



**BÖLGESEL VE SEKTÖREL DAĞILIMLARI ÇERÇEVESİNDE İŞ
KAZALARININ MAKİNE ÖĞRENMESİ KULLANILARAK ANALİZ
EDİLMESİ**

Orçun NEFES

**YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ 2025

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Orçun NEFES

22/07/2025

BÖLGESEL VE SEKTÖREL DAĞILIMLARI ÇERÇEVESİNDE İŞ KAZALARININ MAKİNE ÖĞRENMESİ KULLANILARAK ANALİZ EDİLMESİ

(Yüksek Lisans Tezi)

Orçun NEFES

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2025

ÖZET

Günümüzde iş kazaları, iş sağlığı ve güvenliği (İSG) alanında hem insan hayatı hem de ekonomik kayıplar açısından önemli bir sorun teşkil etmektedir. Özellikle endüstriyel faaliyetlerin yoğun olduğu sektörlerde meydana gelen kazalar, ciddi iş gücü kayıplarına ve maddi zararlara neden olmaktadır. Bu çalışmanın temel amacı, Türkiye’deki iş kazalarının bölgesel ve sektörel dağılımlarını analiz ederek, kazaların nedenlerini daha iyi anlamak ve bu doğrultuda etkili İSG politikalarının geliştirilmesine katkı sağlamaktır. Bu kapsamda, ilk olarak literatürde yer alan iş kazası çalışmaları sistematik bir şekilde incelenmiş; geleneksel istatistiksel yöntemlerin kullanımının yanı sıra makine öğrenmesi temelli yaklaşımlar da değerlendirilmiştir. Daha sonra 2014–2023 yılları arasında Türkiye’de meydana gelen iş kazaları, Sosyal Güvenlik Kurumu verileri temel alınarak analiz edilmiş; madencilik, metal ve inşaat sektörleri özelinde yürütülen analizler geleneksel ve modern veri analizi teknikleri ile karşılaştırmalı olarak incelenmiş, çoklu doğrusal regresyon, LASSO regresyonu, rastgele orman, k-en yakın komşu ve destek vektör makineleri algoritmaları bu analizlerde kullanılmıştır. Analiz sürecinde sektör, kaza zamanı, kazazedinin yaşı, eğitim düzeyi, iş sağlığı ve güvenliği ile mesleki eğitim durumu, kaza olan işletmedeki çalışan sayısı ve kaza şiddeti gibi değişkenler dikkate alınmıştır. Bu veriler üzerinden 2026–2028 yılları için iş kazalarının meydana gelme olasılığı ve sonuçlarının şiddeti tahmin edilmiştir. Hesaplamaların doğrulukları R^2 , MAPE ve MSE performans metrikleri temelinde derinlemesine analiz edilmiştir. Elde edilen bulgular, iş kazalarının öngörülmesinde makine öğrenmesi yöntemlerinin geleneksel yöntemlere kıyasla daha yüksek doğruluk ve öngörü gücü sunduğunu ortaya koymuştur. Çalışmanın sonunda, elde edilen analiz sonuçlarına dayalı olarak çeşitli vadelerde İSG politikalarına yönelik önleyici stratejiler geliştirilmiştir. Bu tez çalışması, iş kazalarının veri temelli analizi ve tahmininde makine öğrenmesinin uygulanabilirliğini ortaya koyarak alan uygulayıcıları için çeşitli vadelerde önleyici stratejiler geliştirmeyi bilimsel olarak ise geleneksel istatistiksel analiz metotlarını modern yöntemler ile kıyaslanması üzerine hesaplama doğruluğunu arttıran bir altyapı oluşturmayı hedeflemektedir.

Bilim Kodu : 90603

Anahtar Kelimeler : Ergonomi, iş kazaları, iş sağlığı ve güvenliği, makine öğrenmesi, veri analizi

Sayfa Adedi : 135

Danışman : Doç. Dr. Burcu YILMAZ KAYA

ANALYSIS OF WORK ACCIDENTS USING MACHINE LEARNING WITHIN THE FRAMEWORK OF REGIONAL AND SECTORAL DISTRIBUTIONS

(M. Sc. Thesis)

Orçun NEFES

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2025

ABSTRACT

Today, occupational accidents constitute a significant problem in terms of both human life and economic losses in the field of occupational health and safety (OHS). Accidents occurring especially in sectors with intensive industrial activities cause serious workforce losses and material damages. The main purpose of this study is to analyze the regional and sectoral distribution of occupational accidents in Turkey, to better understand the causes of accidents and to contribute to the development of effective OHS policies in this direction. In this context, firstly, the studies on occupational accidents in the literature were systematically examined; in addition to the use of traditional statistical methods, machine learning-based approaches were also evaluated. Then, occupational accidents that occurred in Türkiye between 2014 and 2023 were analyzed based on Social Security Institution data; analyses conducted specifically for mining, metal and construction sectors were comparatively examined with traditional and modern data analysis techniques, and multiple linear regression, LASSO Regression, random forest, k-nearest neighbor and support vector machines algorithms were used in these analyses. In the analysis process, variables such as sector, time of accident, age of the victim, education level, occupational health and safety and vocational training status, number of employees in the enterprise where the accident occurred and severity of the accident were taken into consideration. Based on these data, the probability of occupational accidents and the severity of their consequences were estimated for the years 2026–2028. The accuracy of the calculations was analyzed in depth based on R², MAPE and MSE performance metrics. The findings have shown that machine learning methods provide higher accuracy and predictive power compared to traditional methods in predicting occupational accidents. At the end of the study, based on the analysis results obtained, preventive strategies for OHS policies were developed in various terms. This thesis aims to develop preventive strategies for field practitioners in various terms by demonstrating the applicability of machine learning in data-based analysis and prediction of occupational accidents, and to create an infrastructure that increases calculation accuracy by scientifically comparing traditional statistical analysis methods with modern methods.

Science Code : 90603

Key Words : Ergonomics, occupational accidents, occupational health and safety, machine learning, data analysis

Page Number : 135

Supervisor : Assoc. Prof. Dr. Burcu YILMAZ KAYA

TEŞEKKÜR

Bu tez çalışmasının her aşamasında değerli bilgi ve rehberliğiyle bana yol gösteren, akademik gelişimime katkı sağlayan ve yardımını hiçbir zaman esirgemeyen değerli danışmanım Doç. Dr. Burcu YILMAZ KAYA'ya içten teşekkürlerimi sunarım. Bu süreçte desteğiyle yanımda olan sevgili eşim Sultan'a ve varlığıyla hayatımıza renk katan oğlum Can'a teşekkür ederim. Bu süreçte katkı sağlayan ve desteğini esirgemeyen herkese içten teşekkürlerimi sunarım.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ	xii
SİMGELER VE KISALTMALAR.....	xv
1. GİRİŞ.....	1
2. GENEL KAVRAMLAR	5
2.1. İş Kazası	5
2.1.1. İş kazası nedenleri	6
2.1.2. İş kazası verileri.....	8
2.2. Yapay Zeka ve Bilgi Mühendisliği	9
2.2.1. Büyük veri	10
2.2.2. Makine öğrenmesi kavramı.....	11
2.3. Makine Öğrenmesi Algoritmaları	13
2.3.1. Denetimli öğrenme	13
2.3.2. Denetimsiz öğrenme	15
2.3.3. Yarı-denetimli öğrenme.....	16
2.3.4. Pekiştirmeli öğrenme	17
2.3.5. Derin öğrenme	17
2.4. Tahmin	18
2.4.1. Geleneksel tahmin yöntemleri	19

	Sayfa
2.4.2. Makine öğrenmesi temelli tahmin yöntemleri.....	22
3. LİTERATÜR ARAŞTIRMASI	33
3.1. İş Kazası Analizine İlişkin Genel Literatür Taraması.....	34
3.2. İş Kazaları Analizinde Makine Öğrenmesi Uygulamalarına İlişkin Literatür Taraması.....	42
4. MATERYAL VE METODOLOJİ.....	53
4.1. İş Kazaları Analizinde Kullanılacak Geleneksel Tahmin Yöntemleri.....	53
4.1.1. Regresyon analizi	53
4.2. İş Kazaları Analizinde Kullanılacak Makine Öğrenmesi Temelli Tahmin Yöntemleri.....	55
4.2.1. Rastgele orman	56
4.2.2. K-en yakın komşu	58
4.2.3. Destek vektör makineleri.....	60
4.3. Tahmin Doğruluğunun Analizi	64
5. UYGULAMA.....	67
5.1. Problem Tanımı.....	67
5.2. Veriler ve Özellikleri.....	75
5.3. Geleneksel Tahmin Yöntemleri Hesaplamaları	80
5.3.1. Çoklu doğrusal regresyon analizi hesaplamaları.....	80
5.3.2. LASSO regresyon yöntemi hesaplamaları	84
5.4. Makine Öğrenmesi Algoritmaları ve Hesaplamaları	90
5.4.1. Rastgele orman yöntemi hesaplamaları.....	91
5.4.2. KNN yöntemi hesaplamaları.....	96
5.4.3. DVM yöntemi hesaplamaları.....	101
5.5. Bulgular ve Tartışma.....	106

	Sayfa
5.5.1. Korelasyon analizi bulguları	106
5.5.2. Geleneksel yöntemler ve makine öğrenmesi algoritmaları bulgularının karşılaştırmalı yorumlanması.....	109
5.5.3. İSG stratejileri.....	113
6. SONUÇ VE ÖNERİLER	121
KAYNAKLAR	125
ÖZGEÇMİŞ	135



ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. İş kazalarının nedenleri	7
Çizelge 3.1. İş kazaları çalışmalarına ait literatür taraması	40
Çizelge 3.2. İş kazaları analizinde makine öğrenmesi çalışmalarına ait literatür taraması	49
Çizelge 5.1. Analiz edilecek alt sektörler	70
Çizelge 5.2. Bağımlı ve bağımsız değişkenler	75
Çizelge 5.3. Değişkenlerin dağılımları	76
Çizelge 5.4. Özet korelasyon değerleri	79
Çizelge 5.5. Çoklu doğrusal regresyon tahmin sonuçları	81
Çizelge 5.6. Madencilik sektörü çoklu doğrusal regresyon modeli sonuçları	82
Çizelge 5.7. Metal sektörü çoklu doğrusal regresyon modeli sonuçları	82
Çizelge 5.8. İnşaat sektörü çoklu doğrusal regresyon modeli sonuçları	83
Çizelge 5.9. Modellerin alpha değerine göre karşılaştırması	85
Çizelge 5.10. LASSO regresyon modeli tahmin sonuçları	88
Çizelge 5.11. Madencilik sektörü LASSO regresyon modeli sonuçları	88
Çizelge 5.12. Metal sektörü LASSO regresyon modeli sonuçları	89
Çizelge 5.13. İnşaat sektörü LASSO regresyon modeli sonuçları	89
Çizelge 5.14. Modellerin ağaç sayısına göre karşılaştırması	92
Çizelge 5.15. Rastgele orman modeli tahmin sonuçları	93
Çizelge 5.16. Madencilik sektörü rastgele orman modeli sonuçları	94
Çizelge 5.17. Metal sektörü rastgele orman modeli sonuçları	95
Çizelge 5.18. İnşaat sektörü rastgele orman modeli sonuçları	95
Çizelge 5.19. Modellerin komşu sayısına göre karşılaştırması	98

Çizelge	Sayfa
Çizelge 5.20. KNN modeli tahmin sonuçları	99
Çizelge 5.21. Madencilik sektörü KNN modeli sonuçları	99
Çizelge 5.22. Metal sektörü KNN modeli sonuçları	100
Çizelge 5.23. İnşaat sektörü KNN modeli sonuçları	100
Çizelge 5.24. Modellerin çekirdek fonksiyonuna göre karşılaştırması	103
Çizelge 5.25. DVM modeli tahmin sonuçları	103
Çizelge 5.26. Madencilik sektörü DVM modeli sonuçları	104
Çizelge 5.27. Metal sektörü DVM modeli sonuçları	105
Çizelge 5.28. İnşaat sektörü DVM modeli sonuçları	105
Çizelge 5.29. Kullanılan modellerin madencilik sektörü performans kıyaslaması	109
Çizelge 5.30. Kullanılan modellerin metal sektörü performans değerlerinin kıyaslaması	110
Çizelge 5.31. Kullanılan modellerin inşaat sektörü performans değerlerinin kıyaslaması	111

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. Literatür taraması gösterimi.....	34
Şekil 4.1. Rastgele orman örnek yapısı.....	57
Şekil 4.2. Doğrusal olarak ayrılabilen DVM	62
Şekil 5.1. 2014-2023 yılları arasındaki iş kazası sayısı	67
Şekil 5.2. 2014-2023 yılları arasında sektörlere göre yaşanan iş kazası sayısı.....	71
Şekil 5.3. 2014-2023 yılları arasında sektörlere göre yaşanan ölüm sayısı	72
Şekil 5.4. Önerilen çözüm yaklaşımı akış diyagramı	74

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

F	Tahmin değeri
K	Çekirdek fonksiyonu
n	Örnek sayısı
p	Oluş yüzdesi
t	Dönem
tan	Tanjant
w	Ağırlık katsayısı
x	Girdi değeri
y	Çıktı değeri
α	Düzeltilme sabiti

Kısaltmalar

Açıklamalar

AB	Avrupa Birliği
AHP	Analitik hiyerarşi prosesi
AUROC	ROC eğrisi altındaki alan
DVM	Destek vektör makineleri
ESAW	Avrupa iş kazaları istatistikleri
GA	Genetik algoritma
ILO	Uluslararası Çalışma Örgütü
İSG	İş sağlığı ve güvenliği
IoT	Nesnelerin interneti
KNN	K-en yakın komşu
MAD	Ortalama mutlak sapma
MAPE	Ortalama mutlak yüzde hatası
MARS	Çok değişkenli uyarlamalı regresyon eğrileri
ME	Mutlak hata

Kısaltmalar**Açıklamalar****MSE**

Ortalama kare hatası

PSO

Parçacık sürü optimizasyonu

R²

Belirleme katsayısı

RFID

Radyo frekansı ile tanımlama

SGK

Sosyal Güvenlik Kurumu

SPSS

Sosyal bilimler için istatistik paketi

TMMOB

Türk Mühendis ve Mimar Odaları Birliği

XGBOOST

Ekstrem gradyan arttırma ağaçları

YSA

Yapay sinir ağı

1. GİRİŞ

Günümüzde iş kazaları, iş sağlığı ve güvenliği (İSG) alanında önemli bir sorun olarak öne çıkmaktadır. Özellikle endüstriyel faaliyetlerin yoğun olduğu bölgelerde ve sektörlerde meydana gelen iş kazaları hem insan hayatı açısından hem de ekonomik açıdan ciddi kayıplara neden olmaktadır. Uluslararası Çalışma Örgütü'nün (International Labour Organisation-ILO) yaptığı araştırmalara göre, her yıl dünya genelinde yaklaşık üç milyon kişi iş kazaları nedeniyle hayatını kaybetmektedir. Bunun yanı sıra, ölümle sonuçlanmayan 374 milyon iş kazası ve meslek hastalığı meydana gelmekte; bu durumların önemli bir kısmı geçici iş göremezliğe yol açmaktadır (ILO, 2023). Türkiye'de yaşanan iş kazalarının %46,4'ü maden, inşaat ve metal sektörlerinde meydana gelmekte ve buna bağlı ölümlerin %41,1'i bu sektörlerde yaşanan iş kazalarından kaynaklanmaktadır (Delibalta, 2020). İş kazaları, işverenler, çalışanlar ve ülke ekonomisi üzerinde ciddi etkiler yaratan önemli bir sorundur. Bu yüzden işyerlerinde İSG kurallarına uyulmasına azami ölçüde dikkat edilmelidir. Aksi takdirde iş kazaları, kalıcı iş gücü kaybına ve can kayıplarına yol açarken, aynı zamanda ekonomik zararlar ve iş günü kayıpları da yaşanmaktadır (Dündar, A.Bilim ve N.Bilim, 2024). Yeterli ve uygun İSG önlemlerinin alınmaması nedeniyle oluşan iş kazalarının, küresel ekonomiye her yıl Gayri Safi Hasıla'nın %3,94'ü oranında zarar verdiği tahmin edilmektedir (ILO, 2023). Bu da Türkiye'de 2023 yılında iş kazaları nedeniyle yaşanan kayıpların 1 trilyon 51 milyar lirayı bulmakta olduğunu göstermektedir. Türkiye gibi gelişmekte olan ülkelerde, iş kazalarının önlenmesi ve yönetilmesi, sürdürülebilir bir ekonomik kalkınmanın temel unsurlarından biri haline gelmiştir.

Bu bağlamda, iş kazalarının bölgesel ve sektörel dağılımlarının detaylı bir şekilde analiz edilmesi önem arz etmektedir. Bu analiz, iş kazalarının hangi bölgelerde ve hangi sektörlerde daha sık meydana geldiğini belirleyerek kaynakların etkin bir şekilde dağıtılmasına ve önleyici politikaların oluşturulmasına katkı sağlayabilecektir. İş kazaları ve analizi üzerine literatürde birçok farklı çalışma bulunmaktadır. Literatürde yapılan çalışmalarda anket, istatistiksel analiz, kavramsal çalışmalar, veri madenciliği, çok kriterli karar verme gibi çeşitli yöntemler kullanılmış, çalışmalar sektör bazında yapıldığı gibi sektör bazında olmadan genel kapsamlı çalışmalar da bulunmaktadır. Çalışmalarda kaza nedenlerinin analizi, sınıflandırma, tahmin gibi araştırmalar yapılmıştır. Ancak, geleneksel istatistiksel yöntemlerle bu analizleri gerçekleştirmek geleceğe dair tahmin ve sınıflandırmalarda hata

seviyeleri nedeniyle yeterli güvenlik ve geçerlilikte sonuç vermezken uygulamaları da oldukça emek ve zaman gerektirmektedir.

Makine öğrenmesi, büyük veri kümelerinde gizli desenleri ve ilişkileri tespit etme yeteneğiyle bilinmektedir ve bu özelliğiyle iş kazalarının analizinde değerli bir araç olarak kullanılmaya elverişlidir. Makine öğrenmesi, bir bilgisayarın, geçmiş olaylardan öğrenilen deneyim yoluyla bir olayın özelliklerini öğrenerek, daha önce hiç karşılaşılmamış benzer olaylar için kararlar verebilmesi veya çözümler üretebilmesidir (Yılmaz, 2022). Makine öğrenmesi teknolojisinin hedefi, veri yapılarını anlamak ve bu verileri çeşitli uygulamalarda insanlar tarafından kullanılabilir tahminlere dönüştürerek belirli modellere uygun hale getirmektir (Karakaya Özkan, 2023).

Bu tez çalışmasının temel amacı, Türkiye’deki iş kazalarının daha iyi anlaşılmasına katkı sağlamak ve bu doğrultuda daha etkili İSG politikalarının geliştirilmesine zemin hazırlamaktır. Bu amaçla, mevcut literatürde yer alan iş kazası analizlerine ilişkin çalışmalar sistematik literatür taraması yöntemiyle incelenmiştir. İnceleme kapsamında, hem geleneksel istatistiksel yöntemlerin kullanıldığı çalışmalar hem de makine öğrenmesi tabanlı yaklaşımların benimsendiği analizler değerlendirilmiştir. Literatür araştırması bulguları ışığında iş kazası analitiği alanında sıkça başvurulmuş geleneksel yöntemlerin yanı sıra, son yıllarda artan veri hacmi ve veri setlerinin karmaşıklığı dikkate alınarak, daha esnek, öngörücü ve güvenilir analizler sunma potansiyeline sahip olan makine öğrenmesi yöntemlerinin kullanımı ön plana çıkarılmıştır. Bu çerçevede tez amacına yönelik olarak geliştirilecek stratejilerin girdi verisinin hesaplanmasında iş kazalarının analizinde geleneksel istatistiksel yöntemler ve makine öğrenmesi temelli yöntemler ayrı ayrı kullanılmış ve tahmin performansları değerlendirilmiştir. 2014-2023 yılları arasındaki SGK verileri kullanılarak 2026-2028 yılları için hesaplanan bulgular çerçevesinde Türkiye’de insidans ve mortalite oranları en yüksek olan madencilik, metal ve inşaat sektörleri için kısa, orta ve uzun vadeli önleyici stratejileri geliştirilmiştir.

Bu tez çalışması altı bölümden oluşmaktadır. Çalışmanın ikinci bölümünde tez konusu temelinde ele alınan probleme özgü temel kavramlar sunulmuş; iş kazası, yapay zeka, makine öğrenmesi ve tahmin kavramları detaylarıyla araştırılarak incelenmiştir. Daha sonra mevcut ulusal ve uluslararası literatür iş kazaları analizinde kullanılan mevcut yöntemler ve makine öğrenmesinin iş kazaları analizindeki potansiyelinin araştırılması temelinde

incelenmiştir. Gerçekleştirilen literatür araştırması çerçevesinde bilimsel literatürde yer alan çalışmalarda kullanılan veri kaynakları, veri tipi ve miktarları ile analiz süreçleri detaylı olarak araştırılmıştır. Tez çalışmasının bir sonraki bölümünde madencilik, metal ve inşaat sektörleri ve Sosyal Güvenlik Kurumu (SGK) verileri temelinde çoklu doğrusal regresyon ve LASSO regresyon geleneksel tahmin metotları ve rastgele orman, k-en yakın komşu (K-Nearest Neighbor-KNN) ve destek vektör makineleri (DVM) makine öğrenmesi algoritmaları kullanılarak Türkiye’de 2014-2023 yılları arasında gerçekleşen iş kazaları analiz edilmiştir. Bu gerçek hayat uygulamasında ele alınan veri tipleri sektör, kaza zamanı, kazazedenin yaşı, eğitim durumu, iş sağlığı ve güvenliği eğitim durumu ile mesleki eğitim durumu, kaza olan işletmedeki çalışan sayısı ve kaza şiddeti olarak belirlenmiş, öncelikle değişkenler arasındaki ilişki varlığı ve türleri korelasyon analizleri ile incelenmiştir. Bu değişkenler çerçevesinde 2026-2028 yıllarında kazaların meydana gelme olasılığı ile sonuçlarının şiddeti tahmin edilmeye çalışılmıştır. Tez çalışmasının beşinci bölümünde ise ele alınan gerçek hayat problemi üzerindeki araştırmalar çerçevesinde elde edilen sonuçlar ve iş kazası sayıları ve ölüm sayıları boyutlarında ortaya koyulan bulgular detayları ile irdelenerek tartışılmış, çıkarımlar tablolar ile desteklenerek sunulmuş, ayrıca alan uygulayıcıları ve bilimsel araştırmacılar için veri temelli önleyici stratejiler, sektöre özel politikalar ve çalışan profil analizine dayalı stratejiler temelinde çeşitli stratejiler kısa, orta ve uzun vadeli planlama ufuklarına göre geliştirilmiş ve sunulmuştur. Tez çalışmasının son bölümü olan sonuç bölümünde ise çalışmanın ana bulguları özetlenmiş, çalışma sınırlılıkları belirtilerek açıklanmış, geliştirilen uygulama stratejileri derlenmiş ve gelecek araştırmalar için önerilerde bulunulmuştur.

2. GENEL KAVRAMLAR

Bu bölümde, tez çalışmasının kavramsal altyapısını oluşturmak üzere iş kazası, yapay zeka, büyük veri, makine öğrenmesi ve tahmin kavramları ele alınmaktadır.

2.1. İş Kazası

İş kazası kavramı dünyada birçok farklı uluslararası kuruluş tarafından tanımlanmakla birlikte ülkemizde de çeşitli kanunlar ile tanımlanmış ve düzenlenmiştir.

6331 sayılı İş Sağlığı ve Güvenliği Kanunu'na (2012) göre iş kazası “İşyerinde veya işin yürütümü nedeniyle meydana gelen, ölüme sebebiyet veren veya vücut bütünlüğünü ruhen ya da bedenen engelli hâle getiren olay” olarak tanımlanmaktadır.

5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu'nda (2006) ise “İş kazası;

- Sigortalının işyerinde bulunduğu sırada,
- İşveren tarafından yürütülmekte olan iş nedeniyle, sigortalı kendi adına ve hesabına bağımsız çalışıyorsa yürütmekte olduğu iş nedeniyle,
- Bir işverene bağlı olarak çalışan sigortalının, görevli olarak işyeri dışında başka bir yere gönderilmesi nedeniyle asıl işini yapmaksızın geçen zamanlarda,
- Hizmet akdi ile çalışan emziren kadın sigortalının, iş mevzuatı gereğince çocuğuna süt vermek için ayrılan zamanlarda,
- Sigortalıların, işverence sağlanan bir taşıtla işin yapıldığı yere gidiş gelişi sırasında meydana gelen ve sigortalıyı hemen veya sonradan bedenen ya da ruhen engelli hâle getiren olaydır.” olarak tanımlanmıştır.

ILO ise iş kazasını, “ölümcül veya ölümcül olmayan yaralanmayla sonuçlanan, iş sırasında ortaya çıkan olay” şeklinde tanımlamaktadır. Meslek hastalığını ise “bir iş faaliyetinden kaynaklanan tehlikelere maruz kalmanın bir sonucu olarak ortaya çıkan her türlü hastalığı kapsar” şeklinde tanımlamaktadır (ILO, 2015).

İş kazası, İş Kazalarına İlişkin Avrupa İstatistikleri (European Statistics on Accidents at Work-ESAW) metodolojisinde ise iş sırasında fiziksel veya zihinsel zarara yol açan münferit

bir olay olarak tanımlanmaktadır. 'İş sırasında' ifadesi, 'mesleki bir faaliyette bulunurken veya işte geçirilen süre boyunca' anlamına gelmektedir (Eurostat, 2012).

Türkiye'de, iş kazaları bildirimleri 5510 sayılı kanuna göre yapılmaktadır. Daha önce, 2012 yılı ve öncesinde, iş kazaları istatistik kayıtlarında yalnızca ödemeleri tamamlanan iş kazalarının sayıları dikkate alınmaktayken 2013 yılından itibaren, Avrupa Birliği (AB) standartlarına uygun olarak, iş kazası geçiren sigortalı çalışanların sayılarına ilişkin veriler, iş kazası bildirim formuyla elektronik ortamda kaydedilmeye başlanmıştır (Karakaya Özkan, 2023).

2.1.1. İş kazası nedenleri

İş kazalarının birçok nedeni bulunmaktadır ve bunlar bölgeden bölgeye, ülkeden ülkeye ve sektörel bazda değişiklik gösteren olgulardır.

İş kazaları sadece ağır yaralanmalara ve ölümlere neden olmakla kalmamakta, aynı zamanda işin düzenli ilerleyişini de etkilemektedir (Koç ve Gürgün, 2022).

İş kazalarının nedenlerini belirlemeye yönelik birçok çalışma yapılmıştır. Bu çalışmalara göre, iş kazalarının birçok farklı nedene bağlı olarak meydana geldiği ortaya konmuştur. İş kazalarının temel sebepleri genellikle ülkelerin sanayileşme tarzı, işletme yapıları, kaza istatistikleri ve araştırmalarının eksikliği, denetim hizmetlerinin yetersizliği, çalışanların nitelikleri ve iş güvenliği bilincinin oluşturulmamış olmasıdır. İş kazalarının nedenleri genelde benzer özellikler sergilese de farklı ülkelerde kendine özgü bazı özellikler de gösterebilmektedir (Camkurt, 2007).

Çeşitli çalışmalar, iş kazalarının nedenlerini belirlemek amacıyla incelendiğinde, farklı oranlarda ve sonuçlarda kaza nedenlerine ulaşıldığı görülmektedir. Kazaların meydana gelmesinde genellikle tek bir sebebin değil, birçok faktörün bir araya gelerek oluşturduğu olaylar zinciri sonucunda gerçekleştiği anlaşılmaktadır. Araştırmalar, iş kazalarının büyük çoğunluğunun insan kaynaklı (%80), fiziksel ve mekanik çevre koşullarından (%18) ve beklenmedik olaylardan (%2) kaynaklandığını göstermektedir. Başka bir deyişle, iş kazalarının yaklaşık %98'inin, uygun kontrol önlemleri alınarak önlenebileceği ortaya koyulmuştur (Karakaya Özkan, 2023).

İş kazalarının meydana gelmesinde, üretim teknolojisi, üretim araçları, çevresel koşulların yanı sıra sosyolojik, psikolojik ve fizyolojik birçok faktör rol oynamaktadır. Ancak, iş kazalarının nedenleri genellikle iş yerlerindeki güvensiz koşullar ve güvensiz davranışlar olmak üzere iki temel faktöre indirgenebilmektedir (TMMOB, 2015). Benzer şekilde Cascio (1995)'ya göre iş kazalarının iki temel sebebi vardır ve bu sebepleri güvensiz davranışlar ve güvensiz çalışma koşulları olarak tanımlamıştır.

Güvensiz koşullar; üretim sürecinde kullanılan teknoloji ve araçların niteliği, işyeri düzensizlikleri, yetersiz bakım ve kontroller, denetim ve yönetim hataları, sağlıksız çevre koşulları gibi birçok faktörden kaynaklanabilmektedir. Özellikle makine ve tezgahların koruyucu sistemlerinin bulunmaması, amacı dışında veya kapasitesinin üzerinde kullanımı, yerleşim düzenindeki yanlışlıklar ile ürünlerin depolanması, taşınması ve istiflenmesindeki hatalar bu duruma örnek olarak verilebilir. Ayrıca sıcaklık, nem, gürültü ve hava kirliliği gibi fiziksel ve kimyasal çevre koşulları da iş kazası riskini artıran unsurlar arasındadır.

Güvensiz davranışlar ise iş yükünün insanın bedensel ve zihinsel gücünü dikkate almadan düzenlenmesi ve çalışma hızının belirlenmesi, işçinin yapmakla yükümlü olduğu iş için gerekli eğitimi almamış olması, işin kendisine uygun görünmemesi veya kişilik özelliklerinin dikkate alınmaması, çalışma ortamı gibi nedenlerle oluşabilmektedir (TMMOB, 2015). İş kazalarının nedenlerinin yukarıda bahsedilen iki temel faktör altında gösterimi Çizelge 2.1'de verilmektedir.

Çizelge 2.1. İş kazalarının nedenleri (TMMOB, 2015)

Güvensiz Koşullar	Güvensiz Davranışlar
Koruyucusuz Makine ve Tezgahlar	İşi Bilinçsiz Yapmak
Güvensiz Çalışma Yöntemi	Dalgınlık ve Dikkatsizlik
Güvensiz ve Sağlıksız Çevre Koşulları	Makine Koruyucularını Çıkarmak
Topraklanmamış Elektrik Makinaları	Tehlikeli Hızda Çalışmak
İşe Uygun Olmayan El Aletleri	Görevi Dışında İş Yapmak
Kontrol ve Testleri Yapılmamış Basıncı Kaplar,	İş Disiplinine Uymamak
Kaldırma Makinaları	İşe Uygun Makine ve Alet Kullanmamak
Tehlikeli Yükseklikte İstifleme	Yetkisiz ve İzinsiz Olarak Tehlikeli Bölgede
Kapatılmamış Boşluklar	Bulunmak
İşyeri Düzensizliği	Kişisel Koruyucuları Kullanmamak
	Ehliyetsiz ve Tehlikeli Hızda Araç Kullanmak vb.

2.1.2. İş kazası verileri

İş kazası nedenlerine benzer şekilde iş kazası verileri de bölgeye, ülkeye ve sektöre göre değişiklik göstermektedir.

İş kazalarına ilişkin veriler, dünya genelinde çeşitli kurumlar tarafından toplanarak analiz edilmekte ve raporlanmaktadır. Türkiye'de bu süreç SGK tarafından yürütülmekte ve düzenli olarak kamuoyuna sunulmaktadır.

Avrupa'da 1990 yılında, iş kazalarına ilişkin verilerin tutarlılığını sağlamak amacıyla ESAW projesi başlatılmıştır. Avrupa İstatistik Bürosu ve İstihdam ve Sosyal İşler Genel Müdürlüğü tarafından 2001 yılında yayınlanan "Avrupa İş Kazaları İstatistikleri – Metodoloji" adlı rapor, 1990'dan bu yana yapılan metodoloji çalışmalarını içermekte ve hala kullanılmaktadır. Bu metodolojiye göre, iş kazası sonucu üç günden fazla iş göremezlik yaşanan tüm kazalar istatistiklere dahil edilir. ESAW metodolojisi, AB ülkeleri arasında veri karşılaştırması yapılabilmesi için iş kazalarıyla ilgili yaralanma türü, olaya sebep olan sapma gibi birçok alt kategoride bilgilerin standart hale getirilmesini sağlamıştır (Karakaya Özkan, 2023).

ESAW metodolojisine göre iş kazalarına ilişkin sağlanacak uyumlaştırılmış ve ortak veri seti aşağıdaki konuları kapsamaktadır:

- Yaralanan kişinin özellikleri,
- Yaralanmanın özellikleri (Ciddiyet (kayıp gün sayısı) dahil olmak üzere),
- İşletmenin özellikleri (Ekonomik aktivite dahil olmak üzere),
- İşyerinin özellikleri
- Kazanın özellikleri (Kazanın nedenlerini ve koşullarını karakterize eden olay sırası da dahil olmak üzere).

İş kazalarının toplam maliyeti, iş kazalarını önlemek için yapılan harcamalar, iş kazası meydana geldikten sonra yapılan her türlü harcamalar ve bu kazaların yol açtığı kayıplardan oluşmaktadır (Bayram, Ünğan ve Ardıç, 2017).

İş kazalarının analizi, profesyonel popülasyonlardaki kaza modellerini belirlemek için

önemli bir yol sağlar. İş kazaları, birçok faktörün etkilediği karmaşık olaylardır. Bunların önlenmesi yalnızca geçmiş olayların analiz edilmesi ve istatistiksel sonuçların doğru yorumlanmasıyla mümkündür (Bilim ve Çelik, 2023).

2.2. Yapay Zeka ve Bilgi Mühendisliği

Yapay zekâ, insanın zihinsel süreçlerini, özellikle düşünme, anlama ve karar verme yetilerini modelleyerek, bu işlevleri taklit edebilecek bilgisayar sistemlerinin geliştirilmesini amaçlayan bir araştırma alanıdır. Daha geniş bir tanımla, yapay zekâ; bilgi edinme, algılama, görsel tanıma, muhakeme ve karar verme gibi insan zekâsına özgü bilişsel yetkinliklerin bilgisayar sistemleri aracılığıyla simüle edilmesi sürecidir (Kara, 2000).

Yapay zeka, bilgisayar sistemlerine insan benzeri bilişsel yetenekler kazandırmayı amaçlayan, öğrenme, problem çözme, dil işleme ve algı gibi temel süreçlerin matematiksel modeller ve algoritmalar aracılığıyla modellenmesini hedefleyen disiplinler arası bir araştırma alanıdır. Yapay zeka, endüstriyel ve pratik uygulamalarda karmaşık problemlerin çözümünü sağlamak, süreçleri otomatikleştirmek ve verimliliği artırmak amacıyla, makine öğrenmesi, derin öğrenme gibi yöntemlerin entegre edilerek geliştirildiği mühendislik odaklı bir teknolojidir (Russel ve Norvig, 2010).

Yılmaz (2022)'e göre yapay zeka, insan zekasının özelliklerini bir bilgisayara veya makineye aktararak insan gibi düşünebilme, anlam çıkarabilme, genelleme yapabilme ve geçmiş deneyimlerden öğrenme gibi yetileri kazandırmayı hedefler. Yapay zekanın amacı, insan zekasına ihtiyaç duyulan işleri yapabilecek makineler oluşturmaktır.

Yapay zeka kavramının tarihsel gelişimine bakıldığında; yapay zeka düşüncesi çok eski zamanlara dayanmaktadır. Örneğin, Antik Yunan döneminde robotlarla ilgili fikirlerin varlığı gözlemlenmektedir. 1884 yılında Charles Babbage'ın yaptığı, zeki davranışlar sergileyen makineler üzerindeki deneyler yapay zeka için önemli bir kilometre taşıdır. Daha sonra çalışmalar devam ettirilmiş ve 1950 yılında Shannon bilgisayarların satranç oynayabileceğini ortaya koymuştur (Yılmaz, 2022).

1950'li yılların başında Alan Turing'in "Makineler düşünebilir mi?" sorusuyla makine öğrenmesi kavramı ortaya çıkmış, Bilgisayar Bilimci ve Bilişsel Bilimci "John McCarthy"

tarafından “Yapay Zeka” terimi 1956 yılında düzenlenen ilk kez “Dartmouth Konferansı”nda kullanılmıştır (Nilsson, 2009).

1970-1975 yılları arasında yapay zeka çalışmaları hız kazanmıştır. Günümüz yapay zekasının temelleri, hastalık teşhisi gibi konular üzerinde çalışılarak geliştirilen yapay zeka sistemlerinin başarıları sayesinde atılmıştır. Yapay zeka, 1980'lerde pratik uygulamalarda ve büyük projelerde kullanılmaya başlanmıştır. Bu dönemden itibaren, gerçek hayattaki sorunlara çözüm getirmek için yapay zekanın geliştirilmesi ve adapte edilmesi süreci yaşanmıştır. Bu durum, yapay zekanın kullanım alanlarının önemli ölçüde genişlemesini sağlamıştır (Yılmaz, 2022).

Başlıca uzman sistemler, genetik algoritmalar, bulanık mantık, yapay sinir ağları, makine öğrenmesi gibi teknikler, yapay zeka teknolojileri olarak bilinmektedir. Genel anlamda yapay zeka; insan zekasının ve fizyolojik ve nörolojik yapısının ve doğal olayların modellenerek makinelerle aktarılmasıdır. Kısaca yapay zeka, insan gibi düşünen, insan gibi davranan ve canlıların zeki olarak kabul edilen davranışlarına sahip bilgisayar sistemleridir. Makine öğrenmesi ise yapay zekanın son halkası olarak değerlendirilmektedir (Atalay ve Çelik, 2017).

2.2.1. Büyük veri

Büyük veri, mevcut bilgi sistemlerinin işleyemeyeceği kadar yoğun ve karmaşık veri setleri olarak tanımlanır. Başka bir deyişle, bilinen veri tabanı yönetim sistemleri ve yazılım bileşenlerinin, veri toplama, saklama ve analiz yeteneklerini aşan büyüklükteki veri kümeleri için kullanılan bir terimdir (Razbonyalı, 2016).

Büyük veri, geleneksel veri tabanı teknikleriyle işlenemeyen, farklı büyüklükte ve çeşitteki karmaşık veri kümesini tanımlayan bir terimdir ve çeşitli dijital kaynaklardan oluşmaktadır (Aktan, 2018).

Büyük veri, araştırmacıların sorularına yanıt bulmalarını kolaylaştırırken, bireysel davranışları ve topluluk eğilimlerini tahmin etmelerine yardımcı olur. Ekonomik ve ticari faaliyetlerden kamu yönetimine, ulusal güvenlikten bilimsel araştırmalara kadar geniş bir yelpazede büyük veri kullanılmaktadır. Büyük verinin temel uygulama alanları arasında

bankacılık, iletişim, medya ve eğlence, sağlık hizmetleri, eğitim, üretim, kamu hizmetleri, sigortacılık, perakendecilik ve ticaret, ulaşım, enerji sektörü ve kendi kendine ölçüm verisinin analizi bulunmaktadır (Aktan, 2018).

Büyük verinin özellikleri arasında veri miktarı (volume), veri akış hızı (velocity) ve veri çeşitliliği (variety) bulunmaktadır. Bu özellikler, ilk olarak 2001'de Loug Laney tarafından "3V" olarak tanımlanmıştır. Farklı yazarlar, doğruluk (verification) ve değer (value) gibi özellikleri ekleyerek büyük veriyi 5V kavramı üzerinden açıklamışlardır. Daha sonra bu özelliklere değişkenlik ve karmaşıklık boyutlarını eklemiştir (Özcan, 2021).

Büyük veri kavramı, yalnızca veri setinin büyük boyutunu değil, aynı zamanda yüksek hızda veri oluşumunu ve veri çeşitliliğini de vurgular. Büyük verinin kullanımıyla birlikte, bu verilerin depolanması, güvenliği ve gizliliği, işlenmesi ve analizi, veriye dayalı karar verme gibi konularda çeşitli zorluklar ortaya çıkar. Bu süreçler, veri madenciliği, bilgisayar bilimi, makine öğrenmesi, veritabanı yönetimi, matematiksel algoritmalar ve istatistik gibi alanların bir araya gelmesini gerektirir (Atalay ve Çelik, 2017).

Makine öğrenmesi ve sayısal simülasyon alanındaki son gelişmeler, değişkenler arasındaki ilişkileri tanımlamak ve farklı türden olayları tahmin etmek için büyük veri kümeleriyle çalışmayı mümkün kılmaktadır (Fuentes, Serrano, Lasheras, Valverde ve Sanchez, 2022).

2.2.2. Makine öğrenmesi kavramı

Makine öğrenmesi uygulamaları genellikle büyük veri olarak adlandırılan büyük miktarda veriyi analiz etmenin ve bu verilerden algoritmalar oluşturmanın bir yoludur. Makine öğrenmesi, yapay zekanın bir alt dalı olan bir yapay zeka uygulamasıdır ve çoğu zaman yapay zeka teknolojisinin temelini oluşturur (Karakaya Özkan, 2023).

1980'lerden önce, makine öğrenmesi genellikle sinir ağı yöntemleriyle temsil edilirken, bazıları tarafından yapay zekanın sınırlarında yer alıyordu. Ancak, son zamanlarda, makine öğrenmesi modern yapay zekada çok daha merkezi bir rol oynamıştır. 1980'lerde başlayan diğer gelişmeler, makine öğrenmesini yapay zekanın öne çıkan alanlarından biri haline getirmiştir (Nilsson, 2009).

Makine öğrenmesi, bir bilgisayarın, deneyim yoluyla bir olayın özelliklerini öğrenerek, daha önce hiç karşılaşmamış benzer olaylar için kararlar verebilmesi veya çözümler üretebilmesidir. Yapay zekanın bilgisayar bilimlerinde öğrenme ve tanıma için kullanılması amacıyla geliştirilmiştir. Öğrenme yeteneğine sahip ve sonrasında tahmin yapabilen algoritmaların geliştirilmesini hedefler. Bu algoritmalar, geleneksel optimizasyon algoritmalarında olduğu gibi talimatları adım adım takip etmez; bunun yerine eğitim setinden elde edilen veri desenleri ile tahmin yapmayı veya karar vermeyi amaçlarlar (Yılmaz, 2022).

Makine öğrenmesi teknolojisinin hedefi, veri yapılarını anlamak ve bu verileri çeşitli uygulamalarda insanlar tarafından kullanılabilir tahminlere dönüştürerek belirli modellere uygun hale getirmektir (Karakaya Özkan, 2023).

Makine öğrenmesi algoritmalarının geleneksel istatistiki analiz yöntemleri yerine uygulamalarda kullanılmasının birçok avantajı bulunmaktadır. Makine öğrenmesinin en büyük avantajlarından biri, eğilimleri ve modelleri tanımlamayı oldukça basit hale getirmesidir. Verileri izleme, madencilik yapma ve analiz etme yeteneğine sahip bir makine, eğilimleri ve modelleri belirlemeyi büyük ölçüde kolaylaştırır. Makine öğrenmesi uygulamalarıyla bilgisayarlar, yüksek miktarda veriyi tarayarak dakikalar içinde eğilimleri ve modelleri belirleyebilirler. Makine öğrenmesinin bir başka avantajı ise hesaplamalarda insan müdahalesine gerek olmamasıdır. Bu da cihazların otomatik davranışlar kullanarak kendilerini sürekli olarak geliştirebileceği anlamına gelir. İnsan müdahalesi olmadan karmaşık görevleri başarıyla tamamlayabilen ve kendilerini sürekli olarak iyileştirebilen cihazlara sahip olmak, işlem hatalarının enazlanması ve sonuç güvenilirliği açısından büyük bir potansiyel taşır (Karakaya Özkan, 2023).

Makine Öğrenmesi algoritmaları, kendilerine sağlanan verilerden öğrenme konusunda dikkate değer bir yeteneğe sahiptir. Daha fazla veri aldıkça, modellerini sürekli olarak iyileştirirler ve bu da sürekli yeniden programlamaya gerek kalmadan tahminlerde ve karar almada yüksek doğruluk ve tutarlılığa ulaşmasını sağlar. Ayrıca eksik veri veya parçalı/tahrif edilmiş veri durumunda da yüksek doğrulukla çalışabilme özelliğine sahiptir (Das, 2024).

Son yıllarda, makine öğrenmesi teknikleri iş kazalarının analizinde ve önlenmesinde kullanılmaya başlanmıştır. Bu teknikler, potansiyel riskleri ve kazaların nedenlerini

belirlemeye yardımcı olabilecektir.

İş kazaları için tahmin modelleri istatistiksel öğrenmeye dayalı veya makine öğrenmesi temelli olabilmektedir. Mevcut büyük miktardaki veri nedeniyle makine öğrenmesi, mühendislik, tıp bilimi, finans gibi çeşitli alanlarda kullanılan gelecekteki olayları tahmin etmede geleneksel istatistiksel muadillerinin yerini almakta ve çok faydalı sonuçlar vermektedir (Sarkar, Vinay, Raj, Maiti ve Mitra, 2019).

Makine öğrenmesinin potansiyel faydaları yalnızca büyük miktarda veriyi işleme kapasitesinden değil, aynı zamanda geniş çaplı problemlerin üstesinden gelebilme, karmaşıklıklardan bağımsız olarak veri üretim yapısını esnek bir şekilde yeniden oluşturabilme ve kuralların çıkarılması yoluyla öngörü ve yorumlama sağlayabilme yeteneklerinden de kaynaklanmaktadır. Bu güçlü yetenekleri sayesinde, makine öğrenmesi teknikleri iş kazası analizleri de dahil olmak üzere çeşitli alanlarda başarıyla uygulanmaktadır (Sarkar ve diğerleri, 2019b).

2.3. Makine Öğrenmesi Algoritmaları

Makine öğrenmesi algoritmaları, veri öğrenme süreçlerine dayanarak tahminlerde bulunur; bu süreçte denetimli öğrenme ve denetimsiz öğrenme olmak üzere iki ana kategoriye ayrılmaktadır (Donoho, 2017). Makine öğrenmesi denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme gibi algoritmaları içermektedir. Makine öğrenmesi algoritmaları, kullanılan veriler ve algoritmaların işleyişine göre denetimli öğrenme, denetimsiz öğrenme, yarı denetimli öğrenme, pekiştirmeli öğrenme ve derin öğrenme başlıkları altında gruplanabilmektedir (Akdoğan Erkan, 2024).

Çok sayıda çalışma, iş kazalarını ve bunların potansiyel sonuçlarını tahmin etmek için denetimli ve denetimsiz makine öğrenmesi yöntemlerini içeren çeşitli yapay zeka tabanlı algoritmalar geliştirmiştir (Mammadov, Kazar Koç ve Tokdemir, 2023).

2.3.1. Denetimli öğrenme

Denetimli öğrenme, belirli veri kümelerinin matematiksel modellerini oluşturan ve makine için girdileri ve istenen çıktıları içeren algoritmaların bir dizisidir. Öğrenme süreci, girdiler

ile istenen çıktı veya etiketler arasındaki ilişkiyi belirleyerek bir model oluşturmayı içerir. Denetimli öğrenme yöntemindeki veriler, eğitim verileri olarak adlandırılır ve genellikle bir veya daha fazla girdi ile tipik olarak bir istenen çıktı içeren eğitim örneklerinden oluşur. Bu çıktıya genellikle "gözetim sinyali" denir (Karakaya Özkan, 2023).

Denetimli makine öğrenmesi, makine öğrenmesi ve yapay zekânın bir alt dalıdır. Algoritmaları eğitmek için etiketlenmiş veri kümelerinin kullanılmasıyla tanımlanır, bu da verileri sınıflandırmak veya sonuçları doğru bir şekilde tahmin etmek için kullanılır. Denetimli öğrenme, modellere istenen çıktıyı sağlamak için bir eğitim seti kullanır. Bu eğitim veri seti, modelin zaman içinde öğrenmesini sağlayan girdileri ve doğru çıktıları içerir. Algoritma, hata yeterince minimize edilene kadar ayar yaparak, doğruluğu kayıp fonksiyonu aracılığıyla ölçer (Petekci, 2021).

Denetimli öğrenme, işletmelere derin veri iç görüşleri ve gelişmiş otomasyon gibi avantajlar sunsa da sürdürülebilir denetimli öğrenme modelleri oluştururken bazı zorluklarla karşılaşılabilir. Denetimli öğrenme modellerini doğru bir şekilde yapılandırmak belirli bir uzmanlık düzeyi gerektirebilir ve eğitim süreci oldukça zaman alabilir. Ayrıca, veri kümeleri insan hatalarına daha yatkın olabilir, bu da algoritmaların yanlış öğrenmesine yol açabilir. Denetimli öğrenme modelleri, denetimsiz öğrenme modellerinin aksine, verileri kendi başlarına kümeleyemez veya sınıflandıramaz (Petekci, 2021).

Denetimli öğrenme, veri madenciliği sürecinde genellikle iki ana kategoriye ayrılır: sınıflandırma ve regresyon. Sınıflandırma, bir algoritmanın test verilerini belirli kategorilere doğru bir şekilde atamasını sağlar. Veri kümesindeki belirli öğeleri tanıyarak, bu öğelerin nasıl etiketlenmesi veya sınıflandırılması gerektiği konusunda çıkarımlar yapmaya çalışır (Petekci, 2021). Sınıflandırma algoritmaları arasında öne çıkanlar; lineer diskriminant analizi, karar ağaçları, yapay sinir ağları, DVM, lojistik regresyon, KNN, genetik algoritmalar, bellek temelli çıkarım yöntemleri ve naive Bayes algoritmasıdır (Karr ve Freeman, 1999).

Regresyon ise bağımlı ve bağımsız değişkenler arasındaki ilişkiyi anlamak için kullanılır (Petekci, 2021).

Regresyon algoritmaları, makine tarafından belirlenen bir aralıkta sayısal çıktılar üreten

algoritmalarıdır. Makine öğrenmesinde kullanılan regresyon algoritmaları arasında doğrusal ve doğrusal olmayan regresyon ile lojistik regresyon yer alır (Karakaya Özkan, 2023).

2.3.2. Denetimsiz öğrenme

Denetimsiz öğrenme, yalnızca girdilerin sağlandığı bir dizi algoritmadır. Bu algoritmalar, buldukları verilere dayanarak kümeleme yapmak ve ideal çıktılar oluşturmakla sorumludur. Denetimsiz öğrenme algoritmaları genellikle belirli hedeflere sahiptir, ancak kontrol edilmezler. Denetimsiz öğrenmenin hedefi genellikle büyük veri setlerinde gizli bir model bulmaktır. Denetimsiz öğrenme yöntemlerine dayanan algoritmalar, insanlar tarafından etiketlenmemiş, sınıflandırılmamış veya kategorize edilmemiş verilere dayanmaktadır (Karakaya Özkan, 2023).

Denetimsiz öğrenme, etiketlenmemiş veri kümelerini analiz etmek ve kümelemek için makine öğrenmesi algoritmalarını kullanır. Bu algoritmalar, insan müdahalesi olmadan gizli kalıpları veya veri gruplamalarını keşfeder. Benzerlik ve farklılıkları ortaya çıkarma yeteneği, keşifsel veri analizi, çapraz satış stratejileri, müşteri segmentasyonu ve görüntü tanıma gibi alanlarda ideal bir çözüm sağlar. Denetimsiz makine öğrenmesi algoritmaları üç ana kısımda kullanılır: bunlar kümeleme, ilişkilendirme ve boyut azaltma olarak sıralanabilir (Petekci, 2021).

Kümeleme; etiketlenmemiş verileri benzerlik veya farklılık ölçütlerine göre gruplandıran bir veri madenciliği yöntemidir. Kümeleme algoritmaları, ham ve sınıflandırılmamış veri noktalarını, yapılar veya kalıplarla temsil edilen gruplar halinde işlemek için kullanılır. Bu algoritmalar, özel, örtüşen, hiyerarşik ve olasılıksal gibi farklı türlere ayrılabilir (Petekci, 2021).

Denetimsiz makine öğrenmesinin bir alt alanı olan kümeleme tekniklerinde, doğru küme sayısını belirlemek, algoritmaların karşılaştığı önemli bir zorluktur. Gerçek dünya veri setlerinde, özellikle yüksek boyutlu olanlarda, verinin kaç kümeden oluştuğu genellikle bilinmemektedir. Bu belirsizliği ortadan kaldırmak için küme sayısını tahmin edebilen yöntemler geliştirilmiştir (Aydın ve Kayhan, 2021).

Literatürde, küme sayısının önceden belirlenmesine gerek duymayan en yaygın algoritmalar

arasında hiyerarşik kümeleme yöntemleri yer almaktadır. Bu yöntem, kümeleri bir ağaç yapısı içinde hiyerarşik olarak düzenler ve iç içe geçmiş kümeler oluşturur (Ezugwu ve diğerleri, 2021). Klasik kümeleme yöntemlerinden olan K-ortalamlar ve Bulanık C Ortalamalar gibi tekniklerde, küme sayısı kullanıcı tarafından belirlenir (Aydın ve Kayhan, 2021). K-ortalamlar, 50 yıldan fazla bir süredir, belki de en yaygın kullanılan bölme yöntemlerinden biri olmuştur. Diğer taraftan, Bulanık C Ortalamalar kümeleme algoritması, yaygın kullanılan bir yöntem olup, ilk kez Dunn (1974) tarafından önerilmiş ve daha sonra Bezdek (1981) tarafından geliştirilmiştir.

İlişkilendirme; belirli bir veri setindeki değişkenler arasındaki ilişkileri keşfetmek için kural tabanlı bir yöntemdir. Bu teknikler genellikle pazar sepeti analizinde kullanılır ve şirketlere farklı ürünler arasındaki ilişkileri daha iyi anlama imkânı sağlar. Müşterilerin alışveriş alışkanlıklarını anlamak, işletmelerin daha etkili çapraz satış stratejileri ve öneri sistemleri geliştirmelerine yardımcı olabilir. Daha fazla veri genellikle daha doğru sonuçlar sağlar, ancak aynı zamanda makine öğrenmesi algoritmalarının performansını da etkileyebilir (örneğin, aşırı uyum) ve veri kümelerinin görselleştirilmesini zorlaştırabilir (Petekci, 2021).

Boyut azaltma; bir veri setindeki yüksek sayıda özellik veya boyut olduğunda kullanılan bir tekniktir. Veri girişlerinin sayısını azaltarak yönetilebilir bir seviyeye indirirken, veri setinin bütünlüğünü korumaya çalışır (Petekci, 2021).

Yüksek boyutlu veri kümelerinin boyutunu azaltmak için yaygın olarak kullanılan yöntemlerden biri Temel Bileşenler Analizi'dir (Jolliffe, 2002). Bu yöntem, büyük boyutlu veri setlerinde en yüksek varyansı bir arada yakalayarak veri boyutunu indirgemeyi hedefler. Denetimsiz bir boyut küçültme tekniği olan Temel Bileşenler Analizi, meteoroloji, görüntü işleme ve genetik analizler gibi pek çok farklı alanda kullanılmaktadır (Şengöz ve Özdemir, 2016).

2.3.3. Yarı-denetimli öğrenme

Denetimli ve denetimsiz öğrenme yöntemlerinin dışında, yarı denetimli öğrenme adı verilen bir başka algoritma türü de bulunmaktadır. Yarı-denetimli öğrenme, kavramsal olarak denetimli ve denetimsiz öğrenme arasında bir yerde yer alır. Yarı-denetimli öğrenme ile yapılan çalışmalar genellikle sınıflandırma problemleri üzerine yoğunlaşmaktadır. Yarı-

denetimli öğrenme, etiketli verinin kısıtlı olduğu durumlarda etkili bir sınıflandırma yapmak için oldukça yararlı bir yöntemdir. Performansı iyileştirmek amacıyla, sınırlı sayıda etiketli verinin yanı sıra etiketlenmemiş veriler de kullanılarak fayda sağlanmaya çalışılır. Bu algoritmalarda, model hem etiketli hem de etiketsiz verilerle eğitilir ve yeni veri örnekleri için hedef özelliğinin değeri tahmin edilmeye çalışılır (Turan ve Polat, 2024).

Eldeki verinin etiketlenmiş değerleri az sayıda ancak buna karşılık olarak daha fazla sayıda etiketlenmemiş veri mevcutsa denetimli öğrenme de denetimsiz öğrenme de yetersiz kalabilir. Bu durumda en iyi yöntem az sayıdaki etiketlenmiş veriden yola çıkıp bu verileri kullanarak, etiketlenmemiş veriler hakkında bilgi sahibi olmaya çalışmaktır. Bu yöntem yarı denetimli öğrenme denir. Yarı denetimli öğrenmenin denetimli öğrenme ile arasındaki en belirgin fark eldeki etiketlenmiş veri kümesidir. Denetimli öğrenmede etiketlenmiş veri sayısı fazla, tahmin edilmek istenen veri sayısı azken yarı denetimli öğrenmede etiketlenmiş veri sayısı az, tahmin edilmek istenen veri sayısı fazladır (Kızılkaya ve Oğuzlar, 2018).

2.3.4. Pekiştirmeli öğrenme

Pekiştirmeli öğrenme, bir ajanın başlangıçta eğitim verisi olmadan, aldığı geri bildirimlerle öğrenmesini sağlayan bir makine öğrenmesi yöntemidir. Ajan, belirli bir hedefe ulaşmak için içinde bulunduğu ortamla etkileşime girerek çeşitli eylemler gerçekleştirir. Ortamın bu eylemlere verdiği tepkiler, önceden tanımlanmış bir ödül mekanizması ile değerlendirilir. Elde edilen ödüller ile ajan, gerçekleştirdiği eylemin doğruluğunu veya yanlışlığını anlayarak deneyim kazanır ve bu deneyimi daha sonra alacağı kararlarında kullanır. Pekiştirmeli öğrenmenin en büyük zorluklarından biri, keşif ve sömürü arasındaki dengeyi sağlamaktır. Ajan, ödül kazanabilmek için mevcut bilgilerini kullanırken, uzun vadede daha iyi stratejiler geliştirmek için yeni eylemler keşfetmek zorundadır (Akarsu, 2024).

2.3.5. Derin öğrenme

Derin öğrenme, 1950'lerin ortalarından itibaren, yapay zekanın bir alt dalı olan makine öğrenmesi kapsamında gelişmeye başlamış ve son yıllarda birçok alanda yeniliklere imza atmıştır. Sinir ağları, makine öğrenmesinin bir alt disiplini olarak, derin öğrenme algoritmalarının temelini oluşturan yapay zeka tekniklerinden biridir. Derin öğrenme kavramı ortaya çıktığı günden bu yana farklı uygulama alanlarında büyük başarılar elde

etmiştir. Özellikle 2006 yılından sonra, makine öğrenmesinin bir alt kategorisi olarak kabul edilmesiyle birlikte daha fazla dikkat çekmeye başlamıştır. Derin öğrenme süreci, belirli bir görevin yerine getirilmesi için modelin eğitim aşamasını içerir. Bu süreçte, modelin doğru çalışabilmesi için belirli parametrelerin sisteme uygun şekilde tanımlanması gerekmektedir. Derin öğrenme, özellikle öğrenme ve desen tanıma işlemleri için kullanılan, giriş ve çıkış katmanlarını içeren çok katmanlı, hiyerarşik algoritmik yapılarla çalışan doğrusal olmayan veri işleme sistemlerinden oluşur. Bu tür veri odaklı öğrenme yöntemleri, temsili öğrenme olarak adlandırılır. Genel olarak bakıldığında, derin öğrenme görev odaklı bir yaklaşıma sahiptir (Balcıoğlu, 2023).

2.4. Tahmin

Tahmin, geçmişte var olan verilerin kullanılarak, gelecekte ortaya çıkabilecek durumlar ile ilgili yorum ve stratejilerin geliştirilmesi süreci olarak tanımlanabilir. Geçmiş yıllara ilişkin veriler değerlendirilerek, farklı tahmin yöntemleri ile sonraki yıllarda ortaya çıkabilecek verilere ilişkin tahminler yapılabilir. Bu şekilde geleceğin doğru bir şekilde tahmin edilmesi, tahmin edilen alan için uygun stratejilerin belirlenmesini ve doğru planlamaların yapılmasını sağlamaktadır. Doğru tahmin, gelecek dönemlerde gerçekleşecek durumlar hakkında gerekli tedbirlerin alınmasını sağlar (Desticioğlu, 2015).

Tahminin amacı, gelecekte karşılaşılabilecek durumları, geçmişteki veri ve çeşitli teknikler kullanılarak, önlem alınmasını sağlamaktır. Gelecek planlarının düzgün şekilde yapılabilmesi için, iyi bir tahminin yapılması gerekmektedir (Desticioğlu, 2015).

Tahmin, işletme, sanayi, kamu yönetimi, ekonomi, çevre bilimleri, tıp, sosyal bilimler, siyaset ve finans gibi pek çok alanda önemli bir konudur. Tahmin problemleri genellikle kısa, orta ve uzun vadeli olmak üzere üç grupta incelenir. Kısa vadeli tahminler, olayları yalnızca birkaç zaman dilimi içinde (günler, haftalar veya aylar) öngörmeyi amaçlar. Orta vadeli tahminler genellikle bir ila iki yıldan başlayarak daha uzun tahminleri kapsarken, uzun vadeli tahminler bu sürenin daha fazla ilerisini tahmin etmeyi kapsayabilmektedir. Operasyon yönetimi, bütçeleme ve yeni Araştırma-Geliştirme projelerinin belirlenmesi gibi süreçlerde kısa ve orta vadeli tahminler önemli bir rol oynarken, uzun vadeli tahminler daha çok stratejik planlama gibi konularda kullanılır. Tahmin süreçlerinde günlük, haftalık, aylık, üç aylık veya yıllık veriler kullanılabildiği gibi, anlık veriler de analiz edilebilir (Eşidir, Gür,

Yoğunlu ve Çubuk, 2022).

Bu çalışmada tahmin yöntemleri geleneksel tahmin yöntemleri ve makine öğrenmesi tahmin yöntemleri olmak üzere iki kısımda incelenmiştir.

2.4.1. Geleneksel tahmin yöntemleri

Geleneksel tahmin yöntemleri çeşitli şekillerde sınıflandırılabilir. Genel olarak, bu yöntemler nitel (sayısal olmayan) ve nicel (sayısal) tahmin yöntemleri olmak üzere iki gruba ayrılır. Nitel yöntemler, tahmini yapan kişinin kişisel değerlendirmelerine ve uzman görüşlerine dayanırken, nicel yöntemler istatistiksel ve matematiksel analizlere dayalıdır. Karar alıcılar, mevcut verileri inceleyerek hangi tahmin yöntemlerinin uygulanabilir olduğunu belirlemelidir (Desticioğlu, 2015).

Zaman serisi analizi modelleri

Zaman serisi modelleri, tahmin edilmek istenen veriyi yalnızca geçmiş değerlerinin değişimine dayanarak öngören modellerdir. Bu modeller, dışsal faktörleri dikkate almadan çalıştıkları için "kapalı kutu" olarak adlandırılır ve bu özellikleriyle yapay sinir ağlarına benzerlik gösterir (Eşidir ve diğerleri, 2022). Zaman serisi analizi modelleri hareketli ortalamalar, üstel düzeltme, Holt-Winters yöntemi ve otokorelasyon ve mevsimsellik analizi olmak üzere incelenebilmektedir.

Basit hareketli ortalama tahmin yöntemi, öngörü çalışmaları kapsamında en sık kullanılan yöntemlerden biridir. Bu yaklaşımda, yakın geçmişteki veriler daha fazla önem taşır ve genellikle yalnızca bir sonraki dönem için tahmin üretilebilir. Yöntemin uygulanmasında, son m döneme ait talep değerleri toplanarak aritmetik ortalaması alınır. Elde edilen bu ortalama değer, gelecek döneme ilişkin tahmini temsil eder. Basit hareketli ortalama yönteminin matematiksel gösterimi Eş. 2.1'de yer almaktadır.

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-m+1}}{m} \quad (2.1)$$

Bu yöntem, geçmiş dönem talep verilerinin ortalamasına dayanmasına rağmen, zaman

içerisindeki ani değişimleri ya da dalgalanmaları ayrıntılı olarak dikkate alamamaktadır (Desticioğlu, 2015).

Basit hareketli ortalama tahmin yönteminde tüm dönemlere eşit önem verilirken, ağırlıklı hareketli ortalama tahmin yönteminde her döneme farklı ağırlıklar verilir ve son dönem verilerine daha fazla önem verilir. Bu yöntemde, "w" harfiyle gösterilen değerler, her döneme ait belirlenen ağırlıkları ifade eder. Yakın geçmişin daha etkili olduğu bu yaklaşım, yalnızca bir sonraki dönem için satış tahmini yapılmasına olanak tanır. Ağırlıklı hareketli ortalama yöntemine ait matematiksel formül Eş. 2.2'de gösterilmektedir (Desticioğlu, 2015).

$$F_{t+1}^w = \frac{w_m y_t + w_{m-1} y_{t-1} + \dots + w_1 y_{t-m+1}}{w_m + w_{m-1} + \dots + w_1} \quad (2.2)$$

Üstel düzeltme yöntemi, hareketli ortalamaların özel bir türü olarak kabul edilir ve temel farkı, geçmiş verilere eşit değil, giderek azalan ağırlıklar verilmesidir. Bu yöntemde, daha yeni verilere daha yüksek ağırlık tanınırken, eski verilerin etkisi azalır. Ağırlıklandırma işlemi, " α " düzeltme katsayısı ile sağlanır ve bu katsayı sıfır ile bir arasında bir değer alabilir. Böylece, geçmiş gözlem değerleri modele farklı oranlarda katkıda bulunur. Basit üstel düzeltme yönteminin genel matematiksel ifadesi Eş. 2.3'te verilmektedir (Bağcı, 2020).

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t \quad (2.3)$$

Zaman serisi tahmin yöntemleri arasında hareketli ortalamalar ve üstel düzeltme yöntemlerinin yanı sıra Holtz-Winters yöntemi, mevsimsel üstel düzeltme yöntemi gibi diğer yöntemler de bulunmaktadır.

Regresyon yöntemleri

Regresyon analizi, bir bağımlı değişken ile bir bağımsız değişken (basit regresyon) veya birden fazla bağımsız değişken (çoklu regresyon) arasındaki ilişkileri matematiksel bir denklem aracılığıyla ifade etme süreci olarak tanımlanır. Basit doğrusal regresyon modeli pek çok durumda yeterli olabilmektedir. Ancak, gerçek hayatta birçok olgunun açıklanabilmesi için genellikle iki veya daha fazla bağımsız değişkene ihtiyaç duyulur. Birden fazla bağımsız değişken içeren modeller ise çoklu regresyon modeli olarak

adlandırılır (Subaşı, Beycioğlu ve Çullu, 2010).

Regresyon analizi, sebep-sonuç ilişkisine sahip iki ya da daha fazla değişken arasındaki bağlantıyı tespit ederek, bu ilişki üzerinden geleceğe yönelik tahminler yapmayı mümkün kılan bir istatistiksel yöntemdir (Desticioğlu, 2015).

Regresyon yöntemleri arasında basit doğrusal regresyon, çoklu doğrusal regresyon, lojistik regresyon, ridge regresyon ve LASSO regresyon yöntemleri gösterilebilmektedir.

Doğrusal regresyon analizi belirlenmek istenen değişkenden daha kolay veya daha erken saptanabilen değişkenlerden yola çıkarak belirlenmek istenen değişkeni tahmin eden bir model oluşturmaktır. Bilinen normal dağılan sayısal bir değişkenden, bilinmeyen, aralarında ilişki olan bir başka normal dağılan sayısal değişkeni tahmin için uygulanırsa basit doğrusal regresyon, birden çok değişkenden yararlanarak bir değişkeni tahmin etmek amacıyla modelleme yapıldığında ise çoklu doğrusal regresyon olarak tanımlanır. Doğrusal regresyon Eş. 2.4'teki gibi ifade edilir.

$$Y = \beta_0 + \beta_1 \cdot x \quad (2.4)$$

Eğer model için bulunan p değeri <0.05 ise regresyon katsayısı sıfırdan farklıdır yani iki değişken arasındaki ilişki istatistiksel olarak önemlidir, iki değişken arasında doğrusal bir ilişki vardır. Buna karşılık modelin uygunluk göstergesi R^2 ile ifade edilir ve bu değer bire ne kadar yakınsa model o kadar iyidir (Alpar, 2010; Dawson, 2001).

Lojistik regresyon analizi, özellikle sosyal ve biyolojik bilimlerde yaygın bir kullanım alanına sahiptir. Sosyal bilimlerde yürütülen pek çok araştırma, bağımlı değişkenin yalnızca iki olası durumu içerebildiği varsayımına dayanmaktadır. Örneğin bir işçinin istihdam edilip edilmemesi gibi durumlar bu kapsamda değerlendirilebilir. Bu tür ikili durumları yansıtan verilere iki değerli veriler adı verilmektedir. Literatürde, bu değişkenler genellikle sıfır ve bir ile kodlanarak gösterilir. İki değerli bağımlı değişkenlerle çalışan araştırmacıların temel amacı, belirli bağımsız değişkenler kümesi doğrultusunda, belirli bir sonucun gerçekleşme olasılığını tahmin etmektir (Powers ve Xie, 2000).

İlk olarak 1970'te Hoerl ve Kennard tarafından geliştirilen ridge regresyon, çoklu doğrusal

bağlantı sorununu gidermek amacıyla önerilen ve klasik en küçük kareler yöntemine alternatif olarak yaygın şekilde benimsenen güçlü ve yanlı bir tahmin yöntemidir. En küçük kareler, doğrusal ve yansız tahmin ediciler arasında en düşük varyansa sahip olması nedeniyle teorik olarak en iyi performansı sunar. Ancak, uygulamada bağımsız değişkenler arasında yüksek korelasyonun sıkça görülmesi, en küçük kareler yönteminin tahminlerinin istikrarını bozmakta; bu da yüksek varyanslı ve güvenilirliği düşük katsayı tahminlerine yol açmaktadır. Ridge regresyon yöntemi, bu tür çoklu doğrusal bağlantı problemlerinin etkisini azaltmak ve en küçük kareler yaklaşımının tahmin gücünü artırmak amacıyla geliştirilmiş etkili bir regresyon tekniğidir (Mermi, 2024).

Ridge regresyonun çözüm yaklaşımı, klasik en küçük kareler yöntemine benzerdir; ancak temel fark, tahmin işleminden önce tasarım matrisinin ($X'X$) yapısına küçük ve pozitif bir sabit eklenmesidir. Bu sabit, özellikle korelasyonlu açıklayıcı değişkenler nedeniyle oluşan yüksek şartlı sayıyı azaltmak ve matrisin kararlılığını artırmak amacıyla köşegen elemanlara eklenir. $k=0$ durumunda ridge tahmini, en küçük kareler tahminine eşdeğer olur ve bu nedenle ridge regresyon, klasik yöntemin doğrusal bir genellemesi olarak da değerlendirilebilir (Sakallıoğlu ve Kaçıranları, 2008).

En Az Mutlak Büzülme ve Seçim Operatörü Regresyonu (Least Absolute Shrinkage and Selection Operator- LASSO), çoklu bağlantı (multicollinearity) sorununa çözüm getirmek amacıyla Tibshirani (1996) tarafından geliştirilmiştir. Yapısal olarak Ridge regresyona benzese de ceza terimi olarak L1 normunu kullanmasıyla ayrışır. Bu özellik sayesinde, yalnızca regresyon katsayılarını küçültmekle kalmaz, aynı zamanda bazı katsayıları sıfıra indirerek değişken seçimi de yapabilir. Ridge regresyon L2 normuna dayandığı için hiçbir katsayıyı tam olarak sıfırlayamazken, LASSO bu esnekliği sağlar. Bu yönüyle LASSO, hem tahmin gücü hem de değişken seçimi kapasitesi ile Ridge ve klasik en küçük kareler yöntemlerinden ayrılır. Ancak her üç yöntem arasında genel bir üstünlük hiyerarşisi kurmak mümkün değildir; model tercihi, veri yapısına ve analiz amacına göre belirlenmelidir (Vergili, 2023).

2.4.2. Makine öğrenmesi temelli tahmin yöntemleri

Makine öğrenmesi temelli yöntemler, geçmiş verilere dayalı olarak gelecekteki olayları tahmin etmeyi veya verileri önceden tanımlanmış sınıflara ayırmayı amaçlayan

algoritmalarıdır. Bu yöntemler, genel olarak denetimli öğrenme ve denetimsiz öğrenme olmak üzere iki temel gruba ayrılmaktadır. Bu bölümde, temel makine öğrenmesi algoritmaları denetimli ve denetimsiz öğrenme türlerine göre sınıflandırılmış ve algoritmalar ile işlevlerine ilişkin kısa açıklamalara yer verilmiştir.

Denetimli makine öğrenmesi yöntemleri

Karar ağaçları

Karar ağaçları, tek bağımlı değişken ve birden fazla bağımsız değişken içermeleri açısından regresyon modellerine benzemektedir. Ancak, karar ağaçları bunun ötesinde, verilerden regresyon modellerine alternatif olabilecek farklı ve faydalı örüntüleri keşfetme yeteneğine sahiptir (Cox, 2002). Karar ağaçları, bağımlı değişkenin kategorik olduğu durumlarda lojistik regresyona alternatif olabilecek bir yöntemdir.

Karar ağaçları, kurallara dayalı açıklamalar oluşturabilmesi hem sürekli hem de kesikli verilerle çalışabilmesi ve eksik ya da hatalı verilerle tahmin yapabilmesi gibi avantajlara sahiptir. Ayrıca, parametrik olmayan bir yöntem olarak, verinin dağılımına veya sınıflandırıcı yapısına dair herhangi bir varsayımda bulunmaz. Ancak, eksik veya hatalı verilere duyarsızlıkları ve yaprak düğümlerindeki mükerrerlik gibi dezavantajları da vardır (Rokach ve Maimon, 2005).

Rastgele orman

Breiman tarafından 2001 yılında geliştirilen rastgele orman algoritması, klasik karar ağaçlarından farklı olarak, her bir düğümde en iyi bölünmeyi tüm veri kümesinden değil, rastgele seçilen bir alt özellik kümesinden belirler. Bu yapı, her biri farklı kurallara göre çalışan çok sayıda karar ağacının oluşturulmasıyla inşa edilir. Sınıflandırma problemlerinde, her ağacın verdiği sınıf tahmini oylamaya tabi tutulur ve en çok oyu alan sınıf nihai tahmin olarak kabul edilir. Regresyon problemlerinde ise tahmin, tüm ağaçların çıktılarının ortalaması alınarak yapılır. Olasılık kuramındaki büyük sayılar yasasına dayalı olarak, rastgele orman algoritmasının aşırı öğrenmeye karşı dayanıklı olduğu söylenebilir. Bu yöntem, karar ağacı koleksiyonu olarak çalışmasına rağmen ağaç budama işlemi yapmaz ve önceden özellik seçimi gerektirmez. Hiperparametre ayarları; ağaç sayısı, dallanma

derinlikleri, düğüm bölme kriterleri, kullanılacak özellik sayısı ve entropi temelli yapılar için saflık değerleri gibi çeşitli bileşenler üzerinde yapılabilir.

Rastgele orman algoritmasında, eğitim ve test verisi önceden ayrılmamışsa, veri setinin yaklaşık üçte ikisi eğitim, geri kalan üçte biri ise test verisi olarak kullanılır. Her bir karar ağacı için örneklem oluşturulur ve bu örneklemle ağaçlar eğitilir. Her ağacın performansı, kendisine ait test verisiyle test edilerek hata oranı belirlenir. Bu hata oranlarının ortalaması alınarak ormanın genel hata değeri hesaplanır. Elde edilen hata oranlarına göre ağaçlara ağırlık verilir; hata oranı düşük olan ağaçlara yüksek, yüksek hata oranına sahip olanlara ise düşük ağırlık atanır. Bu ağırlıklar doğrultusunda ağaçlar sınıflandırma sürecine katkıda bulunur ve en yüksek toplam oyu alan sınıf, nihai tahmin olarak seçilir (Akman, Genç ve Ankaralı, 2011).

Destek vektör makineleri

DVM, istatistiksel öğrenme teorisine dayanan gözetimli bir öğrenme algoritmasıdır. İlk olarak 1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından ortaya atılmış, 1995 yılında ise Vapnik, Bernhard Boser ve Isabelle Guyon tarafından geliştirilmiştir (Akpınar, 2017).

Makine öğrenmesinde sınıflandırma ve regresyon gibi görevler için sıkça tercih edilen güçlü algoritmalarından biri DVM'dir. Vapnik (1995), bu yöntemi tanımlarken, veri uzayında sınıflar arasında maksimum ayrımı sağlayacak en uygun hiper düzlemi bulmayı amaçladığını belirtmiştir. DVM, çekirdek fonksiyonları sayesinde doğrusal olmayan ilişkileri ve yüksek boyutlu veri kümelerini işleyebilme yeteneğine sahiptir; bu da onu özellikle görsel ve metinsel veri sınıflandırmaları ile biyoinformatik ve kemoinformatik gibi alanlarda etkili hale getirmektedir. Yumuşak marjin ve düzenlileştirme yaklaşımları sayesinde, DVM gürültülü ve dengesiz veri setleriyle de başarılı şekilde başa çıkabilmektedir. Doğal dil işleme, öneri sistemleri ve anomali tespiti gibi pek çok uygulamada başarıyla kullanılan bu yöntem, günümüz makine öğrenmesi yaklaşımlarının temel yapı taşlarından biri olmaya devam etmektedir (Almadani, 2024).

K-en yakın komşu

KNN algoritması, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan, parametrik olmayan bir öğrenme yöntemidir. Bu algoritma, bir veri noktasının tahminini, veri kümesindeki 'k' en yakın komşusunun değerlerine dayanarak gerçekleştirmektedir. Söz konusu tahmin süreci, genellikle komşu noktaların ortalaması (regresyon) ya da modu (sınıflandırma) üzerinden yürütülür. KNN, veriler arasındaki benzerlikleri tanımlamak amacıyla çoğunlukla öklidyen uzaklık gibi ölçütler kullanarak en yakın noktaları belirler (Balcıoğlu ve Sezen, 2024).

KNN algoritması, eğitim gerektirmemesi, uygulamada kolay olması, yerel bilgilere uyum sağlaması ve paralel çalışabilmesi gibi avantajlarla sınıflandırma uygulamalarında tercih edilmektedir. Ayrıca, gürültülü verilere karşı dayanıklıdır. Ancak bu yöntemin, büyük veri setlerinde ve yüksek öznitelik sayılarında işlem yükü ve bellek ihtiyacının artması, performansın k komşu sayısı, uzaklık ölçütü ve öznitelik sayısı gibi parametrelere bağlı olarak değişkenlik gösterebilmesi gibi dezavantajları da vardır (Bhatia ve Vandana, 2010).

Naive bayes

Naive bayes algoritması, adını Bayes teoreminden alan ve tembel öğrenme yaklaşımına dayanan bir sınıflandırma yöntemidir. Az miktarda eğitim verisiyle etkili sonuçlar verebilir ve dengesiz veri setlerinde de kullanılabilir. Her bir olasılığı hesaplayarak en yüksek değere sahip sınıfa göre tahmin yapar. Ancak, test verisinde yer alıp eğitim verisinde hiç görülmeyen değerlerle karşılaştığında, bu değerlere sıfır olasılık atar ve sınıflandırma yapamaz. Bu soruna "sıfır frekans" denir ve çözüm olarak Laplace tahmini gibi düzeltme yöntemleri kullanılabilir. Naive Bayes teoremi Eş. 2.5'teki eşitlik ile ifade edilmektedir.

$$p(x/y) = \frac{P(y/x) * P(x)}{P(y)} \quad (2.5)$$

Burada, $P(x)$, x olayının; $P(y)$ ise y olayının olasılığını belirtir. $P(y/x)$, x gerçekleştiğinde y 'nin olma olasılığı yani y 'nin x 'e bağlı şartlı olasılığıdır. Benzer şekilde, $P(x/y)$ de y gerçekleştiğinde x 'in olma olasılığını ifade eder.

Yapay sinir ağıları

Yapay sinir ağıları (YSA), insan beynindeki sinir hücrelerinin yapısından esinlenilerek geliştirilen ve öğrenme, bilgi işleme ile yeni bilgi üretme gibi bilişsel yetenekleri sayısal olarak modelleyen sistemlerdir. Biyolojik sinir hücrelerinin yapı ve işleyişine dayanan bu yapılar; giriş, gizli ve çıkış katmanlarından oluşan çok katmanlı bir mimariye sahiptir. Yapılarında bulunan yapay nöronlar doğrusal olmayan özellik taşıdıkları için, YSA'lar doğrusal olmayan ve karmaşık problemlerin çözümünde etkili biçimde kullanılmaktadır. Algoritmanın açıkça tanımlanmadığı durumlarda bile, yeterli örnek veriyle öğrenme gerçekleştirebilir; eksik ya da hatalı verileri tolere edebilir ve daha önce karşılaşılmamış durumlar için genelleme yapabilirler. Esnek yapıları sayesinde değişken koşullara uyum sağlayabilen YSA'lar, aynı zamanda yüksek hızda ve doğrulukta sonuçlar üretebilmekte; bu da onları çok çeşitli uygulama alanlarında güçlü bir araç hâline getirmektedir (Çağlayan, 2016).

YSA'lar, sahip oldukları özellikler ve yetenekler sayesinde birçok alanda önemli avantajlar sunmaktadır. Bilgiyi merkezi bir veritabanı yerine ağ boyunca dağıtık olarak depolayan bu yapılar, bilgi kaybına karşı yüksek dayanıklılık gösterir. Paralel işlem yapabilme kapasiteleri sayesinde daha hızlı ve güvenilir sonuçlar üretirler. İnsan beyninden ilham alınarak geliştirilen YSA'lar, eğitim yoluyla öğrenebilir, eksik veri içeren durumlarda bile tahminlerde bulunabilir ve doğrusal olmayan karmaşık ilişkileri modelleyebilir. Hata toleransı sayesinde ağı bazı birimleri işlev dışı kalsa bile genel işleyiş etkilenmez. Ayrıca, dağıtık belleğe sahip olmaları, bilginin hücre bağlantıları aracılığıyla temsil edilmesini sağlar. YSA'lar, zaman içinde kademeli olarak bozulmaları sayesinde ani sistem çöküşlerinden kaçınır; makine öğrenimi yetenekleri sayesinde olaylardan öğrenerek yeni durumlara ilişkin çıkarımlar yapabilir. Bu özellikleri, YSA'ları karmaşık problem çözümünde güçlü ve esnek araçlar hâline getirmektedir (Güller, 2024).

YSA'nın en önemli sınırlılıkları arasında uzun ve karmaşık eğitim süreci, ağ davranışlarının açıklanabilirliğinin düşük olması ve donanım bağımlılığı yer almaktadır. YSA'ların çözüm üretme süreci genellikle "kara kutu" olarak tanımlanır; çünkü elde edilen sonuçların nasıl ve neden ortaya çıktığı net şekilde açıklanamaz. Paralel işlem yapısına dayandıkları için güçlü donanım altyapısına ihtiyaç duyarlar. Ayrıca, uygun ağ yapısının belirlenmesi deneyime dayalıdır ve standart bir yöntemi bulunmamaktadır. Sayısal veri işleme yeteneğine sahip

olmaları nedeniyle problemlerin ağ tarafından işlenebilmesi için önceden sayısal forma dönüştürülmesi gerekir ki bu süreç hem görüntüleme mekanizmalarına hem de kullanıcı yeterliliklerine bağlı olarak değişkenlik gösterir. Eğitim sürecinin tamamlandığı, genellikle hata değerinin belirli bir eşik altına düşmesiyle kabul edilse de bu her zaman en iyi sonuçları garanti etmez. Tüm bu faktörler, YSA uygulamalarının etkinliğini doğrudan etkileyen temel zorluklar olarak öne çıkmaktadır (Güller, 2024).

YSA, çok sayıda yapay sinir hücresinin birbirleriyle bağlantılı şekilde çalışmasıyla oluşur ve bu hücreler genellikle katmanlar hâlinde düzenlenir. YSA'lar temel olarak üç ana katmandan oluşur: girdi katmanı, gizli katman(lar) ve çıktı katmanı. Girdi katmanı, dış çevreden gelen verileri alarak gizli katmana iletir; bu iletim doğrudan ya da dönüştürülerek gerçekleştirilebilir. Gizli katman bir ya da birden fazla katmandan oluşabilir ve aldığı bilgileri işleyerek çıktı katmanına aktarır. Son olarak, çıktı katmanı işlenen bilgileri değerlendirerek ağın nihai çıktısını üretir. Katmanlardaki hücre sayısı ve yapısı, çözülmek istenen probleme göre değişkenlik gösterebilir (Gülmez, 2016).

Gradyan arttırma karar ağaçları

Gradyan artırma karar ağaçları, sınıflandırma ve regresyon gibi makine öğrenimi problemlerinde yaygın olarak kullanılan, güçlü ve esnek bir algoritmadır. İlk olarak Friedman tarafından 1999'da önerilen ve 2002'de geliştirilen bu yöntem, zayıf öğrencilerin genellikle karar ağaçlarının art arda eklenerek hataların kademeli olarak azaltılması prensibine dayanır. Arttırma yaklaşımı, her yeni modelin bir önceki modelin tahmin hatalarını düzeltmesini sağlayarak daha güçlü bir genel tahmin modeli oluşturur (Acar, 2022).

Algoritma, başlangıçta basit bir tahmin fonksiyonuyla başlar ve her iterasyonda gerçek değerle tahmin arasındaki fark üzerinden yeni ağaçlar eklenerek model güncellenir. Bu işlem, kayıp fonksiyonunun türevi alınarak gradyan yönünde optimize edilir. Böylece modelin hata oranı her adımda azaltılır. Bu süreçte parametreler (örneğin, öğrenme oranı, ağırlıklar) modeli daha hassas hâle getirir (Acar, 2022).

Son yıllarda büyük veri kullanımının artmasıyla birlikte gradyan artırma karar ağaçları yönteminde çeşitli iyileştirmeler yapılmış ve ekstrem gradyan artırma karar ağaçları

(extreme gradient boosting-XGBoost), hafif gradyan artırma makinesi, CatBoost gibi gelişmiş algoritmalar ortaya çıkmıştır. Bu modern yaklaşımlar, gradyan artırma karar ağaçları algoritmasının temel prensiplerini korumakla birlikte daha hızlı, daha verimli ve daha düşük kaynak gerektiren çözümler sunmaktadır. Özellikle XGBoost, açık kaynak olması, geniş kullanıcı topluluğu ve dokümantasyon desteği ile literatürde en yaygın kullanılan gradyan artırma yöntemlerinden biri hâline gelmiştir (Acar, 2022).

XGBoost yöntemi

XGBoost, son yıllarda makine öğrenimi alanında öne çıkan en başarılı algoritmalarından biridir. Regresyon ve sınıflandırma problemlerinde boosting tekniğini kullanarak yüksek doğrulukta tahminler üretebilmekte; özellikle spam e-posta tespiti, reklam eşleştirme, dolandırıcılık analizi ve anomalilerin belirlenmesi gibi uygulama alanlarında etkin şekilde kullanılmaktadır (Chen, Guestrin, 2016). Yöntemin temel avantajları arasında hesaplama verimliliği, büyük veri setleriyle uyumlu çalışabilmesi ve tahminlerin doğrudan veriye dayalı olarak elde edilmesi yer almaktadır. Ayrıca, değişkenler arasındaki ilişki bilgi kazancı üzerinden değerlendirilebildiğinden, model yorumlanabilirliği açısından da önemli katkılar sunmaktadır (Yeni, 2024).

XGBoost, yüksek doğruluk, esneklik ve ölçeklenebilirlik özellikleriyle öne çıkan bir makine öğrenimi kütüphanesidir. Özellikle büyük ve karmaşık veri setlerinde sınıflandırma ve regresyon problemleri için yaygın olarak kullanılmaktadır. Temelinde karar ağaçları ile gradient boosting algoritmalarının birleşimi yer alır. Bu yaklaşımda, zayıf öğrenciler olarak tanımlanan karar ağaçları birbiri ardına eklenerek modelin tahmin gücü artırılır. Her yeni ağaç, önceki ağaçların oluşturduğu hataları azaltmayı hedefler (Yeni, 2024).

XGBoost'un başarısında önemli rol oynayan bir diğer unsur da hata fonksiyonunun optimizasyonudur. Model, her iterasyonda tahmin hatalarını azaltmak amacıyla belirli bir kayıp fonksiyonunu (örneğin, ortalama kare hata veya çapraz entropi) minimize eder. Aşırı öğrenmenin önüne geçmek için ise regularizasyon tekniklerinden yararlanır; bu, modelin karmaşıklığını kontrol eden hiperparametreler üzerinden gerçekleştirilir. Paralel işlem desteği ve bellek verimliliği sayesinde büyük veri üzerinde hızlı ve etkili sonuçlar elde edilebilir. Ayrıca, maksimum derinlik, minimum yaprak ağırlığı gibi ağaç yapılandırma parametreleriyle ağacın büyüme süreci hassas biçimde kontrol edilebilir. Bu özellikleri

sayesinde XGBoost, akademik çalışmalardan endüstriyel uygulamalara kadar geniş bir kullanım alanına sahiptir (Yeni, 2024).

Denetimsiz makine öğrenmesi yöntemleri

Yoğunluğa dayalı kümeleme yöntemi

Kümeleme, nesnelerin benzerliklerine göre gruplandırılması sürecidir. Aynı grupta yer alan veriler genellikle bazı ortak özellikleri paylaşır. Kümeleme için pek çok algoritma geliştirilmiştir ve bunlardan biri de yoğunluk tabanlı kümeleme yöntemidir. Bu yaklaşımda, bir veri kümesindeki birbirine yakın konumda bulunan noktalar, belirli bir yoğunluk kriterine göre kümeler oluşturur. Bu tür kümeler, önceden belirlenmiş geometrik şekillere bağlı kalmaksızın serbest biçimlerde olabilirler. Örneğin, bazı kümeler halka şeklinde ya da eğimli bir yapı gösterebilir. Kümelerin dışında kalan ve herhangi bir kümeye dahil edilemeyen veri noktaları, genellikle gürültü olarak değerlendirilir. Kümeleme sürecinde, "çekirdek nokta" kümeye merkezi katkı sağlayan ve yoğunluk oluşturan noktayı ifade ederken, "sınır noktası" kümeyi çevreleyen ve kümeye komşu olan noktaları belirtir. Yoğunluğa dayalı kümeleme yöntemi yöntemi, her bir veri noktasının yoğunluğunu, bu noktanın diğer tüm noktalara göre etki fonksiyonları toplamı üzerinden hesaplayarak kümeleri belirler (Hinneburg ve Keim, 1998). Aykırı değer tespitinde yoğunluk tabanlı kümeleme yöntemi (A Density Clustering Based on Outlier Removal) algoritması ise hem uç noktaların (aykırı değerlerin) ayıklanması hem de kümeleme işlemini birlikte gerçekleştirir. Bu yöntem, 0 ile 1 arasında değişen bir eşik değeri kullanarak farklı boyut, şekil ve yoğunluktaki kümeleri tanımlayabilir (Fahim, Saake, Salem, Torkey ve Ramadan, 2008). Kategorik Verilerin Verimli Katmanlı Yoğunluk Tabanlı Kümeleme Yöntemi (Efficient Layered Density-based Clustering of Categorical Data), özellikle biyomedikal alandaki kategorik veriler için geliştirilmiştir. Bu yöntem, her veriyi özel bir indeksleme yapısıyla temsil eder ve herhangi bir kullanıcı parametresi gerektirmeden çalışır. Böylece yeni bir veri noktası geldiğinde, hangi kümeye ait olduğuna hızlı bir şekilde karar verilebilir. Ayrıca, bu yöntemle bir küme içinde farklı yoğunluklarda başka alt kümeler de tespit edilebilir (Andreopoulos, An, Wang ve Labudde, 2009).

K-ortalamlar yöntemi

K-ortalamlar algoritması, yinelemeli bir kümeleme tekniğidir. Bu yöntemin temel amacı, verileri sahip oldukları özelliklere göre önceden belirlenmiş sayıda kümeye ayırmaktır. Kümeleme işlemi sırasında birimlerle küme merkezleri arasındaki uzaklıkların hesaplanmasında genellikle Öklid uzaklığı kullanılır. K-ortalamlar, kümeler içindeki veri noktaları arasındaki farkı en aza indirmeye ve farklı kümeler arasındaki farkı en fazla seviyeye çıkarmaya çalışır. Algoritma, ilk olarak veri setinden rastgele seçilen k adet merkezle başlar. Ancak bu rastgele başlangıç noktaları, kümeleme sonuçlarının doğruluğunu etkileyebileceği için bir dezavantaj oluşturabilir. Seçilen bu k merkez, verideki tüm birimlerin en yakın merkeze atanmasıyla ilk kümeleri oluşturur. Devam eden adımlarda ise her kümenin merkezi, o kümedeki verilerin ortalaması alınarak güncellenir ve işlem, kümeler sabitlenene kadar sürdürülür (Arslan, 2022).

K-ortalamlar algoritmasının temel işlem adımları aşağıdaki şekilde sıralanabilmektedir:

1. Veri setindeki küme sayısı belirlenir.
2. Belirlenen küme sayısı kadar merkez noktası, başlangıçta rastgele olarak seçilir. Bu noktalar, ilk küme merkezleri olarak kullanılır.
3. Veri kümesindeki her bir gözlem birimi, küme merkezlerine olan Öklid uzaklıkları hesaplanarak en yakın merkeze atanır.
4. Her bir küme için, o kümeye atanan gözlemlerin aritmetik ortalaması alınarak yeni merkez noktaları belirlenir.
5. Küme merkezleri değişmediği sürece 3. ve 4. adımlar tekrarlanır. Merkezlerde değişiklik olmadığı durumda algoritma yakınsama sağlar ve süreç sona erer (Arslan, 2022).

Hiyerarşik kümeleme yöntemi

Hiyerarşik kümeleme algoritması, temel olarak benzer özelliklere sahip verilerin gruplanması veya mevcut bir kümenin alt kümelere ayrılması mantığına dayanmaktadır. Bu yöntemde iki ana yaklaşım söz konusudur: birleştirici ve ayrıştırıcı yöntemler. Birleştirici yaklaşım, tümevarım temellidir ve her veri noktasının başlangıçta ayrı bir küme kabul edilmesiyle sürece başlanır. Ardından, benzer özellikler gösteren kümeler bir araya getirilerek daha büyük kümeler oluşturulur ve süreç, tüm verilerin tek bir kümede toplanmasına kadar devam eder. Buna karşılık, ayrıştırıcı yaklaşım tümünden gelim esasına

dayanır. Bu yöntemde başlangıç noktası tek bir kümedir; tüm veriler bu küme içinde yer alır. Daha sonra, uzaklık ya da benzerlik matrisine göre veriler giderek daha küçük gruplara ayrılarak alt kümeler oluşturulur. Sürecin sonunda her veri noktası tek başına bir küme haline gelir. Hiyerarşik kümeleme analizinde, her adımda veri noktaları arasındaki benzerlik veya uzaklık yeniden hesaplanır. Bu hesaplamalarla oluşturulan matris, hangi bağlantı yönteminin uygulanacağını belirlemek için kullanılır (Ülgen, 2021).

Bağlantı yöntemleri, kümelerin hangi uzaklık ya da benzerlik seviyesinde birleşeceğini veya ayrılacağını belirleyen kurallardır. Bu yaklaşımla veriler adım adım kümelendir ve her aşamada yeni kümeler ortaya çıkar. Anderberg'e göre, araştırmacının veriler üzerinde kaç küme bulunduğunu önceden bilmediği durumlarda hiyerarşik kümeleme oldukça uygun bir yöntemdir. Ayrıca, veri seti üzerinde daha önce fark edilmemiş yapılar ve ilişkilerin keşfedilmesine imkân tanınması yönüyle de değerlidir (Michael, 1973).

Birleştirici Hiyerarşik Kümeleme yönteminde her gözlem birimi başlangıçta ayrı bir küme olarak değerlendirilir. Ardından benzer kümeler kademeli olarak birleştirilerek toplam küme sayısı sırasıyla $n, n-1, n-2, \dots, 2, 1$ şeklinde azaltılır. Süreç şu şekilde özetlenebilir:

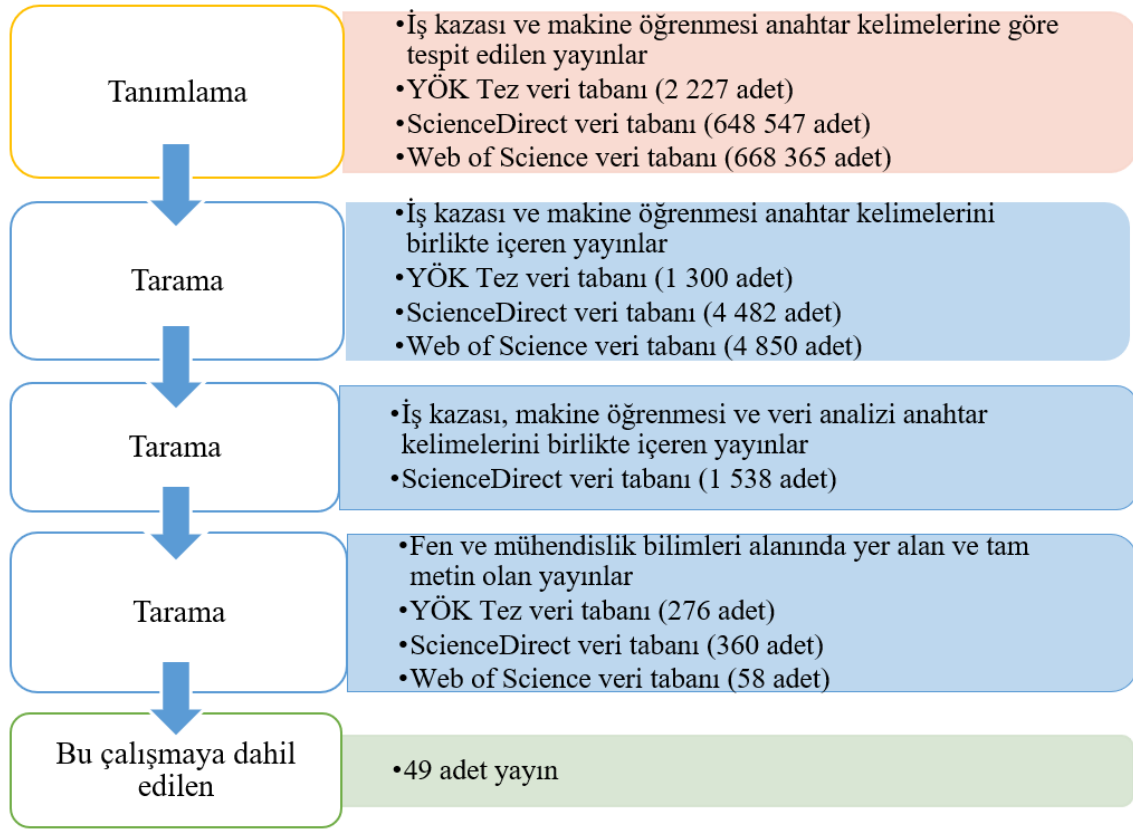
1. Tüm veriler kendi başına birer küme olarak kabul edilir (n tane küme).
2. En küçük uzaklığa sahip iki küme birleştirilir.
3. Küme sayısı bir azaltılır ve uzaklık matrisi güncellenir.
4. 2. ve 3. adımlar, $n-1$ kez tekrarlanır.

Ayrıştırıcı Hiyerarşik Kümeleme yöntemi, birleştirici yaklaşımın tersidir. Başlangıçta tüm veriler tek bir küme içerisinde yer alır. Ardından, benzer veriler bir arada kalacak şekilde ana küme kademeli olarak bölünür. Sürecin sonunda her gözlem birimi kendi başına bir küme haline gelir. Küme sayısı bu süreçte $1, 2, 3, \dots, n$ şeklinde artar (Doğan, 2008).



3. LİTERATÜR ARAŞTIRMASI

Bu bölümde, iş kazaları analizi konusunda geleneksel istatistiksel yöntemler ve makine öğrenmesi yöntemleri kullanımları ile ilgili yapılan çalışmalara yer verilmiştir. Literatür taraması YÖK Tez Veritabanı, Science Direct, Web of Science veri tabanları üzerinde “iş kazaları”, “makine öğrenmesi”, “veri analizi” gibi konuyla ilgili anahtar kelimeler kullanılarak yapılmıştır. YÖK Tez Veritabanı’nda “iş kazası/kazaları” terimleri araması sonucu 585 adet, “makine öğrenmesi” terimi araması sonucu 1 642 adet çalışma bulunmuştur, “iş kazaları” ve “makine öğrenmesi” terimleri birlikte aratıldığında 1 300 adet sonuç bulunmuştur. ScienceDirect veritabanında “iş kazaları” terimi araması sonucu 52 383 adet, “makine öğrenmesi” terimi araması sonucu ise 596 164 adet, “iş kazaları” ve “makine öğrenmesi” terimleri birlikte aratıldığında 4 482 adet yayın bulunmuştur. “İş kazaları”, makine öğrenmesi” ve “veri analizi” terimlerini içeren 1 538 adet yayına ulaşılmıştır. Web of Science veritabanında “iş kazaları” terimi araması sonucu 11 824 adet, “makine öğrenmesi” terimi araması sonucu 656 541 adet, “iş kazaları” ve “makine öğrenmesi” terimleri birlikte aratıldığında 169 adet yayın bulunmuştur. Daha sonra, bu veri tabanlarında fen ve mühendislik bilimleri alanında yapılmış; “iş kazası”, “makine öğrenmesi” ve “veri analizi” anahtar kelimelerini birlikte içeren ve tam metnine erişilebilen çalışmalar ayrılmıştır. Mükerrer yayınlar bu araştırma kümesinden çıkarıldıktan sonra, 694 adet yayından oluşan ve tez çalışması literatür araştırmasında incelenecek çalışmalar kümesi oluşturulmuştur. Bu araştırma kümesinde yer alan tüm çalışmalar incelendikten sonra, tez çalışmasının konusu ve uygulama yöntemleri açısından kavramsal ve kapsam bakımından uygun olduğu belirlenen 49 yayın, tez çalışması kapsamında yürütülen literatür araştırmasında detaylı olarak analiz edilmiş ve takip eden alt bölümlerde açıklanarak sunulmuştur. Yürütülen literatür araştırmasına ilişkin gösterim Şekil 3.1’de gösterilmektedir.



Şekil 3.1 Literatür taraması gösterimi

Bulunan sonuçlardan analitik analiz içeren çalışmalar bu bölümde ayrıntılarıyla incelenmiştir. Bu bölümde iş kazaları analitik analizi çalışmalarında geleneksel yöntemlerin kullanıldığı çalışmalar ve makine öğrenmesi yöntemlerinin kullanıldığı çalışmalar olmak üzere iki başlık altında verilmiştir.

3.1. İş Kazası Analizine İlişkin Genel Literatür Taraması

İş kazalarının analizi alanında farklı disiplinlerden birçok araştırmacı çok sayıda çalışmalarda bulunmuştur. Bu çalışmalar sektörel bazda olduğu gibi bölgesel bazda da olabilmektedir.

Demir (2009), çalışmasında 25 farklı metal iş kolundaki 2007 yılında meydana gelen iş kazalarını istatistiksel olarak inceleyerek, bu kazaların dolaylı ve dolaysız maliyetlerini analiz etmeyi amaçlamıştır. Çalışmada anket yöntemi ile araştırma yapılmıştır. Araştırma sonuçlarına göre, Türkiye'deki tüm iş kazalarının %23'ünün metal sektöründe meydana

geldiği, iş güvenliği yatırımları ile iş kazalarının maliyetleri arasındaki ilişki incelenmiş ve bu maliyetlerin Gayri Safi Milli Hasıla'nın %0,4'üne denk geldiği hesaplanmıştır.

Köseoğlu (2013), çalışmasında Türkiye genelinde iş kazalarının tahminine yönelik modeller geliştirmek amacıyla istatistiksel yöntemlerden yapısal eşitlik modeli kullanmıştır. Çalışmada, SGK'nın 1988-2010 yılları arasındaki verileri kullanılarak sigortalı işçi sayısı, işyeri sayısı, iş kazası, ölüm ve iş göremezlik sayıları model parametreleri olarak değerlendirilmiştir. Üç farklı senaryo çerçevesinde yapılan modellemelerde; iş kazası, geçici iş göremezlik, yaş, çalışan sayısı, mesai türü, ölüm ve sürekli iş göremezlik sayıları gibi değişkenler üzerinden tahmin yapılmıştır. Her bir senaryo kapsamında parametreler arasındaki yapısal ilişkiler analiz edilmiş ve bu ilişkilerin etki düzeyleri ortaya konmuştur.

Desticioğlu (2015), iş kazalarını tahmin etmek amacıyla, ülkemizde en çok iş kazasının yaşandığı sektörler üzerinde çalışma gerçekleştirmiştir. 2001-2012 yılları arasındaki SGK verileri kullanılarak en fazla iş kazasının yaşandığı dört sektör için geçmiş yıllara ait veriler incelenmiş, bu veriler doğrultusunda basit hareketli ortalamalar, ağırlıklı hareketli ortalama, basit üstel düzeltme, Brown'ın üstel düzeltme tekniği, Holt'un iki parametrelili üstel düzeltme tekniği, basit regresyon analizi ve çoklu regresyon analizi olmak üzere farklı istatistiksel tahmin yöntemleri uygulanmıştır. Uygulanan tahmin yöntemlerinin doğruluğu, tahmin hatası analizleri ile test edilmiş ve her sektör için en uygun tahmin yöntemi belirlenmiştir.

Çavuş ve Taçgın (2016), Türkiye'de inşaat sektöründeki 2005-2009 yılları aralığında gerçekleşen iş kazalarını incelemiş, kazaları oluş biçimlerine göre sınıflandırmış ve her bir kaza tipinin sayısal dağılımlarını analiz ederek kazaların nedenleri araştırmış ve değerlendirmiştir.

Tuna (2017), iş kazalarının analizinde bugüne kadar kullanılan yöntemlerin ortaya çıkardığı sonuçların dışında; öngörülemedi, beklenmeyen yeni yapıların keşfi için graf teorisini kullanarak Türkiye'de 2013-2014 yıllarında gerçekleşmiş iş kazaları verilerini analiz etmeyi amaçlamıştır. Analiz ile iş kazalarının kompleks yapısı grafik teorisi ile incelenerek, klasik yöntemlerle tespit edilemeyen gizli yapılar, ilişkiler ve öne çıkan risk alanları ortaya çıkarılmıştır.

Yağımlı ve İzci (2017) yaptıkları çalışmada metal sektörüne ilişkin 2008-2015 yıllarına ait SGK verileri kullanarak 2016 yılına ait iş kazası ve ölümlü iş kazası sayılarının tahmin edilmesini amaçlamışlardır. Tahmin yöntemi olarak uyarlanabilir bulanık zaman serisi ve en küçük kareler yöntemleri kullanılmıştır. İş kazası tahminlerinde uyarlanabilir bulanık zaman serisi, ölümlü iş kazası tahmininde ise en küçük kareler yöntemlerinin daha doğru sonuç verdiğini gözlemlemişlerdir.

Bayraktar, Uyguçgil ve Konuk (2018) yaptıkları çalışmalarında 2002-2015 yılları arasında SGK tarafından elde edilen iş kazası verileri istatistiksel olarak incelenmiş, madencilik sektörü ile diğer sektörlerin kaza oranları karşılaştırılmıştır. Analizler sonucunda madencilik sektöründe çalışan başına düşen iş kazası ve ölüm oranlarının diğer sektörlerle kıyasla daha yüksek olduğu belirlenmiştir.

Demirel (2018), Türkiye’de 2013-2016 yılları arasında gerçekleşen iş kazaları ile Avrupa ülkelerindeki iş kazaları karşılaştırmıştır. Kümeleme ve çok boyutlu ölçekleme analizi ile incelenen ülkelerin 3 gruba ayrılabilceği belirlenmiştir.

Memiş (2018), çalışmasında Türkiye’de 2008–2016 yılları arasında meydana gelen iş kazaları sektörel bazda incelemiş ve en çok kazanın üretim, inşaat ile taşımacılık-dağıtım sektörlerinde gerçekleştiği belirlemiştir. 2008-2016 yılları arasındaki iş kazaları istatistikleri, iş göremezlik süreleri ve ölümlü vaka sayıları doğrultusunda meslek grupları sektörel olarak 12 adet gruba ayrılmış, sektörlerin iş kazası ağırlık oranları ve iş kazası sıklık hızları belirlenmiş ve birbirleri ile kıyaslanmıştır. Hiyerarşik kümeleme analizi ve çok boyutlu ölçekleme analizi uygulanmış, sektörler iki boyutlu uzayda bu kriterlere göre konumlandırılmıştır.

Çakmak (2019), belirli bir sektörle sınırlı olmamakla birlikte sanayi üretimi bağlamında artan iş kazalarının sebepleri analiz edilmiştir. SGK’nın 2016 ve 2017 yıllarına ait verileri kullanılarak, iş kazalarının sayısal artışı ve bu kazalar sonucunda yaşanan ölümler ele alınmıştır. Çalışmada iş kazaları “ciddi” ve “hafif” olmak üzere sınıflandırılmış, bu sınıflandırma doğrultusunda kazaların sebeplerine ilişkin çeşitli değişkenler sorgulanmıştır. Büyük veri içerisinden anlamlı bilgilerin çıkarılması amacıyla veri madenciliği yöntemleri kullanılmıştır.

Kölgeli (2019), SGK istatistik yıllıklarından elde edilen veriler kapsamında 2008-2019 yılları arasında inşaat sektöründe meydana gelen iş kazaları ve ölümlü iş kazaları sektörün alt dallarına göre istatistiksel olarak analiz etmiştir. Araştırmada, 2008–2013 ve 2014–2019 yılları olmak üzere iki örneklem grubu oluşturulmuş; bu dönemler arasında iş kazalarının ve ölümlü iş kazalarının değişimi SPSS programı aracılığıyla karşılaştırmalı olarak incelenmiştir. Elde edilen bulgular, Türkiye’de iş kazalarının en yoğun yaşandığı sektörün inşaat sektörü olduğunu ortaya koymuş, kazaların temel nedeninin ise eksik uygulanan iş sağlığı ve güvenliği önlemleri olduğu vurgulanmıştır.

Özcan (2019) inşaat sektöründe meydana gelen kazaların nedenlerini araştırmış ve araştırma sonucunda nedenleri “yönetimsel kaynaklı nedenler, ekipman ve malzemelerden kaynaklı nedenler ve çalışanlardan kaynaklı nedenler” olmak üzere üç ana başlıkta toplamıştır. İş kazalarına neden olan faktörlerin önem derecelerini bulmak ve nedenler üzerinden firmaları karşılaştırmak için bulanık çok kriterli karar verme yöntemleri kullanılmıştır. İkili karşılaştırma anketlerinden elde edilen veriler, bulanık analitik hiyerarşi prosesi yöntemi ile analiz edilerek iş kazasına neden olan faktörler ağırlıklandırılmıştır. Ana kriterler içerisinde iş kazasının meydana gelmesinde en etkili faktörün ekipman ve malzemeden kaynaklı nedenler olduğu görülmüştür. Söz konusu yöntem sonucunda elde edilen bulgular ışığında bulanık TOPSIS yöntemi kullanılarak inşaat sektöründe faaliyet gösteren üç firma karşılaştırılmış ve firma bazında iş kazalarını önleyici önerilerde bulunulmuştur.

Öztürk (2019), inşaat sektöründeki iş kazalarını ve bu kazaların şiddetini etkileyen faktörleri inceleyerek, özellikle iklim faktörlerinin etkisini belirlemeyi amaçlamıştır. Bu doğrultuda, SGK ve Meteoroloji Genel Müdürlüğü’nden alınan veriler ile iş kazası verileri üzerinde sıklık, ki-kare ve lojistik regresyon yöntemleri kullanılarak istatistiksel analizler yapılmıştır. Kaza şiddeti, hafif, orta ve ağır olmak üzere üç kategoriye ayrılarak, etkileyen faktörlerle birlikte bir kaza şiddeti tahmin modeli oluşturulmuştur. Modelin doğruluğunu artırmak için kaza şiddeti iki kategorili hale getirilerek ikili lojistik regresyon analizi uygulanmıştır. Sonuçlar, iklim faktörlerinin de iş kazalarının şiddetini etkileyen önemli bir etken olduğunu göstermiştir.

Ünal (2019) çalışmasında herhangi bir sektöre özgü olmaksızın Türkiye ve Avrupa Birliği’ne (28 üye ülke) ait iş kazaları istatistiksel olarak karşılaştırmıştır. 2008–2016 yılları arasındaki SGK ve Avrupa İstatistik Ofisi verileri kullanılarak cinsiyet, yaş grupları, ölümlü

ve yaralanmalı kazalar ile işletme büyüklüğü gibi değişkenlere göre analiz yapılmıştır. Veri analizi için ki-kare testi ve lojistik regresyon analizi yöntemleri kullanılmıştır.

Seylan (2021), gıda sektöründe 2010-2019 yılları arasında SGK'dan elde edilen iş kazası ve meslek hastalıkları verilerini istatistiksel olarak analiz edilmiştir. Çalışmada verilerin yıllık dağılımı, cinsiyete göre farklılıklar ve ölümle sonuçlanan vakalar sayısal olarak değerlendirilmiştir. Bulgulara göre, gıda sektörü tüm iş kazalarının ve meslek hastalıklarının yaklaşık %10'unu oluşturmakta ve iş kazalarına bağlı ölümler açısından en riskli ilk 10 sektör arasında yer almaktadır.

Şimşek (2021), Türkiye'deki inşaat sektöründe iş kazalarının sıklığını analiz etmiştir. Çok kriterli karar verme yöntemleriyle bütünleşik garrett sıralama tekniği kullanılarak Türkiye'nin 81 ili, iş kazası yaşanma sıklığına göre sıralanmış ve bir iş kazası risk haritası oluşturulmuştur.

Şimşek Akın (2021), Türkiye'de SGK'nın 2008-2017 arasındaki 10 yıllık çalışma verileri ile çalışanların yaşı, cinsiyeti ve iş kazasının meydana geldiği saat dilimi gibi değişkenleri log-lineer analiz yöntemiyle istatistiksel olarak incelemiştir.

Toptancı (2021) çalışmasında, SGK'dan temin edilen, İç Anadolu Bölgesi'ndeki inşaat iş yerlerine ait 2013-2017 dönemi iş kazası verilerini kullanarak hem iş günü kaybı hem de kaza sonucu tahminine yönelik çeşitli regresyon modelleri (lojistik regresyon, firth, ridge, LASSO, elastik net, grup LASSO, grup elastik net) uygulamıştır. Ayrıca, risk analizi için Fine-Kinney tabanlı aralık değerli nötrosofik AHP ve CoCoSo yöntemlerinin bütünleştiği çok ölçütlü karar verme modeli kullanmış; risk sınıflarının belirlenmesinde k-ortalamalar kümeleme yöntemini tercih etmiştir. Tespit edilen yüksek riskli tehlikeler, ADN-kalite evi modeliyle alınacak tedbirlerle ilişkilendirilerek bu tedbirlerin önceliklendirilmesi sağlanmıştır.

Demir (2022), metal sektöründeki 2012-2018 yılları arasındaki iş kazalarını incelemiş, bu alandaki risk analizi çalışmalarını değerlendirmiş ve sektörde yaşanan iş kazaları istatistiklerini incelemiştir. 5x5 matris ve hata türleri ve etkileri analizi yöntemleri kullanılarak risk değerlendirmesi yapılmıştır.

Demiroğlu (2023), madencilik sektöründe iş kazaları ve meslek hastalıklarını incelemiştir. 2002-2020 yılları arasındaki veriler, Türkiye İstatistik Kurumu ve SGK veri tabanlarından alınarak analiz edilmiştir. Analiz sonucunda 2012 yılından itibaren ölüm oranları azaldığı, ancak 2002-2020 yılları arasında iş kazaları sayısında %50'lik bir artış gözlemlendiği, en çok iş kazasının 2013 yılında yaşandığı ve madencilik faaliyetlerinin en çok kömür, linyit çıkarımı ve taşocaklarında yoğunlaştığı bulunmuştur. Ayrıca, madencilik sektöründeki iş kazalarının her beş yılda bir %10 oranında arttığı belirlenmiştir.

Kosutova (2023), Çekya'da kara taşımacılığı ve ulaşım sektöründe 2012-2021 yılları arasında meydana gelen kayıtlı, ciddi ve ölümcül iş kazalarının sayısını incelemiştir. İşletmeler 250 çalışan altında olan ve üstünde olan olarak ikiye ayrılmış ve iş kazalarının nedenleri istatistiksel olarak analiz edilmiştir. Bulgular, kara taşımacılığı ve ulaşım sektöründe meydana gelen iş kazalarının büyük oranda insan hatasından kaynaklandığını ortaya koymuştur.

Pelit (2023), metal sektöründe yaşanan iş kazalarını ve bu kazalardan kaynaklanan ölümleri incelemiştir. Araştırmada, SGK verileri kullanılarak 2007-2020 yılları arasındaki iş kazaları ve ölümler analiz edilmiştir. Metal sektörünün farklı alt faaliyet alanları incelenmiştir. Çalışmada, veriler iki ayrı döneme (2007-2013 ve 2014-2020) ayrılarak karşılaştırmalı olarak tarama deseni yöntemi ile incelenmiştir. İstatistiksel analizlerle, iş kazalarının ve ölümlü kazaların oranları hesaplanmış ve bu oranların azaltılabilmesi için çözüm önerileri sunulmuştur.

Sevim (2023), 2014-2021 yılları arasında Türkiye'de eğitim sektöründe meydana gelen iş kazaları incelenmiştir. SGK'dan alınan iş kazası istatistikleri, iş göremezlik süreleri ve ölümlü vaka sayıları ile eğitim sektöründeki iş kazalarının sebepleri ve şiddeti analiz edilmiştir. Veriler, kaza sıklık hızı, kaza ağırlık oranı, kaza olabilirlik oranı, mortalite hızı ve fatalite hızı gibi kriterlerde değerlendirilmiştir.

Tanış (2023), 2018–2021 yılları arasında metal sektöründeki bir firmada gerçekleşen iş kazalarına ilişkin veriler verileri kullanarak sıklık analizi ve ki-kare testi uygulayarak iş kazalarını etkileyen faktörler incelemiştir.

Ceylan ve Kaplan (2024), tarafından Türkiye'de 2012-2022 yılları arasında meydana gelen ölümlü iş kazaları, İşçi Sağlığı ve İş Güvenliği Meclisi verileri ile analiz edilmiştir. Veriler tanımlayıcı araştırma modeli ile değerlendirilmiş, hem kayıt dışı ölümlü kazalara dikkat çekilmiş hem de yasal düzenlemelere rağmen iş kazası sonucu ölüm vakalarının arttığı tespit edilmiştir. AB ülkeleri ile karşılaştırıldığında, Türkiye hem iş kazası sonucu ölüm vakalarının sayısı hem de ölümlü iş kazası görülme oranı açısından en kötü performansa sahip olduğu bulunmuştur.

Dündar ve diğerleri (2024), çalışmalarında kişi başına düşen yaralanmalı iş kazası oranı (kaza olabilirlik oranı) ve iş günü kaybı faktörlerini birlikte değerlendirilerek istatistiksel analiz gerçekleştirmiş ve en riskli sektörlerin belirlenmesini amaçlamışlardır. Yapılan değerlendirmeler sonucunda, kömür ve linyit madenciliği sektörünün en riskli sektör olduğu ve en yüksek oranda iş günü kaybına ve yaralanmalara neden olduğu tespit edilmiştir.

Bu bölümde incelenen yayınlar Çizelge 3.1’de özet halinde gösterilmektedir.

Çizelge 3.1. İş kazaları çalışmalarına ait literatür taraması

Yazar ve Yıl	Problem Tipi	Uygulanan Yöntem	Kullanılan Veriler	Sektör	Bölge
Demir (2009)	Maliyet Analizi	Anket	2007 yılı iş kazası verileri	Metal	Türkiye
Köseoğlu (2013)	İş kazası tahmini	Yapısal Eşitlik Modeli	1988- 2010 yılları arasındaki iş kazası verileri	Genel	Türkiye
Desticioğlu (2015)	Tahmin	Hareketli Ortalama, Ağırlıklı Hareketli Ortalama, Basit Üstel Düzeltme, Brown’ın Üstel Düzeltme Tekniği, Holt’un İki Parametrelili Üstel Düzeltme Tekniği, Basit Regresyon Analizi, Çoklu Regresyon Analizi	2001-2012 yılları arasındaki iş kazası verileri	Belirlenen dört sektör	Türkiye

Çizelge 3.1.(devam) İş kazaları çalışmalarına ait literatür taraması

Yazar ve Yıl	Problem Tipi	Uygulanan Yöntem	Kullanılan Veriler	Sektör	Bölge
Çavuş ve Taçgın (2016)	Sınıflandırma	İstatistiksel Analiz	2005-2009 yılları arasındaki iş kazası verileri	İnşaat	Türkiye
Tuna (2017)	İş kazası Analizi	Graf Teorisi	2013-2014 yılları arasındaki iş kazası verileri	Genel	Türkiye
Yağımlı ve İzci (2017)	Kaza Sayısı Tahmini	Uyarlanabilir Bulanık Zaman Serisi, En Küçük Kareler	2008-2015 yılları arasındaki iş kazası verileri	Metal	Türkiye
Bayraktar, Uyguçgil, ve Konuk (2018)	Analiz ve karşılaştırma	İstatistiksel Analiz	2002-2015 yılları arasındaki iş kazası verileri	Madencilik	Türkiye
Demirel (2018)	Avrupa ülkeleri ile Türkiye karşılaştırması	Kümeleme, Çok Boyutlu Ölçekleme Analizleri	2013-2016 yılları arasındaki iş kazası verileri	Genel	Türkiye
Memiş (2018)	Sektörel Analiz	Hiyerarşik Kümeleme Analizleri, Çok Boyutlu Ölçekleme Analizleri	2008-2016 yılları arasındaki iş kazası verileri	Genel	Türkiye
Çakmak (2019)	İş kazası sebep analizi	Veri Madenciliği	2016-2017 yılları arasındaki iş kazası verileri	Genel	Türkiye
Kölgeli (2019)	Sektörel analiz	İstatistiksel Analiz	2008-2019 yılları arasındaki iş kazası verileri	İnşaat	Türkiye
Özcan (2019)	İş Kazası Nedenlerinin önem derecelerini belirleme	Bulanık AHP, Bulanık TOPSIS	İş kazasına neden olan faktörlerin değerlendirilmesi	İnşaat	Türkiye
Öztürk (2019)	İş kazaları üzerindeki iklim faktörlerinin etkisini belirleme	Sıklık, Ki-Kare, Lojistik Regresyon	2015-2017 yılları arasındaki iş kazası verileri	İnşaat	Küçükçekmece / Türkiye
Ünal (2019)	Türkiye ve Avrupa ülkeleri karşılaştırması	Ki-Kare, Lojistik Regresyon	2008-2016 yılları arasındaki iş kazası verileri	Genel	Türkiye ve Avrupa
Seylan (2021)	İş kazası istatistiksel analizi	İstatistiksel Analiz	2010-2019 yılları arasındaki iş kazası verileri	Gıda	Türkiye
Şimşek (2021)	Sıralama	Çok Kriterli Karar Verme, Bütünleşik Garrett Sıralama Tekniği	2013-2019 yılları arasındaki iş kazası verileri	İnşaat	Türkiye

Çizelge 3.1.(devam) İş kazaları çalışmalarına ait literatür taraması

Yazar ve Yıl	Problem Tipi	Uygulanan Yöntem	Kullanılan Veriler	Sektör	Bölge
Şimşek Akın (2021)	İş kazası istatistiksel analizi	İstatistiksel Analiz	2008-2017 yılları arasındaki iş kazası verileri	Genel	Türkiye
Toptancı (2021)	İş kazası sonucu tahmini, risk analizi	Lojistik Regresyon, Firth, Ridge, LASSO, Elastik Net, Grup LASSO, Grup Elastik Net, Fine-Kinney Tabanlı Aralık Değerli Nötrosofik AHP, CoCoSo	2013-2017 yılları arasındaki iş kazası verileri	İnşaat	İç Anadolu Bölgesi / Türkiye
Demir (2022)	Risk Değerlendirmesi	5x5 Matris, Hata Türleri ve Etkileri Analizi	2012-2018 yılları arasındaki iş kazası verileri	Metal	Türkiye
Demiroğlu (2023)	İş kazası istatistiksel analizi	İstatistiksel Analiz	2002-2020 yılları arasındaki iş kazası verileri	Madencilik	Türkiye
Kosutova (2023)	İş kazası nedenlerinin istatistiksel analizi	İstatistiksel Analiz	2012-2021 yılları arasındaki iş kazası verileri	Ulaşım	Çekya
Pelit (2023)	İş kazası istatistiksel analizi	Tarama Deseni	2007-2020 yılları arasındaki iş kazası verileri	Metal	Türkiye
Sevim (2023)	İş kazası istatistiksel analizi	İstatistiksel Analiz	2014-2021 yılları arasındaki iş kazası verileri	Eğitim	Türkiye
Tanış (2023)	Kaza Faktörleri İncelemesi	Sıklık Analizi, Ki-Kare	2018-2021 yılları arasındaki iş kazası verileri	Metal	İstanbul / Türkiye
Ceylan ve Kaplan (2024)	Ölümlü iş kazaları değerlendirme	Tanımlayıcı Araştırma Modeli	2012-2022 yılları arasındaki iş kazası verileri	Genel	Türkiye ve Avrupa
Dündar, Bilim A. ve Bilim N., (2024)	Sektör Belirleme	İstatistiksel Analiz	2021 yılı iş kazası verileri	Genel	Türkiye

3.2. İş Kazaları Analizinde Makine Öğrenmesi Uygulamalarına İlişkin Literatür Taraması

Ceylan ve Avan (2012), Türkiye genelinde iş kazalarıyla ilgili tahmin yapmak amacıyla YSA kullanmışlardır. Çalışmada, sigortalı işçi sayısı, işyeri sayısı, iş kazası, ölüm ve sürekli iş göremezlik verileri model parametreleri olarak belirlenmiş ve bu değişkenlere ait 1970-

2010 yılları arasındaki veriler kullanılmıştır. Modelin eğitimi, ileri beslemeli geri yayılım algoritması ile gerçekleştirilmiş ve 1970-2004 verileriyle eğitilen model, 2005-2010 arası için test edilmiştir. Modelin tahmin sonuçları gerçek verilerle karşılaştırılmış ve başarılı sonuçlar elde edilmiştir. Ayrıca, model performansı ortalama mutlak yüzde hata, ortalama mutlak hata ve ortalama karesel hataların karekökü kriterleriyle değerlendirilmiştir. Bu modeller aracılığıyla, Türkiye’de 2025 yılına kadar olan dönemde iş kazası, ölüm ve sürekli iş göremezlik sayıları üç farklı senaryo altında tahmin edilmiştir.

Ayanoğlu ve Kurt (2019) metal sektöründe meydana gelen iş kazalarının analiz edilerek olası kaza şiddetlerinin tahmin edilmesini amaçlamıştır. Metal sektörüne ait iş kazalarından oluşan bir veri kümesi oluşturulmuş ve bu veri üzerinde çok değişkenli veri analizi ile değişken indirgeme işlemleri yapılmıştır. Ardından, veri madenciliği teknikleri kullanılarak farklı algoritmalar test edilmiş ve en yüksek doğruluk oranına sahip tahmin modelinin yapay sinir ağı olduğu belirlenmiştir. Son olarak, çift katmanlı ileri beslemeli yapay sinir ağı ile oluşturulan kaza tahmin modeli, örnek bir metal sektörü işyeri verisiyle test edilerek, işyerindeki olası kaza riskleri değerlendirilmiştir.

Kang ve Ryu (2019) çalışmasında Güney Kore inşaat sektöründeki 2008-2014 yılları arasındaki 9 796 adet iş kazası verisini kullanarak iş kazası türlerini ayrıntılı bir şekilde sınıflandırmayı ve tahmin etmeyi amaçlamışlardır. Bu amaçla rastgele orman algoritması kullanılarak inşaat sahalarındaki iş kazası türlerini sınıflandıran ve tahmin eden bir model geliştirilmiştir. Modelde, önemli değişkenler özellik önemi yöntemiyle belirlenmiş ve bu değişkenler ile kaza türleri arasındaki ilişkiler analiz edilmiştir. Rastgele orman modeli %71,3 doğruluk oranı ile başarılı sonuçlar vermiştir.

Sarkar ve diğerleri (2019b) araştırmalarında, Hindistan’daki bir çelik fabrikasından elde edilen 2010-2013 yılları arasındaki 1 500 adet iş kazası verileri kullanarak, kaza sonuçlarını (yaralanma, ramak kala ve maddi hasar) tahmin etmek amacıyla DVM ve YSA algoritmalarını kullanmıştır. Bu algoritmaların performansını artırmak için genetik algoritma (GA) ve parçacık sürü optimizasyonu (PSO) kullanılmıştır. Sonuçlar, PSO tabanlı DVM modelinin en yüksek doğruluk ve sağlamlığa ulaştığını göstermektedir. Ayrıca, PSO-DVM modeliyle entegre edilen C5.0 karar ağacı algoritması kullanılarak kaza nedenlerini açıklayıcı dokuz adet kural türetilmiştir. Bu sayede, tahmine ek olarak iş kazalarının temel nedenlerini açıklamaya yönelik anlamlı karar kuralları elde edilmiştir.

Sarkar, Raj, Vinay, Maiti ve Pratihara (2019a) Hindistan'daki bir çelik fabrikasından elde edilen 2010-2013 yılları arasındaki 1 964 adet iş kazası verileri kullanılmıştır. Bu doğrultuda, C5.0, sınıflandırma ve regresyon ağaçları ve rastgele orman algoritmaları kullanılarak bir tahmin modeli oluşturulmuştur. Bu modellerin performansını artırmak amacıyla PSO ve GA teknikleri uygulanmıştır. Deneysel sonuçlara göre, PSO ile optimize edilen rastgele orman modeli en yüksek doğruluk oranını sağlamıştır.

Choi, Gu, Chin ve Lee (2020), Güney Kore'de 2011-2016 yılları arasında inşaat sektöründe yaşanan iş kazası verilerini kullanarak, ölümcül kazaların tahminine yönelik bir makine öğrenmesi modeli geliştirmiştir. Veri setinde; kazazedenin yaşı, cinsiyeti, hizmet süresi, inşaat türü, işverenin büyüklüğü ve kaza tarihi gibi değişkenler yer alınmıştır. Lojistik regresyon, karar ağacı, rastgele orman ve uyarlamalı güçlendirme (adaptive boosting) gibi makine öğrenmesi algoritmaları uygulanmıştır. Elde edilen sonuçlara göre, en yüksek sınıflandırma başarısı %91,98 ROC eğrisi altındaki alan (Area Under the Receiver Operating Characteristic-AUROC) değeriyle rastgele orman algoritmasında görülmüş ve bu yöntemin mevsim, işveren ölçeği, yaş, hafta içi ve hizmet süresi gibi değişkenlerle ölümcül kaza olasılığını en iyi şekilde tahmin ettiği belirlenmiştir.

Koç, Ekmekçioğlu ve Gürgün (2021) yaptıkları çalışmada, inşaat işçilerinin iş kazalarının ardından işçilerin engellilik durumlarını tahmin etmeyi amaçlamışlardır. Türkiye'de 2012-2020 yılları arasındaki 47 938 iş kazası verisi kullanılarak, veri kodlaması, veri ölçeklendirme, boyut indirgeme ve veri yeniden örnekleme gibi çok aşamalı özellik mühendisliği süreçleri uygulanmıştır. Tahminler, dört farklı ağaç tabanlı makine öğrenmesi yöntemi (rastgele orman, XGBoost, AdaBoost, ekstra ağaçlar) ve hiperparametre ayarları için GA teknikleri kullanılmıştır. GA-XGBoost modeli, doğruluk açısından 0.8292 ve AUROC açısından 0.8120 ile en yüksek tahmin doğruluğunu sağlamıştır.

Zhu, Hu, Hou ve Li (2021), Çin inşaat sektöründe meydana gelen kazaların şiddetini tahmin etmek amacıyla makine öğrenmesi algoritmalarını kullanmıştır. Araştırmada, 2016-2017 yıllarındaki 1 326 adet iş kazası verisi kullanılmış, kazaların ciddiyetini etkileyebilecek 16 kritik faktör analiz edilmiş ve bu faktör kombinasyonlarının tahmin performansına etkisi değerlendirilmiştir. Çalışmada lojistik regresyon, karar ağacı, DVM, naive bayes, KNN, rastgele orman, çok katmanlı algılayıcı ve otomatik makine öğrenmesi olmak üzere sekiz farklı algoritma kullanılmıştır. Sonuçlara göre, ham veri seti üzerinden yapılan tahminlerde

naive bayes ve lojistik regresyon modelleri %78,3 F1 skoru ile en iyi performansı göstermiştir.

Fuentes ve diğerleri (2022), İspanya'daki temizlik sektöründe hizmet veren çalışanların maruz kaldığı aşırı zorlanmaya bağlı iş kazalarının iş göremezlik süresi üzerindeki etkilerini analiz etmeyi amaçlamıştır. Araştırmada, 2009-2019 yılları arasındaki 79 000'den fazla iş kazası verisi kullanılarak, Çok değişkenli adaptif regresyon eğrileri (Multivariate adaptive regression splines-MARS) ve DVM makine öğrenmesi teknikleriyle bir iş göremezlik süresi tahmin modeli geliştirilmiştir. Sonuçta DVM modeli MARS modeline göre daha iyi bir tahmin yeteneği göstermiştir.

Koç, Ekmekçioğlu ve Gürgün (2022) çalışmasında, inşaat sektöründe 2012-2020 yılları arasında Türkiye'de kaydedilen 393 160 iş kazası verisi, kısa, orta ve uzun vadeli kaza sayısını tahmin etmek için analiz edilmiştir. Çalışmada ayrık dalgacık dönüşümü ve farklı makine öğrenmesi yöntemlerini birleştirilerek analiz yapılmıştır. Hibrit dalgacık-YSA modeli en iyi performansı göstermiştir.

Özdemir Aydın (2022), yeraltı kömür madenciliği sektöründe meydana gelen iş kazalarının şiddetini tahmin etmeyi amaçlamıştır. Türkiye Taşkömürü Kurumu'na bağlı Amasra, Armutçuk, Karadon, Kozlu ve Üzülmaz bölgelerinde toplanan 8 406 kazaya ait veri kullanılmıştır. Boyut indirgeme yöntemi olarak temel bileşenler analizi uygulanmış, karar ağacı, DVM ve YSA olmak üzere üç farklı makine öğrenimi algoritmasıyla kaza şiddeti tahmin modelleri oluşturulmuştur. Sonuç olarak, %79,2 ile en yüksek doğruluk oranına sahip olan DVM algoritması, bu veri seti için en başarılı tahmin modeli olarak belirlenmiştir. Diğer modellerin doğruluk oranı %78,5 olarak bulunmuştur. Model çıktıları, kazaların büyük kısmının "hafif yaralanma" kategorisinde doğru sınıflandırıldığını göstermiştir.

Recal (2022) çalışmasında, Türkiye'deki iş kazalarını önlemek amacıyla büyük veri analizinden yararlanarak bir karar destek sistemi geliştirmiştir. 2013-2017 yılları arasında meydana gelen 187 910 iş kazası verisi, iki aşamalı bir veri madenciliği yöntemiyle analiz edilmiştir. İlk aşamada, makine öğrenmesi yöntemlerinden Stokastik Gradyan Artırma algoritması kullanılarak iş kazalarının şiddeti üzerinde etkili faktörler belirlenmiş; ikinci aşamada ise, bu faktörlerin oluşturduğu kaza zincirleri birliktelik kural çıkarımı yöntemiyle ortaya konulmuştur.

Alkaissy ve diğerkleri (2023), Avustralya inřaat sektöründe iř kazası sonucu ortaya çıkan yaralanma türlerini tahmin etmek amacıyla makine öğrenmesi modelleri kullanmıştır. Çalışma kapsamında, 2016-2021 yılları arasında meydana gelen 16 878 adet iř kazası kaydı incelenmiş ve XGboost, rastgele orman, DVM ve lojistik regresyon gibi çeşitli makine öğrenmesi algoritmaları uygulanmıştır. Modellerin performansı, doğruluk, hassasiyet, geri çağırma, F1 skoru, AUROC ve hassasiyet hatırlama eğrisi alanı gibi altı metrik ile değerlendirilmiştir. Rastgele orman algoritması en yüksek performansı göstermiştir.

Çamlıyurt, Y.Park, Kim, Kang ve S.Park (2023) çalışmasında, gemilere iniř-biniř sırasında deniz kılavuz kaptanlarının karşılaştığı iř kazalarını incelemiştir. Güney Kore'deki Busan Limanı'nda kaydedilen 500 raporu analiz edilerek, kılavuz kaptanların iř kazalarını tahmin etmek ve açıklamak için bir metodoloji geliřtirmeyi hedeflemişlerdir. Kılavuz kaptanlarının iř kazalarını tahmin etmek ve açıklamak için karar ağacı, rastgele orman ve açıklanabilir makine öğrenmesi yöntemlerini içeren bir metodoloji kullanarak bir deniz pilotu iř kazası tahmin modeli geliřtirmeyi amaçlamıştır. Kılavuz kaptan kazaları ile kaza dışı deęişken önemi ayırt etmek üzere eğitilmiş karar ağaçları ve rastgele orman ile 13 kaza öncesi faktör belirlenmiştir. Sonuçlar işyeri faktörlerinin çevresel faktörlerden daha belirleyici olduğunu göstermiştir. Deniz durumu, rüzgar kuvveti ve operatörün kontrolü dışındaki günün saati gibi deęişkenlerin önemsiz olduğu bulunmuştur.

Karakaya Özkan (2023), 2013-2018 yılları arasında metal sektöründe yaşanmış ulusal iř kazaları için kaza parametrelerini ve risk faktörlerini dikkate alarak istatistiksel bir değerlendirme yapmıştır. Ayrıca, ağır kaza olarak nitelendirilen ölümlü ve uzuv kayıplı iř kazası verilerini kullanarak makine öğrenmesi yöntemiyle bir tahmin algoritması geliřtirmiş ve algoritmaların performans ölçütlerini karşılaştırmıştır. Bu kapsamda, metal sektöründe gerçekleşmiş 1 542 adet ağır kaza ve 103 179 yaralanmalı olmak üzere toplam 104 721 tane iř kazası verisi analiz edilmiştir. İş kazası nedenlerini ve sonuçlarını tahmin etmek amacıyla makine öğrenmesi yöntemlerinden rastgele orman, KNN, CART, gradyan artırma makineleri algoritmaları kullanılmıştır. Rastgele orman (0,9172) algoritması, diğerk algoritmalarından daha iyi performans göstermiştir.

Koç, Ekmekcioğlu ve Gürgün (2023), Türkiye’de inřaat sektöründe yaşanan 76 128 adet iř kazası verisini kullanarak çalışan işçilerin kazalar sonucu en fazla zarar görebilecek vücut bölgelerinin, proaktif olarak tahmin edilmesini amaçlamıştır. Çalışmada makine

öğrenmesine dayalı çok aşamalı bir model önerilmiştir. Modelde, özellikle rastgele orman algoritması ile temel bileşenler analizi birleştirilerek en yüksek doğruluk elde edilmiştir. Analiz sonuçları, işçilerin çalıştığı gün sayısı, çalıştıkları şirketin büyüklüğü (işçi sayısı) ve yaş gibi değişkenlerin, vücut bölgesi yaralanmalarında en etkili faktörler olduğunu göstermektedir.

Luo, Li, Goh, Song ve Liu (2023), Çin’de 2013-2021 yılları arasında inşaat sektöründeki 264 adet çökme kazaları verileri kullanılarak iş kazalarında kaza ciddiyetinin tahmini için yapılandırılmamış metin verilerinden yararlanarak nitelik faktörlerine dayalı bir makine öğrenmesi modeli geliştirilmiştir. Rastgele orman algoritması ile geliştirilen model, üç farklı senaryo altında test edilmiş ve %82’ye varan doğruluk oranına ulaşarak yüksek performans göstermiştir. Ayrıca, özellik önem sıralaması analizleriyle kaza şiddetini etkileyen kritik değişkenler belirlenmiştir.

Mammadov ve diğerleri (2023), çalışmalarında boru hattı inşaatı projelerindeki iş kazalarının sonuçlarının tahmini için yapay zeka temelli bir model geliştirmeyi amaçlamıştır. Bu amaçla, Asya’dan Avrupa’ya uzanan ve 1 184 olay içeren bir veri seti kullanılmıştır. Kaza türü, zaman ve neden türü gibi değişkenlere göre giriş verisi alanları değiştirilerek 12 alt veri seti oluşturulmuş ve her biri 11 farklı makine öğrenmesi algoritmasıyla (YSA, derin öğrenme, hızlı geniş marj, karar ağaçları, rastgele orman, gradyanla güçlendirilmiş ağaç, doğrusal regresyon, genelleştirilmiş doğrusal model, naive bayes algoritması, lojistik regresyon ve DVM) test edilmiştir. Elde edilen bulgulara göre, düşük tehlike farkındalığı ve uygunsuz araç kullanımı en önemli ani nedenlerken, risk tanımında yetersizlik ve saha denetim eksikliği en kritik kök nedenler olarak öne çıkmıştır. Derin öğrenme algoritması, diğerlerine kıyasla sekiz alt veri kümesinde daha iyi performans sergilemiştir.

A.Zermane, Tohir, H.Zermane, Baharudin ve Yusoff (2023), çalışmasında Malezya inşaat sektöründe yükseklikten düşmeye bağlı ölümcül iş kazalarını etkileyen faktörlerin tespitine yönelik makine öğrenmesi tabanlı bir tahmin modeli geliştirilmiştir. Veriler, Malezya İş Sağlığı ve Güvenliği Departmanı’nın endüstriyel kaza veri tabanından alınmış ve toplam 3 321 kazaya ait senaryo incelenmiştir. Çalışmada yedi farklı makine öğrenmesi modeli test edilmiş ve rastgele orman modeli en iyi performansı göstermiştir.

Sanchez-Lite, Fuentes-Bargues, Iglesias ve Gonzalez-Gaya (2024), İspanya'da 2009–2021 yılları arasında işyerinde meydana gelen ölümcül kalp krizi vakalarını inceleyerek, bu tür kazaların oluşumuna etki eden faktörleri belirlemeyi amaçlamışlardır. 15 616 geçerli örnek kullanılarak, 13 değişken beş grupta sınıflandırılmış ve bu değişkenler naive bayes algoritması kullanılarak analiz edilmiştir. Dengesiz veri seti üzerinde iki aşamalı bir algoritma ile 13 farklı tahmin modeli oluşturulmuş, modellerin genel performans puanlarına göre en uygun olanı seçilmiştir. Çalışmada, en fazla ölümcül kalp krizi vakası, sağlık ve sosyal hizmetler, ulaşım ve depolama, imalat ve inşaat sektörlerinde görülmüştür. Ayrıca, kalp krizi ve ölümcül sonuçlarının erkeklerde ve özel sektör çalışanlarında daha yüksek olduğu tespit edilmiştir.

Sena ve diğerleri (2024), çalışmalarında perakende sektöründe kaza olup olmama durumunun sınıflandırılmasını amaçlamıştır. Bir perakende şirketine ait iki veri tabanı kullanılarak her mağaza ve iş birimi için bir etki puanı oluşturulmuş, ardından bu puanlar Ocak-Mayıs 2023 dönemindeki iş kazası gerçekleşip gerçekleşmeme durumu ile ilişkilendirilmiştir. Karar ağaçları, KNN, rastgele orman, lojistik regresyon, DVM olmak üzere beş farklı sınıflandırma algoritmasının uygulanmış, önleyici güvenlik eylemleri veri seti için en yakın DVM, en yüksek doğruluk ve kesinlik değerlerine ulaşırken; eylem planları veri setinde ise tüm performans ölçütlerinde en başarılı sonuçlar lojistik regresyon ile elde edilmiştir.

Tunçman (2024), tekstil sektöründe meydana gelen iş kazalarını analiz ederek, bu kazalara neden olan risk faktörlerini veri madenciliği yöntemleriyle belirlemeyi amaçlamıştır. Çalışmada, 2019-2021 yılları arasında yaşanan 89 963 iş kazasına ait veriler incelenmiş ve kazalara etki eden en önemli değişkenlerin kaza sebebi, kullanılan araç/materyal, alt sektör bilgisi ve firma büyüklüğü olduğu tespit edilmiştir. Yaralanma türlerini tahmin etmek için DVM, ekstra ağaçlar, rastgele orman, gradyan artırma ve XGBoost gibi farklı makine öğrenimi algoritmaları uygulanmıştır. En yüksek başarı, %70 makro F-skoru ile XGBoost algoritmasında elde edilmiştir. DVM (%69) ve ekstra ağaçlar (%68) da yüksek başarı gösteren diğer algoritmalar olmuştur.

Yiğit ve Kazar (2025), inşaat sektöründe özellikle bina bakım projelerinde meydana gelen iş kazalarının tahmin edilmesi üzerinde çalışmışlardır. Çalışmada 2016-2022 yılları arasında 1 807 kamu okulu bakım projesine ait veriler kullanılarak topolojik yaklaşıma dayalı yeni

bir özellik seçimi süreci geliştirilmiş ve bu sürecin iş kazası tahmin modellerine etkisi analiz edilmiştir. Yedi farklı makine öğrenmesi algoritması (DVM, rastgele orman, doğrusal regresyon, stokastik gradyan arttırma, XGBoost, doğrusal ayırma analizi, gaussian naive bayes) ile bunların birleşiminden oluşan hibrit bir model kullanılarak analizler gerçekleştirilmiştir. İş kazası olasılığını en çok etkileyen değişkenler ise şantiye teslimi, maliyet oranı, toplam proje süresi ve yüklenici büyüklüğü olarak belirlenmiştir.

Bu bölümde incelenen yayınlar Çizelge.3.2’de özet halinde gösterilmektedir.

Çizelge 3.2. İş kazaları analizinde makine öğrenmesi çalışmalarına ait literatür taraması

Yazar ve Yıl	Problem Türü	Uygulanan Yöntemler	Kullanılan Veriler	Sektör	Bölge
Ceylan ve Avan (2012)	Tahmin	YSA	1970-2010 yılları arasındaki iş kazası verileri	Genel	Türkiye
Ayanoglu ve Kurt (2019)	İş kazası şiddeti tahmini	YSA	-----	Metal	Türkiye
Kang ve Ryu (2019)	İş kazası türü tahmini	Rastgele Orman	2008-2014 yılları arasındaki iş kazası verileri	İnşaat	Güney Kore
Sarkar ve diğerleri (2019b)	İş kazası sonucu tahmini	DVM, YSA	2010-2013 yılları arasındaki iş kazası verileri	Metal	Hindistan
Sarkar ve diğerleri (2019a)	Kayma, takılma, düşme kazalarının tahmini	C5.0, Sınıflama ve Regresyon Ağaçları, Rastgele Orman	2010-2013 yılları arasındaki iş kazası verileri	Metal	Hindistan
Choi ve diğerleri (2020)	Ölümcül iş kazalarının tahmini	Lojistik Regresyon, Karar Ağacı, Rastgele Orman, AdaBoost	2011-2016 yılları arasındaki iş kazası verileri	İnşaat	Güney Kore
Koç ve diğerleri (2021)	İş kazalarının sonucu tahmini	Rastgele Orman, XGBoost, AdaBoost, Ekstra Ağaçlar.	2012-2020 yılları arasındaki iş kazası verileri	İnşaat	Türkiye
Zhu ve diğerleri (2021)	İş kazası şiddeti tahmini	Lojistik Regresyon, Karar Ağacı, DVM, Naive Bayes, KNN, Rastgele Orman, Çok Katmanlı Algılayıcı, Otomatik Makine Öğrenmesi	2016-2017 yılları arasındaki iş kazası verileri	İnşaat	Çin
Fuentes ve diğerleri (2022)	İş kazalarının sonuçlarının tahmini	Çok Değişkenli Adaptif Regresyon Eğrileri, DVM	2009-2019 yılları arasındaki iş kazası verileri	Temizlik	İspanya

Çizelge 3.2.(devam) İş kazaları analizinde makine öğrenmesi çalışmalarına ait literatür taraması

Yazar ve Yıl	Problem Türü	Uygulanan Yöntemler	Kullanılan Veriler	Sektör	Bölge
Koç ve diğerleri (2022)	İş kazası sayısı tahmini	Hibrit Dalgacık-YSA	2012-2020 yılları arasındaki iş kazası verileri	İnşaat	Türkiye
Özdemir Aydın (2022)	İş kazası şiddeti tahmini	Karar Ağacı, DVM, YSA	-----	Madencilik	Türkiye
Recal (2022)	İş kazası şiddeti tahmini	Stokastik Gradyan Arttırma Algoritması	2013-2017 yılları arasındaki iş kazası verileri	Genel	Türkiye
Alkaissy ve diğerleri (2023)	İş kazası sonucu yaralanma türleri tahmini	XGboost, Rastgele Orman, DVM, Lojistik Regresyon	2016-2021 yılları arasındaki iş kazası verileri	İnşaat	Avustralya
Çamlıyurt ve diğerleri (2023)	İş kazası tahmini	Karar Ağacı, Rastgele Orman, Açıklanabilir Makine Öğrenmesi	Güney Kore Busan Limanı'nda kaydedilen 500 iş kazası raporu	Denizcilik	Güney Kore
Karakaya Özkan, E. (2023)	Ağır iş kazası tahmini	Rastgele Orman, KNN, Sınıflama ve Regresyon Ağaçları, Gradyan Arttırma Makineleri	2013-2018 yılları arasındaki iş kazası verileri	Metal	Türkiye
Koc ve diğerleri (2023)	İş kazası sonucu tahmini	Rastgele Orman	-----	İnşaat	Türkiye
Luo ve diğerleri (2023)	İş kazası şiddeti tahmini	Rastgele Orman	2013-2021 yılları arasındaki iş kazası verileri	İnşaat	Çin
Mammadov ve diğerleri (2023)	İş kazası sonuç tahmini	YSA, Derin Öğrenme, Karar Ağaçları, Rastgele Orman, Gradyanla Güçlendirilmiş Ağaç, Doğrusal Regresyon, Genelleştirilmiş Doğrusal Model, Naive Bayes, Lojistik Regresyon, DVM	1 184 olay içeren bir veri seti	İnşaat	Asya ve Avrupa
Zermane ve diğerleri (2023)	İş kazası faktörleri tahmini	XGBoost, Karar Ağaçları, Rastgele Orman, Çok Katmanlı Algılayıcı, KNN, DVM, Lojistik Regresyon	2020 yılı iş kazası verileri	İnşaat	Malezya
Sanchez-Lite ve diğerleri (2024)	İş kazalarında kalp krizine eki eden faktörlerin tahmini	Naive Bayes	2009–2021 yılları arasındaki iş kazası verileri	Genel	İspanya

Çizelge 3.2.(devam) İş kazaları analizinde makine öğrenmesi çalışmalarına ait literatür taraması

Yazar ve Yıl	Problem Türü	Uygulanan Yöntemler	Kullanılan Veriler	Sektör	Bölge
Sena ve diğerleri (2024)	İş kazası olma sınıflandırması	Karar Ağaçları, KNN, Rastgele Orman, Lojistik Regresyon, DVM	2023 yılı iş kazası verileri	Perakende	-----
Tunçman (2024)	Yaralanma türleri tahmini	DVM, Ekstra Ağaçlar, Rastgele Orman, Gradyan Artırma, XGBoost	2019-2021 yılları arasındaki iş kazası verileri	Tekstil	Türkiye
Yiğit ve Kazar (2025)	İş kazası tahmini	DVM, Rastgele Orman, Doğrusal Regresyon, Stokastik Gradyan Arttırma, Xgboost, Doğrusal Diskiminant Analizi, Naive Bayes	2016-2022 yılları arasındaki iş kazası verileri	İnşaat	Türkiye



4. MATERYAL VE METODOLOJİ

Bu bölümde, bu çalışmada kullanılacak olan materyaller, geleneksel ve makine öğrenmesi yöntemleri incelenecek ve değerlendirilecektir. Uygulanacak yöntemler belirlenirken literatür araştırması sonucunda iş kazalarının analizinde en sık kullanılan yöntemler tespit edilmiş ve bu yöntemlere ek olarak, analizlerin daha geniş bir perspektiften değerlendirilmesi amacıyla farklı yaklaşım gruplarından yöntemler seçilmiştir.

4.1. İş Kazaları Analizinde Kullanılacak Geleneksel Tahmin Yöntemleri

Literatürde, iş kazası analizlerinde regresyon yöntemlerinin yaygın olarak kullanıldığı görülmektedir (Desticioğlu, 2015; Mammadov ve diğerleri, 2023; Öztürk, 2019; Toptancı, 2021; Ünal, 2019; Yiğit ve Kazar, 2025). Bu çalışmada iş kazası verileri üzerinde analiz yapmak amacıyla öncelikle çoklu doğrusal regresyon yöntemi kullanılmıştır. Çoklu doğrusal regresyon, bağımlı değişken ile birden fazla bağımsız değişken arasındaki ilişkiyi modellemede yaygın bir tekniktir (Alpar, 2010). Ancak çok sayıda değişken içeren veri setlerinde çoklu doğrusal bağlantı sorunu görülebilmekte ve modelin güvenilirliği azalabilmektedir. Bu nedenle, LASSO regresyonu kullanılmıştır. LASSO regresyonu bazı değişkenlerin etkisini sıfırlayarak daha sade ve yorumlanabilir modeller sunmaktadır (Vergili, 2023).

4.1.1. Regresyon analizi

Regresyon analizi, iki ya da daha fazla değişken arasındaki neden-sonuç ilişkisini incelemek ve bu ilişkiyi matematiksel olarak ifade etmek amacıyla kullanılan bir istatistiksel yöntemidir. Bu yöntem, bağımlı değişken ile onu etkileyen bağımsız değişkenler arasındaki ilişkinin yönünü ve gücünü belirlemeyi hedefler. Regresyon analizinde, incelenen değişkenler sürekli ya da kesikli olabileceğinden, veri yapısına bağlı olarak farklı regresyon modelleri tercih edilebilir. Bu analiz türü; ekonomi, mühendislik, sağlık bilimleri ve sosyal bilimler gibi çeşitli disiplinlerde yaygın olarak uygulanmaktadır (Akdoğan, 2024).

Çoklu doğrusal regresyon

Bir bağımlı değişken ile birden fazla bağımsız değişkenin yer aldığı modellerin incelenmesine çoklu regresyon analizi denir. Bu analiz türünde, bağımlı değişken ile bağımsız değişkenler arasındaki ilişki matematiksel bir model aracılığıyla ifade edilir ve bağımlı değişkenin değeri, bağımsız değişkenler kullanılarak tahmin edilmeye çalışılır. Çoklu regresyon analizinde temel amaç, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini açıklayabilmektir. Bu süreçte, aralarında yüksek korelasyon bulunan bağımsız değişkenlerin aynı modelde yer alması çoklu doğrusal bağlantı sorununa yol açabileceğinden, bu tür değişkenlerden yalnızca biri seçilmelidir. Bu seçimi yapabilmek için analiz öncesinde değişkenler arasındaki korelasyon matrisi incelenmelidir. Ayrıca, hangi bağımsız değişkenin bağımlı değişken üzerinde daha fazla etkili olduğunu belirlemek de analiz açısından önemlidir. Bu amaçla hem ikili korelasyonlara hem de değişkenin çoklu regresyon modeli içindeki etkisine bakılarak değerlendirme yapılır. Yüksek korelasyon değerleri, daha güçlü doğrusal ilişkiler olduğunu gösterir (Zaman, 2017).

Basit doğrusal regresyon modeline ilave edilen her bir yeni bağımsız değişken, modeli çoklu doğrusal regresyon modeline dönüştürür. Çoklu doğrusal regresyon denklemi, yapısal olarak basit doğrusal regresyon denklemine benzer olmakla birlikte, daha fazla değişken içermesiyle ondan ayrılır (Yüksel, 2023). Çoklu doğrusal regresyon denklemi Eş.4.1’de gösterilmektedir.

$$Y_t = b_1X_1 + b_2X_2 + \dots + b_nX_n + C \quad (4.1)$$

Burada Y bağımlı değişken, X_n bağımsız değişken, b_n regresyon katsayıları ve C kesişme sabiti olarak ifade edilmektedir.

Çoklu doğrusal regresyon modeli, en küçük kareler yöntemine dayanır. Bu yöntem, gözlenen değerler ile modelin tahmin ettiği değerler arasındaki farkların karelerinin toplamını en düşük seviyeye indirmeyi amaçlayarak uygulanır (Yüksel, 2023).

LASSO regresyonu

LASSO regresyon yönteminde çoklu doğrusal regresyona ek olarak Eş. 4.2’de verilen ceza fonksiyonu kullanılır.

$$L = \sum_{i=1}^n (y_i - \hat{Y}_i)^2 + \lambda \sum_{j=1}^n |\beta_j| \quad (4.2)$$

LASSO regresyon yöntemi, ceza fonksiyonunu kullanarak katsayıları sıfıra indirgeme yeteneği sayesinde daha az değişken içeren, daha basit ve kolayca yorumlanabilen modeller oluşturulmasına olanak tanır. Bu özellik, özellikle çok sayıda gözlem ve değişken içeren büyük veri setlerinde LASSO regresyonunun önemli bir avantaj sunmasını sağlamaktadır (Vergili, 2023).

4.2. İş Kazaları Analizinde Kullanılacak Makine Öğrenmesi Temelli Tahmin Yöntemleri

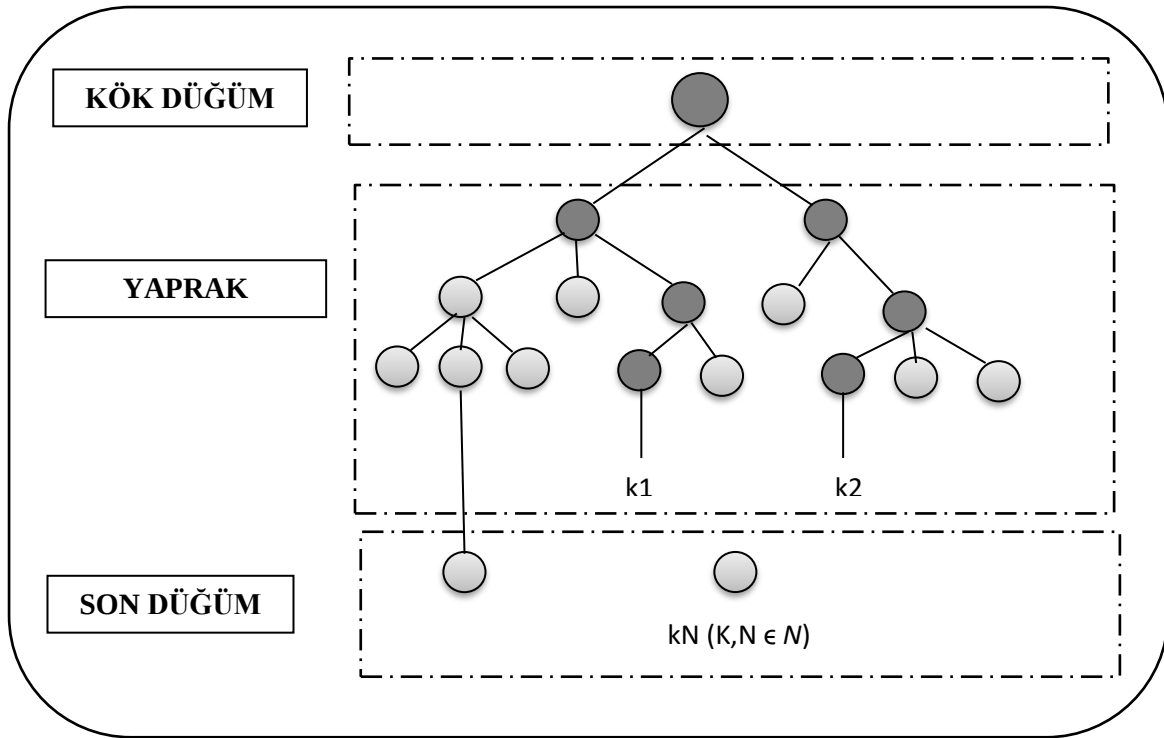
Bu çalışmada iş kazalarının analizi için rastgele orman (Luo ve diğerleri, 2023; Sena ve diğerleri, 2024; Tunçman, 2024; Yiğit ve Kazar, 2025; Zermane ve diğerleri, 2023), KNN (Karakaya Özkan, 2023; Sena ve diğerleri, 2024; Zermane ve diğerleri, 2023; Zhu ve diğerleri, 2021) ve DVM (Alkaissy ve diğerleri, 2023; Mammadov ve diğerleri, 2023; Tunçman, 2024; Yiğit ve Kazar, 2025; Zermane ve diğerleri, 2023) yöntemleri kullanılmıştır. Rastgele orman yöntemi, yüksek tahmin doğruluğu ve yorumlanabilirliği ile öne çıkan bir makine öğrenmesi tekniğidir. Yeni verilerin modele kolayca eklenebilmesi ve değişken önem düzeylerinin hesaplanabilmesi sayesinde, özellikle çok sayıda değişken içeren veri setlerinde daha sade ve etkili tahmin modelleri oluşturulmasına olanak tanır (Görgülü, 2021). Uygulama açısından oldukça sade bir yapıya sahip olması, k-en yakın komşu yönteminin tahmin ve sınıflandırma problemlerinde yaygın olarak tercih edilmesine neden olmaktadır (Alkış, 2017). DVM, yüksek doğruluk oranı, çok sayıda değişkenle çalışabilme yeteneği ve karmaşık problemleri optimize ederek çözme kapasitesiyle öne çıkan bir yöntemdir. Büyük veri setlerinde işlem yükünü azaltarak hızlı ve verimli sonuçlar sunar (Ercan, 2020).

4.2.1. Rastgele orman

Rastgele orman algoritması hem sınıflandırma hem de regresyon problemlerinde kullanılan bir makine öğrenmesi yöntemidir. Bu yöntem, 2001 yılında Breiman tarafından geliştirilmiştir. Temel olarak rastgele orman, birbirinden bağımsız ve farklı veri örnekleriyle oluşturulmuş çok sayıda derin karar ağacının bir araya gelmesiyle oluşur. Algoritmada karar ağaçlarının düğüm ve dalları bulunur. Her düğümde, veri kümesindeki tüm öznitelikler göz önünde bulundurularak en iyi ayrımı yapmak yerine, rastgele seçilen bir öznitelik alt kümesi içinden en uygun olanı belirlenerek veri dallara ayrılır (Kavuk Salış, 2024).

Rastgele orman algoritması, dört temel aşamadan oluşan bir makine öğrenmesi yöntemidir: rastgele alt örneklem oluşturma, rastgele öznitelik seçimi, karar ağaçlarının inşası ve tahmin süreci. İlk aşamada, eğitim verilerinden rastgele yöntemle seçilen alt örneklem kullanılarak her bir karar ağacı için farklı veri alt kümeleri elde edilir. İkinci aşamada, her düğümde tüm öznitelikler yerine, rastgele belirlenen bir öznitelik alt kümesi değerlendirilerek karar ağacının dallanması sağlanır. Üçüncü aşamada, bu örneklem ve seçilen öznitelikler kullanılarak bağımsız karar ağaçları inşa edilir. Son olarak, tüm ağaçlardan elde edilen tahminler birleştirilerek, sınıflandırma problemlerinde çoğunluk oyu, regresyon problemlerinde ise ortalama değer alınmak suretiyle nihai tahmin gerçekleştirilir (Kavuk Salış, 2024).

Şekil 4.1.'de örnek karar ağaçlarının yapısal düzeni sunulmaktadır. Karar ağacının kök düğümünden itibaren, verilerin sınıflandırılmasına yönelik sorular sorularak ağaç yapısı inşa edilmeye başlanır. Bu süreç, dallanma işleminin durduğu ve nihai sınıfların belirlendiği yaprak düğümlere ulaşılan kadar devam eder (Zorlu, 2022).



Şekil 4.1. Rastgele orman örnek yapısı

Rastgele orman yöntemi, makine öğrenmesi teknikleri arasında yüksek tahmin doğruluğu sunması, modelin yorumlanabilirliğini artırması ve çeşitli optimizasyon olanakları sağlaması nedeniyle sıkça tercih edilmektedir. Bu yöntem, modele sonradan yeni verilerin eklenmesine olanak tanıyarak güvenilirlik düzeyi yüksek tahminler yapılmasına imkân verir. Ayrıca, veri setinde yer alan değişkenlerin önem derecelerinin hesaplanması da mümkündür. Bu sayede, özellikle çok sayıda değişken içeren veri setlerinde, bağımlı değişken üzerindeki etkisi en belirgin olan değişkenler belirlenerek model sadeleştirilebilir. Böylece, daha etkin ve açıklayıcı tahmin modelleri geliştirilebilmektedir (Görgülü, 2021).

Rastgele orman yönteminde, eğer modelin doğruluğunu test etmek amacıyla ayrı bir test veri seti mevcut değilse ya da orijinal veri setinden test seti daha önce ayrılmamışsa, veri seti sınıf dağılımı dikkate alınarak rastgele bir şekilde ikiye bölünür. Bu durumda, verilerin yaklaşık üçte ikisi modelin eğitilmesi için öğrenme veri seti olarak kullanılırken, kalan üçte birlik kısmı test veri seti olarak değerlendirilir. Öte yandan, eğer önceden ayrılmış bir test veri seti mevcutsa veya orijinal veri seti eğitim ve test olarak ayrılmışsa, eğitim seti kendi içerisinde benzer şekilde üçte ikisi öğrenme, üçte biri ise test verisi olacak şekilde bölünerek model oluşturulmaktadır (Akman, 2010).

Rastgele orman algoritmasının kaba kodu şu şekilde gösterilebilmektedir.

1. Toplam öznitelik (özellik) sayısı p olan bir veri kümesinden, n adet bootstrap yöntemiyle örneklem alınır. Her bir örneklemin yaklaşık $2/3$ 'ü modelin eğitiminde kullanılır.
2. Eğitim veri seti içinden, rastgele m adet öznitelik seçilir. Bu m değişken arasından, veri kümesini en iyi şekilde bölecek olan değişken belirlenir.
 - a. Regresyon problemi için: $m = p / 3$
 - b. Sınıflandırma problemi için: $m = \sqrt{p}$
3. Seçilen en uygun özniteliğe göre, bölünme noktası belirlenir. Sınıflandırma problemi için bu karar, Gini indeksi gibi dallanma ölçütleriyle yapılır ve düğüm alt kümelerine ayrılır.
4. İkinci ve üçüncü adımlar, her bir ağaçta terminal (yaprak) düğüme ulaşıncaya kadar tekrar edilir.
5. Bu işlemler, toplam n adet karar ağacı oluşturuluncaya kadar tekrarlanır.
6. Eğitilmiş bu rastgele ağaçlar kullanılarak test veri kümesi modele verilir ve her ağaç, tahminde bulunur.
7. Nihai karar şu şekilde verilir:
 - a. Sınıflandırma problemleri için: Her ağaçtan gelen tahminler arasında en fazla oyu alan sınıf, tüm ormanın tahmini olarak seçilir.
 - b. Regresyon problemleri için: Ağaçlardan elde edilen tahminlerin aritmetik ortalaması, modelin çıktısı olarak belirlenir.

Modelin hata oranını tahmin etmek amacıyla öğrenme verisi şu şekilde değerlendirilir:

1. Ana veri kümesinden bootstrap yöntemiyle oluşturulan örneklem, eğitim için kullanılan ve dışta kalan veri olarak iki gruba ayrılır.
2. Eğitim verileri kullanılarak karar ağacı eğitilir.
3. Bu ağaç, modelin eğitiminde kullanılmayan dışta kalan veri ile test edilir ve her ağacın tahmin performansına dayalı olarak bir hata oranı hesaplanır.
4. Tüm ağaçlar için yapılan bu testlerin ardından, bireysel hata oranları birleştirilerek rastgele ormanın genel hata oranı tahmin edilir (Bayraktar, 2019).

4.2.2. K-en yakın komşu

KNN algoritması, veri madenciliğinde özellikle sınıflandırma amaçlı kullanılan ve benzerlik ölçümlerine dayalı çalışan bir yöntemdir. Buradaki “ k ” ifadesi, dikkate alınacak komşu sayısını belirtir. K-en yakın komşu algoritmasının temel amacı, bir örneğin en yakın k komşusunu tespit etmektir. Sınıflandırma sürecinde, veri kümesindeki her bir örneğin diğer örneklerle olan mesafesi hesaplanır ve yalnızca en yakın k adet örnek değerlendirmeye alınır. Bu k komşu, üzerinde işlem yapılan örneğe uzaklık açısından en yakın olanlardır. Algoritma, sınıfı bilinmeyen bir örneği, k komşusu arasında en çok tekrar eden sınıf etiketine göre

sınıflandırır. Uygulama açısından oldukça sade bir yapıya sahip olması, k -en yakın komşu yönteminin sınıflandırma problemlerinde yaygın olarak tercih edilmesine neden olmaktadır (Alkış, 2017).

KNN algoritması, yalnızca sınıflandırma problemlerinde değil, aynı zamanda regresyon problemlerinin çözümünde de kullanılan bir makine öğrenmesi yöntemidir. Özellikle bağımsız değişkenlerin sayısal değerlere sahip olduğu durumlarda tercih edilir ve sınıflandırma işlemi, gözlemler arasındaki mesafelerin hesaplanmasına dayanarak gerçekleştirilir (Deveci, 2020).

KNN algoritmasının işleyiş süreci aşağıdaki adımlarla özetlenebilir (Deveci, 2020).

Adım 1: Öncelikle komşu sayısını belirleyen k değeri tanımlanır.

Adım 2: Hedef veri (örnek) ile veri kümesindeki diğer tüm örnekler arasındaki mesafeler hesaplanır.

Adım 3: Hesaplanan mesafeler küçükten büyüğe doğru sıralanır ve bu sıralamaya göre komşular belirlenir. En küçük mesafeye sahip veri 1. komşu olarak, en büyük mesafeye sahip olan ise toplam veri sayısına göre n . komşu olarak kabul edilir.

Adım 4: Belirlenen k adet en yakın komşunun sınıfları (kategorileri veya durumları) dikkate alınır.

Adım 5: Bu komşular arasından en sık gözlemlenen sınıf seçilerek, hedef veriye bu sınıf atanır.

KNN algoritmasında karşılaşılan iki temel sorun bulunmaktadır. Bunlardan ilki, uygun k değerinin belirlenmesidir. K , her test örneği için kaç komşunun dikkate alınacağını tanımlamaktadır. Bu nedenle, doğru k değerinin seçilmesi, modelin doğruluğu açısından oldukça kritiktir. Çok yüksek bir k değeri, küçük ancak önemli veri desenlerinin gözden kaçmasına yol açabilir.

İkinci önemli sorun ise, test verisi ile onun potansiyel komşuları arasındaki mesafenin nasıl ölçüleceğidir. Bu amaçla en yaygın kullanılan mesafe ölçütleri Öklid, Minkowski ve Manhattan uzaklıklarıdır. Ancak bu yöntemler yalnızca sürekli veri türleri üzerinde etkili sonuçlar verir (Acet, 2022).

Minkowski uzaklığında yer alan p parametresi, uzaklık hesaplamalarında üs olarak kullanılan değeri belirtmektedir. p değerinin bir olarak seçilmesi durumunda, Eş. 4.3'teki Minkowski formülü Manhattan uzaklığını verirken, p değeri iki seçildiğinde ise Öklid uzaklığı ölçümüne karşılık gelmektedir (Acet, 2022).

$$L_p(X, Y) = \sqrt[p]{\sum_{i=1}^n (|x_i - y_i|)^p} \quad (4.3)$$

Manhattan uzaklık ölçütü, literatürde L1 normu, şehir bloğu mesafesi veya doğrusal uzaklık gibi farklı terimlerle de anılmaktadır. Manhattan uzaklığı, iki vektör arasındaki karşılık gelen bileşenlerin mutlak farklarının toplamı olarak tanımlanır. Bu kavramsal tanım, Eş.4.4 ile ifade edilmektedir (Acet, 2022).

$$L_1(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (4.4)$$

Öklid uzaklığı, literatürde sıklıkla L2 normu ya da cetvel uzaklığı olarak adlandırılmaktadır. Bu uzaklık türü, iki vektörün karşılık gelen bileşenleri arasındaki farkların karelerinin toplamının karekökü alınarak hesaplanır. Bu hesaplama yöntemi, Eş.4.5'te gösterilmektedir (Acet, 2022).

$$L_2(X, Y) = \sqrt{\sum_{i=1}^n (|x_i - y_i|)^2} \quad (4.5)$$

Bir veri setinde girdi değişkenleri yükseklik ve genişlik gibi birbirleriyle yüksek oranda ilişkili ise, Öklid uzaklığı ölçütü, buna karşılık, yaş, boy ve cinsiyet gibi aralarında anlamlı ilişki bulunmayan değişkenler söz konusu olduğunda, Manhattan uzaklığı daha uygun bir uzaklık ölçüsü olarak önerilmektedir (Acet, 2022).

4.2.3. Destek vektör makineleri

DVM, yapısal risk minimizasyonu ilkesine dayalı olarak çalışan bir sınıflandırma yöntemidir. 1995 yılında Cortes ve Vapnik tarafından, çeşitli problemlerin çözümüne katkı sağlamak ve bilinmeyen durumları daha anlaşılır hale getirmek amacıyla geliştirilmiştir. DVM'nin teorik temeli, istatistiksel öğrenme kuramı olarak da bilinen Vapnik-Chervonenkis teorisine dayanmaktadır (Akdoğan, 2024).

DVM, temel olarak girdi ve çıktı değişkenleri arasındaki ilişkiyi öğrenerek bir tahmin fonksiyonu oluşturmayı amaçlayan bir modeldir. DVM, modelin doğru formunun önceden bilinmediği durumlarda, potansiyel modeller arasından en uygun tahminleyiciyi seçme prensibine dayanır. Başka bir deyişle, iki sınıfı birbirinden en iyi şekilde ayıran hiperdüzlemin bulunması esasına göre çalışır. Sınıflar arasına birçok farklı hiperdüzlem yerleştirilebilse de DVM yalnızca genelleme yeteneği en yüksek olan, yani doğruluk ve ayırım gücü açısından en optimal olan ayırıcı hiperdüzlemi seçer (Adar, 2023).

DVM birçok avantajı nedeniyle makine öğrenmesi alanında sıkça tercih edilen yöntemlerden biridir. Yüksek doğruluk oranına sahip olması, çok sayıda bağımsız değişkenle çalışabilme yeteneği ve özellikle karmaşık sınıflandırma problemlerini kareli optimizasyon problemleri şeklinde formüle ederek çözebilmesi, en önemli üstünlükleri arasında yer almaktadır. Bu dönüşüm sayesinde, problem çözüm sürecindeki işlem yükü azalmakta ve DVM, diğer yöntemlere kıyasla daha hızlı sonuçlar üretebilmektedir. Bu nedenle, özellikle büyük ölçekli veri setleriyle çalışırken önemli ölçüde avantaj sağlar. Ayrıca, sınıflandırma problemlerinde hesaplama verimliliği ve kullanım kolaylığı açısından da pek çok diğer makine öğrenmesi algoritmasına göre daha başarılıdır (Ercan, 2020).

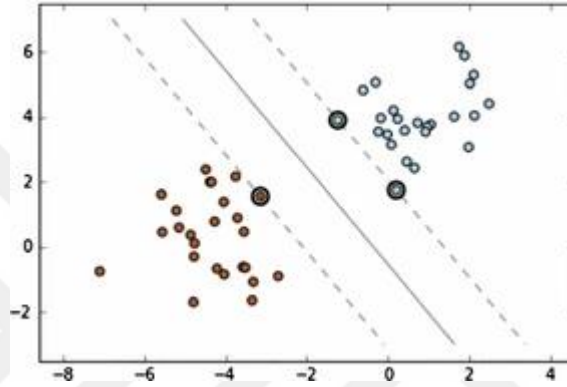
Veri seti iki sınıfa ayrılabiliriyorsa ve bu sınıflar doğrusal olarak ayrılabiliriyorsa, doğrusal DVM yöntemi tercih edilmektedir. Ancak, sınıflar doğrusal bir şekilde ayrılabilir değilse, bu durumda çok sınıflı DVM yaklaşımları kullanılmaktadır (Akdoğan, 2024).

DVM, temelini İstatistiksel Öğrenme Teorisi'nden almaktadır. Başlangıçta bu teori, yalnızca belirli bir veri kümesine uygun fonksiyonların tahminine yönelik teorik analizler amacıyla geliştirilmişti. Ancak zamanla, bu teoriye dayanarak geliştirilen DVM algoritması sayesinde, yalnızca teorik çerçeveler çizmekle kalmayıp, aynı zamanda yeni öğrenme algoritmalarının geliştirilmesinde de etkin bir şekilde kullanılabileceği anlaşılmıştır (Vapnik, 1999: 988).

Bu bölümde iki sınıflı DVM ele alınacak; sırasıyla doğrusal ayrılabilen ve doğrusal ayrılamayan destek vektör makineleri türleri incelenecektir.

Doğrusal ayrılabilen DVM

DVM'nin en basit modelidir ve yalnızca doğrusal olarak ayrılabilen problemleri çözebilen destek vektör türüdür. Bu model, sadece doğrusal ayrılabilir verilere uygulanabildiği için günümüzde yaygın olarak tercih edilmemektedir. Ancak, destek vektör makinesi yapısının temel mantığını kavrayabilmek adına öğretici bir örnek olarak değerlendirilmektedir. Bu yöntemde temel amaç, iki farklı sınıfa ait verileri bir hiper düzlem aracılığıyla sorunsuz bir biçimde birbirinden ayırmaktır (Çapkurt, 2023).



Şekil 4.2. Doğrusal olarak ayrılabilen DVM (Chauhan, Dahiya ve Sharma, 2018)

Doğrusal olarak ayrılabilen veriler için teorik olarak sonsuz sayıda hiper düzlem çizilebilmektedir. Ancak sınıflandırma sürecinde esas hedef, her iki sınıfa ait veri noktalarına eşit uzaklıkta yer alan en uygun hiper düzlemin belirlenmesidir (Çapkurt, 2023).

N adet örnek içeren bir veri kümesi için $X = \{x_i, y_i\}$, $i = 1, 2, 3, \dots, N$ biçiminde bir gösterim varsayalım. Burada $y_i \in \{-1, 1\}$ değerleri ve $x_i \in R$ ise özellik vektörlerini ifade eder. Eğer bu iki veri doğrusal olarak ayrılabilirse, doğrudan hiper düzlem ile ayrılacaktır. Bu durumda, destek vektör makinelerinin temel amacı, söz konusu hiper düzlemin her iki sınıfa eşit uzaklıkta konumlandırılmasını sağlamaktır. Destek vektör makineleri yöntemi, doğrusal olarak ayrılması mümkün olmayan verileri önce daha yüksek boyutlu bir özellik uzayına dönüştürür. Bu yeni uzay, orijinal veri kümesinin boyutundan daha geniştir. Bu dönüşümün temel amacı, verilerin bu yeni uzayda doğrusal bir hiper düzlem aracılığıyla ayrılabilmesini sağlamaktır. Hiper düzlem üzerindeki bir x noktası için, w hiper düzlemin normali ve $|b|/|w|$ ifadesi de bu düzlemin orijine olan dik uzaklığını göstermek üzere, $w \cdot x + b = 0$ koşulu sağlanmalıdır. Burada w , düzlemin normal vektörünü; b , yanlılık terimini; x ise düzlem üzerindeki herhangi bir noktayı temsil etmektedir. Ayrıca, $w \cdot x$ ifadesi, vektörler arasındaki

iç çarpımı göstermektedir. DVM, iki sınıfı en iyi şekilde ayıran hiper düzlemi bulabilmek için, kareli optimizasyon problemlerinin çözümünde Lagrange çarpanlarından faydalanır. Lagrange çarpanları sayesinde, orijinal minimizasyon problemi, çözümü daha kolay olan dual bir probleme dönüştürülür (Akdoğan, 2024).

Doğrusal olmayan DVM

Gerçek dünyadaki pek çok problemde, verilerin doğrusal bir şekilde ayrılması oldukça nadir karşılaşılan bir durumdur. Bu nedenle, doğrusal olarak ayrılabilen DVM'ler birçok karmaşık problemi çözmekte yetersiz kalabilmektedir. Bu tür durumlarla başa çıkmak için DVM, doğrusal olmayan bir haritalama fonksiyonu kullanarak verileri n boyutlu orijinal giriş uzayından daha yüksek boyutlu bir özellik uzayına aktarır. Bu yeni uzayda, veriler arasındaki yapısal farklar daha belirgin hale gelir ve DVM, doğrusal bir ayırıcı hiper düzlem oluşturarak en geniş marjini elde etmeye çalışır. Elde edilen bu doğrusal ayırım kuralı, yüksek boyutlu uzayda doğrusal olsa da, orijinal giriş uzayına geri yansıtıldığında doğrusal olmayan bir yapı kazanır. Böylece, doğrusal olmayan DVM modelleri, yüksek boyutlu uzayda doğrusal bir yöntem gibi çalışarak, karmaşık sınıflandırma problemlerini etkili bir şekilde çözebilmektedir (Akdoğan, 2024).

Destek vektör makinelerinin çekirdek fonksiyonları

DVM'de yaygın olarak kullanılan dört çekirdek fonksiyonu vardır. Bu fonksiyonlar; doğrusal fonksiyon, polinomiyal fonksiyon, sigmoid fonksiyon ve döngüsel temelli fonksiyondur.

Doğrusal çekirdek fonksiyonu

Çekirdek fonksiyon türlerinden en basit olanıdır ve (x,y) iç çarpımı ile ifade edilmektedir. Bu fonksiyon, daha fazla özelliğe sahip verilerde diğer fonksiyonlara göre daha iyi sonuçlar verebilir (Akdoğan, 2024).

$$K(x, y) = x \cdot y \quad (4.6)$$

Polinomiyal çekirdek fonksiyonu

Polinomiyal çekirdek fonksiyonu, diğer çekirdek fonksiyonlarına kıyasla daha fazla parametre içermesine rağmen hem hesaplama açısından daha basit hem de uygulaması kolay bir yapıya sahiptir. Bu çekirdek türü Eş.4.7’de verilen formülle ifade edilir:

$$K(x, y) = [(x \cdot y) + c]^d \quad (4.7)$$

Sigmoid çekirdek fonksiyonu

Sigmoid çekirdek fonksiyonunun formülü Eş.4.8’de gösterilmektedir. Burada c sabit terim, α eğimi, (x,y) ise iç çarpanı ifade etmektedir (Akdoğan, 2024).

$$K(x, y) = \tan(\alpha x \cdot y + c) \quad (4.8)$$

Döngüsel temelli çekirdek fonksiyonu

Döngüsel temelli çekirdek fonksiyonu, sınıflandırma problemlerinde en sık kullanılan çekirdek fonksiyonlarından biridir. DVM, sınıflandırma veya tahmin süreçlerinde döngüsel temelli çekirdek fonksiyonunu kullanarak veri kümesini çeşitli alt bölgelere ayırabilir. Döngüsel temelli çekirdek fonksiyonu çekirdeğinin yayılma parametresi küçük bir değere sahipse, yalnızca birbirine yakın olan veri örnekleri üzerinde etkili olur. Buna karşılık, veri noktalarının birbirinden uzak olduğu durumlarda, çekirdek fonksiyonu bu noktalar arasındaki etkileşimi kademeli olarak sıfıra indirger (Akdoğan, 2024).

4.3. Tahmin Doğruluğunun Analizi

Tahmin hatalarını hesaplamak için çeşitli yöntemler kullanılmaktadır. Tahmin edilen değer ile gerçekleşen değer arasındaki fark, bu hataların temelini oluşturur ve kullanılan tahmin yöntemlerinin doğruluğunun değerlendirilmesine olanak tanır. Bu hataların nedenleri genel olarak iki gruba ayrılabilir: rastlantısal ve rastlantısal olmayan etkenler. Rastlantısal olmayan hatalara örnek olarak; modelde yer alan değişkenlerin eksik veya yanlış seçilmesi, değişkenler arasındaki ilişkilerin hatalı kurulması, yanlış eğilim doğrularının kullanılması ya da mevsimsel etkilerin dikkate alınmaması gibi durumlar gösterilebilir. Tahmin

hatalarının hesaplanmasında farklı yöntemler kullanılmakta olup, her bir yöntemin kendine özgü güçlü ve zayıf yönleri vardır (Desticioğlu, 2015).

T dönemine ilişkin tahmin hatasının hesaplanmasında Eş.4.9'de gösterilen denklemden yararlanılmaktadır:

$$e_t = Y_t - F_t \quad (4.9)$$

Burada, e_t , t dönemine ait tahmin hatasını; Y_t , ilgili dönemde gerçekleşen gerçek talep miktarını; F_t ise aynı dönem için yapılan tahmin değerini ifade etmektedir.

Zaman serisi analizlerinde n dönemlik bir tahmin yapılmışsa, her bir döneme ilişkin tahmin hatalarının ayrı ayrı hesaplanması gerekmektedir. Zaman serilerine yönelik bu hataların ölçülmesinde çeşitli istatistiksel yöntemlerden yararlanılmaktadır. Söz konusu yöntemler arasında; ortalama hata (Mean Error-ME), ortalama mutlak sapma (Mean Absolute Deviation-MAD), ortalama kare hata (Mean Squared Error-MSE) ve ortalama mutlak yüzde hata (Mean Absolute Percentage Error-MAPE) ve belirleme katsayısı (R^2) gibi yaygın performans ölçütleri öne çıkmaktadır (Desticioğlu, 2015). ME, MAD, MSE, MAPE ve belirleme katsayısı formülleri Eş. 4.10, 4.11, 4.12, 4.13 ve 4.14'de gösterilmektedir.

$$ME = \frac{1}{n} \sum_{t=1}^n e_t \quad (4.10)$$

$$MAD = \frac{\sum_{t=1}^n |e_t|}{n} \quad (4.11)$$

$$MSE = \frac{\sum_{t=1}^n (e_t)^2}{n} \quad (4.12)$$

$$MAPE = 100 \times \frac{\sum_{t=1}^n |e_t| / Y_t}{n} \quad (4.13)$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n (P_i - A_i)^2}{\sum_{i=1}^n A_i^2} \right) \quad (4.14)$$

Yukarıda eşitlikleri verilen tahmin hatası yöntemleriyle elde edilen değerlerin sıfıra yakın olması, o yöntemin diğerlerine kıyasla gerçek sonuca daha yakın olduğunu gösterir. Ayrıca, birden fazla yöntemin karşılaştırıldığı durumlarda, en iyi yöntem, diğerlerine göre daha düşük bir tahmin hatası sonucu sağlamaktadır (Desticioğlu, 2015). R^2 yöntemi ise bağımsız değişkenlerin bağımlı değişkeni ne ölçüde açıkladığını gösterir ve 0-1 arasında değişen bir değere sahiptir. Bu değer bire ne kadar yakınsa model o kadar güçlüdür (Kutner, Nachtsheim, Neter ve Li, 2005).

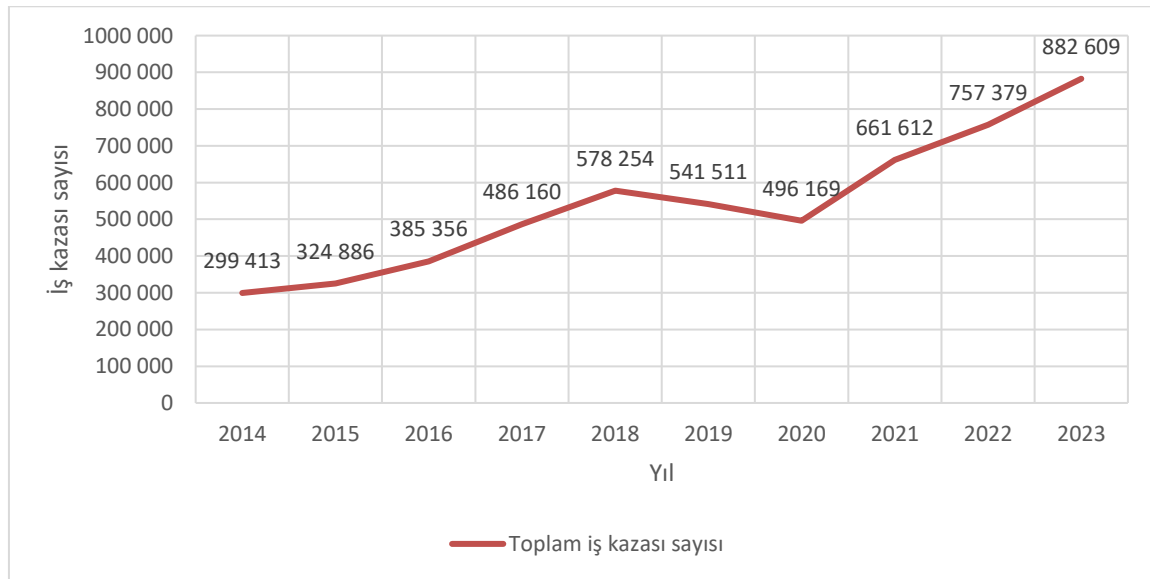


5. UYGULAMA

Bu tez çalışmasında, Türkiye’de meydana gelen iş kazalarının madencilik, metal ve inşaat sektörleri bazında analizi gerçekleştirilmiştir. Bu sektörler; yüksek fiziksel risk içeren faaliyetler, değişken ve zorlu çalışma koşulları, iş gücünün yapısal özellikleri ile birlikte yetersiz İSG önlemleri gibi faktörler nedeniyle iş kazalarına en açık alanlar arasında yer almaktadır. Bu sektörlerdeki 2014-2023 yılları arasındaki iş kazası verileri kullanılarak geleneksel tahmin yöntemlerinden çoklu doğrusal regresyon ve LASSO regresyonu, makine öğrenmesi yöntemlerinden rastgele orman, KNN ve DVM yöntemleri kullanılarak gelecek yılların tahmini ve karşılaştırması gerçekleştirilmiştir.

5.1. Problem Tanımı

Türkiye’de ve dünyada iş kazaları, iş gücü verimliliğini olumsuz olarak etkileyen, ekonomik kayıplara yol açan ve çalışanların can güvenliğini tehdit eden önemli bir sorundur. SGK verilerine göre Türkiye’de her yıl on binlerce iş kazası meydana gelmekte, bu kazaların önemli bir kısmı ciddi yaralanmalara veya can kayıplarına neden olmaktadır. SGK verilerine göre 2014-2023 yılları arasında Türkiye’de tüm sektörlerde meydana gelen toplam iş kazası sayıları Şekil 5.1’de gösterilmektedir.



Şekil 5.1. 2014-2023 yılları arasındaki iş kazası sayısı

Mevcut durumda iş kazalarının önlenmesine yönelik çalışmalar daha çok kaza sonrası

yapılan analizlere dayanmaktadır. Ancak teknolojik gelişmelerin sunduğu imkânlar sayesinde, kazaların önceden tahmin edilmesi ve bu tahminlere göre önleyici tedbirlerin alınması mümkün hale gelmiştir. Bu bağlamda, geçmiş iş kazası verilerinin analiz edilerek risk faktörlerinin belirlenmesi ve bu veriler ışığında iş kazası olasılıklarının öngörülmesi, İSG açısından kritik bir ihtiyaçtır.

Bu çalışmada Türkiye'de farklı sektörlerde meydana gelen iş kazalarının verileri incelenerek, iş kazalarını etkileyen temel faktörler belirlenecek ve bu kazaların oluşma olasılıkları çeşitli makine öğrenmesi algoritmaları ile tahmin edilmeye çalışılacaktır. Çalışma kapsamında Türkiye’de meydana gelen iş kazalarına ilişkin sektörel tahmin analizleri madencilik, metal ve inşaat sektörleri olmak üzere üç farklı sektör ele ele alınarak gerçekleştirilmiştir.

Sektörlerin belirlenmesinde, 2014-2023 yılları arasında SGK tarafından yayımlanan iş kazası istatistikleri esas alınmıştır. Bu aşamada SGK tarafından yayımlanan en güncel verinin 2023 yılına ait olması nedeniyle bu yıla kadar olan veriler baz alınarak on yıllık veri kullanılmıştır. Analiz kapsamında, NACE (Nomenclature des Activités Économiques dans la Communauté Européenne - Ekonomik Faaliyet Sınıflandırması) kodlarına göre sınıflandırılmış sektörlerle ait iş kazası sayıları, ölüm sayıları, mortalite sayıları ve iş kazası oranları (insidans) değerlendirilmiştir.

İnsidans oranı, incelenen yıldaki iş kazası sayısının aynı yıl içerisinde hesaplaması yapılan sektördeki çalışan sayısına bölünerek 100 000 ile çarpılmasıyla hesaplanmaktadır (Karakaya Özkan, 2023).

$$\text{İnsidans Oranı} = \frac{\text{İş Kazası Sayısı}}{\text{Sektördeki Çalışan Sayısı}} \times 100\,000 \quad (5.1)$$

Mortalite oranı ise, incelenen yıldaki ölümlü kaza sayısının o yıl için hesaplaması yapılan sektördeki çalışan sayısına bölünmesi ve 100 000 ile çarpılmasıyla hesaplanmaktadır (Karakaya Özkan, 2023).

Bu değerlendirme sonucunda, 2014-2023 yılları arasında en yüksek insidans oranı %25,01 ile “05-Kömür ve Linyit Çıkartılması” sektöründe gözlemlenmiştir. Bu sektörü %10,14 oranı ile “24-Ana Metal Sanayii” sektörü takip etmiştir. Aynı dönemde, toplam 288 706 iş kazası

ile “41-Bina İnşaatı” sektörü en fazla iş kazasının yaşandığı alan olmuştur. Ayrıca, bu sektörde meydana gelen 2 622 ölümle sonuçlanan iş kazası, inşaat sektörünü ölümlü kazaların en çok yaşandığı alan hâline getirmiştir.

Madencilik sektörü hem madenlerin bulunması hem de işlenerek yer altından çıkarılması sürecini ekonomik açıdan verimli ve çevresel etkileri en aza indirecek şekilde gerçekleştiren temel sanayi dallarından birisidir. Ulusların ekonomik gelişimi ve savunma ihtiyaçları için gerekli olan güvenilir mineral ve malzeme kaynaklarına ulaşabilmeleri açısından büyük önem taşır. Tarih öncesi dönemlerden beri varlığını sürdüren madencilik, günümüzde hemen her ülkede uygulanmaktadır. Modern madencilik; cevherin aranması, potansiyel maden sahasının kârlılık ve sürdürülebilirlik açısından değerlendirilmesi, malzemelerin çıkarılması, bu malzemelerin işlenmesi (kıırma, öğütme, yıkama ve konsantre etme gibi işlemler) ve son olarak maden kapandıktan sonra arazinin çevresel açıdan uygun şekilde rehabilite edilmesini kapsayan kapsamlı bir süreçtir (Demiroğlu, 2023).

Metal sektörü sahip olduğu geniş faaliyet alanı ile diğer sanayi kollarının önünde yer almakta ve birçok sektöre doğrudan katkı sağlamaktadır. Otomotivden beyaz eşyaya, ağır sanayi ürünlerinden elektrik-elektronik cihazlara kadar pek çok alanda temel bir rol üstlenmektedir. Bu sektörde eritme, döküm, dövme, presleme, kaynak, kesme ve tornalama gibi çeşitli üretim ve işleme yöntemleri yaygın olarak kullanılmaktadır. Üretim kapasitesinin yüksekliği ve süreçlerin karmaşıklığı, sektörü iş kazaları açısından riskli hale getirmektedir. İmalat süreçlerinin doğası gereği, iş akışının her aşamasında sürekli denetim ve dikkat gereklidir. Bu yönüyle metal sektörü, İSG açısından kritik bir öneme sahiptir (Şahiner, 2024).

İnşaat sektörü hem yarattığı istihdam hem de ekonomiye sağladığı katma değer açısından ülkelerin en canlı ve önemli sektörlerinden biridir. Büyük oranda yerli sermaye ile faaliyet gösteren bu sektör, birçok meslek grubunu içinde barındırmakta ve üretim sürecine doğrudan katkı sunmaktadır. Ancak sektör aynı zamanda iş kazalarının en sık yaşandığı alanlardan biridir. Her yıl birçok inşaat çalışanı bu kazalar nedeniyle hayatını kaybetmekte ya da ciddi şekilde yaralanmaktadır. Bu durum sadece fiziksel kayıplara değil, aynı zamanda önemli psikolojik, sosyal ve ekonomik sorunlara da yol açmaktadır (Çınar, 2018).

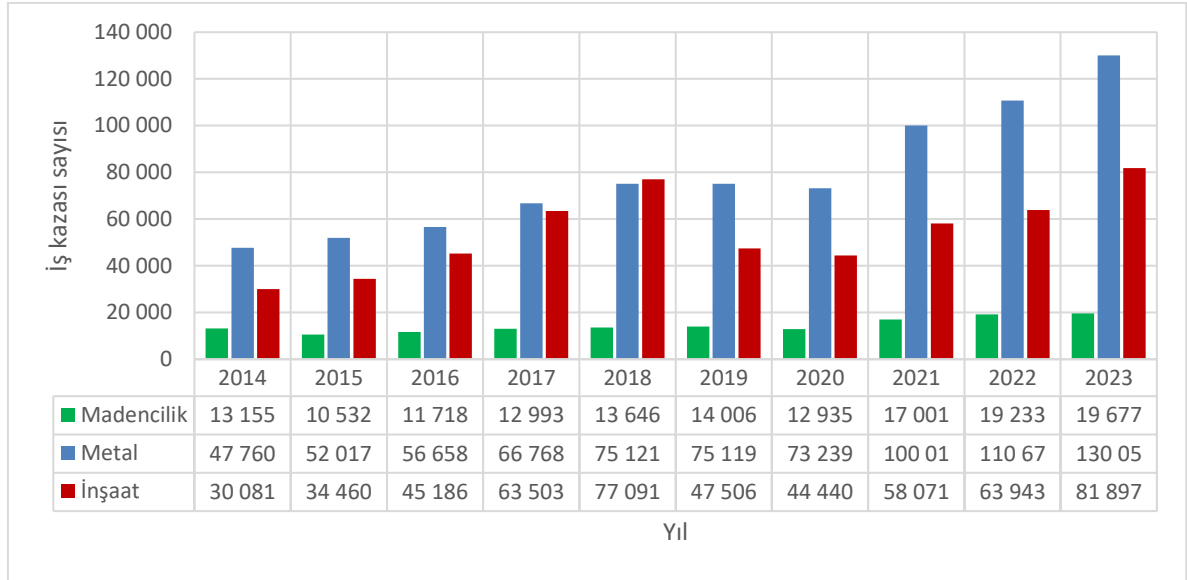
Sektörel bazda daha kapsamlı ve bütüncül analizler yapılabilmesi amacıyla, İş Sağlığı ve Güvenliğine İlişkin İşyeri Tehlike Sınıfları Tebliği'nde yer alan NACE kodları incelenmiş

ve yukarıda yer alan sektörlerin de bulunduğu madencilik, metal ve inşaat sektöründe analizlere dâhil edilecek diğer alt sektörler belirlenmiştir (T.C. Çalışma ve Sosyal Güvenlik Bakanlığı, 2025). Buna göre, madencilik, metal ve inşaat sektörü kapsamında analiz edilecek alt sektörler belirlenmiş ve Çizelge 5.1’de derlenerek sunulmuştur.

Çizelge 5.1. Analiz edilecek alt sektörler

Madencilik sektörü kapsamında analiz edilecek alt sektörler
05-Kömür ve Linyit Çıkartılması 06-Ham Petrol ve Doğalgaz Çıkarımı 07-Metal Cevheri Madenciliği 08-Diğer Madencilik ve Taş Ocakçılığı 09-Madenciliği Destekleyici Hizmet Faaliyetleri
Metal sektörü kapsamında analiz edilecek alt sektörler
24-Ana Metal Sanayii 25-Fabrikasyon Metal Ürünleri İmalatı (Makine ve Teçhizat Hariç) 28-Başka Yerde Sınıflandırılmamış Makine ve Ekipman İmalatı 29-Motorlu Kara Taşıtı, Treyler ve Yarı Treyler İmalatı 30-Diğer Ulaşım Araçlarının İmalatı 33-Makine ve Ekipmanların Kurulumu ve Onarımı
İnşaat sektörü kapsamında analiz edilecek alt sektörler
41-Bina İnşaatı 42-Bina Dışı Yapıların İnşaatı 43-Özel İnşaat Faaliyetleri olarak belirlenmiştir.

Belirlenen alt sektörlerle göre yıllar bazında madencilik, metal ve inşaat sektörlerinde yaşanan iş kazası sayıları Şekil 5.2’de, iş kazaları sonucunda yaşanan ölüm sayıları ise Şekil 5.3’te gösterilmektedir. Şekil 5.2 ve Şekil 5.3, SGK tarafından yayınlanan 2014-2023 yılları arasındaki iş kazası ve ölüm sayıları temel alınarak hazırlanmış ve bu veriler kuruma yapılan yasal bildirimlerden derlenmiştir, dolayısıyla bildiri yapılmayan durumların ayrıca bu verileri arttırabileceği ihtimali dikkate alınmalıdır.

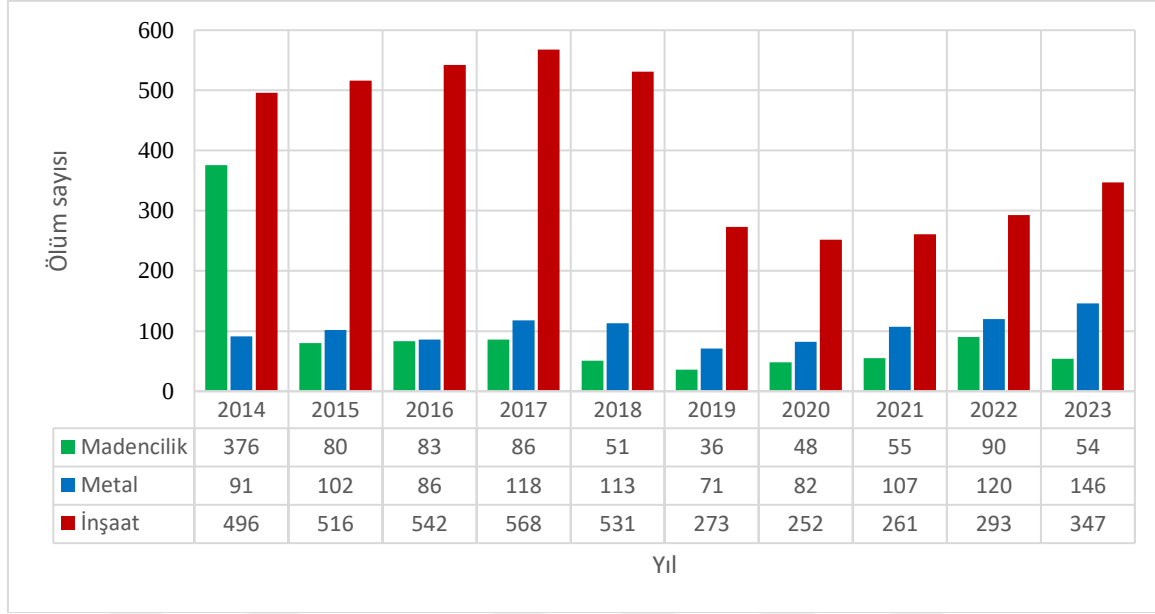


Şekil 5.2. 2014-2023 yılları arasında sektörlere göre yaşanan iş kazası sayısı

Madencilik sektörünün iş kazası sayıları incelendiğinde 2014 yılında 13 155 adet olan iş kazası sayısı 2015 yılında 10 532 adete düşmüş, daha sonraki yıllardaki artış eğilimiyle birlikte 2023 yılında 19 677 adet iş kazası yaşanmıştır. Diğer sektörlere kıyasla madencilik sektöründe yaşanan iş kazalarının artış oranı daha düşük bir seviyededir.

Metal sektörünün iş kazası sayıları incelendiğinde 2014 yılında yaşanan 47 760 iş kazasının 2023 yılında 130 054'e çıktığı görülmektedir. Bu artış yaklaşık olarak %170'lik bir artış anlamına gelmektedir. Özellikle 2020 yılından sonra ciddi bir sıçrama yaşandığı görülmektedir. Bu artış metal sektörünü en riskli sektörlerden birisi haline getirmektedir.

İnşaat sektörünün iş kazası sayıları incelendiğinde iş kazası sayılarının dalgalı bir seyir izlediği görülmektedir. 2014 yılında 30 081 adet iş kazası her yıl artarak 2018 yılında 77 091 adete çıkarak zirve yapmış daha sonra 2019 ve 2020 yıllarında yaşanan düşüşün ardından tekrar artış eğilimine girerek 2023 yılında 81 897 adete çıkmıştır.



Şekil 5.3. 2014-2023 yılları arasında sektörlere göre yaşanan ölüm sayısı

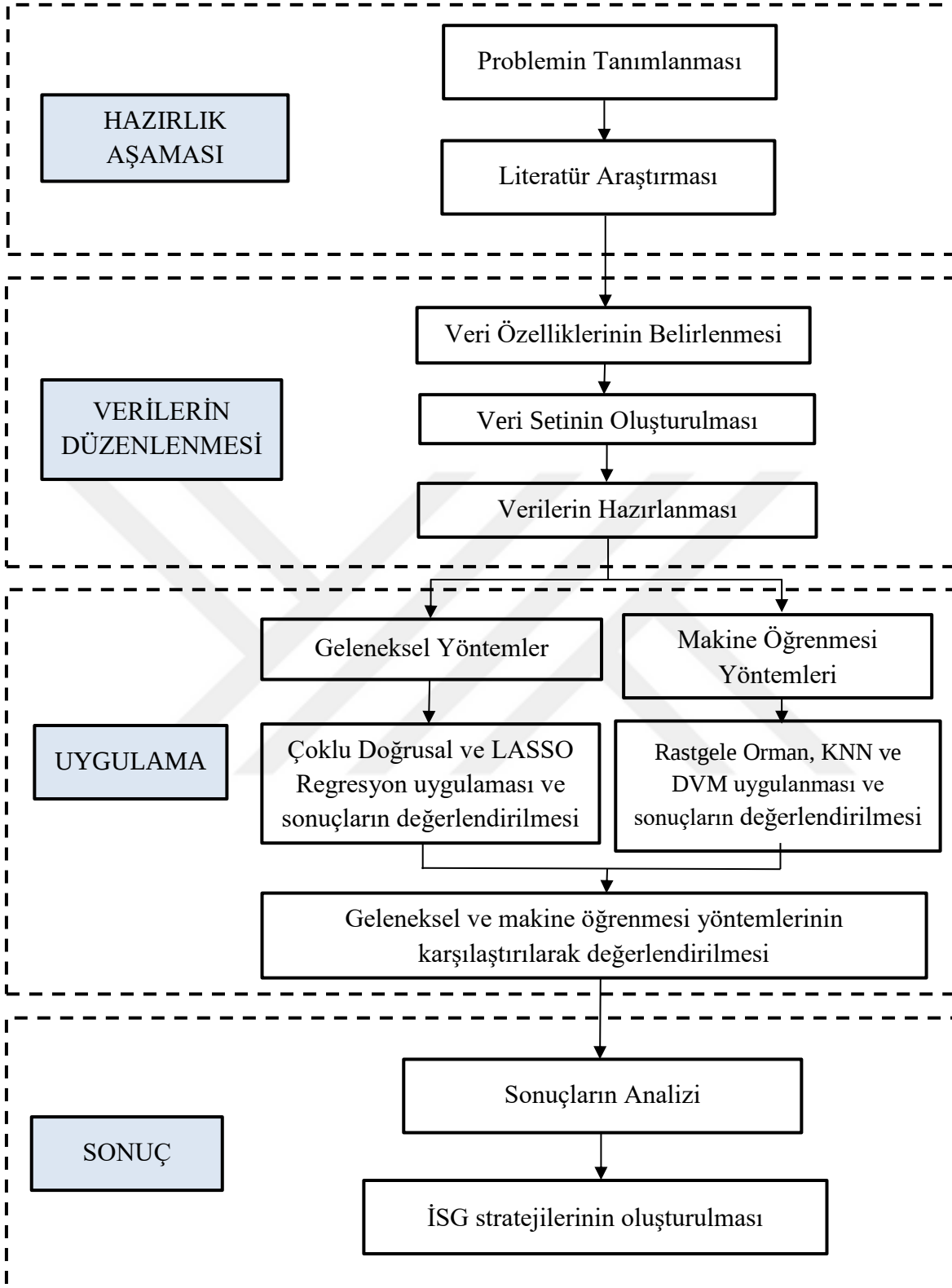
Madencilik sektörü verilerine göre 2014 yılında Manisa'nın Soma ilçesinde yaşanan maden kazası nedeniyle ölüm sayıları çok yüksek olarak gözükmekte ancak daha sonraki yıllarda dalgalı bir seyir izlemektedir.

Metal sektörü verilerine göre genel olarak sabit bir seyir izlemekle birlikte 2019 yılından beri yükselme eğilimi göstermektedir. 2023 yılındaki 146 ölüm sayısı bu sektör için 2014-2023 yılı aralığının en yüksek değeri olarak gözükmektedir.

İnşaat sektörü verilerine göre diğer sektörlere oranla en yüksek ölümlü iş kazalarının yaşandığı sektör olarak görülmektedir. 2020 yılında 252 ölüm ile en düşük seviyeye indiği gözlemlenmiştir. Bu duruma, Covid-19 pandemisinin etkisiyle inşaat faaliyetlerinin yavaşlamasının neden olabileceği değerlendirilmektedir. Sonraki yıllarda ise ölümlü iş kazası sayıları tekrar artış göstererek 2023 yılında ölüm sayısı 347 adet olmuştur.

Bu tez çalışmasında önerilen çözüm yaklaşımı, Şekil 5.4'te gösterildiği üzere hazırlık, verilerin düzenlenmesi, uygulama ve sonuçların değerlendirilmesi olmak üzere dört temel aşamadan oluşmaktadır. İlk aşama olan hazırlık aşamasında, problemin tanımı yapılmış, 2014-2023 yıllarına ilişkin SGK iş kazası verileri kullanılarak insidans ve mortalite değerlerine göre araştırma yapılacak sektörler belirlenmiştir. Ayrıca iş kazaları ile ilgili matematiksel analiz içeren geleneksel yöntemlerin kullanıldığı ve makine öğrenmesi

yöntemlerinin kullanıldığı ilgili literatür taranmış çalışmada kullanılacak algoritmalar belirlenmiş ve hazır hale getirilmiştir. İkinci aşamada, kullanılacak veriler değerlendirilmiş, analizde yer alan ve hesaplamalarda kullanılacak olan SGK verileri kullanılacak tahmin yöntemleri için hazırlanmıştır. Bu aşamada bağımlı ve bağımsız değişkenler belirlenmiş, SGK verilerinin kendi aralarında dağılımları oluşturulmuş, analiz yapılacak olan iş kazası sayısı ve ölüm sayısı bazında oranları hesaplanmıştır. Daha sonra korelasyon analizi yapılarak veri özellikleri analiz edilmiştir. Uygulama aşamasında, hazırlanan veri seti geleneksel ve modern tahmin yöntemleri ile ayrı ayrı analiz edilmiş, çoklu doğrusal regresyon ve LASSO regresyonu geleneksel tahmin yöntemleri ve rastgele orman, KNN ve DVM makine öğrenmesi algoritmaları kullanılarak 2026-2028 yılları için iş kazası ve ölüm sayıları tahmini yapılmıştır. Hesaplamalarda Python'da Jupyter Notebook platformu kullanılmış, pandas ve scikitlearn kütüphaneleri uygulanmıştır. Yürütülen tahmin çalışmalarında elde edilen gerçek hayat bulguları tahmin doğruluğu açısından analiz edilmiş ve yorumlanmış, daha sonra geleneksel ve modern tahmin yöntemlerinin karşılaştırmalı analizi ile ek araştırmalar gerçekleştirilmiştir. Son aşamada ise tahmin ve karşılaştırmalı analiz sonuçlarına dayanan kısa, orta ve uzun vadede İSG stratejileri geliştirilerek önerilerde bulunulmuştur.



Şekil 5.4. Önerilen çözüm yaklaşımı akış diyagramı

5.2. Veriler ve Özellikleri

Türkiye’de iş kazaları ve meslek hastalıklarına ilişkin veriler, işverenlerin iş kazası veya meslek hastalığı olması durumunda yaptığı bildirimler aracılığıyla SGK tarafından toplanmaktadır. SGK, bu verilerin özetlerini her yıl internet sitesinde yayımlamaktadır. Ancak bu istatistikler büyük oranda genel bilgiler içermekte ve sektör bazında bir ayrım sunmamaktadır. Bu durum, sektör bazında analiz yapılmasını zorlaştırmaktadır. Bu durumun önüne geçmek amacıyla insidans ve mortalite değerlerine göre belirlenen Türkiye’deki madencilik, metal ve inşaat sektörleri için 2014-2023 yıllarına ait gerçek hayat verileri SGK’dan resmi olarak talep edilmiştir. Elde edilen verilere göre belirlenen bağımlı ve bağımsız değişkenleri gösteren liste Çizelge 5.2’de verilmiştir. Veri kümesinin büyüklüğü nedeniyle verilerin tamamı bu tez çalışmasının ekinde sunulamamış, araştırma şeffaflığının sağlanabilmesi için tüm veriler Zenodo açık arşiv platformuna yüklenmiş olup, izin dahilinde kullanıcılara erişim yetkisi verilebilmektedir (Nefes, 2025).

Çizelge 5.2 Bağımlı ve bağımsız değişkenler

Değişken Türü	Değişkenler	
Bağımlı değişkenler	İş kazası sayısı	Ölüm sayısı
Bağımsız değişkenler	Kazanın yaşandığı yıl Kazanın yaşandığı ay Kazanın yaşandığı gün Kazanın yaşandığı saat Kazazedenin öğrenim durumu Kazazedenin yaşı Kazazedenin İSG eğitimi durumu Kazazedenin mesleki eğitim durumu Kaza olan işletmedeki çalışan sigortalı sayısı Kazanın yaşandığı il Kaza şiddeti	Kazanın yaşandığı yıl Kazanın yaşandığı ay Kazanın yaşandığı gün Kazanın yaşandığı saat Kazazedenin öğrenim durumu Kazazedenin yaşı Kazazedenin İSG eğitimi durumu Kazazedenin mesleki eğitim durumu Kaza olan işletmedeki çalışan sigortalı sayısı Kazanın yaşandığı il İş kazası sayısı

Heinrich tarafından 1931 yılında yapılan araştırma sonucunda, bir işyerinde 330 iş kazası gerçekleştiğinde, bunların 300’ünün ramak kala olayları olduğunu, 29’unun hafif yaralanmalarla sonuçlandığını ve birinin büyük bir kaza ya da ölümle neticelendiğini belirtilmiştir (Koçali, 2022). Bu çalışma dikkate alınarak ölüm sonucunun iş kazası olayına bağlı olarak geliştiği önermesine dayanarak ölüm sayısı tahmin modelinde iş kazası sayısı

da bağımsız değişken olarak modele dahil edilmiştir.

Belirlenen değişkenlerin dağılımlarına ilişkin kendi içlerindeki dağılımlarının gösterimi Çizelge 5.3’de verilmektedir.

Çizelge 5.3 Değişkenlerin dağılımları

Değişkenler	Alt Kategoriler	
Kazanın yaşandığı yıl	2014 2015 2016 2017 2018	2019 2020 2021 2022 2023
Kazanın yaşandığı ay	Ocak Şubat Mart Nisan Mayıs Haziran	Temmuz Ağustos Eylül Ekim Kasım Aralık
Kazanın yaşandığı gün	Pazartesi Salı Çarşamba Perşembe	Cuma Cumartesi Pazar
Kazanın yaşandığı saat	00:00-08:00 08:00-12:00 12:00-16:00	16:00-20:00 20:00-24:00
Kazazedenin yaş aralığı	14-24 25-34 35-44	45-54 55-64 65+
Kazazedenin öğrenim durumu	Okur-yazar değil Okur-yazar İlkokul Ortaokul	Lise Yüksek okul Lisans Lisansüstü
Kazazede İSG Eğitimi	Var	Yok
Kazazede Mesleki Eğitimi	Var	Yok
İş yerinde çalışan sigortalı sayısı	0-49 50-99 100-249	250-499 500-1000 1 000+
Kazanın yaşandığı il	Tüm iller (Türkiye)	
Kaza şiddeti	Ölümlü Yaralanmalı	Uzuv Kayıplı Diğer

Çalışmada kullanılan veriler üzerinden sektörel bazda korelasyon analizi gerçekleştirilmiştir. Gerçekleştirilen korelasyon analizi, bağımlı değişkenler olan iş kazası sayısı ve ölüm sayısı ile bağımsız değişkenler arasındaki ilişkilerin gücü hakkında önemli

bulgular sunmaktadır. Korelasyon analizi Microsoft Office Excel programı kullanılarak yapılmıştır. Çalışmalar, 16 GB RAM, 2,6 GHz Intel Core i9 işlemci ve Intel UHD Graphics 8 GB grafik kartına sahip, Windows 10 işletim sistemi yüklü bir Dell dizüstü bilgisayarda gerçekleştirilmiştir.

Madencilik sektöründe yapılan korelasyon analizine göre iş kazası sayısı ile en yüksek pozitif korelasyona sahip değişkenler kazazede yaşı (14-24), kaza şiddeti (diğer), öğrenim durumu (lise) olmuştur. Bu bulgular, genç çalışanların daha fazla kazaya karışma eğiliminde olduğunu ve lise mezunu çalışanların daha fazla kazaya karıştığını göstermektedir. Ayrıca kazaların büyük kısmı ölüm, yaralanma veya uzuv kaybı gibi ağır sonuçlar taşımayan türlerde gerçekleşmiştir. Kahramanmaraş, İzmir ve Hakkari illerinin madencilik sektöründeki iş kazaları ile ilgili pozitif korelasyona sahip olduğu görülmüştür.

Madencilik sektörü ölüm sayısı değişkeni ile ilgili yapılan korelasyon analizine göre, işyerinde çalışan sigortalı sayısı (1000+), gün (Salı), saat (12:00-16:00) değişkeni arasında en yüksek pozitif korelasyon görülmüştür. Bu da büyük ölçekli işyerlerinde (1000'den fazla çalışanı olan), veya Salı günü veya saat 12:00-16:00 arasında yaşanacak ölümle sonuçlanan iş kazası oranının daha yüksek olabileceği sonucunu göstermektedir. İl analizi bakımından en yüksek pozitif korelasyon Manisa ilinde görülmüştür. Bu durumun temel nedeni, 2014 yılında Manisa'nın Soma ilçesinde yaşanan maden kazasıdır.

Metal sektöründe yapılan korelasyon analizine göre iş kazası sayısı ile en yüksek pozitif korelasyona sahip değişkenler sırasıyla kaza şiddeti (diğer), kazazede yaşı (14-24) ve öğrenim düzeyi (Lisans) olmuştur. Bu bulgular, 14-24 yaş grubundaki çalışanların daha fazla kazaya karışma eğiliminde olduğunu ve lisans mezunlarının yoğunlaştığı bölgelerde kazaların daha sık yaşandığını göstermektedir. Ayrıca, 1000 ve üzeri çalışanı olan büyük ölçekli işletmelerdeki kazaların daha fazla olması da yüksek korelasyon değeriyle desteklenmiştir. Adana, Konya, Yalova gibi illerin iş kazası sayısı ile yüksek pozitif korelasyon veya Bursa, Hatay, Kocaeli illerinin yüksek negatif korelasyonlara sahip olması, sanayi yoğunluğu ve sektörel faaliyetlerin bölgesel dağılımı ile ilişkilendirilebilir.

Metal sektöründe ölüm sayısı değişkeni ile yapılan korelasyon analizine göre, en yüksek pozitif ilişki saat (00:00–08:00), Çarşamba günü ve kazazede yaşı (45–54) değişkenleri ile gözlenmiştir. Bu bulgular, gece vardiyasında (özellikle 00:00–08:00 saatlerinde) ve hafta

ortasında, meydana gelen iş kazalarının ölümle sonuçlanma olasılığının daha yüksek olabileceğine işaret etmektedir. Ayrıca, 45-54 yaş grubundaki çalışanların karıştığı iş kazalarında ölüm oranının görece daha fazla olduğu da anlaşılmaktadır. İller bazında değerlendirildiğinde ise Kahramanmaraş, Adıyaman, Ankara ve Bolu gibi illerin ölümlü iş kazalarıyla pozitif korelasyon içerisinde olduğu görülmüştür.

İnşaat sektöründe yapılan korelasyon analizine göre iş kazası sayısı değişkeni ile en yüksek pozitif ilişki kazazede yaşı (14-24) değişkeninde gözlenmiştir. Bu bulgu, genç çalışanların kazalara daha fazla maruz kaldığını göstermektedir. Ayrıca, işyerinde çalışan sigortalı sayısı (1000+) ve İSG eğitimi var değişkenleriyle olan pozitif korelasyon, büyük ölçekli işletmelerin hem daha fazla kazaya açık olduğunu hem de bu kazaların daha etkin biçimde raporlandığını düşündürmektedir.

İnşaat sektöründe ölümle sonuçlanan kazaların analizinde en yüksek pozitif korelasyonlar kazazede yaşı (14-24), kazazede yaşı (25-34) ve mesleki eğitim (yok) değişkenlerinde gözlemlenmiştir. Bu durum, ölümle sonuçlanan kazaların özellikle genç ve orta yaşlı çalışanlar arasında daha yaygın olduğunu ve mesleki eğitim eksikliğinin ölümcül riskleri artırabileceğini göstermektedir. İl bazında ise Bolu, Sinop ve Malatya illeri en yüksek pozitif korelasyon gösteren illerdir.

Korelasyon sonuçlarına göre, tahmin modellerinde özellikle kazazede yaşı, öğrenim durumu, İSG ve mesleki eğitim durumu, işyerinde çalışan sigortalı sayısı gibi değişkenlerin dikkate alınması, anlamlı sonuçlar üretme potansiyeline sahiptir. Ayrıca, ölüm sayısına ilişkin analizde elde edilen bulgular, değişkenler arasındaki korelasyon katsayılarının düşük olması nedeniyle ilişkilerin zayıf ve sınırlı düzeyde olduğunu göstermektedir. Bu durum, sektörler bazında ilgili grupların ölümcül kazalara daha fazla maruz kalabileceğini işaret etse de bu değişkenlerin güçlü veya belirleyici risk faktörleri olduğunu söylemek mümkün değildir. Ayrıca korelasyon analizine göre iş kazalarının il bazında dağılımı iş kazası sayılarının tahmini için anlamlı bir sonuç üretme potansiyeli göstermediği için tahmin modellerinde il bazında iş kazası sayıları değişkeni dikkate alınmamıştır. Sektörlere ilişkin özet korelasyon değerleri Çizelge 5.4'te verilmektedir.

Çizelge 5.4. Özet korelasyon değerleri

Madencilik sektörü korelasyon değerleri			
Değişken	İş kazası sayısı	Değişken	Ölüm sayısı
Kazazede yaşı (14-24)	0,671	İşyerinde çalışan sigortalı sayısı (1000+)	0,504
Kaza şiddeti (diğer)	0,639	Gün (Salı)	0,386
Öğrenim durumu (Lise)	0,605	Saat (12:00-16:00)	0,291
Metal sektörü korelasyon değerleri			
Değişken	İş kazası sayısı	Değişken	Ölüm sayısı
Kaza şiddeti (diğer)	0,863	Saat (00:00-08:00)	0,209
Kazazede yaşı (14-24)	0,801	Gün (Çarşamba)	0,137
Öğrenim durumu (Lisans)	0,792	Kazazede yaşı (45-54)	0,132
İnşaat sektörü korelasyon değerleri			
Değişken	İş kazası sayısı	Değişken	Ölüm sayısı
Kazazede yaşı (14-24)	0,858	Kazazede yaşı (14-24)	0,300
İşyerinde çalışan sigortalı sayısı (1000+)	0,632	Kazazede yaşı (25-34)	0,274
İSG eğitimi var	0,529	Mesleki eğitimi yok	0,225

Literatürde, eğitim ve test veri setlerinin oranlarının belirlenmesine ilişkin çeşitli yaklaşımlar mevcuttur. Araştırmacıların büyük bir kısmı, genellikle %90'a %10, %80'e %20 ya da %70'e %30 gibi oranlara dayanan veri bölme yöntemlerini tercih etmektedir (Akdoğan, 2024). Gholamy, Kreinovich ve Kosheleva (2018) tarafından yapılan ampirik araştırmada, veri setinin %80'lik kısmını eğitim, %20'sini test için ayırmanın genellikle en iyi sonuçları verdiği gösterilmiştir. Bu doğrultuda bu çalışmada veri seti kaza tarihine göre %80'i eğitim için ve %20'si test için kullanılmak üzere eğitim ve test veri setlerine bölünmüştür.

Oluşturulan geleneksel ve makine öğrenmesi temelli modellerin etkinliğini değerlendirmek amacıyla modelin doğruluğu test edilmiştir. Bu kapsamda MAPE, MSE ve R^2 parametreleri temelinde karşılaştırmalı analizler gerçekleştirilmiştir.

Model geliştirme aşamasında Python'da Jupyter Notebook kullanılmış, pandas ve scikitlearn kütüphaneleri uygulanmıştır. Çalışmalar, 16 GB RAM, 2,6 GHz Intel Core i9 işlemci ve

Intel UHD Graphics 8 GB grafik kartına sahip, Windows 10 işletim sistemi yüklü bir Dell dizüstü bilgisayarda gerçekleştirilmiştir.

5.3. Geleneksel Tahmin Yöntemleri Hesaplamaları

Literatürde, iş kazası analizlerinde regresyon yöntemlerinin yaygın olarak kullanıldığı görülmektedir (Toptancı, 2021; Ünal, 2019; Öztürk, 2019; Desticioğlu, 2015). Bu çalışmada, 2014-2023 yılları arasında Çizelge 5.2’de verilen değişkenler kullanılarak madencilik, metal ve inşaat sektörlerine ait SGK verileri analiz edilmiştir. Analiz sürecinde, tüm değişkenlerin etkilerini değerlendirmek için çoklu doğrusal regresyon, gereksiz değişkenleri eleyerek daha sağlam ve yalın bir model elde etmek amacıyla ise LASSO regresyon yöntemi kullanılmıştır.

5.3.1. Çoklu doğrusal regresyon analizi hesaplamaları

Çoklu doğrusal regresyon analizi hesaplama algoritmasına dair oluşturulan kaba kod aşağıda sunulmaktadır.

```
# 1. Veriyi Yükleme
veri = EXCEL_DOSYASINI_OKU("Veri_dosyası.xlsx")

# 2. Bağımlı ve Bağımsız Değişkenlerin Belirlenmesi
X = veri'den "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu çıkarılır
Y = veri'nin "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu seçilir

# 3. Veriyi Eğitim ve Test Setlerine Ayırma
[X_train, X_test, Y_train, Y_test] = VERIYI_AYIR(X, Y, test_orani=0.2, rastgelelik=42)

# 4. Model Oluşturma ve Eğitim
model = DOGRUSAL_REGRESYON_MODELİ_OLUSTUR()
model.EGIT(X_train, Y_train)

# 5. Test Verisi ile Tahmin
Y_pred = model.TAHMIN_ET(X_test)

# 6. Performans Değerlendirme
mse = MEAN_SQUARED_ERROR(Y_test, Y_pred)
r2 = R2_SCORE(Y_test, Y_pred)
mape = MEAN_ABSOLUTE_PERCENTAGE_ERROR(Y_test, Y_pred)

EKRANA_YAZ("Model Performansı:")
EKRANA_YAZ("MSE: ", mse)
EKRANA_YAZ("R2: ", r2)
```

```
EKRANA_YAZ("MAPE: ", mape)
```

```
# 7. 2026–2028 Yılları için Tahmin
ortalama_satir = X.ORALAMA_DEGERLERI()
yillik_kaza_tahminleri = BOS_LISTE()
FOR yil IN [2026, 2027, 2028]:
    toplam = 0
    FOR ay IN 1 TO 12:
        satir = ortalama_satir.KOPYALA()
        satir["Yıl"] = yil
        satir["Ay"] = ay
        tahmin = model.TAHMIN_ET([satir])
        toplam = toplam + tahmin
    yillik_kaza_tahminleri.EKLE(toplam)
```

```
# 8. Sonuçları Yazdırma
EKRANA_YAZ("2026–2028 İş Kazası Sayısı Tahminleri:")
FOR i IN 0 TO 2:
    yil = 2026 + i
    toplam = yillik_kaza_tahminleri[i]
    EKRANA_YAZ("Yıl: ", yil, "Tahmini Toplam İş Kazası: ", toplam)
```

Çalışmada ilk olarak çoklu doğrusal regresyon yöntemi uygulanarak, 2014–2023 yılları arasındaki aylık veriler kullanılarak 2026–2028 yılları için iş kazası sayısı ve ölüm oranları tahmin edilmiştir. Model, 120 aylık veri üzerinden oluşturulmuş, %80’i eğitim, %20’si test seti olarak ayrılmıştır. Burada modelin eğitimi aylık veriler kullanılarak yapılırken, tahminler 2026-2028 yılları için aylık olarak yapılmış daha sonra yıllık olarak toplanarak yıllık tahminler bulunmuştur. Madencilik, metal ve inşaat sektörleri kapsamında 2026-2028 yılları için elde edilen iş kazası sayıları ve ölüm sayılarını gösteren tablo Çizelge 5.5’te verilmektedir.

Çizelge 5.5. Çoklu doğrusal regresyon tahmin sonuçları

Sektör	Yıl	İş Kazası Sayısı	Ölüm Sayısı
Madencilik	2026	22 710	62
	2027	23 832	61
	2028	24 954	61
Metal	2026	137 671	141
	2027	145 579	147
	2028	153 486	152
İnşaat	2026	38 657	429
	2027	36 422	431
	2028	34 187	433

Madencilik, metal ve inşaat sektörleri kapsamında geliştirilen çoklu doğrusal regresyon modelinin performansı bu çalışma için seçilen MAPE, MSE ve R² parametreleri çerçevesinde değerlendirilmiş ve elde edilen sonuçlar madencilik sektörü için Çizelge 5.6'da, metal sektörü için Çizelge 5.7'de ve inşaat sektörü için Çizelge 5.8'de sunulmuştur.

Çizelge 5.6 Madencilik sektörü çoklu doğrusal regresyon modeli sonuçları

Tahmin Tipi	MAPE	MSE	R ²
İş kazası sayısı	0,154	86 807,73	-0,265
Ölüm sayısı	1,619	3 798,42	-0,058

Madencilik sektöründeki iş kazası sayısı tahmin modelinin MAPE değerinin 0,154 olması modelin tahminlerinin ortalama olarak kabul edilebilir düzeyde bir hata ile gerçekleştiğini göstermektedir. Ancak, modelin R² değerinin -0,265 olması modelin açıklayıcılığının oldukça düşük olduğunu ve tahminlerinin, sabit ortalama bir tahminden daha kötü performans gösterdiğini ortaya koymaktadır. Bu durum, modelin veri setindeki iş kazası sayısı değişkenini yeterli düzeyde temsil edemediğini, dolayısıyla tahmin gücünün sınırlı olduğunu göstermektedir. Sonuç olarak, iş kazası sayısı tahminine yönelik oluşturulan model, madencilik sektörü verilerine uygulandığında güçlü ve güvenilir bir tahmin aracı olarak değerlendirilemez.

Benzer şekilde, ölüm sayısı modeline ilişkin performans metrikleri de bu değişkenin tahmin edilmesinde yetersiz ve istatistiksel açıdan güvenilir sonuçlar verdiğini göstermektedir. Özellikle 1,619 seviyesinde gerçekleşen MAPE değeri ve negatif R² katsayısı, modelin hem tahmin doğruluğu hem de açıklayıcılık açısından beklentilerin oldukça altında kaldığını ortaya koymaktadır.

Genel olarak değerlendirildiğinde, çoklu doğrusal regresyon modeli hem iş kazası sayısının hem de ölüm sayısının tahmininde veri setiyle uyumlu ve etkili bir yöntem olarak değerlendirilememektedir.

Çizelge 5.7 Metal sektörü çoklu doğrusal regresyon modeli sonuçları

Tahmin Tipi	MAPE	MSE	R ²
İş kazası sayısı	0,099	612 999,96	0,843
Ölüm sayısı	0,863	22,88	0,183

Metal sektöründe iş kazası sayısı tahmin modelinin R² değeri 0,843 olup, modelin verinin

%84,3'ünü doğru şekilde açıkladığını göstermektedir. Bu, modelin güçlü bir tahmin gücüne sahip olduğunu ve iş kazası sayısını oldukça iyi bir şekilde tahmin ettiğini işaret eder. Ancak, MSE değeri 612 999,96 gibi yüksek bir değere sahip olup, bu durum modelin tahminlerinde büyük hatalar yaptığını ve iyileştirilmesi gerektiğini göstermektedir. MAPE değeri ise %9,94 olarak çıkmıştır, bu da modelin tahminlerinin gerçek verilere ortalama %9,94 yakın olduğunu ve bu hata oranının iş kazası gibi sayısal tahminlerde kabul edilebilir olduğunu göstermektedir.

Modelin ölüm sayısı tahmin performansı incelendiğinde, modelin zayıf bir performans sergilediği görülmektedir. MAPE değeri %86,3 gibi yüksek bir seviyededir; bu da modelin tahminlerinin, gerçek değerlere göre neredeyse iki katı kadar sapma gösterdiğini ortaya koymaktadır. R^2 değeri ise 0,183 olup, modelin verideki örüntüleri yeterince tanıyamadığını ve bu nedenle zayıf bir performans sergilediğini göstermektedir. Bu veriler ışığında, mevcut modelin ölüm sayısını tahmin etmede başarısız olduğu söylenebilir.

Metal sektöründe oluşturulan çoklu doğrusal regresyon modeli, iş kazası sayısının tahmininde genel olarak güvenilir ve kullanılabilir bir model olarak değerlendirilmektedir. Ancak ölüm sayısı tahmin modeli performansı, yetersiz bulunmuştur.

Çizelge 5.8 İnşaat sektörü çoklu doğrusal regresyon modeli sonuçları

Tahmin Tipi	MAPE	MSE	R^2
İş kazası sayısı	0,154	679 401,71	0,610
Ölüm sayısı	0,410	227,58	-0,040

İnşaat sektöründe iş kazası sayısını tahmin etmeye yönelik oluşturulan çoklu doğrusal regresyon modelinin performans değerlendirmesi, doğruluk düzeyinin sınırlı olduğunu göstermektedir. R^2 değeri modelin verideki toplam dağılımın yalnızca %61'ini açıklayabildiğini ortaya koymakta bu da modelin açıklayıcılık gücünün görece düşük olduğunu işaret etmektedir. MAPE değeri %15,4 ile hata oranının nispeten yüksek olduğunu göstermektedir. Ayrıca, 679 402 civarında hesaplanan MSE değeri, hata büyüklüğünün önemli düzeyde olduğunu göstermektedir. Bu metrikler birlikte değerlendirildiğinde, modelin doğruluk düzeyinin iyileştirilmeye açık olduğu değerlendirilmektedir.

Ölüm sayısını tahmin etmeye yönelik oluşturulan modelin performansı, zayıf bir tahmin

doğruluğuna işaret etmektedir. Modelin R^2 değeri negatif olup, bu durum modelin verideki yapıyı yakalayamadığını ve öğrenmede başarısız olduğunu ortaya koymaktadır.

Sonuç olarak, çoklu doğrusal regresyon modeli, inşaat sektöründe iş kazası sayısının tahmininde kısıtlı da olsa kullanılabilir bir araç olarak değerlendirilebilirken, ölüm sayılarının tahmininde ise güvenilir bir yöntem olarak kullanılamayacağı anlaşılmaktadır.

5.3.2. LASSO regresyon yöntemi hesaplamaları

Bu çalışmada LASSO regresyon modeli, değişken seçimi özelliği sayesinde yüksek boyutlu veri setlerinde daha sade ve dengeli modellerin oluşturulmasına olanak sağlaması nedeniyle tercih edilmiştir (Vergili, 2023). LASSO regresyon analizi hesaplama algoritmasına dair oluşturulan kaba kod aşağıda sunulmaktadır.

```
# 1. Veriyi Yükleme
veri = EXCEL_DOSYASINI_OKU("Veri_dosyasi.xlsx")

# 2. Bağımlı ve Bağımsız Değişkenlerin Belirlenmesi
X = veri'den "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu çıkarılır
Y = veri'nin "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu seçilir

# 3. Veriyi Eğitim ve Test Setlerine Ayırma
[X_train, X_test, Y_train, Y_test] = VERIYI_AYIR (X, Y, test_orani=0.2, rastgelelik=42)

# 4. Lasso Modelini Oluşturma ve Eğitim
model = LASSO_REGRESYON_MODELİ_OLUSTUR (alpha=n,
maksimum_iterasyon=10000)
model.EGIT(X_train, Y_train)

# 5. Test Verisi ile Tahmin
Y_pred = model.TAHMIN_ET(X_test)

# 6. Performans Değerlendirme
mse = MEAN_SQUARED_ERROR(Y_test, Y_pred)
r2 = R2_SCORE(Y_test, Y_pred)
mape = MEAN_ABSOLUTE_PERCENTAGE_ERROR(Y_test, Y_pred)
EKRAANA_YAZ("Model Performansı (Lasso, alpha=n):")
EKRAANA_YAZ("MSE: ", mse)
EKRAANA_YAZ("R2: ", r2)
EKRAANA_YAZ("MAPE: ", mape)

# 7. 2026–2028 Yılları için Tahmin
ortalama_satir = X.ORTALAMA_DEGERLERI()
```

```

yillik_kaza_tahminleri = BOS_LISTE()
FOR yıl IN [2026, 2027, 2028]:
    toplam = 0
    FOR ay IN 1 TO 12:
        satir = ortalama_satir.KOPYALA()
        satir["Yıl"] = yıl
        satir["Ay"] = ay
        tahmin = model.TAHMIN_ET([satir])
        toplam = toplam + tahmin
    yillik_kaza_tahminleri.EKLE(toplam)

```

8. Sonuçları Yazdırma

```

EKRAANA_YAZ("2026–2028 İş Kazası Sayısı Tahminleri (Lasso):")
FOR i IN 0 TO 2:
    yıl = 2026 + i
    toplam = yillik_kaza_tahminleri[i]
    EKRAANA_YAZ("Yıl: ", yıl, "Tahmini Toplam İş Kazası: ", toplam)

```

En iyi çıktı sonucunu veren alpha değerinin belirlenebilmesi için “0,01” (LASSO_Model_1), “0,1” (LASSO_Model_2), “1” (LASSO_Model_3), “10” (LASSO_Model_4) ve “100” (LASSO_Model_5) değerlerini kullanan LASSO regresyon modelleri oluşturulmuş ve ayrı ayrı hesaplamalar yapılmıştır. Analizde yer alan her sektöre ilişkin bahsedilen beş farklı alpha değerine sahip modellerin üretmiş olduğu sonuçların karşılaştırmasının gösterimi Çizelge 5.9’da sunulmaktadır.

Çizelge 5.9. Modellerin alpha değerine göre karşılaştırması

Sektör	Tahmin Tipi	Alpha değeri	Model	MAPE	MSE	R ²
Madencilik	İş kazası sayısı	0,01	LASSO_Model_1	0,142	70 822,03	-0,031
		0,1	LASSO_Model_2	0,098	23 624,21	0,655
		1	LASSO_Model_3	0,105	34 878,04	0,491
		10	LASSO_Model_4	0,128	56 054,25	0,183
		100	LASSO_Model_5	0,130	51 416,75	0,250
	Ölüm sayısı	0,01	LASSO_Model_1	1,535	3 788,64	-0,056
		0,1	LASSO_Model_2	0,791	3 697,51	-0,030
		1	LASSO_Model_3	0,692	3 676,94	-0,024
		10	LASSO_Model_4	0,764	3 669,85	-0,022
		100	LASSO_Model_5	0,658	3 678,62	-0,025
Metal	İş kazası sayısı	0,01	LASSO_Model_1	0,100	588 905,58	0,848
		0,1	LASSO_Model_2	0,101	596 394,36	0,847
		1	LASSO_Model_3	0,071	403 696,77	0,896
		10	LASSO_Model_4	0,087	523 205,53	0,865
		100	LASSO_Model_5	0,100	595 073,26	0,847
	Ölüm sayısı	0,01	LASSO_Model_1	0,601	16,32	0,417
		0,1	LASSO_Model_2	0,483	13,71	0,510
		1	LASSO_Model_3	0,453	14,34	0,488

Çizelge 5.9.(devam) Modellerin alpha değerine göre karşılaştırması

Sektör	Tahmin Tipi	Alpha değeri	Model	MAPE	MSE	R ²
Metal	Ölüm sayısı	10	LASSO_Model_4	0,418	17,74	0,366
		100	LASSO_Model_5	0,399	18,02	0,356
İnşaat	İş kazası sayısı	0,01	LASSO_Model_1	0,153	689 570,21	0,604
		0,1	LASSO_Model_2	0,134	566 251,61	0,675
		1	LASSO_Model_3	0,111	482 085,76	0,723
		10	LASSO_Model_4	0,126	396 310,03	0,772
		100	LASSO_Model_5	0,202	966 208,18	0,446
	Ölüm sayısı	0,01	LASSO_Model_1	0,373	197,06	0,098
		0,1	LASSO_Model_2	0,244	124,48	0,430
		1	LASSO_Model_3	0,216	110,36	0,495
		10	LASSO_Model_4	0,234	119,10	0,455
		100	LASSO_Model_5	0,243	126,13	0,423

Çizelge 5.9 göstermektedir ki madencilik sektöründe iş kazası sayısı tahmini için LASSO regresyon alpha değeri “0,1” ve ölüm sayısı tahmini için LASSO regresyon alpha değeri “10”, metal sektöründe iş kazası sayısı tahmini için LASSO regresyon alpha değeri “1” ve ölüm sayısı tahmini için LASSO regresyon alpha değeri “0,1” ve inşaat sektöründe iş kazası sayısı tahmini için LASSO regresyon alpha değeri “10” ve ölüm sayısı tahmini için LASSO regresyon alpha değeri “1” olan modeller en iyi sonuçları vermiştir. Buna göre madencilik sektöründe iş kazası sayısı tahmini için LASSO_Model_2, ölüm sayısı tahmini için LASSO_Model_4, metal sektöründe iş kazası sayısı tahmini için LASSO_Model_3, ölüm sayısı tahmini için LASSO_Model_2 ve inşaat sektöründe iş kazası sayısı için LASSO_Model_4, ölüm sayısı tahmini için LASSO_Model_3 modelleri kullanılmıştır.

Buna göre sektörel bazda en iyi tahmin değerlerini veren LASSO modellerinin genelleme yeteneğini artırmak amacıyla ceza katsayısı uygulanmış ve önemsiz görülen değişkenler modelden çıkartılmıştır. Madencilik sektöründe iş kazası tahmini için uygulanan LASSO_Model_2 sonucunda, gün (Salı, Perşembe), saat (08:00-12:00), kazazede yaşı (25-34, 35-44, 55-64, 65+), öğrenim düzeyi (okur yazar değil, lise, lisans, lisansüstü), İSG eğitimi (var, yok), mesleki eğitim (yok), işyerinde çalışan sigortalı sayısı (0-49, 100-249), kaza şiddeti (ölüm, yaralanma, uzuv kayıplı) değişkenleri modelden çıkarıldığı belirlenmiştir. Ölüm sayısı tahmini için uygulanan LASSO_Model_4 sonucunda ise iş kazası sayısı değişkeni dışında tüm değişkenlerin modelden çıkarıldığı belirlenmiştir.

Metal sektöründe iş kazası tahmini için uygulanan LASSO_Model_3 sonucunda, gün

(Pazartesi, Çarşamba, Perşembe), saat (00:00–08:00, 08:00–12:00, 12:00–16:00, 20:00–24:00), kazazede yaşı (25–34, 35–44, 45–54, 55–64, 65+), öğrenim düzeyi (okur yazar değil, okur yazar, ilkokul, ortaokul, yüksekokul, lisans, lisansüstü), İSG eğitimi (var, yok), mesleki eğitim (var, yok), işyerinde çalışan sigortalı sayısı (50–99, 250–499, 1000+) ve kaza şiddeti (ölüm, yaralanma, uzuv kayıplı) değişkenlerinin katsayıları sıfıra indirgenerek modelden çıkarılmıştır. Ölüm sayısı tahmini için uygulanan LASSO_Model_2 sonucunda ise yıl, ay ve iş kazası sayısı değişkeni dışında tüm değişkenlerin modelden çıkarıldığı belirlenmiştir.

İnşaat sektöründe iş kazası tahmini için uygulanan LASSO_Model_4 sonucunda, yıl, ay, kazazede yaşı (14–24) ve işyerinde çalışan sigortalı sayısı (1000+) değişkenleri modelin oluşturulmasında kullanılmış; diğer tüm bağımsız değişkenlerin katsayılarının sıfıra indirgenerek modelden çıkarıldığı belirlenmiştir. Ölüm sayısı tahmini için uygulanan LASSO_Model_3 sonucunda ise ay ve iş kazası sayısı değişkenleri haricindeki tüm değişkenlerin katsayılarının sıfır olduğu ve modelden çıkarıldığı gözlemlenmiştir.

Alt bölüm 5.3.1 temelinde oluşturulan regresyon modeli sonuçları regresyon modellerinin ele alınan veriyi temsil etmede yetersiz kaldığı ve tahmin performansının düşük olduğu (Çizelge 5.6, Çizelge 5.7, Çizelge 5.8) görülmektedir. Benzer şekilde, en yaygın tabanda tahmin analizi yürütmek amacıyla farklı ceza katsayıları denenerek oluşturulan LASSO regresyon modellerinin de yüksek performanslı tahmin verebilmek için pek çok girdi verisini göz ardı ederek problem uzayının uygun temsilini sağlamada yetersiz kaldığı (Çizelge 5.9) görülmektedir. Bu sonuçlardan hareketle problem uzayının detaylı ve tam analizini yapabilmek ve tüm girdi verilerini dikkate alırken yüksek performanslı tahmin geliştirebilmek amaçlarıyla tez çalışması uygulamasının ilerleyen bölümlerinde iş kazası ve ölüm sayısı tahminleri için makine öğrenmesi modelleri de kullanılarak sonuçlar incelenecektir.

Madencilik, metal ve inşaat sektörleri kapsamında 2026-2028 yılları için elde edilen iş kazası sayıları ve ölüm sayılarını gösteren tablo Çizelge 5.10'da verilmektedir.

Çizelge 5.10. LASSO regresyon modeli tahmin sonuçları

Sektör	Yıl	İş Kazası Sayısı	Ölüm Sayısı
Madencilik	2026	23 355	66
	2027	24 547	66
	2028	25 738	66
Metal	2026	119 255	139
	2027	124 683	144
	2028	130 111	149
İnşaat	2026	73 553	406
	2027	76 054	406
	2028	78 554	406

Madencilik, metal ve inşaat sektörleri kapsamında geliştirilen LASSO regresyon modelinin performansı bu çalışma için seçilen MAPE, MSE ve R² parametreleri çerçevesinde değerlendirilmiş ve elde edilen sonuçlar madencilik sektörü için Çizelge 5.11’de, metal sektörü için Çizelge 5.12’de ve inşaat sektörü için Çizelge 5.13’de sunulmuştur.

Çizelge 5.11. Madencilik sektörü LASSO regresyon modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	LASSO_Model_2	0,098	23 624,21	0,655
Ölüm sayısı	LASSO_Model_4	0,764	3 669,85	-0,022

Madencilik sektöründe iş kazası sayısı tahmini için oluşturulan LASSO modelinin R² değeri iş kazası sayısı değişkeninin dağılımının yaklaşık %65,5’ini açıklayabilmektedir. Bu durum, modelin yüksek düzeyde bir tahmin yeteneğine ulaşamadığını göstermektedir. MAPE değeri tahminlerin gerçek değerlere ortalama %9,8 oranında sapma gösterdiğini ve bu hata oranının tahminlerde kabul edilebilir düzeyde olduğunu göstermektedir. Modelin MSE değeri 23 624,21 olarak hesaplanmıştır, bu değer mutlak olarak yüksek görünse de bu sapma, tahmin edilen büyüklükle karşılaştırıldığında görece olarak düşüktür.

Ölüm sayısı için kurulan LASSO modelinin R² değeri “-0,022” olarak hesaplanmıştır ve bu sonuç ölüm sayısı değişkeninin tahmininde LASSO modelinin verideki örüntüleri açıklamakta başarısız olduğunu göstermektedir. Ayrıca, modelin MAPE değeri %76,4 olarak hesaplanmıştır; bu oran, tahminlerin gerçek değerlere kıyasla çok yüksek oranda sapma gösterdiğini ve modelin tahmin gücünün zayıf olduğunu göstermektedir.

Sonuç olarak, iş kazası sayısı tahmini için LASSO regresyon modeli kabul edilebilir düzeyde

güvenilir bir araç olarak değerlendirilebilir, ancak daha yüksek doğruluk için iyileştirmeye açık yönleri bulunmaktadır. Ölüm sayısı tahmini için ise modelin genel performans metrikleri doğrultusunda güvenilir ve kullanılabilir bir tahmin aracı olarak değerlendirilemeyeceği görülmektedir.

Çizelge 5.12. Metal sektörü LASSO regresyon modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	LASSO_Model_3	0,072	403 696,77	0,896
Ölüm sayısı	LASSO_Model_2	0,483	13,71	0,510

Metal sektöründe, iş kazası sayısı modelinin performansı oldukça iyi bir genel doğruluk göstermektedir. R² değeri modelin verideki dağılımın %89,6'sını başarıyla açıkladığını göstermektedir. Bu değer, modelin iş kazası sayısını büyük ölçüde doğru tahmin ettiğini ve genel olarak iyi bir uyum sağladığını ifade eder. Ancak, MSE değeri 403 696,77 gibi yüksek bir değere sahip olup, modelin tahminlerinde büyük hatalar yapıldığını ve daha fazla iyileştirilmesi gerektiğini göstermektedir. Bununla birlikte, MAPE değeri tahminlerin gerçek verilere ortalama %7,2 yakın olduğunu göstermektedir.

Ölüm sayısı tahmin modelinin performansı, değerlendirilen metrikler doğrultusunda orta düzeyde performans göstermiştir. R² değeri 0,510 olup, bu modelin tahminlerde ortalama bir açıklama gücüne sahip olduğunu göstermektedir. MAPE değeri ise modelin tahminlerinin ortalama olarak %48,3 oranında hatalı olduğunu göstermektedir. Genel olarak model, ölüm sayısını tahmin etmede orta düzeyde bir performans sergilemiştir.

Çizelge 5.13. İnşaat sektörü LASSO regresyon modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	LASSO_Model_4	0,126	396 310,03	0,772
Ölüm sayısı	LASSO_Model_3	0,216	110,36	0,495

İnşaat sektöründeki iş kazalarının tahmini için geliştirilen LASSO regresyon modelinin performansına ilişkin metrikler, modelin iş kazası sayısını tahmin etme konusunda orta düzeyde bir başarı sağladığını göstermektedir. R² değeri modelin iş kazası sayısındaki değişimin yaklaşık %77,2'sini açıkladığını ifade etmektedir. Bu oran kabul edilebilir bir açıklayıcılık düzeyine işaret etmektedir. MAPE değeri modelin tahminlerinin ortalama olarak gerçek değerden %12,6 civarında sapma gösterdiğini belirtmektedir. Bu sonuç,

modelin tahminlerinde belirli bir sapma payı olduğunu ve bazı durumlarda iyileştirme yapılabileceğini göstermektedir. MSE değeri ise kabul edilebilir düzeydedir. Bu sonuçlar, modelin belirli bir tahmin kapasitesine sahip olduğunu ve daha ileri düzey optimizasyonlar ile daha yüksek başarı elde edilebileceğini ortaya koymaktadır.

Modelin ölüm sayısını tahmin performansı incelendiğinde, MAPE değeri modelin tahminlerinin gerçek değerlere göre %21,6'lık bir sapma gösterdiğini ifade etmektedir. Bu, modelin tahminlerinin makul düzeyde doğru olduğunu ve yeterli bir doğruluğa sahip olduğunu gösterir. R^2 değeri modelin veriye yaklaşık %50'lik bir açıklayıcılıkla uyum sağladığını gösterir. Bu değer, modelin veriyi orta seviyede açıklayabildiğini, ancak iyileştirme potansiyeline sahip olduğunu ortaya koymaktadır.

Sonuç olarak, iş kazası ve ölüm risklerinin öngörülmesinde LASSO regresyon modeli faydalı olabileceği görülmektedir.

5.4. Makine Öğrenmesi Algoritmaları ve Hesaplamaları

2014-2023 yılları arasında Çizelge 5.1'de verilen değişkenler doğrultusunda hazırlanan veriler kullanılarak madencilik, metal ve inşaat sektörleri için makine öğrenmesi yöntemlerinden rastgele orman, KNN ve DVM kullanılarak ayrı ayrı analizler gerçekleştirilmiştir. Literatür araştırması sonucunda, iş kazalarının analizinde yaygın olarak kullanılan makine öğrenmesi yöntemleri belirlenmiştir. Ayrıca analizin daha geniş bir perspektiften yapılabilmesi için bu yöntemler yaklaşım gruplarına göre değerlendirilmiş ve analizlerde rastgele orman (Luo ve diğerleri, 2023; Sena ve diğerleri, 2024; Tunçman, 2024; Yiğit ve Kazar, 2025; Zermane ve diğerleri, 2023), KNN (Karakaya Özkan, 2023; Sena ve diğerleri, 2024; Zermane ve diğerleri, 2023; Zhu ve diğerleri, 2021) ve DVM (Alkaissy ve diğerleri, 2023; Mammadov ve diğerleri, 2023; Tunçman, 2024; Yiğit ve Kazar, 2025; Zermane ve diğerleri, 2023) yöntemleri kullanılmıştır.

5.4.1. Rastgele orman yöntemi hesaplamaları

Rastgele orman yöntemi hesaplamalarına dair oluşturulan kaba kod aşağıda sunulmaktadır.

```
# 1. Veriyi Yükleme
veri = EXCEL_DOSYASINI_OKU("Veri_dosyasi.xlsx")

# 2. Bağımlı ve Bağımsız Değişkenlerin Belirlenmesi
X = veri'den "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu çıkarılır
Y = veri'nin "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu seçilir

# 3. Veriyi Eğitim ve Test Setlerine Ayırma
[X_train, X_test, Y_train, Y_test] = VERIYI_AYIR(X, Y, test_orani=0.2, rastgelelik=42)

# 4. Rastgele Orman Modelini Oluşturma ve Eğitim
model = RASTGELE_ORMAN_REGRESYON_MODELI_OLUSTUR(agac_sayisi=n,
rastgelelik=42)
model.EGIT(X_train, Y_train)

# 5. Test Verisi ile Tahmin
Y_pred = model.TAHMIN_ET(X_test)

# 6. Performans Değerlendirme
mse = MEAN_SQUARED_ERROR(Y_test, Y_pred)
r2 = R2_SCORE(Y_test, Y_pred)
mape = MEAN_ABSOLUTE_PERCENTAGE_ERROR(Y_test, Y_pred)
EKRANA_YAZ("Model Performansı (Rastgele Orman):")
EKRANA_YAZ("MSE: ", mse)
EKRANA_YAZ("R2: ", r2)
EKRANA_YAZ("MAPE: ", mape)

# 7. 2026–2028 Yılları için Tahmin
ortalama_satir = X.ORTALAMA_DEGERLERI()
yillik_kaza_tahminleri = BOS_LISTE()
FOR yil IN [2026, 2027, 2028]:
    toplam = 0
    FOR ay IN 1 TO 12:
        satir = ortalama_satir.KOPYALA()
        satir["Yıl"] = yil
        satir["Ay"] = ay
        tahmin = model.TAHMIN_ET([satir])
        toplam = toplam + tahmin
    yillik_kaza_tahminleri.EKLE(toplam)

# 8. Sonuçları Yazdırma
EKRANA_YAZ("2026–2028 İş Kazası Sayısı Tahminleri (Rastgele Orman):")
FOR i IN 0 TO 2:
    yil = 2026 + i
```

toplam = yıllık_kaza_tahminleri[i]

EKRANA_YAZ("Yıl: ", yıl, "Tahmini Toplam İş Kazası: ", toplam)

Kaba kodu verilmiş olan rastgele orman algoritmasında en iyi değeri veren ağaç sayısının seçilmesi için 50 (Rastgele_Orman_Model_1), 100 (Rastgele_Orman_Model_2), 200 (Rastgele_Orman_Model_3), 300 (Rastgele_Orman_Model_4), 500 (Rastgele_Orman_Model_5) ve 1 000 (Rastgele_Orman_Model_6) ağaçlı rastgele orman modelleri oluşturularak modellerin performansları ayrı ayrı analiz edilmiştir. Rastgele orman algoritması hesaplamalarında kullanılacak modelin tayini için gerçekleştirilen performans karşılaştırma temelli analiz sonuçları Çizelge 5.14'te sunulmaktadır.

Çizelge 5.14. Modellerin ağaç sayısına göre karşılaştırması

Sektör	Tahmin tipi	Ağaç	Model	MAPE	MSE	R ²
Madencilik	İş kazası sayısı	50	Rastgele_Orman_Model_1	0,103	33 593,74	0,510
		100	Rastgele_Orman_Model_2	0,100	31 133,37	0,546
		200	Rastgele_Orman_Model_3	0,097	30 530,58	0,555
		300	Rastgele_Orman_Model_4	0,097	31 109,71	0,546
		500	Rastgele_Orman_Model_5	0,097	30 654,90	0,553
		1 000	Rastgele_Orman_Model_6	0,096	29 724,24	0,566
	Ölüm sayısı	50	Rastgele_Orman_Model_1	0,274	3 473,27	0,031
		100	Rastgele_Orman_Model_2	0,271	3 423,32	0,045
		200	Rastgele_Orman_Model_3	0,248	3 440,56	0,041
		300	Rastgele_Orman_Model_4	0,235	3 433,26	0,043
		500	Rastgele_Orman_Model_5	0,246	3 445,02	0,039
		1 000	Rastgele_Orman_Model_6	0,236	3 442,85	0,040
Metal	İş kazası sayısı	50	Rastgele_Orman_Model_1	0,105	722 958,01	0,814
		100	Rastgele_Orman_Model_2	0,107	761 700,91	0,804
		200	Rastgele_Orman_Model_3	0,099	698 681,22	0,820
		300	Rastgele_Orman_Model_4	0,100	707 876,62	0,818
		500	Rastgele_Orman_Model_5	0,099	695 422,25	0,821
		1 000	Rastgele_Orman_Model_6	0,098	697 002,23	0,820
	Ölüm Sayısı	50	Rastgele_Orman_Model_1	0,361	13,81	0,507
		100	Rastgele_Orman_Model_2	0,333	13,34	0,524
		200	Rastgele_Orman_Model_3	0,351	13,19	0,529
		300	Rastgele_Orman_Model_4	0,345	13,03	0,535
		500	Rastgele_Orman_Model_5	0,340	12,90	0,539
		1 000	Rastgele_Orman_Model_6	0,348	12,92	0,538
İnşaat	İş kazası sayısı	50	Rastgele_Orman_Model_1	0,103	404 790,86	0,767
		100	Rastgele_Orman_Model_2	0,109	444 629,71	0,745
		200	Rastgele_Orman_Model_3	0,109	431 111,92	0,752
		300	Rastgele_Orman_Model_4	0,106	415 077,30	0,762
		500	Rastgele_Orman_Model_5	0,104	417 187,13	0,760
		1 000	Rastgele_Orman_Model_6	0,105	425 590,54	0,756

Çizelge 5.14.(devam) Modellerin ağaç sayısına göre karşılaştırması

Sektör	Tahmin tipi	Ağaç	Model	MAPE	MSE	R ²
İnşaat	Ölüm sayısı	50	Rastgele_Orman_Model_1	0,206	113,00	0,483
		100	Rastgele_Orman_Model_2	0,212	119,40	0,453
		200	Rastgele_Orman_Model_3	0,207	115,78	0,470
		300	Rastgele_Orman_Model_4	0,202	115,90	0,470
		500	Rastgele_Orman_Model_5	0,202	116,54	0,467
		1 000	Rastgele_Orman_Model_6	0,200	115,80	0,470

Çizelge 5.14'te sunulan hesaplama sonuçlarına göre madencilik sektöründe iş kazası sayısı tahmini için 1 000 ağaçlı ve ölüm sayısı tahmini için 100 ağaçlı rastgele orman modelleri, metal sektöründe iş kazası sayısı ve ölüm sayısı tahmini için 500 ağaçlı rastgele orman modelleri ve inşaat sektöründe iş kazası sayısı ve ölüm sayısı tahmini için 50 ağaçlı rastgele orman modelleri en iyi sonuçları vermiştir. Buna göre madencilik sektöründe iş kazası sayısı tahmini için Rastgele_Orman_Model_6, ölüm sayısı tahmini için Rastgele_Orman_Model_2, metal sektöründe iş kazası ve ölüm sayısı tahmini için Rastgele_Orman_Model_5 ve inşaat sektöründe iş kazası ve ölüm sayısı tahmini için Rastgele_Orman_Model_1 modelleri kullanılmıştır. Madencilik, metal ve inşaat sektörleri kapsamında 2026-2028 yılları için elde edilen iş kazası ve ölüm sayıları tahminleri takip eden Çizelge 5.15'de verilmektedir.

Çizelge 5.15. Rastgele orman modeli tahmin sonuçları

Sektör	Yıl	İş Kazası Sayısı	Ölüm Sayısı
Madencilik	2026	15 728	48
	2027	15 728	48
	2028	15 728	48
Metal	2026	64 853	127
	2027	64 853	127
	2028	64 853	127
İnşaat	2026	54 926	319
	2027	54 926	319
	2028	54 926	319

Madencilik, metal ve inşaat sektörleri gerçek hayat verileri ile iş kazası analizi problemi için parametrik özellikleri temelinde seçilen en uygun rastgele orman modelinin tahmin performansı MAPE, MSE ve R² parametreleri çerçevesinde değerlendirilmiş ve elde edilen sonuçlar madencilik sektörü için Çizelge 5.16'da, metal sektörü için Çizelge 5.17'de ve inşaat sektörü için Çizelge 5.18'de sunulmuştur.

Çizelge 5.16. Madencilik sektörü rastgele orman modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	Rastgele_Orman_Model_6	0,096	29 724,24	0,566
Ölüm sayısı	Rastgele_Orman_Model_2	0,271	3 423,32	0,045

Madencilik sektöründeki iş kazası sayısı tahmini için kurulan rastgele orman modelinin R² değeri değişkenliğin yaklaşık %56,6'sını açıklayabilmektedir. Bu durum, modelin verideki örüntüleri orta düzeyde bir başarıyla yakalayabildiğini göstermektedir. R² değerinin 0,01-0,29 arasındaki değerler alması düşük düzeyde ilişki, 0,3-0,7 arasında değerler alması orta düzeyde ilişki, 0,70 ve üzeri değerler alması yüksek düzeyde ilişki anlamlarına gelmektedir (Korkmaz ve Kacar, 2024). Modelin MAPE değeri tahminlerin gerçek değerlere ortalama %9,6 oranında sapma gösterdiğini ve bu hata payının iş kazası sayısı gibi yüksek büyüklükteki değişkenler için kabul edilebilir düzeyde olduğunu ortaya koymaktadır. Ayrıca, MSE değeri modelin hata düzeyinin görece düşük olduğunu göstermektedir. Genel olarak değerlendirildiğinde, rastgele orman modeli madencilik sektöründe iş kazası sayısının tahmininde orta düzeyde güvenilir bir araç olarak kullanılabilir.

Madencilik sektöründeki ölüm sayısı tahmininde rastgele orman modeli açısından R² değeri ölüm sayısı verisinin yalnızca yaklaşık %4,5'lik kısmını açıklayabilmektedir. Bu oldukça düşük değer, modelin verideki dağılımı yakalama kapasitesinin son derece sınırlı olduğunu göstermektedir. Ayrıca modelin MAPE değeri %27,1 seviyesindedir; bu da tahminlerin gerçek değerlere kıyasla ortalama olarak neredeyse yarı yarıya sapma gösterdiğini ve modelin tahmin doğruluğunun zayıf olduğunu ortaya koymaktadır. Ölçülen MSE değeri önemli ölçüde saptmaya işaret etmektedir. Tüm bu bulgular birlikte değerlendirildiğinde, rastgele orman modeli ölüm sayısı tahmininde yetersiz bir performans sergilemiştir.

Sonuç olarak, rastgele orman algoritması madencilik sektöründe iş kazası sayısının tahmininde orta düzeyde açıklayıcılığa ve kabul edilebilir hata oranına ulaşarak, makul düzeyde güvenilir bir tahmin aracı olarak değerlendirilebileceği görülmüştür. Ancak, ölüm sayısı tahmininde modelin yetersiz performans sergilediği görülmüştür.

Çizelge 5.17. Metal sektörü rastgele orman modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	Rastgele_Orman_Model_5	0,099	695 422,25	0,821
Ölüm sayısı	Rastgele_Orman_Model_5	0,340	12,90	0,539

Metal sektöründe iş kazası sayısını tahmin etmeye yönelik geliştirilen 500 ağaçlı rastgele orman modelinin performans metrikleri, modelin başarılı bir tahmin gücüne sahip olduğunu göstermektedir. Modelin R² değeri 0,821 olup, bu değer bağımsız değişkenlerin hedef değişken olan iş kazası sayısındaki değişimin %82,1'ini açıkladığını ifade etmektedir. R² değerinin %70 ile %90 arasında olması modelin verileri iyi bir şekilde açıkladığını ve iyi bir uyum sağladığını göstermektedir (Türkan, Bozdağ, Karkınlı ve Ulucan, 2023). MAPE değeri ise %9,9 olarak hesaplanmıştır. Bu oran, modelin tahminlerinin ortalama olarak gerçek değerden %10'dan daha az saptığını göstermekte olup, çoğu uygulama için kabul edilebilir hatta iyi düzeyde bir hata oranıdır. MSE değeri ilk bakışta yüksek gibi dursa da tahmin edilen iş kazası sayılarına göre bu hata %1'in biraz üzerinde bir sapmaya karşılık gelmektedir. Bu da mutlak hata düzeyinin, tahmin edilen miktarın büyüklüğüyle kıyaslandığında makul bir düzeyde olduğunu göstermektedir. Sonuç olarak; model, iş kazası sayısını tahmin etmede güçlü ve güvenilir bir performans sergilemektedir.

Metal sektöründeki ölüm sayısı tahmin modeli orta düzeyde bir performans göstermektedir. MAPE değeri modelin tahminlerinin ortalama olarak gerçek değerlerden %34 kadar sapma gösterdiğini göstermektedir. R² değeri ise modelin verideki değişimin yaklaşık %54'ünü açıkladığını göstermektedir. Bu oran, modelin veriyi anlamada ve tahmin üretmede orta seviyede bir açıklayıcılığa sahip olduğunu ortaya koymaktadır. Sonuç olarak, bu modelle elde edilen ölüm sayısı tahminleri belirli bir doğruluk seviyesine sahip olmakla birlikte, daha iyi performans için geliştirilmesi gerekmektedir.

Çizelge 5.18. İnşaat sektörü rastgele orman modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	Rastgele_Orman_Model_1	0,104	404 790,86	0,767
Ölüm sayısı	Rastgele_Orman_Model_1	0,206	113,00	0,483

İnşaat sektöründeki iş kazası sayısını tahmin etmek üzere oluşturulan 50 ağaçlı rastgele orman modelinin doğruluk düzeyi oldukça yüksektir. Modelin R² değeri modelin bağımsız değişkenler aracılığıyla iş kazası sayısındaki dağılımın %76,7'sini açıkladığını göstermekte

olup güçlü bir açıklayıcılık düzeyine işaret etmektedir. MAPE değeri tahminlerin ortalama olarak gerçek değerin yaklaşık %10,4'ü kadar sapma gösterdiğini ve genellikle kabul edilebilir sınırlar içinde olduğunu ortaya koymaktadır. MSE değeri ise %1'in biraz üzerinde bir ortalama kare hataya karşılık gelmektedir. Bu oran, değer mutlak olarak yüksek görünse de tahmin edilen toplam büyüklükle karşılaştırıldığında makul kabul edilebilir. Model genel olarak başarılı bir performans sergilemektedir.

Ölüm sayısı tahmininde ise modelin MAPE değeri tahmin edilen ölüm sayılarının gerçek değerlere göre ortalama %20,6 sapma gösterdiğini ifade etmektedir. Bu oran genellikle orta düzeyde doğruluk anlamına gelmektedir. MSE değeri 113'tür. Bu değer modelin bazı tahminlerde önemli sapmalar yapabildiğini göstermektedir. R^2 değeri ise modelin ölüm sayısındaki dağılımın yaklaşık %48,3'ünü açıkladığını göstermektedir. Bu değer, doğrusal ilişki açısından modelin orta düzeyde bir açıklayıcılığa sahip olduğunu gösterir. Sonuç olarak model, genel eğilimleri yakalayabiliyor olsa da hata oranı ve açıklayıcılığı orta seviyededir.

Sonuç olarak, rastgele orman bu sektörde en dengeli ve başarılı modellerden biri olarak değerlendirilebilir.

5.4.2. KNN yöntemi hesaplamaları

KNN yönteminin kullanılması ile yürütülecek tahmin hesaplamalarına dair oluşturulan kaba kod aşağıda sunulmaktadır.

```
# 1. Veriyi Yükleme
veri = EXCEL_DOSYASINI_OKU("Veri Dosyasi.xlsx")

# 2. Bağımlı ve Bağımsız Değişkenlerin Belirlenmesi
X = veri'den "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu çıkarılır
Y = veri'nin "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu seçilir

# 3. Veriyi Eğitim ve Test Setlerine Ayırma
[X_train, X_test, Y_train, Y_test] = VERIYI_AYIR(X, Y, test_orani=0.2, rastgelelik=42)

# 4. KNN Modelini Oluşturma ve Eğitim
model = KNN_REGRESYON_MODELI_OLUSTUR(komsu_sayisi=n)
model.EGIT(X_train, Y_train)
```

5. Test Verisi ile Tahmin

```
Y_pred = model.TAHMIN_ET(X_test)
```

6. Performans Değerlendirme

```
mse = MEAN_SQUARED_ERROR(Y_test, Y_pred)
```

```
r2 = R2_SCORE(Y_test, Y_pred)
```

```
mape = MEAN_ABSOLUTE_PERCENTAGE_ERROR(Y_test, Y_pred)
```

```
EKRANA_YAZ("Model Performansı (KNN):")
```

```
EKRANA_YAZ("MSE: ", mse)
```

```
EKRANA_YAZ("R2: ", r2)
```

```
EKRANA_YAZ("MAPE: ", mape)
```

7. 2026–2028 Yılları için Tahmin

```
ortalama_satir = X.ORTALAMA_DEGERLERI()
```

```
yillik_kaza_tahminleri = BOS_LISTE()
```

```
FOR yıl IN [2026, 2027, 2028]:
```

```
    toplam = 0
```

```
    FOR ay IN 1 TO 12:
```

```
        satir = ortalama_satir.KOPYALA()
```

```
        satir["Yıl"] = yıl
```

```
        satir["Ay"] = ay
```

```
        tahmin = model.TAHMIN_ET([satir])
```

```
        toplam = toplam + tahmin
```

```
yillik_kaza_tahminleri.EKLE(toplam)
```

8. Sonuçları Yazdırma

```
EKRANA_YAZ("2026–2028 İş Kazası/Ölüm Sayısı Tahminleri (KNN):")
```

```
FOR i IN 0 TO 2:
```

```
    yıl = 2026 + i
```

```
    toplam = yillik_kaza_tahminleri[i]
```

```
EKRANA_YAZ("Yıl: ", yıl, "Tahmini Toplam İş Kazası/Ölüm: ", toplam)
```

Kaba kodu verilmiş olan KNN algoritmasında en iyi değeri veren komşu sayısının seçilmesi için üç (KNN_Model_1), beş (KNN_Model_2), yedi (KNN_Model_3), on (KNN_Model_4) ve on beş (KNN_Model_5) komşulu KNN modelleri oluşturulmuş ve bu modellerin performansları ayrı ayrı hesaplamalar ile analiz edilmiştir. Sektörlere ilişkin farklı komşu sayısına sahip model performansları karşılaştırmalı olarak analiz edilerek ele alınan problemin çözümünde kullanılacak KNN modeli parametreleri belirlenmiştir. KNN algoritması hesaplamalarında kullanılacak modelin tayini için gerçekleştirilen performans karşılaştırma temelli analiz sonuçları Çizelge 5.19’da sunulmaktadır.

Çizelge 5.19. Modellerin komşu sayısına göre karşılaştırması

Sektör	Tahmin tipi	Komşu sayısı	Model	MAPE	MSE	R ²
Madencilik	İş kazası sayısı	3 komşu	KNN_Model_1	0,108	36 671,15	0,465
		5 komşu	KNN_Model_2	0,115	39 758,47	0,420
		7 komşu	KNN_Model_3	0,108	37 680,66	0,451
		10 komşu	KNN_Model_4	0,106	34 959,93	0,490
		15 komşu	KNN_Model_5	0,097	30 653,93	0,553
	Ölüm sayısı	3 komşu	KNN_Model_1	0,207	3 669,18	-0,022
		5 komşu	KNN_Model_2	0,198	3 419,13	0,047
		7 komşu	KNN_Model_3	0,224	3 474,67	0,031
		10 komşu	KNN_Model_4	0,233	3 522,53	0,018
		15 komşu	KNN_Model_5	0,270	3 565,67	0,006
Metal	İş kazası sayısı	3 komşu	KNN_Model_1	0,079	468 924,93	0,879
		5 komşu	KNN_Model_2	0,075	441 235,22	0,886
		7 komşu	KNN_Model_3	0,093	695 711,89	0,821
		10 komşu	KNN_Model_4	0,086	504 302,55	0,870
		15 komşu	KNN_Model_5	0,089	411 406,17	0,894
	Ölüm Sayısı	3 komşu	KNN_Model_1	0,296	11,82	0,296
		5 komşu	KNN_Model_2	0,310	12,05	0,569
		7 komşu	KNN_Model_3	0,338	12,59	0,550
		10 komşu	KNN_Model_4	0,358	13,64	0,513
		15 komşu	KNN_Model_5	0,437	15,78	0,436
İnşaat	İş kazası sayısı	3 komşu	KNN_Model_1	0,119	421 228,50	0,758
		5 komşu	KNN_Model_2	0,133	500 006,74	0,713
		7 komşu	KNN_Model_3	0,166	794 798,27	0,544
		10 komşu	KNN_Model_4	0,169	662 705,67	0,620
		15 komşu	KNN_Model_5	0,201	871 190,70	0,500
	Ölüm sayısı	3 komşu	KNN_Model_1	0,232	97,90	0,552
		5 komşu	KNN_Model_2	0,240	100,21	0,541
		7 komşu	KNN_Model_3	0,227	109,57	0,498
		10 komşu	KNN_Model_4	0,214	101,07	0,537
		15 komşu	KNN_Model_5	0,212	103,54	0,526

Buna göre madencilik sektöründe iş kazası sayısı tahmini için on beş komşulu ve ölüm sayısı tahmini için beş komşulu model, metal sektöründe iş kazası sayısı tahmini için on beş komşulu, ölüm sayısı tahmini için beş komşulu ve inşaat sektöründe iş kazası ve ölüm sayısı tahmini için üç komşulu KNN modelleri en iyi sonuçları vermiştir. Buna göre madencilik sektöründe iş kazası sayısı tahmini için KNN_Model_5, ölüm sayısı tahmini için KNN_Model_2, metal sektöründe iş kazası sayısı tahmini için KNN_Model_5, ölüm sayısı tahmini için KNN_Model_2 ve inşaat sektöründe iş kazası ve ölüm sayısı tahmini için KNN_Model_1 modelleri kullanılmıştır. Bu modellere göre madencilik, metal ve inşaat sektörleri kapsamında 2026-2028 yılları için elde edilen iş kazası ve ölüm sayıları tahminlerini gösteren tablo Çizelge 5.20’de verilmektedir.

Çizelge 5.20. KNN modeli tahmin sonuçları

Sektör	Yıl	İş Kazası Sayısı	Ölüm Sayısı
Madencilik	2026	19 188	36
	2027	19 226	36
	2028	19 345	36
Metal	2026	128 595	120
	2027	129 285	120
	2028	130 093	120
İnşaat	2026	82 478	454
	2027	82 478	454
	2028	82 690	454

Madencilik, metal ve inşaat sektörleri gerçek hayat verilerinin analizi için uygunluğu belirlenen KNN modellerinin tahmin performansları MAPE, MSE ve R^2 parametreleri çerçevesinde değerlendirilmiş ve elde edilen sonuçlar madencilik sektörü için Çizelge 5.21’de, metal sektörü için Çizelge 5.22’de ve inşaat sektörü için Çizelge 5.23’de sunulmuştur.

Çizelge 5.21. Madencilik sektörü KNN modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R^2
İş kazası sayısı	KNN_Model_5	0,097	30 653,94	0,553
Ölüm sayısı	KNN_Model_2	0,198	3 419,13	0,047

Bu çalışmada madencilik sektörü iş kazası sayısı tahminlerinde kullanılan on beş komşulu KNN modeli R^2 değeri analiz edildiğinde iş kazası sayısı verisinin yaklaşık %55,3’ünü açıklanabilmekte olduğu görülmektedir ve bu da modelin verideki örüntüleri orta düzeyde başarıyla temsil ettiğini göstermektedir. MAPE değeri modelin tahminlerinin gerçek değerlere ortalama olarak yaklaşık %10 oranında sapma gösterdiğini ortaya koymaktadır. MSE değeri ise modelin hata büyüklüğünün kabul edilebilir sınırlar içinde kaldığını göstermektedir. Tüm bu metrikler değerlendirildiğinde, KNN modeli iş kazası sayısının tahmininde kullanılabilir ve güvenilir bir alternatif olarak öne çıkmaktadır.

Öte yandan, ölüm sayısı tahminlerinde kullanılan KNN modelinin R^2 değeri, ölüm sayısı verisinin varyansını açıklamada yetersiz bir performans göstermiştir. MAPE değeri modelin tahminlerinin gerçek değerlere ortalama olarak %20’ye yakın sapma gösterdiğini ortaya koymakta ve bu hata oranı önemli bir sapmaya işaret etmektedir. KNN modeli madencilik sektöründe ölüm sayısı tahmininde istatistiksel olarak yetersiz ve güvenilirliği düşük bir

model olarak değerlendirilmektedir.

Sonuç olarak, KNN modeli madencilik sektöründe iş kazası sayısının tahmininde kısmen başarılı bir performans sergilemiş, orta düzeyde açıklayıcılığa ve kabul edilebilir hata oranına ulaşarak kullanılabilir bir tahmin aracı olarak öne çıkmıştır. Ancak, aynı model ölüm sayısı tahminlerinde gerek açıklayıcılık gücü gerekse hata düzeyi açısından yetersiz kalmış, yüksek sapma oranları nedeniyle istatistiksel olarak güvenilir bir sonuç üretmemiştir.

Çizelge 5.22. Metal sektörü KNN modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	KNN_Model_5	0,089	411 406,17	0,894
Ölüm sayısı	KNN_Model_2	0,310	12,05	0,569

Tez çalışmasında metal sektörü için kullanılan on beş komşulu KNN modeli metal sektöründeki iş kazası sayısı tahminlerinde genel olarak iyi bir doğruluk sergilemektedir. R² değeri modelin verilerdeki dağılımın %89,4'ünü doğru şekilde açıkladığını göstermektedir. MSE değeri ise oldukça makul bir düzeydedir.

Ölüm sayısı tahmininde, beş komşulu KNN modelinin R² değeri verideki değişimin yaklaşık %57'sini açıkladığını göstermektedir. Bu, modelin tahminlerinin orta seviyede doğru olduğunu göstermektedir. MAPE değeri ise modelin tahminlerinin ortalama olarak gerçek verilere %31 kadar sapma göstermektedir. Sonuç olarak, model orta düzeyde başarılı bir şekilde tahminler yapmaktadır.

Çizelge 5.23. İnşaat sektörü KNN modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	KNN_Model_1	0,119	421 228,50	0,758
Ölüm sayısı	KNN_Model_1	0,232	97,90	0,552

İnşaat sektörü için kullanılan üç komşulu KNN modelinin performans metrikleri incelendiğinde, modelin genel olarak başarılı bir tahmin gücüne sahip olduğu görülmektedir. MAPE değeri %11,9 olup, tahminlerin gerçek değerlere oldukça yakın olduğunu göstermektedir. R² skoru 0,758 ile modelin verideki değişkenliğin yaklaşık %75,8'ini açıkladığını ortaya koymaktadır. MSE değeri büyük bir sayı olarak görünse de tahminlerin büyüklüğü ile kıyaslandığında oldukça makul düzeydedir. Genel olarak model, iş kazası

sayısı tahmini açısından kabul edilebilir bir performans sergilemektedir.

Ölüm sayısı tahmini modeli performansı genel olarak orta düzeydedir. MAPE değeri modelin tahminlerinin ortalama olarak gerçek değerlerden %23,2 kadar sapma gösterdiğini göstermektedir. MSE değeri 97,90 ise, ortalama ölüm sayısının 454 olduğu göz önüne alındığında, modelin orta düzeyde bir hata payına sahip olduğunu gösterir. R^2 değeri modelin verideki dağılımın %55,2'sini açıkladığını göstermektedir. Sonuç olarak, model orta düzeyde bir performans sergilemekte olup, daha doğru tahminler için modelin iyileştirilmesi ve girdi özelliklerinin gözden geçirilmesi gerekmektedir.

5.4.3. DVM yöntemi hesaplamaları

DVM yöntemi ile iş kazası tahmini hesaplamalarına dair oluşturulan kaba kod aşağıda sunulmaktadır.

```
# 1. Veriyi Yükleme
veri = EXCEL_DOSYASINI_OKU("Veri Dosyasi.xlsx")

# 2. Bağımlı ve Bağımsız Değişkenlerin Belirlenmesi
X = veriden "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu çıkarılır
Y = verinin "İş Kazası Sayısı"/"Ölüm Sayısı" sütunu seçilir

# 3. Veriyi Ölçeklendirme (StandardScaler)
X_OLCEKLEYICI = STANDARD_SCALER_OLUSTUR()
Y_OLCEKLEYICI = STANDARD_SCALER_OLUSTUR()
X_olcekli = X_OLCEKLEYICI.EGIT_VE_DONUSTUR(X)
Y_olcekli = Y_OLCEKLEYICI.EGIT_VE_DONUSTUR(Y.RESHAPE(-
1,1)).DUZLESTIR()

# 4. Veriyi Eğitim ve Test Setlerine Ayırma
[X_train, X_test, Y_train, Y_test] = VERIYI_AYIR(X_olcekli, Y_olcekli, test_orani=0.2,
rastgelelik=42)

# 5. DVM Modelini Oluşturma ve Eğitim
model = DVM_MODELİ_OLUSTUR(cekirdek="linear"/"rbf")
model.EGIT(X_train, Y_train)

# 6. Test Verisi ile Tahmin ve Geri Ölçekleme
Y_pred_olcekli = model.TAHMIN_ET(X_test)
Y_pred = Y_OLCEKLEYICI.GERI_DONUSTUR(Y_pred_olcekli)
Y_test_orijinal = Y_OLCEKLEYICI.GERI_DONUSTUR(Y_test)

# 7. Performans Değerlendirme
```

```

mse = MEAN_SQUARED_ERROR(Y_test_orijinal, Y_pred)
r2 = R2_SCORE(Y_test_orijinal, Y_pred)
mape = MEAN_ABSOLUTE_PERCENTAGE_ERROR(Y_test_orijinal, Y_pred)
EKRANA_YAZ("Model Performansı (DVM):")
EKRANA_YAZ("MSE: ", mse)
EKRANA_YAZ("R2: ", r2)
EKRANA_YAZ("MAPE: ", mape)

```

```

# 8. 2026–2028 Yılları için Tahmin
ortalama_satir = X.ORTALAMA_DEGERLERI()
yillik_kaza_tahminleri = BOS_LISTE()
FOR yil IN [2026, 2027, 2028]:
    toplam = 0
    FOR ay IN 1 TO 12:
        satir = ortalama_satir.KOPYALA()
        satir["Yıl"] = yil
        satir["Ay"] = ay
        satir_olcekli = X_OLCEKLEYICI.DONUSTUR([satir])
        tahmin_olcekli = model.TAHMIN_ET(satir_olcekli)
        tahmin = Y_OLCEKLEYICI.GERI_DONUSTUR(tahmin_olcekli)
        toplam = toplam + tahmin
    yillik_kaza_tahminleri.EKLE(toplam)

# 9. Sonuçları Yazdırma
EKRANA_YAZ("2026–2028 İş Kazası/Ölüm Sayısı Tahminleri (DVM):")
FOR i IN 0 TO 2:
    yil = 2026 + i
    toplam = yillik_kaza_tahminleri[i]
    EKRANA_YAZ("Yıl: ", yil, "Tahmini Toplam İş Kazası/Ölüm: ", toplam)

```

Kaba kodu verilmiş olan DVM algoritmasını kullanacak model parametrelerinin belirlenebilmesi için en iyi değeri veren çekirdek fonksiyonunun belirlenmesi için lineer (DVM_Model_1), döngüsel temelli (DVM_Model_2), polinom (DVM_Model_3) ve sigmoid (DVM_Model_4) çekirdek fonksiyonlarına sahip DVM modelleri oluşturulmuş ve bu modellerin performansları ayrı ayrı hesaplamalar ile analiz edilmiştir. DVM algoritması hesaplamalarında kullanılacak modelin tayini için gerçekleştirilen performans karşılaştırma temelli analiz sonuçları Çizelge 5.24'te sunulmaktadır. DVM hesaplamalarında madencilik sektörü için iş kazası tahmininde en iyi sonucun döngüsel temelli çekirdek fonksiyonu, ölüm sayısı tahmininde polinom çekirdek fonksiyonu, metal sektörü için iş kazası tahmininde en iyi sonucun lineer çekirdek fonksiyonu, ölüm sayısı tahmininde sigmoid çekirdek fonksiyonu çekirdek fonksiyonunun, inşaat sektörü için her iki değişkenin tahmininde ise döngüsel temelli çekirdek fonksiyonu çekirdek fonksiyonunun verdiği bulunmuştur. Buna göre madencilik sektöründe iş kazası sayısı tahmini için DVM_Model_2, ölüm sayısı

tahmini için DVM_Model_3, metal sektöründe iş kazası sayısı tahmini için DVM_Model_1, ölüm sayısı tahmini için DVM_Model_4 ve inşaat sektöründe iş kazası ve ölüm sayısı tahmini için DVM_Model_2 modelleri kullanılmıştır.

Çizelge 5.24. Modellerin çekirdek fonksiyonuna göre karşılaştırması

Sektör	Tahmin tipi	Fonksiyon	Model	MAPE	MSE	R ²
Madencilik	İş kazası sayısı	Lineer	DVM_Model_1	0,158	73 696,75	-0,073
		Döngüsel temelli	DVM_Model_2	0,082	24 965,00	0,636
		Polinom	DVM_Model_3	0,154	107 069,55	-0,559
		Sigmoid	DVM_Model_4	0,158	123 542,67	-0,799
	Ölüm sayısı	Lineer	DVM_Model_1	1,377	3 695,48	-0,030
		Döngüsel temelli	DVM_Model_2	0,926	3 675,99	-0,024
		Polinom	DVM_Model_3	0,914	3 634,42	-0,013
		Sigmoid	DVM_Model_4	1,111	3 738,45	-0,042
Metal	İş kazası sayısı	Lineer	DVM_Model_1	0,088	464 406,96	0,881
		Döngüsel temelli	DVM_Model_2	0,116	724 704,50	0,814
		Polinom	DVM_Model_3	0,149	1 368 672,58	0,648
		Sigmoid	DVM_Model_4	0,204	1 883 714,03	0,516
	Ölüm Sayısı	Lineer	DVM_Model_1	0,845	23,56	0,159
		Döngüsel temelli	DVM_Model_2	0,415	23,15	0,173
		Polinom	DVM_Model_3	0,720	25,80	0,079
		Sigmoid	DVM_Model_4	0,345	22,67	0,190
İnşaat	İş kazası sayısı	Lineer	DVM_Model_1	0,148	656 536,63	0,623
		Döngüsel temelli	DVM_Model_2	0,105	283 731,44	0,837
		Polinom	DVM_Model_3	0,179	706 048,48	0,595
		Sigmoid	DVM_Model_4	0,176	727 273,37	0,583
	Ölüm sayısı	Lineer	DVM_Model_1	0,347	165,75	0,242
		Döngüsel temelli	DVM_Model_2	0,216	133,83	0,388
		Polinom	DVM_Model_3	0,418	230,70	-0,055
		Sigmoid	DVM_Model_4	0,381	213,79	0,022

Belirlenen bu modellere göre madencilik, metal ve inşaat sektörleri kapsamında 2026-2028 yılları için elde edilen iş kazası sayıları ve ölüm sayılarını gösteren tablo Çizelge 5.25’de verilmektedir.

Çizelge 5.25. DVM modeli tahmin sonuçları

Sektör	Yıl	İş Kazası Sayısı	Ölüm Sayısı
Madencilik	2026	16 272	91
	2027	16 323	91
	2028	16 344	91
Metal	2026	113 684	117
	2027	118 363	119
	2028	123 042	122

Çizelge 5.25.(devam) DVM modeli tahmin sonuçları

Sektör	Yıl	İş Kazası Sayısı	Ölüm Sayısı
İnşaat	2026	56 254	500
	2027	56 420	477
	2028	56 541	454

Madencilik, metal ve inşaat sektörleri kapsamında geliştirilen DVM modellerinin performansı MAPE, MSE ve R² parametreleri çerçevesinde değerlendirilmiş ve elde edilen sonuçlar madencilik sektörü için Çizelge 5.26’da, metal sektörü için Çizelge 5.27’de ve inşaat sektörü için Çizelge 5.28’de sunulmuştur.

Çizelge 5.26. Madencilik sektörü DVM modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	DVM_Model_2	0,082	24 965,00	0,636
Ölüm sayısı	DVM_Model_3	0,914	3 634,42	-0,013

Bu çalışmada madencilik sektöründe iş kazası sayısı tahmininde kullanılan DVM modelinin R² değeri modelin iş kazası sayısı verisinin yaklaşık %63,6’sını açıklayabildiğini ve veri içindeki örüntüleri orta ile yüksek arasında bir doğruluk düzeyiyle temsil ettiğini göstermektedir. Modelin MAPE değeri tahminlerin gerçek değerlere %8,2 civarında sapma gösterdiğini ortaya koymakta ve bu hata oranı, yüksek hacimli değişkenler için oldukça kabul edilebilir düzeydedir. Ayrıca, MSE değeri hata büyüklüğünün kabul edilebilir sınırlar içinde kaldığını göstermektedir.

Madencilik sektöründe ölüm sayısını tahmin etmeye yönelik olarak kullanılan DVM modelinin R² değerinin negatif olması, modelin ölüm sayısı verisindeki varyasyonu açıklamakta başarısız olduğunu göstermektedir. MAPE değeri modelin tahminlerinin gerçek değerlere kıyasla ortalama olarak iki kata yakın bir sapma içerdiğini ve tahmin doğruluğunun son derece zayıf olduğunu ortaya koymaktadır. Ayrıca, modelin MSE değeri önemli boyutta hata içerdiğini göstermektedir.

Bu bulgular ışığında, DVM modeli, madencilik sektöründe iş kazası sayısının tahmininde güvenilir ve etkili bir tahmin aracı olarak kullanılabilecek düzeyde performans sergilemiştir. Ölüm sayısı tahmininde ise kullanılabilirlik düzeyinin oldukça altında kalmış ve güvenilir bir öngörü aracı olarak değerlendirilememektedir.

Çizelge 5.27. Metal sektörü DVM modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	DVM_Model_1	0,088	464 406,97	0,881
Ölüm sayısı	DVM_Model_4	0,345	22,67	0,190

Metal sektöründe iş kazası sayısı tahmin modeli performans metrikleri, modelin yüksek düzeyde bir doğruluk sağladığını göstermektedir. R² değeri modelin verilerdeki dağılımın %88,1'ini açıkladığını göstermektedir. MSE değeri yüksek olup, modelin tahminleri ile gerçek değerler arasında önemli bir hata bulunduğunu göstermektedir. MAPE değeri ise %8,8 olup, bu hata oranı genellikle iyi kabul edilir.

Ölüm sayısı tahmin modelinin MAPE değeri modelin tahminlerinin gerçek değerlere göre ortalama %34,5 hata ile sapma gösterdiğini ortaya koymaktadır. Bu, oldukça yüksek bir hata oranıdır ve modelin tahmin gücünü sınırlamaktadır. R² değeri ise modelin toplam dağılımın yalnızca yaklaşık %19'unu açıklayabilmekte olduğunu göstermektedir. Bu, modelin verideki desenleri açıklama ve genelleme kapasitesinin oldukça düşük olduğunu göstermektedir.

Çizelge 5.28. İnşaat sektörü DVM modeli sonuçları

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	DVM_Model_2	0,105	283 731,44	0,837
Ölüm sayısı	DVM_Model_2	0,216	133,83	0,388

İnşaat sektöründe iş kazası sayısının tahminine yönelik geliştirilen modelin performans metrikleri değerlendirildiğinde, modelin genel anlamda tatmin edici sonuçlar verdiği görülmektedir. Modelin doğruluk düzeyini gösteren MAPE değeri %10,5 olup, bu oran genellikle kabul edilebilir sınırlar içerisindedir. R² değeri 0,837 olarak hesaplanmış ve bu da modelin, iş kazası sayısındaki değişkenliğin yaklaşık %83,7'sini açıklayabildiğini ortaya koymaktadır. Bu oran, modelin verideki genel desenleri iyi yakaladığını ve tahminlerde güçlü bir performans sergilediğini göstermektedir. Ayrıca MSE değeri 283 731 gibi yüksek bir sayı vermesine rağmen bu değer doğrudan tahmin edilen iş kazası sayılarının büyüklüğüyle ilişkili olması nedeniyle bu bağlamda yorumlanmalıdır. Sonuç olarak model, tahmin gücü yüksek, hata oranı makul düzeyde olan ve metal sektörü verisine uygunluk gösteren bir yapı sunmaktadır.

Ölüm sayısı tahmin modelinin MAPE değeri tahminlerin gerçek değerlere %21,6 sapma gösterdiğini ortaya koymaktadır. Bu oran, hedef değişkenin küçük değerler içermesi göz önünde bulundurulduğunda nispeten yüksek bir hata payına işaret etmekte ve modelin öngörülerinde tutarsızlık riski bulunduğunu göstermektedir. MSE değeri 133,83 olup, düşük ölçekli bir hedef değişken için anlamlı bir hata büyüklüğü ifade etmektedir. R^2 değeri %38,8 düzeyinde gerçekleşmiş ve bu da modelin veri içindeki değişkenliğin yalnızca sınırlı bir kısmını açıklayabildiğini göstermektedir. Bu durum, girdi değişkenlerinin hedef değişken üzerindeki etkisinin sınırlı kaldığını ortaya koymaktadır. Sonuç olarak, model temel düzeyde bir tahmin gücü sunmakla birlikte, özellikle hata oranları açısından geliştirmeye açık bir yapıya sahiptir.

5.5. Bulgular ve Tartışma

5.5.1. Korelasyon analizi bulguları

Bu tez çalışması kapsamında sektörel bazdaki aylık veriler kullanılarak korelasyon analizleri gerçekleştirilmiştir. Analizler, iş kazası sayısı ve ölüm sayısı olmak üzere iki temel bağımlı değişken üzerinden yürütülmüş; bunların çeşitli sosyo-demografik, kurumsal ve bölgesel değişkenlerle olan ilişkileri incelenmiştir.

Madencilik sektöründe yapılan korelasyon analizleri en yüksek pozitif korelasyon kazazede yaşı (14–24), kaza şiddeti (diğer) ve öğrenim durumu (lise) değişkenleri arasında gözlemlenmiştir. Genç çalışanlar ve lise mezunu işçiler arasında daha fazla kaza yaşanması, eğitim ve deneyim eksikliklerinin kazaları artıran önemli faktörler arasında yer alabileceğini göstermektedir. Bu bulgu, sektördeki genç işçilerin daha yüksek kaza riski taşıdığına işaret etmektedir. Ayrıca, analizde kazaların çoğunun ölüm, yaralanma veya uzuv kaybı gibi ağır sonuçlar taşımayan türlerde gerçekleştiği bulunmuştur. Bu durum, sektörün genelde daha hafif kazalarla karşı karşıya olduğunu ancak yine de önemli bir İSG sorununa işaret ettiğini göstermektedir. Bölgesel analizlerde ise Kahramanmaraş, İzmir ve Hakkari illerinin madencilik sektöründeki iş kazaları ile yüksek pozitif korelasyona sahip olduğu görülmüştür. Bu illerdeki kazaların daha sık görülmesi, bölgesel sanayi yoğunluğu ve yerel İSG uygulamalarına dair farklılıklarla ilişkilendirilebilir. Bu bulgu, her bir bölgenin yerel dinamiklerinin ve sanayi yoğunluğunun iş kazası oranları üzerinde etkili olabileceğini düşündürmektedir.

Madencilik sektöründeki ölüm sayısı ile yapılan korelasyon analizlerinde, işyerinde çalışan sigortalı sayısı (1000+), gün (Salı), saat (12:00-16:00) değişkenleri arasında en yüksek pozitif korelasyonlar gözlemlenmiştir. Bu bulgular, büyük ölçekli işletmelerde (1000 ve üzeri çalışanı olan) çalışanların ölümle sonuçlanan kazalara daha fazla maruz kaldığını, özellikle Salı günleri ve saat 12:00-16:00 arasında ölümle sonuçlanan kazaların daha sık görüldüğünü göstermektedir. Bu durum, çalışma saatlerinin ve günlerinin İSG riskleri üzerinde etkili olabileceğini ortaya koymaktadır. Literatürde yer alan Pazartesi etkisinin (yüksek iş kazası sayısı beklentisi) madencilik sektöründe Salı gününe kayması, sektörde sürekli çalışma modeli (yedi gün-24 saat) uygulanmasına dayanarak açıklanabilecektir. İl düzeyinde ise Manisa ilinin ölüm sayısı ile en yüksek pozitif korelasyona sahip olduğu tespit edilmiştir. Bu bulgu, 2014 yılında Manisa'nın Soma ilçesinde yaşanan büyük maden kazasının etkilerini yansıtmaktadır.

Metal sektöründe yapılan korelasyon analizleri, özellikle genç çalışanlar ve büyük ölçekli işletmelere dair önemli bulgular sunmaktadır. İş kazası sayısı ile en yüksek pozitif korelasyon; kaza şiddeti (diğer), kazazede yaşı (14–24) ve öğrenim düzeyi (lisans) değişkenlerinde gözlemlenmiştir. Madencilik sektörüne benzer şekilde genç çalışanların deneyim eksikliği ve risk algısının düşük olması, iş kazası riskini artıran etkenler arasında sayılmaktadır. Ayrıca, metal sektörü için işgücü demografisi incelendiğinde çalışanlar arasında lisans mezunlarının çoğunlukta olması nedeniyle iş kazaları ve eğitim durumunun ilişkisinde lisans mezunu çalışan ve iş kazası sayısı ilişkisinin kuvvetli olması normal olarak değerlendirilebilmektedir. Bununla birlikte, lisans mezunlarının görev aldığı teknolojik iş ortamlarının ve karmaşık iş süreçlerinin de iş kazası sayısını artırabileceği düşünülebilir. İşyeri büyüklüğü açısından değerlendirildiğinde, 1000 ve üzeri sigortalı çalışanı bulunan büyük ölçekli işletmelerde iş kazası sayısının daha yüksek olduğu görülmektedir. Bu bulgu, büyük işletmelerde daha fazla iş gücünün bulunması ve iş süreçlerinin karmaşıklığı ile ilişkili olabilir. Öte yandan, iller düzeyinde Adana, Konya ve Yalova gibi illerde pozitif korelasyon değerlerinin yüksek olması, sanayi yoğunluğunun ve sektörel dağılımın kaza sayıları üzerindeki etkisini ortaya koymaktadır.

Metal sektörüne ilişkin gerçekleştirilen korelasyon analizi sonuçlarına göre, ölüm sayısı değişkeni ile saat 00:00–08:00 aralığı, Çarşamba günü ve kazazede yaşı 45–54 arasında en yüksek pozitif korelasyon değerleri gözlemlenmiştir. Bu bulgular, özellikle gece vardiyası koşullarında ve hafta ortasında gerçekleşen iş kazalarının ölümle sonuçlanma olasılığının

daha yüksek olabileceğine işaret etmektedir. Metal sektöründe iş kazası sayısının artışında incelenebilecek bir faktör olan Pazartesi etkisinin sektördeki ağır iş tanımları, kanunen zorunlu İSG uygulamaları ve hafta başı mesai kontrollerinin etkisiyle Pazartesi günü yerine, işin monotonluğu ve yorgunluk nedeniyle Çarşamba gününe kaydığı sonucuna varılabilir. Bunun nedeni, denetim ve kontrollerin hafta başında yoğun olması, işin monotonluğu ve yorgunluk nedeniyle hafta ortasında iş kazası riskinin artmasıdır. Bu durumu önlemek amacıyla, hafta başında yapılan kontrol ve denetimlerin hafta ortasında da yapılması önerilebilir. Ayrıca, 45–54 yaş aralığındaki çalışanların, fiziksel efor, iş yükü ya da dikkat düzeyi gibi faktörlere bağlı olarak ölümle sonuçlanan iş kazalarına daha fazla maruz kalabileceği anlaşılmaktadır. İl bazında yapılan değerlendirmelerde ise, Kahramanmaraş, Adıyaman, Ankara ve Bolu illerinin ölümle sonuçlanan iş kazaları ile pozitif yönde ilişkili olduğu tespit edilmiştir. Bu illerdeki iş kazalarının nedenleri, sektörel yoğunluk, iş güvenliği uygulamalarının seviyesi ve çalışma koşulları gibi değişkenler üzerinden detaylı biçimde incelenmeye açıktır.

İnşaat sektöründe yapılan korelasyon analizi bulgularına göre iş kazası sayısı ile en yüksek pozitif korelasyon, kazazede yaşı (14–24) değişkeninde gözlenmiştir. Bu bulgu, madencilik ve metal sektörüne benzer şekilde genç işçilerin sektörde daha fazla risk altında olduğunu ve deneyim eksikliğinin kazalara zemin hazırladığını ortaya koymaktadır. Ayrıca, işyerinde çalışan sigortalı sayısı (1000+) ve İSG eğitimi var değişkenleriyle olan pozitif korelasyon, büyük ölçekli işletmelerde hem kazaların daha sık yaşandığını hem de bu kazaların daha sistematik biçimde raporlandığını göstermektedir.

İnşaat sektöründe ölümle sonuçlanan iş kazalarına ilişkin yapılan korelasyon analizinde, kazazede yaşı 14–24, 25–34 ve mesleki eğitim alınmamış değişkenleri ile en yüksek pozitif korelasyonlar tespit edilmiştir. Bu bulgular, ölümle sonuçlanan kazaların özellikle genç ve orta yaş gruplarındaki çalışanlar arasında daha yaygın olduğunu ve mesleki eğitim eksikliğinin ölümcül iş kazası riskini artıran önemli bir faktör olabileceğini göstermektedir. Mesleki bilgi ve becerilerin yetersizliği, güvenlik protokollerinin doğru uygulanamaması gibi nedenlerle bu çalışan gruplarının daha fazla risk altında olduğu değerlendirilmektedir. Coğrafi dağılım açısından bakıldığında ise, Bolu, Sinop ve Malatya illerinin, ölümle sonuçlanan iş kazalarıyla pozitif yönde anlamlı ilişkiler gösterdiği saptanmıştır. Bu illerdeki sektörel yapı, iş güvenliği kültürü, denetim sıklığı ve çalışan profili gibi etkenlerin, ölümlü iş kazalarının oluşumuna katkı sunabileceği düşünülmektedir.

Bu korelasyon analizinin bulguları, oluşturulan tahmin modelleri açısından da yol gösterici niteliktedir. Özellikle kazazede yaşı, öğrenim durumu, İSG eğitimi durumu ve işyeri büyüklüğü gibi değişkenlerin iş kazası ve ölüm sayıları ile olan ilişkilerinde anlamlı örüntüler ortaya koymuştur. Bu gibi değişkenlerin tahmin modellerinde dikkate alınması, modelin açıklayıcılığını artırmakta ve daha anlamlı sonuçlar elde edilmesini sağlamaktadır. Bu bağlamda, korelasyon analizi yalnızca değişkenler arası ilişkileri değil, aynı zamanda hangi değişkenlerin modelleme sürecinde daha etkili olabileceğini belirlemede de önemli bir araç işlevi görmüştür.

5.5.2. Geleneksel yöntemler ve makine öğrenmesi algoritmaları bulgularının karşılaştırmalı yorumlanması

Tez çalışması kapsamında ele alınan sektörlerdeki iş kazası ve ölüm sayılarının tahminine yönelik geliştirilen modellerin performans sonuçları derlenmiş ve model tahmin performanslarının karşılaştırılması amacıyla madencilik sektöründeki iş kazası ve ölüm sayısı tahminine ilişkin sonuçlar Çizelge 5.29’te, metal sektöründeki iş kazası ve ölüm sayısı tahminine ilişkin sonuçlar Çizelge 5.30’da, inşaat sektöründeki iş kazası ve ölüm sayısı tahminine ilişkin sonuçlar Çizelge 5.31’de verilmiştir.

Çizelge 5.29. Kullanılan modellerin madencilik sektörü performans kıyaslaması

Tahmin tipi	Model	MAPE	MSE	R ²
İş kazası sayısı	Çoklu doğrusal regresyon	0,154	86 807,73	-0,265
	LASSO Regresyonu (LASSO_Model_2)	0,098	23 624,21	0,655
	Rastgele Orman (Rastgele_Orman_Model_6)	0,096	29 724,24	0,566
	KNN (KNN_Model_5)	0,097	30 653,94	0,553
	DVM (DVM_Model_2)	0,082	24 965,00	0,636
Ölüm Sayısı	Çoklu doğrusal regresyon	1,619	3 798,42	-0,058
	LASSO Regresyonu (LASSO_Model_4)	0,764	3 669,85	-0,022
	Rastgele Orman (Rastgele_Orman_Model_2)	0,271	3 423,32	0,045
	KNN (KNN_Model_2)	0,198	3 419,13	0,047
	DVM (DVM_Model_3)	0,914	3 634,42	-0,013

Madencilik sektörü verilerine dayalı yapılan modelleme çalışmasında, iş kazası sayısı tahminleri orta düzeyde doğrulukla elde edilebilmiş, ancak ölüm sayısı tahminlerinde ciddi yetersizlikler gözlemlenmiştir. İş kazası sayısı tahminlerinde en başarılı model LASSO regresyon modeli olmuştur. DVM modeli de LASSO regresyona çok yakın sonuçlar vermiştir. ($R^2 = 0,636$, MAPE = %8,2). Bu modeller, sektörel değişkenlerle kazalar arasında güçlü ve doğrusal bir ilişkiyi temsil edebilmektedir. Bu modellerin ardından rastgele orman

ve KNN modelleri birbirlerine çok yakın sonuçlar vermiştir. Çoklu Doğrusal Regresyon modeli ise başarısız olmuştur. Ölüm sayısı tahminleri açısından bakıldığında tüm modeller genel olarak başarısız olmuştur. Rastgele orman ve KNN modelleri birbirlerine yakın pozitif R^2 değerleri ile tahmin yapabilmiş ancak yine de modeller yetersiz bir performans göstermiştir.

Çizelge 5.30. Kullanılan modellerin metal sektörü performans değerlerinin kıyaslaması

Tahmin tipi	Model	MAPE	MSE	R^2
İş kazası sayısı	Çoklu doğrusal regresyon	0,099	612 999,96	0,843
	LASSO Regresyonu (LASSO_Model_3)	0,072	403 696,77	0,896
	Rastgele Orman (Rastgele_Orman_Model_5)	0,099	695 422,25	0,821
	KNN (KNN_Model_5)	0,089	411 406,17	0,894
	DVM (DVM_Model_1)	0,088	464 406,97	0,881
Ölüm Sayısı	Çoklu doğrusal regresyon	0,863	22,88	0,183
	LASSO Regresyonu (LASSO_Model_2)	0,483	13,71	0,510
	Rastgele Orman (Rastgele_Orman_Model_5)	0,340	12,90	0,539
	KNN (KNN_Model_5)	0,310	12,05	0,569
	DVM (DVM_Model_4)	0,345	22,67	0,190

Metal sektörü verileri üzerinde uygulanan tahmin modelleri, genel olarak iş kazası sayısının tahmini konusunda başarılı bir performans sergilemiş, ölüm sayısı tahminlerinde ise orta düzeyde bir performans göstermiştir.

İş kazası sayısı tahmini açısından en başarılı performansı, LASSO regresyon modeli göstermiştir. Bu model, %7,2'lik MAPE değeriyle en düşük tahmin hatasına sahiptir. Ayrıca 403 696,77 seviyesindeki MSE değeri ve 0,896'lık R^2 değeri ile veriyi açıklamada yüksek bir başarı sergilemiştir. Bu sonuçlar, LASSO regresyonun metal sektöründe iş kazalarının tahmini açısından hem hata oranı düşük hem de veriyle güçlü bir ilişki kurabildiğini ortaya koymaktadır. KNN ve DVM modelleri ise LASSO regresyona oldukça yakın bir performans sergilemiştir. KNN modelinin %8,9'luk MAPE ve 0,894 R^2 değerleri ile doğruluğu yüksektir. DVM modelinin ise MAPE değeri %8,8 ve R^2 değeri 0,881'dir. Çoklu doğrusal regresyon ve rastgele orman modeli de metal sektöründe iş kazası tahmininde başarılı sonuçlar vermiştir. Rastgele orman modeli iş kazası sayısının tahmini konusunda başarılı olmakla birlikte değerlendirilen modeller arasında son sırada yer almıştır.

Ölüm sayısının tahmini için kullanılan geleneksel yöntemlerden çoklu doğrusal regresyon ve makine öğrenmesi yöntemlerinden DVM düşük R^2 değerleri ile yetersiz performans

sergilemiştir. Öte yandan, makine öğrenmesi yöntemlerinden KNN modeli R^2 (0,569) ve MAPE (0,310) değerleri ile tüm yöntemler arasında en iyi performansı gösteren yöntem olmuştur. KNN modelinin ardından sırasıyla rastgele orman modeli ve LASSO regresyonu en iyi performansı göstermiştir. Sonuç olarak KNN, rastgele orman ve LASSO regresyon modelleri metal sektöründe ölüm sayısının tahmininde orta seviyede bir performans göstermiştir.

Çizelge 5.31. Kullanılan modellerin inşaat sektörü performans değerlerinin kıyaslaması

Tahmin tipi	Model/Performans Metrikleri	MAPE	MSE	R^2
İş kazası sayısı	Çoklu doğrusal regresyon	0,154	679 401,71	0,610
	LASSO Regresyonu (LASSO_Model_4)	0,126	396 310,03	0,772
	Rastgele Orman (Rastgele_Orman_Model_1)	0,104	404 790,86	0,767
	KNN (KNN_Model_1)	0,119	421 228,50	0,758
	DVM (DVM_Model_2)	0,105	283 731,44	0,837
Ölüm Sayısı	Çoklu doğrusal regresyon	0,410	227,58	-0,040
	LASSO Regresyonu (LASSO_Model_3)	0,216	110,36	0,495
	Rastgele Orman (Rastgele_Orman_Model_1)	0,206	113,00	0,483
	KNN (KNN_Model_1)	0,232	97,90	0,552
	DVM (DVM_Model_2)	0,216	133,83	0,388

İnşaat sektörü verileri üzerinde uygulanan beş farklı tahmin yöntemi karşılaştırıldığında, tahmin başarılarının değişkenlik gösterdiği gözlemlenmiştir. Özellikle iş kazası sayısı tahminlerinde en yüksek doğruluk DVM modelinde gözlenmiştir ($R^2 = 0,837$, MAPE = %10,5). Bu model, doğrusal olmayan ilişkileri başarıyla temsil etmiştir. DVM modelinden sonra en iyi ikinci model LASSO regresyon modeli olmuştur. Rastgele orman ($R^2 = 0,767$ ve MAPE = 0,105) ve KNN ($R^2 = 0,758$ ve MAPE = 0,119) modelleri ise LASSO regresyona çok yakın sonuçlar elde etmiştir. Çoklu doğrusal regresyon modeli ise kullanılan yöntemler arasında en yetersiz model olmuştur.

Ölüm sayısının tahminine yönelik uygulanan farklı modellerin performans ölçütleri karşılaştırıldığında, en başarılı sonuçları KNN modeli vermiştir. Bu model, MAPE değeri olan 0,232 ve en yüksek R^2 değeri olan 0,552 ile değişkenler arasındaki ilişkiyi en iyi şekilde açıklamıştır. LASSO regresyon modeli (MAPE = 0,216, $R^2 = 0,495$) ve rastgele orman modeli (MAPE = 0,206, $R^2 = 0,483$) ise KNN modelinin ardından birbirlerine yakın olacak şekilde orta düzeyde performans göstermiştir. DVM algoritması (MAPE=0,216 ve $R^2=0,388$) bu modellerin ardından gelirken, çoklu doğrusal regresyon modeli ise negatif R^2 değeri ile verideki varyasyonu açıklamakta başarısız olmuştur. İnşaat sektöründe ölüm sayısı

tahmininde en başarılı model KNN modeli olmuştur.

Geliştirilen modeller ve yürütülen hesaplamalar sonucunda elde edilen iş kazası sayısı ve ölüm sayısı tahminleri ile bu tahminlere ilişkin performans ölçütleri, farklı geleneksel yöntemler ve makine öğrenmesi yöntemleri ile ele alınan probleme uygulanabilirliğini değerlendirmek amacıyla analiz edilmiştir. Bu sayede, iş kazası sayısı ve ölüm sayısı hesabında kullanılan yöntemlerin etkinlik ve güvenilirliği karşılaştırmalı olarak incelenmiş, her bir algoritmanın ilgili veri seti üzerindeki başarımı nesnel ölçütlerle ortaya konmuştur.

Bu tez çalışmasında incelenen gerçek hayat verileri doğrultusunda, makine öğrenmesi tabanlı modellerin veri işleme ve anlamlı sonuçlar üretme açısından geleneksel yöntemlere göre daha etkin olduğu görülmektedir. Özellikle makine öğrenmesi algoritmalarının sağladığı yüksek performans, tahmine dayalı analizlerde bu yöntemlerin klasik yaklaşımlara kıyasla daha başarılı ve güvenilir çıktılar sunduğunu göstermektedir. Bu durum, yalnızca modelleme başarısı açısından değil, aynı zamanda iş kazası sayısı ve ölüm sayısı tahminine yönelik literatüre de önemli bir katkı niteliği taşımaktadır.

İş kazası sayısı tahminlerinde genel olarak en iyi performans LASSO ve DVM modelleri tarafından elde edilmiştir. Makine öğrenmesi yöntemlerinden DVM modelinin performans metrikleri madencilik sektöründe ($R^2 = 0,636$, MAPE = %8,2), metal sektöründe ($R^2 = 0,881$, MAPE = %8,8) ve inşaat sektöründe ($R^2 = 0,837$, MAPE = %10,5) olarak elde edilmiştir. Rastgele orman ve KNN modelleri genel olarak daha düşük doğruluk sağlamakla birlikte yine de yeterli bir performans göstermiştir. Özellikle KNN modeli metal sektöründe en başarılı sonuçları vermiştir. Geleneksel yöntemlerden çoklu doğrusal regresyon yöntemi madencilik ve inşaat sektörlerinde makine öğrenmesi yöntemlerinin gerisinde kalmıştır. LASSO regresyonu ise metal ve madencilik sektörlerinde tüm yöntemler arasında en iyi performansı göstermiştir.

Ölüm sayısı tahminlerinde madencilik sektöründe tüm yöntemler yetersiz performans göstermiştir. Diğer sektörlerde ise KNN modeli metal sektöründe ($R^2 = 0,569$, MAPE = %31) ve inşaat sektöründe ($R^2 = 0,552$, MAPE = %23,2) değerleri ile kullanılan yöntemler arasında en başarılı yöntem olarak gözükmektedir. Ölüm olaylarının rastlantısal, düşük frekanslı ve çok değişken yapısı nedeniyle, geleneksel regresyon yöntemleri ve makine öğrenmesi yöntemleri kullanılan veri setiyle birlikte güçlü bir performansa ulaşamamıştır.

Ölüm sayısı gibi olayların tahmini geleneksel regresyon modelleriyle çoğu zaman mümkün olmamaktadır. Modellerin başarısı, modelin yetersizliğinden ziyade, verinin yapısal sınırlılıklarından kaynaklanmaktadır. Bu nedenle daha özelleştirilmiş, olay odaklı ve kararlı modelleme stratejilerine geçiş yapılması gerekmektedir.

Gerçekleştirilen çalışma neticesinde, yüksek doğruluk oranlarıyla tahmin edilen iş kazası sayıları sayesinde, işletmelerin İSG risklerini daha gerçekçi bir şekilde öngörebilmeleri ve buna uygun önleyici stratejiler geliştirebilmeleri mümkün hâle gelecektir. Bu öngörülebilirlik, iş gücü kesintilerinin minimize edilmesine, güvenli çalışma koşullarının tesis edilmesine, operasyonel sürekliliğin sağlanmasına, iş kazası kaynaklı maliyetlerin azaltılmasına katkı sunarak, iş kazalarının önüne geçilmesini sağlayacaktır.

5.5.3. İSG stratejileri

İş kazalarının önlenmesi, işletme verimliliği ve ekonomik sürdürülebilirlik açısından kritik bir öneme sahiptir. Bu tez çalışmasında elde edilen bulgular, iş kazalarının sektörel etkenlerle güçlü ilişkiler içinde olduğunu ve bu ilişkilerin veriye dayalı modeller aracılığıyla öngörülebilir hale getirilebileceğini göstermiştir. Bu doğrultuda, İSG uygulamalarında proaktif ve veri temelli yaklaşımların benimsenmesi önemlidir.

Özellikle bu tez çalışması kapsamında incelenen madencilik, metal ve inşaat sektörleri gibi iş kazası riski yüksek sektörlerde öncelikli olarak önleyici stratejilerin geliştirilmesi gereklidir. Bu stratejiler, model tahminlerine dayalı olarak riskli dönem ve alanların belirlenmesiyle desteklenmelidir. Örneğin, kazaların sıklıkla meydana geldiği zaman dilimleri, işletme türleri ya da çalışan profilleri tespit edilerek bu alanlara yönelik eğitim programları, teknik anlamda yapılabilecek iyileştirmeler ve denetim mekanizmaları uygulanabilir. Bu kapsamda, veriye dayalı olarak riskin yüksek olduğu durumlarda işletmelere önleyici müdahale imkânı sunulabilir.

Ayrıca, analizlerde kullanılan değişkenlerin çeşitliliği dikkate alındığında, İSG stratejilerinin yalnızca çalışma ortamıyla sınırlı kalmaması gerektiği ortaya çıkmaktadır. Çalışanların yaşı, eğitim düzeyi, mesleki geçmişi ve işletmedeki deneyimi gibi bireysel faktörlerin iş kazalarıyla ilişkili olduğu görülmektedir. Bu nedenle, çalışan profiline dayalı bireyselleştirilmiş eğitim programları, mesleki yeterliliklerin artırılması ve yeni işe alınan

personelin kademeli entegrasyonu gibi politikalar, iş kazalarının önlenmesinde önemli rol oynayacaktır.

Diğer yandan, ölüm olayları gibi nadir ve rastlantısal gelişen durumların mevcut modelleme teknikleriyle öngörülmesindeki sınırlılıklar, İSG stratejilerinde farklı türden yaklaşımların geliştirilmesi gerektiğini ortaya koymaktadır. Bu tür olaylara ilişkin analizler, vaka incelemeleri ve sektörel simülasyonlar aracılığıyla daha detaylı değerlendirmeler yapılabilir. Aynı zamanda bu durum, tehlike bildirim sistemlerinin yaygınlaştırılmasının ve kazaların öncesine dair belirtilerin sistematik olarak kaydedilmesinin önemini de göstermektedir.

Bu tez çalışması kapsamında elde edilen bulgulara göre genel ve sektörel bazda kısa, orta ve uzun vadeli İSG stratejileri geliştirilmiştir.

Kısa vadeli İSG stratejileri

- Çalışan profiline dayalı olarak eğitimler verilmelidir. Makine öğrenmesi yöntemleri ile tespit edilen kazaların yüksek olduğu gruplar için çalışan profiline uygun özel içerikli eğitim modülleri hazırlanmalı özellikle geçici ve yeni personel için zorunlu kısa oryantasyon programları ile desteklenmelidir.
- 6331 sayılı İş Sağlığı ve Güvenliği Kanunu'nda eğitim konusuna çok fazla değinilmekle birlikte yaşanan ve tahmin edilen iş kazası ve ölüm sayıları değerlendirildiğinde uygulamada denetim ile ilgili zayıflıkların bulunabileceği kanısı oluşmaktadır. İş kazalarının engellenememesinin en önemli nedenlerinden birinin denetim eksikliği olduğu değerlendirilmektedir. Sektörler özelinde kazaların yoğunlaştığı zaman, yaş ve eğitim profillerine göre denetimler artırılmalı, denetim verileri makine öğrenmesi modellerine geri besleme olarak entegre edilmelidir.
- 6331 sayılı kanunun 4. ve 5. maddeleri uyarınca işverenlerin tehlikeleri önleme ve risk değerlendirmesi yükümlülüğü çerçevesinde ve İş Sağlığı ve Güvenliği Risk Değerlendirmesi Yönetmeliği uyarınca makine öğrenmesi temelli modellerle oluşturulan sektörel risk haritaları kullanılarak, riskli bölgeler ve dönemler tespit edilmeli, bu alanlara özel otomatik denetim ve müdahale planları geliştirilmelidir.

Madencilik sektörü kısa vadeli İSG stratejileri

- Makine öğrenmesi ile tespit edilen kazaların yüksek olduğu gruplar için özel içerikli eğitim modülleri hazırlanmalıdır. Bu eğitimler, yer altı güvenliği, ekipman kullanımı ve acil durum prosedürleri gibi konuları kapsamalıdır. Ayrıca, yüksek riskli bölgelerde çalışanlar için, kazaların sık yaşandığı lokasyonlar ve vardiya saatlerine göre özelleştirilmiş eğitimler verilmelidir. Ayrıca, makine öğrenmesi tabanlı veri analizleri ile en sık karşılaşılan kazaların türleri belirlenerek, çalışanların bu tür acil durumlara yönelik eğitimleri güncellenmelidir.
- Madencilik sektöründe, İSG risklerini en aza indirmek için makine öğrenmesi tabanlı risk haritaları kullanılmalıdır. Bu haritalar kullanılarak, yer altı sahalarındaki kazaların sıklaştığı lokasyonlar, vardiya saatleri ve diğer faktörler analiz edilerek, yüksek risk taşıyan alanlar belirlenmelidir. Bu tespitler doğrultusunda, riskli bölgeler için özel denetim programları oluşturulmalı ve güvenlik prosedürleri daha sıkı uygulanmalıdır. Ayrıca, ekipmanlar detaylı şekilde denetlenmeli, çalışanlar daha fazla eğitime tabi tutulmalı ve yoğun vardiya saatlerinde ek denetimler yapılmalıdır.
- Makine öğrenmesi kullanarak madencilik ekipmanlarının performansı izlenebilir. Ekipmanların belirli bir süre veya yoğunlukla kullanımı sonrası arızalanma olasılıkları tespit edilerek, proaktif bakım ve güvenlik önlemleri alınmalıdır.

Metal sektörü kısa vadeli İSG stratejileri

- Metal sektöründe, çalışan sayısı ve üretim hacmine dayalı olarak oluşturulan risk skorları ile yüksek riskli tesisler için makine öğrenmesi algoritmaları ve İSG yönetim yazılımları kullanılarak otomatik risk raporları hazırlanmalıdır. Bu raporlar, tesislerdeki risk faktörlerini detaylı bir şekilde analiz ederek, belirli önlemler ve çözüm önerileri sunmalıdır. Ayrıca, bu raporlar düzenli olarak güncellenmeli ve güvenlik ekiplerine hızlı bir şekilde iletilerek, gerekli güvenlik önlemleri ve acil durum planları oluşturulmalıdır.
- Makine öğrenmesi algoritmalarıyla analiz edilen geçmiş verilerden, iş kazaları ve kaza şiddeti tahmin edilmelidir. Böylece, belirli durumlarda daha fazla dikkat ve önlem alınabilir. Örneğin, belirli makineler veya vardiya saatleri için yüksek riskli durumlar önceden öngörülerek gerekli önlemlerle kaza ve ölüm durumları engellenmelidir.

İnşaat sektörü kısa vadeli İSG stratejileri

- İnşaat sektöründe, makine öğrenmesi ile mevsimsel ve bölgesel risk analizleri yapılarak, bu risklere özel denetim ekipleri ve takvimleri oluşturulmalıdır. Örneğin, yağışlı mevsimlerde zemin kayması veya kışın donma riski gibi faktörler göz önünde bulundurulmalı ve her mevsime özgü yerel verilerle desteklenen tahminler ile güvenlik önlemleri alınmalıdır. Yerel hava koşulları, zemin yapısı ve geçmiş kazalar gibi veriler analiz edilerek, her bölgeye özel güvenlik planları ve takvimler hazırlanmalı, bu takvimlere göre denetimler düzenlenmelidir.

Orta vadeli İSG stratejileri

- Orta vadede İSG, sadece yasal zorunluluklar çerçevesinde değil, ISO 45001:2018 İSG Yönetim Sistemi Standardı da referans alınarak kurum kültürüne entegre edilmelidir. Tüm çalışanları kapsayan kurumsal güvenlik kültürü inşa edilmelidir. Çalışanların yalnızca eğitilmesi değil, aynı zamanda İSG süreçlerine aktif katılımı sağlanmalı, güvenli davranışlar teşvik edilmelidir.
- Bu dönemde ayrıca, kısa vadede sunulan eğitimlerin sürdürülebilirliğinin sağlanması hedeflenmelidir. Eğitim içerikleri, sektöre özel kaza desenlerine, iş yeri büyüklüğüne veya çalışanın pozisyonuna göre çeşitlendirilerek güncellenmelidir. 6331 sayılı Kanun'un 17. maddesi, çalışanlara düzenli eğitim verilmesini zorunlu kılarken, işverenin performans izlemesine dayalı bireysel eğitim yükümlülüklerini de tanımlamaktadır. Bu doğrultuda, çalışan performansının İSG bağlamında izlenmesi, eksiklik tespit edilen alanlara yönelik müdahalelerin zamanında yapılmasına olanak tanır.
- İSG stratejileri, sürekli olarak güncellenmeli ve gelişen verilere dayalı olarak iyileştirilmelidir. Modellerin tahminleri, teknik önlemler ve denetim sıklıklarının belirlenmesine rehberlik etmelidir.

Madencilik sektörü orta vadeli İSG stratejileri

- Madencilik sektöründe, İSG kurum kültürü, yeraltı ve tehlikeli alanlarda çalışan işçilere yönelik davranışsal güvenlik eğitimleriyle desteklenmelidir. Bu eğitimler, işçilerin riskli durumları daha iyi fark etmelerini, güvenlik prosedürlerine uymalarını ve güvenli

davranışlar geliřtirmelerini saęlamak amacıyla düzenlenmelidir. Eęitimlerin, teoriyle birlikte pratik uygulamalarla da desteklenmesi, güvenlik kùltürünün yerleřmesine ve iř kazalarının önlenmesine katkı saęlayacaktır.

- Makine öğrenmesi yöntemleri ile kazaların örüntüleri analiz edilerek, gelecekteki riskler tahmin edilmelidir. Bu tahminlere dayanarak, öncelikli teknik iyileřtirme alanları belirlenmeli, ekipman kontrolleri sıklařtırılmalı ve yüksek riskli bölgelerde düzenli bakım yapılmalıdır. Ayrıca, vardiya düzenlemeleri, çalışan yorgunluęu göz önünde bulundurularak optimize edilmeli ve daha güvenli çalışma saatleri planlanmalıdır.

Metal sektörü orta vadeli İSG stratejileri

- Metal sektöründe, kurumsal güvenlik kùltürü ISO 45001 standardına uygun süreçlerle güçlendirilmelidir. Bu süreçler, risk deęerlendirmeleri, güvenlik protokollerinin uygulanması ve çalışan eęitimi gibi temel unsurları içermelidir. Ayrıca, güvenli üretim hedefleri belirlenerek, güvenli çalışma alışkanlıkları gösteren çalışanlar ödüllendirilmelidir. Bu ödüllendirme, güvenlik bilincini artırıp ve çalışanları güvenli davranışları sürdürmeye teşvik eder.
- Yoęun üretim dönemlerinde artan kaza risklerini azaltmak için vardiya ve iř planlaması, makine öğrenmesi tahminlerine dayalı olarak yapılmalıdır. Makine öğrenmesi algoritmaları ile, geçmiş verileri analiz ederek, yüksek riskli dönemler tahmin edilmeli ve buna göre vardiya düzenlemeleri optimize edilmelidir.

İnřaat sektörü orta vadeli İSG stratejileri

- İnřaat sektöründe, İSG kùltürünün yaygınlařtırılması, sadece ana yüklenici firmalarla sınırlı kalmamalıdır; tařeron firmalar da bu kùltürün içine dahil edilmelidir. Tařeron firmalarla birlikte çalışan tüm personelin, İSG kurallarına uygun davranması, güvenlik eęitimlerine katılması ve güvenli iř yapma alışkanlıklarını benimsemesi saęlanmalıdır. Bu amaçla, tařeron firmalara yönelik özel güvenlik eęitimleri düzenlenmeli, iř güvenlięi denetimleri ve kontrolleri daha sık yapılmalıdır. Ayrıca, tařeronlarla yapılan sözleşmelerde, İSG gerekliliklerine uyum saęlanması bir şart olarak belirlenmeli ve bu kurallara uyulmadıęı takdirde gerekli yaptırımlar uygulanmalıdır.

- Makine öğrenmesi destekli analizler kullanılarak mevsimsel, hava ve saha koşullarına göre dinamik denetim planları oluşturulmalıdır. Makine öğrenmesi algoritmaları, farklı hava koşulları ve mevsimsel değişikliklerin iş güvenliği üzerindeki etkilerini belirleyerek, bu faktörlere göre denetim sıklıklarını ve önlemleri optimize etmelidir. Örneğin, yağışlı havalarda zemin kayması riski artarken, aşırı sıcaklarda çalışanlar için sıcak çarpması riski yükselir. Bu koşullara göre, güvenlik önlemleri ve denetimler artırılmalı, çalışanların uygun ekipman kullanması sağlanmalıdır.

Uzun vadeli İSG stratejileri

- İSG stratejilerinin sürdürülebilirliği ve kapsayıcılığı açısından veri temelli karar verme mekanizmalarının kurumsallaşması gerekmektedir. Bu çerçevede, yapay zeka destekli, saha sensörleri, ekipman verileri ve personel hareketliliğini izleyen gerçek zamanlı risk izleme sistemleri devreye alınmalıdır. Bu tür sistemler, iş yerlerinde saha sensörlerinden, iş ekipmanı verilerinden ve personel hareketliliğinden beslenerek proaktif önlemlerin alınmasını sağlayacaktır.
- İSG politikalarının dinamik hâle getirilmesi de uzun vadeli bir öncelik olmalıdır. Bu doğrultuda, modellerle güncellenen sektörel risk haritaları yardımıyla, her yıl yenilenen İSG planları oluşturulmalı; denetim stratejileri, risk skorlarına göre farklılaştırılmalıdır. Özellikle madencilik, metal ve inşaat gibi yüksek riskli alanlarda, mevzuatın ötesine geçen veri destekli uygulamalar hayata geçirilmelidir.
- Çalışan sağlığını sürekli iyileştirmeye yönelik uygulamalar da uzun vadede kalıcı hâle getirilmelidir. Özellikle, İSG profesyonellerinin etkinliği, psikososyal risklerin takibi, periyodik sağlık taramaları ve çalışan memnuniyeti analizleri ile davranışsal İSG boyutu kurumsal sistemlere entegre edilmelidir.
- İSG stratejileri, sektörler arası iş birlikleriyle geliştirilmeli ve küresel düzeyde iş kazalarını azaltmaya yönelik ortak standartlar oluşturulmalıdır. Sektörler arası iş birlikleri ve uluslararası standartlarla uyum, İSG alanında ulusal başarının küresel düzeyde sürdürülebilirliğini sağlayacaktır. Bu kapsamda, ISO 45001 ve AB İSG Çerçeve Direktifi (89/391/EEC) temel alınarak ortak standartlar geliştirilmeli, AB İSG uygulamaları ve ILO çalışma belgeleri doğrultusunda ulusal stratejiler yönlendirilmelidir.

Madencilik sektörü uzun vadeli İSG stratejileri

- Gaz, sıcaklık ve hareketlilik gibi sensör verileriyle gerçek zamanlı bir risk izleme altyapısı kurulmalıdır. Bu altyapı, IoT cihazları ile desteklenerek, sahadaki tüm veriler sürekli olarak toplanmalı ve merkezi bir sisteme iletilmelidir. Bu veriler, makine öğrenmesi modelleriyle entegre edilerek, potansiyel tehlikeler anlık olarak tespit edilmeli ve riskler analiz edilerek tahmin edilmelidir. Makine öğrenmesi, sensör verilerini analiz ederek gaz sızıntıları, aşırı sıcaklıklar veya tehlikeli bölgelerdeki hareketlilik gibi riskleri önceden belirlemelidir. IoT cihazları sayesinde tüm veriler anlık olarak izlenmeli ve bu sayede daha hızlı müdahale edilmesi sağlanmalıdır. Elde edilen bu veriler, kurumsal karar süreçlerine entegre edilerek, güvenlik önlemleri daha bilinçli ve etkin bir şekilde alınmalıdır.
- İSG kültürünün inşa edilmesi için, uzun vadede çalışan eğitimine odaklanılmalıdır. Eğitimler yalnızca başlangıç seviyesinde değil, sürekli gelişim temelli olmalıdır. Bu eğitimler, sanal gerçeklik ve artırılmış gerçeklik gibi teknolojilerle zenginleştirilerek, çalışanların gerçekçi ve interaktif bir ortamda güvenlik prosedürlerini öğrenmeleri sağlanmalıdır.

Metal sektörü uzun vadeli İSG stratejileri

- Çalışan yoğunluğu, iş ekipmanı verileri ve üretim zamanlamaları makine öğrenmesi yöntemleriyle entegre edilerek anlık tehlike izleme sistemleri kurulmalıdır. Bu sistemler, üretim sürecindeki her aşamayı izleyerek, çalışan yoğunluğuna bağlı güvenlik riskleri, ekipman arızaları ve üretim hızının neden olduğu yorgunluk gibi tehlikeleri tespit etmelidir. Makine öğrenmesi algoritmaları, bu verileri analiz ederek, potansiyel riskleri önceden tahmin etmeli ve anında uyarılar göndermelidir. Bu sayede, güvenlik ekipleri hızlı müdahale etmeli, kazalar önlenmeli ve üretim süreçleri daha güvenli hale getirilmelidir.
- Metal sektöründe uzun vadede, IoT ve RFID tabanlı entegre bir güvenlik ve izleme sistemi kurulmalıdır. Bu sistemde IoT cihazları ve RFID sensörleri, tesis içerisindeki ekipman, makineler ve çalışan hareketlerini gerçek zamanlı olarak izlemelidir. RFID etiketleri, çalışanların konumlarını ve kişisel koruyucu donanımların (baret, eldiven, gözlük vb.) kullanım durumunu sürekli takip ederek, İSG protokollerine uyulup

uyulmadığını denetlemelidir. Ayrıca IoT sensörleri, makinelerden sıcaklık, titreşim, basınç gibi kritik verileri toplayarak olası arızaları ve risk unsurlarını önceden tespit etmelidir. Bu veriler, makine öğrenmesi algoritmaları ile analiz edilerek ekipman arızalarının olasılığı tahmin edilmeli ve proaktif bakım süreçleri planlanmalıdır. Tüm bu veriler merkezi bir platformda toplanmalı ve yapay zekâ destekli analizler ile güvenlik ihlalleri, arıza riski veya acil durumlar için anlık uyarılar oluşturulmalıdır.

İnşaat sektörü uzun vadeli İSG stratejileri

- İnşaat sektöründe, makine öğrenmesi modelleri ile mevsimsel ve coğrafi risk tahminleriyle desteklenen, personel hareketliliği ve hava koşullarına bağlı anlık uyarı veren saha risk sistemleri geliştirilmelidir. Makine öğrenmesi algoritmaları, çevresel faktörleri, zemin durumu ve çalışan yoğunluğunu analiz ederek, riskli durumları tespit etmeli ve uyarılar göndermelidir. Ayrıca, hava koşulları ve saha koşulları ile ilgili veriler gerçek zamanlı olarak izlenmeli, bu verilerle güvenlik önlemleri zamanında alınmalıdır. Bu sistemler, iş kazalarını azaltarak, güvenli ve verimli çalışma ortamları sağlamalıdır.
- İnşaat sektöründe uzun vadede, IoT tabanlı entegre bir güvenlik ve izleme sistemi kurulmalıdır. IoT cihazları, inşaat alanındaki her türlü ekipman, araç ve çalışan hareketliliğini gerçek zamanlı olarak izlemeli, bu sensörler ekipmanlardan veri toplayarak, olası arızaları veya güvenlik açıklarını tespit etmelidir. Ayrıca, çalışanların giydiği güvenlik ekipmanlarının doğru kullanımı takip edilmeli, bu sayede güvenlik protokollerine uyulup uyulmadığı sürekli izlenmelidir. Tüm bu veriler merkezi bir sistemde toplanarak, güvenlik ihlalleri veya acil durumlar için anlık uyarılar gönderilmelidir.

Sonuç olarak, bu çalışmada geliştirilen veri temelli tahmin modelleri, İSG stratejilerinin bilimsel temellere dayandırılarak daha etkin hale getirilmesine olanak sağlamaktadır. İş kazalarının yalnızca gerçekleştiğinde değil, gerçekleşmeden önce tahmin edilerek yönetilmesi yaklaşımı benimsenmeli, sektörel risk haritaları, çalışan profili analizleri ve zamana dayalı trend tahminleri ile desteklenen dinamik İSG politikaları geliştirilmelidir. Böylece hem insan yaşamının korunması hem de ekonomik kayıpların azaltılması yönünde somut ve sürdürülebilir adımlar atılabilecektir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, Türkiye'de madencilik, metal ve inşaat sektörlerinde yaşanan iş kazalarının ve bu kazalara bağlı ölüm sayılarının tahmin edilebilmesi amacıyla farklı makine öğrenmesi yöntemleri ile geleneksel tahmin yöntemleri karşılaştırmalı olarak analiz edilmiştir. Çalışmada, 2014–2023 yılları arasındaki SGK verileri kullanılarak, ilgili sektörlerle ait iş kazası ve ölüm sayıları geçmiş veriler ışığında modellenmiş ve 2026–2028 yıllarına ilişkin tahminler gerçekleştirilmiştir. Modelleme sürecinde geleneksel yöntemlerden çoklu doğrusal regresyon ve LASSO regresyonu, makine öğrenmesi yöntemlerinden rastgele orman, KNN ve DVM gibi farklı algoritmalar kullanılmış; her bir modelin tahmin doğruluğu MAPE, MSE ve R^2 gibi performans metrikleriyle değerlendirilmiştir.

Gerçekleştirilen analizler sonucunda, iş kazası sayılarının tahmininde makine öğrenmesi yöntemlerinde en yüksek doğruluk genel olarak DVM algoritması tarafından sağlanmıştır. Madencilik, metal ve inşaat sektörlerinde elde edilen başarı düzeyi, bu modelin sektörel değişkenleri etkili bir biçimde işleyerek, iş kazalarının oluşma dinamiklerini yüksek isabetle öngörebildiğini göstermiştir. KNN ve rastgele orman yöntemleri ise genel olarak geleneksel yöntemlerden çoklu doğrusal regresyon yönteminden daha iyi sonuçlar vermiştir. LASSO ve çoklu doğrusal regresyon yöntemleri ise özellikle metal sektöründe tutarlı ve açıklayıcı sonuçlar üretmiştir.

Buna karşılık, ölüm sayısı tahminine yönelik oluşturulan modeller arasında en başarılı model KNN olarak öne çıkmaktadır. Madencilik sektöründe tüm yöntemler yetersiz bir performans sergilemişken, metal ve inşaat sektörlerinde KNN modeli, en iyi performansı gösterse de orta düzeyde bir doğruluk sağlamıştır. Çoklu doğrusal regresyon ise tüm sektörlerde yetersiz performans göstermiştir. Bu sonuçlar, ölüm olaylarının doğası gereği nadir, düzensiz ve öngörülmesi zor yapılar taşıdığını göstermektedir. Bu nedenle, bu çalışmada kullanılan veri setiyle hem geleneksel regresyon yöntemleri hem de makine öğrenmesi algoritmaları, ölüm sayısı gibi seyrek olayların tahmininde güçlü bir tahmin performansı sergileyememiştir.

Bu tez çalışmasında iş kazası ve ölüm sayılarına ilişkin tahmin modelleri geliştirilirken, analizler belirli yapısal ve metodolojik sınırlılıklar çerçevesinde ele alınmıştır. Veri setinde yalnızca SGK kayıtlarına dayalı sınırlı sayıda değişkenin bulunması, kazaların oluşumuna

etki edebilecek çevresel, operasyonel ve davranışsal faktörlerin modele dâhil edilmesini engellemiştir. Örneğin, çalışanların tecrübe düzeyi, ekipman kullanımı veya denetim sıklığı gibi değişkenler modele yansıtılamamıştır. Özellikle ölümle sonuçlanan kazalar gibi seyrek olayların düzensiz ve rastlantısal dağılım göstermesi, geliştirilen modellerin tahmin performansını önemli ölçüde sınırlamıştır. Son olarak, elde edilen bulgular yalnızca 2014–2023 dönemi için Türkiye’de SGK kapsamında kayıtlı olan madencilik, inşaat ve metal sektörlerini kapsamaktadır. Dolayısıyla, çalışmanın bulguları yalnızca kullanılan veri seti ve dönemiyle sınırlı olup, farklı sektörler, ülkeler veya veri kaynakları için doğrudan genellenemez niteliktedir.

Bu çalışmada elde edilen düşük MAPE, MSE ve R^2 değerleri modellerin tahmin performanslarının sınırlı olduğunu göstermektedir. Bu durum, modelde kullanılan değişkenlerin sayısı, veri kalitesi ve model parametrelerinin optimizasyonundaki kısıtlamalardan kaynaklanmış olabilir. Gelecek çalışmalarda, daha kapsamlı veri setlerinin kullanılması, değişken seçiminin iyileştirilmesi ve model hiperparametrelerinin optimize edilmesi yoluyla MAPE ve MSE değerlerinin düşürülmesi, R^2 değerinin ise artırılması hedeflenmelidir. Ayrıca, farklı makine öğrenmesi algoritmalarının denenmesi ve model karmaşıklığının dikkatli yönetilmesiyle daha yüksek tahmin doğruluğuna ulaşılabilir.

Çalışmanın bulguları doğrultusunda, iş kazalarının ve ölüm olaylarının tahmin edilmesine yönelik yapılacak ileri araştırmalar ve uygulamalar için aşağıdaki öneriler sunulmaktadır:

- İş kazası sayılarının tahmini için DVM, KNN, rastgele orman ve LASSO gibi algoritmalar kullanılabilir. Özellikle model seçimi yapılırken sektör bazlı veri yapısına göre tercih yapılmalıdır.
- Modellerin genelleme yeteneği, kullanılan veri setinin kapsamı ve çeşitliliği ile doğrudan ilişkilidir. Daha geniş ve çeşitli veri kümeleri ile çalışmak, modellerin tahmin doğruluğunu ve uyarlanabilirliğini önemli ölçüde artırabilir. Bununla birlikte, veri setine anlamlı ve açıklayıcı yeni değişkenlerin dahil edilmesi, modellerin tahmin kapasitesini daha da güçlendirecektir. Bu kapsamda mevcut veriye kazazedeyle ilgili (işletmedeki ve iş hayatındaki tecrübesi gibi) ve iş yeri ile ilgili farklı ek değişkenler entegre edilmesi ve kaza şiddeti sınıflandırmasının çeşitlendirilmesi ile modellerin performansı analiz edilebilir. Modellerin geçerliliğinin ve güvenilirliğinin artırılabilmesi adına, elde edilen tahminlerin hem gerçek yaşam verileriyle hem de geleneksel yöntemlerle karşılaştırılması

büyük önem taşımaktadır. Bu tür karşılaştırmalar, geliştirilen modellerin pratikteki başarısını ortaya koyarak, güvenilir bir tahmin aracı olup olmadığını somut biçimde değerlendirme olanağı sunar.

- Gelecek çalışmalarda modellerin performansını daha sağlıklı ve güvenilir bir biçimde ölçebilmek adına, çapraz doğrulama yöntemlerinin kullanılması faydalı olacaktır. Bunun yanı sıra, modellerin bağımsız ve daha önce görülmemiş test verileri üzerinde değerlendirilmesi, genelleme kapasitesinin daha net bir şekilde anlaşılmasına olanak tanıyacaktır.
- Bu çalışma kapsamında geliştirilen tahmin modelleri, iş kazası sayılarının önceden öngörülmesini mümkün kılarak, İSG alanında daha etkili stratejilerin oluşturulmasına katkı sağlamıştır. Özellikle iş kazası riskinin yüksek olduğu madencilik, metal ve inşaat sektörlerinde, makine öğrenmesi algoritmalarının tahmin gücü sayesinde, önleyici İSG politikalarının hedef odaklı biçimde planlanması mümkün hale gelmiştir. Sektörel iş kazası tahminleri, risk değerlendirme süreçlerinde önceden uyarı mekanizması olarak kullanılabilir. İş kazası riski yüksek sektörlerde önleyici eğitimler, denetimler ve teknik iyileştirmeler planlanarak kaynaklar daha verimli kullanılabilir.
- Daha büyük ve dengeli veri setlerinin kullanımı, farklı makine öğrenmesi modellerinin performansını da değerlendirme imkanı sunacaktır. Farklı ülkelerdeki iş kazası verileriyle karşılaştırmalı analizler yapılabilir, bu sayede evrensel risk desenleri ortaya konabilir.
- Belirtilen kısıtlamalar ve sunulan öneriler doğrultusunda, gelecekte yapılacak çalışmaların kapsamı genişletilerek elde edilen bulguların geçerliliği ve güvenilirliği artırılabilir. Bu unsurlar göz önünde bulundurularak geliştirilecek yeni araştırmalarda, daha kapsamlı veri setleri ve gelişmiş modelleme yaklaşımlarıyla daha tutarlı ve güçlü tahmin modelleri oluşturulması mümkün olacaktır.

Sonuç olarak bu tez çalışması, iş kazalarının veriye dayalı yöntemlerle modellenmesinin mümkün olduğunu ve makine öğrenmesi algoritmalarının bu alanda önemli avantajlar sunduğunu göstermiştir. DVM, KNN ve rastgele orman gibi algoritmalar, özellikle iş kazalarının ve ölüm sayılarının tahmininde geleneksel tahmin yöntemleri ile karşılaştırıldığında yüksek başarı göstermiştir. Bu çalışmada yalnızca var olan veriler üzerinden hem geleneksel ve modern yöntemler kullanılarak tahmin doğruluğunu artırıcı yöntemler geliştirilmiş hem de sektör uygulayıcılarına bir yol haritası sunulmuştur.



KAYNAKLAR

- Acar, G. (2022). *Zaman Serisi Yöntemleri Ve XGBoost Algoritması İle Kömür Satış Tahmini: Türkiye Tarım Kredi Kooperatifleri Uygulaması*, Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Acet, A. (2022). *SVM, NB, KNN, ADABOOST ve Random Forest Sınıflandırma Algoritmaları Kullanılarak Meme Kanserinin Tahmini*, Yüksek Lisans Tezi, İnönü Üniversitesi Fen Bilimleri Enstitüsü, Malatya.
- Adar, E. (2023). *Destek Vektör Makinesi ve Çekirdek Fonksiyonları Üzerine Bir Çalışma*, Yüksek Lisans Tezi, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.
- Akarsu, C.H. (2024). *Konteyner Yükleme Problemleri İçin Tabu Arama ve Pekiştirmeli Öğrenme Tabanlı Bir Hibrit Yaklaşım*, Doktora Tezi, İstanbul Üniversitesi-Cerrahpaşa Lisansüstü Eğitim Enstitüsü, İstanbul.
- Akdoğan, H.E. (2024). *Standart Süre Hesabında Makine Öğrenmesi Yöntemlerinin Kullanılması ile Üretim Planlama Faaliyetlerinin Etkinliğinin Arttırılması*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Akman, M. (2010). *Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama*, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Akman, M., Genç, Y., Ankaralı, H. (2010). Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama. *Türkiye Klinikleri Biostatik Dergisi*, 3(1), 36-48.
- Akpınar, H. (2017). *Data: Veri Madenciliği Veri Analizi* (2. Baskı). İstanbul: Papatya Yayıncılık.
- Aktan, E. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu. *Bilgi Yönetimi Dergisi*, 1(1), 1-22.
- Alkaissy, M., Arashpour, M., Golafshani, E.M., Hosseini, M.R., Khanmohammadi, S., Bai, Y., Feng, H. (2023). Enhancing construction safety: Machine learning-based classification of injury types. *Safety Science*, 162, 106102.
- Alkış, B.N. (2017). *Çoklu Lojistik Regresyon ve K-En Yakın Komşu Yöntemleri ile BİST100 Endeks Getiri Yönünün Tahmini*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Almadani, A. (2024). *Trafik Kazası Yaralanma Şiddeti Tahmini İçin Makine Öğrenmesi Yöntemlerinin Karşılaştırılması*, Yüksek Lisans Tezi, Bursa Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Bursa.
- Alpar R. (2010). Basit Doğrusal Regresyon Çözümlemesi. Spor, Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik Güvenirlik, *Detay Yayıncılık*, 285-304.

- Andreopoulos, B., An, A., Wang, X., Labudde, D. (2009). Efficient layered density-based clustering of categorical data. *Journal of Biomedical Informatics*, 42(2), 365-376.
- Arslan, S. (2022). *Temel Bileşenler Analizi ve K-Ortalamlar Yöntemi Kullanarak Ülkelerin Suç Analizi*, Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara.
- Atalay, M., Çelik, E. (2017). Büyük Veri Analizinde Yapay Zeka ve Makine Öğrenmesi Uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.
- Ayanoğlu, C.C., Kurt, M. (2019). Metal Sektöründe Veri Madenciliği Yöntemleri ile Bir İş Kazası Tahmin Modeli Önerisi, *Ergonomi*, 2(2), 78-87.
- Aydın, N., Kayhan, G. (2021). Jensen Shannon Mesafesi Temelli Uyarlanmış Bulanık C Ortalamalar Kümeleme Yöntemi. *Avrupa Bilim ve Teknoloji Dergisi*, 29, 58-64.
- Bağcı, B. (2020). Hareketli Ortalamalar ve Üssel Düzeltme Yöntemlerinin Tahmin Gücünün Artırılması: Türkiye’de Döviz Kuru Tahmini. *Turkuaz Uluslararası Sosyo-Ekonomik Stratejik Araştırmalar Dergisi*, 2(1).
- Balcıoğlu, Y.S. (2023), *Derin Öğrenme ile Görüntü İşleme: Endüstriyel Parça Üretiminin Derin Öğrenme Sayesinde Yüzey Hata Tespiti (Görüntüleme Üzerine Uygulama)*, Doktora Tezi, Gebze Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Kocaeli.
- Balcıoğlu, Y.S., Sezen, B. (2024). K- En Yakın Komşu (KNN) ve MLP Yöntemi ile Hibrit Bir Sistem: Trafik Akış Tahmini. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 26(4), 1801-1816.
- Bayraktar, B. (2019). *Rastgele Ormanlar Ve Aşırı Öğrenme Makineleri Teknikleri İle Oltalama Saldırılarının Tespiti*, Yüksek Lisans Tezi, İstanbul Üniversitesi-Cerrahpaşa, Lisansüstü Eğitim Enstitüsü, İstanbul.
- Bayraktar, B., Uyguçgil, H., Konuk, A. (2018). Türkiye Madencilik Sektöründe İş Kazalarının İstatistiksel Analizi. *Madencilik*, 57, 85-90.
- Bayram, M., Ünğan, M.C., Ardıç, K. (2017). The relationships between OHS prevention costs, safety performance, employee satisfaction and accident costs. *International Journal of Occupational Safety and Ergonomics*, 23(2), 285-296.
- Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press.
- Bhatia, N., Vandana. (2010). Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security*, 8 (2), 302-305.
- Bilim, A., Çelik, O.N. (2023). Estimating the possibility of workday loss accidents in road construction, *Građevinar*, 75, 1183-1192.

- Çağlayan, N. (2016). *Türkiye'nin Alt Sektörler Bazında Yapay Sinir Ağları İle Elektrik Enerjisi Tüketiminin 2030 Yılına Kadar Tahmini*, Yüksek Lisans Tezi, Atatürk Üniversitesi Fen Bilimleri Enstitüsü, Erzurum.
- Çakmak, H. (2019). *Güncel İş Kazaları Ham Verilerinin, Veri Madenciliği Süreci ile Analizi*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Camkurt, M.Z. (2007). İşyeri çalışma sistemi ve iş yeri fiziksel faktörlerin iş kazaları üzerindeki etkisi. *TUHİS İş Hukuku ve İktisat Dergisi*, 20(6), 80-106.
- Camliyurt, G., Park, Y., Kim, D., Kang W.S., Park, S. (2023). Machine Learning with Multi-Source Data to Predict and Explain Marine Pilot Occupational Accidents. *Journal of Marine Science and Technology*, 31, 348-364.
- Cascio, F.W. (1995). *Managing Human Resources, Productivity, Quality of Work Life, Profits*. (5). New York: McGraw Hills.
- Ceylan, H., Avan, M. (2012). Türkiyede'ki İş Kazalarının Yapay Sinir Ağları ile 2025 Yılına Kadar Tahmini. *International Journal of Engineering Research and Development*, 4(1), 46-54.
- Ceylan, H., Kaplan, A. (2024). Türkiye'de Şehir ve Ülke Perspektifinden Ölümlü İş Kazaları. *Kent Akademisi*, 17(1), 231-254.
- Chauhan, V.K., Dahiya, K., Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. *Artificial Intelligence Review*, 52, 803-855.
- Choi, J., Gu, B., Chin, S., Lee, J. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. *Automation in Construction*, 110, 102974.
- Cox, L. A. (2002). Data Mining and Causal Modeling of Customer Behaviors. *Telecommunication Systems*, 21(2), 349-381.
- Çavuş, A., Taçgın, E. (2016). Türkiye'de İnşaat Sektöründeki İş Kazalarının Sınıflandırılarak Nedenlerinin İncelenmesi. *Academic Platform Journal of Engineering and Smart Systems*, 4(2), 13-24.
- Delibalta M.S. (2020). Lojistik Yönetiminin Madencilik Sektörü İş Sağlığı Ve Güvenliği Üzerine Etkileri, *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 9(1), 480-487.
- Demir, E. (2009). *Metal İş Kolunda Meydana Gelen İş Kazaları ve İş Kazalarının Oluşturduğu Kayıpların Ekonomik Yönden Analizi*, Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Demir, N. (2022), *Makina İmalat Sektöründe İş Kazaları İstatistikleri ve İSG Uygulamaları*, Yüksek Lisans Tezi, İstanbul Esenyurt Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.

- Demirel, E. (2018). *Türkiye’deki İş Kazaları İstatistiklerinin Bazı Avrupa Ülkeleri ile Karşılaştırılması*, Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.
- Demiroğlu, O. (2023), *Madencilik Sektöründe Meydana Gelen İş Kazalarının ve Meslek Hastalıklarının İstatistiksel Analizi*, Yüksek Lisans Tezi, Hasan Kalyoncu Üniversitesi Lisansüstü Eğitim Enstitüsü, Gaziantep.
- Desticioğlu, B. (2015). *Türkiye’de Sektörel Bazda Gelecek Yıllar İçin İş Kazalarının Projeksiyonu*, Yüksek Lisans Tezi, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Deveci, E. (2020). *K-En Yakın Komşu Algoritması ile Ekonomik Özgürlük Endeksinin Tahmin Edilmesi: OECD Ülkelerinde Örnek Uygulama*, Yüksek Lisans Tezi, İstanbul Aydın Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.
- Doğan, B. (2008). *Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü İçin Bir Uygulama*, Doktora Tezi, Kadir Has Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766.
- Dündar, S., Bilim, A., Bilim N. (2024). Türkiye’deki yaralanmalı iş kazalarının sektörel dağılımı ve analizi, *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 13(1), 28-33.
- Dunn, J.C. (1974), A fuzzy relative ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybern*, 3, 32-57.
- Ercan, S. (2020). *Destek Vektör Makinelerini Kullanan Teknik Seçmeli Öneri Sistemi*, Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Eşidir, K.A., Gür, Y.E., Yoğunlu, A., Çubuk, M. (2022). Yapay Sinir Ağları (YSA) ve ARIMA Modelleri ile Türkiye’de Aylık Sıfır km Otomobil Satış Adetlerinin Tahmin Edilmesi. *Pamukkale Üniversitesi İşletme Araştırmaları Dergisi*, 9(2), 260-277.
- Ezugwu, A.E., Shukla, A.K., Nath, R., Akinyelu, A.A., Agushaka, J.O., Chiroma, H., Muhuri, P.K. (2021). Metaheuristics: a comprehensive overview and classification along with bibliometric analysis. *Artificial Intelligence Review*, 24, 4237-4316.
- Fahim, A.M., Saake, G., Salem, A.M., Torkey, F.A., Ramadan, M.A. (2008). DCBOR: A Density Clustering Based on Outlier Removal. *World Academy of Science, Engineering and Technology*, 45, 171-176.
- Fuentes, A.G., Serrano, N.M.B., Lasheras, F.S., Valverde, G.F., Sanchez, A.S. (2022). Work-related overexertion injuries in cleaning occupations: An exploration of the factors to predict the days of absence by means of machine learning methodologies, *Applied Ergonomics*, 105, 103847.

- Gholamy, A., Kreinovich, V., Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Reports (CS)*, 1209.
- Görgülü, V. (2021). *Rastgele Ormanlar Yöntemi ile Özellik Seçimi Kullanılarak Van İlinde Yaşayanların Trafikte Algı ve Tutumlarının Belirlenmesi*, Yüksek Lisans Tezi, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.
- Güller, S. (2024). *Bir Üretim İşletmesinde Yapay Sinir Ağları Ve Anfis Metoduyla Talep Tahminlemesi*, Yüksek Lisans Tezi, İstanbul Beykent Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.
- Hinneburg, A., Keim, D.A. (1998). An Efficient Approach to Clustering in Large Multimedia Databases with Noise, *KDD-98*, 58-65.
- İnternet: 5510 Sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu. (2006). Resmi Gazete, 26200 (16 Haziran 2006). URL: <https://www.mevzuat.gov.tr/MevzuatMetin/1.5.5510.pdf>. Son Erişim Tarihi: 13.06.2024.
- İnternet: 6331 sayılı İş Sağlığı Ve Güvenliği Kanunu. (2012). Resmi Gazete, 28339 (30 Haziran 2012). URL: <https://www.mevzuat.gov.tr/mevzuatmetin/1.5.6331.pdf>. Son Erişim Tarihi: 13.06.2024.
- İnternet: Das, S. (2024). Advantages and Disadvantages of Machine Learning, Ello., URL: <https://ello.io/advantages-and-disadvantages-of-machine-learning/> Son Erişim Tarihi: 13.06.2024.
- İnternet: Eurostat. (2012). *European Statistics on Accidents at Work (ESAW) Summary methodology*. URL: <https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/ks-ra-12-002>. Son Erişim Tarihi: 24.04.2025
- İnternet: Nefes, O. (2025). *Maden, Metal, İnşaat İş kazası ve ölüm veri seti (2014-2023)*, URL: <https://doi.org/10.5281/zenodo.15771579>. Son Erişim Tarihi: 02.06.2025.
- İnternet: Uluslararası Çalışma Örgütü. (2015). *Investigation of Occupational Accidents and Diseases A Practical Guide for Labour Inspectors*. URL: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_dialogue/@lab_admin/documents/publication/wcms_346714.pdf. Son Erişim Tarihi: 24.04.2025.
- İnternet: Uluslararası Çalışma Örgütü. (2023). Safety and health at work. Web: <https://www.ilo.org/global/topics/safety-and-health-at-work/lang--en/index.htm>. Son Erişim Tarihi: 24.04.2025.
- İnternet: Ülgen, E.K., (2021), Hiyerarşik Kümeleme, <https://www.veribilimiokulu.com/hiyerarsik-kumeleme/> Son Erişim Tarihi: 09.05.2025.
- Jolliffe, I.T. (2002). *Principal Component Analysis*. ABD: Springer New York.

- Kang, K., Ryu, H. (2019). Predicting types of occupational accidents at construction sites in Korea using random forest model. *Safety Science*, 120, 226-236.
- Kara, Y. (2000). *Hücresel Üretim Sistemi Tasarımında Kullanılan Yapay Zeka Teknikleri ile Sezgisel Yöntemlerin Karşılaştırılması ve Uygulamalı Analizleri*, Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Karakaya Özkan, E. (2023). *Metal Sektöründe Yaşanan İş Kazalarının İstatistiksel Olarak İncelenmesi Ve Makine Öğrenim Algoritmaları İle Tahmin Çalışmaları*, Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Karr, C.L., Freeman, L.M. (1999). *Industrial Applications of Genetic Algorithms*, ABD: CRC Press.
- Kavuk Salış, D. (2024). *Temel ve Teknik Analize Dayalı Makine Öğrenmesi ile Hisse Senedi Fiyat Tahmini: Rastgele Orman Sınıflandırması Yaklaşımı*, Yüksek Lisans Tezi, Gebze Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Kocaeli.
- Kızılkaya, Y.M., Oğuzlar, A. (2018). Bazı Denetimli Öğrenme Algoritmalarının R Programlama Dili ile Kıyaslanması. *Karadeniz*, 37, 90-98.
- Koç, K., Ekmekcioğlu, Ö., Gurgun, A.P. (2021). Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers. *Automation in Construction*, 131, 103896.
- Koç, K., Ekmekcioğlu, Ö., Gurgun, A.P. (2022). Accident prediction in construction using hybrid wavelet-machine learning. *Automation in Construction*, 133, 103987.
- Koç, K., Ekmekcioğlu, Ö., Gurgun, A.P. (2023). Determining susceptible body parts of construction workers due to occupational injuries using inclusive modelling. *Safety Science*, 164, 106157.
- Koç, K., Gurgun, A.P. (2022). Scenario-based automated data preprocessing to predict severity of construction accidents. *Automation in Construction*, 140, 104351.
- Koçali, K. (2022). İş Sağlığı ve Güvenliğinde “Ramak Kala” Terimi Üzerine Değerlendirme. *Avrasya Terim Dergisi*, 10(2), 34-39.
- Korkmaz, C., Kacar, İ. (2024). Modelleme ve Tahmin Amaçlı Veri Ön İşleme Yöntemlerinin Ürün Kurutma Örneği ile Açıklanması. *Tekirdağ Ziraat Fakültesi Dergisi*, 21(2), 482-500.
- Košútová, K. (2023). Analysis of occupational accidents in the land transport and transport sector, *Transportation Research Procedia*, 74, 1428-1435.
- Kölgeli, N. (2021). *Türkiye’de İnşaat Sektöründe, 2008-2019 Yılları Arasında Meydana Gelen İş Kazalarının Analizi*, Yüksek Lisans Tezi, İzmir Ekonomi Üniversitesi Lisansüstü Eğitim Enstitüsü, İzmir.

- Köseoğlu, S.C. (2013). *İş Kazaları ve Sonuçları Arasındaki İlişkinin Yapısal Eşitlik Modeli ile İncelenmesi ve Bir Uygulama*, Yüksek Lisans Tezi, Atatürk Üniversitesi Fen Bilimleri Enstitüsü, Erzurum.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. (2005). *Applied Linear Statistical Models*, 5th Edition, ABD: McGraw-Hill/Irwin.
- Luo, X., Li, X., Goh, Y.M., Song, X., Liu, Q. (2023). Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents. *Safety Science*, 163, 106138.
- Mammadov, A., Kazar, G., Koc, K., Tokdemir, O.B. (2023). Predicting Accident Outcomes in Cross-Border Pipeline Construction Projects Using Machine Learning Algorithms. *Arabian Journal for Science and Engineering*, 48, 13771–13789.
- Memiş, O. (2018). *Türkiye’deki İş Kazalarının Bazı Meslek Gruplarına Göre Kaza Sıklık Hızı ve Ağırlık Oranları İle Karşılaştırılması*, Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.
- Mermi, S. (2024). *Ridge Regresyonda Ridge Parametresi İçin Önerilen Tarama Yöntemine Dayalı Yeni Bir Tahmin Edici*, Doktora Tezi, Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü, Muğla.
- Nilsson, J.N. (2009). *The Quest for Artificial Intelligence*, ABD: Cambridge University Press.
- Özcan, A. (2021). Büyük Veri: Fırsatlar ve Tehditler. *TRT Akademi*, 6(11), 12-30.
- Özcan, S.G. (2019). *İnşaat Sektöründe İş Kazalarına Neden Olan Faktörlerin Çok Kriterli Karar Verme Yöntemleriyle Değerlendirilmesi*, Yüksek Lisans Tezi, Ankara Yıldırım Beyazıt Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Özdemir Aydın, M. (2022). *Accident Severity Prediction of Zonguldak District Underground Coal Mines by Machine Learning Techniques*, Doktora Tezi, Ortadoğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Öztürk, T. (2019). *İş Kazalarına Etki Eden Faktörlerin Çok Değişkenli İstatistiksel Yöntemlerle Değerlendirilmesi*, Doktora Tezi, İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Pelit, Ö. (2023). *Türkiye’de 2007-2020 Yılları Arasında Metal Sektöründe Meydana Gelen İş Kazalarının İstatistiksel Analizleri*, Yüksek Lisans Tezi, Ağrı İbrahim Çeçen Üniversitesi Lisansüstü Eğitim Enstitüsü, Ağrı.
- Petekci, A.R. (2021). *Teorik Olarak Makine Öğrenimi*, Türkiye: KDY.
- Powers, D.A., Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*, Londra: Academic Press.

- Razbonyalı, C. (2016). *Big Data (Büyük Veri) ve Geleneksel Veri Saklama ve İşleme Yöntemlerine Etkisi Üzerine Bir Araştırma*, Doktora Tezi, Okan Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Recal, F. (2022). *Ulusal İş Kazaları Verisinden Karar Destek Sistemi Oluşturmada Makine Öğrenmesi Yaklaşımı*, Doktora Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Rokach, L., Maimon, O. (2005). *Clustering Methods. Data Mining and Knowledge Discovery Handbook*. Boston: Springer.
- Russel, S.J., Norvig, P. (2010). *Artificial Intelligence A Modern Approach*, ABD: Prentice Hall Series.
- Sakallıoğlu, S., Kaçıranlar, S. (2008). A new biased estimator based on ridge estimation, *Stat Papers*, 49, 669-689.
- Sanchez-Lite, A., Fuentes-Bargues, J.L., Iglesias I., Gonzalez-Gaya, C. (2024). Proposal of a workplace classification model for heart attack accidents from the field of occupational safety and health engineering. *Heliyon*, 10, e37647.
- Sarkar, S., Raj, R., Vinay, S., Maiti, J., Pratihari, D.K. (2019a). An optimization-based decision tree approach for predicting slip-trip-fall accidents at work. *Safety Science*, 118, 57-69.
- Sarkar, S., Vinay, S., Raj, R., Maiti, J., Mitra, P. (2019b). Application of optimized machine learning techniques for prediction of occupational accidents. *Computers and Operations Research*, 106, 210-224.
- Sena, I., Braga, A.C., Novais, P., Fernandes, F.P., Pacheco, M.F., Vaz, C.B., Lima, J., Pereira, A.I. (2024). Exploring Features to Classify Occupational Accidents in the Retail Sector, Optimization, Learning Algorithms and Applications. *OL2A 2023*, 49-62.
- Şengöz, N., Özdemir, G. (2016). Temel Bileşenler Analizi ve K-Ortalama Kümeleme Yönteminin Birlikte Kullanımı: Bir Örnek Uygulama. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(15), 85-94.
- Sevim, A. (2023). *Türkiye’de 2014-2021 Yılları Arasında Eğitim Alanında Yaşanan İş Kazalarının Analizi*, Yüksek Lisans Tezi, Şırnak Üniversitesi Lisansüstü Eğitim Enstitüsü, Şırnak.
- Seylan, R. (2021). *Gıda Sektöründe Meydana Gelen İş Kazaları ve Meslek Hastalıklarının İstatistiksel Analizi*, Yüksek Lisans Tezi, Hasan Kalyoncu Üniversitesi Lisansüstü Eğitim Enstitüsü, Gaziantep.
- Şimşek Akın, D. (2021). *Türkiye’de Son 10 Yılda Görülen İş Kazalarının Çalışma Saati, Cinsiyet ve Yaşa Göre İncelenmesi*, Yüksek Lisans Tezi, İzmir Demokrasi Üniversitesi Sağlık Bilimleri Enstitüsü, İzmir.

- Şimşek, R. (2021). *İnşaat Sektöründe İş Kazalarının Çok Kriterli Karar Verme Yöntemleriyle Bütünleşik Garrett Sıralama Tekniği İle Değerlendirilmesi*, Yüksek Lisans Tezi, Akdeniz Üniversitesi Fen Bilimleri Enstitüsü, Antalya.
- Subaşı, S., Beycioğlu, A., Çullu, M. (2010). Bulanık Mantık ve İstatistiksel Analiz Yöntemleri ile Revibrasyon Uygulanmış Betonlarda Basınç Dayanımı Tahmini. *SDU International Journal of Technologic Sciences*, 2(3), 46-52.
- Tanış, M.C. (2023). *İş Kazalarını Etkileyen Faktörlerin İstatistiksel Yöntemlerle İncelenmesi*, Yüksek Lisans Tezi, İstanbul Esenyurt Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.
- Toptancı, Ş. (2021). *İnşaat Sektöründe İş Kazalarının Tahminlemesi, Risk Analizi ve Kalite Evi ile Alınacak Tedbirlerin Önceliklendirilmesi*, Doktora Tezi, Eskişehir Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Eskişehir.
- Tuna, G. (2017). *Türkiyede Meydana Gelen İş Kazalarının Veri Madenciliği Kümeleme Modelinde Graf Teori ile Analizi*, Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Tunçman, B. (2024). *Tekstil Sektöründe İş Kazalarına Neden Olan Risk Faktörlerinin Veri Madenciliği Yöntemleriyle Değerlendirilmesi*, Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü, Bursa.
- Turan, A.K., Polat, H. (2024). Yarı denetimli makine öğrenmesi yöntemini kullanarak müzik türlerinin tespiti. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 12(1), 92-107.
- Türk Mühendis ve Mimar Odaları Birliği Makina Mühendisleri Odası. (2015). İşçi Sağlığı ve İş Güvenliği. *MMO/617*, 2015.
- Türkan, M., Bozdağ, A., Karkınlı, A. E., & Ulucan, A. G. (2023). Kent ölçeğinde konutlara ilişkin toplu değer değişiminin makine öğrenim algoritmaları ile analizi. *Türkiye Arazi Yönetimi Dergisi*, 5(2), 66-77.
- Ünal, İ.S. (2019). *Türkiye'deki İş Kazalarının Analizi ve Avrupa'daki Durum ile Karşılaştırılması*, Yüksek Lisans Tezi, Tarsus Üniversitesi Lisansüstü Eğitim Enstitüsü, Mersin.
- Vergili, M. (2023). *Doğrusal Regresyonda Ridge, Lasso Ve Elastik Net Yöntemlerinin Sağlık Alanında Uygulanması*, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Sağlık Bilimleri Enstitüsü, Isparta.
- Yağımlı, M., İzci, F.B. (2017). Türkiye'de Makine ve Teçhizatı Hariç Fabrikasyon Metal Ürünleri İmalatı Sektöründe Yaşanan İş Kazaları ve Ölümlü İş Kazası Sayılarının Tahmini. *Karaelmas İş Sağlığı ve Güvenliği Dergisi*, 1(1), 9-15.
- Yeni, A. (2024). *Bir Otomotiv Firmasında Bireysel Performans Puanlarının Yapay Sinir Ağı Yaklaşımı ve Xgboost Yöntemi ile Tahminlenmesi*, Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.

- Yiğit, U., Kazar, G. (2025). Project characteristics-based predicting the likelihood of occupational accidents in public school maintenances using a topological approach. *Safety Science*, 184, 106764.
- Yılmaz, A. (2022). *Yapay Zeka*. İstanbul: Kodlab Yayın Dağıtım Yazılım ve Eğitim Hizmetleri San. ve Tic. Ltd. Şti.
- Yüksel, F.Ş. (2023). *Türkiye'nin Havayolu Taşıyıcı Modellerine Göre Yolcu Talebinin Çoklu Doğrusal Regresyon, ANFIS ve YSA Teknikleri ile Tahminlenmesi*, Doktora Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.
- Zaman, T. (2017). *Çoklu Doğrusal Regresyon Analizinde Çakı Tahmin Yöntemi ve Test Problemlerine Katkılar*, Doktora Tezi, T.C. On Dokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.
- Zermane, A., Tohir, M.Z.M., Zermane, H., Baharudin, M.R., Yusoff, H.M. (2023). Predicting fatal fall from heights accidents using random forest classification machine learning model. *Safety Science*, 159, 106023.
- Zhu, R., Hu, X., Hou, J., Li, X. (2021). Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection*, 145, 293-302.
- Zorlu, Ö. (2022). *Mersin İline Ait Kıyı Alanlarının Zamansal Değişim Oranının Rastgele Orman Algoritması Kullanarak İncelenmesi*, Yüksek Lisans Tezi, Mersin Üniversitesi Fen Bilimleri Enstitüsü, Mersin.



Gazili olmak ayrıcalıktır