

**TÜRKİYE'DE OTOMOBİL SİGORTASI
SAHTEKÂRLIKLARININ MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE TESPİT EDİLMESİ**

**DETECTING AUTOMOBILE INSURANCE FRAUDS IN
TURKEY USING MACHINE LEARNING APPROACHES**

EZGİ GÜNBATAR

DR. ÖĞR. ÜYESİ BARBAROS YET

Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
Endüstri Mühendisliği Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2019

EZGİ GÜNBATAR'ın hazırladığı “Türkiye’de Otomobil Sigortası Sahtekârlıklarının Makine Öğrenmesi Yöntemleri İle Tespit Edilmesi” adlı bu çalışma aşağıdaki jüri tarafından ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Doç. Dr. Furkan BAŞER

Başkan


.....

Dr. Öğr. Üyesi Barbaros YET

Danışman


.....

Prof. Dr. Özlem Müge TESTİK

Üye


.....

Doç. Dr. Oumout CHOUSEINOGLU

Üye


.....

Dr. Öğr. Üyesi Diclehan TEZCANER ÖZTÜRK

Üye


.....

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından YÜKSEK LİSANS TEZİ olarak / / tarihinde onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Fen Bilimleri Enstitüsü Müdürü

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

30. / 05 / 2019

Ezgi GÜNBATAR

YAYINLANMA FİKRİ MÜLKİYET HAKKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*” kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.
- Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ay ertelenmiştir.
- Tezim ile ilgili gizlilik kararı verilmiştir.

30. /05/2019

Ezgi GÜNBATAR

ÖZET

TÜRKİYE'DE OTOMOBİL SİGORTASI SAHTEKÂRLIKLARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TESPİT EDİLMESİ

Ezgi GÜNBATAR

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Barbaros YET

Mayıs 2019, 97 sayfa

Sigortacılık, güven ilişkisine dayalı doğası nedeniyle sahtekârlığa elverişli bir sektördür. Kişilerin sigorta sahtekârlığını diğer alanlara göre daha düşük riskli görmesi, kolay yoldan kazanç sağlayabilmeleri, yakalanma riskinin az olması, soygun ve kaçakçılık gibi suçlara göre daha güvenli bulmaları da bu alandaki sahtekârlığın yaygın olmasına sebep olmuştur. Bu alandaki sahtekârlığın yaygın ve tespitinin zor olması, hem sigorta şirketini hem de diğer sigortalılar gibi ekonomideki farklı aktörleri olumsuz etkilemektedir. Sigortalı tarafından basit gözükken bir eylem bile aslında sigortacının öngöremediği finansal bir yükün altına girmesine yol açmakta ve sigortacılık sektörünün mali dengesini bozmaktadır. Bunun sonucunda yapılan aktüeryal hesaplamalarda sapmalar meydana gelir, hasar için ayrılan karşılıklar yetersiz kalır, günün sonunda da sigorta maliyetlerinin artması nedeniyle bu maliyetin bir kısmı sigorta primlerine yüklenerek suçsuz olan sigortalılar da daha fazla prim ödemek zorunda kalır.

Sigorta sahtekârlıkları dünyada olduğu gibi Türkiye’de de sektörün en önemli sorunlarından biridir. Türkiye’de sigorta sahtekârlığıyla mücadele konusunda hem şirketler hem de

düzenleyici otoriteler özelinde farkındalık artmaya başlamış bununla ilgili birimler kurulmaya ve denetimler yapılmaya başlanmıştır.

Buradan yola çıkarak bu çalışmada, Türkiye'deki sigorta verileri üzerinden bir sahtekârlık tespiti çalışması yapılması amaçlanmıştır. Otomobil sigortası verilerine ilişkin kayıt sistemlerinin daha gelişmiş ve verilerin daha sağlıklı olması, ayrıca otomobil sigortalarının sektörün önemli bir payını kapsamaması sebebiyle ülkemizdeki kasko sigortası verileri üzerinden bir çalışma yapılmıştır. Türkiye'de herhangi bir sigorta branşında tüm sigorta şirketi verileri üzerinden yapılmış sahtekârlık tespitine yönelik bir çalışma bulunmamaktadır. Bu çalışma ile elde edilen büyük veriyle, Türkiye'de kasko sigortasındaki tazminata dayalı sahtekârlıklarda önemli olan değişkenlerin, konuların tespit edilmesi ve bu alanda sahtekârlık tespiti yapan en iyi makine öğrenmesi yöntemini bulmak amaçlanmıştır.

Bu kapsamda, literatürde benzer problemlere uygulanan istatistik ve makine öğrenme yöntemleri (Lojistik Regresyon, Karar Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri) ele alınarak büyük veri analizi uygulamaları yapılmıştır. Bunlara ilave olarak, bu alanda literatürde uygulama yeri bulamadığı düşünülen Bayes Ağlarına da yer verilmiştir. Bu çerçevede Bayes Ağları öğrenen üç farklı algoritma ile çalışılmıştır. Uygulanan yöntemlerden sahte hasarların tespiti noktasında tüm yöntemlerin yüksek performans gösterdiği ancak hiçbirinin diğerine çok üstün gelmediği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Otomobil Sigortası, Sahtekârlık Tespiti, Makine Öğrenmesi, Büyük Veri, Bayes Ağları, Yapay Sinir Ağları.

ABSTRACT

DETECTING AUTOMOBILE INSURANCE FRAUDS IN TURKEY USING MACHINE LEARNING APPROACHES

Ezgi GÜNBATAR

Master of Science, Department of Industrial Engineering

Supervisor: Assistant Professor Barbaros YET

May 2019, 97 pages

The insurance sector has inherent high risk of fraudulent activity due to its trust-based nature. Insurance fraud is often considered to be less risky and easier than other criminal activities like robbery and smuggling. This has also caused the fraudulent behavior to be widespread in this domain. Since insurance frauds are relatively more common and difficult to detect, it has significant impact on insurance companies and other stakeholders in this sector. Even simple frauds may lead to a financial burden that is not expected by the insurer, and this impairs the financial balance of the insurance sector. As a result, deviations occur in the actuarial calculations, claim reserves become insufficient. At the end of the day, increased insurance costs are paid as insurance premiums by the insured who are not guilty.

Insurance fraud is one of the most important problems of the sector in Turkey as well as in other countries. Awareness about fighting against insurance fraud has increased for both insurance companies and regulatory authorities, as they have started to establish units related to fraud and increased the controls on this subject.

This study aims to detect insurance frauds in Turkey by using a large dataset. Since the automobile insurance covers a significant share of the sector, the registration systems for automobile insurance data are more advanced in Turkey, and data can be recorded more accurately. However, previous studies that use comprehensive automobile insurance data to detect fraud in Turkey are not available. We aim to understand the important variables related to automobile insurance fraud, and evaluate the performance of different machine learning algorithms in this domain.

We have applied various statistical and machine learning methods that are used for similar problems in the previous literature (Logistic Regression, Decision Trees, Artificial Neural Networks, Support Vector Machines). In addition, we have also used Bayesian Networks with three different learning algorithms, which has not been applied for automobile insurance fraud detection in previous studies. All methods had high performance for fraud detection in our evaluations but none of them were superior to the other.

Keywords: Automobile Insurance, Fraud Detection, Machine Learning, Big Data, Bayesian Networks, Artificial Neural Networks.

TEŐEKKÜR

Yüksek lisans çalışması süresince sağladığı tüm katkı ve desteęi için tez danışmanım sayın Dr. Öğr. Üyesi Barbaros YET'e,

bu süreçte yardımlarını esirgemeyen Genel Müdür Yardımcım Mehmet HÖBEK'e,

başta Salih TAŞYON, Fatma Perihan NEFES BİLİCİ ve Şehmus ŞEN olmak üzere bu çalışmada bana yardımcı olan ve kıymetli vakitlerini ayıran Sigorta Bilgi Merkezi çalışanlarına,

bilgileri ve manevi desteęiyle her zaman yanımda olan sevgili eşim Çınar GÜNBATAR'a

ve hayatımın her noktasında sonsuz destek ve sevgileriyle yanımda olan aileme

en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	iii
TEŞEKKÜR.....	v
ÇİZELGELER.....	viii
ŞEKİLLER	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ	1
2. SİGORTA VE TEMEL KAVRAMLAR.....	4
2.1. Risk Tanımı ve Çeşitleri.....	4
2.1.1. Saf ve Spekülatif Risk.....	5
2.1.2. Temel ve Özel Risk.....	5
2.1.3. Fiziksel ve Moral (Ahlaki) Risk.....	6
2.2. Risk Yönetim Teknikleri	6
2.2.1. Risk Kontrolü	7
2.2.2. Risk Finansmanı	7
2.3. Sigorta Kavramı ve Temel Prensipleri	8
2.4. Sigorta Türleri.....	10
2.4.1. Mal (Zarar) Sigortaları	10
2.4.2. Can Sigortaları.....	10
2.4.3. Sorumluluk Sigortaları	11
3. SİGORTA SAHTEKÂRLIĞI KAVRAMI.....	12
3.1. Sigorta Sahtekârlığı Nedir?	12
3.2. Sahtekârlık Riski ve Sahtekârlık Üçgeni	14

3.3.	Sahtekârlık Türleri	16
3.3.1.	Otomobil Sigortasında Karşılaşılan Sahtekârlık Örnekleri	18
3.3.2.	Türkiye’de Sigorta Sahtekârlıkları	19
4.	SİGORTA SAHTEKÂRLIĞI TESPİTİNDE YAPILAN ÇALIŞMALAR	24
4.1.	Sigorta Sahtekârlığı Tespitinde Kullanılan Yöntemler	25
4.1.1.	Gözetimli (Supervised) Yöntemler	25
4.1.2.	Gözetimsiz (Unsupervised) Yöntemler	32
4.1.3.	Yarı Gözetimli (Hibrit) Yöntemler	34
4.2.	Sahtekârlık Tespiti Çalışmalarındaki Önemli Değişkenler	36
4.3.	Yapılan Çalışmanın Literatürdeki Konumu	38
5.	ÇALIŞMADA KULLANILAN İSTATİSTİKSEL YÖNTEMLER	40
5.1.	Lojistik Regresyon Analizi	40
5.2.	Bayes Ağları	42
5.3.	Karar Ağaçları	45
5.4.	Yapay Sinir Ağları	47
5.5.	Destek Vektör Makineleri	49
5.6.	Modellerin Performansını Ölçmeye Yönelik Çalışmada Kullanılan Kavramlar	51
6.	TÜRKİYE’DE OTOMOBİL SİGORTALARINDAKİ SAHTEKÂRLIK TESPİTİNE YÖNELİK BİR UYGULAMA	54
6.1.	Veri ve Verideki Değişkenler	54
6.2.	Veri Düzeltme İşlemleri	61
6.3.	Veri Analizi ve Sonuçlar	66
7.	SONUÇLAR VE ÖNERİLER	87
8.	KAYNAKLAR	92
	ÖZGEÇMİŞ	97

ÇİZELGELER

Çizelge 4.1. Otomobil sigortasında sahtekârlık tespitine yönelik yapılan çalışmalar.....	36
Çizelge 5.1. Hata Matrisi (Confusion Matrix) değerleri	51
Çizelge 6.1. Veri düzenlenmeden önceki mevcuttaki iş kuralları.....	59
Çizelge 6.2. Başka bir değişkenle karşılanması sebebiyle çıkarılan iş kuralları.....	64
Çizelge 6.3. Çalışmaya dâhil edilen iş kurallarının eski ve yeni isimleri	64
Çizelge 6.4. Veri analizinde kullanılan değişkenler.....	66
Çizelge 6.5. Lojistik Regresyon analizinden elde edilen önemli değişkenler, katsayıları ve p değerleri	68
Çizelge 6.6. Lojistik Regresyon analizinden elde edilen performans ölçütleri.....	70
Çizelge 6.7. Lojistik Regresyon analizinden elde edilen hata matrisi değerleri	71
Çizelge 6.8. HC Bayes Ağından elde edilen performans ölçütleri.....	72
Çizelge 6.9. HC Bayes Ağından elde edilen hata matrisi değerleri	72
Çizelge 6.10. MMHC Bayes Ağından elde edilen performans ölçütleri	73
Çizelge 6.11. MMHC Bayes Ağından elde edilen hata matrisi değerleri	73
Çizelge 6.12. Naive Bayes Sınıflandırıcılı Bayes Ağından elde edilen performans ölçütleri ..	74
Çizelge 6.13. Naive Bayes Sınıflandırıcılı Bayes Ağından elde edilen hata matrisi değerleri .	75
Çizelge 6.14. C4.5 Karar Ağacından elde edilen performans ölçütleri.....	76
Çizelge 6.15. C4.5 Karar Ağacından elde edilen hata matrisi değerleri	76
Çizelge 6.16. Bagging Karar Ağacından elde edilen performans ölçütleri.....	77
Çizelge 6.17. Bagging Karar Ağacından elde edilen hata matrisi değerleri	77
Çizelge 6.18. Random Forest Karar Ağacından elde edilen performans ölçütleri.....	78
Çizelge 6.19. Random Forest Karar Ağacından elde edilen hata matrisi değerleri	78

Çizelge 6.20. Yapay Sinir Ağlarından elde edilen performans ölçütleri.....	80
Çizelge 6.21. Yapay Sinir Ağlarından elde edilen hata matrisi değerleri	80
Çizelge 6.22. Destek Vektör Makinelerinden elde edilen performans ölçütleri	83
Çizelge 6.23. Destek Vektör Makinelerinden elde edilen hata matrisi değerleri.....	83
Çizelge 6.24. Çalışmada uygulanan yöntemlerde elde edilen performans ölçütleri	85
Çizelge 6.25. Çalışmada uygulanan yöntemlerin çalışma süreleri.....	86

ŞEKİLLER

Şekil 3.1. Sahtekârlık üçgeni.....	15
Şekil 3.2. 2018 yılında otomobil sigortası hasar dosyalarındaki sahtekârlıkların dağılımı	20
Şekil 3.3. Nisan 2018-Nisan 2019 tarihleri arasında otomobil sigortası hasar dosyalarındaki sahtekârlıkların dağılımı	21
Şekil 5.1. Bayes ağının yapısı	43
Şekil 5.2. Bagging yöntemi gösterimi.....	46
Şekil 5.3. Yapay sinir ağı yapısı ve katmanları.....	48
Şekil 5.4. R^2 ve R^3 'teki doğrusal hiper düzlem ve destek vektörleri.....	50
Şekil 5.5. R^2 ve R^3 'teki doğrusal olmayan hiper düzlem	50
Şekil 5.6. ROC eğrisi.....	53
Şekil 6.1. Lojistik Regresyon analizinden elde edilen ROC eğrisi.	71
Şekil 6.2. HC Bayes Ağından elde edilen ROC eğrisi	73
Şekil 6.3. MMHC Bayes Ağından elde edilen ROC eğrisi.....	74
Şekil 6.4. Naive Bayes Sınıflandırıcılı Bayes Ağından elde edilen ROC eğrisi.....	75
Şekil 6.5. C4.5 Karar Ağacından elde edilen ROC eğrisi.....	77
Şekil 6.6. Bagging Karar Ağacından elde edilen ROC eğrisi	78
Şekil 6.7. Random Forest Karar Ağacından elde edilen ROC eğrisi	79
Şekil 6.8. Yapay Sinir Ağlarından elde edilen ROC eğrisi.....	81
Şekil 6.9. Çalışmada uygulanan Yapay Sinir Ağı yapısı	82
Şekil 6.10. Destek Vektör Makinelerinden elde edilen ROC eğrisi.....	84

SİMGELER VE KISALTMALAR

Kısaltmalar

ABD	Amerika Birleşik Devletleri
ACFE	The Association of Certified Fraud Examiners
AIC	Akaike Information Criterion
AUC	Area Under Curve
Bagging	Bootstrap Aggregating
BIC	Bayesian Information Criterion
CHAID	Chi-Squared Automatic Interaction Detector
CT	Consolidated Trees
C&RT	Classification and Regression Trees
DVM	Destek Vektör Makineleri
DN	Doğru Negatif
DP	Doğru Pozitif
GS	Grow-Shrink
HC	Hill Climbing
HIC	Health Insurance Commission
H2PC	Hybrid HPC
IAMB	Incremental Association
MARS	Multivariate Adaptive Regression Splines
MMHC	Max-Min Hill Climbing
MMPC	Max-Min Parents and Children
ROC	Receiver Operator Characteristics
RSMAX2	2-phase Restricted Maximization
SBM	Sigorta Bilgi Merkezi
SİSBİS	Sigorta Suistimalleri Bilgi Paylaşım Sistemi
SİSEB	Sigorta Sahteciliklerini Engelleme Bürosu
SPRINT	Scalable Parallelizable Induction of Decision Trees
YN	Yanlış Negatif

YP

Yanlış Pozitif

QUEST

Quick, Unbiased, Efficient Statistical Tree

1. GİRİŞ

Sigorta, bir riskin belirli bir ödeme karşılığında (prim), bir tarafça (sigortalı) başka bir tarafa (sigortacı) belirli koşullar altında transfer edilmesidir [1]. Sigorta sözleşmeleri doğası gereğince karşılıklı güven üzerine kurulur, tarafların imkânları ve sorumlulukları çerçevesinde bu güveni sarsmaması gerekir. Dolayısıyla tarafların özellikle de sigortalının azami iyi niyet kuralına riayet ederek, gerek sözleşme kurulurken gerekse sözleşmenin devamında ve hasar anında bu tutumunu koruması gerekir. Sigortalının yanlış beyanda bulunması ve sigortacıyı yanıltıcı hareketlerde bulunarak azami iyi niyet kuralına aykırı davranması, sigorta sahtekârlığına zemin hazırlamaktadır. Sigorta sahtekârlığı, haksız kazanç elde etmek amacıyla, sigortalı tarafından sigortacının bilerek ve kasıtlı olarak aldatılmasıdır [2]. Sigortacılık, güvene dayalı yapısı nedeniyle sahtekârlığa elverişli bir sektördür. İnsanların sigorta sahtekârlığına daha kolay yönelmeleri; kolay yoldan kazanç sağlayabilmeleri, yakalanma riskinin az olması ve yakalansa bile sonuçlarının çok ağır olmayacağı düşüncesinden kaynaklanmaktadır. Bu durum da bu alandaki sahtekârlığın yaygın olmasına ve normalde suça yönelmeyen kişiler tarafından bile denenmesine yol açmaktadır. Sigorta sahtekârlıkları hem sigorta şirketini hem de diğer sigortalılar gibi ekonomideki farklı aktörleri olumsuz etkilemektedir. Gerek başka ülkelerde gerekse Türkiye’de sahte hasarların tüm hasar ödemeleri içerisindeki payının %10-15 arasında olduğu tahmin edilmektedir. Türkiye’de 2014 yılında kasko ve trafik sigortası branşlarında ödenen ve muallak olan toplam 11 milyar TL’lik hasarın, 1.4 – 1.7 milyar TL’sinin suistimal olduğu düşünülmektedir [3-5]. Ekonomik ve sosyal açıdan büyük bir sorun yaratan sigorta sahtekârlıklarının tespiti ve mücadele edilmesi konusunda gün geçtikçe farkındalık artmakta ve çalışmalar yapılmaktadır.

Sigorta sahtekârlıklarının sektöre ve paydaşlarına verdiği zararlar düşünüldüğünde bu alanda çalışma yapmanın elzem olduğu düşünülmektedir. Buradan yola çıkarak Türkiye’deki sigorta verileri üzerinden bir sahtekârlık tespiti çalışması yapılması amaçlanmış, otomobil sigortası verilerine ilişkin kayıt sistemlerinin daha gelişmiş ve verilerin daha sağlıklı olması sebebiyle ülkemizdeki kasko sigortası verileri üzerinden bir çalışma yapılmıştır. Çalışma kapsamında kullanılan veriler Türkiye’deki sigorta verilerinin toplandığı bir merkez konumundaki Sigorta Bilgi Merkezi’nin (SBM) altında kurulan Sigorta Sahteciliklerini Engelleme Bürosu’ndan

(SİSEB) temin edilmiştir. SİSEB'ten alınan veri, tüm sigorta şirketlerinin kasko sigortası hasar dosyasına ilişkin bilgileri içermekte ve bunların arasında hasar suistimallerine ilişkin bazı değişkenler de bulunmaktadır. Türkiye'de herhangi bir sigorta branşında tüm sigorta şirketi verileri üzerinden yapılmış sahtekârlık tespitine yönelik bir çalışma bulunmamaktadır. Bu çalışma ile eldeki büyük veriyle, Türkiye'de kasko sigortasındaki tazminata dayalı sahtekârlıklarda önemli olan değişkenlerin, konuların tespit edilmesi ve bu alanda sahtekârlık tespiti yapan en iyi makine öğrenmesi yöntemini bulmak amaçlanmıştır.

Sigorta sahtekârlıklarının istatistiksel yöntemlerle tespit edilmesi noktasında özellikle de otomobil sigortalarıyla ilgili birçok çalışma bulunmaktadır. Bu çerçevede; Lojistik Regresyon, Karar Ağaçları, Yapay Sinir Ağları (YSA), Bulanık Mantık, Destek Vektör Makineleri (DVM), Naive Bayes Sınıflandırıcısı ve K-En Yakın Komşu Sınıflandırıcısı gibi gözetimli öğrenme yöntemlerinin kullanıldığı çalışmaların yanı sıra, Öz Düzenleyici Haritalar, Temel Bileşen Analizi ve Veri Görselleştirme gibi gözetimsiz öğrenme yöntemlerini kullanan çalışmalar da bulunmaktadır (bu çalışmaların detaylı literatür taraması Bölüm 4'te sunulacaktır).

Sınıflandırma algoritmaları, 0 ve 1 şeklinde etiketlenmiş ve sınıflandırıcı eğitimini içeren bir veri kümesi gerektirir. Bu, en yaygın öğrenme yaklaşımıdır. Gözetimli öğrenmenin en büyük avantajı, bu yaklaşımın algoritma tarafından işlenen tüm sınıf çıktılarının insanlar için anlamlı olması ve sınıflandırma ile regresyon problemleri için kolayca kullanılabilir olmasıdır. Ancak gözetimli öğrenmenin bazı kısıtlamaları vardır. Bunların en önemlisi, verilerin etiketlenmesinde yaşanan zorluklardır. Büyük miktardaki verilerle çalışıldığında, verinin hepsinin etiketlenmesi maliyetli ve fiziksel olarak zor olabilir. Diğer bir problem ise, bazen ayırt edici bir etiket bulmanın son derece zor olması veya etiket bilgilerinde belirsizlikler olmasıdır [6]. Gözetimsiz yöntemler, gözetimli yöntemlerde bahsedilen bilgilere ve etiketlere gereksinim duymaz ve normal davranışı temsil eden dağılımı modelleyerek daha sonra bu modelden en büyük sapmayı gösteren gözlemleri, yani aykırı değerleri bulmaya çalışır [7]. Tez çalışmasında kullanılacak öğrenme yöntemi seçilirken bu özellikler göz önünde bulundurulmuş, eldeki mevcut verinin sahtekârlığa ilişkin hali hazırda etiketlenmiş olması sebebiyle çalışmaya ilave bir maliyet getirmeyeceğinden, bu bilgilerin avantajı göz önünde bulundurularak gözetimli öğrenme yöntemleri tercih edilmiştir.

Bu kapsamda çalışmada, otomobil sigortası sahtekârlık tespitinde literatürdeki yöntemler (Lojistik Regresyon, Karar Ağaçları, YSA, DVM) ele alınarak büyük veri analizi yapılmıştır. Ancak literatürdeki yöntemlere ilave olarak bu alanda uygulama yeri bulamadığı düşünülen Bayes Ağlarına da yer verilmiş ve üç farklı algoritma ile çalışılmıştır. Özetle, bu çalışma ile hem literatürde yer bulamayan bir yöntem kullanıldığı hem de Türkiye’de uygulaması yapılmamış bir alan ve veriyle çalışıldığından literatüre yeni bir katkı yapıldığı değerlendirilmektedir.

Bu tez çalışmasının diğer bölümleri şu şekilde düzenlenmiştir. 2. Bölüm’de çalışmanın temel kavramları olan risk ve sigorta gibi terimler anlatılmaktadır. Aynı zamanda bunların çeşitlerine değinilmekte, çalışmaya temel olan iyi niyet prensibi ve moral risk gibi kavramlardan bahsedilmektedir. 3. Bölüm’de sigorta sahtekârlığı kavramının, sahtekârlık türlerinin neler olduğu, çalışmada tespit edilmesi hedeflenen sahtekârlık türünün hangi sınıfa girdiği ve otomobil sigortalarında sık görülen sahtekârlık şekilleri anlatılmaktadır. 4. Bölüm’de ağırlıklı olarak otomobil sigortaları olmak üzere sigorta sahtekârlık tespitine yönelik literatürde yapılan istatistiksel çalışmalara yer verilmektedir. 5. Bölüm’de tez çalışmasında kullanılan gözetimli öğrenme yöntemlerine ilişkin teorik bilgilere yer verilmektedir. 6. Bölüm’de çalışmada kullanılan verinin yapısına, yapılan veri temizleme işlemlerine ve uygulanan yöntemlerin sonuçlarına değinilmektedir. Son olarak 7. Bölüm’de elde edilen sonuçlar ortaya konularak yorumlanmakta ve gelecek çalışmalar için öneriler sunulmaktadır.

2. SİGORTA VE TEMEL KAVRAMLAR

Bu bölümde, çalışmanın temel kavramları olan risk ve sigorta gibi terimler anlatılmaktadır. Aynı zamanda bunların çeşitlerine değinilmekte, çalışmaya temel olan iyi niyet prensibi ve moral risk gibi kavramlardan bahsedilmektedir.

2.1. Risk Tanımı ve Çeşitleri

Risk olgusu, hayatımızda birçok alanda görülebilir olmakla beraber değişik sektörler ve bakış açılarına göre farklı anlamlar kazanabilmesi sebebiyle tanımlanması zor bir kavramdır. Ekonomistlerin, davranış bilimcilerinin, risk teorisyenlerinin, istatistikçilerin ve aktüerlerin her biri kendi içinde farklı bir risk kavramına sahiptir.

Bununla birlikte, tarihsel olarak risk, belirsizlik açısından tanımlanmıştır. Bu kavrama dayanarak, risk, bir kayıp oluşumuna ilişkin belirsizlik olarak tanımlanır. Ekonomi ve finans literatüründe, çoğu zaman risk ve belirsizlik arasında ayırım yapılmamakla birlikte, “risk” kavramı, muhtemel sonuçların olasılığının belirli bir doğrulukta tahmin edilebildiği durumlarda kullanılırken, "belirsizlik" ise bu olasılıkların tahmin edilemediği durumlarda kullanılır. Bu nedenle, birçok yazar kendi kavramını geliştirmiştir [1].

Risk, beklenen bir sonuçtan olumsuz bir sapma ihtimalinin olduğu durumdur. Örneğin, ev sahibiyken evinizde yangın çıkmamasını umarsınız, ancak sonuç olarak beklentinizin dışındaki durumun olma ihtimali sizin risk olasılığınızı oluşturur [8]. Başka bir tanımlamayla; “Risk, gerçekleşebilecek ancak gerçekleşmesi ya da ne zaman gerçekleşeceği kesin olmayan zarar verici olaylardır” [9].

Risk kavramı sigortacılığın temel unsurlarından biridir ve konusuna göre pek çok alt tür altında incelenebilir. Riskler, tanımlanabilme, değerlendirilebilme, ölçülebilme ve yönetilebilme amaçlarıyla bazı sınıflara ayrılmaktadır. Bu kapsamda riskler aşağıdaki gibi sınıflandırılabilir.

2.1.1. Saf ve Spekülatif Risk

Saf ve spekülatif risk ayrımındaki temel husus kaybetme riskine karşı kazanma olasılığının olup olmamasıdır. Spekülatif risk, kayıp ihtimaliyle birlikte kazanç ihtimalinin de olduğu durumu ifade eder. Spekülatif risklere örnek olarak borsa yatırımları, at yarışı bahisleri ve gayrimenkul yatırımları sayılabilir. Bu gibi durumlarda hem kazanç hem de kayıp söz konusudur. Saf risk ise, kazanç elde etme olasılığının olmadığı yani sadece kaybın varlığından veya yokluğundan bahsedildiği durumu ifade eder. Saf risk örnekleri arasında erken ölüm, iş kazaları, yangın, yıldırım, sel ve deprem sayılabilir [1].

Sigorta bir zenginleşme aracı değil, güvence aracıdır. Sigorta yoluyla risk yönetmek için sigortalının menfaatinin zarara uğraması gerekir. Ancak spekülatif riskte kazanma ve kaybetme olasılığı birlikte. Sonuç belirsizdir, her iki taraf için de kazanma veya kaybetme olasılığı vardır. Saf riskte ise kâr elde edilmesi söz konusu değildir, risk meydana geldiğinde kesin olarak kişinin ekonomik gücünü azaltan bir zarar doğurur. Bu nedenle spekülatif risklerin yönetimi genellikle sigorta konusunun dışında olup yalnızca saf riskler sigortalanabilir. Saf riskler ise kendi içinde üç grupta incelenebilir:

- **Kişiyeye Yönelik Riskler:** Kişinin gelir kazanma yeteneğinden yoksun kalmasına bağlı olarak gelirin ya da mal varlığının azalmasına sebep olan risklerdir. Buna örnek olarak; işsiz kalmak, hastalanmak, maluliyet hali, emeklilikte yeterli gelire sahip olamamak, erken veya zamansız gerçekleşen ölüm gibi riskler verilebilir.
- **Mala Yönelik Riskler:** Çeşitli olaylar neticesinde, bir malın ekonomik değerinin tamamen ortadan kaybolması veya azalması gibi risklerdir.
- **Sorumluluğa Yönelik Riskler:** Kişinin kendi davranışlarının veya yetki alanına giren bir olayın sonucunda ortaya çıkan risklerdir [10].

2.1.2. Temel ve Özel Risk

Temel ve özel riskler arasındaki ayrım, kayıpların nedenleri ve sonuçlarındaki farka dayanır. Temel riskler, neden ve sonuç açısından kişisel olmayan kayıpları içerir. Bunlar, fiziki olaylardan kaynaklanabileceği gibi, çoğunlukla ekonomik, sosyal ve siyasi etkenlerin sebep

olduğu grup riskleridir. Büyük kesimleri hatta nüfusun tamamını etkilerler. Özel riskler ise, bireysel olaylardan kaynaklanan kayıpları içerir ve belirli bir grup tarafından değil, bireyler tarafından hissedilmektedir [8]. İşsizlik, savaş, enflasyon, deprem ve sel temel risklerdir. Bir konutu su basması ve bir işyerinde yangın çıkması gibi olaylar özel risklere örnek olarak verilebilir.

2.1.3. Fiziksel ve Moral (Ahlaki) Risk

Riskin sağlıklı yönetilebilmesi için, riskin gerçekleşmesinde etkisi olan olayların bilinmesi büyük önem taşımaktadır. Riskler fiziksel bir nedenle ortaya çıkabileceği gibi, kişilerin moralitesinden yani ahlaki durumunda da kaynaklanabilmektedir. Fiziksel risklerde riskin gerçekleşmesine etki eden faktörler fizikseldir. Örneğin, bir yapı malzemesinin yangına veya depreme karşı dayanıklılığı, kişinin mesleğinin madenci ya da öğretmen olması, ailede kalıtsal bir hastalığın olup olmaması gibi durumlar fiziksel risklerin konusudur [10].

Moral riskler ise, sigortalının ahlaki durumu ile ilişkili risklerdir. Moral risklerin konusunu, sigortalının karakteri, değerleri, yetiştirilme tarzı ve ailesi gibi sigortalıya bağlı olan soyut kavramlar oluşturmaktadır [10]. Bu risk, sosyal ya da ahlaki risk olarak da adlandırılabilir. Moral risk, sigortalının dikkatsizliği, düzensizliği, bilgisizliği, ruhsal ya da fiziksel bozukluğu nedenlerinden kaynaklanabildiği gibi sigortalının kasıtlı olarak hile ve haksız kazanç sağlama isteği, riske karşı kayıtsız kalması gibi nedenlerden de kaynaklanabilir. Çalışmanın temel kavramlarından biri olan moral riskin bir sonucu olarak ortaya çıkan sigorta sahtekârlıkları Bölüm 3’te daha detaylı anlatılmaktadır.

Risk tanımını ve risk çeşitlerini anlattıktan sonra bu kavramların sigortayla olan bağlantısını kurabilmek için öncelikle risk yönetim tekniklerine değinmek gerekmektedir.

2.2. Risk Yönetim Teknikleri

Risk yönetimi; maruz kalınan risklerin tespit edilmesi, bu risklerin etkilerinin belirlenmesi, risklerle nasıl başa çıkılacağıın tespit edilmesi, belirlenen bu yöntemlerin uygulanması ve en

sonunda elde edilen sonuçların incelenmesini kapsayan bir süreçtir [9]. Modern risk yönetimi terminolojisinde riskle başa çıkma teknikleri iki temel yaklaşıma ayrılmıştır: Risk Kontrolü ve Risk Finansmanı.

2.2.1. Risk Kontrolü

Risk kontrolü, maruz kalınan zarar riskini en aza indirmeye odaklanır ve riskten kaçınma ile risk azaltma tekniklerini içerir.

- **Riskten Kaçınmak:** Risk meydana geldiğinde ortaya çıkacak olumsuz durumlarla uğraşılmak istenmemesi ya da riski yönetmeye uygun bir tekniğin bulunamaması gibi durumlarda, riski meydana getiren temel olaydan kaçınmak ve böylece riski tamamen ortadan kaldırmaktır. Bir işletmenin tehlikeli bir maddenin üretimini yapmamayı tercih etmesi riskten kaçınmaya örnek olarak verilebilir. Riskten kaçınmak, bazı risklerle baş etmenin tek alternatifi olmasına rağmen olumsuz görülen bir yaklaşımdır. Bireyin kişisel ilerlemesi ve ekonomideki ilerleme, risk almayı gerektirir, sistematik bir şekilde riskten kaçınmanın kullanıldığı durumlarda bireylerin hedeflerine ulaşması mümkün olmayabilir.
- **Riski Azaltmak:** Riskin ortaya çıkma olasılığının azaltılmasına ya da ortaya çıktığında olumsuz etkilerinin azaltılmasına veya önlenmesine yönelik eylemler risk azaltma yöntemleridir. Bir fabrikada yangın çıkma olasılığı azaltmak amacıyla yangına dayanıklı malzeme kullanmak ya da yangın sonrasında devreye girecek bir fiskeye sistemi kurmak riski azaltmaya örnek verilebilir [8].

2.2.2. Risk Finansmanı

Risk finansmanı, risk kontrol tekniklerinin uygulanmasından sonra kalan risklerden kaynaklanan kayıpları karşılamak için fonların mevcudiyetinin düzenlenmesine yoğunlaşmakta ve riski üzerinde tutma ile risk transferi araçlarını içerir.

- **Riski Üzerinde Tutmak:** Bu yöntem, riski kabullenmek olarak da adlandırılabilir. Bir başka deyişle, riski yönetmek için hiçbir yöntemin uygulanmaması durumunda risk,

bireyin riski kendi üzerinde tutmasıyla yönetilmiş olur. Riski üzerinde tutmak, riskin başka yöntemler kullanılarak yönetilmesinin yüksek maliyetler getirmesi ve risk gerçekleştikten sonra meydana gelecek zararın büyük boyutlarda olmaması gibi durumlarda tercih edilebilir [8].

- **Riski Transfer Etmek:** Bir risk gerçekleştiğinde ortaya çıkabilecek kaybın maliyetinin bir başka kişi ya da kurumla paylaşılmasına risk transferi denir. Bu transfer çeşitli yollarla gerçekleştirilebilir. Sigorta sözleşmelerinin satın alınması, genellikle risk transferinde birincil yaklaşımdır. Bu yöntemde, risk bu konuda uzmanlaşmış bir başka organizasyona bir sözleşme aracılığıyla devredilir. Sigorta sözleşmelerinde, risk belirli bir ödeme karşılığında (prim), bir tarafça (sigortalı) bir başka tarafa (sigortacı) belirli koşullar dâhilinde transfer edilir. Riski transfer etmek, kayıp olasılığının düşük ancak zararın şiddetinin yüksek olduğu kayıp risklerine karşı uygun bir risk yönetim tekniğidir [1].

2.3. Sigorta Kavramı ve Temel Prensipleri

Risk kavramında olduğu gibi sigorta kavramında da birçok tanımlama bulunmaktadır. Vaughan ve Vaughan [8] sigortayı, yeteri kadar çok sayıda homojen olayı gruplandırarak kayıpları bir bütün olarak öngörülebilir hale getirme yoluyla riskin azaltılması ve ortadan kaldırmasında kullanılan ekonomik bir araç olarak tanımlamıştır. Sigortanın tanımını “Aynı türden tehlikeyle karşı karşıya olan kişilerin, belirli bir miktar para ödemesi yoluyla toplanan tutarın sadece o tehlikenin gerçekleşmesi sonucu fiilen zarara uğrayanların zararını karşılamada kullandığı bir risk transfer sistemidir.” şeklinde yapılabılır [2]. Bu tanımları daha geniş bir çerçevede değerlendirecek olursak, sigorta, aynı türden risklerin bir araya getirilerek paylaşılması yöntemiyle etkilerinin azaltılmasını sağlayan ve bu sayede normal şartlarda yönetilemeyecek kayıpların tazmin edilmesine yönelik ekonomik bir araçtır.

Teminat sunan sigorta şirketleri ve teminat talebinde bulunan sigortalılar arasında yapılacak sigorta sözleşmesi kapsamında, genel anlamda kabul gören altı tane temel sigorta prensibi bulunmaktadır:

- Azami İyi Niyet
- Sigortalıların Menfaat

- Yakın Sebep
- Tazminat
- Halefiyet
- Hasara Katılım

Bu prensipler bir sigorta sözleşmesiyle oluşturulan ilişkide gerekli olan temel kuralları kapsamaktadır. Bu çalışma kapsamında asıl üzerinde duracağımız konuların temelini “azami iyi niyet prensibi” oluşturduğundan söz konusu prensibi aşağıdaki gibi detaylandırabiliriz.

Azami iyi niyet, sigorta sözleşmesi taraflarının birbirlerine mutlak iyi niyetle davranma zorunluluğunu ifade eder. Sigorta sözleşmeleri güven üzerine kurulur, tarafların imkânları ve sorumlulukları çerçevesinde bu güveni sarsmaması gerekir. Azami iyi niyet prensibi her iki tarafı ilgilendirse de uygulamada çoğunlukla sigortalıyı ilgilendiren bir yükümlülüktür. Bu prensip gereği, sigorta sözleşmesinde sigortalının beyanının doğru olması gerekmektedir. Çünkü sigortacı sigortalının beyan ettiği şartlarda riski üstlenmeyi kabul eder. Eğer sigorta sözleşmesindeki beyan yanlış ya da eksikse, sigortalı bu prensibe aykırı davranmış olur. Sigortalı bunu kasıtlı olarak yapmışsa, sigortacının poliçeyi iptal etme ya da hasarı ödememe hakkı bulunmaktadır. Aynı zamanda bu prensip gereğince sigortalı, hasar meydana geldikten sonra sigortalı değilmişçesine hareket ederek zararın azaltılmasına yardımcı olmak ve hasar tespiti süresince sigortacıya her türlü kolaylığı göstermekle yükümlüdür. Kısaca bu prensip moral risk dediğimiz sigortalının ahlaki karakteriyle bağlantılı olan riskle doğrudan ilişkilidir.

Sigortacı da aynı şekilde bu ilkeye uygun olarak davranmalı sigortalıyı kötü niyetli olarak yanıltmamalıdır. Yüksek değerde prim alma, olmayan riskler için teminat verme, tazminat ödemesini geciktirme gibi davranışlardan kaçınmalıdır.

Sigorta sözleşmesinin kurulması esnasında hem sigortalı hem de sigortacı, karşı tarafın verdiği bilgiye güvenmek durumundadır. Taraflardan birinin azami iyi niyet prensibine uymayarak yanlış bir bilgi vermesi, karşı tarafı yanıltmasına ve öngöremediği bir sözleşmeye girmesine sebep olmaktadır ve bu prensibin ihlal edilmesi halinde karşı tarafa sözleşmeyi feshetme hakkı doğmaktadır [9].

2.4. Sigorta Türleri

Sigorta türleri risk türlerinin artmasına bağlı olarak, zaman içinde sürekli çeşitliliğini arttırmıştır. Sigorta türleri yaygın olarak sigorta konusuna göre üç grupta sınıflandırılmaktadır:

2.4.1. Mal (Zarar) Sigortaları

Bir kişi veya kuruluşa ait maddi değeri olan ve parayla ölçülebilen mal varlığını tehdit eden risklerin sonuçlarına yönelik yapılan sigortalardır. Bu sigortayla, sigortalının uğradığı maddi zararı tazmin etmek amaçlanmaktadır. Bu sebeple mal sigortaları, zarar ya da tazminat sigortaları olarak da adlandırılmaktadır. Mal sigortalarının konusunu, parayla ölçülebilen ve maddi değeri önceden belirlenebilen her türlü hak, mal ve alacaklar oluşturmaktadır. Mal sigortalarının altında; yangın, mühendislik, kaza, nakliyat ve tarım sigortaları gibi çeşitli birçok alt branş bulunmaktadır [9].

Bu çalışmada da üzerinde çalışılan veriler “Kara Araçları Kasko Sigortası” branşına aittir. Bu sigorta branşı kısaca “Kasko Sigortası” olarak da bilinmektedir. Kasko sigortası, sigortalının, poliçede belirtilen ve karayolunda kullanma izni olan motorlu ve motorsuz kara araçlarından, römork veya karavanlardan iş makinelerinden, lastik tekerlekli traktörler ve diğer zirai tarım makinelerinden doğan menfaatinin çeşitli risklerin (çarpma, çarpışma, yanma, çalınma vs.) gerçekleşmesi sonucunda uğrayacağı maddi zararları teminat altına almaktadır [11]. Kasko sigortası zorunlu bir sigorta türü olmayıp, sigortalının isteği doğrultusunda teminatları genişletilebilir.

2.4.2. Can Sigortaları

Can sigortaları; ölüm, hastalanma, sakatlık, yaşlanma, kaza geçirme gibi insan hayatına yönelik riskleri teminat altına almaktadır. Ferdi kaza, hayat ve sağlık sigortaları bu branş altında yer almaktadır.

2.4.3. Sorumluluk Sigortaları

Sigorta ettirenin sorumluluğundaki eylemler veya kazalar neticesinde üçüncü şahısların mallarında ve canlarında meydana gelen zararları teminat altına alan sigortalardır. Sorumluluk sigortaları ile sigorta ettiren üçüncü kişilere karşı sorumluluğunu kabul etmekte, ancak sorumluluğun sonucunda meydana gelecek zararları karşılama veya tazminat ödeme borcunu devretmektedir [9]. Zorunlu sigortalar genellikle bu sigorta türü altında yer almaktadır. Bu tür sigortalara örnek olarak, hukuki sorumluluk sigortası, tehlikeli maddeler sorumluluk sigortası ve karayolları motorlu araçlar zorunlu mali sorumluluk (trafik) sigortası verilebilir.

Bu bölümde, risk ve sigorta gibi temel kavramların yanı sıra çalışmanın konusu olan mal (zarar) sigortaları altında yer alan kasko sigortasından bahsedilmiştir. Bu sayede ilerleyen bölümlerde çalışmada kullanılan veri yapısının anlaşılabilmesi için kasko sigortasının yapısına ilişkin kısa da olsa bilgi verilmesi amaçlanmıştır. Aynı zamanda, iyi niyet prensibine olan aykırılığın ve moral riskin varlığının çalışmanın asıl odak noktası olan sigorta sahtekârlıklarıyla olan bağlantısı kurulmuştur. Bir sonraki bölümde de sigorta sahtekârlığı kavramı detaylandırılmaktadır.

3. SİGORTA SAHTEKÂRLIĞI KAVRAMI

Bu bölümde sigorta sahtekârlığı kavramının, sahtekârlık riskinin ne olduğu ve sigortacılıkta sahtekârlık türlerinin neler olduğuna değinilerek çalışmada tespit edilmesi hedeflenen sahtekârlık türünün hangi sınıfa girdiği ve otomobil sigortasında sık görülen sahtekârlık şekilleri anlatılmaktadır. Aynı zamanda sahtekârlıkların sigorta sektörüne olan etkisi ve neden bu alanda bir çalışma yapılması ihtiyacının olduğundan bahsedilmektedir.

3.1. Sigorta Sahtekârlığı Nedir?

Genel anlamıyla sahtekârlık, hukuka aykırı ve haksız bir biçimde fayda elde etme amacıyla yapılan eylem olarak tanımlanabilir. The Association of Certified Fraud Examiners (ACFE) sahtekârlığı; kişinin, bir kuruluşun kaynaklarını veya varlıklarını kendini zenginleştirmeye yönelik kasıtlı olarak ve kötüye kullanarak yapmış olduğu uğraşlardır şeklinde tanımlamıştır [12].

Sigorta özelinde düşünüldüğünde sahtekârlık; haksız kazanç sağlanması, sigortacıya kasti olarak gerçeğe aykırı beyanda bulunulması, sigortacının alacağı finansal kararları etkileyecek gerçeklerin gizlenmesi, sorumluluğun veya güven ilişkisinin kötüye kullanılması gibi yollarla yapılabilmektedir [13]. Bir başka deyişle sigorta sahtekârlığı; “Haksız kazanç elde etmek amacıyla, sigortalı tarafından sigortacının bilerek, kasıtlı olarak aldatılmasıdır.” şeklinde tanımlanabilir [2]. Bu durumda sigortalı, sigortanın temel prensiplerinden “azami iyi niyet” ilkesine uymamış olur. Bir önceki bölümde de anlatıldığı gibi sigorta sözleşmeleri karşılıklı güven üzerine kurulduğundan bu prensibin ihlal edilmesi sözleşmeye temelinden zarar vermekte, sigortalı özelinde tazminata hak kazanamamaya kadar sonuçları olabilmektedir.

Sahtekârlık hareketlerinin arkasındaki temel güdü kısaca, yasal olmayan yollardan kazanç sağlamaktır diyebiliriz. Bu durum da ekonomi, hukuk ve ahlaki değerler üzerinde büyük olumsuz etkiler yaratmaktadır. Sigortacılık, güven ilişkisine dayalı doğası nedeniyle sahtekârlığa elverişli bir sektördür.

Ayrıca, kişilerin sigorta sahtekârlığını diğer alanlara göre daha düşük riskli görmesi; soygun, kaçakçılık vb. suçlara göre daha güvenli bulmaları da bu alandaki sahtekârlığın yaygın

olmasının sebepleri arasındadır. İnsanların sigorta sahtekârlığına daha kolay yönelmeleri; kolay yoldan kazanç sağlayabilmeleri, yakalanma riskinin az olması ve yakalansa bile sonuçlarının çok ağır olmayacağı düşüncesinden kaynaklanmaktadır [14]. Aynı zamanda bu alanda yapılan hafif sahtekârlıklardaki içgüdülerden bir tanesi de “verilen primin veya paranın karşılığını alabilmek” olduğundan çoğu zaman sigortalılar tarafından bir sahtekârlık olarak bile görülmeyebilir. Kasko sigortasından örnek verecek olursak, kişi bir kaza sonrasında sigortadan karşılanacak hasarı için tamirhaneye başvurduğunda kaza öncesi hasarlarını da göstererek sigortanın olduğundan fazla hasarı karşılamasını talep edebilmekte ve bu durum kişi tarafından çoğunlukla bir sahtekârlık olarak değerlendirilmemektedir.

Sigorta sahtekârlıkları hem sigorta şirketini hem de diğer sigortalılar gibi ekonomideki farklı aktörleri olumsuz etkilemektedir. Haksız yere tazminat alan kişilerin sigorta şirketinin mali dengesini olumsuz yönde etkilemesi sebebiyle, yapılan aktüeryal hesaplamalarda sapmalar meydana gelir, hasar için ayrılan karşılıklar yetersiz kalabilir, günün sonunda da sigorta maliyetlerinin artması nedeniyle bu maliyetin bir kısmı sigorta primlerine yüklenerek suçsuz olan sigortalılar da daha fazla prim ödemek zorunda kalabilir. Benzer şekilde sigorta şirketleri, bu tazminat talepli sahtekârlıkları önlemek amacıyla hasar departmanlarında daha fazla personel çalıştırarak yine sigorta maliyetlerinin artmasına sebep olabilir.

2012 yılında Amerika Birleşik Devletleri’nde (ABD) otomobil sigortalarında bedeni yaralanmalara ilişkin sahte hasar tazminatlarının 5,6-7,7 milyar dolar civarında olduğu ve bunun tüm bedeni hasar tazminatlarının yaklaşık %15’ine karşılık geldiği tahmin edilmektedir. Poliçe düzenlenmesi esnasında gerçeğe aykırı beyan sonucunda olması gerekenden daha düşük bir prim ödeyen sigortalılar sebebiyle otomobil sigorta şirketlerinin 2009 yılında 15,9 milyar dolar kaybettiği tahmin edilmekte, bu da bu branşta yazılan primlerin yaklaşık %10’unu oluşturmaktadır. Hayat dışı sigorta branşında ise toplam sigorta sahtekârlığı maliyetinin yılda 40 milyar dolardan fazla olduğu tahmin edilmektedir [3]. Avrupa sigorta şirketleri birliği olan “Insurance Europe” ise Avrupa’daki sahte hasarların maliyetinin toplam hasarların %10’u kadar olduğunu açıklamıştır [5].

Bir önceki bölümde de detaylı anlatıldığı üzere sigorta sözleşmesi; sigortalının sigortacıya prim ödeyerek riskini transfer ettiği ve sigortacının da bu ödeme karşılığında teminat altına aldığı

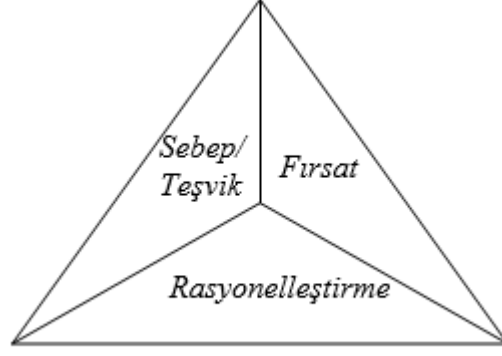
riskin meydana gelmesi halinde tazminat ödemeyi taahhüt ettiği bir sözleşmedir. Bu sözleşme kapsamında işlem yapan tüm tarafların yasal olarak karşılıklı bilgi paylaşımını azami iyi niyet ilkesine uygun bir şekilde yapması gerekir. Taraflar için eşit şekilde bağlayıcılığı olan azami iyi niyet ilkesi, sözleşmenin sonuna kadar sürdürülmelidir. Diğer taraftan, iyi niyet eksikliği, her zaman sahtekârlık anlamına gelmez. Kesin özellikleri yasal sistemlere göre değişse de, sigorta sahtekârlığından bahsedilebilmesi için literatürde kabul edilen şekliyle şu unsurların var olması beklenir:

- Gerçeğe aykırı beyanda bulunma.
- Aldatma kastının bulunması.
- Yasal olmayan fayda elde etme amacı olması [14].

Bu unsurlardan bir ya da birkaçının eksik olması halinde ortaya çıkan durum sahtekârlık değil, sigorta suistimali ya da yanlış sigorta uygulaması şeklinde adlandırılabilir. Ancak, literatürde sahtekârlık kavramı için gereken unsurlar belirtilmesine karşın, tüm durumlar; sahtekârlık, suistimal, yanlış sigorta uygulaması gibi ayrımlara gidilmeden “sahtekârlık” başlığı altında çalışmalar yapılmıştır. Çalışma kapsamında yapılan analizde de, sigortanın kötüye kullanımına ilişkin terimlerde bir ayrıma gidilmeyerek “sigorta sahtekârlığı” kavramı çerçevesinde genel bir değerlendirme yapılmıştır.

3.2. Sahtekârlık Riski ve Sahtekârlık Üçgeni

İnsanların neden sahtekârlığa karıştıklarına yönelik 1950’li yıllarda “sahtekârlık üçgeni hipotezi” ortaya atılmıştır. Bu hipoteze göre, sahtekârlığın meydana gelmesi için üç unsurun bir arada bulunması gerekmektedir. Bunlar: Sebep/Teşvik, Fırsat ve Rasyonelleştirme’dir. (Şekil 3.1)



Şekil 3.1. Sahtekârlık üçgeni

İnsanları sahtekârlığa yöneltecek birçok sebep olabilir. Sahtekârlığa yönelten bu teşvik, maddi zorluk içerisinde olmak ve çeşitli nedenlerle daha fazla gelir elde etme isteği gibi finansal sebeplerin yanı sıra hırs ve aile baskısı gibi sosyal sebeplerden de kaynaklanabilir. Bununla birlikte suça karışacak kişiler sahtekârlık yapma şansına yani fırsata sahip olmalıdır. Sigortacılık sektörünün gelişmesi, sigorta poliçelerinin çeşitlenmesi ve yeni ürünlerin çıkmasıyla birlikte, sahtekârlık için ortaya yeni fırsatların doğmasına bu da artan miktarlarda ve farklı türlerde sahtekârlıkların meydana gelmesine neden olmaktadır. Tespit olasılığının düşük olduğu düşünülen alanlarda insanların sahtekârlık yapma fırsatları artmaktadır. Sigorta şirketlerinin ve denetleyici otoritelerin; tarafların sahtekârlık yapma teşvikiyle ya da durumu rasyonelleştirmeleri konusunda yapabilecekleri fazla bir işlem bulunmazken, fırsatları sınırlandırarak sahtekârlığı belli bir ölçüde azaltabilmeleri mümkündür [13].

Rasyonelleştirme ise kişilerin yaptığı sahtekârlığı akla uygun hale ve haklı duruma çıkarmanın zihinsel sürecidir. Sigortalının poliçe dönemi içerisinde devrettiği risklerin gerçekleşmemesi sebebiyle ödediği primlerden beklediği karşılığı ya da tazminatı alamaması ve “herkes aynı şeyi yapar” mantığıyla diğer insanların yanlış yaptığı davranışları kopyalaması gibi durumlar yapılan sahtekârlığı mantıklı bir çerçeveye sokmak için üretilen bahanelerdir [13].

Bu temel unsurlar çerçevesinde bakıldığında, iki tip sahtekâr profili bulunur. İlki “fırsatçı sahtekârlık” yapan kişilerdir. Bu tür sahtekârlığa karışan kişiler, normalde yasalara saygılı ve suç işlemeyen bireyler olmasına rağmen sigortanın sunmuş olduğu fırsat ortamından

yararlanırlar. Örneğin, sigorta şirketinin kendisine göre daha fazla mali güce sahip olduğu ve önceki yıllarda tazminat talebinde bulunmadığı düşüncesiyle ödediği primleri telafi etmek için meydana gelen bir hasarı poliçe teminat kapsamındaymış gibi göstererek haksız kazanç elde eder. Bir diğer profil olan “profesyonel sahtekarlık” ise bu yolla gelir elde eden, yakalanana kadar birden fazla sigorta şirketini hedef alan, daha karmaşık ve kapsamlı sahtecilik yapabilen ve çoğunlukla organize suç örgütleri şeklinde hareket eden kişilerdir. Bu tür sahtekârlıkta elde edilen kaynaklar çoğunlukla başka suçları finanse etmede kullanılır [13]. Azaltılması en zor olan sahtekârlık türü de fırsatçı sahtekârların yaptığı “hafif dolandırıcılık” olarak da adlandırılan eylemlerdir. Çünkü her bir şüpheli olayı tespit etmenin maliyeti sahtekârlıkların maliyetinden daha yüksektir ve kişilerin daha önceden bir suça karışmamış olması eylemlerin tespit edilmesini güçleştirir. Profesyonel sahtekârların yaptığı “ağır dolandırıcılık” olarak adlandırılan eylemler ise sahtekârlık önlemeye yönelik alınan tedbirlerle önlenemezler [15].

3.3. Sahtekârlık Türleri

Literatürde farklı açıdan sınıflandırmalar bulunmakla birlikte, genel olarak International Association of Insurance Supervisors (IAIS) tarafından yapılan sınıflandırma esas alınmaktadır. Diğer sahtekârlık türleri ve örneklerine ilişkin çeşitli başlıklar altında bilgi verilmektedir. IAIS tarafından yapılan sınıflandırma aşağıdaki gibidir.

- **Dâhili (Şirket İçi) Sahtekârlık:** Sigorta şirketinde çalışan veya burada yetkili kişilerce haksız menfaat elde etmeye yönelik yapılan hileli hareketler dâhili sahtekârlık olarak adlandırılmaktadır [14]. Yönetim kurulu üyeleri, yöneticiler ve çalışanların işleri gereğince sahip oldukları bilgilere erişim ve onay yetkilerini kötüye kullanılması sonucunda ortaya çıkmaktadır. Bütün sektörlerde görülebilecek şirket kaynaklarının çalınması, rüşvet ve gayri resmi komisyon alınması gibi idari suistimallerin yanı sıra sigorta şirketlerinde en sık karşılaşılan sahtekârlık örneklerden birisi de hasar departmanında çalışanların dışarıdan kişilerle de işbirliği yaparak sahte hasar ödemesi yapması durumudur [13].

- **Tazminat Talebine Yönelik (Sigortalı Kaynaklı) Sahtekârlık:** Bu sigorta sahtekârlığı, sigortalılar tarafından sigorta sözleşmesi düzenlenirken, sigorta sözleşmesinin yürürlükte olduğu dönemde veya tazminat talebi ödemeleri sırasında gerçekleşebilmektedir [13]. Sigortalı poliçe düzenlenirken, hasar geçmişi veya daha önce başvurduğu sigortacılar tarafından teminat talebinin reddedilmesi nedeniyle, sigortacı tarafından talep edilen bazı bilgileri kasten gizlemekte veya hatalı bilgi verebilmektedir. Bu durum sigortacının, normal koşullarda yapmayacağı bir sözleşme ilişkisine girmesine veya daha düşük bir fiyatla riski üstlenmesine neden olarak, sigortacının öngöremediği finansal bir yükün altına girmesine yol açmakta ve sigortacılık sektörünün mali dengesini bozmaktadır. Hasar sonrasında ya da tazminat talebi esnasında yapılan sahtekârlık ise sigortalının kasıtlı olarak hasar meydana getirmesi, meydana gelmiş hasarı olduğundan fazla göstermesi, sigorta teminatı kapsamında olmayan bir hasarı yalan beyanda bulunarak ya da kurgulayarak teminat içerisinde göstermesi, sahtekâr kişinin hasara uğrayan taraf hakkında gerçeğe aykırı beyan vermesi gibi birçok örnekte görülebilmektedir. Ayrıca, tazminat talebi kaynaklı sahtekârlık, diğer sahtekârlık çeşitleriyle birlikte ortaya çıkabilmektedir. Örneğin yukarıda sıralanan durumlar sigorta şirketinde çalışan biriyle veya bir sigorta aracısıyla organize şekilde de gerçekleşebilir.
- **Aracı Kaynaklı Sahtekârlık:** Sigorta aracıları, 5648 sayılı Sigortacılık Kanunu'nda sigorta acentesi ve brokeri olarak tanımlanmıştır. Sigorta acentesi, "Sigorta şirketine bağlı olarak mesleki faaliyette bulunan, yetki ve sorumlulukları, yasa, yönetmelikler veya sözleşme ile belirlenmiş sigorta aracısıdır." Sigorta brokeri ise, "Kişilerin sigorta ihtiyaçlarına yönelik olarak onlara finansal tavsiyede bulunmak, onlar adına sigorta şirketiyle görüşmeler yapmak, hasarın belirlenmesinde sigortalı adına hareket etmek gibi hizmetlerde bulunan yetki ve sorumlulukları yasa, yönetmelikler veya sözleşme ile belirlenmiş sigorta aracısıdır." [2]. Kısaca acente sigortacının, broker ise sigortalının temsilcisidir denilebilir. Aracılar sigortacılık işlemlerinde ve süreçlerindeki yetkileri ve işlemleri bakımından önemli ve güven teşkil eden bir konuma sahiptirler. Ancak güvenin ve yetkilerin kötüye kullanılması riski de bulunmaktadır. Aracı kaynaklı sahtekârlıkta sık karşılaşılan örnek; acentenin sigortalıdan aldığı primi sigortacıya aktarmayarak bunun neticesinde teminatın başlamamasıdır. Sigortalı çoğu zaman, bir hasar gerçekleşmediği sürece bu durumun farkına varmaz. Bunun yansırı

acenteler, sigortalı veya üçüncü taraflarla işbirliği yaparak tazminat talebine yönelik de sahtekârlığa karışabilmektedir. Örneğin, sigortacıya yanlış bilgi vererek hasar meydana geldikten sonra poliçe düzenlemesi ve geriye dönük işlem yapması gibi [13].

Bu çalışma kapsamında tespit edilmesi amaçlanan sahtekârlık, otomobil sigortalarındaki tazminat talebine yönelik olan sahtekârlıklardır. Bu sebeple yukarıda bahsedilen temel sınıflandırmanın yanı sıra otomobil sigortasında sık karşılaşılan sahtekârlık örneklerine değinilmektedir.

3.3.1. Otomobil Sigortasında Karşılaşılan Sahtekârlık Örnekleri

Otomobil sigortaları iki temel menfaatin korunmasına yönelik yapılır. Bu menfaatler kişilerin kendi araçlarının göreceği fiziksel zararların (mal sigortası) ve kaza nedeniyle üçüncü kişilerin araçlarına vereceği fiziksel zararlar ile bu araçtaki kişilerde oluşacak bedeni yaralanmalara bağlı olarak ortaya çıkan zararların (sorumluluk sigortası) telafisi için yapılan sigortalardır. Bu alandaki sahtekârlıklar, söz konusu durumlar suistimal edilerek haksız kazanç elde etmeye yönelik yapılmaktadır. Ayrıca bu alanda sigortacılık sektörüne verilen zararlara, sahip olduğu yetkiyi kötüye kullanan acenteler, tamirhaneler, doktorlar ve avukatlar da katkıda bulunabilmektedir.

- **Hırsızlık bahanesiyle gerçekleştirilen sahtekârlık:** Aracın çalındığı iddia edilerek sigortadan haksız tazminat talebinde bulunması eylemidir. “Araçtan vazgeçme” diye de adlandırılabilen bu örnekte, kişinin nakde ihtiyacı olması durumunda, sigorta poliçesinden bunu karşılamak amacıyla otomobilini bir yerde terk ederek başka kişiler tarafından aracın çalınmasını ya da parçalarına ayrılarak imha edilmesini bekleyerek kişinin sigorta şirketine durumu aracı çalınmış gibi rapor etmesidir. Aynı zamanda kişiler araçlarını bir şekilde saklayarak da aracının çalındığına dair sigorta şirketine tazminat başvurusu yapabilmektedir.
- **Sonradan sigorta yaptırma:** Sık karşılaşılan bir örnekse, sonradan sigorta yaptırma sahtekârlığıdır. Kişi, sigortası olmayan otomobiliyle bir kaza yapar ve bu durumu sigortacıya bildirmeden hasarlı aracına sonradan sigorta yaptırır. Bir süre bekledikten sonra hasar ihbarında bulunur ve kaza sigortadan sonra gerçekleşmiş gibi durumu rapor

ederek tazminat talebinde bulunur [12]. Söz konusu durum bireysel olarak yapılabildiği gibi kötü niyetli acente yardımıyla da yapılabilmektedir.

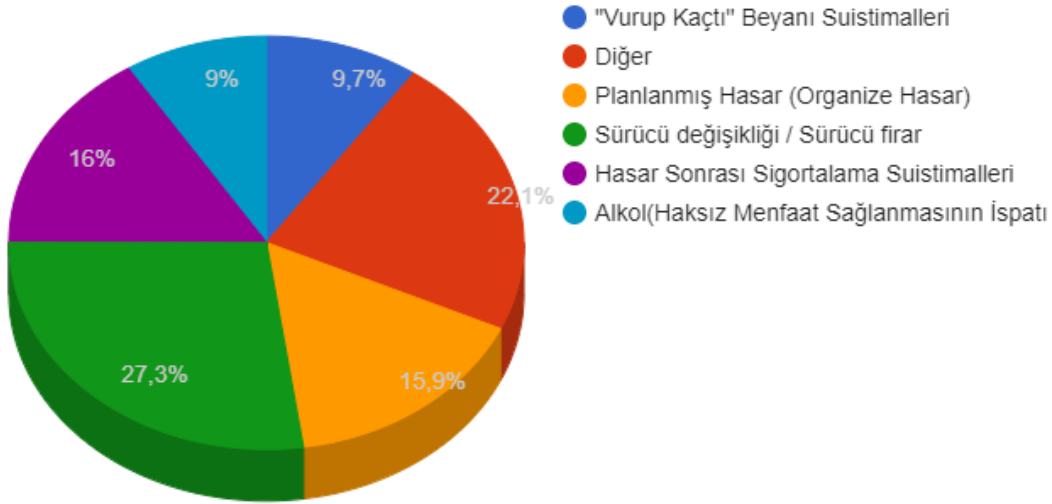
- **Kazaya sebep olarak yapılan sahtekârlık:** Kötü niyetli kişilerin kendi sigortalarından ziyade karşı tarafın (kurbanın) trafik sigortasını hedef almaları durumunda sık başvurdukları yöntemlerdir. Bunun için farklı yollar kullanarak kurbanın kendi aracına çarpmasına neden olmaya çalışırlar. Örneğin, aniden fren yaparak arkadaki aracın kendisine çarpmasına sebep olur ya da trafikte daha avantajlı durumda gözükebilecekleri hallerde diğer aracın önüne kırarak çarpışmaya neden olurlar [16]. Bu şekilde karşı tarafı hatalı göstererek onların trafik sigortasından yararlanıp araçlarındaki eski hasarları tamir ettirebilirler, daha kötüsü araçta dahi olmayan kişilerin bulunduğunu iddia ederek bu kişilerin yaralanmalarına ilişkin yüksek tutarlı tazminat taleplerinde de bulunabilirler.
- **Kendi aracına zarar vererek yapılan sahtekârlık:** Hırsızlık bahanesiyle yapılan sahtekârlıklarla benzer amaçla; yüklü nakit ihtiyacı, büyük bir krediyi kapatmak ya da satılamayacak durumda olan araçtan kurtulma güdüsüyle yapılan sahtekârlıklardır. Çoğunlukla kendi aracını kundaklama ya da bunun için başka biriyle anlaşma şeklinde gerçekleştirilir. Kişi sonrasında aracının “kendiliğinden yanmaya başladığı” ya da “aracının çalınarak başkaları tarafından yakıldığı” bahaneleriyle sigorta şirketinden haksız tazminat talebinde bulunurlar [16].
- **Kâğıt üzerinde yapılan sahtekârlık:** Bu tür sahtekârlıkta gerçekte meydana gelen herhangi bir kaza veya yaralanma olmamasına rağmen kişiler, bazen tamirhaneye de anlaşarak kâğıt üzerinde düzenlenilen sahte raporlarla kaza gerçekleşmiş gibi haksız tazminat talebinde bulunmaktadır. Bu tür tazminat talepleri genelde, polis raporu gerektirmeyecek olması ya da sigorta şirketlerince eksper görevlendirilmemesi sebebiyle çok yüksek tutarlı değildir [17].

3.3.2. Türkiye’de Sigorta Sahtekârlıkları

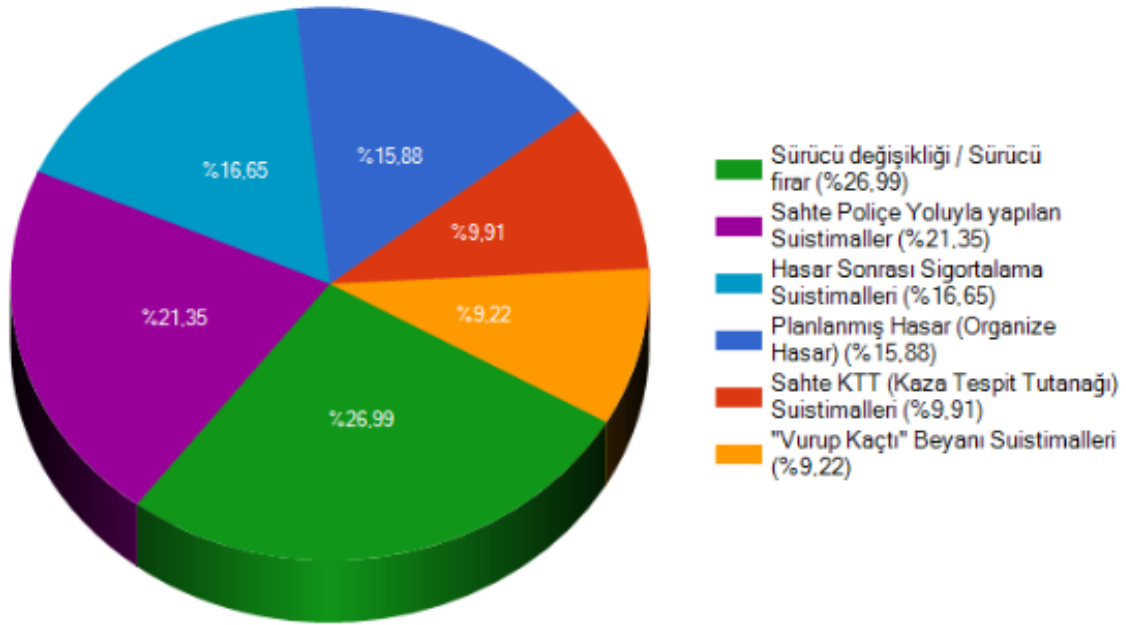
Sigorta sahtekârlıkları, Türkiye’de sigortacılık sektöründeki önemli sorunlardan biridir. Dünya’da olduğu gibi ülkemizde de bu alanda yapılan sahtekârlıklar ekonomiyi olumsuz yönde etkiledikleri gibi bunların maddi külfeti sigorta şirketlerine ve dürüst sigortalılara yansımaktadır. Sigorta sahtekârlıklarının boyutu ülkelere ve sigorta branşlarına göre farklılık

göstermekle birlikte, sahte hasarların ödenen tüm hasarlar içindeki oranının hayat dışı branşta % 10 civarında olduğu tahmin edilmektedir. Türkiye’de 2014 yılında kasko ve trafik sigortası branşlarında ödenen ve muallak olan toplam 11 milyar TL’lik hasarın, 1.4 – 1.7 milyar TL’sinin suistimal olduğu düşünülmektedir [4].

Şekil 3.2.’de 2018 yılında ve Şekil 3.3.’te Nisan 2018-Nisan 2019 tarihleri arasında otomobil sigortasındaki sahtekârlıkların nedenlerine göre dağılımı gösterilmektedir. Buna göre, 2018 yılındaki otomobil sigortalarındaki sahtekârlıkların %27,3’ü sürücü değişikliği veya sürücünün firar etmesi şeklinde, %16’sı hasar sonrasında poliçe yaptırma, %15,9’u planlanmış hasar şeklinde gerçekleşmiştir. Nisan 2018-Nisan 2019 tarihleri arasındaki otomobil sigortası sahtekârlıklarının %26,99’u sürücü değişikliği veya sürücünün firar etmesi şeklinde, %21,35’i sahte poliçeyle, %16,65’i hasar sonrasında poliçe yaptırma, %15,88’i planlanmış hasar şeklinde gerçekleşmiştir.



Şekil 3.2. 2018 yılında otomobil sigortası hasar dosyalarındaki sahtekârlıkların dağılımı[18].



Şekil 3.3. Nisan 2018-Nisan 2019 tarihleri arasında otomobil sigortası hasar dosyalarındaki sahtekârlıkların dağılımı[18].

Türkiye’de sigorta sahtekârlığıyla mücadele konusunda sigorta şirketleri kendi bünyelerinde birimler kurmakta, sigortacılık alanında düzenleyici ve denetleyici otorite olan Hazine ve Maliye Bakanlığı tarafından ilgili mevzuat çıkarılmakta ve denetim yapılmaktadır. Bu noktada sigorta şirketlerinin verilerinin toplandığı SBM’nin daha etkin bir rolü vardır.

SBM’de Türkiye’de faaliyet gösteren sigorta şirketlerine ait; “hayat, hastalık/sağlık, zorunlu trafik, yeşilkart, zorunlu karayolu taşımacılık mali sorumluluk, otobüs zorunlu koltuk ferdi kaza, motorlu kara taşıtları (kasko), tıbbi kötü uygulama zorunlu mali sorumluluk sigortası, tehlikeli maddeler ve tüpgaz zorunlu sorumluluk sigortası, mesleki sorumluluk sigortaları” branşlarına ilişkin veriler tutulmaktadır. Bu şekilde, söz konusu veriler tek merkezde toplanarak, sigortacılık faaliyetlerinin daha etkin bir şekilde yürütülmesi, sigorta şirketleri tarafından sağlıklı fiyatlandırma yapılabilmesi, sahtekârlıkların önlenmesi, sektör genelinde uygulama birliğinin sağlanması, sigortaya olan güvenin artırılması, sağlıklı istatistiklerin oluşturulması ve kamu gözetim-denetiminin etkinleştirilmesi hedeflenmektedir [19].

Sigorta sektöründeki sahtekârlıklarla mücadeleye destek vermek ve bu konuda ortaya çıkacak finansal kayıpları engellemek amacıyla SBM bünyesinde 2015 yılında SİSEB kurulmuştur. Sigorta şirketleri tarafından açılan hasar dosyalarına ilişkin bilgiler belirli aralıklarda buraya ulaştırılmakta ve çeşitli kriterlere göre yapılan skora işlemi sonucunda şüpheli olabilecek hasarlar sigorta şirketlerine bildirilmektedir. Aynı zamanda, üçüncü şahıslardan ve sigorta şirketlerinden sağlanan sigorta sahtekârlığına konu olabilecek veriler ise “Sigorta Suistimalleri Bilgi Paylaşım Sistemi” (SİSBİS) veri tabanında tutulmaktadır. Sigorta şirketleri şüpheli gördüğü hasarları SİSBİS’e bildirmekle yükümlüdür.

Sigorta şirketlerinin SİSBİS’e bildireceği sigorta sahtekârlığı örnekleri SİSEB tarafından aşağıdaki gibi örneklendirilmiştir:

- “Mahkemede karara bağlanmış sahtekârlık dosyaları.
- Savcılık soruşturmasına konu hasar başvuruları.
- Şirketçe ortaya çıkarılan sahtekârlık durumlarında sigortalıdan feragatname alındığı durumlar. (Sahte belgeli hasar başvuruları, gerçek dışı beyana dayalı şüpheli hasar başvuruları, hasardan sonra poliçe düzenlenmesi)
- Sahte hasar şüphesi ile reddedilen ve sigortalının şirketi dava ettiği durumlar.
- Alkollü araç kullanımına ilişkin durumlar. (Alkollü araç kullanımı nedeniyle reddedilen durumlarda sürücü riski bilgileri paylaşımı)
- Sürücü bilgilerine ilişkin durumlar. (Yetersiz sürücü belgesi nedeniyle reddedilen durumlarda sürücü riski bilgileri paylaşımı)
- Şirketlerin sahtekârlık şüphesiyle ek araştırma gereksinimi duyduğu durumlar” [20].

Özetle, sigorta şirketleri kendi bünyelerinde karşılaştıkları ya da tespit ettikleri sahtekârlıkları SBM’ye bildirmekte, buna ilave olarak sigorta şirketlerinden SBM’ye aktarılan poliçe ve hasar dosyası bilgilerine dayanarak SİSEB’te yapılan analizlerle elde edilen tespitler ilgili sigorta şirketleriyle paylaşılmaktadır. Bu şekilde hem şirketlerin kendi uzmanlıklarından yararlanılmakta hem de büyük veriyle sektör bazında analiz yapılmaktadır.

Bu alıřma kapsamında SBM'den alınan tm sigorta Őirketlerine ait kasko sigortası hasar dosyası verilerinden yararlanılmıřtır. Verilerin analizinden nce (Blm 6), sigorta sahtekrlıęının tespitinde kullanılan alıřmalara ynelik literatr (Blm 4) ve analiz iin kullanılan yntemler (Blm 5) incelenecektir.

4. SİGORTA SAHTEKÂRLIĞI TESPİTİNDE YAPILAN ÇALIŞMALAR

Günümüze kadar araştırmacılar tarafından sahtekârlık tespitine yönelik bankacılık sektöründen sonra en çok çalışma yapılan alan sigortacılık sektörüdür. Sigorta sahtekârlığı, konusu itibariyle diğer alanlarla da ilişki içerisinde olması sebebiyle ilgi gören bir çalışma alanı olmuştur. Bu çalışmalar sağlık sigortası, otomobil sigortası, konut sigortası ve tarım sigortası gibi alt branşlardaki sahtekârlık tespit yöntemleri olmak üzere çeşitlenmiştir. Bu bölümde, ağırlıklı olarak otomobil sigortası olmak üzere, sigorta alanında yapılan sahtekârlıkların tespitine yönelik kullanılan yöntem ve yaklaşımlar incelenecektir.

1980'lerin başında sigorta sahtekârlıklarının tespiti ilgi çeken bir konu olmaya başlamış, bu alandaki ilk akademik çıktı 1992 yılında Cummins ve Tennyson'ın otomobil sigortalarındaki sahtekârlıklar üzerine yaptıkları çalışma ile başlamıştır. Bunu, işsizlik sigortası, sağlık sigortası, mal ve sorumluluk sigortalarındaki sahtekârlıklara ilişkin yapılan akademik çalışmalar izlemiştir [21]. Aynı zamanda bu yıllarda veritabanı organizasyonu ve seçim stratejileri, bulanık kümeleme, basit regresyon skorlama modelleri ve probit modelleri ile sahte hasar taleplerini belirlemek için pratik modeller ortaya çıkmaya başlamıştır.

Yapılan ilk çalışmalarda, otomobil sigortasındaki hasar ihbarlarının büyük bir bölümünde sahtekârlık ya da hasarın olduğundan fazla gösterilmesi gibi sonuçların olduğu çıkarılmıştır. Salsas-Forn ve Tennyson'ın [22] çalışmasında otomobil sigortasındaki sahtekarlıkların tespitine yönelik genel bir inceleme yapılmış, bu konularda ampirik kanıt sağlamak için, otomobil sigortası piyasasındaki hasar denetimini analiz etmiştir. Çalışmada, bireysel otomobil sigortası ihbarlarından ve bu ihbarların sigorta şirketi tarafından incelenmesi sonrasındaki bilgileri içeren bir veri setinden yararlanılmış ve otomobil sigortası kapsamında bedeni yaralanma teminatına ilişkin Massachusetts Otomobil Sigortacıları Bürosu'ndan elde edilen veriler ile hasar ihbarları incelenmiştir. Çalışmada, tazminatların ortalaması, minimum-maksimum aralığı, standart sapması gibi basit istatistiksel dağılımı çıkarıldığında bile hak sahiplerinin tazminata hak kazanabilmek amacıyla, tedavi masraflarının geçmesi gereken eşik değer kriterini sağlaması için masrafları yüksek göstermesi gibi moral risk içeren hareketlere büyük bir eğiliminin olduğu

görülmüştür. Buna ek olarak, hasar ihbarları üzerine kurulan lojistik model ile hasar ihbarlarında yapılan denetimlere ilişkin bir hasar hangi özelliklere sahipse denetime girmeli, denetime giren hasarlarda ne oranda haksız kazanç elde etme amacı bulunuyor gibi sigortacılara tavsiye niteliğinde sonuçlar sunulmuştur.

Bu alanda yapılan ilk çalışmalarda yukarıda da bahsedildiği gibi manuel ya da otomatik yaklaşımlar kullanılmış, ancak zamanla bu yöntemlerin etkili olmadığı, çeşitli maliyet ve kısıtlar getirdiği görülmüştür. Dolayısıyla bu alanda gelişmiş ve etkin bir yaklaşım kullanılması ihtiyacı doğmuştur. Gelişmiş ve etkin bir sahtekârlık tespit sisteminin oluşturulabilmesi için çeşitli istatistiksel yöntemleri içeren daha karmaşık yapıların kullanılması gereklidir. Sahtekârlık tespitinde kullanılan makine öğrenmesi ya da örüntü algılama yöntemlerine dayalı istatistiksel yöntemlerin başlıca avantajları aşağıdaki gibidir:

- 1) Sahtekârlık desenlerinin veriden otomatik olarak öğrenilmesi.
- 2) Her bir olay özelinde sahtekârlık olasılığının belirlenerek şüpheli durumların araştırma süreçlerinde önceliklendirilmesinin sağlanması.
- 3) Daha önce tespit edilemeyen yeni dolandırıcılık türlerinin belirlenmesi [23].

4.1. Sigorta Sahtekârlığı Tespitinde Kullanılan Yöntemler

Sahtekârlıkların tespitinde istatistiksel olarak kullanılan yöntemler üç sınıfa ayrılmaktadır: gözetimli (supervised) yöntemler, gözetimsiz (unsupervised) yöntemler ve yarı gözetimli (hibrit) yöntemler. Literatüre baktığımızda otomobil sigortasındaki sahtekârlık tespitine yönelik kullanılan istatistiksel yöntemler aşağıdaki gibi incelenebilir.

4.1.1. Gözetimli (Supervised) Yöntemler

Gözetimli yöntemlerde, veri tabanındaki sahtekârlıklara ait mevcutta bulunan etiketlenmiş veriler ve uzman bilgisi kullanılarak tespit yapabilecek bir sistem oluşturulur. Sistem bu verilerden bir öğrenme gerçekleştirerek sonuçlar sunar.

Bu yöntem, "sahtekârlık" ve "sahtekârlık olmayan/gerçek" olarak etiketlenmiş ve sınıflandırıcı eğitimini içeren bir veri kümesi gerektirir. Bu, en yaygın öğrenme yaklaşımıdır. Gözetimli öğrenmenin en büyük avantajı, bu yaklaşımın algoritma tarafından işlenen tüm sınıf çıktılarının insanlar için anlamlı olması ve sınıflandırma ile regresyon modelleri için kolayca kullanılabilir olmasıdır [6].

Gözetimli yöntemler verinin yapısına göre regresyon veya sınıflandırma problemlerinin çözümlenmesi için kullanılır. Regresyon problemlerinde, yanıt değişkeni sürekliyken, sınıflandırma problemlerinde yanıt değişkeni kategoriktir. Sahtekârlık tespiti de bir sınıflandırma problemi olarak ele alınabilir. En bilinen ve sık başvurulan sınıflama yöntemleri; Lojistik Regresyon, Karar Ağaçları, Yapay Sinir Ağları, Bulanık Mantık, Destek Vektör Makineleri, Naive Bayes Sınıflandırıcısı, Bayes Ağları, K-En Yakın Komşu Sınıflandırıcısı olarak sayılabilir.

Bununla birlikte, gözetimli öğrenmenin bazı kısıtlamaları vardır. Bunların en önemlisi, verilerin etiketlenmesinde yaşanan zorluklardır. Büyük miktardaki verilerle çalışıldığında, verinin hepsinin etiketlenmesi imkânsız olmasa bile pahalıya mal olur. Diğer bir problem ise, bazen ayırt edici bir etiket bulmanın son derece zor olması veya etiket bilgilerinde belirsizlikler olmasıdır. Bahsedilen bu kısıtlamalar bazı durumlarda gözetimli öğrenme yaklaşımlarının uygulanmasına engel olabilir [6]. Literatürde otomobil sigortası sahtekârlık tespiti çalışmalarına yönelik gözetimli öğrenme yöntemleriyle yapılan çalışmalar aşağıdaki gibi sıralanabilir.

- **Lojistik Regresyon**

Artis ve arkadaşları [24], sigorta sahtekârlıklarının tespit edilebilmesi için sigortalı davranışlarının analiz etmenin gerekli olduğundan yola çıkarak, sigortalı ve hasar özelliklerinin sahtekârlık yapma ihtimaline olan etkisini tahmin etmek için çalışmada ayrık-seçimli modeller kullanmıştır. İspanya'daki otomobil sigortasındaki (zorunlu trafik sigortası) hasar verilerinin kullanıldığı çalışmada, veriler temin edildiği sigorta şirketi tarafından hali hazırda gerçek ve sahte hasar olarak sınıflandırılmış şekildedir. Sahte hasarlar, kişilerin durumu kabul/itiraf etmesi neticesinde elde edilen bilgiler sonucunda etiketlenmiştir. Kurulan çoklu lojistik regresyon modeli çerçevesinde, her bir hasar ihbarı "gerçek hasar, kişisel çıkara yönelik

sahtekârlık ve üçüncü taraflar çıkarına sahtekârlık” olmak üzere 3 sınıfa ayrılmıştır. Çalışmada gerçek hasar ihbarlarını, İspanyol pazarında en çok görülen bu iki sahtekârlıkla karşılaştırmak için sınırlandırma yapılmıştır. Sonuçlar, incelenen sahtekârlık davranışları etkisinin, sahtekârlığın sınıfına göre değiştiğini göstermiştir. Polis raporunun varlığının sahte uygulamaları önlediği, buna karşılık sigortalıların geçmiş hasarlardaki davranışları ve hasar sürecindeki tutumu üzerinde büyük bir etkiye sahip olduğu görülmüştür. Metodolojik bir nitelik taşıyan bu çalışmada, dolandırıcılık davranışının ekonometrik bir modele dayanan deneysel sonuçları sunulmuştur.

Artis ve arkadaşları [25] ile Belhadji ve Dionne [26], sigortalıların davranışlarına ilişkin önceki bilgilere dayanarak hasar dosyasında sahtekârlık olma olasılığını tahmin ettikleri çalışmalarında bu hasar dosyalarının yanlış sınıflandırılmasını da dikkate aldıkları bir lojistik regresyon modeli önermişlerdir. Artis ve arkadaşları [25], çalışmalarında sahtekârlık algılama için ikili seçim modellerinin performansını gösterir ve yanıt değişkenindeki yanlış sınıflandırma için bir model uygular. Gerçek ve sahte hasar ihbarları içeren İspanyol sigorta pazarından bir veri tabanı kullanılır. Bu makalede ele alınan soru: “Gerçek olarak etiketlenen tüm hasarlar doğru sınıflandırılmış mı?” dir. Sahtekârlık tespiti için uygulanan denetim teknolojisi kusursuz olmayabilir ve hasarlar nihayetinde yanlış sınıflandırılabilir. Yapılan çalışmalar çoğunlukla çeşitli modeller kullanarak, şüpheli koşulların varlığını tanımak ve denetçilerin sahtekârlıkları tespit edebilmesi için yeni araçlar sağlamak üzerine yoğunlaşmıştır. Ancak, kullanılan bu yöntemlere yönelik eleştirilerden biri, daha önceki hasar dosyalarına ilişkin bilgilerin kusursuz olmamasıdır. “Gerçek hasar” olarak sınıflandırılan dosyalarının, mevcut metodolojinin tanımlayamadığı sahte hasarlar olma ihtimali mevcuttur. Bu makalede, sahtekârlık tespiti için ikili (binary) seçim modellerinin, yanıt değişkeninin yanlış sınıflandırılmasındaki performansı gösterilmekte ve lojistik regresyon modeli tarafından tespit edilemeyen sahte ihbarların yüzdesi tahmin edilmektedir.

Belhadji ve Dionne [26], Quebec’teki en büyük 20 sigorta şirketinin otomobil sigortası verileriyle yaptığı çalışmada, sahtekârlık tespitinde anlamlı çıkan değişkenlerini belirledikten sonra oluşturduğu matematiksel denklem ve probit model içeren yazılımla, hasar dosyalarının sahtekârlık olasılığını belirleyen bir yöntem sunmaktadır. Veriler sigorta şirketi çalışanlarının

hasar dosyalarına ilişkin doldurduğu anketlerden elde edilmiştir. Veride sigorta sahtekârlığı yönünde işaretlenen dosyalar, kanıtlanmış veya eksperlerin bu anket doğrultusunda güçlü sahtekârlık şüphesi duyduğu dosyalar üzerinden belirlenmiştir. Sigorta eksperlerinin sahte hasar dosyalarını tespit etmesine yardımcı olması amaçlanan bu yazılımın sigorta şirketlerinin kendi özelinde daha doğru sonuçlar vermesi için kendi verilerini kullanarak yeniden önemli değişkenlerin ağırlıklarının belirlenmesi ve yazılımın çalıştırılması önerilmektedir.

Weisberg ve Derrig'in [27]; otomobil sigortalarındaki bedensel hasarlarla ilgili verileri kullandığı çalışmada, farklı sigorta sahtekârlığı türlerinin göstergelerini belirlemek için çoklu doğrusal regresyon modeli önerilmiştir. Çalışmada, "sigorta sahtekârlığı" kavramı netleştirilerek; "planlı sahtekârlık", "fırsatçı sahtekârlık" ve "mevcut hasarın yüksek gösterilmesi" olmak üzere 3 grup altında incelenmiştir.

Wen ve arkadaşları [28], sigortalıların ve hasar özelliklerinin sigorta sahtekârlığı yapma olasılığı üzerindeki etkisini tahmin etmek için Çoklu Logit ve İç İçe (Nested) Logit modelleri kullanmışlardır. Tayvan'da bir hayat dışı sigorta şirketinin otomobil (kasko) sigortası verileri kullanılan çalışmada daha çok kişilerin satın alma alışkanlıkları üzerinde durulmuş, buradan yola çıkarak sigortacılıktaki bilgi asimetrisi sonrasında ortaya çıkan "ters seçim"e neden olan etkenler ve sigortalı özelliklerine değinilmiştir. Örneğin; fazla sayıda hasar ihbarı yapan sigortalıların (yüksek riskli), teminat kapsamı geniş poliçeler olan A ve B paketlerini tercih ettikleri görülmüştür. Bu durum da sigortacılıkta ters seçim problemi olduğunu destekler niteliktedir. Çalışma sonucunda iki tür sigortalı davranışı gözlemlenmiştir: Birincisi araçlarının değeri çok yüksek olan ancak hiç hasar ihbarı bulunmayan sigortalılar. İkincisi ise poliçe bitiş tarihine yakın bir zamanda, bir tamirhaneye anlaşarak "kötü niyetli zarar verme ve vandalizm" sebepleriyle hasar ihbarında bulunan sigortalılar.

Otomobil sigortası sahtekârlıkları alanında, Artis ve arkadaşları [24, 25] ile Belhadji ve Dionne [26], Bayesci olmayan bir bakış açısıyla ayrık seçim modellerini tahmin etmiştir. Bu çalışmalarda [24–26], simetrik bağlara dayanan ağırlıklandırılmış bir olasılık fonksiyonunun maksimize edilmesiyle tahmin edilen ayrık modeller kullanılmıştır. Buna karşılık bu alanda Bayes yaklaşımıyla yapılan çalışma sayısı azdır [29].

Bermúdez ve arkadaşları [29] ise çalışmalarında, otomobil sigortası tazminat taleplerinde sahtekârlık tespiti için asimetrik ve çarpık Bayesyen Logit model kullanmıştır. Modelde, İspanya’da rastgele seçilmiş 10.000 otomobil sigortası tazminat talebi verisi üzerinden çalışılmıştır. Çalışmada, Standart Logit modelle buna karşılık gelen Bayes Logit modeli karşılaştırılmış, her iki modelin de parametre tahminleri ve uyum bakımından benzer sonuçlar verdiği sonucu elde edilmiştir. Ancak çalışmada önerilen ve yeni bir yaklaşım olan asimetrik bağıın kullanılmasının, Standart Logit ve Simetrik Bayes yaklaşımına göre, model tahmininde doğru sınıflandırılan vakaların yüzdesini önemli ölçüde artırdığı ve genel uyumu iyileştirdiği sonucuna ulaşılmıştır. Bu durumun en önemli sebebi, verinin sahte ve gerçek hasar verileri bazında eşit ağırlıklı olmamasıdır, genelde sigorta sahteciliğine ilişkin etiketlenmiş verilerde gerçek (sahtecilik olmayan) hasar dosyalarının ağırlığı sahte hasar dosyalarına oranla çok yüksektir.

- **Karar Ağaçları**

Pérez ve arkadaşları [30]; bir otomobil sigortası şirketinde sahtekârlık tespiti yapmak için, çalışmada “Consolidated Trees” (CT) diye bahsi geçen bir Karar Ağacı yöntemi kullanmıştır. Makine öğrenmesi yöntemlerinin gerçek hayat problemlerinde doğru çalışması ve sınıflandırıcıların doğruluğu için verinin dağılımının önemli olduğu hususu üzerinde durulmuştur. Sınıflandırıcılar çarpık (dengesiz) verilerle eğitildiklerinde çok sağlıklı çalışmazlar, örneğin veri kümesindeki örneklerin çok büyük bir kısmı bir sınıfa ait olduğunda sınıflandırıcı hata oranını en aza indirgeyecek şekilde çalıştığından doğruluğu da oldukça yüksek çıkacak ve sınıflandırıcılar öğrenirken örnek sayısının fazla olduğu sınıfa doğru bir eğilim gösterecektir. Dolayısıyla her alanın kendine özgü optimal bir sınıf dağılımı bulunmaktadır. Veri setini bu dağılıma uygun hale getirmek için “oversampling” (örneklerin tekrarlanması) ve “undersampling” (örneklerin elenmesi) gibi örnekleme teknikleri kullanılabilir. Bu çalışmada da, özellikle undersampling yönteminin çarpık veri setlerinde kullanılan etkin bir örnekleme metodu olmasına karşın bilgi kaybına yol açtığı, bu noktada kullanılabilecek “bagging” ve “boosting” gibi öğrenme yöntemlerinin de karar ağacının karmaşılaşması ve anlaşılabilirliğinin azalmasına neden olduğuna dikkat çekilmiştir. Çalışmada bu problemlerle başa çıkabildiği belirtilen CT yöntemiyle, istenen sınıf dağılımına sahip

örnekleme alt veri gruplarıyla oluşturarak aynı zamanda karmaşık olmayan ancak açıklayıcılığı yüksek tek bir karar ağacı sunulmaktadır. Bu yöntem ile bir otomobil sigortası şirketinin verileri üzerinde çalışılarak, hasar departmanındaki dosyaları inceleyen kişilere dosyada ne kadarlık bir sahtekârlık olasılığı olduğunu veren, yardımcı bir araç önerilmektedir.

- **Yapay Sinir Ağları**

Viaene ve arkadaşları [31] ile Bermúdez ve arkadaşlarının [29] çalışmalarında bazı hibrit algılama modelleri de kullanılmıştır. Tahmin doğruluğunu geliştirmek için logit modeller gibi geleneksel istatistiksel yöntemlerle, Bayes ağı gibi bazı yeni akıllı yöntemler geliştirilmiştir. Viaene, sigortacılıkta sahtekârlık tespiti için Bayes öğrenme sinir ağlarını kullanmış, regresyon algoritmalarına ilişkin bölümde çalışmasına değinilen Bermudez ise asimetrik bağlantılı Bayesyen Logit model geliştirmiştir.

Viaene ve arkadaşları [31]; Bayesyen yöntemle eğitilmiş YSA'ların açıklayıcılık kapasitesini göstermek için otomatik ilişki belirleme ağırlık düzenlemesinin kullanıldığı bir model geliştirmiştir. Modelin deneysel kısmında 2002 yılındaki çalışmalarındaki [32] otomobil sigortası bedeni yaralanma hasarı verileri kullanılmıştır.

Xu ve arkadaşları [33]; sigorta sahtekarlık tespiti için bir topluluk (ensemble) öğrenme yöntemi olan Rastgele Uzaylar (Random Rough Subspace) tabanlı YSA modeli önermiştir. Bu yöntemde öncelikle veri setinin tutarlılığını koruyabilmek için indirgenmiş alt veri kümeleri oluşturulmuş ve farklı özelliklerin yer aldığı bu alt veri kümeleri sinir ağı sınıflandırıcısını eğitmesi için kullanılmıştır, sonrasında bu eğitilmiş sinir ağı sınıflandırıcıları birleştirilmiştir. Yöntemin doğrulanması amacıyla gerçek otomobil sigortası verileri ile test edilmiş, ROC (Receiver Operator Characteristics) eğrisi ve doğru sınıflandırma oranlarıyla, yöntemin tek sınıflandırıcılara ve diğer modellere göre daha hızlı ve doğru sonuçlar verdiği gözlemlenmiştir.

- **Naive Bayes Sınıflandırıcısı**

Viaene ve arkadaşları [32], otomobil sigortalarındaki sahtekârlıkların tespitine yönelik, son yıllarda ortaya çıkan modern yöntemleri ele almış; Naive Bayes sınıflandırıcısı, Bayes

Öğrenicili Sınır Ağları, Lojistik Regresyon, Karar Ağaçları, K-En Yakın Komşuluk ve DVM gibi gözetimli yöntemlerin performanslarını doğru sınıflandırma oranı ve ROC eğrisi sonuçlarına göre karşılaştırmıştır. Bu çalışmada “Automobile Insurers Bureau of Massachusetts”ten alınan otomobil sigortası bedeni yaralanma hasar verileri kullanılmıştır. Ele alınan yöntemlerinin hiçbirinin performansının diğerlerine göre tutarlı ve önemli ölçüde üstün olmadığı sonucuna ulaşılmıştır. Doğrusal Lojistik Regresyon ve Doğrusal Kernel En Küçük Kareler DVM gibi nispeten basit ve verimli sınıflandırıcı tekniklerinin genel tahmin performansının diğer yöntemlere göre kayda değer bir başarısı gözlenmiştir. Buna ilave olarak Naive Bayes sınıflandırıcısı da iyi performans göstermiş ancak C4.5 Karar Ağacı istenen başarıyı gösterememiştir. Bu yazarların, bu konularda benzer yöntemlerin işlendiği başka çalışmaları da bulunmaktadır. Viaene ve arkadaşlarının [34] 2004’te yaptığı çalışmada ise yine aynı veri setiyle sahtekârlık tespiti için popüler bir öğrenme yöntemi olan “AdaBoost” tabanlı Naive Bayes Skorlandırma modeli önermektedir. AdaBoost algoritmasında bir tahmin edici, kendinden önceki tahmin edicinin eksik öğrendiği eğitim verilerine daha çok dikkat etmektedir. Bu algoritmada sınıflandırıcı, eğitim seti üzerinde eğitilir ve tahmin yapılır. Sonrasında, yanlış sınıflandırılan eğitim verilerinin göreceli ağırlığı artırılarak model eğitilir ve tekrar tahmin yapılır, sonrasında bu işlemler tekrar edilir.

Bhowmik [15], otomobil sigortası verilerinden sahtekârlıklara ilişkin veri desenlerini tespit etmek için Naive Bayes ve Karar Ağacı sınıflandırma yöntemlerini kullanmış ve bu sınıflandırıcı tahminlerini analiz ederek birleştirilmiştir.

- **Destek Vektör Makineleri**

Tao ve arkadaşları [35], DVM’yi kullanarak sigorta verilerindeki yanlış etiketleme sorunu üzerine bir çalışma yapmıştır. Sigorta şirketinin; sahte bir hasarı tespit edememesi sebebiyle gerçek hasar (sahtekârlık içermeyen) olarak ya da gerçek bir hasarın yanlış değerlendirilmesi sebebiyle sahte olarak etiketleme sorunu, veri setlerinde karşılaşılan bir durumdur ve buna ilişkin kurulan modelin performansını da kötü yönde etkilemektedir. Çalışmada, sigorta sahtekârlığı tespitine yönelik bu problemi çözmek için “İkili Bulanık DVM modeli” önerilmektedir. Bu yöntemde geleneksel DVM’den farklı olarak model öğrenmesi için iki eğitim verisi örnekleme kullanılabilir, bu nedenle model sadece aşırı uyum (overfitting)

problemin üstesinden gelmekle kalmaz, aynı zamanda sınırlı örneklerden de tam olarak faydalanabilir. İkili Bulanık DVM modelinin performansını test etmek için Pekin’deki 800 adet otomobil sigortası verisiyle çalışılmış, aynı zamanda Doğrusal Diskriminant Analizi, Karesel Diskriminant Analizi, Naive Bayes, YSA ve DVM gibi yöntemler de test edilerek en iyi performansa sahip yöntemin İkili bulanık DVM model olduğu gösterilmiştir.

- **Bulanık Mantık**

Doğada nasıl ki karmaşık davranışlar, şekiller bulunuyor, her durumda bir kesinlik olmayabiliyor, kalıplar çok net ve doğrusal değilse bunların makinelere de aktarılışındaki bu bulanıklık ve belirsizlik “Bulanık Mantık” (Fuzzy Logic) kavramının ortaya çıkışında etkili olmuştur. Klasik mantıktaki “0” ve “1” gruplarına ayırma mantığının aksine iki değer arasındaki diğer ihtimallere de fırsat tanınır, bunlar mutlak değerlerden ziyade gri bölge için sınır değerleri olarak görülür. Bulanık Mantık; elektronik aletler, endüstri, robotik, otomasyon, görüntü işleme ve örüntü tanıma gibi birçok alanda kullanılmaktadır.

Bulanık Mantığın, mal ve sorumluluk sigortaları ile sağlık sigortaları alanında örnek çalışmaların yapılmaya başlandığı dönemde Derrig ve Ostaszewski [36] de otomobil sigortalarında şüpheli risk ve hasarların sınıflandırılmasına yönelik bir örüntü tanıma modeli geliştirmiştir. Sahtekârlık tespitinin analizi için en iyi yöntem olduğu iddia edilmemekle birlikte bulanık mantık yeni ve alternatif bir teknik olarak sunulmaktadır.

Pathak ve arkadaşları [37] tarafından geliştirilen bulanık tabanlı bir algoritma ile denetçilerin sigorta şirketine ulaşan tazminat taleplerinde dolandırıcılık unsurlarını tespit edebilmesi için bir yapı sunulmaktadır. Çalışmada tasarlanan model, sigortacılık verilerine erişimde yaşanan zorluklar sebebiyle gerçek veriler üzerinde denenememiş olup modelin performansına yönelik bir değerlendirme yapılamamıştır.

4.1.2. Gözetimsiz (Unsupervised) Yöntemler

Gözetimsiz yöntemler, gözetimli yöntemlerde bahsedilen bilgilere ve etiketlere gereksinim duymaz. Mevcutta konuya ilişkin herhangi bir sahtekârlık verisi olmadığında kullanılır ve bu

durumda normal davranışı temsil eden dağılım modellenerek daha sonra bu modelden en büyük sapmayı gösteren gözlemler, yani aykırı değerler bulmaya çalışılır [7]. Böyle yöntemlerde, sistemin herhangi bir veriyi gözlemlemeksizin kendini eğittiği söylenebilir. Bu yöntemlerde sıklıkla kullanılan algoritmalar ve yöntemler şu şekilde sıralanabilir:

- Kümeleme Algoritmaları: K- Ortalama Tekniği, Özdüzenleyici Haritalar
- Aykırı Değer Tespitine Yönelik Algoritmalar
- Boyut Azaltıcı Algoritmalar: Temel Bileşen Analizi, Doğrusal Diskriminant Analizi

Literatürde otomobil sigortası sahtekârlık tespiti çalışmalarına yönelik gözetimsiz öğrenme yöntemleriyle yapılan çalışmalar aşağıdaki gibi sıralanabilir.

Brockett ve arkadaşları [38], otomobil sigortalarındaki bedeni yaralanma teminatına ilişkin sahte hasar taleplerini sınıflandırmak veya tespit etmek için gözetimsiz bir öğrenme yöntemi olan “Kohonen Öz Düzenleyici Harita (Self-Organizing Maps)” yöntemi önermektedir. Bu yöntem uygulanırken ileri beslemeli sinir ağları ve geriye yayılım öğrenme algoritması kullanılmıştır. Karşılaştırmalı deneylerle, bu tekniğin, sigorta eksperlerinin ve araştırmacılarının sahte hasar dosyalarına ilişkin değerlendirmelerine göre tutarlılık ve güvenilirlik açısından daha iyi bir performansa sahip olduğu gösterilmektedir.

Brockett ve arkadaşlarının [39] çalışmasında ise, otomobil sigortacılığındaki bedeni yaralanma teminatı hasarlarında, RIDIT adı verilen skorlama yönteminden elde edilen skorların Temel Bileşen Analizinde kullanılmasıyla sahtekârlıkları tespit eden gözetimsiz bir yöntem önermektedir. PRIDIT olarak ifade edilen bu yöntemin, önceki çalışmalarında [38] kullanılan ve daha karmaşık bir yapı olan Kohonen Öz Düzenleyici Harita yöntemine göre daha basit ve yeni bir parametrik olmayan bir yöntem olduğu belirtilmektedir. Bu yapıyla eksperlere ve sigorta araştırmacılarına harcanan maliyetin çok daha azıyla benzer çıktılar elde edilebileceği öne sürülmektedir.

Šubelj ve arkadaşları [40], otomobil sigortalarında sahtekârlık tespiti için bir sosyal ağ yöntemi olan Yinelemeli Değerlendirme Algoritması'nı kullanmıştır. Diğer çalışmalardan farklı olarak daha organize şekilde yapılan sigorta sahtekârlıklarının tespitine ve devamında

soruşturulmasına yönelik kullanılabilir bu yöntemde, verilerin temsili ağlar ile yapılmış ve sistem gerçek veriler üzerinde değerlendirilmiştir. Önerilen yöntemde model, sadece varlıkların özelliklerini analiz etmekle kalmaz, aynı zamanda birbiri arasındaki karmaşık ilişkileri de araştırmaktadır. Bu yaklaşımda kullanılan algoritma; etiketlenmiş ve büyük veriye ihtiyaç duymaz, diğer yöntemlere karşı en büyük avantajının bu olduğu da çalışmada belirtilmektedir. Sonuçlar, otomobil sigortacılığı sahtekârlığının önerilen sistemle etkili bir şekilde tespit edilebileceğini ve uygun veri temsilinin hayati önem taşıdığını göstermektedir.

4.1.3. Yarı Gözetimli (Hibrit) Yöntemler

Yarı gözetimli öğrenme, az sayıda etiketlenmiş örnek ve çok sayıda etiketsiz numune içeren verilerde kullanıldığından, gözetimli ve gözetimsiz öğrenme arasında yer almaktadır. Yarı gözetimli yaklaşımın asıl amacı hem etiketlenmiş hem de etiketlenmemiş verilerden bir sınıflandırıcı yetiştirmektir. Yarı gözetimli öğrenme, gözetimli öğrenme ile karşılaştırıldığında daha fazla avantaja sahiptir; çünkü hem etiketli hem de etiketsiz veriyi kullanarak daha iyi performans sağlar. Ayrıca yarı gözetimli öğrenme, girdilerin çoğunun kendiliğinden etiketsiz olduğu insan davranışlarını anlamaya yönelik bir model sağlar [6].

Yarı gözetimli öğrenme araştırmacılar için cazip bir yöntemdir, çünkü etiketli ve etiketlenmemiş verilerin tümünü kullanarak gözetimli öğrenmeden daha iyi bir performans elde edilebilir. Diğer bir bakış açısıyla, yarı gözetimli öğrenme daha az etiketli veriyle gözetimli öğrenmeyle aynı performansa ulaşabilir. Bu durum, ek bilgi gereksinimini azaltarak maliyetin azalmasını sağlar[41]. Literatürde yarı gözetimli öğrenme yöntemlerinin kullanıldığı sigorta sahtekârlıklarının tespiti konusunda yapılan çalışmalar, sağlık sigortası alanında bulunmaktadır.

He ve arkadaşları [42], çalışmalarında Öz Düzenleyici Haritaları (gözetimsiz) ve YSA (gözetimli) yöntemlerini, sınıf sayısının seçimine rehberlik etmesi amacıyla kullanmıştır. Çalışmada sağlık alanındaki hizmet sağlayıcılarının yaptığı sahtekârlıkların tespit edilmesi amacıyla bir yöntem sunulmakta ve Avustralya Sağlık Sigortası Komisyonu'ndan (Health Insurance Commission- HIC) alınan veriler kullanılmaktadır. Eğitim verileri başlangıçta YSA'yla farklı sahtekârlık olasılıklarını temsil eden dört sınıfa ayrılmış ve yetersiz

sınıflandırma sonuçları sebebiyle veride gömülü kümeleri bulmak için Öz Düzenleyici Haritalar kullanılmıştır. Bu işlem sonrasında mevcuttaki dört sınıfı ikişer olarak birleştirmenin verileri temsil etme noktasında daha doğru olduğu sonucuna ulaşılmıştır.

Aynı yıl, sigortalıların sahtekarlıklarını tespit etmek için yine HIC'ten alınan verilerin kullanıldığı William ve Huang'ın [43] çalışmasında ise Kümeleme Algoritmaları ve Karar Ağaçları entegre edilmiştir. Kullanılan verinin çok büyük boyutlarda olması; karar ağacının çok karmaşık ve yorumlamasının zor olmasına neden olmuş, bu sorunla başa çıkmak için kümeleme algoritmalarına başvurulmuştur. Bu sayede sigortalılar özelliklerine göre gruplandırılmış, her bir grup için ayrı karar ağacı oluşturulmuştur.

Musal [44], sağlık sigortası hizmet sağlayıcılarının sahtekârlıklarını tespit etmek üzere, kümeleme algoritması ve regresyon analizinin kullanıldığı iki model önermektedir. İlk modelde sağlık sigortası hasar verileri, posta koduna göre homojen gruplar halinde kümelendirilmiş, bu bölgeler için sağlık hizmeti kullanımı ve fatura bilgilerine göre regresyon analizi yapılmıştır. Sonrasında her bir bölge için bu hizmet kullanımları ve fatura bedellerine göre aykırı değerler tespit edilmiştir. İkinci kurulan modelle, sigortalıların ve hizmet sağlayıcılarının posta kodları bilgilerine bakarak, seyahat mesafelerine ilişkin bir değişken oluşturulmuş, belirli bir mesafenin üzerindeki veriler şüpheli durum olarak işaretlenmiştir.

Bu bölümde anlatılan, otomobil sigortası alanında sahtekârlık tespitine yönelik yapılan çalışmalar Çizelge 4.1'deki gibi özet bir tabloyla gösterilebilir. Yarı gözetimli yöntemlere ilişkin otomobil sigortası alanında yapılan bir çalışma bulunmadığı için bu yöntemler özet tabloya eklenmemiştir.

Çizelge 4.1. Otomobil sigortasında sahtekârlık tespitine yönelik yapılan çalışmalar

Öğrenme yaklaşımı	Veri madenciliği kategorisi	Teknik	Çalışmalar
Gözetimli	Sınıflandırma	Lojistik Modeller	Dionne (1997), Artis vd. (1999), Weisberg vd. (1998), Artis vd. (2002), Wen ve Wang (2005), Bermúdez vd. (2008)
		Karar Ağaçları	Pérez vd. (2005), Bhowmik (2011)
		Yapay Sinir Ağları	Viaene vd. (2005), Xu vd. (2011)
		Naive Bayes	Viaene vd. (2004), Bhowmik (2011)
		Destek Vektör Makineleri	Tao vd. (2012)
		Bulanık Sinir Ağları	Derrig ve Ostaszewski (1995), Pathak vd. (2005)
Gözetimsiz	Kümeleme	Öz Düzenleyici Haritalar	Brockett vd. (1998)
	Boyut Azaltıcı	Temel Bileşen Analizi	Brockett vd. (2002)
	Görselleştirme	Veri Görselleştirme	Šubelj vd. (2011)

Sonuç olarak, literatürde önerilen yöntemler, sigorta araştırmacılarının, eksperlerin ve sigorta şirketlerinin manuel olarak yaptıkları değerlendirmelere göre, tutarlılık ve güvenilirlik açısından daha iyi performans göstermekte ve sigortacılık sektöründeki bu aktörlere sahtekârlıkların tespiti konusunda yardımcı olmak üzere araçlar sunmaktadır.

4.2. Sahtekârlık Tespiti Çalışmalarındaki Önemli Değişkenler

Otomobil sigortasında sahtekârlık tespitine yönelik yapılan çalışmaların bazılarında modelde önemli çıkan değişkenlere de yer verilmiştir. Bu değişkenlerin neler olduğunu görmek ve kuracağımız modelde dikkate almak için tek bir başlık altında toplanmıştır.

Bu çalışmalarda hasar dosyalarındaki sahtekârlık şüphesini artıran önemli değişkenler aşağıdaki gibi derlenmiştir.

- Sigortalı aracın eski ve değerinin düşük olması [32, 34],
- Kaza yapan tarafların ilişkili olması [24, 32, 34],
- Sigortalının yüksek ücretli bir avukat tutmuş olması [32, 34],
- Sigortalının ya da tazminat talebinde bulunan kişinin sigorta ve hasar konusunda çok bilgili olması, [22, 32, 34],
- Kazadaki bedeni yaralanmalarının sadece burkulma ya da incinme boyutunda olması [27, 32, 34],
- Polisin olay yeri raporu olmaması [27, 32, 34],
- Kazanın makul bir açıklamasının olmaması [27],
- Sigortalının hatayı reddetmesi ve işbirliği yapmak istememesi [22, 27, 32, 34] (bunun tersi olarak, sigortalının işbirliği yapmasının sahtekârlık şüphesini artırdığı sonucuna ulaşan çalışmalar da olmuştur [25, 29],
- Bedeni yaralanmaya ilişkin nesnel bir kanıt bulunmaması [27],
- Polis raporuyla uyuşmayan yaralanmaların olması [27],
- Poliçe teminatının artırılması [25] (bunun tam tersi olarak sigorta teminatının genişletildiği poliçelerde sahtekârlık şüphesinin azaldığı sonucuna ulaşan çalışmalar da bulunmaktadır [29]),
- Araç kullanım türünün özel olması [25],
- Kazanın gece veya hafta sonu gerçekleşmesi [25],
- Kazanın kırsal bölgelerde gerçekleşmesi [25],
- Kazaya ilişkin tanık bulunması [25],
- Sürücünün genç olması [25],
- Hasarın geç bildirilmesi [27, 29],
- Kazaya karışan otomobillerin sigortasının aynı sigorta şirketinden olması [27, 29],
- Poliçe düzenleme tarihi ile poliçe yürürlük tarihi arasında kaza yapılması [27, 29],

Çalışmalardan elde edilen yukarıdaki önemli değişkenlerin birçoğu hali hazırda sigorta şirketleri tarafından da sahtekârlık şüphesini artıran durumlar arasında kabul edilmektedir, ancak bunlar şirketlerin geçmiş tecrübelerine dayanan ve çoğunlukla istatistiksel bulgularla kanıtlanmayan durumlardır.

Ancak bunları içlerinde toplumların davranışlarına göre değişkenlik gösteren ve çalışmalara da yansıyan durumlar bulunmaktadır. Örneğin; sigortalının hatayı reddetmesi ve işbirliği yapmak istememesi bazı çalışmalarda şüpheli durum olarak saptanırken bunun tam tersi olarak, sigortalının işbirliği yapmasını şüpheli olduğu sonucuna ulaşan çalışmalar da olmuştur. Bu durumun toplumsal farklılıklardan kaynaklanabileceği gibi hasar dosyasını inceleyen ve kayıtlarını tutan kişilerin sübjektif yargılarından kaynaklandığı da düşünülebilir. Yine, sigortalının sigorta ve hasar konusunda çok bilgili olması durumunun da sübjektif yargılar içerdiği değerlendirilmektedir.

Benzer bir durum ise poliçe teminatının artırılması konusunda da görülmektedir. Teminatın genişletilmesini şüpheli bulan çalışmalar olduğu gibi, sigorta teminatının genişletildiği poliçelerde sahtekârlık şüphesinin azaldığı sonucuna ulaşan çalışmalar da bulunmaktadır. İki durum da muhtemeldir: Sigorta poliçesi düzenlenirken aklında sahte hasar planlamak olan kötü niyetli bir sigortalı teminatlarını geniş tutmak isteyebilir. Ancak tam tersi bilinçli ve iyi niyetli bir sigortalı da maruz kalabileceği her türlü riske karşı kendini sigortalatarak teminatını genişletmek isteyebilir. Bu çalışmada da poliçe teminatının artırılmasının sahtekârlık şüphesi üzerindeki etkisine yönelik bir değerlendirme yapılması imkânı bulunmaktadır. Önemli değişkenlere ilişkin sonuçlara 6. Bölüm’de yer verilmektedir.

4.3. Yapılan Çalışmanın Literatürdeki Konumu

Bu bölümde ağırlıklı olarak otomobil sigortası sahtekârlığı tespiti konusundaki uluslararası literatürdeki istatistiksel yöntemlere ilişkin çalışmalar incelenmiştir. Bu çalışmada da literatürdeki yöntemler (Lojistik Regresyon, Karar Ağaçları, YSA, DVM) ele alınarak büyük veri analizi uygulamaları yapılmıştır. Ancak literatürdeki yöntemlerden farklı olarak bu alanda

uygulama yeri bulamadığı düşünölen Bayes Ağlarına da yer verilmiştir. Bu kapsamda Bayes Ağlarında üç farklı algoritma ile çalışılmıştır.

Buna ek olarak Türkiye’de yeni yeni sigorta sahtekârlığına yönelik farkındalık artmakta ve sigorta şirketleri ile düzenleyici otoriteler bu konuya ağırlık vermektedir. Dolayısıyla bu alandaki veri tabanları ve kayıt sistemleri gün geçtikçe gelişmektedir. Başlangıç olarak otomobil sigortası verilerine ilişkin kayıt sistemleri daha gelişmiş ve sağlıklı olduğundan bu alandaki verilerle çalışmak tercih edilmiştir. Aynı zamanda herhangi bir sigorta branşında Türkiye’deki tüm sigorta şirketi verileri üzerinden yapılmış sahtekârlık tespitine yönelik bir çalışma bulunmamaktadır. Sigorta sahtekârlıklarının sektöre ve paydaşlarına verdiği zararlar düşünüldüğünde bu alanda çalışma yapmanın elzem olduğu düşünülmektedir. Buradan yola çıkılarak ölkemizdeki kasko sigortası verileri üzerinden sahtekârlık tespitine yönelik bir analiz yapılması amaçlanmıştır.

Özetle, bu çalışma ile hem literatürde yer bulamayan bir yöntem kullanıldığı hem de Türkiye’de uygulaması yapılmamış bir alan ve veriyle çalışıldığından literatüre yeni bir katkı yapıldığı değerlendirilmektedir.

5. ÇALIŞMADA KULLANILAN İSTATİSTİKSEL YÖNTEMLER

Bu bölümde, çalışmada kullanılan istatistiksel yöntemlere ilişkin teorik bilgilere yer verilmiştir. Bölüm 4'te literatürde otomobil sigortası sahtekârlık tespitinde yapılan çalışmalar anlatılırken gözetimli ve gözetimsiz öğrenme yöntemlerinin sınıflandırılmasına değinilmiştir. Kısaca gözetimli yöntemler, veri tabanında mevcutta bulunan etiketlenmiş veriler ve uzman bilgisinden öğrenme gerçekleştirerek yeni karşılaşılan verilere yönelik tahmin ve sonuçlar sunar. Gözetimsiz yöntemlerde ise sistemin herhangi bir veriyi gözlemlemeksizin kendini eğittiği ve örüntüleri tespit ettiği söylenebilir. Çalışmada gözetimli öğrenme yöntemlerinden aşağıda yer alan teknikler ve algoritmalar kullanılması sebebiyle bunlara ilişkin bilgi verilmiştir.

- Lojistik Regresyon Analizi
- Bayes Ağları (Naive Bayes, Hill Climbing, Max-Min Hill Climbing)
- Karar Ağaçları (C.4.5, Bagging, Random Forest)
- Yapay Sinir Ağları
- Destek Vektör Makineleri

5.1. Lojistik Regresyon Analizi

İki ya da daha fazla değişkenin aralarındaki ilişkiyi matematiksel olarak açıklayan yöntem regresyon analizidir. Doğrusal regresyonda bu ilişki doğrusal bir fonksiyonla tanımlanır. Regresyon fonksiyonunda tahmin edilen yanıt değişkenlerine bağımlı değişken, tahmin etmek için kullanılan açıklayıcı değişkenlere de bağımsız değişken denir. Tek bağımsız değişkenli regresyon modellerine basit regresyon, birden fazla bağımsız değişkenli modellere ise çoklu regresyondan denir. Basit doğrusal regresyon modeli, Y bağımlı değişken, x_1 bağımsız değişken, β_0 ve β_1 bu değişkenin bilinmeyen parametreleri, ε_i hata terimi olmak üzere aşağıdaki gibi yazılabilir.

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Çoklu doğrusal regresyon modeli ise p adet bağımsız değişken için aşağıdaki gibi yazılabilir [45].

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

En Küçük Kareler tahmin yönteminin kullanıldığı doğrusal regresyonda bağımsız değişkenler nicel (sürekli veya kesikli) olmalıdır. Doğrusal regresyonda yapılan tahminlerin güvenilir olması için bazı varsayımların sağlanması gereklidir. Bunlar; hata terimlerinin şansa bağlı, bağımsız ve normal dağılım göstermesi, hata terimlerinin varyanslarının homojen ve σ^2 'ye eşit olması, hata terimleriyle bağımsız değişkenler arasında ilişki olmaması ve bağımsız değişkenler arasında ilişki olmamasıdır [45].

Sınıflandırma problemlerinde, bağımlı değişkenin iki ya da daha fazla sınıfla gösterilmesi durumunda, bağımlı değişkenle bağımsız değişkenler arasındaki ilişkiyi açıklamak için **Lojistik Regresyon** kullanılır. Bağımlı değişken iki kategorili olduğunda (evet-hayır, var-yok) ikili (binary) lojistik regresyon, ikiden fazla kategorik değer aldığı anda ise çoklu (multinomial) lojistik regresyondan bahsedilir. Lojistik Regresyonda, En Çok Olabilirlik tahmin yöntemi kullanılmaktadır. Ayrıca Doğrusal Regresyonda bağımlı değişkenin değerinin tahmin edilmesi amaçlanırken, lojistik regresyonda bağımlı değişkenin alabileceği değerlerin gerçekleşme olasılığı tahmin edilir [46].

Bağımsız değişkenleri $\mathbf{x} = (x_1, x_2, \dots, x_n)$ vektörü ile gösterirsek, bağımlı değişken $Y = 1$ değerini aldığı zaman koşullu olasılık $P(Y = 1 | \mathbf{x}) = \pi(\mathbf{x})$ şeklinde olur. Bu durumda çoklu lojistik regresyonun logit modeli,

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_{pi} + \varepsilon_i$$

ve lojistik regresyon modeli de,

$$P(Y = 1 | \mathbf{x}) = \pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}}$$

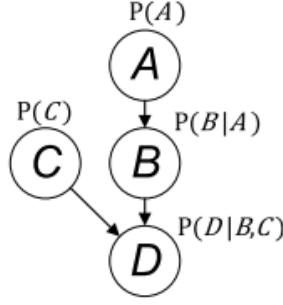
şeklinde gösterilir [47].

Çoklu lojistik regresyon modellerinde değişkenlerin seçiminde olabilirlik oran ölçütü, değişkenlerin katsayılarının anlamlılığını ölçen Wald ölçütü ve AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) gibi skor ölçütleri kullanılabilir. Bu ölçütlere göre ileriye doğru seçim (forward selection), geriye doğru eleme (backward elimination) ve adimsal seçim (stepwise selection) gibi yöntemlerle değişkenler modele eklenmekte ve çıkarılmaktadır [48].

İleri doğru seçim yönteminde analize, ilk adımda önemli kovaryansları kapsayan değişkenlerle başlanır. Sonraki adımlarda diğer değişkenler eklenerek aynı zamanda en etkin modeldeki değişkenler arasındaki etkileşimler de belirlenerek modele eklenir. Geriye seçim yönteminde, başlangıçta tüm bağımsız değişkenler modele alınır, sonrasında değişkenlerin modele sağladığı katkıya göre modelin anlamlılığını düşüren ($p > \alpha$) değişkenler her adımda çıkarılır. Elde edilen son model anlamlılığı en yüksek olan modeldir. Adimsal seçim yöntemi ise ileriye doğru seçim ve geriye doğru eleme yöntemlerinin birleşimini içerir. Model anlamlılığına göre değişkenler eklenir ya da çıkarılır. Bu yöntemin sakıncası ise, modelde olmayan bütün değişken katsayılarının maksimum olabilirlik tahminlerinin her adımda tek tek hesaplanmasının gerekmesidir. Bu durum da, çok değişkenli büyük verilerin analizinde hem zaman hem de para bakımından bu yöntemin tercih edilmemesine sebebiyet verebilir.

5.2. Bayes Ağları

Bayes Ağları uzman bilgisi ile istatistiksel veriyi birleştirmek için uygun grafiksel olasılıksal modellerdir. Bayes Ağının grafiksel yapısı, değişkenleri temsil eden düğümlerden ve bu değişkenler arasındaki nedensel ilişkileri temsil eden oklardan oluşur. Bir Bayes Ağında, A ve B değişkenleri $A \rightarrow B$ deki gibi bağlandıysa, A ebeveyn (parent) değişken, B ise çocuk (child) değişken olarak adlandırılır. Bir Bayes Ağında ebeveyni olan her değişkenin ebeveynleriyle olan koşullu olasılık dağılımı, ebeveyni olmayan değişkenlerin ise marjinal olasılık dağılımı vardır. Kesikli Bayes Ağlarında, bu olasılık dağılımlarının parametreleri, koşullu olasılık tabloları ile gösterilir. Şekil 5.1’de Bayes Ağının yapısı ve olasılık dağılımlarına ilişkin bir örnek görebiliriz. Burada $P(A,B,C,D)$ bileşik dağılımı, $P(D|B,C) P(C) P(B|A) P(A)$ şeklinde bulunur [49, 50].



Şekil 5.1. Bayes ağının yapısı [50].

Bayes ağlarının yapısı ve koşullu olasılık dağılımlarını veriden öğrenen algoritmalar bulunmaktadır. Bayes Ağlarını öğrenme algoritmaları üç kategoriye ayrılabilir: Skorlama (search and scoring-based) yöntemlerine dayalı yöntemler, bağımlılık analizine (constraint-based) dayalı yöntemler ve hibrit yöntemler. Skorlama algoritmaları Bayes ağında AIC veya BIC gibi bir skoru maksimize eden yapıyı bulmaya çalışır. Bağımlılık analizine dayanan algoritmalar, verideki koşullu bağımsızlık durumlarını en doğru tespit eden yapıyı bulmaya çalışır. Hibrit yöntemler ise skorlama ve bağımlılık analizine dayanan algoritmaları dönüşümlü olarak kullanır [51].

Bayes ağlarındaki sık kullanılan bazı algoritmalar:

- Skorlamaya dayalı yöntemler: Hill Climbing (HC), TABU
- Bağımlılık analizine dayalı yöntemler: Max-Min Parents and Children (MMPC), Incremental Association (IAMB), Grow Shrink (GS)
- Hibrit yöntemler: Max-Min Hill Climbing (MMHC), 2-phase Restricted Maximization (RSMAX2), Hybrid HPC (H2PC)

Bu çalışma kapsamında 3 Bayes Ağ algoritması denenmiştir: Naive Bayes, HC ve MMHC.

- **Hill Climbing (HC):** Skorlama yöntemine dayalı bir yöntem olarak sınıflandırılabilir. HC, her ağ yapısını değerlendirmek için K2, AIC ve BIC gibi puanlama ölçütlerine göre sıralayan ve örnek verilere en uygun optimum yapıyı aramaya çalışan bir sezgisel optimizasyon algoritmasıdır. HC, Bayes Ağında ok ekleme, çıkarma veya tersine çevirme gibi yerel değişiklikler yaparak ağı yinelemeli olarak geliştirir.

- **Max-Min Hill Climbing (MMHC):** Her iki kategoriye ait algoritmalarından kavramlar ve teknikler kullanan hibrit bir yöntem olarak sınıflandırılabilir. MMHC, önce MMPC adı verilen yerel bir keşif algoritması kullanarak Bayes Ağının yapısını/iskeletini öğrenir. Daha sonra, bir Bayes skorlaması olan HC algoritmasını kullanarak bu yapıyı yönlendirir. Temel fikir, her bir düğümün ebeveyn ve çocuk düğümlerini bulmak ve arama alanını azaltmak için koşullu bağımsızlık testini kullanmaktır. Ardından, yapıdaki en yüksek skoru bulmak için HC algoritmasını kullanır [51].
- **Naive Bayes:** “Naive” ya da “Basit” Bayes sınıflandırıcısı en çok bilinen ve kullanılan Bayes ağı sınıflandırıcısıdır. Naive Bayes sınıflandırıcısı; makine öğrenmesi, veri madenciliği ve örüntü tanıma gibi disiplinlerde kullanılan, koşullu olasılık temelli bir sınıflandırıcıdır. Naive Bayes sınıflandırıcısında, sınıflandırılmak istenen değişkenin tüm özellikleri, sınıflandırılan değişkene koşullu olarak birbirinden bağımsızdır. Mevcuttaki sınıflandırılmış verileri kullanarak sisteme yeni dâhil olan verilerin hangi sınıfa ait olacağını olasılığını hesaplar[34]. Bayes teoreminin denklemi aşağıdaki gibidir,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Burada,

$P(A|B)$: B olayı gerçekleştiğinde A olayının meydana gelme olasılığı,

$P(B|A)$ = A olayı gerçekleştiğinde B olayının meydana gelme olasılığı,

$P(A)$, $P(B)$; A ve B olaylarının önsel olasılıklarıdır.

Veri setindeki her bir gözlem n özelliğe sahip olduğunda ve hangi sınıfa ait olduğu bilinmeyen yeni bir gözlem Naive Bayes modeli ile sınıflandırılmak istendiğinde, Bayes denklemi kullanılarak, bu gözlemin verilen t sınıfına ait olma olasılığı aşağıdaki denklemle hesaplanır.

$$P(x|t) = \prod_{m=1}^n P(x_m|t)$$

5.3. Karar Ağaçları

Karar Ağaçları, verileri belirli özelliklere göre sınıflandırmaya yarar. Kök, dallar ve yapraklardan oluşan ağaca benzer yapısıyla basit karar verme adımları uygulayarak, çok sayıdaki veri kümesini küçük gruplara böler ve her başarılı bölme işlemiyle, alt grupların üyeleri birbiriyle çok daha benzer hale gelmektedir. Bu kökten dallara doğru büyüyen ağaç yapısında her ayrışma bir “düğüm”dür, düğümler üzerinde değişken özelliklerinin test işlemi yapılmakta ve bunun sonucunda ağaç veri kaybetmeden dallara ayrılmaktadır. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşerek nihayetinde ağaç sınıflar ile son bulmaktadır. Karar Ağaçlarının oluşturulurken en önemli husus ağaçtaki dallanmanın hangi kritere göre ağaç yapısının oluşturulacağıdır. Buna ilişkin geliştirilmiş çeşitli yaklaşımlar vardır, bunlardan en önemlileri bilgi kazancı, Gini indeksi, Towing kuralı ve Ki-kare olasılık tablo istatistiği yaklaşımlarıdır.

Karar Ağaçları tarafından üretilen sonuçlar kolayca anlaşılabilir, sonuçlar kurallara dönüştürülebilmektedir. Ayrıca Karar Ağaçları çok sayıda işlem yapmaya gerek duymadan sınıflandırma işlemini gerçekleştirebilir, hem kategorik hem de nicel veriler üzerinde işlem yapabilmektedir. Bu özellikler Karar Ağaçlarını sık başvurulan bir sınıflandırma yöntemi haline getirmektedir [52].

Karar Ağaçlarında zaman içerisinde çok çeşitli algoritmalar geliştirilmiştir, bunlardan bazıları: C&RT (Classification and Regression Trees), CHAID (Chi-Squared Automatic Interaction Detector), Rastgele Orman (Random Forest), ID3, C4.5, Bagging (Bootstrap Aggregating), Hızlandırılmış Ağaçlar (Boosted Trees), QUEST (Quick, Unbiased, Efficient Statistical Tree), MARS (Multivariate Adaptive Regression Splines), C5.0, SPRINT (Scalable Parallelizable Induction of Decision Trees) [52].

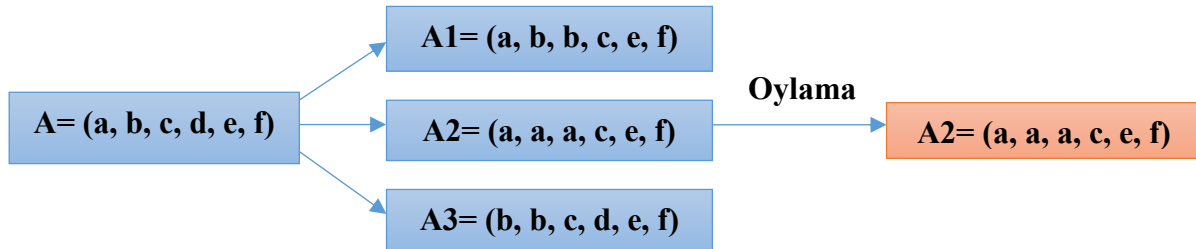
Çalışmada yukarıda bahsi geçen algoritmalarından C4.5, “Bagging” ve “Random Forest” yöntemlerinden yararlanılmıştır. Bu algoritmaların çalışma mantıklarına aşağıda değinilmektedir.

- **C4.5:** Bilgi kazancına dayalı bir Karar Ağacıdır. Bilgi kazancının ölçümünde entropi kullanılır. Entropi, rassal bir değişkenin belirsizliğinin ölçütü olarak ifade edilebilir. C4.5 öncelikle hedef (yanıt) değişken için entropi değerini hesaplar, sonrasında ise her bir tahmin edici değişkenin bilgi kazancını hesaplar. Bu hesaplamaların sonunda en yüksek bilgi kazancını sağlayan tahmin edici değişken tespit edilerek dallanma gerçekleşir. Bir düğümdeki bütün örnekler tek bir sınıfa ait olana kadar algoritma her bir alt düğüme tekrar tekrar uygulanır.

C.4.5, kategorik verilerin yanı sıra nicel verilerle de çalışabilmesi, kayıp, gürültülü ve büyük verilerle baş edebilmesi, yorumlamasının kolay olması sebebiyle sıklıkla başvurulan bir Karar Ağacı algoritmasıdır. Çalışmada da bu sebepler göz önünde bulundurularak tercih edilmiştir.

- **Bagging (Bootstrap Aggregating/ Torbalama):** Tek bir karar ağacından daha iyi performans elde etmek amacıyla çeşitli karar ağaçlarını birleştiren ağaç topluluk (ensemble) algoritmaları adı verilen yöntemler bulunmaktadır. “Bagging” algoritması bu yöntemlerden biridir. Ağaç topluluk algoritmalarının temel amacı zayıf sınıflandırıcılardan güçlü sınıflandırıcılar oluşturmaktır. Bagging yönteminde, bir eğitim verisinin farklı kombinasyonları üretilerek birden fazla eğitim örneğinin oluşturulması amaçlanır. Bu durum da varyansı düşürdüğü için aşırı uyumun engellenmesine yardımcı olur, aynı zamanda kayıp veriye karşı güçlü bir yapısı vardır.

N adet bir eğitim verisinden yine N örneğe sahip bir eğitim verisi rastgele seçimle yerine koymalı olarak üretilir. Bu durumda seçilen yeni örnekte eğitim verisindeki bazı örnekler yer almazken, bazıları birden fazla yer alır. Her bir karar ağacı bu şekilde üretilmiş birbirinden farklı örnekler içeren eğitim verileri ile eğitilirler ve sonuçta her örnek için olasılık değerleri ve sayısal tahminler toplanarak bir oylama yapılır, en yüksek oya sahip sonuç seçilir (Şekil 5.2.) [53].



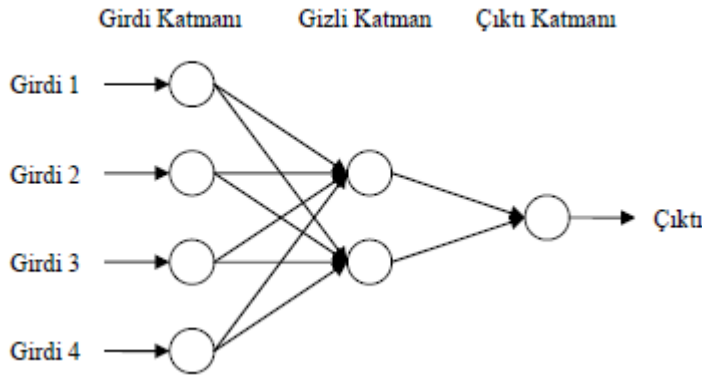
Şekil 5.2. Bagging yöntemi gösterimi

- **Random Forest (Rastgele Orman):** Random Forest hem sınıflandırma hem de regresyon problemlerinde kullanılan, makine öğrenmesinin popüler ağaç topluluk algoritmalarından biridir. Ayrıca hem kategorik hem de nicel veriler üzerinde işlem yapabilir. Random Forest, veriden birden fazla karar ağacını oluşturarak gözlemleri ve özellikleri rastgele seçer ve sonrasında ağaçları birleştirerek sonuçların ortalamasını alır. Bu sayede özelliklerin alt kümelerini kullanarak daha küçük ağaçlar oluşturur ve çoğu zaman aşırı uyumu engeller. Random Forest, Bagging yönteminin rastgelelik özelliği eklenerek geliştirilmiş halidir. Tüm değişkenler içinden en iyi dalı kullanarak düğümleri dallara ayırmak yerine, rastgele olarak seçilmiş değişkenlerden en iyisini kullanarak düğümleri dallara ayırır. Kısacası rastgele özellik seçimi kullanılarak ağaçlar geliştirilir ve geliştirilen ağaçlar budanmaz [54]. Hem regresyon hem sınıflandırmada kullanılması, çoğu veri tipiyle çalışabilmesi, kolay yorumlanması ve aşırı uyumu azaltması gibi avantajlarının yanında bazı dezavantajları bulunmaktadır. Bunlardan bir tanesi çok fazla ağaç üretmesi ve işlem yapması sebebiyle gerçek zamanlı tahminler için yavaş çalışmaktadır. Gerçek hayattaki uygulamalarda random forest algoritması yeterince hızlıdır, ancak çalışma zamanı performansının önemli olduğu durumlarda diğer yaklaşımların tercih sebebi olabilir. Diğer bir durum ise, Random Forest tahmin edici bir modelleme aracı olup açıklayıcı bir araç değildir. Eğer verideki ilişkilerin tanımıyla ilgileniliyorsa yine başka yaklaşımlar tercih edilmelidir. Random Forest, tıp, bankacılık, finans, e-ticaret, borsa gibi birçok farklı alanda tercih edilmektedir. Örneğin bankanın hizmetlerini daha sık kullanacak müşterileri tespit etmek ve borçlarını zaman içerisinde ödenebilirliğini anlamak, e-ticaret sitelerinde müşterinin ürünü beğenip beğenmeyeceğini belirlemek, finans alanında sahtekârlıkları tespit etmek gibi amaçlarla kullanılmaktadır.

5.4. Yapay Sinir Ağları

YSA, beynin bir işi veya fonksiyonu yerine getirme yöntemini modelleyen makinelerdir. İnsan beynine benzerliği, bilgiyi öğrenme yoluyla elde etmesi ve bilginin depolanması için sinir hücreleri (nöron) arası bağları kullanmasıdır. YSA kısaca, deneysel bilgiyi saklama ve kullanıma hazır hale getirme yeteneğine sahip, basit işleme birimlerinden oluşan, çok yoğun, paralel ve dağıtılmış düzende çalışan bir işlemci olarak tanımlanabilir [55].

Bir YSA'nın yapısı Şekil 5.3'te görüleceği üzere 3 katmandan oluşur. İlk katman “Girdi Katmanı”dır. Girdi katmanında veriler herhangi bir işleme tabi tutulmadan girdiler ile aynı değerde çıktı üretirler. İkinci katman, ağın ne girdi ne de çıktısına bağlı olmaması sebebiyle “Gizli Katman” olarak adlandırılır. Ağ yapılarında gizli katmanın kullanılması zorunluluğu bulunmamaktadır, ağ sadece girdi ve çıktı katmanlarından oluşabilir. Ancak gizli katman, daha fazla örüntünün tanınması ve daha fazla öğrenmenin sağlanmasından dolayı ağı daha güçlü kılar. Son katman ise, “Çıktı Katmanı”dır. Çıktı katmanı en az bir çıktıdan oluşur ve çıktı ağ yapısında bulunan fonksiyona bağlıdır. Bu birimlerde girdi katmanında olanın aksine işlem gerçekleştirilir ve birimler kendi çıktılarını üretirler [56, 57].



Şekil 5.3. Yapay sinir ağı yapısı ve katmanları [56].

YSA, öncelikle bir veri kümesi üzerinde öğrenme algoritmaları çalıştırılarak eğitilir. Bu eğitim sonrasında YSA'nın içerisindeki bazı ağırlıklar belirlenir. Bu ağırlıklar kullanılarak yeni gelen veriler işlenir ve bir sonuç üretilir. Ancak bu ağırlıkların neden ilgili değerleri aldıklarının bilinmemekte ve çıkan sonucun nedenleri açıklanamamaktadır. Bu sebeple bir YSA'nın girdi ile çıktı katmanları arasında yaptıkları işlemler içi bilinmeyen bir şekilde çalışan “kara kutu” olarak düşünülmektedir. Bu olumsuz özellik sebebiyle ABD’de, YSA’nın, kredi taleplerinin değerlendirilmesinde kullanılması yasaklanmıştır. Çünkü sistemin bir kişiye kredi vermeme sebebi açıklanamamaktadır [56].

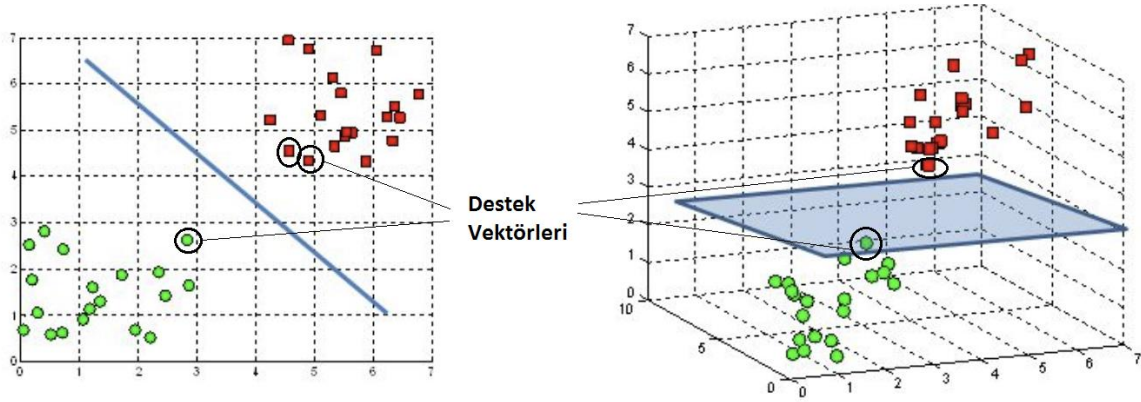
YSA'nın hesaplama ve bilgi işleme gücünü, paralel dağılmış yapısından, öğrenbilme ve genelleme yeteneğinden aldığı söylenebilir. Ayrıca, çok çeşitli alanlardaki karmaşık

problemlerin çözümünde kullanılabilir, hem kategorik hem de nicel veriler üzerinde işlem yapabilir.

5.5. Destek Vektör Makineleri

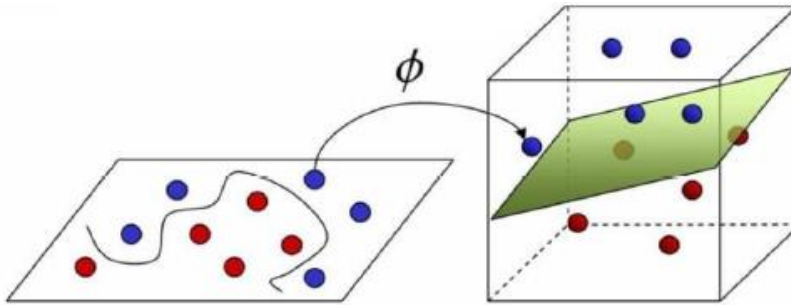
DVM, verileri optimal olarak ikiye ayıran, çok boyutlu bir hiper düzlem oluşturan ve örüntü tanıma problemlerinde sınıflandırma yapmak için geliştirilmiş yaygın bir öğrenme yöntemidir. DVM’de kullanılan algoritmalar başlangıçta iki sınıflı doğrusal verilerin sınıflandırılması amacıyla tasarlanmış, sonrasında çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması için geliştirilmiştir [58].

DVM, karmaşık veri setlerinde ve çözümlemesi zor örüntülerin tanımlanmasında kullanılan bir öğrenme yöntemi olup hem sınıflandırma problemlerinde (Support Vector Classification) hem de regresyon problemlerinde (Support Vector Regression) uygulanabilmektedir. DVM; veri seti üzerinde ortalama hata karesini minimize ederek türetilen yapısal risk minimizasyonu prensibine dayalıdır ve temel varsayımlarından biri, eğitim kümesindeki tüm örneklerin bağımsız ve benzer olarak dağılmış olmasıdır. DVM’nin veriyi sınıflandırmak/bölmek amacıyla oluşturduğu “hiper düzlem” (hyper plane) probleme göre doğrusal ya da doğrusal olmayan şekillerde olabilir. DVM’nin amacı bu hiper düzlemin yeni katılacak olan veriye karşı dayanıklı olabilmesi için iki ayrı sınıfta bulunan örnek grubuna eş uzaklıkta olmasını sağlamaktır. İki sınıflı verileri gruplara ayırabilen birden çok hiper-düzlem çizilebilir, ancak DVM’nin amacı kendisine en yakın noktalar arasındaki uzaklığı maksimuma çıkaran hiper düzlemi bulabilmektir. Bu uzaklığı maksimuma çıkararak en uygun ayrımı yapan hiper düzleme “optimum hiper-düzlem” ve bu grupların sınır çizgisine en yakın noktalar ise “destek vektörleri” olarak adlandırılmaktadır [59, 60]. Doğrusal ve doğrusal olmayan hiper düzlem ve destek vektör gösterimleri Şekil 5.4. ve 5.5.’teki gibidir.



Şekil 5.4. R^2 ve R^3 'teki doğrusal hiper düzlem ve destek vektörleri

DVM, kernel fonksiyonu sayesinde doğrusal olmayan dönüşümler yaparak verilerin yüksek boyutta doğrusal olarak ayırılmasına imkân sağlar. Literatürde en çok kullanılan kernel fonksiyonları; radyal tabanlı fonksiyon, polinom, Pearson VII (PUK) fonksiyonu ve normalleştirilmiş polinom kernelleridir [59].



Şekil 5.5. R^2 ve R^3 'teki doğrusal olmayan hiper düzlem

Günümüzde DVM; yüz tanıma, metin tanıma, ekonomik tahmin, hastalıkların teşhisi, finansal sektörlerde risk gruplarının saptanması ve örüntü tespiti gibi birçok alanda kullanılmaktadır.

5.6. Modellerin Performansını Ölçmeye Yönelik Çalışmada Kullanılan Kavramlar

Bir öğrenme algoritması kurduktan ve verileri eğittikten sonraki adım kurulan modelin başka verilere karşı ne kadar etkili olduğunu bulmaktır. Modelin eğitilirken hiç görmediği verilerle test edilmesi çıktının yanlı olması ve modelin aşırı uyum (overfitting) yapmaması açısından önemlidir. Bu sebeple veriyle çalışmadan önce eğitim ve test verisi olarak bölerek analiz yapmak gereklidir. Bu kapsamda veri iki parçaya ayrılarak, birinci parça sınıflandırma yapacak modelin oluşumundaki eğitim verisi, ikinci parça ise modelin performansının belirlenmesindeki test verisi olarak kullanılır. Bu şekilde oluşturulan modelin test verisine uygulanması sonucunda doğru sınıflandırılan örneklerin oranı sınıflandırıcının performansını gösterir [59].

Sınıflandırma modellerinin performansının ölçülmesinde en sık kullanılan yöntemler Hata Matrisi (Confusion Matrix) ve ROC eğrisidir. Hata matrisinde, bir tarafta model tahminine ilişkin diğer tarafta ise gerçek verilere ilişkin değerler bulunmakta ve modelin ne kadar doğru ne kadar yanlı tahmin yapıldığına ilişkin bilgi vermektedir. (Çizelge 5.1.)

Çizelge 5.1. Hata Matrisi (Confusion Matrix) değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	Doğru Pozitif	Yanlış Pozitif
	Negatif (0)	Yanlış Negatif	Doğru Negatif

Burada; doğru pozitif (DP) gerçek değerleri 1 olan modelin de 1 tahmin ettiği örnekler, doğru negatif (DN) gerçek değerleri 0 olan modelin de 0 tahmin ettiği örnekler, yanlış pozitif (YP) gerçek değeri 0 olan ancak modelin 1 tahmin ettiği örnekler, yanlış negatif (YN) gerçek değeri 1 olan ancak modelin 0 tahmin ettiği örnekleri ifade etmektedir. Buradan yola çıkarak da modelin performansında kullanılacak ölçütler oluşmaktadır.

Doğruluk oranı (Accuracy rate): Sınıflayıcının ne kadar doğru tahmin ettiğinin oranıdır.

$$\frac{DP + DN}{\text{Toplam}}$$

Yanlış sınıflandırma oranı= Hata oranı (Misclassification rate, Error rate): Sınıflayıcının ne kadar yanlış tahmin ettiğinin oranıdır.

$$\frac{YP + YN}{\text{Toplam}} = 1 - \text{Doğruluk oranı}$$

Gerçek pozitif değerlerin oranı=Duyarlılık= Hassasiyet (Sensitivity, Recall): Sınıflayıcının tüm pozitif sınıftan ne kadar doğru tahmin ettiğinin oranıdır.

$$\frac{DP}{DP + YN}$$

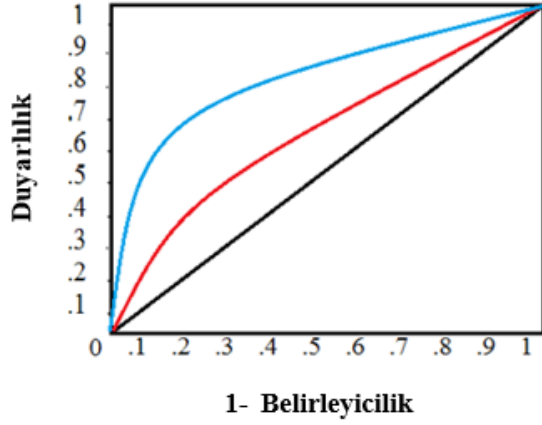
Gerçek negatif değerlerin oranı= Belirleyicilik= Seçicilik (Specificity, Selectivity): Sınıflayıcının tüm negatif sınıftan ne kadar doğru tahmin ettiğinin oranıdır.

$$\frac{DN}{DN + YP}$$

Yanlış pozitif değerlerin oranı= Sınıflayıcının gerçek değeri negatif olduğu halde pozitif olarak tahmin ettiği gözlemlerin oranıdır.

$$\frac{YP}{DN + YP} = 1 - \text{Belirleyicilik}$$

ROC eğrisi, Şekil 5.6.'da da gösterildiği üzere, x ekseninde gerçek pozitif değerlerin oranına (duyarlılık) karşılık, y ekseninde yanlış pozitif değerlerin oranının (1-belirleyicilik) olduğu bir eğridir. Tüm olası değerler üzerinde performansı özetlemek için kullanılır. İyi bir sınıflandırıcıda eğri sol üst köşeye yakındır.



Şekil 5.6. ROC eğrisi

ROC eğrisinin altında kalan alan ise AUC (Area Under Curve) modelin iki sınıf arasında ne kadar iyi ayırt edici olduğunun bir ölçüsüdür. ROC eğrisi sol üst köşeye yaklaştıkça da eğrinin altında kalan alan artacağından AUC değerinin 1'e yakın olması da yine sınıflandırıcının performansının iyi olduğunu gösterir.

Bu bölümde çalışmada uygulanan yöntemlere ve kullanılan performans ölçülerine ilişkin bilgi verilmiştir. Bir sonraki bölümde ise analiz öncesindeki verinin yapısına, yapılan düzeltme işlemlerine değinilmekte ve modellerin sonuçlarına yorumlamalar yapılmaktadır.

6. TÜRKİYE’DE OTOMOBİL SİGORTALARINDAKİ SAHTEKÂRLIK TESPİTİNE YÖNELİK BİR UYGULAMA

Bu çalışma kapsamında Türkiye’de otomobil sigortalarındaki hasar dosyalarında yapılan sahtekârlıkları büyük veriden tespit eden bir model kurmak amaçlanmıştır. Bu amaçla literatür taraması yapılmış ve bu alanda kullanılan makine öğrenmesi yöntemleri incelenmiştir. Bu yöntemlerden mevcut veriye en çok uyan ve daha iyi performans sağlayan gözetimli yöntemler tercih edilmiştir. Bu Doğrultuda Lojistik Regresyon Analizi, Bayes Ağları, Karar Ağaçları, YSA ve DVM uygulamaları yapılmıştır.

6.1. Veri ve Verideki Değişkenler

Çalışmada kullanılan veriler, SBM altında bulunan SİSEB’ten temin edilmiştir. Tutulan kayıtların daha sağlıklı olması ve sigorta sahtekârlıklarının bu alanda sık yaşanması sebebiyle çalışmada otomobil (kasko) sigortasındaki maddi hasarlara ilişkin bir analiz yapılmıştır. SİSEB’ten temin edilen 2017 yılına ait kasko sigortalarındaki maddi hasar dosyalarına ilişkin bilgileri içeren 512.627 adet ham veri, kişisel bilgileri içermeyecek şekilde olup aşağıdaki değişkenlerden oluşmaktadır.

- **Fraud (Yanıt değişkeni):** Sigorta sahtekârlığına ilişkin durumu gösteren yanıt değişkenidir. İspatlanmış bir sahtekârlık (mahkeme süreci, somut deliller, feragat) veya somut delil gösterilemediği halde sigorta şirketince sahtekârlık olarak tespit edilmiş hasar dosyası. (0=hayır, 1=evet)
- **Dosya_durum:** Hasar dosyasına ilişkin yapılan işlem.
 - ✓ **Açık:** Hasar dosyasına ilişkin işlemler henüz sonuçlandırılmamış.
 - ✓ **Kapalı-ödeme:** Tazminat ödemesi yapılarak hasar dosyası kapanmış.
 - ✓ **Kapalı-red:** Hasar dosyası çeşitli sebepler ile tazminat ödemesi yapmadan kapatılmış. Bu sebep sahtekârlık kaynaklı olabileceği gibi sigorta poliçesi kapsamında olmayan sebeplerle hasarın gerçekleşmesinden kaynaklanmış diğer durumlar da olabilir.
 - ✓ **Kapalı-iptal:** Sigorta şirketi tarafından yanlışlıkla açılan ve kapatılan hasar dosyası.
 - ✓ **Yeniden açıldı:** Hasar dosyası çeşitli sebeplerle açılarak tekrar incelemeye alınmış.

- **Dosya_alt_durum:** “Dosya_durum” deęişkeninde hasar dosyasına ilişkin belirtilen işlemin detayı.
 - ✓ **Alkollü araç kullanımı:** Alkollü araç kullanımı sebebiyle hasar dosyasına ilişkin tazminat ödemesinin reddedilmesi.
 - ✓ **Çifte sigorta:** Deęerinin tamamı sigortalanan aracın, aynı rizikolara karşı aynı süreler için sigorta ettirilmiş olması durumunda sonradan yapılan poliçeye ilişkin tazminat talebinin reddedilmesi.
 - ✓ **Ehliyetsiz kullanım:** Aracın sürücüsünün ehliyetsiz olması sebebiyle hasar dosyasına ilişkin tazminat ödemesinin reddedilmesi.
 - ✓ **Eksik evrak bekleniyor:** İncelemede olan dosyayla ilgili eksik evrak beklenen hasar dosyası.
 - ✓ **Feragat:** Hasar dosyasına ilişkin sahtekârlığın tespit edilmesi sonrasında yargı sürecine girmeksizin sigortalıyla anlaşarak sigortalının tazminattan feragat etmesi sebebiyle tazminat talebinin reddedilmesi veya sigorta şirketinin hasar dosyasını yanlışlıkla açması sonrasında iptal etmesi.

Bu noktada, veri setinde dosyanın sahtekârlık olup olmadığını belirleyici nokta “dosya_durum” deęişkenidir. “Feragat” alt durumundaki olan dosyalardan dosya durumu “kapalı-red” olanlar sahtekârlık şüphesi olanlar (fraud=1), “kapalı-iptal” durumunda olanlar ise yanlışlıkla açılan (fraud=0) dosyalardır.
 - ✓ **Garanti kapsamında olma nedeniyle:** Garanti kapsamında gerçekleşen bir hasar dosyasının reddedilmesi.
 - ✓ **Hasarın sigorta süresi dışında olması:** Hasar sigorta poliçe süresi dışında gerçekleştięi için tazminat ödemesinin reddedilmesi.
 - ✓ **Hatalı açılma:** Sigorta şirketi tarafından yanlışlıkla açılan hasar dosyası.
 - ✓ **Hukuk dosyası:** Hasarla ilgili süreçte taraflar arasındaki ihtilaftan kaynaklı olarak hukuk yoluna gidilen hasar dosyası.
 - ✓ **İlk dosya açılması:** Yeni açılmış ve herhangi bir işlem yapılmamış hasar dosyası.
 - ✓ **İncelemede:** Üzerindeki inceleme devam eden açık durumdaki hasar dosyası.
 - ✓ **Kısmi ödeme yapıldı:** Kısmi tazminat ödemesi yapılan hasar dosyası.
 - ✓ **Muafiyet nedeniyle red:** Hasar tutarının poliçede belirtilen muafiyet tutarının altında kalması sebebiyle tazminat talebinin reddedilmesi.

- ✓ **Ödeme günü verildi:** Hasar dosyasına ilişkin incelemeler tamamlanarak tazminat ödeme kararının alınması ancak henüz ödeme gerçekleşmemesi.
 - ✓ **Ödeme yapıldı:** Hasar dosyasına ilişkin tazminat ödemesi yapılması.
 - ✓ **Özel şartlara istinaden red:** Poliçe özel şartlarında yer alan teminat dışı bir halden dolayı tazminat ödemesinin reddedilmesi.
 - ✓ **Primin ödenmemesi:** Sigorta ettiren tarafından poliçe priminin ödenmemesi sebebiyle hasar dosyasına ilişkin tazminat talebinin reddedilmesi.
 - ✓ **Suistimal tespiti:** Hasar dosyasına ilişkin sahtekârlığın tespit edilmesi nedeniyle tazminat talebinin reddedilmesi.
 - ✓ **Sigortalı kusursuz:** Sigortalının bir kusuru olmadan yanlışlıkla açılan hasar dosyası.
 - ✓ **Tahsilat problemlili hasar dosyası:** Poliçe priminin tahsilatıyla ilgili sorun yaşanması sebebiyle hasar dosyasına ilişkin tazminat talebinin reddedilmesi.
 - ✓ **Talepten vazgeçildi:** Sigortalının çeşitli sebeplerden dolayı tazminat talebinden vazgeçmesi.
 - ✓ **Teminat limit yetersizliği:** Poliçede belirtilen teminat limitlerinin yetersiz olması sebebiyle tazminat talebinin reddedilmesi.
 - ✓ **Olay yeri terk:** Kaza sonrasında sigortalının olay yerini herhangi bir tutanak düzenlemeden terk etmesi.
 - ✓ **Uyumsuz hasar:** Sigortalı beyanıyla hasar bulgularının uyuşmaması ve sahtekârlık olarak değerlendirilmesi sebebiyle tazminat talebinin reddedilmesi.
 - ✓ **Zamanaşımı:** Zamanaşımı sebebiyle hasar dosyasına ilişkin tazminat talebinin reddedilmesi.
- **Ktt_var_mi:** Kaza sonrasında kaza tespit tutanağın tutulmuş mu? (0=hayır, 1=evet)
 - **Eksper_var_mi:** Hasar dosyası incelenirken eksper görevlendirilmiş mi? (0=hayır, 1=evet)
 - **Hasar_ihbar_tarihi:** Hasarın sigorta şirketine ihbar edilme/bildirilme tarihi.
 - **Police_baslangic_tarihi:** Poliçe vadesinin başlangıç tarihi.
 - **Police_bitis_tarihi:** Poliçe vadesinin bitiş tarihi.
 - **Hasar_il:** Hasarın gerçekleştiği il.
 - **Hasar_ilce:** Hasarın gerçekleştiği ilçe.

- **Kaza_arac_sayisi:** Kazaya karışan araç sayısı.
- **Surucu_farkli_mi:** Kazayı gerçekleştiren sürücü ile poliçede yer alan sürücü isimleri farklı mı? (0=hayır, 1=evet)
- **Arac_model_yil:** Sigortalı aracın model yılı.
- **Kullanim_sekli:** Sigortalı aracın ne amaçla kullanıldığı.
 - ✓ Özel
 - ✓ Resmi
 - ✓ Ticari
 - ✓ Bilinmiyor
- **Hasar_olus:** Hasar dosyasına konu kazanın oluş şekli.
 - ✓ Aksesuar yanması
 - ✓ Araç çarpışması
 - ✓ Anahtar çalınması
 - ✓ Bakımsızlık nedeniyle oluşan zararlar
 - ✓ Cam kırılması
 - ✓ Çalınma
 - ✓ Çarpma
 - ✓ Çarpışma
 - ✓ Çekme-Çekilme
 - ✓ Deprem
 - ✓ Devrilme
 - ✓ Diğer ek cihazlar çalınma
 - ✓ Dolu/ yıldırım/ fırtına/ don
 - ✓ Far/ stop/ sinyal/ ayna kırılması
 - ✓ GLKHH (Grev,lokavt,kötü niyetli hareketler ve halk hareketleri)
 - ✓ Görsel işitsel cihazlar (Radyo, Navigasyon, LCD)
 - ✓ Hayvana çarpma
 - ✓ İnsana çarpma
 - ✓ Kemirgen hayvanların vereceği zararlar
 - ✓ Mekanik, elektrik, elektronik donanım arızaları
 - ✓ Mini onarım

- ✓ Park halinde çarpma
- ✓ Sabit bir cisme çarpma
- ✓ Sel, su baskını
- ✓ Yanma
- **Rucu_var_mi:** Hasar dosyasında rücu¹ işlemi yapıldı mı? (0=hayır, 1=evet)
- **Teminat_artis:** Sigortalı poliçe teminatlarını artırmış mı? (0=hayır, 1=evet)
- **Police_bosluk:** Son iki poliçesi arasında kaç günlük boşluk olduğu.
- **Servisler_farkli_mi:** Hasar ihbarının yapıldığı servis ile onarım işleminin yapıldığı servis farklı mı? (0=hayır, 1=evet)
- **Sigortalilik_suresi:** Kaza anı itibariyle sigortalının o şirketteki sigortalılık süresi. (yıl)
- **Ortak_kaza_sayisi:** Hasar dosyasındaki kazaya karışan tarafların toplamda ortak olarak karıştıkları kaza sayısı.
- **Dosya_tutari:** Hasar dosyasında sigorta şirketinin ödeyeceği maddi hasarlara ilişkin tazminat tutarı.
Dosya tutarı değişkeni doldurulurken sigorta şirketleri tarafından izlenen yöntem şu şekildedir: Eğer eksper raporu varsa; hasara ilişkin rapordaki eksper tespiti dosya tutarı olarak kaydedilmiştir. Ancak eksper raporu olmadığında; dosya kapalıysa ödenen tazminat tutarı, dosya açıksa muallak hasar tutarı ve varsa kısmi ödenen tazminatın toplamı dosya tutarı olarak alınmıştır. Eğer hasar dosyasına ilişkin yukarıda bahsedilen bilgilerden hiçbiri mevcut değilse aynı sigorta türü ve kaza oluş şekli bazında yapılan ödemelere ilişkin ortalama değer alınmıştır.
- **Servis_yetkili_mi:** Kaza sonrasında onarım yapan yer yetkili servis mi? (0=hayır, 1=evet)
- **ihbar_baslangic:** (Kaza ihbar tarihi - poliçe başlangıç tarihi) kaç gün?
- **ihbar_bitis:** (Poliçe bitiş tarihi - kaza ihbar tarihi) kaç gün?
- **Arac_yasi:** Kaza anındaki aracın yaşı.

¹ “Rücu: Sigorta konusu olan şeyin; sigortalının herhangi bir kusuru, ihmali veya poliçe şartlarını ihlali gibi bir fiili olmaksızın, tamamen üçüncü bir şahsın eylemi ve teminat kapsamındaki tehlikelerden herhangi biri nedeniyle hasarlanması durumunda, sigortacının hasarı sigortalıya ödedikten sonra, sigortalının yerine geçmesi ve sigortalının bu hasara ilişkin olarak üçüncü şahıs nezdindeki tüm yasal alacaklarının yeni sahibi olmasıdır.” [2]

İş kuralları: Bu değişkenlere ilave olarak, sigorta şirketleri bugüne kadarki hasar tecrübelerine dayanarak sahtekârlık şüphesini artıran bazı durumlar belirlemiş, bu durumlar da iş kuralı haline getirilerek SİSEB'in veri tabanına değişken olarak eklenmiştir. İlgili durumların görüldüğü dosyalar 1, görülmediği dosyalar 0 olarak işaretlenmiştir. Sigorta şirketleri tarafından belirlenen bu iş kuralları Çizelge 6.1.'de görülmektedir.

Çizelge 6.1. Veri düzenlenmeden önceki mevcuttaki iş kuralları

K1: Olay cuma akşamı ve pazartesi sabahı arasında meydana gelmiştir.
K2: Olay tarihi ile ihbar tarihi arasında büyük zaman farkı bulunmaktadır.
K3: Olay olağandışı bir zamanda meydana gelmiştir.
K4: Hasar ihbar tarihi poliçenin bitiş tarihine ya da yenilenme tarihine yakındır.
K5: Birden çok tazminat talebinde bulunan taraf ve ödeme yapılan bir tane yetkili olmayan servis bulunmaktadır. (Trafik)
K6: Tazminat talebinde bulunan tarafın aracı eskidir.
K7: Hasar dosyasında yer alan araçlardan biri perte çıkmış, çalınmış veya büyük oranda hasar görmüş olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır.
K8: Kazada yer alan taraflardan herhangi biri kazadan önce SİSBİS şüpheli veri tabanına girmiştir.
K9: Sigortalı, geçmişte aynı veya farklı poliçeler üzerinden birçok (son 2 yılda 3'ten fazla) hasar dosyasında sigortalı olarak yer almıştır. (Trafik)
K10: Tazminat talebinde bulunan taraf olaydan önceki birçok (son 2 yılda 3'ten fazla) hasar dosyasında sigortalı olarak yer almıştır. (Trafik)
K11: Kazaya karışmış olan tazminat talebinde bulunan veya sigortalı son 6 ay içerisinde farklı bir sigorta şirketi tarafından reddedilmiş ya da suistimal tespit edilmiş hasar dosyası bulunmaktadır.
K12: Kazaya karışmış olan aracın son 6 ay içerisinde farklı bir sigorta şirketi tarafından reddedilmiş ya da suistimal tespit edilmiş hasar dosyası bulunmaktadır.
K13: Kazaya karışmış olan tazminat talebinde bulunan veya sigortalı son 1 yıl içerisinde farklı bir sigorta şirketi tarafından reddedilmiş ya da suistimal tespit edilmiş hasar dosyası bulunmaktadır.

K14: Kazaya karışmış olan aracın son 1 yıl içerisinde farklı bir sigorta şirketi tarafından reddedilmiş ya da suistimal tespit edilmiş hasar dosyası bulunmaktadır.
K15: Kazaya karışmış olan tazminat talebinde bulunan veya sigortalı son 2 yıl içerisinde farklı bir sigorta şirketi tarafından reddedilmiş ya da suistimal tespit edilmiş hasar dosyası bulunmaktadır.
K16: Kazaya karışmış olan aracın son 2 yıl içerisinde farklı bir sigorta şirketi tarafından reddedilmiş ya da suistimal tespit edilmiş hasar dosyası bulunmaktadır.
K17: Kazaya karışan araçlardan biri motosiklettir.
K18: Sigortalının aracı geçmişte şimdikinden farklı bir plaka ile birçok hasar dosyasında yer almıştır. (plaka değişikliği olmuş ama şasi no aynı)
K19: Poliçe üzerindeki hasar dosyalarının sayısı oldukça yüksektir.
K20: Olay poliçe başlangıç tarihinden kısa bir süre sonra olmuştur.
K21: Olaydan kısa bir süre önce teminat miktarını artıran bir zeyilname yapılmıştır.
K22: Tazminat talebinde bulunan taraflardan birinin aracında ciddi hasar olmasına rağmen kaza tarihi ile ihbar tarihi arasında gecikme olmuştur.
K23: Hasar dosyasındaki tazminat talebinde bulunanların sayısı diğer hasar dosyalarına oranla yüksektir.
K24: Açık dosyalar için muallak tutarı, kapalı dosyalar içinse ödenen tutar bu tür kaza/kayıp türü için yüksektir.
K25: Tazminat talebinde bulunan tarafın geçmişte yapmış olduğu birçok (son 2 yılda 3'ten fazla) tazminat talebi bulunmaktadır.
K26: "Rent a car" şüphesi vardır.
K27: Hasar dosyasında yer alan araçlardan biri perte çıkmış olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır. (Kasko)
K28: Hasar dosyasında yer alan araçlardan biri çalınmış olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır.
K29: Hasar dosyasında yer alan araçlardan biri büyük oranda hasar görmüş olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır.
K30: Tamirhane kazadan önce SİSBİS şüpheli veri tabanına girilmiştir.

K31: Sigortalı, sürücü ya da tazminat talebinde bulunan taraf kazadan önce SİSBİS şüpheli veri tabanına girilmiştir.
K32: Hasarın bağlı olduğu poliçede boşluk vardır.
K33: Sigortalının aracının yaşı eskidir ve ihbar yetkili tamirhane tarafından yapılmıştır.
K34: İhbar yapan tamirhane ile işlemi gerçekleştiren tamirhane farklıdır.
K35: Hasar dosyasında yer alan taraflar geçmişteki bir hasar dosyasında da birlikte taraf olarak yer almıştır.
K36: Kaza tek araçlı ve yüksek tutarlıdır.

Ancak bu iş kurallarının hepsi modelde kullanılacak veriye dâhil edilmemiş, farklı sebeplerle uygun olmayanlar çıkarılarak aralarından belirli iş kuralları çalışmaya girmiştir. İş kurallarında yapılan elemelere ve kalan değişkenlere ilişkin detaylı açıklamalara bir sonraki bölümde değinilmiştir.

6.2. Veri Düzeltme İşlemleri

SBM'den temin edilen ham veride sigorta şirketlerinin farklı ve hatalı veri girişleri, değişkenlerin yaptığımız çalışmayla uyumu ve değişkenlerin tutarlığı gibi konular göz önünde bulundurularak modele geçmeden önce bazı veri düzeltme işlemleri yapılmıştır. Bu başlık altında yapılan veri ve değişken düzeltmeleri değinilmektedir.

Fraud yanıt değişkeni, sigorta şirketleri tarafından bir sahtekârlığın tespiti ya da güçlü şüphesi sonrasında “1” olarak etiketlenebildiği gibi SBM’de gerçekleştirilen bir inceleme sonrasında da yetkililerce etiketlenebilmektedir. Eğer sigorta şirketi dosyayı suistimal vardır şeklinde etiketlediyse, dosya alt durumu sigorta şirketinin belirlediği doğrultuda kalır ve veri tabanında SBM tarafından değiştirilmez. Ancak sigorta şirketi hasar dosyasının suistimal durumuyla ilgili bir etiketleme yapmadıysa dosya alt durumlarının “feragat (kapalı-red), uyumsuz hasar, suistimal tespiti” olması durumunda ilgili dosyalara ilişkin fraud değişkeni “1” olarak etiketlenmektedir. Özetle, yanıt değişkeni olan “fraud” hem sigorta şirketlerinin etiketlemeleriyle hem de SBM tarafından “dosya_durum” ve “dosya_alt_durum” değişkenleri kullanılarak oluşturulmuş bir değişkendir. Dolayısıyla yanıt değişkeni, bahsi geçen bu iki

değişkenle ilişkili olduğundan kurulan modelde bu durum göz önünde bulundurularak, bu değişkenler modele dâhil edilmemiştir.

Ayrıca, “hasar_il”, “hasar_ilce” ve “hasar_olus” kategorik değişkenleri çok fazla sınıf içermesi, verilerin bu sınıflardaki frekanslarının çok küçük olması, yanlış ve eksik girdiler içermesi sebepleriyle çalışmaya olumlu yönde bir katkı sağlamayacağından modele dâhil edilmemiştir.

Buna ek olarak sigorta şirketi çalışanlarının veri girişi yaparken hatalı seçimler yapması durumları olabilmektedir. Örneğin dosya alt durumu “uyumsuz hasar” ve “suistimal tespiti” olan hasar dosyalarıyla ilgili bir sahtekârlık durumu şirket tarafından tespit edildiği halde “fraud” değişkeni “0” olarak etiketlenmiştir. Çalışma kapsamında bu alandaki bilgi kaybının önlenmesi amacıyla söz konusu duruma sahip hasar dosyalarında “fraud” etiketi “1” yapılarak veri düzeltilmiştir. Genel olarak veride sigorta şirketi, tamirhaneler ve SBM çalışanları tarafından girilen mükerrer satırlar olduğu tespit edilmiş, bu satırlar temizlenmiştir. “araç_yasi” değişkeninde 0’dan küçük değere sahip olan satırlar, “ihbar_baslangic” ve “ihbar_bitis” değişkenleri altında 0-366 gün aralığı dışında değer alan satırlar anlam olarak tutarsızlıklar içermesi sebebiyle silinmiştir. Tüm alt sınıfları “bilinmiyor” şeklinde işaretlenmesi sebebiyle “kullanım_sekli” değişkeni, başka değişkenlerle nümerik olarak ifade edilebilmeleri sebebiyle “police_baslangic_tarihi”, “police_bitis_tarihi”, “hasar_ihbar_tarihi” ve “araç_model_yili” değişkenleri modele dâhil edilmemiştir.

Bunlara ilave olarak, veri setinde yer alan iş kurallarının tümü çalışmaya dâhil edilmemiş, bu iş kurallarından nicel ya da nitel olarak başka bir değişkenle ifade edilebilen, belirsizlik ya da yorum taşıyan kurallar çalışmaya değişken olarak eklenmemiştir. Veri setinde bulunan 36 sayıdaki iş kuralları aşağıda belirtilen sebepler çerçevesinde 9’a düşürülmüştür. İş kuralları değişkenlerinin veriden çıkartılma sebeplerini üç grupta inceleyebiliriz.

- Öznel yargılar içermesi ve sigorta şirketleri tarafından kayıt oluşturulurken farklılıklara ve hatalara açık olması sebebiyle veriden çıkartılan iş kuralları: **K2, K3, K4, K19, K22, K23, K24.**

Örneğin, “K4:Hasar ihbar tarihi, poliçenin bitiş tarihine ya da yenileme tarihine yakındır.” iş kuralı hem durumu sayısal olarak ifade eden bir değişkenin olmasından ve hem de öznel yorum içermesinden dolayı veriden çıkarılmıştır. Verideki “ihbar_baslangic” ve “ihbar_bitis” değişkenleriyle söz konusu iş kuralının içerdiği bilgiye sayısal olarak erişilebilir. Ayrıca tarihin yakın olması hususu sigorta şirketlerine göre farklılık gösterebilecek bir uygulama olup veri girişi yapan kişiye göre değişebilmektedir. Poliçe bitişine kaç gün kala ya da poliçe başladıktan ne kadar sonra hasarın olması şüpheli bir durum yaratıyorsa, bu konu model kurulduktan sonra daha doğru bir şekilde varılacak bir bilgidir. Benzer şekilde K2 ve K19 kurallarına da baktığımızda, “büyük zaman farkı” ve “oldukça yüksektir” gibi ifadeler kişisel yorumlamaya açık olduğundan veri girişi yapan kişiye ve sigorta şirketine göre farklılık gösterebilir.

- Sigorta şirketlerinin, trafik sigortası (Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası) için belirledikleri ve kullandıkları iş kurallarının kasko sigortasının çalışma mantığına uymaması sebebiyle çıkarılan iş kuralları: **K5, K9, K10.**
- İlgili durumu gösteren nicel ya da nitel başka bir değişkenin olması sebebiyle veriden çıkartılan iş kuralları ise Çizelge 6.2’de gösterilmektedir.

Çizelge 6.2. Başka bir değişkenle karşılanması sebebiyle çıkarılan iş kuralları

Veriden Çıkarılan İş Kuralları	Durumu Karşılıyan Verideki Değişkenler
K4	ihbar_baslangic, ihbar_bitis
K6	arac_model_yil, araç_yasi
K7	IK7, IK8, IK9
K11, K12, K13, K14, K15, K16 K30, K31	IK2
K20	ihbar_baslangic
K21	teminat_artis
K32	police_bosluk, ihbar_baslangic, ihbar_bitis
K33	arac_model_yil, araç_yasi, servis_yetkili_mi
K34	servisler_farkli_mi
K35	ortak_kaza_sayisi
K36	kaza_arac_sayisi, dosya_tutari

Veride kalan iş kuralı değişkenleri, yeni isimleri ve açıklamaları Çizelge 6.3'teki gibidir.

Çizelge 6.3. Çalışmaya dâhil edilen iş kurallarının eski ve yeni isimleri

Değişkenin Eski İsmi	K1	K8	K17	K18	K25	K26	K27	K28	K29
Değişkenin Yeni İsmi	IK1	IK2	IK3	IK4	IK5	IK6	IK7	IK8	IK9

- **IK1:** Olay cuma akşamı ve pazartesi sabahı arasında meydana gelmiştir.
- **IK2:** Kazada yer alan taraflardan herhangi biri kazadan önce SİSBİS şüpheli veri tabanına girmiştir.
- **IK3:** Kazaya karışan araçlardan biri motosiklettir.
- **IK4:** Sigortalının aracı geçmişte şimdikinden farklı bir plaka ile birçok hasar dosyasında yer almıştır. (plaka değişikliği olmuş ama şasi no aynı)

- **İK5:** Tazminat talebinde bulunan tarafın geçmişte yapmış olduğu birçok (son 2 yılda 3'ten fazla) tazminat talebi bulunmaktadır.
- **İK6:** “Rent a car”² şüphesi vardır.
İK6 değişkeni, “Rent a car” yani araç kiralama şirketi şüphesini barındıran, aşağıda belirtilen durumlardan hepsini sağlıyorsa 1, sağlamıyorsa 0 değerini alır.
 - Araç yaşı 3 ve 3'ten küçükse.
 - Sürücü ve poliçe sahibi farklıysa.
 - Kişiye ait yürürlükteki kasko poliçe sayısı 4 ve 4'ten büyükse.
- **İK7:** Hasar dosyasında yer alan araçlardan biri perte³ çıkmış olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır.
- **İK8:** Hasar dosyasında yer alan araçlardan biri çalınmış olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır.
- **İK9:** Hasar dosyasında yer alan araçlardan biri büyük oranda hasar görmüş olarak geçmişte oluşmuş bir tazminat talebinde yer almıştır.

Yapılan bu düzeltmeler sonrasında veri setinde; 342.918 adet “0” etiketli, 3753 adet “1” etiketli olmak üzere toplamda 342.918 adet veri bulunmaktadır. Ayrıca değişkenlerde yapılan elemelerden sonra veri analizinde kullanılan değişkenler Çizelge 6.4.'te gösterilmektedir.

² “Rent a car” yani araç kiralama şirketleri için sigortaya kabul şartları ve prim fiyatları kişisel (özel) kullanım araçlarına göre farklılık göstermektedir. Bu şirketler daha yüksek prim ödememek amacıyla durumu sigorta şirketine bildirmeyip araçlarını özel kullanım için sigortalıyormuş gibi hareket etme yoluna gidebilirler. Bu durum da sigortacıyı aldatmaya yönelik bir hareket olduğundan suistimal şüphesini artırıcı bir değişken olarak düşünülerek çalışmaya dâhil edilmiştir.

³ Pert terimi, halk ağzında ve sigortacılık uygulamalarında sıkça kullanılan bir kelime olup herhangi bir trafik kazasında aracın çok ağır hasar alarak (tam hasar) kullanılamaz hale gelmesi anlamına gelmektedir.

Çizelge 6.4. Veri analizinde kullanılan değişkenler

Değişken adı	Türü	Değişken adı	Türü
Fraud(Yanıt değişkeni)	Kategorik	ihbar_baslangic	Nicel (Kesikli)
Ktt_var_mi	Kategorik	ihbar_bitis	Nicel (Kesikli)
Eksper_var_mi	Kategorik	Arac_yasi	Nicel (Kesikli)
Kaza_arac_sayisi	Nicel (Kesikli)	IK1	Kategorik
Surucu_farkli_mi	Kategorik	IK2	Kategorik
Rucu_var_mi	Kategorik	IK3	Kategorik
Teminat_artis	Kategorik	IK4	Kategorik
Police_bosluk	Nicel (Kesikli)	IK5	Kategorik
Servisler_farkli_mi	Kategorik	IK6	Kategorik
Sigortalilik_suresi	Nicel (Sürekli)	IK7	Kategorik
Ortak_kaza_sayisi	Nicel (Kesikli)	IK8	Kategorik
Dosya_tutari	Nicel (Sürekli)	IK9	Kategorik
Servis_yetkili_mi	Kategorik		

6.3. Veri Analizi ve Sonuçlar

Çalışmada, kasko sigortası hasar verilerindeki özelliklerden yola çıkarak ve sahtekârlık olarak etiketlenmiş geçmiş verilerden yararlanarak gözetimli öğrenme yöntemlerinden en iyi performansı veren yönteme ulaşmak amaçlanmıştır. Literatürde otomobil sigortasında sahtekârlık tespitine yönelik farklı ülke verileri üzerinden çeşitli yöntemlerle yapılan çalışmalar incelenmiş, ancak ülkemizde tüm sigorta şirketlerinin kasko sigortası hasar verilerini kapsayacak genellikte bir çalışmanın olmadığı görülmüştür. Buradan yola çıkarak SBM'den alınan verilerle literatürde incelenen yöntemler Türkiye verilerine uygulanmıştır. Bu kapsamda,

- Lojistik Regresyon Analizi
- Bayes Ağları (Naive Bayes, Hill Climbing, Max-Min Hill Climbing)
- Karar Ağaçları (C.4.5, Bagging, Random Forest)
- Yapay Sinir Ağları
- Destek Vektör Makineleri

yöntemleri uygulanmıştır. Çalışmada veri analizi için “R” programı kullanılmıştır.

Bu yöntemlerin tahmin performanslarını karşılaştırmak için Çapraz Doğrulama yapılmıştır. Çapraz Doğrulamada amaç, oluşturulan modelin performansının belirlenmesidir. Bu kapsamda veri iki parçaya ayrılarak, birinci parça sınıflandırma yapacak modelin oluşumundaki eğitim verisi, ikinci parça ise modelin performansının belirlenmesindeki test verisi olarak kullanılır. Bu şekilde oluşturulan modelin test verisine uygulanması sonrasında doğru sınıflandırılan örneklerin oranı sınıflandırıcının performansını gösterir.

Çalışmada rastgele şekilde verilerin %80'i eğitim ve %20'si test verisi olarak ayrılmıştır. Modeller eğitim verileri ile öğrenmeyi gerçekleştirerek, hiç görmedikleri test verileri üzerinden tahmin yapmış ve bu işlem 50 kez tekrar edilmiştir. DVM'de, modelin çalışmasının çok uzun sürmesi ve mevcut fiziksel kaynakların modeli çok sayıda çalıştırmada yetersiz kalması sebebiyle sadece bu yöntem için 10 deneme yapılabilmektedir. Her tekrarda rastgele şekilde eğitim ve test verileri seçilmiş, bu şekilde örneklem seçiminin modeller üzerindeki etkisinin azaltılması ve çok tekrar yapılarak Büyük Sayılar Kanunu gereği performansların ortalamasının gerçek değere yaklaşması hedeflenmiş, aynı zamanda da aşırı uyumun önüne geçilmiştir. Sonrasında elde edilen performans ölçütlerinin ortalaması alınarak modeller karşılaştırılmıştır.

- **Lojistik Regresyon Analizi:**

Öncelikle literatürde de sıklıkla başvurulan bir yöntem olan lojistik regresyon sonuçlarına bakacak olursak, çalışmada geriye doğru eleme (backward elimination) metoduyla lojistik regresyon çalışması yapılmıştır. Başlangıçta tüm değişkenlerin dâhil olduğu ilk model gösterim olarak aşağıdaki gibi oluşturulmuştur:

$$\begin{aligned} \text{Fraud} = & \text{Ktt_var_mi} + \text{Eksper_var_mi} + \text{Kaza_arac_sayisi} + \text{Surucu_farkli_mi} \\ & + \text{Rucu_var_mi} + \text{Teminat_artis} + \text{Police_bosluk} + \text{Servisler_farkli_mi} \\ & + \text{Sigortalilik_suresi} + \text{Ortak_kaza_sayisi} + \text{Dosya_tutari} \\ & + \text{Servis_yetkili_mi} + \text{ihbar_baslangic} + \text{ihbar_bitis} + \text{Arac_yasi} + \text{IK1} \\ & + \text{IK2} + \text{IK3} + \text{IK4} + \text{IK5} + \text{IK6} + \text{IK7} + \text{IK8} + \text{IK9} \end{aligned}$$

Adımsal olarak (9 adım) değişkenler çıkarıldıktan sonra anlamlılığı en yüksek çıkan son modeldeki değişkenlerin gösterimi;

$$\begin{aligned} \text{Fraud} = & \text{Ktt_var_mi} + \text{Surucu_farkli_mi} + \text{Rucu_var_mi} + \text{Teminat_artis} \\ & + \text{Police_bosluk} + \text{Servisler_farkli_mi} + \text{Sigortalilik_suresi} \\ & + \text{Dosya_tutari} + \text{Servis_yetkili_mi} + \text{ihbar_baslangic} + \text{ihbar_bitis} \\ & + \text{Arac_yasi} + \text{IK1} + \text{IK2} \end{aligned}$$

şeklinde olup değişken katsayıları, anlamlılıkları Çizelge 6.5.'te gösterilmektedir.

Çizelge 6.5. Lojistik Regresyon analizinden elde edilen önemli değişkenler, katsayıları ve p değerleri

Değişkenler	Katsayısı	P değeri
Sabit	1,935	0,000*
Ktt_var_mi	-6,665	0,000*
Surucu_farkli_mi	0,095	0,084**
Rucu_var_mi	-1,204	0,000*
Teminat artis	-0,395	0,069**
Police_bosluk	0,008	0,001*
Servisler_farkli_mi	0,237	0,000*
Sigortalilik_suresi	-0,104	0,000*
Dosya_tutari	0,010	0,000*
Servis_yetkili_mi	-0,696	0,008*
İhbar_baslangic	-0,0282	0,000*
İhbar_bitis	-0,024	0,000*
Arac_yasi	0,661	0,000*
IK1	0,052	0,002*
IK2	0,347	0,086**

Modelin anlamlılığına bakmak için “Hosmer Lemeshow Uyum İyiliği Testi” yapılmış p değeri: 0,07 olarak Ki-kare test değeri:14.573 olarak bulunmuştur. Modelin anlamlılığına ilişkin hipotezi aşağıdaki gibi kuracak olursak,

H_0 = Verilerin kurulan modele uyumu vardır.

$p > \alpha = 0,05$ olduğundan yokluk hipotezi kabul edilir. Verilerin modele uyumunun olduğu %95 güven düzeyinde söylenebilir.

Değişkenlerin anlamlılığına bakacak olursak,

$$H_0: x = 0$$

$H_1: x = 0$ şeklinde hipotezler kurulduğunda,

*: ($p < \alpha = 0,05$) için yokluk hipotezi reddedilir ve değişken anlamlıdır. (%95 güven düzeyi)

** : ($p < \alpha = 0,1$) için yokluk hipotezi reddedilir ve değişken anlamlıdır. (%90 güven düzeyi)

Bu kapsamda lojistik regresyon denklemi aşağıdaki gibi yazılabilir:

$$P(Y = 1) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}$$
$$= \frac{e^{1,935 - 6,665x_1 + 0,095x_2 - 1,204x_3 - 0,395x_4 + 0,008x_5 + 0,237x_6 - 0,104x_7 + 0,01x_8 + \dots + 0,0282x_{14}}}{1 + e^{1,935 - 6,665x_1 + 0,095x_2 - 1,204x_3 - 0,395x_4 + 0,008x_5 + 0,237x_6 - 0,104x_7 + 0,01x_8 + \dots + 0,0282x_{14}}}$$

Bilindiği üzere “Fraud” bağımlı değişkeni için referans=1 olup sahtekârlığın varlığını ifade etmekteydi. Benzer şekilde diğer kategorik bağımsız değişkenlerde de 0=hayır ve referans olan 1=evet anlamına gelmekteydi. Bu bilgiler ışığında, değişkenlerin katsayılarını yorumlayacak olursak; bağımsız nicel değişkenin önünde pozitif katsayı varsa, bu değişkenin değeri arttıkça bağımlı değişkenin referans değerini alma olasılığı $P(y=1)$ artar; bağımsız kategorik değişkenlerin önünde pozitif katsayı varsa, bu değişkenin referans olan 1 değerini alması halinde bağımlı değişkenin 1 değerini alma olasılığı artar. Bu kapsamda otomobil sigortalarında,

- Kaza sonrasında kaza tespit tutanağının olmaması,
- Poliçede yazan sürücü ismiyle kazaya karışan sürücünün isminin farklı olması,
- Hasar dosyasında rücu imkânının olmaması,
- Poliçede ilave teminat artışının yapılmaması,
- Sigortalının iki poliçesi arasındaki sigortasız geçen sürenin fazla olması,
- Hasar ihbarının yapıldığı servis ile onarımın yapıldığı servisin farklı olması,
- Sigortalılık süresinin düşük olması,
- Hasar dosyası tutarının yüksek olması,

- Kaza sonrasında onarımı gerçekleştiren servisin yetkili olmaması,
- Kaza ihbar tarihinin poliçe başlangıç tarihine veya bitiş tarihine yakın olması,
- Araç yaşının yüksek olması,
- Kazanın cuma akşamı ve pazartesi sabahı arasında meydana gelmesi,
- Kazada yer alan taraflardan birinin daha önceden SİSBİS şüpheli veri tabanına girmesi

durumları hasar dosyalarındaki sahtekarlık olasılığını artıran değişkenler olarak yorumlanabilir.

Literatür taramasında yapılan çalışmalardan çıkan önemli değişkenlere bakacak olursak; sigortalı aracın eski ve değerinin düşük olması, polislin olay yeri raporunun olmaması, poliçe teminatının artırılmaması, kazanın gece veya hafta sonu gerçekleşmesi, poliçe düzenleme tarihi ile poliçe yürürlük tarihi arasında kaza yapılması gibi değişkenler çalışmamızın sonucunda çıkan önemli değişkenleri destekler niteliktedir.

Eğitim verisiyle oluşturulan lojistik regresyon modeli test verisi üzerinde tahmin işlemi yapılarak bu işlem 50 kere tekrarlanmıştır. Bu tekrarlardan elde edilen performans ölçütlerinin ortalama, standart sapma, minimum ve maksimum değerleri (Çizelge 6.6) aşağıdaki gibidir.

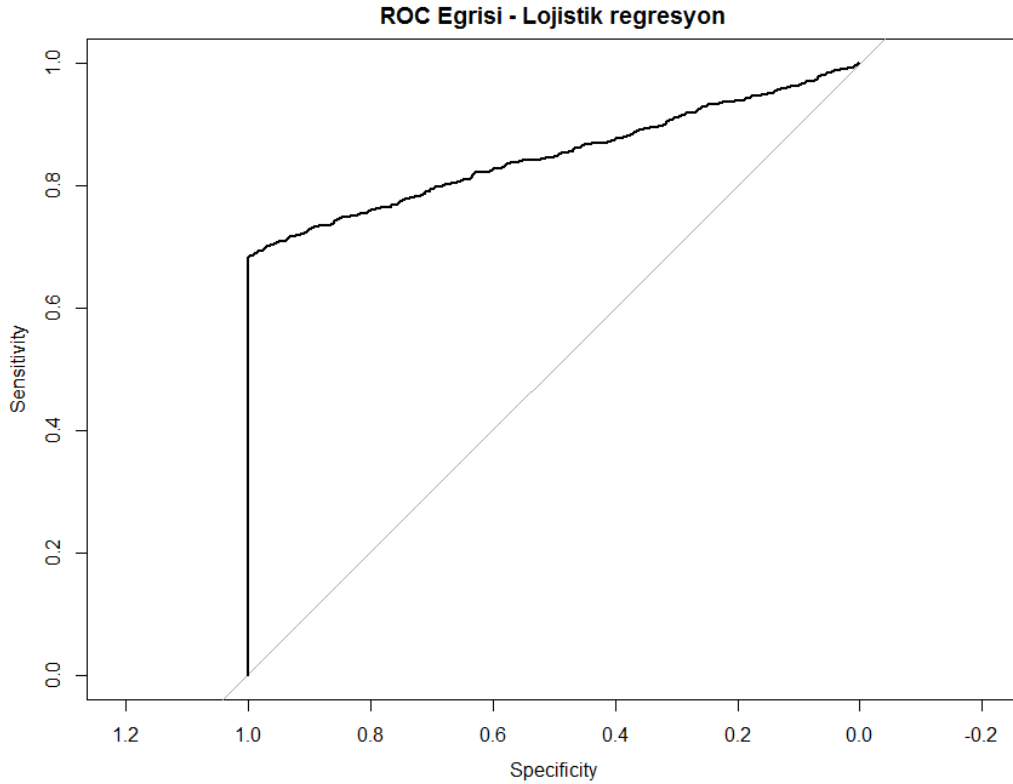
Çizelge 6.6. Lojistik Regresyon analizinden elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,995361	0,709196	0,998514	0,863787
Std. Sapma	0,000415	0,013592	0,000303	0,009435
Minimum	0,99305	0,672677	0,99688	0,842029
Maksimum	0,995821	0,732484	0,998887	0,884436

Buna ilave olarak bir deneme için elde edilen hata matrisi (Çizelge 6.7) ve ROC eğrisi (Şekil 6.1) aşağıda gösterilmektedir. Modellerin performansını karşılaştırmak için hata matrislerini karşılaştırmak doğru bir ölçüt olmamakla birlikte çalışmada temsili olarak gösterilmektedir.⁴

Çizelge 6.7. Lojistik Regresyon analizinden elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	541	44
	Negatif (0)	206	67.793



Şekil 6.1. Lojistik Regresyon analizinden elde edilen ROC eğrisi.

⁴ Modeller için yapılan tekrarların her birinde rasgele olarak örneklem seçimi (%80 eğitim, %20 test verisi) yapıldığından, toplamda 68.584 gözlem içeren test verilerindeki 0 ve 1 etiketli gözlemlerin sayısı her deneme için değişebilmektedir. Hata matrislerindeki gerçek değerlerin toplamındaki farklılıklar bu durumdan kaynaklanmaktadır.

Yukarıda görüldüğü üzere, lojistik regresyon modelinin %99,54 oranında gerçek ve sahte hasarlara ilişkin doğru sınıflandırma yaptığı, sahte hasarları %70,92 oranında yakaladığı, sahte olmayan hasarları da %99,85 oranında doğru tespit ederek masum kişileri suçlamadığı söylenebilir. ROC eğrisine ve AUC=0,864 değerine bakıldığında da modelin performansının iyi olduğu şeklinde yorumlanabilir.

- **Bayes Ağları:**

Eğitim verisiyle oluşturulan Bayes Ağları algoritmalarının (HC, MMHC ve Naive Bayes) test verisi üzerinde tahmin işlemi yapılarak bu işlemin 50 kere tekrarlanması sonrasında elde edilen performans ölçütlerinin ortalama, standart sapma, minimum ve maksimum değerleri (Çizelge 6.8, Çizelge 6.10, Çizelge 6.12) aşağıdaki gibidir.

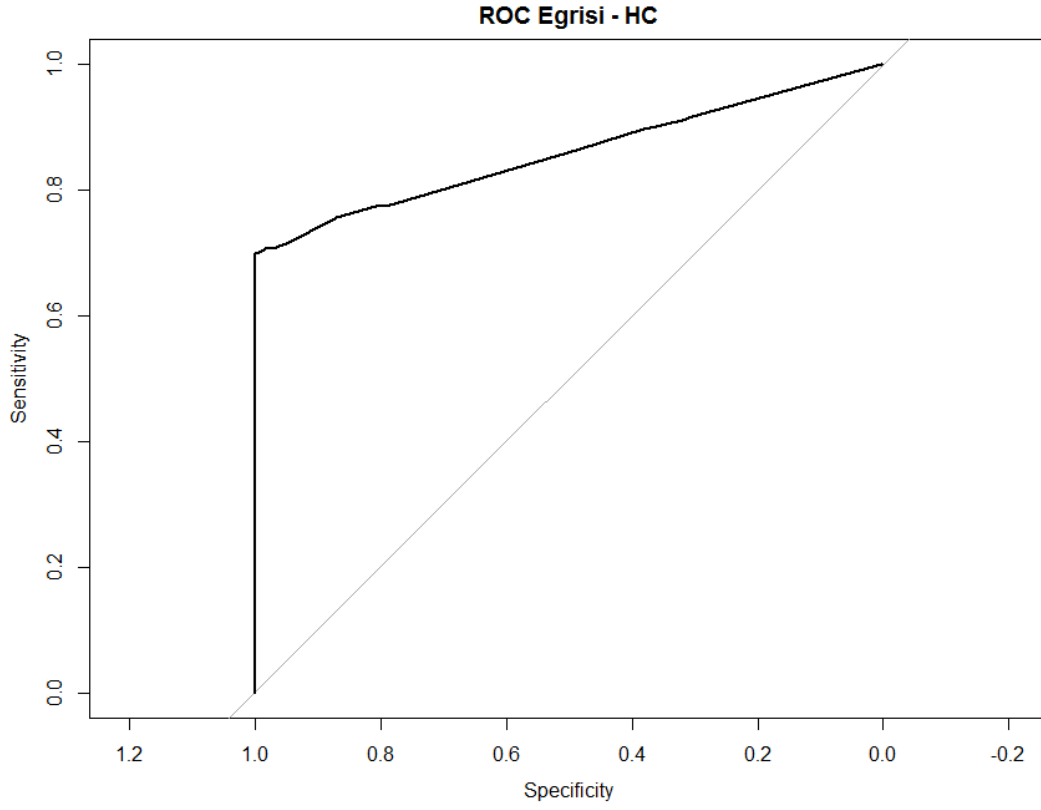
Buna ilave olarak bir deneme için elde edilen hata matrisleri (Çizelge 6.9, Çizelge 6.11, Çizelge 6.13) ve ROC eğrileri (Şekil 6.2, Şekil 6.3, Şekil 6.4) aşağıda gösterilmektedir. Modellerin performansını karşılaştırmak için hata matrislerini karşılaştırmak doğru bir ölçüt olmamakla birlikte çalışmada temsili olarak gösterilmektedir.

Çizelge 6.8. HC Bayes Ağından elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,996078	0,707821	0,999257	0,862460
Std. Sapma	0,000206	0,01355	0,000088	0,007649
Minimum	0,995684	0,680649	0,999042	0,844816
Maksimum	0,996646	0,741848	0,999499	0,876856

Çizelge 6.9. HC Bayes Ağından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	519	49
	Negatif (0)	243	67.773



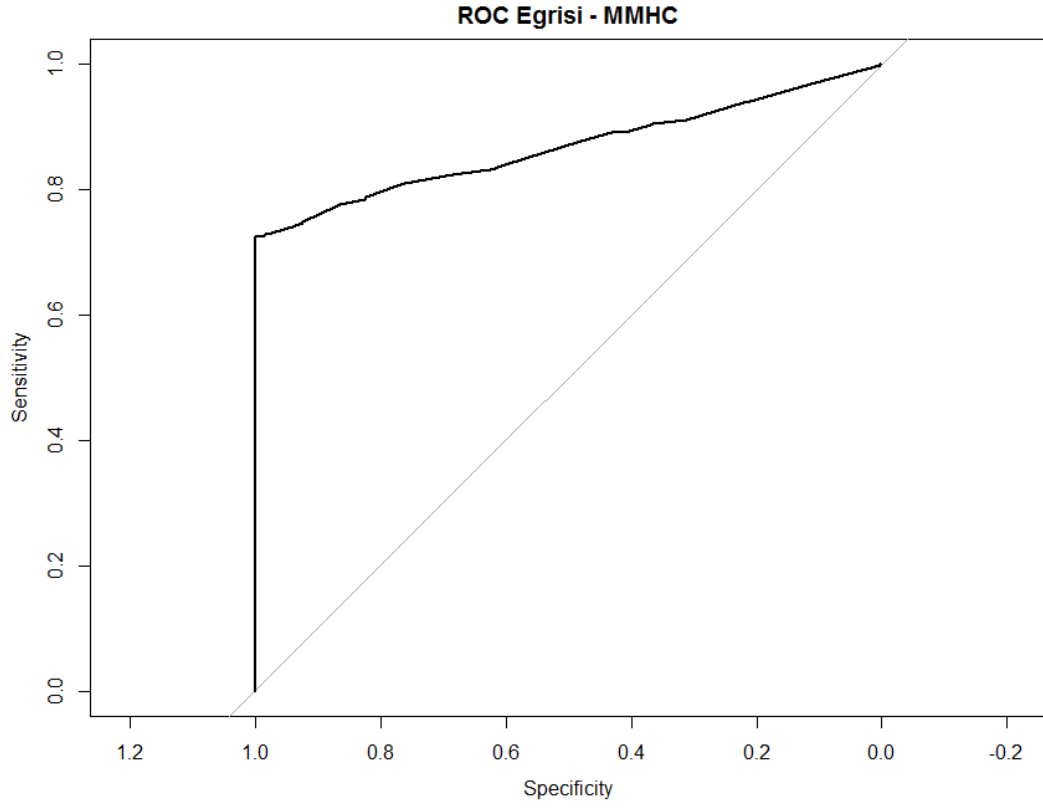
Şekil 6.2. HC Bayes Ağından elde edilen ROC eğrisi

Çizelge 6.10. MMHC Bayes Ağından elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,996071	0,708797	0,999280	0,856717
Std. Sapma	0,000197	0,017933	0,000099	0,009672
Minimum	0,995611	0,678161	0,999012	0,835263
Maksimum	0,996544	0,788024	0,999484	0,873443

Çizelge 6.11. MMHC Bayes Ağından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	537	45
	Negatif (0)	194	67.808



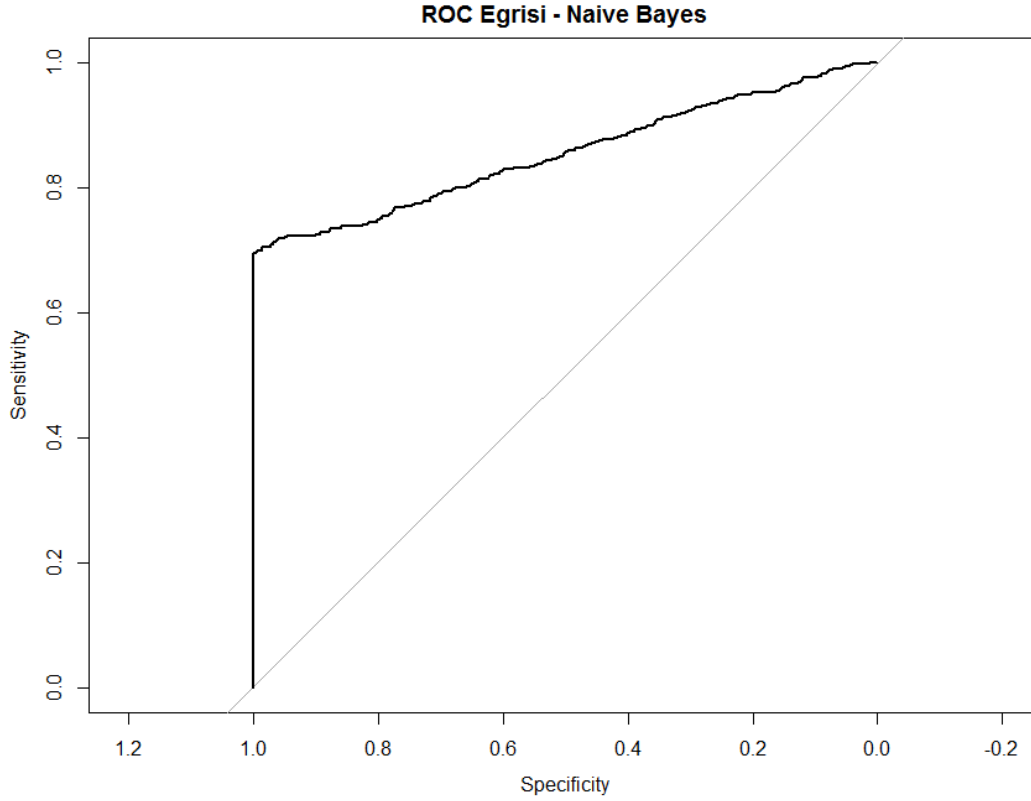
Şekil 6.3. MMHC Bayes Ağından elde edilen ROC eğrisi

Çizelge 6.12. Naive Bayes Sınıflandırıcılı Bayes Ağından elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,995686	0,707127	0,998898	0,857820
Std. Sapma	0,000229	0,014816	0,000108	0,007359
Minimum	0,995028	0,663239	0,998645	0,840095
Maksimum	0,996122	0,742404	0,99916	0,875756

Çizelge 6.13. Naive Bayes Sınıflandırıcılı Bayes Ağından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	551	36
	Negatif (0)	218	67.779



Şekil 6.4. Naive Bayes Sınıflandırıcılı Bayes Ağından elde edilen ROC eğrisi

Bayes ağlarının performans ölçütlerine bakarsak birbirlerine çok yakın sonuçlar verdikleri, %99,57- 99,61 oranında gerçek ve sahte hasarlara ilişkin doğru sınıflandırma yaptıkları, sahte hasarları %70,71-70,87 oranında yakaladıkları, sahte olmayan hasarları da %99,89-99,83 oranında doğru tespit ederek masum kişileri suçlamadıkları söylenebilir. ROC eğrilerine ve 85,7-86,2 aralığındaki AUC değerlerine bakıldığında da modelin performansının iyi olduğu şeklinde yorumlanabilir.

- **Karar Ağaçları:**

Eğitim verisiyle oluşturulan Karar Ağaçları algoritmalarının (C4.5, Bagging, Random Forest) test verisi üzerinde tahmin işlemi yapılarak bu işlemin 50 kere tekrarlanması sonrasında elde edilen performans ölçütlerinin ortalama, standart sapma, minimum ve maksimum değerleri (Çizelge 6.14, Çizelge 6.16, Çizelge 6.18) aşağıdaki gibidir.

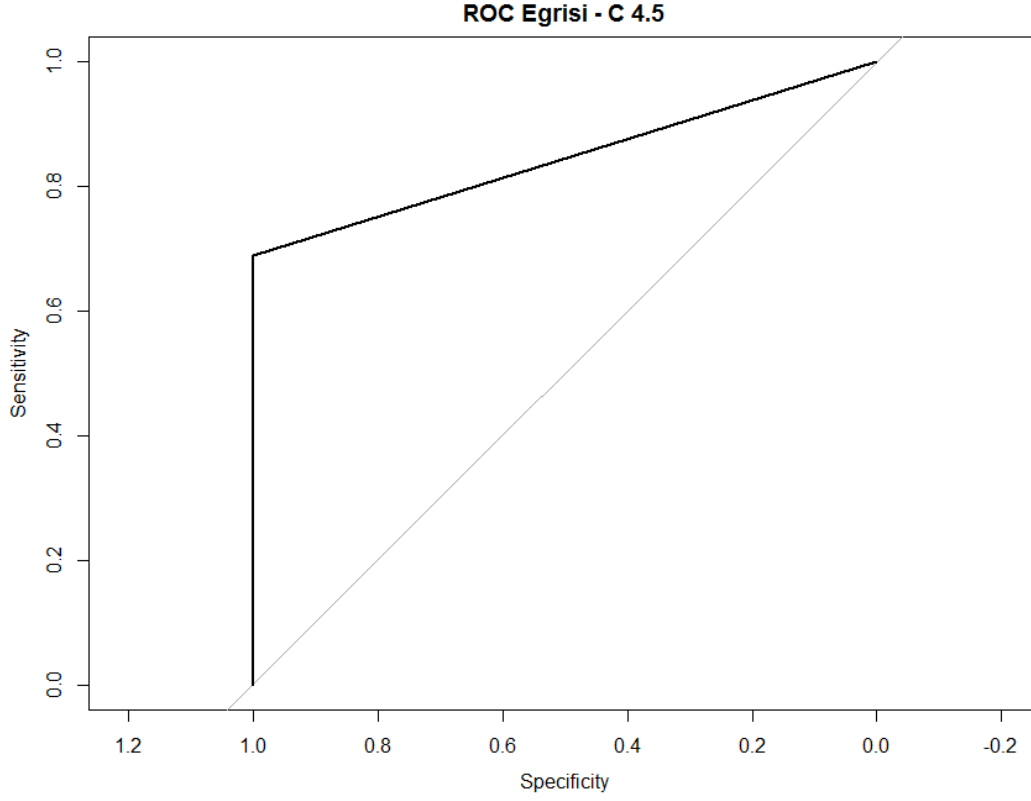
Buna ilave olarak bir deneme için elde edilen hata matrisleri (Çizelge 6.15, Çizelge 6.17, Çizelge 6.19 ve ROC eğrileri (Şekil 6.5, Şekil 6.6, Şekil 6.7) aşağıda gösterilmektedir. Modellerin performansını karşılaştırmak için hata matrislerini karşılaştırmak doğru bir ölçüt olmamakla birlikte çalışmada temsili olarak gösterilmektedir.

Çizelge 6.14. C4.5 Karar Ağacından elde edilen performans ölçütleri

	Doğru sınıflandırma oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,996129	0,710962	0,999298	0,852044
Std. Sapma	0,000109	0,008552	0,000058	0,008229
Minimum	0,995947	0,699229	0,999204	0,825118
Maksimum	0,996326	0,726923	0,999381	0,866555

Çizelge 6.15. C4.5 Karar Ağacından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	544	44
	Negatif (0)	234	67.762



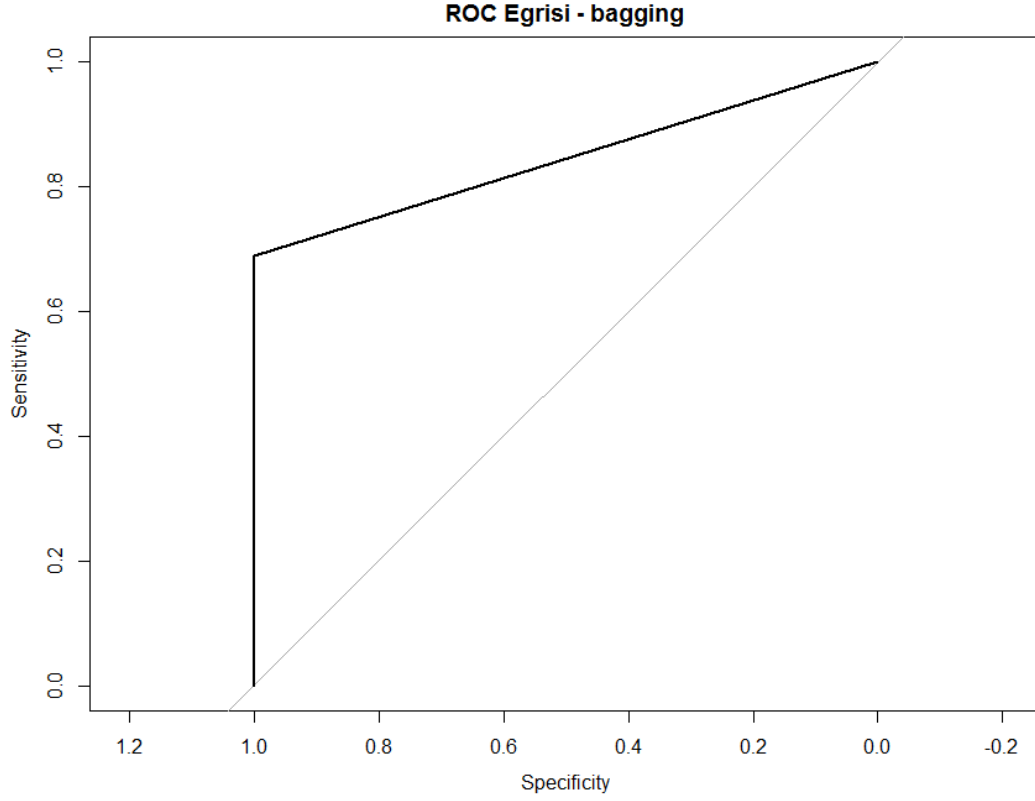
Şekil 6.5. C4.5 Karar Ağacından elde edilen ROC eğrisi

Çizelge 6.16. Bagging Karar Ağacından elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,996033	0,702599	0,999274	0,850936
Std.Sapma	0,000239	0,014401	0,000089	0,007211
Minimum	0,995203	0,661673	0,999131	0,830409
Maksimum	0,996399	0,728117	0,999484	0,863749

Çizelge 6.17. Bagging Karar Ağacından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	534	43
	Negatif (0)	221	67.786



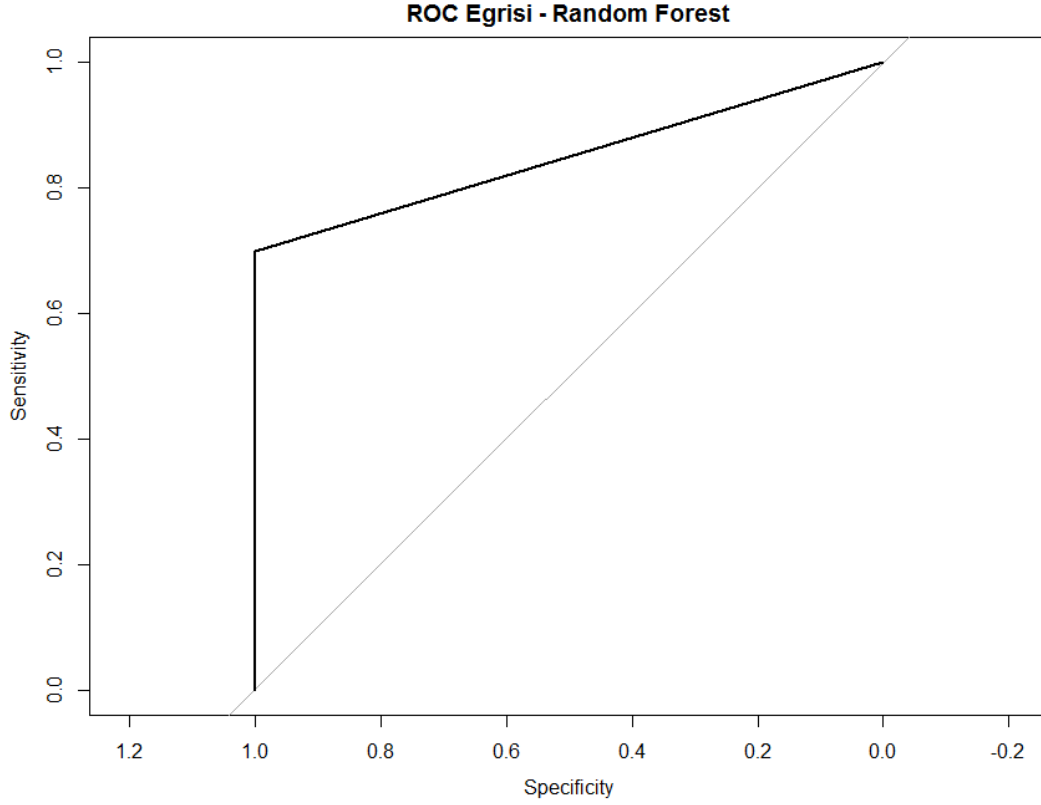
Şekil 6.6. Bagging Karar Ağacından elde edilen ROC eğrisi

Çizelge 6.18. Random Forest Karar Ağacından elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,996132	0,712277	0,999274	0,855775
Std. Sapma	0,000174	0,011764	0,00009	0,005874
Minimum	0,995742	0,688276	0,999042	0,843784
Maksimum	0,996559	0,74069	0,999469	0,869991

Çizelge 6.19. Random Forest Karar Ağacından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	513	54
	Negatif (0)	242	67.775



Şekil 6.7. Random Forest Karar Ağacından elde edilen ROC eğrisi

Karar ağaçlarının performans ölçütlerine bakarsak birbirlerine çok yakın sonuçlar verdikleri, %99,60- 99,61 oranında gerçek ve sahte hasarlara ilişkin doğru sınıflandırma yaptıkları, sahte hasarları %70,3-71,2 oranında yakaladıkları, sahte olmayan hasarları da %99,89-99,93 oranında doğru tespit ederek masum kişileri suçlamadıkları söylenebilir. ROC eğrilerine ve 85,2- 85,6 aralığındaki AUC değerlerine bakıldığında da modelin performansının iyi olduğu şeklinde yorumlanabilir.

- **Yapay Sinir Ağları:**

Eğitim verisiyle oluşturulan YSA'nın test verisi üzerinde tahmin işlemi yapılarak bu işlemin 50 kere tekrarlanması sonrasında elde edilen performans ölçütlerinin ortalama, standart sapma, minimum ve maksimum değerleri (Çizelge 6.20) aşağıdaki gibidir.

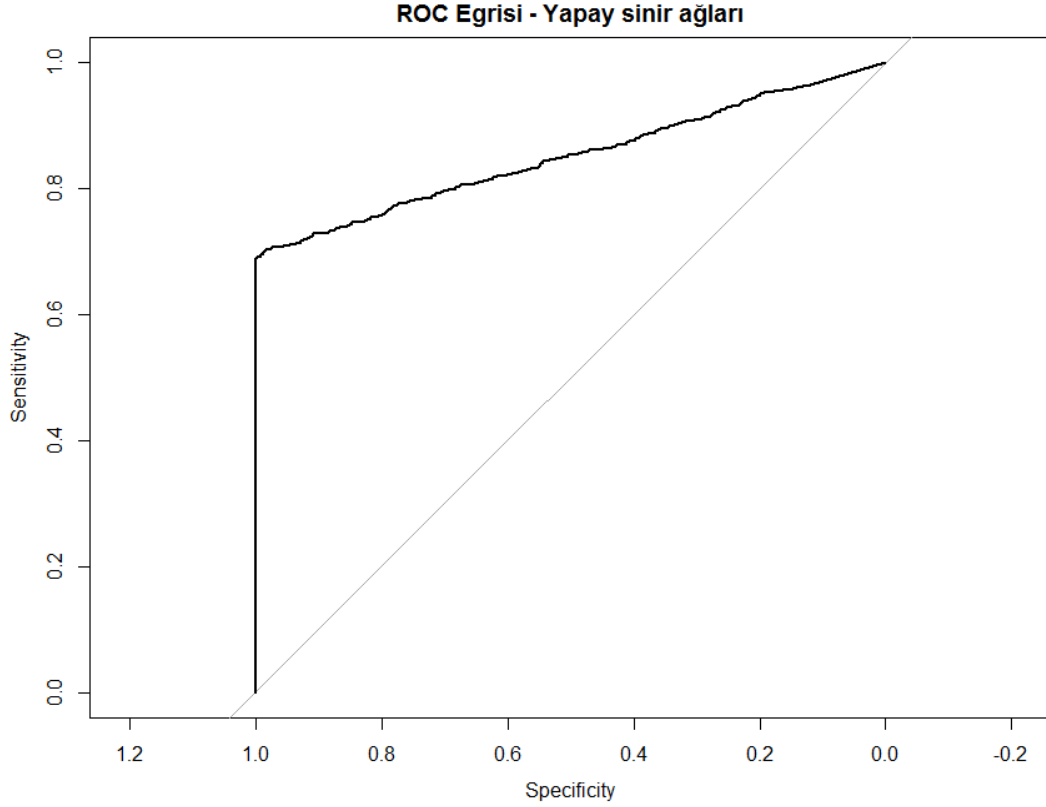
Çizelge 6.20. Yapay Sinir Ağlarından elde edilen performans ölçütleri

	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,995998	0,703853	0,999222	0,858588
Std.Sapma	0,000256	0,016365	0,000096	0,008933
Minimum	0,995567	0,672	0,998998	0,839983
Maksimum	0,996690	0,740692	0,999455	0,881366

Buna ilave olarak bir deneme için elde edilen hata matrisi (Çizelge 6.21), ROC eğrisi (Şekil 6.8) ve oluşturulan YSA yapısı (Şekil 6.9) aşağıda gösterilmektedir. Modellerin performansını karşılaştırmak için hata matrislerini karşılaştırmak doğru bir ölçüt olmamakla birlikte çalışmada temsili olarak gösterilmektedir.

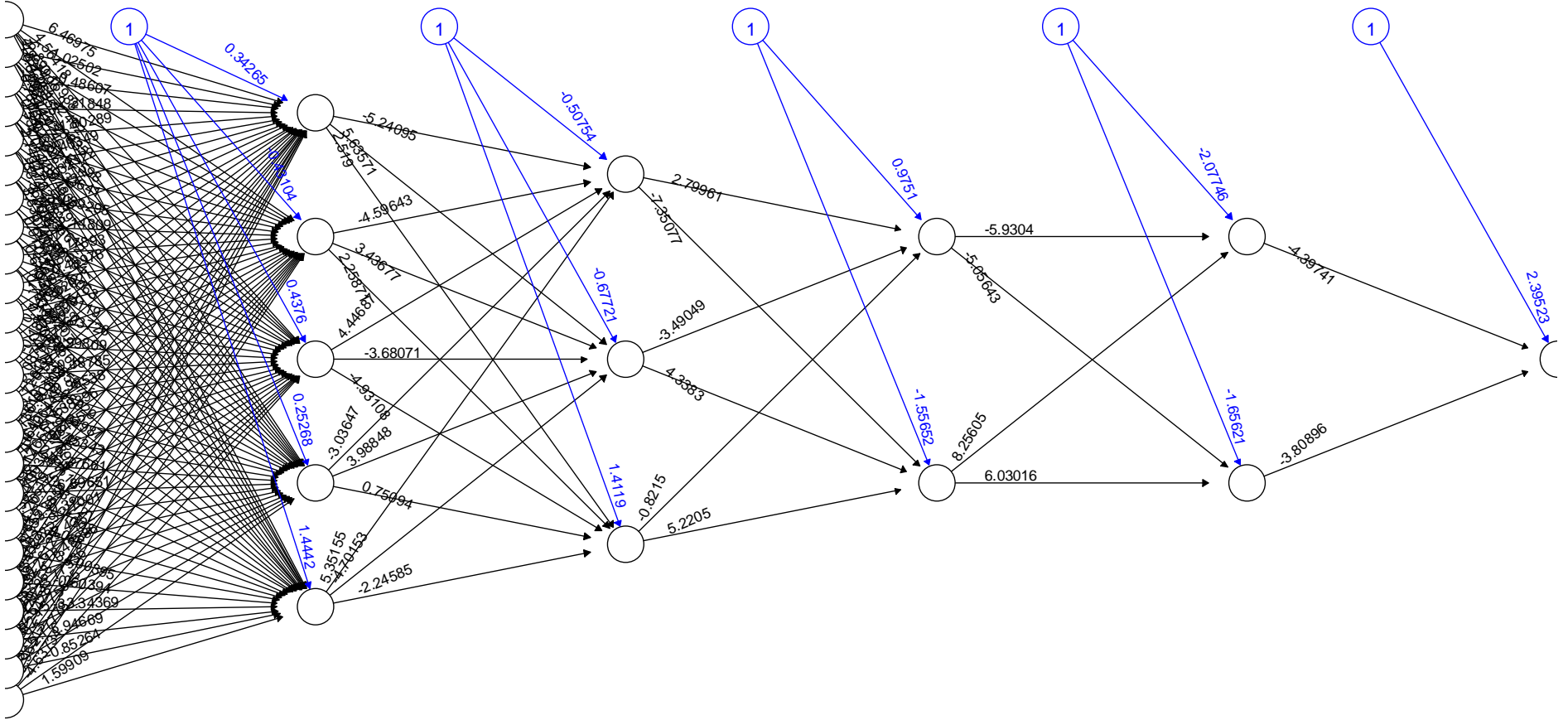
Çizelge 6.21. Yapay Sinir Ağlarından elde edilen hata matrisi değerleri

		Gerçek değerler	
		Pozitif (1)	Negatif (0)
Tahmin değerleri	Pozitif (1)	521	44
	Negatif (0)	247	67.772



Şekil 6.8. Yapay Sinir Ağlarından elde edilen ROC eğrisi

Yukarıda görüldüğü üzere, YSA'nın %99,6 oranında gerçek ve sahte hasarlara ilişkin doğru sınıflandırma yaptığı, sahte hasarları %70,38 oranında yakaladığı, sahte olmayan hasarları da %99,92 oranında doğru tespit ederek masum kişileri suçlamadığı söylenebilir. ROC eğrisine ve $AUC=0,858$ değerine bakıldığında da modelin performansının iyi olduğu şeklinde yorumlanabilir. Oluşturulan YSA yapısı Şekil 6.9.'daki gibidir.



Şekil 6.9. Çalışmada uygulanan Yapay Sinir Ağı yapısı

- **Destek Vektör Makineleri:**

Eđitim verisiyle oluřturulan DVM'nin test verisi üzerinde tahmin iřlemi yapılarak bu iřlemin 10 kere tekrarlanması⁵ sonrasında elde edilen performans ölçütlerinin ortalama, standart sapma, minimum ve maksimum deęerleri (Çizelge 6.22) ařađıdaki gibidir.

Çizelge 6.22. Destek Vektör Makinelerinden elde edilen performans ölçütleri

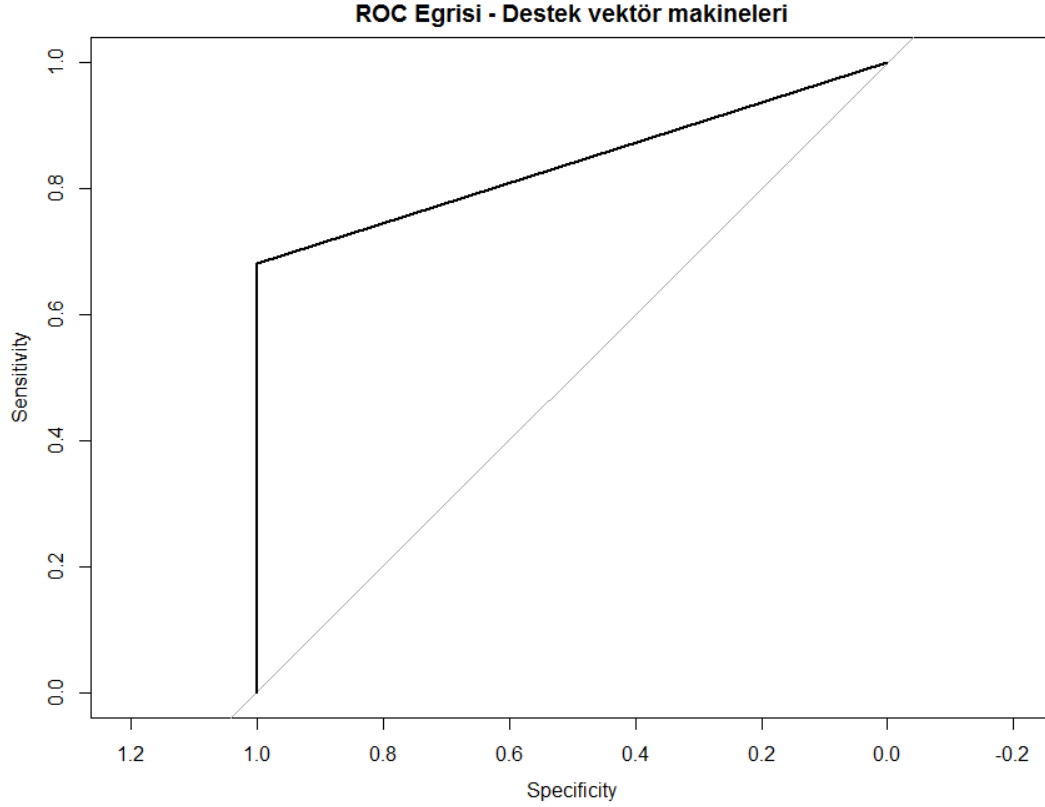
	Doęru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Ortalama	0,995941	0,697921	0,999258	0,850275
Std. Sapma	0,00018	0,012586	0,000105	0,00358
Minimum	0,995728	0,680965	0,998968	0,8443
Maksimum	0,996282	0,72372	0,999395	0,8536

Buna ilave olarak bir deneme için elde edilen hata matrisi (Çizelge 6.23) ve ROC eęrisi (řekil 6.10) ařađıda gösterilmektedir. Modellerin performansını karřılařtırmak için hata matrislerini karřılařtırmak doęru bir ölçüt olmamakla birlikte çalıřmada temsili olarak gösterilmektedir.

Çizelge 6.23. Destek Vektör Makinelerinden elde edilen hata matrisi deęerleri

		Gerçek deęerler	
		Pozitif (1)	Negatif (0)
Tahmin deęerleri	Pozitif (1)	501	50
	Negatif (0)	226	67.807

⁵ Destek Vektör Makinelerinde, modelin çalıřmasının çok uzun sürmesi ve mevcut fiziksel kaynakların modeli çok sayıda çalıřtırmada yetersiz kalması sebebiyle sadece bu yöntem için 10 deneme yapılabilmektedir. Modellerin çalıřma sürelerine Çizelge 6.25'te yer verilmiřtir.



Şekil 6.10. Destek Vektör Makinelerinden elde edilen ROC eğrisi

Yukarıda görüldüğü üzere, DVM'nin %99,6 oranında gerçek ve sahte hasarlara ilişkin doğru sınıflandırma yaptığı, sahte hasarları %69,79 oranında yakaladığı, sahte olmayan hasarları da %99,93 oranında doğru tespit ederek masum kişileri suçlamadığı söylenebilir. ROC eğrisine ve $AUC=0,85$ değerine bakıldığında da modelin performansının iyi olduğu şeklinde yorumlanabilir.

Uygulanan tüm yöntemlere ilişkin performans ölçütlerine tamamına Çizelge 6.24'te görülebilir.

Çizelge 6.24. Çalışmada uygulanan yöntemlerde elde edilen performans ölçütleri

Kullanılan Yöntem	Doğru Sınıflandırma Oranı	Duyarlılık	Belirleyicilik	AUC
Lojistik Regresyon	0,995361	0,709196	0,998514	0,863787
Bayes Ağları (HC)	0,996078	0,707821	0,999257	0,862460
Bayes Ağları (MMHC)	0,996071	0,708797	0,999280	0,856717
Bayes Ağları (Naive Bayes)	0,995686	0,707127	0,998898	0,857820
Karar Ağacı (C.4.5)	0,996129	0,710962	0,999298	0,852044
Karar Ağacı (Bagging)	0,996033	0,702599	0,999274	0,850936
Karar Ağacı (Random Forest)	0,996132	0,712277	0,999274	0,855775
Yapay Sinir Ağları	0,995998	0,703853	0,999222	0,858588
Destek Vektör Makineleri	0,995941	0,697921	0,999258	0,850275

Yöntemlerin performanslarına bakıldığında tüm yöntemlerin yüksek performans gösterdiği ve hiçbirinin diğerine çok üstün gelmediği söylenebilir. Viane ve arkadaşlarının çalışmasında da [32], Naive Bayes sınıflandırıcısı, YSA, Karar Ağaçları, Lojistik Regresyon yöntemleri uygulanmış, bu çalışmadaki sonuca benzer olarak ele alınan yöntemlerinin hiçbirinin performansının diğerlerine göre tutarlı ve önemli ölçüde üstün olmadığı sonucuna ulaşılmıştır.

Otomobil sigortası sahtekârlık tespitinde uygulanan yöntemlerden herhangi birinin kullanılabileceği yorumlanmakla birlikte sadece çalışma süresi göz önünde bulundurularak destek vektör makineleri tercih sebebi olmayabilir. Çalışmada kullanılan yöntemlerin çalışma süresine ilişkin bilgiler Çizelge 6.25’de görülebilir.

Çizelge 6.25. Çalışmada uygulanan yöntemlerin çalışma süreleri⁶

Kullanılan Yöntem	Yöntemin Çalışma Süresi
Lojistik Regresyon	30 dakika
Bayes Ağları (HC)	5 dakika
Bayes Ağları (MMHC)	3 dakika
Bayes Ağları (Naive Bayes)	1 saniye
Karar Ağacı (C.4.5)	1 dakika
Karar Ağacı (Bagging)	23 dakika
Karar Ağacı (Random Forest)	18 dakika
Yapay Sinir Ağları	20 dakika
Destek Vektör Makineleri	3,5 saat

⁶ Çalışmada uygulanan modeller R programında ve Intel(R) Core (TM) i5-4300 CPU işlemci ve 8 GB RAM özelliklerine sahip bir bilgisayarda çalıştırılmış ve yaklaşık olarak çalışma süreleri verilmiştir. Söz konusu fiziksel koşulların ve kaynakların farklılık göstermesi halinde bu süreler de değişecektir.

7. SONUÇLAR VE ÖNERİLER

Sigorta sahtekârlığı, haksız kazanç elde etmek amacıyla, sigortalı tarafından sigortacının bilerek ve kasıtlı olarak aldatılmasıdır. Birçok şekilde görülebildiği gibi en sık rastlanan ve ekonomiyeye zarar veren türü tazminat taleplerine yönelik olan sahtekârlıklardır. Haksız yere tazminat alan kişilerin sigorta şirketinin mali dengesini olumsuz yönde etkilemesi sebebiyle, yapılan aktüeryal hesaplamalarda sapmalar meydana gelir, hasar için ayrılan karşılıklar yetersiz kalabilir ve bunların sonucunda artan sigorta maliyetlerinin bir kısmı sigorta primlerine yüklenerek suçsuz olan sigortalılar daha fazla prim ödemek zorunda kalabilir. Bu ve bunun gibi ekonomide birçok olumsuz etki yaratan sigorta sahtekârlığının tespit edilmesi sigorta şirketleri için önemli bir hale gelmiştir.

Sunulan tez çalışmasında, otomobil sigortasındaki tazminata yönelik yapılan sahtekârlıkların tespit edilmesine ilişkin bir makine öğrenme yönteminin oluşturulması ve bu alanda şüpheli değişken diyebileceğimiz durumların bulunması amaçlanmıştır. Buradan yola çıkarak literatür taraması yapılmış, otomobil sigortasında sahtekarlık tespitine yönelik kullanılan yöntemler ve çalışmalardan elde edilen önemli değişkenler incelenmiştir. Çalışmanın analiz bölümü için mevcut verilere en uygun yaklaşım olması sebebiyle gözetimli öğrenme tercih edilmiş, literatürde bu alanda sıklıkla kullanılan Lojistik Regresyon, Karar Ağaçları (C4.5., Bagging, Random Forest), YSA, DVM yöntemleri veriye uygulanmıştır. Bunlara ilave olarak, bu alanda literatürde kullanılmamış bir yöntem olarak Bayes Ağlarına uygulanmış; HC, MMHC ve Naive Bayes olmak üzere 3 farklı algoritmayla çalışılmıştır.

Türkiye’de sigortacılık alanındaki kayıt sistemleri ilk olarak otomobil sigortaları (trafik ve kasko) ve zorunlu sigortalar ile geliştirilmeye başlanmış, bunu hayat ve sağlık sigortaları takip etmiştir ve günümüzde diğer hayat dışı sigorta branşları için de çalışmalar sürdürülmektedir. Dolayısıyla bir çalışma yürütmek için en doğru ve sağlıklı veriler otomobil sigortası alanında bulunmaktadır. Bu sebeple tez çalışmasında SBM’den temin edilen tüm sigorta şirketlerinin hasar dosyası bilgilerini içeren kasko sigortası verileri kullanılmıştır.

Bu kapsamda, yukarıda bahsi geçen öğrenme yöntemleri kullanılmadan önce modellerin aşırı öğrenmesini engellemek amacıyla rastgele olarak verinin %80'i eğitim, %20'si test verisi olarak ayrılmıştır. Eğitim verisiyle öğrenerek oluşturulan modeller daha önce görmedikleri test verileri üzerinden tahmin yapmış, her bir yöntem 50 tekrara tabi tutulmuştur. Yalnızca DVM'de, modelin çalışmasının çok uzun sürmesi ve mevcut fiziksel kaynakların modeli çok sayıda çalıştırmada yetersiz kalması sebebiyle 10 deneme yapılabilmektedir. Her tekrarda rastgele şekilde eğitim ve test verileri seçilmiş, bu şekilde örneklem seçiminin modeller üzerindeki etkisinin azaltılması ve çok tekrar yapılarak Büyük Sayılar Kanunu gereği performansların ortalamasının gerçek değere yaklaşması hedeflenmiştir. Son olarak da modellerin performanslarının ortalamaları alınarak doğru sınıflandırma oranı, duyarlılık, belirleyicilik, AUC gibi ölçütler üzerinden karşılaştırılmıştır.

Çıkan sonuçlara bakıldığında tüm yöntemlerin iyi performans gösterdiği ve hiçbirinin diğerine çok üstün gelmediği söylenebilir. Genel olarak yöntemlerin doğru sınıflandırma oranları ve belirleyicilikleri %99'un üzerinde, duyarlılıkları % 69-71 arasında, AUC değerleri %85-86 arasında bulunmuştur. Yöntemlerin sahte ve gerçek hasarları %99 oranında doğru sınıflara ayırdığı, sahte hasarları ise % 69-71 gibi bir oranla yakaladığı söylenebilir. Aralarında çok önemsiz farklar olmasına rağmen duyarlılıklara bakıldığında nispeten C4.5 ve Random Forest Karar Ağacı algoritmalarının diğerlerine göre daha iyi performans gösterdiği, DVM'nin daha başarısız olduğu söylenebilir. Lojistik Regresyon ve HC algoritmaları Bayes Ağlarının ise AUC değerlerine göre yine çok küçük farklarla diğerlerine nispeten daha iyi performans gösterdiği söylenebilir. Yöntemlerin doğru sınıflandırma oranı ve sahte hasarları yakalama oranlarına bakıldığında sigorta şirketleri tarafından kullanılmaya uygun olduğu, tamamen otomatik bir sistem oluşturulmasa da hasar departmanlarındaki inceleme ekiplerini yönlendirici ve destekleyici argümanlar sunabileceği düşünülmektedir.

Tahmin performanslarının yanı sıra Lojistik Regresyon yöntemiyle otomobil sigortalarındaki sahte hasar dosyalarında şüpheleri artıran önemli değişkenlerin tespit edilmesi de amaçlanmıştır. Bu kapsamda,

- Kaza sonrasında kaza tespit tutanağının olmaması,
- Poliçede yazan sürücü ismiyle kazaya karışan sürücünün isminin farklı olması,
- Hasar dosyasında rücu imkânının olmaması,
- Poliçede ilave teminat artışının yapılmaması,
- Sigortalının iki poliçesi arasındaki sigortasız geçen sürenin fazla olması,
- Hasar ihbarının yapıldığı servis ile onarımın yapıldığı servisin farklı olması,
- Sigortalılık süresinin düşük olması,
- Hasar dosyası tutarının yüksek olması,
- Kaza sonrasında onarımı gerçekleştiren servisin yetkili olmaması,
- Kaza ihbar tarihinin poliçe başlangıç tarihine veya bitiş tarihine yakın olması,
- Araç yaşının yüksek olması,
- Kazanın cuma akşamı ve pazartesi sabahı arasında meydana gelmesi,
- Kazada yer alan taraflardan birinin daha önceden SİSBİS şüpheli veri tabanına girmesi

durumları hasar dosyalarındaki sahtekarlık şüphesini artıran değişkenler olarak ortaya çıkmıştır. Literatür taramasında yapılan çalışmalardan da; sigortalı aracın eski ve değerinin düşük olması, polis olay yeri raporunun olmaması, poliçe teminatının artırılmaması, kazanın gece veya hafta sonu gerçekleşmesi, poliçe düzenleme tarihi ile poliçe yürürlük tarihi arasında kaza yapılması gibi değişkenler önemli olarak belirlenmiş olup bu da çalışmamızın sonucunu destekler niteliktedir.

Sigorta poliçesi düzenlendikten kısa bir süre sonra ya da poliçe süresinin bitimine az kala yapılan hasar ihbarları sigorta şirketleri için her zaman “kırmızı bayraklı” dosyalar olmuştur. Bu hasarlar masum sigortalılar tarafından yapılan tamamen tesadüf eseri kazalardan kaynaklanabileceği gibi, kaza yaptıktan sonra poliçe yaptıran ve ardından çok geçmeden hasar ihbarında bulunan ya da eski hasarlarını poliçesi bitmeden karşılatmaya çalışan kötü niyetli sigortalılardan da kaynaklanabilmektedir. Hasar ihbarının yapıldığı servis ile onarımın yapıldığı servisin farklı olması ve kaza sonrasında onarımı gerçekleştiren servisin yetkili olmaması durumlarının şüpheli olması ise sigortalıların tamirhaneye işbirliği sonucu yaptığı sahtekârlıklara işaret etmektedir. Sigortalının iki poliçesi arasındaki sigortasız geçen sürenin fazla olması ve sigortalılık süresinin düşük olması, kişinin düzenli bir sigorta alışkanlığının

olmadığı, geçmiş bir hasarını sonradan poliçe yaptırarak karşılatmaya ya da poliçe yaptırırken bir sahtekârlık planladığını gösterebilir. Bu örnekler her bir durum için çoğaltılabilir ancak sahtekârlığın varlığı nihai olarak hasar konusunda uzmanlaşmış kişilerin dosya üzerinden yapacağı araştırma ve eksperlerin teknik bilgisiyle anlaşılacaktır. Bu sebeple sadece veri analizi üzerinden sigortalılara karşı bir suçlama yöneltilemeyeceği ancak veri analizlerinin hasar dosyası üzerinden yapılan araştırmalara büyük bir katkı sağlayacağı söylenebilir.

Bunlara ilave olarak, Kişisel Verilerin Korunması Kanunu'na uyum kapsamında çalışmadan kişisel verilere ilişkin çıkarılan değişkenler modellerin performansını olumsuz yönde etkilemiştir. Söz konusu verilere erişim ve kullanım hakkı olan kurumlar tarafından bu modellerin uygulanması sonucunda daha yüksek performansların elde edilmesi mümkündür. Model performanslarını etkilediği düşünülen bir diğer husus ise “yanlış etiketleme” durumudur. Model tarafından sahtekârlık olarak tahmin edilen ancak veride sahte olmayan (gerçek) hasar şeklinde etiketlenmiş dosyalar modellerin performansını düşürmektedir. Ancak veri setinde gerçek hasar olarak etiketlenmiş bir dosyanın, sigorta şirketinin veya SBM'nin tespit edemeyerek yanlış etiketlediği ve gerçekte sahte bir hasar dosyası olma ihtimali vardır. Bu durumda aslında çalışmadaki modelin doğru tahmin ettiği ancak gerçek hayatta tespit edilemeyen hasar dosyaları modelin performansını düşürmektedir. Dolayısıyla çalışmadan elde edilen model performanslarının hesaplandığından daha yüksek olduğu düşünülmektedir.

Sonuç olarak, bu tez çalışması ile hem sigorta literatüründe yer bulamayan bir yöntem kullanıldığı hem de Türkiye'de uygulaması yapılmamış bir alan ve veriyle çalışıldığından literatüre yeni bir katkı yapıldığı değerlendirilmektedir.

Gelecekte bu konunun geliştirilmesi amacıyla yapılabilecek çalışmalar şu şekilde sıralanabilir: Öncelikle birden fazla yöntem denenmiş hepsinin performansları ayrı ayrı karşılaştırılmıştır. Bunun yerine tüm modellerdeki sınıflandırıcıların performanslarını birleştirecek model istifleme (stacking) veya ortalama (averaging) temelli toplu bir öğrenme yöntemi geliştirilebilir.

Kasko sigortası üzerine yapılan bu çalışmanın devamında trafik sigortası verileriyle de çalışılarak hem önemli değişkenler hem de modellerin tahmin performansları bulunabilir. Çoğu

zaman ikisi insanlar tarafından aynı gibi düşünülse de trafik sigortası bir sorumluluk sigortası, kasko ise mal sigortası olması sebebiyle çalışma mantıkları çok farklı olup özellikle de önemli değişkenler açısından farklı sonuçlara ulaşılabilir.

Çalışmada kullanılan “R” programı büyük veri analizinde son derece başarılı olsa da kullanım kolaylığı açısından herkese hitap etmemektedir. Dolayısıyla bu öğrenme yöntemlerinin sigorta şirketlerince kullanılması halinde daha kullanıcı dostu programların geliştirilmesi düşünülebilir.

8. KAYNAKLAR

- [1] G. E. Rejda, M. J. McNamara, Principles of Risk Management and Insurance, 12. Baskı Pearson, **2014**.
- [2] Milli Reasürans T.A.Ş., Açıklamalı Sigorta ve Reasürans Terimleri Sözlüğü, İstanbul, **2004**.
- [3] <https://www.insurancefraud.org/statistics.htm> (Erişim tarihi: **10 Mayıs 2019**).
- [4] <http://www.sigortacigazetesi.com.tr/karlilikin-onundeki-engel-sigorta-suistimalleri/> (Erişim tarihi: **23 Ekim 2018**).
- [5] <https://www.atlas-mag.net/en/article/insurance-fraud-detection-and-cost-to-industry>. (Erişim tarihi: **10 Mayıs 2019**).
- [6] A. Abdallah, M. A. Maarof, A. Zainal, Fraud Detection System: A Survey, Journal of Network and Computer Applications, Vol. 68, 90–113, **2016**.
- [7] R. J. Bolton, D.J. Hand, Statistical Fraud Detection: A Review, Statistical Science, Vol. 17, No. 3, 235–255, **2002**.
- [8] E. J. Vaughan, T. M. Vaughan, Fundamentals of Risk and Insurance, 11. Baskı, Kendallville, Wiley, **2013**.
- [9] M. Çipil, Risk Yönetimi ve Sigorta, Ankara, Nobel Yayıncılık, **2008**.
- [10] Ş. G. Uralcan, Temel Sigorta Bilgileri ve Sigorta Sektörünün Yapısal Analizi, 3. Baskı, Hiperlink Yayınları, **2011**.
- [11] <https://www.tsb.org.tr/Default.aspx?pageID=1077> (Erişim tarihi: **02.05.2019**)
- [12] Association of Certified Fraud Examiners, Insurance Fraud Handbook, **2018**.
- [13] International Association of Insurance Supervisors, Guidance Paper on Preventing, Detecting and Remedying Fraud in Insurance, No.12, **2006**.
- [14] S. Viaene, G. Dedene, Insurance Fraud: Issues and Challenges, Geneva Papers on Risk and Insurance, Issues and Practice, Vol. 29, No.2, 313–333, **2004**.
- [15] R. Bhowmik, Detecting Auto Insurance Fraud by Data Mining Techniques, Journal of Emerging Trends in Computing and Information Sciences, Vol.2, No.4, 156–162, **2011**.
- [16] B. T. Mengi, Araç Sigortası Hileleri ve Bu Hilelere Yönelik Önlemler, Finans. Araştırmalar ve Çalışmalar Dergisi, Vol.4, No.8, 71–86, **2014**.
- [17] Association of Certified Fraud Examiners, Fraud Examiners Manuel, **2013**.

- [18] <https://siseb.sbm.org.tr/tr/istatistikler> (Erişim tarihi: **10 Mayıs 2019**).
- [19] <https://www.sbm.org.tr/tr/Sayfalar/Anasayfa.aspx>. (Erişim tarihi: **08 Eylül 2018**).
- [20] <https://siseb.sbm.org.tr/tr/sisbis>. (Erişim tarihi: **23 Ekim 2018**).
- [21] R. A. Derrig, Insurance Fraud, Vol.69, No.3, 271–287, **2002**.
- [22] P. Salsas-Forn, S. Tennyson, Claims Auditing in Automobile Insurance. **2001**.
- [23] J. Li, K.-Y. Huang, J. Jin and J. Shi, A Survey on Statistical Methods for Health Care Fraud Detection, Health Care Manag. Sci., Vol.11, No.3, 275–287, **2008**.
- [24] M. Artís, M. Ayuso, M. Guillén, Modelling Different Types of Automobile Insurance Fraud Behaviour in the Spanish Market, Insurance: Mathematics and Economics, Vol.24, No.1–2, 67–81, **1999**.
- [25] M. Artís, M. Ayuso, M. Guillén, Detection of Automobile Insurance Fraud with Discrete Choice Models and Misclassified Claims, Vol.69, No.3, 325–340, **2002**.
- [26] E. B. Belhadji, G. Dionne, Development of an Expert System for the Automatic Detection of Automobile Insurance Fraud, Vol.6, **1997**.
- [27] H. I. Weisberg, R. A. Derrig, Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims, Vol. 35, 75–101, **1998**.
- [28] C.H. Wen, M.J. Wang, L. W. Lan, Discrete Choice Modeling for Bundled Automobile Insurance Policies, Journal of the Eastern Asia Society for Transportation Studies, Vol.6, 1914–1928, **2005**.
- [29] L. Bermúdez, J. M. Pérez, M. Ayuso, E. Gómez, and F. J. Vázquez, A Bayesian Dichotomous Model with Asymmetric Link For Fraud in Insurance, Insurance: Mathematics and Economics, Vol.42, No.2, 779–786, **2008**.
- [30] J. M. Pérez, J. Muguerza, O. Arbelaitz, I. Gurrutxaga, and J. I. Martín, Consolidated Tree Classifier Learning in A Car Insurance Fraud Detection Domain with Class, **2005**.
- [31] S. Viaene, G. Dedene, R. A. Derrig, Auto Claim Fraud Detection Using Bayesian Learning Neural Networks, Expert Systems with Applications, Vol.29, No.3, 653–666, **2005**.
- [32] S. Viaene, R. A. Derrig, B. Baesens, G. Dedene, A Comparison of State-of-The-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection, The Journal of Risk and Insurance, Vol.69, No.3, 373–421, **2002**.
- [33] W. Xu, S. Wang, D. Zhang, B. Yang, Random Rough Subspace Based Neural Network Ensemble for Insurance Fraud Detection, 2011 Fourth International Joint Conference on Computational Sciences and Optimization, 1276–1280, **2011**.

- [34] S. Viaene, R. A. Derrig, G. Dedene, A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis, *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No.5, 612–620, **2004**.
- [35] H. Tao, L. Zhixin, S. Xiaodong, Insurance Fraud Identification Research Based on Fuzzy Support Vector Machine with Dual Membership, *2012 International Conference on Information Management, Innovation Management and Industrial Engineering*, 457–460, **2012**.
- [36] R. A. Derrig, K. M. Ostaszewski, Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification, *Journal of Risk and Insurance*, Vol.62, No.3, 447–482, **1995**.
- [37] J. Pathak, N. Vidyarthi, S. L. Summers, A Fuzzy-Based Algorithm for Auditors to Detect Elements of Fraud in Settled Insurance Claims, *Managerial Auditing Journal*, Vol.20, No. 6, 632–644, **2005**.
- [38] P. L. Brockett, X. Xia, R. A. Derrig, Using Kohonen’s Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, *Journal of Risk and Insurance*, Vol.65, No.2, **1998**.
- [39] P. Brockett, R. Derrig, L. Golden, A. Levine, M. Alpert, Fraud Classification Using Principal Component Analysis of RIDITs, *Journal of Risk and Insurance*, Vol.69, No.3, 341–371, **2002**.
- [40] L. Šubelj, Š. Furlan, M. Bajec, An Expert System for Detecting Automobile Insurance Fraud Using Social Network Analysis, *Expert Systems with Applications*, Vol.38, No.1, 1039–1052, **2011**.
- [41] X. Zhu, A. B. Goldberg, Introduction to Semi-Supervised Learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol.3, No.1, **2009**.
- [42] H. He, J. Wang, W. Graco, S. Hawkins, Application of Neural Networks to Detection of Medical Fraud, *Expert Systems with Applications*, Vol.13, No.4, 329–336, **1997**.
- [43] G. J. Williams, Z. Huang, Mining the Knowledge Mine: The Hot Spots Methodology for Mining Large Real World Databases, *Lecture Notes in Artificial Intelligence*, Vol. 1342, 340–348, **1997**.
- [44] R. M. Musal, Two Models to Investigate Medicare Fraud Within Unsupervised Databases, *Expert Systems with Applications*, Vol.37, 8628–8633, **2010**.
- [45] S. Weisberg, *Applied Linear Regression*, 3. Baskı, New Jersey, John Wiley & Sons, **2005**.
- [46] B. Hüdaverdi, Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama, *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Vol.2, 185–208, **2004**.

- [47] D. W. Hosmer, S. Lemeshow, Applied Logistic Regression, John Wiley & Sons Inc, **1989**.
- [48] H. Tatlıdil, Uygulamalı Çok Değişkenli İstatistiksel Analiz, Ankara, Ziraat Matbaacılık, **2002**.
- [49] B. Yet, Bayesian Networks for Evidence Based Clinical Decision Support, Doktora Tezi, Queen Marry University of London, **2013**.
- [50] M. Karabıyık, B. Yet, Bayes Ağları ile Futbol Analitiği: Futba Modeli, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, Vol.25, No.1, 121–131, **2019**.
- [51] Y. Wang, Analysis of the Max-Min Hill-Climbing Algorithm, Advances in Intelligent Systems Research, Vol.161, 509–511, **2018**.
- [52] H. Akpınar, Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi İşletme Fakültesi Dergisi, Vol.1, 1–22, **2000**.
- [53] Ü. H. Atasever, C. Özkan, Arazi Örtüsünün Belirlenmesinde Torbalama-Karar Ağaçları Yönteminin Kullanımı, **2012**.
- [54] L. Breiman, Random Forests, Statistics Department University of California Berkeley, **2001**.
- [55] L. V. Fausett, Fundamentals of Neural Networks. New Jersey, Printice-Hall Inc, **1994**.
- [56] M. A. Göral, Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, **2007**.
- [57] Y. E. Kuyucu, Lojistik Regresyon Analizi (LRA), Yapay Sinir Ağları (YSA) ve Sınıflandırma ve Regresyon Ağaçları (C&RT) Yöntemlerinin Karşılaştırılması ve Tıp Alanında Bir Uygulama, Yüksek Lisans Tezi, Gaziosmanpaşa Üniversitesi, **2012**.
- [58] V. N. Vapnik, The Nature of Statistical Learning Theory, 2. Baskı, New York, Springer, **2000**.
- [59] T. Kavzoğlu, İ. Çölkesen, Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi, Harita Dergisi, No.144, 73–82, **2010**.
- [60] F. Başer, A. Apaydın, “Sınıflandırma Amaçlı Destek Vektör Makinelerinin Lojistik Regresyon ile Karşılaştırılması,” Anadolu Üniversitesi Bilim ve Teknoloji Dergisi B-Teorik Bilimler, Vol.3, No.2, pp. 53–65, **2015**.



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 30/05/2019

Tez Başlığı / Konusu: TÜRKİYE'DE OTOMOBİL SİGORTASI SAHTEKÂRLIKLARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TESPİT EDİLMESİ

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 87 sayfalık kısmına ilişkin, 24/05/2019. tarihinde tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 7 'dir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç
- 3- 5 kelimededen daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

30.05.2019

Tarih ve İmza

Adı Soyadı: Ezgi GÜNBATAR
Öğrenci No: N13123273
Anabilim Dalı: Endüstri Mühendisliği
Programı: Endüstri Mühendisliği
Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Dr. Öğr. Üyesi Barbaros YET

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Ezgi GÜNBATAR
Doğum yeri : Ankara
Doğum tarihi : 26.08.1990
Medeni hali : Evli
Elektronik posta adresi : ezgiskn@gmail.com

Eğitim

Lisans : Hacettepe Üniversitesi İstatistik Bölümü
Anadolu Üniversitesi İşletme Bölümü
Yüksek Lisans : Hacettepe Üniversitesi Endüstri Mühendisliği

Yabancı Dil ve Düzeyi

İngilizce, İyi derecede

İş Deneyimi

Aralık 2013- Mart 2014 Türkiye Halk Bankası A.Ş. (Management Trainee)
Mart 2014- Kasım 2017 T.C. Başbakanlık Hazine Müsteşarlığı
Sigortacılık Genel Müdürlüğü (Hazine Uzman Yardımcısı)
Kasım 2017- ... Hazine ve Maliye Bakanlığı
Sigortacılık Genel Müdürlüğü (Hazine ve Maliye Uzmanı)

Deneyim Alanları

Hayat Dışı Sigortalar, Tarım Sigortaları, Veri Analizi

Tezden Üretilmiş Projeler ve Bütçesi : -

Tezden Üretilmiş Yayınlar : -

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar: -