ihbarına hasar tespiti için bir veya birden fazla eksper veya aktüer ve ihbarı takip etmesi için dosya inceleme görevlisi atanmaktadır. Eksperler veya aktüerler sahte hasar olup olmadığını ve eğer ihbar edilen hasar gerçek ise durumu incelemektedir. İncelemeleri sonucunda oluşturdukları raporu, hasarın farklı açılardan çektikleri fotoğrafları ile birlikte sigorta teminatı için belirlediği tazminat bedelini sigorta şirketine bildirir. Şirket bu yapılan inceleme karşılığında eksper veya aktüere belirli bir ücret ve ödeme yapılacak şahıs veya kurumlara tazminat bedelini öder. Bunların yanında hasar ihbarının incelenmesi aşamasında anlaşmazlık çıkması ve ihbarın mahkemeye intikal etmesi gibi durumlar söz konusu olduğunda şirket ihbar için avukat masrafı gibi ek ödemeler de yapmak zorunda kalabilmektedir.

Açılan hasar ihbarının sonucunda meydana gelen hasar poliçe sınırları dâhilinde değil veya sahte hasar ihbarı söz konusu olduğunda dosyayı takip eden çalışanlara, inceleme yapan birimlere zaman kaybı, şirket tarafından çalışanlara, eksperlere veya aktüerlere ödenen ücretler göz önüne alındığında şirket için önemli kayıplara sebep olmaktadır.

Şirket ihbar geldiği aşamada ihbarın kabul edilebilir bir ihbar olup olmadığına karar verebilirse bu şirket önemli bir kazanç olacaktır.

5.2.2 Verilerin hazırlanması

Verilerin hazırlanması aşamasında uygulanması gereken adımlardan bahsetmiştik. Bu adımları uygulayarak veri madenciliği uygulama için gerekli olan düzenli verileri elde etmiş olacağız.

- **Veri toplama**

Birçok sektörde olduğu gibi sigortacılık sektörü de teknolojik gelişmelere ayak uydurmuştur. Şirketler yaptıkları poliçelerin ve hasar ihbarlarının kayıtlarını veri tabanlarında saklayarak içsel veriler oluşturulmaktadır. Bunun yanında sorgulama merkezleri ile dışsal verilerden de faydalanmaktadırlar. Örneğin 2003 yılında trafik sigortasında uygulama birliğinin sağlanması, sigorta sahtekarlıklarının önlenmesi, sigorta sistemine olan güvenin artırılması, sigortasını yaptırmamış motorlu araç işletenlerinin tespiti, sağlıklı fiyatlandırma yapılması ve benzeri amaçları gerçekleştirmek üzere, bu sigortaya ilişkin verilerin merkezi bir veritabanında toplanmasını, günlük olarak güncellenmesini ve bu verilerin belirlenen yetki düzeyine göre ilgili tarafların kullanımına sunulması için Trafik Sigortası Bilgi Merkezi (TRAMER) oluşturulmuştur. Başka bir örnek verecek olursak şahıs bilgilerine ulaşmak için kimlik veya vergi numarası ile sorgulama yapılan veri tabanları kullanılmaktadır.

Veri madenciliği için elimizde bir sigorta şirketine ait geçmiş tarihli hasar ihbar kayıtları içinden rastgele alınmış yaklaşık 2000 hasar ihbarının verileri bulunmaktadır.

- **Değer Bıçme**

Veriler şirketin içsel veri tabanlarında tutulmakla birlikte, ihbar kaydı oluşturulurken çeşitli dışsal kaynaklardan kişi veya araç bilgilerinin temini için yardım alınmaktadır.

Kayıt aşamasında veriler ortak yapıya çevrilip kaydedildiği için bizim farklı veritabanlarından temin edilen veriler içi uygulanan verilere değer biçme işlemi uygulamamıza gerek duyulmamaktadır.

- **Birleştirme ve temizleme**

Uygulama geliştirilecek olan mevcut verilerde bazı bilgilerin NULL olarak eksik bırakıldığı tespit edildi. Bu eksik verilere çeşitli yöntemlerle uygun bir değer ataması yapıldı.

Çizelge 5.1’de görülen cinsiyet, medeni durum, eğitim durumu verilerinde bulunan NULL değerlerine N sabiti atandı. Veriler incelendiğinde ihbar poliçesinin müşterisi gerçek kişi değilse yani tüzel kişi ise bu durumlarda cinsiyet, medeni durum, eğitim durumu bilgilerinin karşılıkları olmadığı için NULL olarak bırakıldığı anlaşılmıştır. Geliştirilecek model sonrası atanan N sabiti anlamlı bir sonuç ya da ayırt edici bir özellik olarak karşımıza çıkarsa modelin yorumlanması aşamasında bu bilgiye dikkat edilmesi gerekecektir.

Çizelge 5.1 : Uygulama verileri NULL değeri.

cinsiyet	medeni durum	eğitim durumu
NULL	NULL	NULL
NULL	NULL	NULL
E	E	I
E	E	L
E	B	L
NULL	NULL	NULL
E	B	L
E	B	U
NULL	NULL	NULL
K	E	I
NULL	NULL	NULL
NULL	NULL	NULL
K	E	L
E	E	L
E	E	L
NULL	NULL	NULL
NULL	NULL	NULL
E	E	I
K	E	L
E	B	Y
K	B	Y
E	B	I
NULL	NULL	NULL
NULL	NULL	NULL
E	B	U

Çizelge 5.2 : Sabit değer atanmış uygulama verileri.

cinsiyet	medeni durum	eğitim durumu
N	N	N
N	N	N
E	E	I
E	E	L
E	B	L
N	N	N
E	B	L
E	B	U
N	N	N
K	E	I
N	N	N
N	N	N
K	E	L
E	E	L
E	E	L
N	N	N
N	N	N
E	E	I
K	E	L
E	B	Y
K	B	Y

Verileri incelediğimizde poliçe teminat bedeli 0 görünen hasar ihbarları mevcut olduğu görüldü. Bu eksik veriler ortalama değeri atama işlemi ile değer atama işlemi ile değerleri hesaplandı.

Teminat bedelleri branş ve teminat türüne göre farklılıklar göstermektedir. Elimizdeki verileri incelediğimizde; Örneğin İşyeri Paket branşının Hırsızlık teminatı için teminat bedellerinin ortalama değeri 96.555 YTL iken Cam Kırılması teminatı için teminat bedellerinin ortalaması ise 3.535 YTL’dir. Bunun yanında Yangın branşının Yıldırım teminatı için teminat bedellerinin ortalama değeri 3.932.375 YTL iken Bizim Konut branşının Seylan teminatı için teminat bedelleri ortalaması 207.698YTL’dir. Bu durumda hiçbir ayırım yapmaksızın teminat bedellerinin ortalamaları alınarak elde edilecek ortalama değer gerçeğe yakın bir değer olmayacağı için verinin bütünlüğünde bozulmalara sebep olacaktır. Sadece branş ayırımı yapılması düşünüldüğünde teminat türüne göre teminat bedel ortalama değeri değişkenlik gösterdiği için elde edilecek ortalama değer sağlıklı olmayacaktır. Teminat bedeli eksik olan ihbarlara atanacak ortalama değer hesaplanırken anlamlı değerlerin elde edilmesi için bedeli 0 olan ihbarlar branşlarına ve teminatlarına göre gruplandırıldı. Oluşan grupların ortalama değerleri hesaplanarak değer ataması yapıldı.

Örneğin Çizelge 5.3'te İşyeri Paket branşının Cam Kırılması teminat değerleri verildi.

Çizelge 5.3 : Teminat bedeli 0 olan uygulama verileri.

teminat adı	branş adı	teminat tutarı
CAM KIRILMASI	İŞYERİ PAKET	0
CAM KIRILMASI	İŞYERİ PAKET	0
CAM KIRILMASI	İŞYERİ PAKET	5000
CAM KIRILMASI	İŞYERİ PAKET	0
CAM KIRILMASI	İŞYERİ PAKET	5000
CAM KIRILMASI	İŞYERİ PAKET	0
CAM KIRILMASI	İŞYERİ PAKET	0
CAM KIRILMASI	İŞYERİ PAKET	9000
CAM KIRILMASI	İŞYERİ PAKET	2000
CAM KIRILMASI	İŞYERİ PAKET	0
CAM KIRILMASI	İŞYERİ PAKET	1000
CAM KIRILMASI	İŞYERİ PAKET	2500
CAM KIRILMASI	İŞYERİ PAKET	250

Eksik verilerin hesaplaması şu aşağıdaki şekilde yapıldı:

$$5000+5000+9000+2000+1000+2500+250=24750$$

$$24750/7 \cong 3535 \text{ YTL}$$

Çizelge 5.4 : Ortalama değeri hesaplanan uygulama verileri.

teminat adı	branş adı	teminat tutarı
CAM KIRILMASI	İŞYERİ PAKET	3535
CAM KIRILMASI	İŞYERİ PAKET	3535
CAM KIRILMASI	İŞYERİ PAKET	5000
CAM KIRILMASI	İŞYERİ PAKET	3535
CAM KIRILMASI	İŞYERİ PAKET	5000
CAM KIRILMASI	İŞYERİ PAKET	3535
CAM KIRILMASI	İŞYERİ PAKET	3535
CAM KIRILMASI	İŞYERİ PAKET	9000
CAM KIRILMASI	İŞYERİ PAKET	2000
CAM KIRILMASI	İŞYERİ PAKET	3535
CAM KIRILMASI	İŞYERİ PAKET	1000
CAM KIRILMASI	İŞYERİ PAKET	2500
CAM KIRILMASI	İŞYERİ PAKET	250

- **Seçme**

Modelde kullanılacak niteliklerin ve hedef değerinin neler olduğu, veri kümesinin hangi alanlarından faydalanılacağıının belirlememiz gerekmektedir. Öncelikle kod ve ad şeklinde iki alan ile tanımlanmış verilerin ad veya kod bilgisinden biri modelde kullanmak için yeterli olacağı için diğer alanlar veri kümesinden çıkarıldı. Bunun yanında poliçe numarası, ihbar numarası gibi tekilliği sağlamak için kullanılmış olan alanlardan model kurmada anlamlı sonuçlar çıkmayacağı için bu alanlarda veri kümesinden çıkarıldı. Yine müşteri kimlik numarası ve vergi numarası gibi tekilliği sağlayan alanlar ve yine açıklama tipindeki müşteri adresi, mesleği, ekspertiz adresi alanlar veri kümesinden çıkarıldı.

Hedef değer olarak ihbar sonucu alanını belirlendi. İhbar sonucu K, I, N, D, O, M olarak kodlanmış 6 farklı durum içermektedir. Bunlardan sonuçlardan N ile ifade edileni hasar ihbarının neticesiz olduğu, D ile ifade edileni ise reddedilmiş ihbarı ifade etmektedir. Diğer dört sonuç ise ihbarın sonuçlanması durumuna kadar olan tüm ihbarın süreçlerini ifade etmektedir. Bu durumda ihbar sonucu N veya D olması olumsuz sonuç, diğer sonuçlar ise olumlu sonuç olarak tanımlandı. Uygulanacak modelin sonucunun daha kolay yorumlanabilmesi için D ve N durumlarını H, diğer durumları E sabiti ile genelleştirildi.

Hedef değer, seçilen 18 adet nitelik alanları ve bu alanların alacağı değerler Çizelge 5.5'te gösterildi.

Bu işlemler sonucunda model uygulanması için en uygun sade veriler elde edildi.

Çizelge 5.5 : Nitelik değerleri.

Nitelikler	Değerler	Açıklama
hasar il adı		
hasar sebep adı		
bölge adı		
menfaat adı		
riziko adı		
teminat adı		
branş adı		
kur kodu	TL,EUR,USD	TL:Türk Lirası - EUR:Euro - USD:Amerikan Doları
şüpheli hasar	0,1	1:ihbar şüpheli - 0:şüpheli değil
zabıt yok	0,1	1:zabıt var - 0:zabıt yok
faturalı	0,1	1:ihbar fatura ile açıldı - 0:faturasız ihbar
cam onarımı	0,1	1:cam onarımı var - 0:cam onarımı yok
eski hasar var	0,1	1:eski hasarı var - 0:eski hasarı yok
teminat tutarı		
araç yada değil	A,D	A:Araç - D: Araç değil
cinsiyet	E,K,N	E:Erkek - K:Kadın - N:Belirsiz
medeni durum	E,B,N	E:Evli - B:Bekar - N:Belirsiz
eğitim durumu	I,L,U,Y,N	I:İlkokul - L:Lise - U:Üniversite -Y:Yüksek Okul - N:Belirsiz
Hedef Değer	Değerler	Açıklama
ihbar sonucu	E,H	E: İhbar sonuçlandı - H: İhbar sonuçlanmadı

5.2.3 Modelin kurulması ve değerlendirilmesi

Uygulamamızda geçmiş verilerden yola çıkarak bir model geliştirilmesi ve bu modeli gelecek verilere uygulayarak tahmin sonucu elde etmeyi planladığımız için tahmin edici veri madenciliği modellerinden faydalanılacaktır. Nitelik değerlerinin hepsi nümerik olarak ifade edilemediği için nitelik değerleri kategorik olarak kullanılacaktır. Bu sebeple tahmin yöntemlerinden biri olan sınıflandırma modelinden yararlanılacaktır. Sınıflandırma yöntemlerinden karar ağacı yöntemi kullanıldı.

Karar ağacı yöntemi için çok sayıda yazılım bulunmaktadır. Biz QT Orange Canvas yazılımını kullanıldı.

5.2.3.1 QT Orange yazılımı hakkında genel bilgi

Orange C++ tabanlı bir yazılım olup; makine öğrenimi ve veri madenciliği ile ilgili çok fazla algoritma içermektedir. Buna ek olarak veri girdisi ve bu verinin işlenmesini sağlayan modüllere sahiptir.

Orange bileşen tabanlı bir yazılımdır. Bileşen tabanlı olması ile var olan bileşenlerin yanında kendi bileşenlerimizi de yaratıp kullanabiliriz.

Orange’ın özellikleri:

- Veri giriş/çıkış: Orange C4.5 dosyalarını okuyup ve yazabilir. Bununla birlikte bazı değişik formatları da desteklemektedir(txt, dat, arff..gibi).
- Önişleme: Altküme seçme özelliği, tahminsel görevler için yararlı tahminler üretme
- Tahminsel Modelleme: Sınıflandırma ağaçları, Naive Bayes sınıflandırıcı, mantıksal regresyon, kural tabanlı sınıflandırıcı
- Veri Tanımlama Metodu: Çeşitli görselleştirmeler, hiyerarşik kümeleme, çok boyutlu ölçekleme

Orange’ın görsel programlama arayüzüne “Orange Widgets” denir. İki bileşen arasında bağlantı oluşturarak bir bileşenden gelen veriyi diğer bileşene Orange Widgets’i kullanarak rahatlıkla geçirebiliriz. Orange’ın bu uygulmasına “Orange Canvas” adı verilir.

Widget’ler Orange üzerinde yapabileceğimiz işlemlerin kullanıcıya rahatlık sağlaması bakımından oluşturulan bileşenlerdir. Şu an yaklaşık 40 adet bileşen vardır. Programın her yeni sürümünde programın gelişmesiyle paralel olarak bileşen sayısı artmaktadır.

Eğer program ile birlikte otomatik olarak gelen bileşenler tam olarak kullanıcının gereksinimlerini karşılamıyor ise kullanıcı kendi gereksinimleri doğrultusunda kendi bileşenini yazar. Burada Orange’ın çok büyük kolaylığı vardır.

Orange Python programlama dili ile tamamen uyumlu bir yazılımdır. Python programlama dili büyük ve esnek bir dildir. Orange’ı Python ile kontrol edebilmekteyiz, böylece kendi bileşenlerimizi çok kolay bir şekilde Python ile yazabilmekte ve programa adapte edebiliriz.

Orange için Python ile yazılmış birkaç örnek:

```
import orange
data = orange.ExampleTable('voting')

classifier = orange.BayesLearner(data)
for i in range(5):
    print data[i].getclass(), 'classified as', classifier(data[i])
```

Şekil 5.1 : Orange için Python il yazılmış kod örneği.

Şekil 5.1’de gösterilen Python kodu örneğinde çalışan adımlar şunlardır:

- Veri dosyası okunur
- Naive Bayesian sınıflandırıcı oluşturulur
- Çıktı belirlenir

```
import orange, orngTest, orngStat
data = orange.ExampleTable('voting')

bayes = orange.BayesLearner()
tree = orange.TreeLearner()

results = orngTest.crossValidation([bayes,tree], data, folds=5)
print 'Classification accuracy: ', orngStat.CA(results)
print 'Brier Score: ', orngStat.BrierScore(results)
```

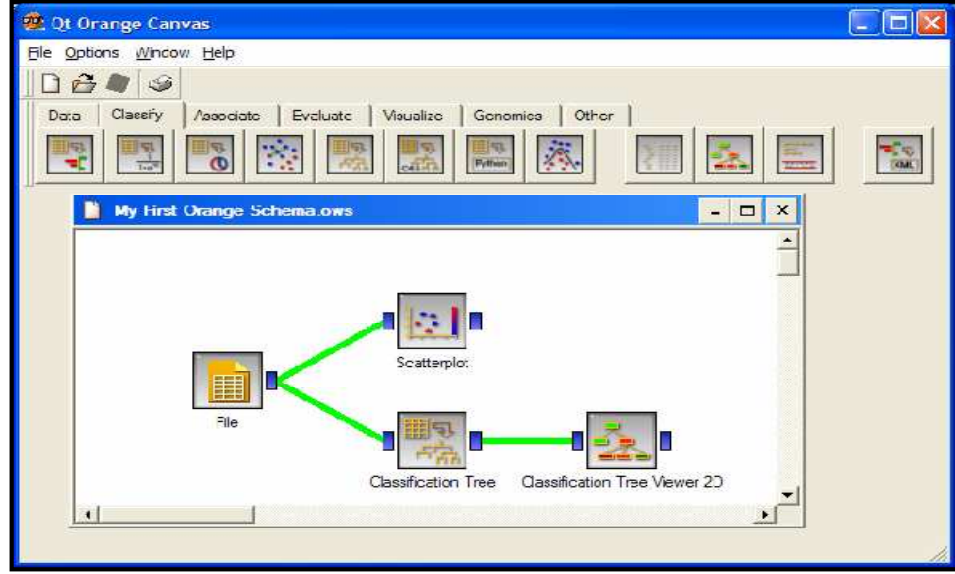
Şekil 5.2 : Orange için Python il yazılmış kod örneği.

Şekil 5.2 örneğinde çalışan adımlar şunlardır:

- İki Orange modülü oluşturulur
- Veri okunur
- Çapraz doğrulama (Cross-validation) kullanarak iki sınıflandırıcı karşılaştırılır.

Tüm bileşenler verilerini bileşen iletişimi sayesinde alırlar. İletişim ise bileşenlerin sahip olduğu kanallardan alınıp verilen veriler ile sağlanır.

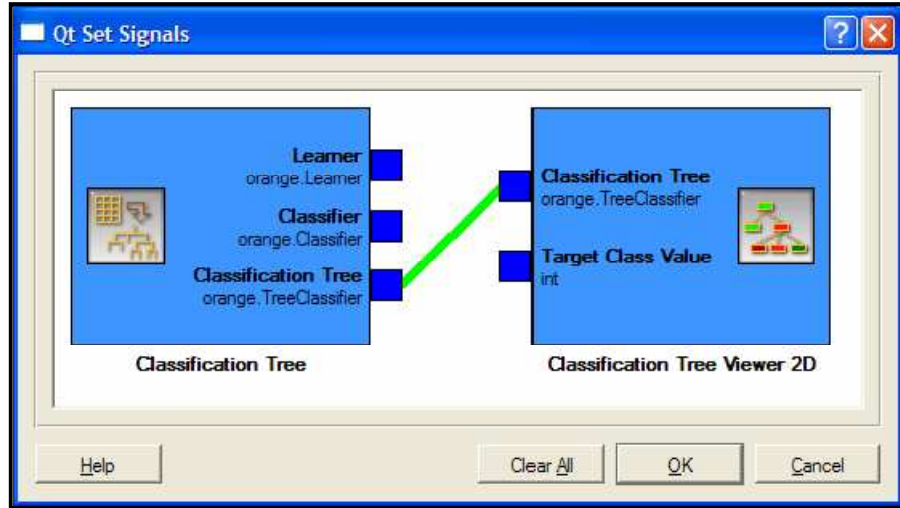
Orange Canvas; iletişim için yapılan arayüzdür. Orange Canvas’ta sürükle-bırak yöntemi ile bileşenler arası iletişim kolaylıkla sağlanır.



Şekil 5.3 : QT Orange Canvas bileşen iletişim örneği.

Yukarıdaki resimde görüldüğü üzere bileşenler Orange Canvas üzerine sürük-le-bırak ile yerleştiriliyor ve daha sonra bileşenler arasındaki bağlantı sağlanarak veri üzerinde istediğimiz işlemi yapabiliyoruz.

Tüm bileşenler iletişim kanallarına sahiptir. Her bir bileşen birden fazla kanala sahip olabilir. İki bileşen arası iletişimin olabilmesi için bileşenlerin kanalları uyuşmalıdır.



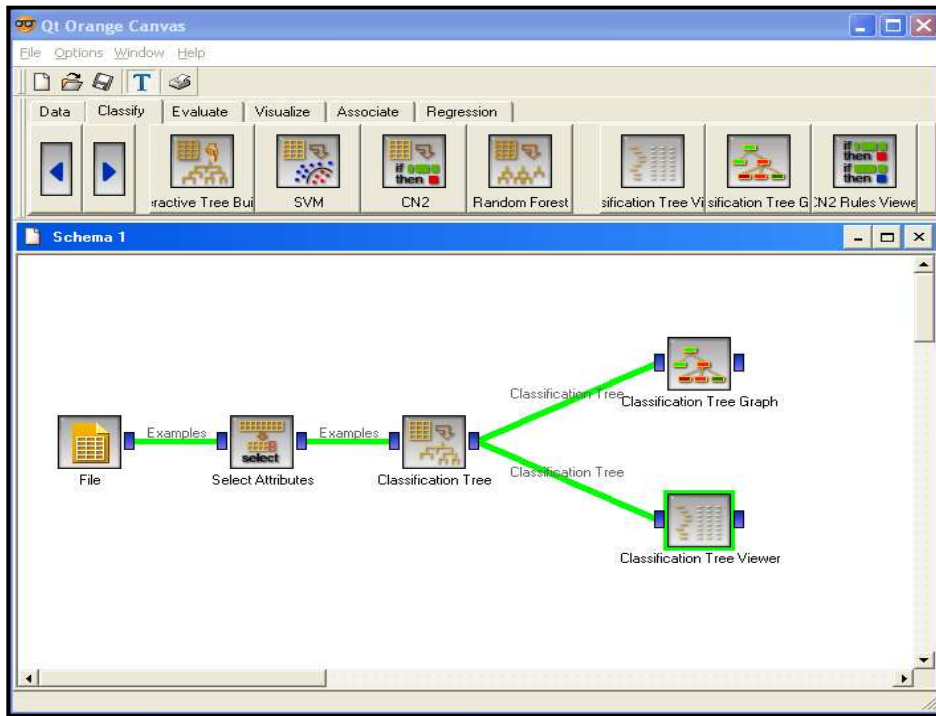
Şekil 5.4 : Bileşen iletişim kanalı örneği.

Şekil 5.4’te gösterilen örneğe göre Sınıflandırma Ağacından gelen veri 2-boyutlu Sınıflandırma ağacı göstericisine geçecektir. Her iki Widget ta birden fazla iletişim kanalına sahiptir. Burada seçilecek olan kanal, her ikisinde de aynı olan kanal olan “Classification Tree” kanalıdır [51].

5.2.3.2 QT Orange Canvas sınıflandırma ağacı uygulaması

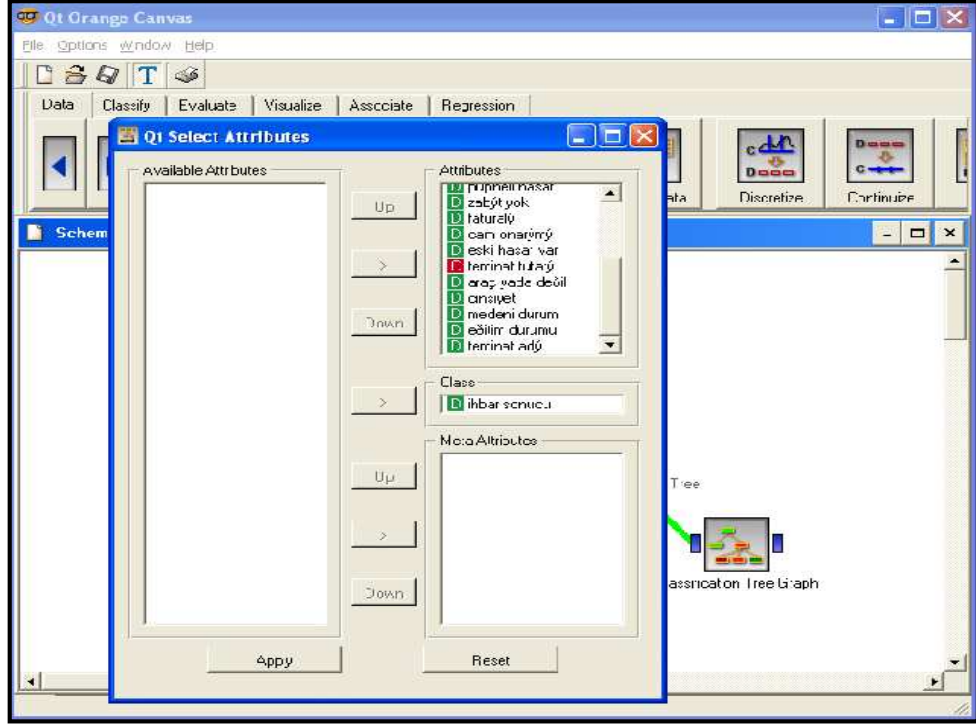
Orange programı ile sınıflandırma ağacı yapmak açılan şemaya uygun bileşenler eklenmelidir. Uygulamada verileri yüklemek için File bileşeni, nitelik ve hedef değişkenlerini belirlemek için Select Attribute bileşeni, sınıflandırma ağacını oluşturmak için Classification Tree bileşeni, oluşan ağacın grafik gösterimi için Classification Tree Graph bileşeni ve son olarak ta grafik gösteriminden daha detaylı bilgiyi hiyerarşik metin formatında gösteren Classification Tree Viewer bileşeni eklendi.

Daha öncede belirttiğimiz gibi File bileşeni belli formatta verileri desteklemektedir. Bu sebeple verileri programa uygun formata çevirmek gerekmektedir. Programa uygun hale çevrilen veriler File bileşeni tarafından seçilip programa yüklendi.



Şekil 5.5 : Uygulama bileşen seçimi.

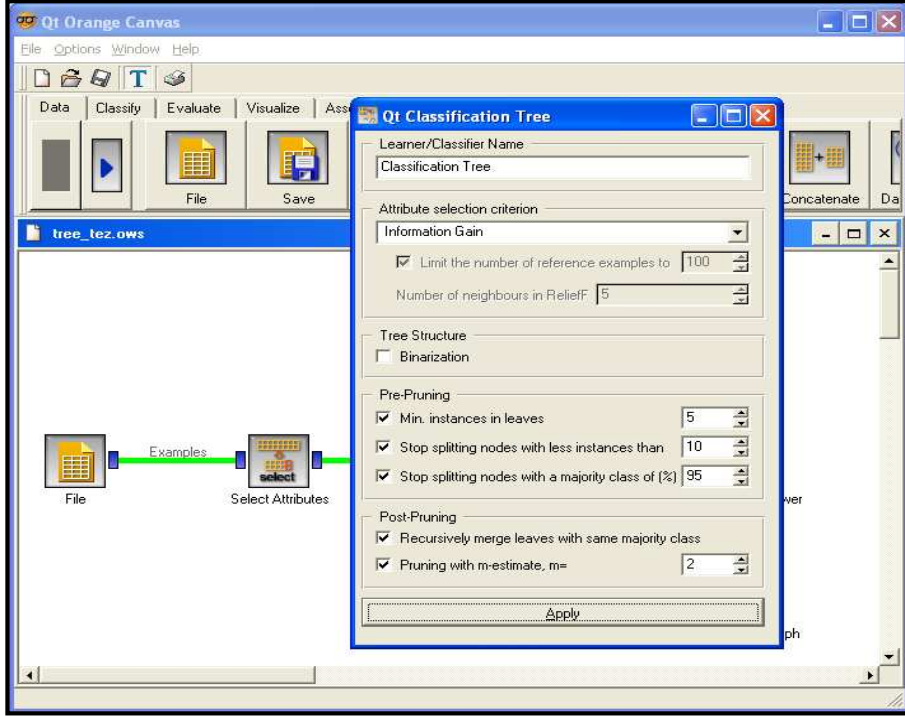
Sınıflandırma yapabilmemiz için niteliklerin ve hedef sınıfın ne olduğunu belirtmek gerektiği için Select Attribute bileşeni eklendi. Bu bileşen ile veri içerisindeki hangi alanların nitelik değeri olarak kullanılacağını ve sınıflandırmanın hangi alana göre yapılacağını tanımlanması yapılmaktadır. Şekil 5.6’da mevcut veriler için belirlenen hedef değeri ve nitelik değerleri tanıtıldı.



Şekil 5.6 : Uygulama nitelik seçimi.

Classification Tree bileşeninde nitelik seçme kriteri tanımlaması yapılmaktadır. Sınıflandırma ağacı yapmak için Orange yazılımı 4 tür karar ağacı algoritması seçeneği sunmaktadır. Bunlar: kazanç bilgisi, kazanç oranı, gini indeksi veya relif’dir. Uygulamada önceki veri madenciliği bölümünde bahsedilen kazanç bilgisi algoritmasını kullanıldı.

Bunun yanında ağacın yapısı ve ağacın budama kriterleri bu bileşende tanımlanır. Yapraklarda olması istenen en az durum sayısı, dallanmanın durdurulması istenen durum sayısı ve sınıf yoğunluğunun yüzde değeri de bu bileşende tanımlanır. Uygulamada yapraklarda en az durum sayısı 5 olarak, dallanmanın durdurulacağı en az durum sayısı 10 ve sınıf yoğunluğu %95 olarak tanımlandı.



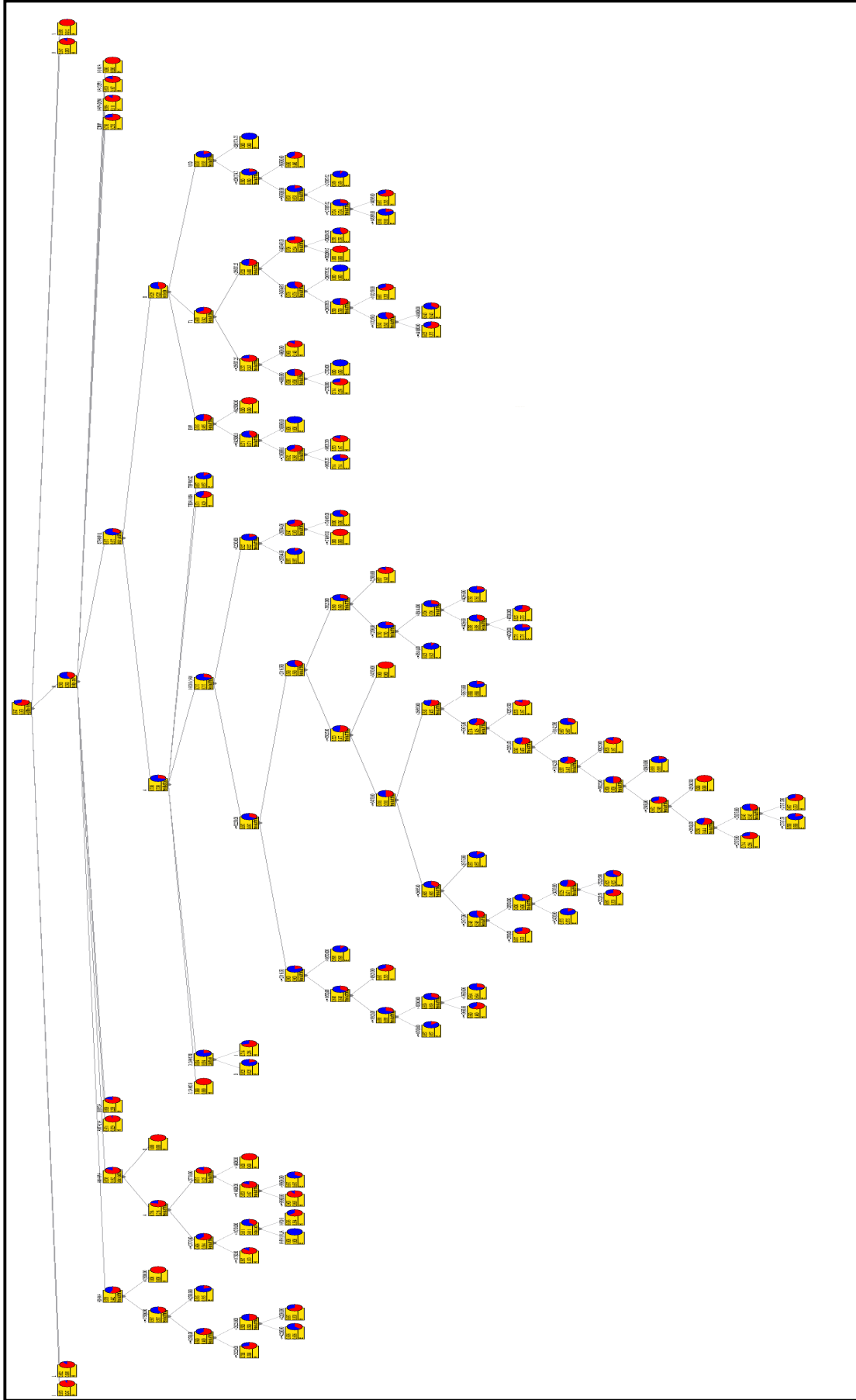
Şekil 5.7 : Ağaç kriterlerinin seçimi.

Yukarıda belirtilen tanımlamalar yapıldığında oluşan ağacın hiyerarşik metin hali Şekil 5.8'deki gibidir. Hiyerarşik metin gösteriminde oluşan sınıfın hedef değişkene göre yüzde değeri, bu sınıfa dâhil olan örnek sayısı, örneğin hedef değişkene göre dağılımı adedi ve dağılım yüzdesi olarak sonuçları gösterilir.

Uygulama sonucunda oluşan ağacın grafik gösterimi Şekil 5.9'da gösterilmiştir. Grafikte dalların ve yaprakların üzerindeki oranlar pasta dilimi ile görselleştirilmiştir. İhbarın olumlu sonuçlanması kırmızı ve olumsuz sonuçlanması ise mavi renk ile ifade edilmektedir.

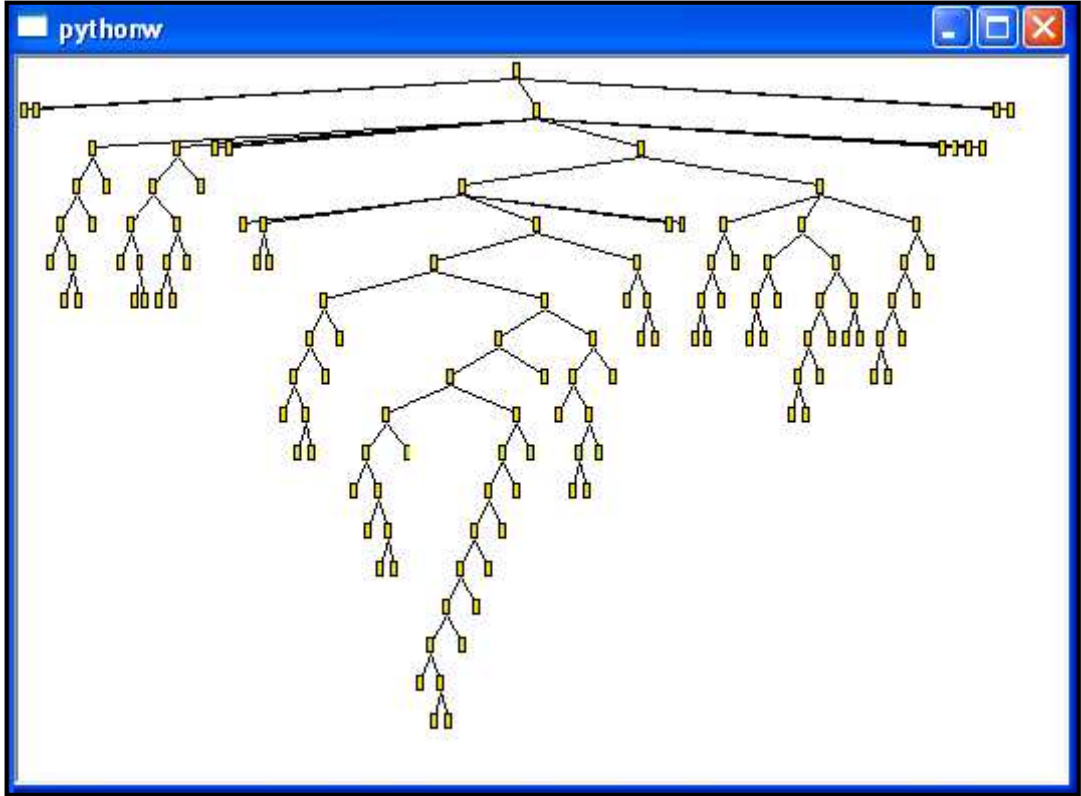
Classification Tree	Class	P(Class)	P(Target)	#Inst	Distribution (rel)	Distribution (abs)
<root>	H	0.647	0.647	2097	0.353:0.647	740:1357
egitim durumu = Y	H	0.988	0.988	82	0.012:0.988	1:81
egitim durumu = U	H	0.917	0.917	228	0.083:0.917	19:209
egitim durumu = N	E	0.560	0.440	1201	0.560:0.440	672:529
bolge adi = KONYA	H	1.000	1.000	6	0.000:1.000	0:6
bolge adi = KAYSERI	H	0.833	0.833	6	0.167:0.833	1:5
bolge adi = KARADENİZ	H	0.889	0.889	9	0.111:0.889	1:8
bolge adi = İZMİR	H	0.730	0.730	37	0.270:0.730	10:27
bolge adi = İSTANBUL	E	0.677	0.323	883	0.677:0.323	598:285
arac yada degil = D	E	0.528	0.472	231	0.528:0.472	122:109
kur kodu = USD	E	0.833	0.167	78	0.833:0.167	65:13
teminat tutari > 2081774.750	E	1.000	0.000	13	1.000:0.000	13:0
teminat tutari <= 2081774.750	E	0.800	0.200	65	0.800:0.200	52:13
teminat tutari > 865000.000	H	0.600	0.600	5	0.400:0.600	2:3
teminat tutari <= 865000.000	E	0.833	0.167	60	0.833:0.167	50:10
teminat tutari > 217067.625	E	0.939	0.061	33	0.939:0.061	31:2
teminat tutari <= 217067.625	E	0.704	0.296	27	0.704:0.296	19:8
teminat tutari > 148368.000	H	0.667	0.667	6	0.333:0.667	2:4
teminat tutari <= 148368.000	E	0.810	0.190	21	0.810:0.190	17:4
kur kodu = TL	H	0.658	0.658	120	0.342:0.658	41:79
teminat tutari > 286507.250	H	0.520	0.520	50	0.480:0.520	24:26
teminat tutari > 3405468.500	E	0.706	0.706	17	0.294:0.706	5:12
teminat tutari > 5002069.500	E	0.556	0.444	9	0.556:0.444	5:4
teminat tutari <= 5002069.500	H	1.000	1.000	8	0.000:1.000	0:8
teminat tutari > 3405468.500	E	0.576	0.424	33	0.576:0.424	19:14
teminat tutari > 2069785.625	E	1.000	0.000	5	1.000:0.000	5:0
teminat tutari <= 2069785.625	E	0.500	0.500	28	0.500:0.500	14:14
teminat tutari > 1132150.000	H	0.667	0.667	6	0.333:0.667	2:4
teminat tutari <= 1132150.000	E	0.545	0.455	22	0.545:0.455	12:10
teminat tutari > 441000.000	E	0.643	0.357	14	0.643:0.357	9:5
teminat tutari <= 441000.000	H	0.625	0.625	8	0.375:0.625	3:5
teminat tutari <= 286507.250	H	0.757	0.757	70	0.243:0.757	17:53
teminat tutari > 4000.000	H	0.860	0.860	50	0.140:0.860	7:43
teminat tutari <= 4000.000	E	0.500	0.500	20	0.500:0.500	10:10
teminat tutari > 2750.000	E	1.000	0.000	6	1.000:0.000	6:0
teminat tutari <= 2750.000	H	0.714	0.714	14	0.286:0.714	4:10
kur kodu = EUR	H	0.515	0.515	33	0.485:0.515	16:17
teminat tutari > 6425000.000	H	1.000	1.000	5	0.000:1.000	0:5
teminat tutari <= 6425000.000	E	0.571	0.429	28	0.571:0.429	16:12
teminat tutari > 2100000.000	E	1.000	0.000	9	1.000:0.000	9:0
teminat tutari <= 2100000.000	H	0.632	0.632	19	0.368:0.632	7:12
teminat tutari > 99835.359	H	0.833	0.833	12	0.167:0.833	2:10
teminat tutari <= 99835.359	E	0.714	0.286	7	0.714:0.286	5:2
arac yada degil = A	E	0.730	0.270	652	0.730:0.270	476:176
teminat adi = TEFRİKSİZ (BEDENİ VE MADDİ)	E	0.857	0.143	7	0.857:0.143	6:1
teminat adi = TEDAVİ MASRAFLARI	H	0.571	0.571	7	0.429:0.571	3:4
teminat adi = KASKO (YANMA, ÇALINMA, CARPISMA)	E	0.717	0.283	474	0.717:0.283	340:134
teminat tutari > 82230.000	E	0.827	0.173	150	0.827:0.173	124:26
teminat tutari > 285844.000	H	0.647	0.647	17	0.353:0.647	6:11
teminat tutari > 1749163.000	E	0.600	0.400	10	0.600:0.400	6:4
teminat tutari <= 1749163.000	H	1.000	1.000	7	0.000:1.000	0:7
teminat tutari <= 285844.000	E	0.887	0.113	133	0.887:0.113	118:15
teminat tutari <= 82230.000	E	0.667	0.333	324	0.667:0.333	216:108
teminat tutari > 23116.500	E	0.560	0.440	182	0.560:0.440	102:80
teminat tutari > 35625.000	E	0.690	0.310	71	0.690:0.310	49:22
teminat tutari > 72500.000	H	0.857	0.857	7	0.143:0.857	1:6
teminat tutari <= 72500.000	E	0.750	0.250	64	0.750:0.250	48:16
teminat tutari > 38414.000	E	0.706	0.294	51	0.706:0.294	36:15
teminat tutari <= 38414.000	E	0.923	0.077	13	0.923:0.077	12:1
teminat tutari <= 35625.000	H	0.523	0.523	111	0.477:0.523	53:58
teminat tutari > 34750.000	H	1.000	1.000	7	0.000:1.000	0:7
teminat tutari <= 34750.000	E	0.510	0.490	104	0.510:0.490	53:51
teminat tutari > 26685.000	H	0.545	0.545	66	0.455:0.545	30:36
teminat tutari <= 26685.000	E	0.605	0.395	38	0.605:0.395	23:15
teminat tutari <= 23116.500	E	0.803	0.197	142	0.803:0.197	114:28
teminat tutari > 19550.000	E	0.898	0.102	88	0.898:0.102	79:9
teminat tutari <= 19550.000	E	0.648	0.352	54	0.648:0.352	35:19
teminat tutari > 18925.000	H	0.667	0.667	6	0.333:0.667	2:4
teminat tutari <= 18925.000	E	0.688	0.312	48	0.688:0.312	33:15
teminat tutari > 10700.000	E	0.659	0.341	41	0.659:0.341	27:14
teminat tutari <= 10700.000	E	0.857	0.143	7	0.857:0.143	6:1
teminat adi = 3. SAHİS MADDİ ZARARLAR	E	0.804	0.196	158	0.804:0.196	127:31
zabit yok = 1	H	0.714	0.714	7	0.286:0.714	2:5
zabit yok = 0	E	0.828	0.172	151	0.828:0.172	125:26
teminat adi = 3. SAHİS BEDENİ ZARARLAR	H	1.000	1.000	6	0.000:1.000	0:6
bolge adi = BURSA	H	0.800	0.800	10	0.200:0.800	2:8
bolge adi = ANTALYA	H	0.971	0.971	34	0.029:0.971	1:33
bolge adi = ANKARA	H	0.838	0.838	130	0.162:0.838	21:109
arac yada degil = D	H	1.000	1.000	33	0.000:1.000	0:33
arac yada degil = A	H	0.784	0.784	97	0.216:0.784	21:76
teminat tutari > 23770.000	H	0.875	0.875	64	0.125:0.875	8:56
teminat tutari > 114000.000	H	1.000	1.000	16	0.000:1.000	0:16
teminat tutari <= 114000.000	H	0.833	0.833	48	0.167:0.833	8:40
teminat tutari > 98900.000	E	0.667	0.333	6	0.667:0.333	4:2
teminat tutari <= 98900.000	H	0.905	0.905	42	0.095:0.905	4:38
teminat tutari <= 23770.000	H	0.606	0.606	33	0.394:0.606	13:20
teminat tutari > 11750.000	E	0.611	0.389	18	0.611:0.389	11:7
brans adi = KASKO	H	0.636	0.636	11	0.364:0.636	4:7
brans adi = KARAYOLLARI MOTORLU ARAÇLAR ZORUNLU MALİ SORUML...	E	1.000	0.000	7	1.000:0.000	7:0
teminat tutari <= 11750.000	H	0.867	0.867	15	0.133:0.867	2:13
bolge adi = ADANA	H	0.558	0.558	86	0.442:0.558	38:48
teminat tutari > 175000.000	H	1.000	1.000	19	0.000:1.000	0:19
teminat tutari <= 175000.000	E	0.567	0.433	67	0.567:0.433	38:29
teminat tutari > 62500.000	E	0.815	0.185	27	0.815:0.185	22:5
teminat tutari <= 62500.000	H	0.600	0.600	40	0.400:0.600	16:24
teminat tutari > 29229.000	E	0.500	0.500	20	0.500:0.500	10:10
teminat tutari > 42200.000	H	0.667	0.667	9	0.333:0.667	3:6
teminat tutari <= 42200.000	E	0.636	0.364	11	0.636:0.364	7:4
teminat tutari <= 29229.000	H	0.700	0.700	20	0.300:0.700	6:14
egitim durumu = L	H	0.902	0.902	396	0.098:0.902	39:357
egitim durumu = İ	H	0.953	0.953	190	0.047:0.953	9:181

Şekil 5.8 : Ağacın hiyerarşik metinsel çıktısı.



Şekil 5.9 : Ağacın grafik gösterimi.

Programda ağacın istenilen kısmına yoğunlaşılabilmesi için dallanma ön izlemesi bulunmaktadır. Uygulamanın önizleme Şekil 5.10'daki gibidir.



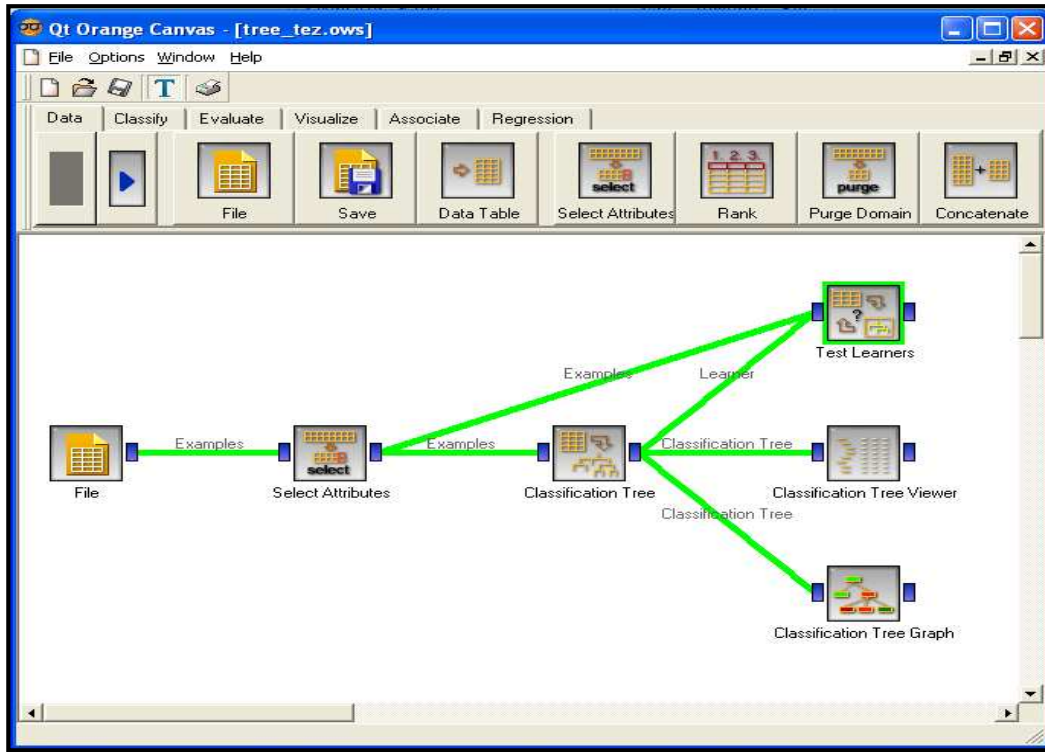
Şekil 5.10 : Ağacın önizlemesi.

Modelin değerlendirilmesinde sınıflandırma ağacının doğruluk oranını tespit etmek için çapraz doğrulama (Cross-validation) testi kullanıldı.

Bu testi yapabilmemiz için Evaluate tabındaki Test Learner bileşeni eklendi. Bu bileşen Şekil 5.11'de gösterildiği gibi Select Attributes ve Classification Tree bileşenleri ile ilişkilendirildi.

Bileşende içinde yöntem olarak cross validation seçildi. Önceki bölümlerde çapraz doğrulamada sabit sayıda kata karar verildiğinden bahsetmiştik. Genellikle bu kat sayısı olarak 10 seçilir bu sebeple bileşende kat sayısı (number of folds) 10 olarak seçildi.

10 sayıda katın kullanıldığı durumda, veri seti, öğrenme algoritması tarafından rassal olarak birbirine eşit 10 parçaya bölünür ve herbir bölüntü sırayla sınamaya, kalan 9 bölüntü ise eğitim için kullanılır. Ulaşılan sonuçların ortalama değeri alınarak toplam sınıflandırma sonucu elde edilir.



Şekil 5.11 : Çapraz doğrulama testi için bileşen seçimi.

Uygulamada hesaplanması istenen performans değerleri seçildi. Bu değerler sınıflandırma doğruluk derecesi, duyarlılık derecesi, özgüllük derecesi ve ROC (Receiver Operating Curve) eğrisi altında kalan alan değerleri seçildi.

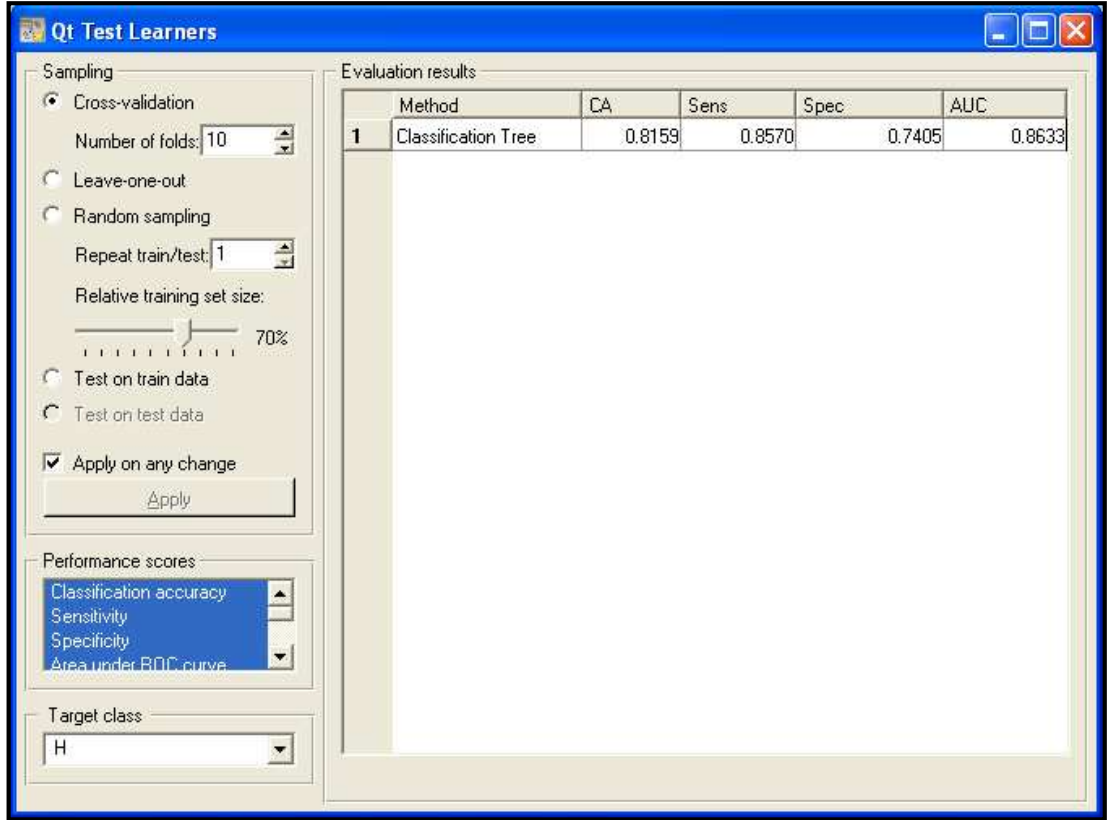
Bu değerlerin anlamı:

Sınıflandırma Doğruluk Oranı (*Classification Accuracy*): Doğru olarak sınıflanan örneklerin oranını gösterir.

Duyarlılık (*Sensitivity*): Doğru pozitif oranı (*True Positive Rate-TPR*) olarak ta adlandırılır. Gerçekleşen pozitif örneklerin tüm pozitif örnekler içindeki oranını gösterir. Testin hedeflenen durumu saptama kabiliyeti olarak ta tanımlanabilir.

Özgüllük (*Specificity*): Doğru negatif oranı (*True Negative Rate-TNR*) olarak ta adlandırılır. Gerçekleşen negatif örneklerin tüm negatif örnekler içindeki oranını gösterir [51]. Testin negatif durumu tanımlama kabiliyeti olarak ta tanımlanabilir.

ROC altında kalan alan (*Area Under ROC Curve- AUC*): ROC (*Receiver-operating curve*) özgüllük değerine karşı duyarlılık değerinin eğrisidir olan eğrinin altında kalan alanın değerini gösterir [46]. Alanın değeri 1 ile 0.5 aralığında değişir. Değerin 1 olması en iyi durumdur [51].



Şekil 5.12 : Çapraz doğrulama testi sonucu.

Bunun sonucunda Şekil 5.12’de gösterildiği gibi modelin doğruluk oranı %81.59, hata oranı ise %18.41 olarak hesaplandı. Bunun anlamı öğrenme seti dışında incelenen yeni bir örnek eğer olumsuz olarak sonuçlanan bir yapraktaki koşulları sağlıyorsa %81.59 oranında olumsuz, %18.41 oranında olumlu olarak sınıflandırılması beklenir. Duyarlılık oranı %85.70, özgüllük oranı %74.05 ve AUC değeri ise %86.33 olarak hesaplandı.

5.2.4 Modelin kullanılması

Bu bölümde model sonucunda elde edilenleri doğrudan bir uygulama olarak veya başka bir uygulamanın alt uygulaması olarak kullanabileceğimiz belirtilmişti. Uygulamamızda oluşan ağacı doğrudan uygulama olarak kullanıldı ve karar ağacı ile risk analizi yapıldı.

Oluşan ağacın yapraklarındaki ihbarın olumsuz sonuçlanması durumlarının <eğer-sonra> koşulları aşağıda kurallar halinde çıkarıldı ve hangi koşullar sonucunda hangi yüzde oranında hasar ihbarının olumsuz olarak sonuçlanmasının risk olarak değerlendirilebileceğini incelendi. Bu çıkarımlardan yola çıkarak yeni bir hasar

ihbarı sırasında ihbarın olumlu veya olumsuz sonuçlanması olasılık değerlerine ulaşılabilirliği değerlendirildi.

Kural 1:

IF eğitim durumu = I

THEN ihbar sonucu = H

190 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 181 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %95.3'dür.

Kural 2:

IF eğitim durumu = U

THEN ihbar sonucu = H

228 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 209 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %91.7'dir.

Kural 3:

IF eğitim durumu = L

THEN ihbar sonucu = H

396 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 357 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %90.2'dir.

Kural 4:

IF eğitim durumu = Y

THEN ihbar sonucu = H

82 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 81 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %98.8'dir.

Kural 5:

IF egitim durumu = N AND

bolge adi = IZMIR

THEN ihbar sonucu = H

37 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 27 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %73'dür.

Kural 6:

IF egitim durumu = N AND

bolge adi = KARADENIZ

THEN ihbar sonucu = H

9 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 8 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %88.9'dur.

Kural 7:

IF egitim durumu = N AND

bolge adi = KAYSERI

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 5 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %83.3'dür.

Kural 8:

IF egitim durumu = N AND

bolge adi = BURSA

THEN ihbar sonucu = H

10 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 8 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %80'dir.

Kural 9:

IF eğitim durumu = N AND

bolge adi = KONYA

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 10:

IF eğitim durumu = N AND

bolge adi = ANTALYA

THEN ihbar sonucu = H

34 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 33 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %97.1'dir.

Kural 11:

IF eğitim durumu = N AND

bolge adi = ADANA AND

teminat tutari <=29229.000

THEN ihbar sonucu = H

20 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve bu ihbarlardan 14 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %70'dir.

Kural 12:

IF eğitim durumu = N AND

bolge adi = ADANA AND

teminat tutari <=62500.000 AND

teminat tutari >42200.000

THEN ihbar sonucu = H

9 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve bu ihbarlardan 6 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 13:

IF egitim durumu = N AND

bolge adi = ADANA AND

teminat tutari >175000.000

THEN ihbar sonucu = H

19 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 14:

IF egitim durumu = N AND

bolge adi = ANKARA AND

arac yada degil = A AND

teminat tutari <=11750.000

THEN ihbar sonucu = H

15 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve bu ihbarlardan 13 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %86.7'dir.

Kural 15:

IF egitim durumu = N AND

bolge adi = ANKARA AND

arac yada degil = A AND

teminat tutari >23770.000 AND

teminat tutari <=98900.000

THEN ihbar sonucu = H

42 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve bu ihbarlardan 38 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %90.5'dir.

Kural 16:

IF eğitim durumu = N AND

bolge adi = ANKARA AND

arac yada degil = A AND

teminat tutari >114000.000

THEN ihbar sonucu = H

16 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 17:

IF eğitim durumu = N AND

bolge adi = ANKARA AND

arac yada degil = A AND

teminat tutari <=23770.000 AND

teminat tutari >11750.000 AND

brans adi = KASKO

THEN ihbar sonucu = H

11 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve bu ihbarlardan 7 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %63.6'dır.

Kural 18:

IF eğitim durumu = N AND

bolge adi = ANKARA AND

arac ya da degil = D

THEN ihbar sonucu = H

33 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 19:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari >10700.000 AND

teminat tutari <=13650.000

THEN ihbar sonucu = H

5 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 3 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %60'dır.

Kural 20:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=19550.000 AND

teminat tutari >18925.000

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 4 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 21:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari >23116.500 AND

teminat tutari <=23850.000

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 4 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 22:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari >24535.000 AND

teminat tutari <=25220.500

THEN ihbar sonucu = H

9 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 6 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 23:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari >26685.000 AND

teminat tutari <=27075.000

THEN ihbar sonucu = H

7 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 5 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %71.4'dür.

Kural 24:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=28100.500 AND

teminat tutari >27517.500

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 4 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 25:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=28450.000 AND

teminat tutari >28100.500

THEN ihbar sonucu = H

5 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dir.

Kural 26:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=31142.500 AND

teminat tutari >30025.000

THEN ihbar sonucu = H

12 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 10 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %83.3'dür.

Kural 27:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=33675.000 AND

teminat tutari >32851.500

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 5 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %83.3'dür.

Kural 28:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=35625.000 AND

teminat tutari >34750.000

THEN ihbar sonucu = H

7 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 29:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=44266.000 AND

teminat tutari >40700.000

THEN ihbar sonucu = H

8 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 5 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %62.5'dir.

Kural 30:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=82230.000 AND

teminat tutari >72500.000

THEN ihbar sonucu = H

7 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 6 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %85.7'dir.

Kural 31:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari >285844.000 AND

teminat tutari <=1749163.000

THEN ihbar sonucu = H

7 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 32:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = 3. SAHIS MADDI ZARARLAR AND

zabit yok = 1

THEN ihbar sonucu = H

7 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 5 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %71.4'dür.

Kural 33:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac yada degil = A AND

teminat adi = 3. SAHIS BEDENI ZARARLAR

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 34:

IF egitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = TEDAVI MASRAFLARI

THEN ihbar sonucu = H

7 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 4 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %57.1'dir.

Kural 35:

IF egitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = EUR AND

teminat tutari <=2100000.000 AND

teminat tutari >99835.359

THEN ihbar sonucu = H

12 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 10 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %83.3'dir.

Kural 36:

IF egitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = EUR AND

teminat tutari >6425000.000

THEN ihbar sonucu = H

5 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Kural 37:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = USD AND

teminat tutari <=217067.625 AND

teminat tutari >148368.000

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 4 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 38:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = USD AND

teminat tutari <=2081774.750 AND

teminat tutari >865000.000

THEN ihbar sonucu = H

5 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 3 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %60'dır.

Kural 39:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = TL AND

teminat tutari <=2750.000

THEN ihbar sonucu = H

14 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 10 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %71.4'dür.

Kural 40:

IF egitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = TL AND

teminat tutari <=286507.250 AND

teminat tutari >4000.000

THEN ihbar sonucu = H

50 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 43 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %86'dır.

Kural 41:

IF egitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = TL AND

teminat tutari >286507.250 AND

teminat tutari <=441000.000

THEN ihbar sonucu = H

8 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 5 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %62.5'dir.

Kural 42:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = TL AND

teminat tutari <=2069785.625 AND

teminat tutari >1132150.000

THEN ihbar sonucu = H

6 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerden 4 tanesi olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %66.7'dir.

Kural 43:

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = D AND

kur kodu = TL AND

teminat tutari >3405468.500 AND

teminat tutari <=5002069.500

THEN ihbar sonucu = H

8 hasar ihbar örneği aşağıdaki koşulu sağlıyor ve örneklerin tümü olumsuz sonuçlanmış. Bu durumda bu koşulları sağlayan ihbarların olumsuz sonuçlanma ihtimali %100'dür.

Ağaç sonucunda oluşan kurallar risk maddesi olup olmayacağının değerlendirildi. Kural 1, kural 2, kural 3, kural 4 incelendiğinde eğitim durumu I, L, U, Y olan ihbarlar yüksek oranda olumsuz sonuçlanmış görüldü. Bu noktada dikkat edilmesi

gereken nokta dallanmalar N değeri üzerinden devam ediyor olmasıdır. Daha öncede belirtildiği gibi eğitim durumunun tanımsız yani N olması durumundaki müşterilerin tüzel kişilik olmalarıdır. Bu yapraklardan çıkan kurala göre gelen özel kişi ihbarlarının hepsinin risk olarak değerlendirilmesi gerekmektedir. Bu durum yeterince anlamlı olmadığı için risk maddesi olarak değerlendirmeye alınamadı.

Bölge bazında oluşan yaprakların kuralları incelendi. Kural 5, kural 6, kural 7, kural 8, kural 9 ve kural 10 göre Antalya, İzmir, Karadeniz, Kayseri, Bursa, Konya bölgelerinde dallanma olmadığı ihbarların yüksek oranda olumsuz sonuçlandığı görüldü. Fakat tek başına bölge bilgisi ihbarın risk maddesi değerlendirmesi yapmak için yeterli bulunmadığı için değerlendirmeye alınamadı.

Diğer bölgeler sırasıyla incelendiğinde Adana bölgenin olumsuz sonuçlanan ihbarları teminat tutarlarına göre değişkenlik gösterdiği gözlemlendi. Teminat tutarının 175000YTL'den büyük veya 42200YTL-62500YTL aralığında veya 29229YTL'den küçük olması durumunda olumsuz sonuçlanabilmektedir. Bu durumlar risk olarak değerlendirilebilir. Risklerin yüzde değerleri aşağıda gösterildiği gibidir.

Risk 1: kural 11 - %70

Risk 2: kural 12 - %66.7

Risk 3: kural 13 - %100

Ankara bölgesi incelendiğinde olumsuz sonuçlanan hasar ihbarlarının araç hasarları olduğu durumda ve belirli teminat tutarları aralığında dallanma meydana geldiği ve araç dışındaki ihbarların tamamı olumsuz sonuçlandığı gözlemlendi. Bu durumda Ankara bölgenin tüzel kişilere ait ve teminat tutarı 11750 YTL'den küçük veya 23770YTL-98900YTL aralığında veya 114000YTL'den büyük olan araç hasar ihbarlarının olumsuz sonuçlanabilmektedir ve dikkat edilmesi gereken, risk teşkil edebilecek ihbarlardır. Kuralların yüzde değerleri aşağıdaki gibidir.

Risk 4: kural 14 - %86.7

Risk 5: kural 15 - %90.5

Risk 6: kural 16 - %100

Bunun yanında bir diğer durumda Ankara bölgenin tüzel kişilere ait Kasko branşındaki araç ihbarlarında teminat tutarı aşağıdaki 11750YTL-23770YTL

aralığında olması durumundaki ihbarlar olumsuz olarak sonuçlanabilmektedir. Bu durum da risk maddesi olarak değerlendirmeye alındı.

Risk 7: kural 17 - %63.6

Bir diğer yaprak sonucu oluşan kural 18 incelendiğinde, Ankara bölgesindeki tüzel kişilere ait olan araç dışı ihbarların hepsinin risk olarak değerlendirilmesi sonucu çıkmaktadır. Bu risk maddesi anlamlı olmayacağı için bu kural dikkate alınmadı.

İstanbul bölgesi incelendiğinde Kasko(Yanma, Çalınma, Çarpışma) teminatında, tüzel kişiliklerin poliçeleri üzerinden açılan araç hasar ihbarlarında teminat tutarı değeri 10700YTL-13650YTL, 18925YTL-19950YTL, 23116.5YTL-23850YTL, 24535YTL-25220.5YTL, 26685YTL-27075YTL, 27517.5YTL-28100.5YTL, 28100.5YTL-28450YTL, 30025YTL-31142.5YTL, 32851.5YTL-33675YTL, 34750YTL-35625YTL, 40700YTL-44266YTL, 72500YTL-82230YTL, 174163YTL-285844YTL aralıklarında ise ihbarların olumsuz sonuçlanmaları söz konusudur. Olumsuz sonuçlanma yüzdeleri aşağıdaki gibidir. Bu durumlar risk maddeleri olarak ele alınmalıdır.

Risk 8: kural 19 - %60

Risk 9: kural 20 - %66.7

Risk 10: kural 21 - %66.7

Risk 11: kural 22 - %66.7

Risk 12: kural 23 - %71.4

Risk 13: kural 24 - %66.7

Risk 14: kural 25 - %100

Risk 15: kural 26 - %83.3

Risk 16: kural 27 - %83.3

Risk 17: kural 28 - %100

Risk 18: kural 29 - %62.5

Risk 19: kural 30 - %85.7

Risk 20: kural 31 - %100

Yine İstanbul bölgesinde 3. Şahıs Maddi Zararlar teminatında, tüzel kişiliklerin poliçeleri üzerinden açılan zabıt bilgisi olan araç hasar ihbarlarının olumsuz sonuçlanması söz konusudur. Bu kural da risk maddesi olarak değerlendirilmeye alındı.

Risk 21: kural 32 - %71.4

Bunun yanında uygulama verilerin İstanbul bölgesinde 3. Şahıs Bedeni Zararlar teminatında tüzel kişiliklerin poliçeleri üzerinden açılan araca ait hasar ihbarlarının tamamında olumsuz sonuçlanmalar söz konusudur. Bu kural da risk maddesi olarak değerlendirmeye alındı.

Risk 22: kural 33 - %100

Kural 34'e göre İstanbul bölgesinde tüzel kişiliklerin araç hasarları Tedavi Masrafları teminatında ise olumsuz sonuçlanmalar söz konusu fakat yüzde olarak düşük olduğu için dikkate alınmadılar.

İstanbul bölgesinde tüzel kişiliklerin araç dışı ihbarlarında poliçenin kur kodunun EUR olması durumunda açılan ihbarların teminat tutarı 99835.395YTL-2100000YTL aralığında ise veya 6425000YTL'den büyük ise olumsuz sonuçlanabilmektedir. Bu koşullar risk maddesi olarak değerlendirildi.

Risk 21: kural 35 - %83.3

Risk 22: kural 36 - %100

İstanbul bölgesinde tüzel kişiliklerin araç dışı ihbarlarında poliçenin kur kodunun USD olması durumunda teminat tutarı 865000YTL-2081774.75YTL aralığında ise veya 148368YTL-217067.625YTL aralığında ise ihbarlar olumsuz sonuçlanabilmektedir. Bu koşullar risk maddesi olarak değerlendirildi.

Risk 23: kural 37 - %66.7

Risk 24: kural 38 - %60

İstanbul bölgesinde tüzel kişiliklerin araç dışı ihbarlarında poliçenin kur kodunun TL olması durumunda teminat tutarı 0YTL-2750YTL, 4000YTL-286507.25YTL, 286507.25YTL-441000YTL, 1132150YTL-2069785.625YTL, 340468.5YTL-5002069.5YTL aralık değerleri içerisinde ise ihbarlar olumsuz sonuçlanabilmektedir

ve bu koşullar da birer risk maddesidir. Kuralların olumsuz sonuçlanma yüzdeleri aşağıdaki gibidir.

Risk 25: kural 39 - %71.4

Risk 26: kural 40 - %86

Risk 27: kural 41 - %62.5

Risk 28: kural 42 - %66.7

Risk 29: kural 43 - %100

Ek 1’de 5 tane yeni hasar ihbarından oluşan test verileri verilmiştir. Oluşan risk maddeleri ile yeni hasar ihbarların sonuçları tahmin edilmeye çalışılacaktır:

İhbar 1: İstanbul bölgesinde ve ilinde, branşı Kasko, teminatı Kasko (Yanma, Çalınma, Çarpışma), menfaati Araç+Aksesuar ve teminat tutarı 30200YTL olan hususi tüzel kişiliğe ait aracın ve TL poliçesinde çarpışma nedeni ile meydana geldiği hasar ihbarı alınmıştır. İhbarın zabıt bilgisi vardır. Şüpheli bir hasar değildir. Cam onarımı yoktur ve farklı sigorta şirketinden fatura talebi ile açılmamıştır.

İhbar 1 değerlendirildiğinde risk 15’e dâhil olmaktadır ve olumsuz sonuçlanma yüzdesi % 83.3’dür. Riskin kuralı aşağıdaki gibidir.

IF eğitim durumu = N AND

bolge adi = ISTANBUL AND

arac ya da degil = A AND

teminat adi = KASKO(YANMA,CALINMA,CARPISMA) AND

teminat tutari <=31142.500 AND

teminat tutari >30025.000

THEN ihbar sonucu = H

İhbar 2: İstanbul bölgesinde ve ilinde, branşı İşveren Mali Mesuliyet, teminatı İşveren Mali Mesuliyet, menfaati 3. Şahıs ve teminat tutarı 500000YTL olan tüzel kişiliğe ait USD kurunda işletme poliçesinde tedavi nedeni ile meydana gelen hasar ihbarı alınmıştır. İhbarın zabıt bilgisi yoktur. Şüpheli bir hasar değildir. Cam onarımı yoktur ve farklı sigorta şirketinden fatura talebi ile açılmamıştır.

İhbar 2 değerlendirildiğinde hiçbir risk grubuna dâhil olmadığı gözlemlenmiştir. Bu durumda ihbar kabul edilebilir olarak değerlendirilebilir.

İhbar 3: Ankara bölgesinde ve Düzce ilinde, branşı Kasko, teminatı Kasko (Yanma, Çalınma, Çarpışma), menfaati Araç+Aksesuar ve teminat tutarı 20000YTL olan tüzel kişiliğe ait yurtiçi römork ve TL poliçesinde araç çarpması nedeni ile meydana gelen hasar ihbarı alınmıştır. İhbarın zabıt bilgisi yoktur. Şüpheli bir hasar değildir. Cam onarımı yoktur ve farklı sigorta şirketinden fatura talebi ile açılmamıştır.

İhbar 3 değerlendirildiğinde risk 5'e dâhil olmaktadır ve olumsuz sonuçlanma yüzdesi % 90.5'dir. Riskin kuralı aşağıdaki gibidir.

IF eğitim durumu = N AND

bolge adi = ANKARA AND

arac yada degil = A AND

teminat tutari >23770.000 AND

teminat tutari <=98900.000

THEN ihbar sonucu = H

İhbar 4: Karadeniz bölgesinde ve Trabzon ilinde, branşı Kasko, teminatı Kasko (Yanma, Çalınma, Çarpışma), menfaati Araç ve teminat tutarı 22400YTL olan evli bir erkeğe ait hususi araçta ve TL poliçesinde araç çarpması nedeni ile hasar ihbarı alınmıştır. İhbarın zabıt bilgisi vardır. Şüpheli bir hasar değildir. Cam onarımı yoktur ve farklı sigorta şirketinden fatura talebi ile açılmamıştır.

İhbar 4 değerlendirildiğinde hiçbir risk grubuna dâhil olmadığı gözlemlenmiştir. Bu durumda ihbar kabul edilebilir olarak değerlendirilebilir.

İhbar 5: İstanbul bölgesinde ve ilinde, branşı Abonman, teminatı All Risk, menfaati Emtea ve teminat tutarı 144226.5YTL olan tüzel kişiliğe ait yurtdışı kamyon ve EUR poliçesinde kırılma nedeni ile meydana gelen hasar ihbarı alınmıştır. İhbarın zabıt bilgisi vardır. Şüpheli bir hasar değildir. Cam onarımı yoktur ve farklı sigorta şirketinden fatura talebi ile açılmamıştır.

İhbar 5 değerlendirildiğinde risk 21'e dâhil olmaktadır ve olumsuz sonuçlanma yüzdesi % 83.3'dür. Riskin kuralı aşağıdaki gibidir.

IF egitim durumu = N AND
bolge adi = ISTANBUL AND
arac ya da degil = D AND
kur kodu = EUR AND
teminat tutari <=2100000.000 AND
teminat tutari >99835.359
THEN ihbar sonucu = H

6. SONUÇ VE ÖNERİLER

Veri hacmi arttıkça bunların islenmesi ve istenilen bilgilere ulaşmak oldukça güçleşecektir. Veri madenciliği burada devreye girer; Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların aranmasıdır. Yani, veri madenciliği büyük miktarlardaki verinin içinden anlamlı ve yararlı ilişki ve kuralların bilgisayar programları aracılığıyla aranması ve analizidir. Veri madenciliğinde ne kadar çok veri kullanılırsa, sistemin tüm sürecini kapsayacak veri kümesi ile daha geçerli kurallar çıkarılabilir.

Veri madenciliği yaklaşımları sağlık, bankacılık, sigortacılık gibi birçok sektörde yaygın olarak kullanılmaktadır. Sigortacılık sektöründe veri madenciliği riskli müşterilerin, sigorta dolandırıcılıklarının veya müşteri poliçe taleplerinin tespit edilebilmesi gibi karar destek sistemi tasarımı ve gerçekleştiriminde kullanılmaktadır. Tasarımda veri madenciliği sınıflandırma yöntemlerinden karar ağacı modeli kullanılmıştır.

Veri Madenciliği tekniklerinden biri olan sınıflandırma tekniği, verinin önceden belirlenen çıktıları uygun olarak ayrıştırılmasını sağlayan bir tekniktir. Çıktılar, önceden bilindiği için sınıflama, veri kümesini denetimli olarak öğrenir. Temel sınıflandırma tekniklerinden biri olan karar ağaçları da; verileri belli nitelik değerlerine göre sınıflandırmaya yarar. Hazırlanan veri paketleri Orange'daki sınıflandırma ağacı ile analiz edilmiştir.

Bu tezde sigortacılık sektöründe hasar ihbarlarının neticesiz veya reddedildi olarak sonuçlandırması için veri madenciliği ile karar destek sistemi tasarımı ve gerçekleştirimi yapılmıştır. Uygulama sonucunda oluşan karar ağacının kuralları incelenip risk maddesi olarak değerlendirip değerlendirilmeyeceğine karar verilmiştir. Bu inceleme sonunda 29 risk maddesi tanımlanmıştır. Oluşan risk maddeleri ile yeni hasar ihbarlarının sonucu tahmin edilmeye çalışılarak sistemin uygulanabilirliği araştırılmıştır.

Oluřan risk maddeleri üzerinde genelleme yapmak gerekirse İstanbul bölgesinde tzel kiřiliklerin ara ihbarlarının olumsuz sonulanmaları tahmini olduka yksek ıkmıřtır. alıřma sonucunda sigortacılık sektrnde řirket stratejilerinin belirlenmesinde gelecek tahmini iřlemlerinde veri madencilięi ile elde edilen karar destek sistemleri etkin olarak kullanılabileceęi grlmřtır.

Geliřtirilen karar destek sistemi zaman ile deęiřim gstereceęi iin belirli dnemlerde yeni veri kmeleri ile uygulama geliřtirilebilir. Bylece zaman ile deęiřkenlik riskler tespit edilip mdahale iin stratejiler belirlenebilir.

KAYNAKLAR

- [1] Shmueli, G., Patel, N.R., Bruce, P.C., 2007. Data Mining for Business Intelligence - John Wiley and Sons, Ltd., Hoboken, NJ, USA.
- [2] Özkan, Y., 2008. Veri Madenciliği Yöntemleri - Papatya Yayıncılık Eğitim A.Ş., İstanbul, Türkiye.
- [3] Silahtaroglu, G., 2008. Kavram ve Algoritmalarıyla Temel Veri Madenciliği - Papatya Yayıncılık Eğitim A.Ş., İstanbul, Türkiye.
- [4] Alpaydın, E., 2004. Introduction to Machine Learning, MIT Press, Cambridge, MA.
- [5] Larose, D.T., 2005. Discovering Knowledge In Data - John Wiley and Sons, Ltd., Hoboken, NJ, USA.
- [6] Berry, M.J.A., Linoff, G.S., 2004. Data Mining Techniques: For Marketing Sales and Customer Relationship Management, Wiley Publishing, Inc., Indianapolis, USA.
- [7] Witten, I.H., Frank, E., 2005. Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, California, USA.
- [8] Weiss, S. I., Kulikowski, C., 1991. Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems. San Francisco, Calif.: Morgan Kaufmann.
- [9] Larose, D. T., 2005. Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley and Sons, Ltd., Hoboken, N.J., USA.
- [10] Doğan Ş., 2007. Veri Madenciliği Kullanarak Biyokimya Verilerinden Hastalık Teşhisi, *Yüksek Lisans Tezi*, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ. (ornek_tez3)
- [11] Aydoğan F., 2003. E-ticarette Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirimi, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul. (ornek_tez2)
- [12] Alpaydın, E., 2000. Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri, Bilişim 2000 Eğitim Semineri.
- [13] Uskurlu Ü.B., 2006. Veri Tabanı, Veri Madenciliği, Veri Ambarı, Veri Pazarı, *Yüksek Lisans Tezi*, Başkent Üniversitesi, Bilgisayar Mühendisliği, Yönetim Bilişim Sistemleri, Ankara.
- [14] Vose D., 2008. Risk Analysis: A Quantitative Guide, 3.d Edition, John Wiley and Sons, Ltd, New York.

- [15] **Tevfik, T. A.**, 1997. Risk Analizine Giriş, Alfa Basım Yayım Dağıtım A.Ş., Yayın No : 337, İstanbul.
- [16] **Peltier, T.R.**, 2005. Information Security Risk Analysis, Taylor & Francis Group, USA.
- [17] **Karacan, A. İ.**, 2000. “Bankacılık Ve Kriz”, Creative Yayıncılık, İstanbul.
- [18] **Ansell, J., Wharton, F.**, 1992. Risk: Analysis, Assessment and Management, John Wiley and Sons. s: 4-5.
- [19] **Çıtak, S.**, 1999. “Geleneksel Risk Yönetiminden Programlanmış Menkul Kıymet İşlemlerine”, Dünya Yayıncılık, Ekonomi Dizisi:7, İstanbul.
- [20] **Galway L.**, 2004. Quantitative Risk Analysis for Project Management, Rand Working Paper, WR-112-RC.
- [21] **Akpınar, H.**, 2000. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, *İ.Ü. İşletme Fakültesi Dergisi*, C:29,s: 1-22.
- [22] **Kaya E., Bulun M., Arslan A.**, 2003. Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları, Akademik Bilişim Konferansları.
- [23] **Ramkumar G.D., Swami A.**, 1998. Clustering Data Without Distance Functions, *IEEE Bulletin Of the Technical Committee on Data Engineering*, Vol.:21, No.1.
- [24] **Kusiak A.**, Data Mining Intelligent Systems Laboratory, The University of Iowa
- [25] **Vahaplar A., İnceoğlu M. M.**, 2001. Veri Madenciliği ve Elektronik Ticaret, VII. Türkiye’de İnternet Konferansı
- [26] **Alataş B., Akın E.**, 2004. Sınıflandırma Kurallarının Karınca Koloni Algoritmasıyla Keşfi, *ELECO'2004*, 357-361, Bursa
- [27] **Ayaz R.**, 2007. Web Madenciliğine Bir Bakış, Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği, İstanbul.
- [28] **Takçı H., Soğukpınar İ.**, 2002. Kütüphane Kullanıcılarının Erişim Örüntülerinin Keşfi, *Bilgi Dünyası*, 3(1), s: 12-26.
- [29] **Kuşan H., Aytakin O., Özdemir İ.**, 2008. İnşaat Projelerinde Bilgisayar Destekli Risk Yönetim Sistemleri, Çanakkale Onsekiz Mart Üniversitesi, *Akademik Bilişim*.
- [30] **Ekemen K. S.**, 2008. İş Sağlığı Güvenliğinde Risk Yaklaşımı Paneli.
- [31] **Çelebi N., Seyrantepe T.**, 1990. Türkiye Kömür Madenciliği Yatırımları İçin Bir Risk Analiz Modeli, Türkiye 7. Kömür Kongresi, Zonguldak, s:147-160.
- [32] **Nasuf E., Orun E.S.**, 1990. Madencilik Projelerinde Mikro-Bilgisayar Destekli Risk Analizi, Madencilik, C:3 s:19-30.
- [33] **Karabacak B., Soğukpınar İ.**, 2003. Risk Analizi ve Yönetimi Süreçleri, *A New Electronic Journal in Computer Science*, III.
- [34] **Sever H. Oğuz B.**, 2002. Veri Tabanlarında Bilgi Keşfine Formel Bir Yaklaşım Kısım I: Eşleştirme Sorguları ve Algoritmalar, *Bilgi Dünyası*, 3(2), s: 173-204, Ankara.

- [35] Çetinyokuş T. Gökçen H., 2008. Bütünleşik Veri Kupu Sistemi (BVKS):Satış Kupu Uygulaması, *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, C: 23, No: 2, s: 477-484, Ankara.
- [36] Kayalı C. A., 2007. 2000-2006 Döneminde Türkiye’de Faaliyet Gösteren Sigorta Şirketlerinin Etkinlik Değerlendirmesi, *Yönetim ve Ekonomi*, C: 14, No: 2, s: 104-114, Manisa.
- [37] Duygulu E., 2006. Türkiye’de Sigorta İşletmeleri ve Sigorta İşletmelerinin Mali Durumuna Genel Bir Bakış, *Sigorta Araştırma Dergisi*, Sayı 2, s: 185-193.
- [38] Dikmen C., 2001. İşletme Ders Notları, Türk Sigorta Enstitüsü Vakfı Yayınları, TSEV 18, Sayı 2, İstanbul.
- [39] Yıldız T., Yıldırım S., Altılar D.T., 2008. İstenmeyen İletilerin Paralleştirilmiş KNN Algoritması ile Tespiti, *Akademik Bilişim*, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale.
- [40] Güven E. N., Onur H., Sağiroğlu Ş., 2008. Yapay Sinir Ağlar ile Web İçeriklerini Sınıflandırma, *Bilgi Dünyası*, 9(1), s: 158-178.
- [41] Kayaalp K., 2007. Asenkron Motorlarda Veri Madenciliği ile Hata Tespiti, *Yüksek Lisans Tezi*, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.
- [42] Uğur A., Kınacı A.C., 2008. Yapay Zeka Teknikleri ve Yapay Sinir Ağları Kullanılarak Web Sayfalarının Sınıflandırılması, *İnternet Teknolojileri Derneği*, inet-tr’08.
- [43] Koç M.L., Balas C.E., Arslan A., 2004. Taş Dolgu Dalgakıranların Yapay Sinir Ağları ile Ön Tasarımı, *İMO Teknik Dergi*, s: 3351-3375, Yazı: 225.
- [44] Yıldırım P., Uludağ M., Görür A., 2008. Hastane Bilgi Sistemlerinde Veri Madenciliği, *Akademik Bilişim*, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale.
- [45] Kılıç N., Gümüş E., Kusayeva N., 2007. Veri Madenciliği, *Lisans Tezi*, Selçuk Üniversitesi, Mühendislik-Mimarlık Fakültesi, Konya.
- [46] Sezgin Ö., 2006. Statistical Methods In Credit Rating, *Yüksek Lisans Tezi*, Orta Doğu Teknik Üniversitesi, Finansal Matematik Bölümü, Ankara.
- [47] Dirican A., 2001. Tanı Testi Performanslarının Değerlendirilmesi ve Kıyaslanması, *Cerrahpaşa Tıp Dergisi*, 32 (1), s: 25-30.
- [48] Hillson, D., 2005. Effective Project Risk Management: Beyond the PMBOK, Risk Doctor and Partners Internal Report, Hampshire, England.
- [49] Aryeetey, K., 2003. Data Analysis and Predictive Modelling Using The Variable Precision Rough Set Approach, *Master Thesis*, Faculty of Graduate Study and Research of University of Regina, Canada.
- [50] <<http://ocw.kfupm.edu.sa/user062/CEM51601/L03%20Risk%20Identification.pdf>>, alındığı tarih 01.09.2008.
- [51] <<http://www.ailab.si/orange>>, alındığı tarih 02.11.2008.
- [52] <<http://yunus.hacettepe.edu.tr/~hcengi/ist376a/6Bolum.doc>>, alındığı tarih 01.09.2008.

- [53] <http://www.assetpartnership.com/downloads/QRA_Final.pdf>, alındığı tarih 12.09.2008.
- [54] <<http://www.uofaweb.ualberta.ca/mece/pdfs/QuantitativeRiskAnalysis.pdf>>, alındığı tarih 12.12.2008.

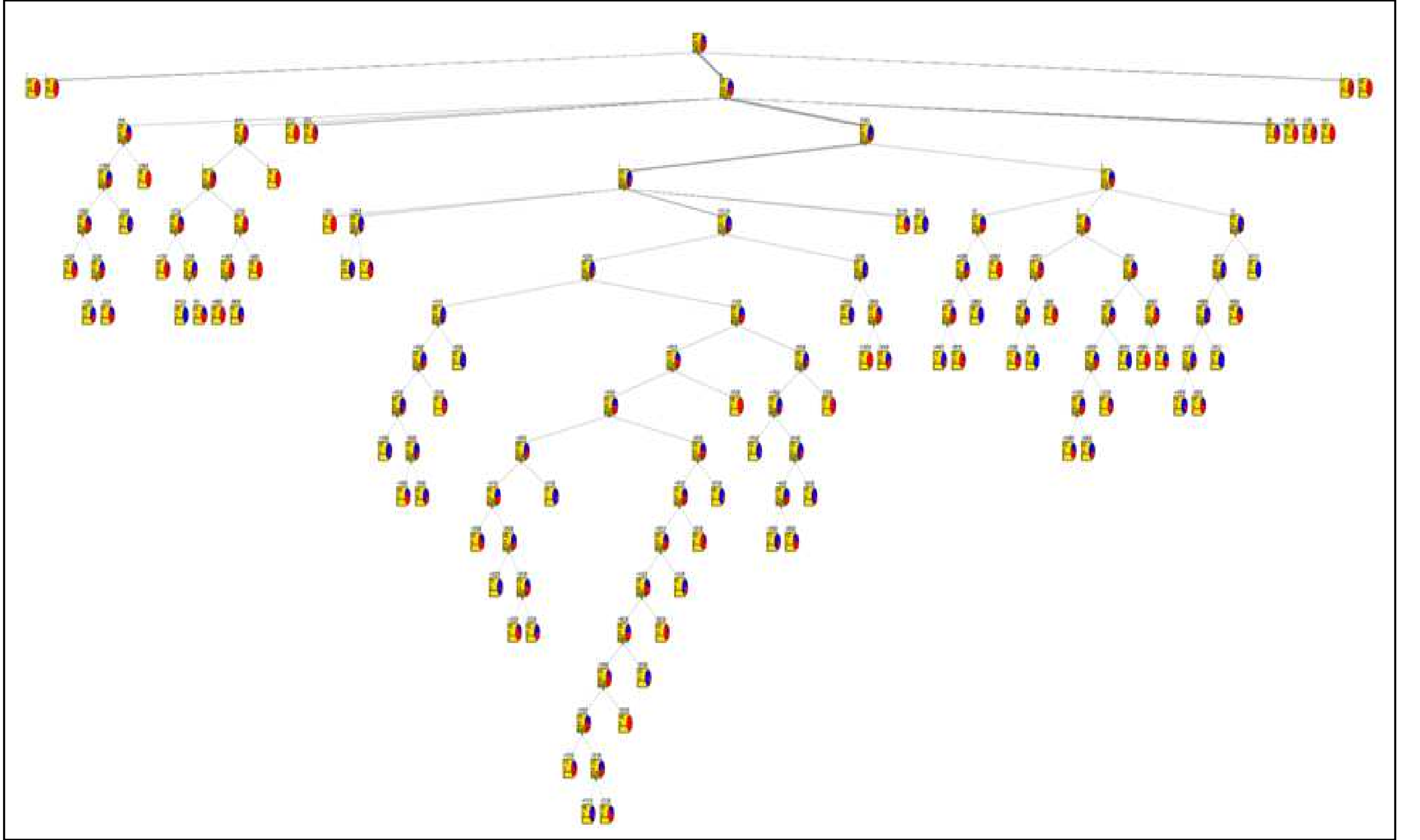
EKLER

EK A.1: Uygulama testi için ihbar verileri

EK A.2: Uygulama karar ağacının grafik gösterimi

Çizelge A.1 : Uygulama testi için ihbar verileri.

hasar il adi	hasar sebebi	bolge adi	menfaat adi	riziko adi	teminat adi	brans adi	kur kodu	supheli hasar	zabit yok	faturali	cam onarimi	eski hasar var	ihbar sonucu	teminat tutari	arac degil	cinsiyet	medeni durum	egitim durumu
ISTANBUL	CARPISMA	ISTANBUL	ARAC+AKS	HUSUSI	KASKO (YANMA, CALINMA, CARPISMA)	KASKO	TL	0	1	0	0	0	H	30200	A	N	N	N
ISTANBUL	KIRILMA	ISTANBUL	EMTEA	KAMYON Y.DISI	ALL RISK	ABONMAN	EUR	0	1	0	0	0	E	144226,5	D	N	N	N
DUZCE	ARAC CARPMASI	ANKARA	ARAC+AKS	ROMORK YURTICI	KASKO (YANMA, CALINMA, CARPISMA)	KASKO	TL	0	0	0	0	0	H	20000	A	N	N	N
ISTANBUL	TEDAVI	ISTANBUL	3.SAHIS	ISLETME	ISVEREN MALI MESULIYET	ISVEREN MALI MESULIYET	USD	0	1	0	0	0	E	500000	D	N	N	N
TRABZON	ARAC CARPMASI	KARADENIZ	ARAC	HUSUSI	KASKO (YANMA, CALINMA, CARPISMA)	KASKO	TL	0	1	0	0	0	H	22400	A	E	E	Y



Şekil A.1 : Uygulama karar ağacının grafik gösterimi

ÖZGEÇMİŞ

Ad Soyad: Duygu Muslu

Doğum Yeri ve Tarihi: İstanbul 29.05.1982

Adres: Latilokum Sok. No: 10/19 Mecidiyeköy

Lise: Sakıp Sabancı Anadolu Lisesi

Lisans Üniversitesi: Yıldız Teknik Üniversitesi - Matematik Mühendisliği

