T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ULUSAL İŞ KAZALARI VERİSİNDEN KARAR DESTEK SİSTEMİ OLUŞTURMADA MAKİNE ÖĞRENMESİ YAKLAŞIMI

Füsun RECAL

DOKTORA TEZİ
Endüstri Mühendisliği Anabilim Dalı
İş Güvenliği ve İşçi Sağlığı Doktora Programı

Danışman
Doç. Dr. Tufan DEMİREL

Haziran, 2022

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ULUSAL İŞ KAZALARI VERİSİNDEN KARAR DESTEK SİSTEMİ OLUŞTURMADA MAKİNE ÖĞRENMESİ YAKLAŞIMI

Füsun RECAL tarafından hazırlanan tez çalışması 02.06.2022 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı İş Güvenliği ve İşçi Sağlığı Doktora Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. Tufan DEMİREL
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Tufan DEMİREL, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Ali Fuat GÜNERİ, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Semih ÖNÜT, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Çağlar AKSEZER, Üye
Işık Üniversitesi _____

Dr. Öğr. Üye. Selden CEPNİ, Üye
Işık Üniversitesi _____

Danışmanım Doç. Dr. Tufan DEMİREL sorumluluğunda tarafımca hazırlanan Ulusal İş Kazaları Verisinden Karar Destek Sistemi Oluşturmada Makine Öğrenmesi Yaklaşımı başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Füsun RECAL

İmza

Bu tez kızım Defne'ye ve gelecekte eşit şartlarda, özgürce üretmelerini dilediğim tüm kız çocuklarına ithaf edilmiştir.

TEŞEKKÜR

Tez çalışmamın başından sonuna kadar, bana motivasyonumda en çok ihtiyaç duyduğum morali sağlayan, yakınlığı ve hoşgörüsü ile destek veren, bana bu yolculukta liderlik eden saygıdeğer danışman hocam Doç. Dr. Tufan DEMİREL'e katkılarından dolayı en içten teşekkürlerimi sunarım.

Lisans, lisansüstü ve doktora eğitim hayatımın her aşamasında, mesleki olarak bulunduğum noktaya gelmemde çok büyük emekleri ve katkıları olan, tez sürecimde hiçbir zaman desteğini esirgemeyen, öğrencisi olmaktan onur duyduğum, değerli hocam Prof. Dr. Çağlar AKSEZER'e teşekkürlerimi sunarım.

Tez izleme jürimde yer alan değerli hocam Prof. Dr. Ali Fuat GÜNERİ'ye kıymetli önerileri ve katkıları için teşekkürlerimi sunarım.

Tez sürecimde tıkandığım yerlerde bana vakit ayırıp yol gösteren Doç. Dr. Gürkan Emre GÜRCANLI'ya kıymetli katkılarından ötürü teşekkürü borç bilirim.

Tez savunma sınavı jüri üyelerinden Prof. Dr. Semih Önüt'e ve geri bildirimleriyle bana destek olan Dr. Öğr. Üyesi Selden CEPNİ'ye ayrı ayrı teşekkürlerimi sunarım.

Ve elbette doktora çalışmalarım süresince bana her yönden, her durumda destek olan eşim Reşit RECAL'e, benden ümidini kesmeden beni sürekli motive eden anneme ve babama ve sabırla mezun olmamı bekleyen kızıma çok teşekkür ederim.

Tez sürecimde, hedefe odaklanmak yerine, öğrenmenin ve gelişmenin gereği olan sürece odaklanmamı sağlayan ve asıl başarının yolculuğu tamamlamak olmadığını, öğrenme yolculuğunda karşılaşılan engellere rağmen yola devam edebilme heyecanını hiç vazgeçmeden sürdürmek olduğunu bana sürekli fısıldayan iç sesime ve bu iç sesi besleyen herkese teşekkür ederim.

Füsun RECAL

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ x

TABLO LİSTESİ xi

ÖZET xiii

ABSTRACT xv

1 GİRİŞ 1

1.1 Literatür Özeti . . . 3

1.1.1 Türkiye'de ve Dünyada İş Kazaları . . . 3

1.1.2 Kaza Analizlerinde Veri Madenciliği Yöntemleri Uygulamaları . 7

1.2 Tezin Amacı . . . 21

1.3 Orijinal Katkı . . . 22

2 MAKİNE ÖĞRENMESİ YÖNTEMİ 23

2.1 Veri Ön İşleme . . . 23

2.2 Tahminleme . . . 26

2.2.1 Stokastik Gradyan Artırma (SGB) Algoritması . . . 26

2.2.2 Tahmin Doğruluğu Yaklaşımı . . . 28

2.2.3 Sınıf Dengesizliği Yaklaşımı . . . 34

3 BİRLİKTELİK KURALI ÇIKARIMI 35

3.1 Apriori Algoritması . . . 35

4 ÖN ARAŞTIRMA 40

4.1 Yöntem . . . 40

4.2 Bulgular . . . 43

4.3 Tez Çalışmasına Katkısı . . . 44

5 ÖNERİLEN YÖNTEM 45

5.1 Veri Ön İşleme 46
5.1.1 Veri Temizleme 48
5.1.2 Boyut İndirgeme 54
5.2 Tahminleme 58
5.2.1 Şiddet Gruplarının Oluşturulması 58
5.2.2 Makine Öğrenmesi ile Model Oluşturma ve Sınıf Tahminleme . 58
5.2.3 Önemli Değişkenlerin Belirlenmesi 62
5.3 Kural Oluşturma 73
5.3.1 Dolaylı Değişken Etkilerini İçeren Kaza Zincirlerinin Çıkarılması 74
5.3.2 Şiddet ve Yaralanma Sonucu İçeren Kaza Zincirlerinin Ortaya Çıkarılması 90
5.3.3 Kaza Zinciri Sorgulama 93

6 SONUÇ VE ÖNERİLER **100**

KAYNAKÇA **103**

A DEĞİŞKEN KATEGORİLERİ **113**

B KURALLAR **135**

TEZDEN ÜRETİLMİŞ YAYINLAR **139**

SİMGE LİSTESİ

$\hat{F}(x)$	Girdi değişken vektörü x ile ortaya çıkan sonucu çıktı değişkeni y'ye en iyi yaklaştıran fonksiyon
$L(y, F(x))$	Kayıp fonksiyonu
$F_0(x)$	Gradyan arttırma başlangıç fonksiyonu
r_{im}	i gözleminin m ağacındaki hata değeri (sözde rezidual)
R_{jm}	m ağacındaki j yaprağının hata değeri (rezidual)
γ_{jm}	m ağacındaki j yaprağının rezidual tahmin değeri
$F_m(x)$	m ağacının gradyan arttırma model Fonksiyonu
$F_M(x)$	M sayıda ağaca sahip gradyan arttırma model fonksiyonu

KISALTMA LİSTESİ

ANFIS	Uygulamalı Ağtabanlı Bulanık Çıkarım Sistemi (Adaptive-Network Based Fuzzy Inference Systems)
ANN	Yapay Sinir Ağları (Artificial Neural Network)
BN	Bayes Ağları (Bayesian Network)
C5.0	C5.0 Karar Ağacı
CART	Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees)
CBR	Vaka Temelli Akıl Yürütme (Case-Based Reasoning)
CHAID	Ki-kare Otomatik Etkileşim Algılama (Chi-square Automatic Interaction Detection)
DT	Karar Ağaçları (Decision Tree)
ELM	Aşırı Öğrenme Makinaları (Extreme Learning Machines)
ESAW	Avrupa İş Kazaları istatistiği
EUROSTAT	AB İstatistik Ofis
ILO	Uluslararası Çalışma Örgütü (International Labour Organization)
ILOSTAT	Uluslararası Çalışma İstatistikleri Veritabanı (International Labour Statistics Database)
İSG	İş Sağlığı ve Güvenliği
K-Means	K-Ortalama Kümeleme
KNN	K en yakın Komşu (K-Nearest Neighbor)
LCCA	Örtük Sınıf Küme Analizi (Latent Class Cluster Analysis)
LDA	Örtülü Diriklet Yerayırması (Latent Dirichlet Allocation)
LR	Doğrusal Regresyon (Linear Regression)
MAR	Rastgele Kayıp (Missing at Random)

MCAR	Tamamen Rastgele Kayıp (Missing Completely at Random)
MICE	Zincirleme Denklemlerle Çok Değişkenli Atama (Multivariate Imputation by Chained Equations)
MLR	Çok terimli Lojistik Regresyon (Multiple Linear Regression)
NACE	İşyeri Faaliyet Kodu
NB	Naif Bayes Ağları (Naive Bayes)
NN	Sinir Ağı (Neural Network)
OSHA	İş Sağlığı ve Güvenliği İdaresi (Occupational Safety and Health Administration)
PCA	Temel Bileşenler Analizi (Principal Component Analysis)
RBF	Radyal Tabanlı Fonksiyon (Radial Basis Function)
RF	Rassal Orman Algoritması (Random Forest)
SGB	Stokastik Gradyan Artırma (Stochastic Gradient Boosting)
SGK	Sosyal Güvenlik Kurumu
SOM	Özdüzenleyici Haritalar (Self Orginising Maps)
SVM	Destek Vektör Makineleri (Support Vector Machines)
WHO	Dünya Sağlık Örgütü

ŞEKİL LİSTESİ

Şekil 1.1 SGK verilerine göre yıllık ölümlü iş kazası sayıları 5
Şekil 1.2 SGK verilerine göre yıllık iş kazası sayıları 5
Şekil 2.1 Farklı eşik değerlere göre F_β skoru 31
Şekil 2.2 ROC Eğrisi 32
Şekil 2.3 Kesinlik-Duyarlılık Eğrisi 33
Şekil 3.1 Apriori algoritması akış şeması 39
Şekil 5.1 Metodoloji akış şeması 46
Şekil 5.2 Değişkenlerin tahmin gücü 55
Şekil 5.3 Dolaylı etkili faktörlerin etkilediği kazaların kaldıraç değeri dağılımları 75
Şekil 5.4 İşyeri özelinde kural sorgusu grafiği 96

TABLO LİSTESİ

Tablo 1.1 Literatür taraması özet tablosu - Gözetimli Makine Öğrenmesi . . . 13
Tablo 1.2 Literatür taraması özet tablosu - Gözetimsiz Makine Öğrenmesi . . . 18
Tablo 1.3 Önerilen Yöntem . . . 20
Tablo 2.1 Karışıklık matrisi . . . 29
Tablo 3.1 Birliktelik kuralları metrikleri . . . 36
Tablo 4.1 Ön araştırma veri seti değişken sayıları . . . 41
Tablo 4.2 Ön araştırma eğitim/test veri seti kaza sayıları . . . 42
Tablo 4.3 Ön araştırma optimum modeller . . . 42
Tablo 5.1 SGK veri tabanında yer alan değişkenler . . . 47
Tablo 5.2 Çalışmadan sayılan hallere ait iş kazası sayıları . . . 49
Tablo 5.3 Başlangıç veri setinde bulunan eksik veri yüzdeleri . . . 49
Tablo 5.4 CC ve SG_MESLEK_ISCO08 değişkenleri eksik veri sayıları . . . 51
Tablo 5.5 Değişkenlerin eksik veri sayıları . . . 52
Tablo 5.6 Eşleştirme öncesi ve sonrası eksik veri sayıları . . . 53
Tablo 5.7 Materyal/ekipman değişkenlerinin aynı seçilme yüzdeleri . . . 54
Tablo 5.8 Analiz sonuçlarına göre veriden silinen bağımlı değişkenler . . . 55
Tablo 5.9 Bağımlı değişken sayılarındaki değişimler . . . 56
Tablo 5.10 Şiddet gruplamasına göre kaza sayıları . . . 58
Tablo 5.11 Kaza karışıklık matrisi . . . 60
Tablo 5.12 Şiddet grupları tahmin performansları . . . 61
Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler . . . 63
Tablo 5.14 Dolaylı değişkenli kural sayıları . . . 75
Tablo 5.15 Etkili dolaylı değişkenler ve ortalama kaldıraç değerleri . . . 78
Tablo 5.16 Dolaylı değişkenlerin en etkili olduğu kaza zincirleri . . . 80
Tablo 5.17 Birden fazla dolaylı değişkenden etkilenen kaza zincirleri . . . 84
Tablo 5.18 Kural sayıları . . . 91
Tablo 5.19 İşyeri özelinde kural sorgusu sonuç tablosu . . . 97
Tablo A.1 NACE_K . . . 113
Tablo A.2 KAZA_IL_K . . . 115
Tablo A.3 CC_K . . . 117
Tablo A.4 SG_KAZAGERYER_K . . . 118

Tablo A.5 KAZA_KS_K 119
Tablo A.6 GENELFAA_K 120
Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K 121
Tablo A.8 OZELFAA_K 127
Tablo A.9 SG_KAZAGERORTAM_K 128
Tablo A.10 YS_KOD_K 129
Tablo A.11 SG_MESLEK_ISCO08SG_GOREVI_KOD 130
Tablo A.12 SG_CINSIYET 131
Tablo A.13 SG_MEDENIHAL 131
Tablo A.14 OGR_DURUMKOD 131
Tablo A.15 SAAT_ARALIK 132
Tablo A.16 KAZA_MEVSİM 132
Tablo A.17 YT_KOD 132
Tablo A.18 YVY_KOD 133
Tablo B.1 Dolaylı etkili değişkenlerin ortalama kaldıraç değerleri 135

ÖZET

Ulusal İş Kazaları Verisinden Karar Destek Sistemi Oluşturmada Makine Öğrenmesi Yaklaşımı

Füsun RECAL

Endüstri Mühendisliği Anabilim Dalı
Doktora Tezi

Danışman: Doç. Dr. Tufan DEMİREL

Her gün meydana gelen iş kazaları sonucu çalışanlar yaralanmakta, çalışamaz hale gelmekte ve hatta yaşamlarını kaybetmektedirler. İş kazalarının ülkelerin toplum yaşamları üzerindeki psikolojik ve yaşamsal etkileri yanı sıra ülke refahını etkileyen ekonomik sonuçları da literatürde yer almaktadır. Artan iş kazalarının önlenmesi için mevzuat çıkarma atılımlarına ek olarak İSG alanında daha geniş araştırma çalışmalarına ihtiyaç bulunmaktadır. İş kazalarının önlenmesinde doğru kararlar alınabilmesi için, iş kazalarının oluşumundaki etkenlerin belirlenmesi ve karar alıcılara sunulması önemli bir rol oynamaktadır. Bu etkenlerin belirlenmesinde kullanılabilecek en önemli kaynaklardan birisi ulusal çapta toplanan iş kazaları verisidir. Ancak, kazaları tanımlayan çok sayıda parametre ile birlikte toplanan veri boyutunun her geçen gün büyümesi, geleneksel istatistiksel yöntemlerin hesaplama yükünü arttırmış ve analizlerin uygulanmasını zorlaştırmıştır. Bu sebeple, büyük veri analizinde veri madenciliği adı altında disiplinlerarası bir alanda toplanan bazı teknikler, geleneksel istatistiksel yöntemlere güçlü bir alternatif oluşturmaktadır.

Bu tez çalışmanın amacı, karmaşık yapıdaki ulusal iş kazası ham verisini, iş kazalarını önlemede etkili karar alıcılara sunmak için bir karar destek sistemine dönüştürmektir. Önerilen yöntemde, kazanın şiddeti üzerinde etkili faktörlerin makine öğrenmesi ile belirlenmesi ve belirlenen faktörlerin oluşturduğu kaza zincirlerinin birliktelik kural çıkarımı yöntemi ile ortaya çıkarılması olmak üzere iki aşamalı bir veri madenciliği yöntemi geliştirilmiştir. Önerilen yöntem, 2013-2017 yılları arasında Türkiye'de meydana gelen 187.910 kazadan oluşan ulusal kaza veri setini kullanan bir uygulama

ile geliştirilmiştir. Önerilen yöntemin uygulamasında aynı veri seti 4 farklı ikili şiddet sınıflandırması kullanılarak analiz edilmiş ve bu şiddet sınıflandırmalarını içeren toplam 316.557 kaza zinciri ortaya çıkarılmıştır.

Ortaya çıkan kaza zincirlerinin oluşturduğu karar destek sistemi, hem ulusal düzeyde faaliyetlerin belirlenmesinde hem de işyerlerinde gerçekleştirilen risk değerlendirme çalışmaları esnasında karar alıcılara yol gösterici olacaktır. Ulusal veritabanına yeni kazalar eklendikçe de, önerilen yöntemin tekrarı ile, karar destek sistemi mevcut koşulları yansıtan bir şekilde karar vericilere sunulabilir.

Anahtar Kelimeler: İş kazaları, büyük veri, makine öğrenmesi, birliktelik kuralları, karar destek sistemi

ABSTRACT

Creating a Decision Support Framework from National Occupational Accidents Data with Machine Learning Approach

Füsun RECAL

Department of Industrial Engineering
Doctor of Philosophy Thesis

Supervisor: Assoc. Prof. Tufan DEMİREL

As a result of work accidents that occur every day, employees are injured, become unable to work and even lose their lives. Besides the psychological and vital effects of occupational accidents on the social lives of countries, the economic consequences that affect the welfare of the country are also mentioned in the literature. In addition to the legislative initiatives to prevent occupational accidents increase, there is a need for wider research studies in the field of OHS. In order to make the right decisions in the prevention of occupational accidents, determining the factors behind the occupational accidents and presenting them to decision makers plays an important role. One of the most important sources that can be used in the determination of these factors is the nationally collected work accident data. However, the increase in the size of the data collected and the large number of parameters describing the accidents, increased the computational load of traditional statistical methods and make the analysis difficult. For this reason, some techniques gathered in an interdisciplinary field under the name of data mining in big data analysis constitute a strong alternative to traditional statistical methods.

The aim of this thesis is to transform the complex national occupational accident raw data into a decision support system to present it to decision makers who are effective in preventing occupational accidents. In the proposed method, a two-stage data mining method has been developed: determining the factors affecting the severity of the accident with machine learning and revealing the accident chains formed by the determined factors with the association rule extraction method. The proposed method

developed with an application using a national accident dataset consisting of 187,910 accidents that took place in Turkey between 2013 and 2017. In the application of the proposed method, the same data set was analyzed using 4 different binary severity classifications, and at the end total of 316,557 accident chains containing these severity classifications were revealed.

The decision support system created by the resulting chains of accidents will guide the decision makers both in determining the activities at the national level and during the risk assessment studies carried out in the workplaces. As new accidents are added to national database, by repeating the proposed method, the decision support system can be presented to the decision makers in a way that reflects the current conditions.

Keywords: Occupational accidents, big data, machine learning, association rules, decision support system

1
GİRİŞ

ILO, dünya genelinde yılda en az 4 gün işten ayrılmayı gerektiren 330 milyonun üzerinde iş kazası meydana geldiğini ve bu kazalar sonucu günde yaklaşık 860.000 kişinin yaralandığını veya sağlığının zarar gördüğünü, 6.300 kişinin de öldüğünü tahmin etmektedir [1]. Bu durum dünyada gayri safi milli hasıla'nın ortalama yüzde 4'ünün kaybına neden olmaktadır. Bu kayıp zaman zaman sadece gelişmekte olan ülkelerde değil, aynı zamanda iyi gelişmiş ülkelerin ekonomilerinde de gayri safi milli hasılanın yüzde 10'una, hatta daha fazlasına ulaşabilmektedir [2].

İş kazalarının incelenmesi ve elde edilen çıkarımların sonucunda atılacak adımlarla engellenmesi amacıyla, ulusal ve uluslararası boyutlarda iş kazası verileri toplanmaktadır. Ancak, toplanan veri boyutunun büyümesi, çok değişkenli analiz veya doğrusal regresyon gibi geleneksel istatistiksel yöntemlerin hesaplama yükünü arttırmış ve analizlerin uygulanmasını zorlaştırmıştır. Literatürde, büyük verilerin analizinde veri madenciliği adı altında disiplinlerarası bir alanda toplanan bazı teknikler kullanılmaktadır ve kazaların analizinde de kullanımının son yıllarda arttığı görülmektedir.

Ülkemizde toplanan ulusal iş kazası verilerinden oluşan veri tabanı 2013 yılında ESAW standartlarına dönüştürülmüştür. Yeni düzenleme ile birlikte iş nedeni ile doğan tüm kazaların üç iş günü içinde çevrimiçi bir platform aracılığıyla Sosyal Güvenlik Kurumu'na (SGK) bildirme zorunluluğu gelmiştir. Çevrimiçi raporlamanın bir sonucu olarak, rapor edilen kaza sayısı artmıştır. Artan kaza kayıt oranları her bir kazayı tanımlayan çok sayıda parametre ile birlikte geniş ve karmaşık bir veri tabanı oluşturmaktadır. Toplanan bu ham veri SGK tarafından kamu ile paylaşılmamakta, sadece ham veriden çıkarılan yıllık özet istatistikler SGK çevirimiçi sayfasından yayınlanmaktadır. Ancak yayınlanan özet istatistikler ile kaza tipi, yaralanma tipi gibi detay istatistiklerin elde edilmesi mümkün olmamaktadır. Ulusal iş kazası veri tabanında yer alan büyük kaza verisinin sadece betimsel istatistik yıllıkları hazırlamak için kullanılması, iş kazası sayılarını düşürecek etkili önleyici eylemler oluşturmak için yeterli değildir. İş kazalarına ait büyük verinin, kazaların örtüşen özelliklerini ortaya

çıkaran ve önleme uygulamalarına ışık tutan çıkarımsal yöntemler kullanılarak analizi gerekmektedir. Veri madenciliği, iş kazalarının ardındaki faktörlerin tanımlanmasında ve karmaşık bir veri tabanından anlamlı bilgileri çıkarılmasında geleneksel yöntemlere sağlam ve güçlü bir alternatif oluşturmaktadır. Bu nedenle, bu çalışmada, ulusal iş kazaları verisinin bir bütün olarak ele alınması, iş kazalarının oluşumunda ve şiddetinde etkili faktörlerden kaza zincirlerinin oluşturulması hedeflenmektedir. Böylelikle ulusal iş kazası veri tabanının karmaşık yapısı, kaza zincirlerinden oluşan bir karar destek sistemi yapısına dönüşerek, ilgili devlet kurumları ve özel sektördeki karar alıcılara alınacak önlemler hususunda ışık tutacaktır.

Önerilen yöntem iki aşamadan oluşmaktadır. İlk aşamayı iş günü kaybı değişkenini kullanarak analist tarafından belirlenen ikili kaza şiddeti gruplamaları üzerinde etkili faktörlerin makine öğrenmesi yöntemi ile belirlenmesi oluşturmaktadır. Bu aşamada, önce kaza veri tabanında yer alan tüm kategorik değişkenlerin alt kategorileri 0 veya 1 değeri alan ayrı ayrı birer değişken olarak tanımlanmış ve sayısal değişkenlerle birlikte şiddet üzerine etkileri tahminleme gücü üzerinden makine öğrenmesi yöntemi ile değerlendirilmiştir. İkinci aşamada ise birinci aşamada ortaya çıkarılan güçlü değişkenlerin bulunduğu, başka bir deyişle 1 değeri aldığı kazalar ile belirli bir kaza şiddetliyle sonuçlanan kaza zincirleri birliktelik kuralları yöntemi ile ortaya çıkarılmaktadır. Önerilen yöntem 2013-2017 yılları arasında toplanan ulusal kaza verisi üzerinde uygulanarak geliştirilmiştir. Ancak, iş kazaları verisinin dinamik yapısına uygun olarak, önerilen yöntemi ulusal kaza verilerine yeni kazaların eklenmesiyle tekrarlamak ve güncel koşulları yansıtan yeni karar destek sistemleri oluşturmak mümkündür.

Çalışmanın ilk bölümünde literatüre iş kazaları ile ilgili çalışmalara odaklanarak geniş anlamda yer verilmiş ve incelenen çalışmalar ile ilgili özet tablolar oluşturulmuş, tezin amacı ve özgün katkısı ile ilgili açıklayıcı bilgiler verilmiştir. İkinci bölümde, önerilen yöntemde uygulanan makine öğrenmesi, veri önişleme aşamaları, stokastik gradyan arttırma (SGB) algoritması, tahmin doğruluğu ve sınıf dengesizliği yaklaşımı başlıklarıyla açıklanmaktadır. Üçüncü bölümde birliktelik kuralları çıkarımında uygulanan apriori algoritması anlatılmıştır. Dördüncü bölümde, geliştirilen yöntemde kullanılacak makine öğrenmesi algoritmasının araştırılması ve ikili üçlü sınıflandırma farkının incelenmesi amacıyla gerçekleştirilen ön çalışma sunulmaktadır. Beşinci bölümde, geliştirilen yöntem, veri ön işleme, tahminleme ve kural oluşturma olmak üzere üç aşamada anlatılmıştır. Önerilen yöntemin uygulama çıktıları kural oluşturma aşamasında verilmektedir. Son olarak altıncı bölümde önerilen yöntem uygulamasından elde edilen sonuçlar tartışılmış, çalışmanın katkıları ve sınırlılıkları anlatılmış, gelecek çalışma önerilerinden bahsedilmiştir.

1.1 Literatür Özeti

1.1.1 Türkiye'de ve Dünyada İş Kazaları

Sanayi devrimi ile birlikte gelişen teknolojiler, artan dünya nüfusunun ihtiyaçlarını karşılamak amacıyla, fabrikaların seri üretim kapasitesini arttırmış ve yüksek karlılık amacının da etkisiyle fabrikalarda birçok insan bir arada ve uzun mesai saatleri boyunca çalıştırılmaya başlanmıştır. Değişen ve zorlaşan çalışma şartları ile çalışma hayatında artan kazalar ve hastalıklar, sadece çalışanı değil, ailesini, işyerini ve ülkeyi de ciddi şekilde etkileyerek görünür etkilerinin yanı sıra dolaylı ve görünmez bir çok etki de oluşturmaktadır. Bu nedenle daha güvenli ve sağlıklı çalışma şartları için dünyada hem ulusal hem de uluslararası düzeyde hukuksal düzenlemeler ve iş sağlığı ve güvenliği (İSG) uygulamalarının yaygınlaştırılması için çeşitli çalışmalar gerçekleştirilmektedir.

Birleşmiş Milletler uzmanlık kuruluşlarından birisi olan ve insan haklarıyla ilgili düzenlemeler yapan Uluslararası Çalışma Örgütü (ILO), İSG bakımından da önemli çalışmalar yapmaktadır. ILO, çalışma hayatındaki sorunların aşılması yönünde çözümler oluşturulmasında, bu kapsamda önerdiği reformların uygulamaya geçirilmesinde ve çalışma hayatına ilişkin standartların belirlenmesi konularında çalışmaktadır. Türkiye de ILO üyesi bir ülke olarak, 155 no'lu iş Sağlığı ve Güvenliği ve Çalışma Ortamına İlişkin Sözleşmeyi imzalamıştır. Bu sözleşmeye göre her üye ülke, en fazla temsil kabiliyetine sahip işçi ve işveren kuruluşlarına danışarak, ulusal koşullar ve uygulamaya göre, iş güvenliği, iş sağlığı ve çalışma ortamına ilişkin tutarlı bir ulusal politika geliştirecek, uygulayacak ve periyodik olarak gözden geçirecektir [3].

ILO ilkelerine göre İSG, çalışanların sağlığının korunması, çalışma koşullarının ve ortamının iyileştirilmesi ve geliştirilmesi için birçok farklı meslekten (örn. mühendisler, doktorlar, hijyenistler, hemşireler) üyeleri bulunan bir disiplin olarak tanımlanmaktadır. Bu ilkeye dayanan tüm faaliyetlerin temel sonucu ise iş kazalarının ve meslek hastalıklarının önlenmesidir [4].

Dünya Sağlık Örgütü (WHO) iş kazasını, "önceden planlanmamış, çoğu zaman yaralanmalara, makine ve teçhizatın zarara uğramasına veya üretimin bir süre durmasına yol açan olay" olarak tanımlamaktadır. ILO ise iş kazasını, "işten kaynaklanan ve ölümcül veya ölümcül olmayan bir yaralanmayla sonuçlanan beklenmedik olay" olarak tanımlamaktadır [1]. 2012 yılında yürürlüğe giren ve Türkiye'deki ilk iş sağlığı ve güvenliği kanunu olma özelliği taşıyan 6331 Sayılı İş Sağlığı ve Güvenliği Kanununun 3. maddesinde ise iş kazası "işyerinde veya işin yürütümü nedeniyle meydana gelen, ölüme sebebiyet veren veya vücut bütünlüğünü

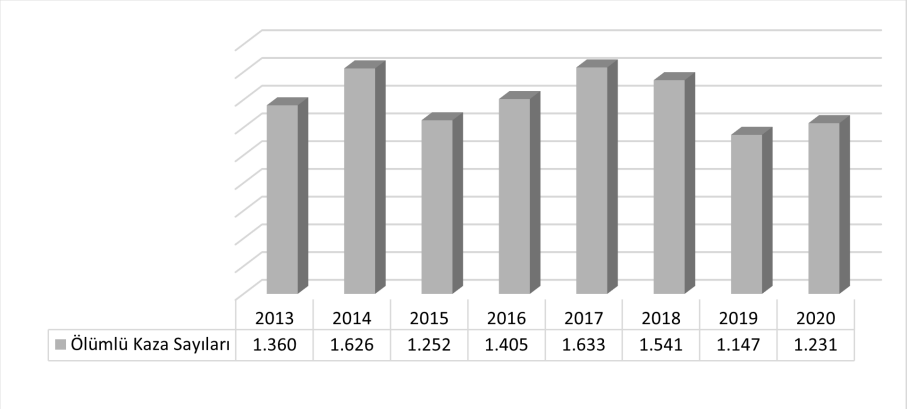
ruhen ya da bedenen özre uğratan olay" şeklinde ifade edilmektedir [5].

İş kazalarının önlenmesine yönelik faaliyetlerin belirlenmesinde iş kazası verisinin kapsamlı ve standart bir yöntemle kayıt altına alınması ve kaza oluşumu ile ilgili ışık tutabilecek istatistiksel analizlerin yapılması en önemli adımları oluşturmaktadır. Bu kapsamda yapılacak araştırmalarda uluslararası bazı önemli kuruluşlar görev almaktadır. ILOSTAT, ILO'nun faaliyetleri kapsamında ülkelerin çalışma hayatı ile ilgili verilerine ulaşmaya çalışan istatistik birimidir. ILOSTAT, kullanıcılara iş kazalarının yanı sıra işgücü ile ilgili geniş bir alanda kapsamlı uluslararası veriler sağlar [6].

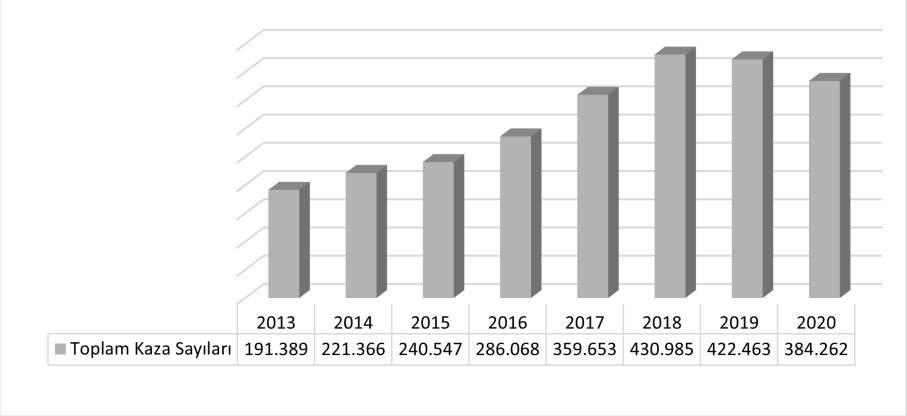
AB İstatistik Ofisi (EUROSTAT) ise uluslararası düzeyde iş kazaları verisi ve istatistikleri sunan bir diğer önemli uluslararası kuruluştur. Ulusal ve uluslararası verilerin tutarlılığı, istatistiksel çalışmalarla ilgili önemli bir husustur. EUROSTAT 2013 yılında, verilerin tutarlığının arttırılması amacıyla, ülkelerin iş kazası verisi toplaması yönteminde ESAW adı verilen bir standart belirlemiştir. ESAW standartları ile birlikte her ülke kendini uluslararası düzeyde bir diğer ülke ile karşılaştırarak diğer uygulamaları örnek alma imkânı bulmaktadır. Türkiye'nin ulusal iş kazası veri tabanı da 2013 yılında ESAW standartlarına dönüştürülmüştür. Yeni düzenleme ile birlikte iş nedeni ile oluşan tüm kazaların, üç iş günü içinde, çevrimiçi bir platform aracılığıyla, SGK'ya bildirme zorunluluğu gelmiştir. ESAW metodolojisinde ise 'iş kazası', çalışma sırasında (mesleki bir faaliyette bulunurken veya işte geçirilen süre boyunca) meydana gelen fiziksel veya zihinsel zarara yol açan olay olarak tanımlanır [7].

Türkiye, iş sağlığı ve güvenliği alanında önemli düzenlemelerden ilki olan 6331 Sayılı İş Sağlığı ve Güvenliği Kanununu [5], 2012 yılında yayınlayarak, İSG mevzuatı çalışmalarına çok daha öncelerde başlamış gelişmiş ülkelere kıyasla gelişmeleri geriden takip etmiş ancak o günden bügüne kadar da bir çok tamamlayıcı yönetmelik hazırlayıp yürürlüğe koyarak İSG alanındaki gelişmelerinde ivme kazanmıştır. Bu kanun ile birlikte stajyer, çırak farketmeksizin tüm çalışanlar kanun kapsamına almış, tüm işletmelerde risk değerlendirmelerinin ve acil durum planlarının yapılması zorunlu hale getirilmiştir. Bu kanun ile birlikte Türkiyedeki işyerleri, gerçekleştirdikleri faaliyetlere göre, çok tehlikeli, tehlikeli ve az tehlikeli olarak üç tehlike sınıfına ayrılmıştır. Tehlikeli ve çok tehlikeli sınıfta yer alan işyerleri ile az tehlikeli sınıfta yer alıp 50'den fazla çalışanı olan işyerlerine, 6331 sayılı kanun yükümlülüğü gereği gerçekleştirilmesi gereken faaliyetlerde danışılmak üzere iş sağlığı güvenliği hizmeti alma zorunluluğu getirilmiş ve hizmet süreleri, faaliyetlerin yenilenme periyotları gibi uygulamalar işyerlerinin çalışan sayısı ile tehlike sınıfına göre belirlenmiştir. Ayrıca iş kazaları ve meslek hastalıklarının bakanlığa bildiriminde, çevirimiçi bir platform kullanımına başlanmıştır. Bu sayede, iş kazaları ve meslek hastalıkları verilerinin hem bakanlığa ulaştırılması kolaylaştırılmış, hem de bakanlığın daha etkin bir şekilde veri

tabanı oluşturması sağlanmıştır. Tüm bu yenilikler ve uygulamalar devletin bu konuya verdiği önemi ortaya koyması açısından önemlidir. Ancak Şekil 1.1'de görüldüğü üzere Türkiye'de gerçekleşen ölümlü iş kazaları sayılarında 2013'ten bu yana belirgin bir iyileşme görülmediği gibi, Şekil 1.2'de görüldüğü üzere, gerçekleşen tüm iş kazaları sayılarının da arttığı gözlenmektedir [8].



Şekil 1.1 SGK verilerine göre yıllık ölümlü iş kazası sayıları



Şekil 1.2 SGK verilerine göre yıllık iş kazası sayıları

Bu artışın nedenleri arasında elbette göçmen statüsündeki kişilerin çalışma hayatına dahil olmaları gibi farklı sebepler sayılabilecek olsa da ana sebebin Türkiye'nin İSG alanında kazandığı ivmenin henüz gelişmiş ülkelerin seviyesini yakalamakta yeterli olmayışı olduğu, dünyada yayınlanan verilerle kıyaslama yapılarak görülebilmektedir. 2014 yılında Türkiye'de iş kazası ölüm hızı, AB ortalamasına göre yaklaşık olarak 6.8 kat daha fazla gerçekleşmiştir [9]. Ayrıca 2016 yılı verilerine göre Türkiye ölümlü olmayan kaza sayısı dünya sıralamasında 73 ülke arasından en kötü 19. ve ölümlü kaza sayısı dünya sıralamasında en kötü 9. sırada yer almaktadır [6]. Ölümlü olmayan kazaların istatistiğe yansıma oranlarının gelişmiş ülkelerde daha fazla olduğu gerçeği göz önünde bulundurulacak olunursa, Türkiye'nin gerçekte tüm iş kazaları açısından

da dünya sıralamasında ölümlü kazaların sıralamasına yani 9. sıraya daha yakın olduğu düşünülebilir.

Gelişmiş ülkelerin dünya kaza sıralamalarında düşük iş kazası oranlarına sahip olmalarının ana sebebi alınan önlemler olsa da, kendi ülkelerindeki tehlikeli teknolojik faaliyetlerini, dolayısıyla iş kazası ve meslek hastalıklarını az gelişmiş ülkelere transfer ediyor olmaları gibi farklı sebepler de araştırmalar esnasında göz önünde bulundurulmalıdır [10].

İş sağlığı ve güvenliği alanındaki literatür incelendiğinde, iş kazalarının ve hastalıklarının önlenmesine yönelik araştırma çalışmalarının dünyada son yıllarda oldukça ivme kazanmış olduğu görülmektedir. Türkiye'de İSG ilgili çalışmalar, ölümlü kazaların sıklıkla yaşandığı inşaat [11–16] ve maden sektörlerinde [17] yoğunlaşmaktadır. İSG alanında yapılan araştırmaların büyük çoğunluğunu, kayıt altına alınan kaza verisinin, iş kazaları ve meslek hastalıklarının nedenlerinin tespitine yönelik incelenmesiyle gerçekleştirilen istatistiksel çalışmalar oluşturmaktadır. Nedenlerin ortaya çıkarılmaya çalışılmasındaki ana amaç, gerekli düzeltmeleri yaparak benzer olumsuzlukların tekrarlanması önlemektir. İşletme düzeyinde yapılan veri analizleri, işletme özelindeki tehlike kaynaklarını ortaya çıkarmayı amaçlarken, ulusal veri tabanı üzerinde yapılan analiz, ulusal düzeyde İSG mekanizmasının aksayan yönlerini bulmayı ve karar alıcılar için güvenilir bir karar destek sistemi oluşturmayı amaçlar. Ancak ulusal düzeyde ele alınan veri, işletme düzeyine göre çok daha karmaşık ve kirli olabilmektedir.

Bu sorunun üstesinden gelmek amacıyla, son yıllarda istatistiklerde kullanılan iş kazası ve meslek hastalıkları verisinin toplanılmasında standartlaşmaya gidilmesi ile birlikte, veritabanlarına kayıt daha sistematik hale gelmiş, bu da bilimsel çalışmalara daha az kirli veri sağlamıştır. Türkiye de ulusal iş kazası veri tabanını 2013 yılı itibariyle yürürlüğe giren 6331 sayılı İş Sağlığı ve Güvenliği Kanunu ile birlikte ESAW standartlarına dönüştürmüştür [7]. Bu sayede toplanan veriler akademik çalışmalarda diğer ülkelerle ortak dil oluşturması açısından önemlidir. Ancak veri tabanı standartında iyileşme olmasına rağmen bildirimlerin çoğunlukla iş kazalarının araştırılması konusunda bilinçsiz olan kişiler tarafından doldurulması sebebiyle veri tabanındaki kazalar büyük ölçüde hafifletilmiş, eksik veya yanlış bilgiler içermektedir ve hafif atlatılan ramak kala olaylar ise bildirilmediği için veri tabanında bulunamaz. Bu açıdan da iş kazaları ve meslek hastalıklarının gerçek sayıları yansıtmaması iş sağlığı ve güvenliği alanındaki bilimsel çalışmalar için önemli bir sorun olarak karşımıza çıkmaktadır [18].

1.1.2 Kaza Analizlerinde Veri Madenciliği Yöntemleri Uygulamaları

Literatürde, iş kazaları, trafik kazaları ve uçuş kazalarına ait veri tabanlarına ilişkin birçok sınıflandırma ve analiz probleminde, kazaların oluşumunun ardında birbirleri ile etkileşen faktörlerin belirlenmesi için veri madenciliği kullanımı geniş bir uygulama alanı bulmuştur. Sadece iş kazaları değil tüm kazalarla ilgili literatür incelendiğinde, kazalara neden olan faktörlerin tespiti için, sınıflandırmada kullanılan gözetimli makine öğrenmesi algoritmaları ile kümelemede kullanılan gözetimsiz makine öğrenmesi algoritmaları sıklıkla kullanıldığı görülmektedir. İş kazaları özelindeki literatür incelendiğinde ise, iş kazalarının rastgele meydana gelmediğini ve bir örüntüyü takip ettiğini, bu nedenle de kaza verilerinin toplanması ve analizi ile tahmin edilebileceğini ve bu sayede gelecekte aynı tür kazaların meydana gelmesinin de önlenebileceğini gösteren çalışmalar mevcuttur [19, 20].

Teknolojinin gelişmesi ile bilikte iş kazalarına ait verinin düzenli olarak kaydedilmesi ve miktarının büyümesi sonucu, iş kazaları verisi büyük veri olarak tanımlanmaya başlanmıştır. İş kazaları kapsamında bakıldığında, büyük veri, bir çok kaynaktan hızlı bir şekilde girdiye dönüşen, doğruluğu/güvenilirliği tam bilinemeyen, büyük hacimli farklı veri türlerinin toplamı olarak ifade edilmektedir. Derin madencilik vb. analizler, insanların kazaları toplu olarak tanımlamada ve ayrıntıları daha iyi anlamalarına yardımcı olacak bilgilere erişmede ve bu sayede de gelecekteki kazaları daha iyi tahmin edebilmelerinde kullanılmaktadır [21]. Türkiye'de İSG ile ilgili çalışmalar incelendiğinde, veri madenciliği ve makine öğrenmesi yöntemlerini kullanan çalışmaların henüz az sayıda olduğu görülmektedir [22–30].

Literatürde, gözetimli makine öğrenmesi algoritmaları, trafik, tren, havacılık ve iş kazalarının arkasındaki faktörlerin ortaya çıkarılmasında, kazalar arasındaki ortak örüntünün analizinde ve bu sayede kazaların sonucunun tahmin edilmesinde kullanılmaktadır. Literatürde gözetimli makine öğrenmesi algoritmalarının en yaygın olarak trafik kazaları alanında [31–40] kullanıldığı görülmektedir. Örneğin, Alikhani vd. [32] İran'daki trafik kazalarının şiddetinin sınıflandırılması için hibrit kümeleme-sınıflandırma metodu kullanmıştır. Yapay Sinir Ağları (ANN) ve Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) ile K-Ortalama (K-Means) ve Özdüzenleyici Haritalar (SOM) yöntemlerini karşılaştırmış ve analiz sonucunda çıkarımının kendi analiziyle sınırlı olmasıyla birlikte sapması yüksek verilerde kümeleme yerine sınıflandırmanın daha iyi sonuç veren bir metot olduğunu ortaya koymuştur. Literatürde trafik kazalarının yanısıra tren kazalarını [41] ve havacılık kazalarını [42] makine öğrenmesi yöntemleriyle inceleyen çalışmalar da bulunmaktadır. Literatür iş kazaları odaklı incelendiğinde, daha az yaygın olmakla birlikte, iş kazalarının kümelenmesi, sonuçlarının tahminlenmesinde de veri

madenciliğinin kullanıldığı görülmektedir[20, 26, 27, 43–46]. Örneğin yaralanma başlığındaki çalışmaları inceleyen Valmuur'un [47], spor ve şiddet sonucu oluşan tüm çeşitlerdeki yaralanmalar ile kıyaslandığında makine öğrenmesi yöntemlerini kullanan araştırmaların yarısından çoğunun mesleki yaralanmalar üzerine olduğunu ortaya koyan bir çalışması mevcuttur. Tüm bu çalışmalar, tahminleme, nitelik seçimi, veri üretimi ve daha da önemlisi kuralların oluşturulması yoluyla verinin yorumlanması gibi kabiliyetler açısından makine öğrenmesi tekniklerinin geleneksel istatistiksel yöntemlere olan üstünlüğünü göstermiştir. Literatürde tüm kazalar kapsamında çok sayıda çalışma olmakla birlikte bu tez çalışmadaki literatür taraması iş kazaları odağında gerçekleştirilmiştir.

1.1.2.1 İş Kazaları Analizinde Gözetimli Makine Öğrenmesi Uygulamaları

Gözetimli Makine Öğrenmesi algoritmaları, sınıflandırma ve tahmin modellemelerinde, bir gözlemin sınıfını veya kategorisini, öğrenmenin gerçekleştiği veriden çıkarılan bilgilere dayanarak tahmin etmek için yaygın olarak kullanılır [48]. İş kazaları kapsamındaki çalışmalarda yaygın yaklaşım, makine öğrenmesi algoritmalarının oluşturduğu istatistiksel modellemelerde, bağımlı değişken olarak kaza sonucunun ve bağımsız değişken olarak kaza, işyeri, iş, mağdur vb. özelliklerin kullanılmasıdır.

İş kazalarının sonuçları açısından hem ikili hem de çok sınıflı sınıflandırmalar kullanılabilmektedir. Bağımlı değişken olarak kullanılan kaza şiddeti, bazı araştırmalarda ölümlü/ölümlü olmayan, ciddi/ciddi olmayan [48], yaralanmalı/yaralanmasız [49], yüksek şiddetli/düşük şiddetli [50] olmak üzere ikili düzeyde kullanılırken, bazı araştırmalarda bağımlı değişkenin ölümlü/major/minör vb. olmak üzere ikiden fazla değer aldığı multinominal modeller kullanılmıştır [20].

Çok sınıflı sınıflandırma örneklerinde ise çalışmalar çoklu kategorilere sahip şiddet düzeyinin tahminine [20, 26, 27, 51], belirli bir kaza türünün tahminine [19, 52, 53] veya kaza sonucu, yaralanma türü vb. türlerinin tahminine odaklanmaktadır [20, 26, 27, 54, 55].

Matias vd. [53] İspanya'daki sanayi, madencilik, inşaat ve hizmet sektörlerinde yer alan 1500 şirkette, 2003-2006 döneminde kaydedilen 148 yer seviyesinde düşme kazalarına ait veriyi kullanarak bir kaza zincirinin sonucunun yer seviyesindeki düşme olup olmayacağının tahmininde, Bayes Ağları (BN), Karar Ağaçları (DT), Destek Vektör Makinaları (SVM) ve Aşırı Öğrenme Makineleri (ELM) algoritmalarını uygulamışlar ve performanslarını değerlendirmişlerdir.

Bevilacqua vd. [43] 1994–2004 yılları arasında, bir petrol rafinerisinde gerçekleşen 200 kazayı inceleyerek veriye dönüştürmüş, ağaç tabanlı sınıflandırma yöntemlerini kullanarak en tehlikeli çalışma durumlarını ve nedenlerini belirlemek için risk kategorisini, kaza tipini ve kaza periyodunu (yasalar öncesi ve sonrası) tahminleyen faktörleri incelemişlerdir. Analiz sonucuna göre Sınıflandırma ve Regresyon Ağaçları (CART) algoritması risk değerinin tahmininde en yüksek performansı göstermiştir.

Cheng vd. [44] çalışmalarında, Tayvan'da inşaat sektöründe 2000-2009 yılları arasında gerçekleşen 1.542 ciddi kazanın sebep sonuç ilişkilerini CART algoritmasını kullanarak incelemiş ve kazaların oluşumda etkili kuralları çıkararak düşme ve çökme olaylarının yaralanma tahmininde anahtar faktör olduğunu bulmuşlardır.

Mıstıkoğlu vd. [22], inşaat sektöründe gerçekleşen çatıdan düşme kazalarının niteliklerini ortaya koymak ve bu nitelikler arasındaki ilişkileri gösteren kuralları çıkarmak için İş Sağlığı ve Güvenliği İdaresi (OSHA) verilerinden elde ettikleri 594 ölümlü toplam 1413 kaza için iki farklı karar ağacı algoritması (C5.0 ve CHAID) kullanmışlar ve bu algoritmaların yaralanma derecesini (ölümcül ve ölümcül olmayan yaralanma) tahminleme performanslarını incelemişlerdir. Analiz sonucuna göre, çatı düşmelerinin yaralanma derecesini belirleyen en güçlü değişkenlerin, düşme mesafesi, ölüm/yaralanma nedeni, güvenlik eğitimi ve düşmeye neden olan inşaat operasyonu olduğunu bulmuşlardır.

Hajakbari [56] işyerlerinin risk seviyelerinin ölçümünde frekans, şiddet, kaza riski ve iki kaza arasındaki zaman farkı gibi değerleri kullanan DFSRM isimli skorlama sistemini önermişlerdir. k-Ortalama ile işyerleri kümelendirilmiş ve her bir küme için karar ağaçları temelli kurallar çıkarılmıştır.

Amiri vd. [57] çalışmalarında, 2007-2011 yılları arasında İran'da gerçekleşen 21.864 inşaat iş kazasını iki gruba ayırmışlar ve kazaların oluşumunda etkili faktörlerin belirlenmesi için Topluluk Modeli Temelli Karar Ağaçları metotlarını uygulayarak performanslarını karşılaştırmış ve daha iyi performans gösteren modelin ortaya çıkardığı faktörlerle birliktelik kuralları ortaya koymuşlardır.

Chen ve Luo [51], OSHA'nın veri tabanından çekilen 2005-2015 yılları arasındaki 1.161 düşme ile ilgili kayıtlar üzerinde K-modları, Radyal Temelli İşlev Ağı (RBF) ve karar ağaçları (DT) algoritmaları kullanılarak üç risk tahmin modeli oluşturmuşlar ve önerilen bu üç modelin performansını değerlendirerek, DT tabanlı düşme riski tahmin modelinin %75 ile en iyi performansa sahip model olduğunu bulmuşlardır. Ayrıca, düşme olayının şiddetinde kritik faktörlerin ilk üçününün yerden uzaklık, işçinin mesleği ve düşmenin kaynağı olduğunu göstermişlerdir.

Altunkaynak [58] Türkiye'de 2012 yılında üretim alanında gerçekleşen 39.327 iş kazasınının oluşumda etkili olabilecek kayıp gün, kaza sonrası durum ve kaza tipi arasındaki ilişkiyi incelmede, Karar Ağaçları ve Bayes Ağları veri madenciliği yöntemleri ile klasik Ki-kare test metodlarını karşılaştırmış, en sık gerçekleşen kaza tipinin "iki obje arasında vucüdun sıkışması" tipindeki kaza olduğunu ve daha çok "makine dışı metal üretimi" sektöründe, 500'den az çalışanı olan işyerlerinde, kazalının yaşı 40'dan küçük olduğunda görüldüğünü ve bu tip kazaların 30-59 gün arası kayıpla sonuçlandığını bulmuştur.

Abubakar vd. [25] çalışmalarında metal döküm endüstrisinden topladıkları 306 anketi kullanarak işyeri yaralanmaları, güvenlik iklimi ve güvenli davranışlar arasındaki ilişkiyi test etmek için Yapay Sinir Ağları (ANN) kullanarak tahmine dayalı bir analiz yapmışlar ve sonuç olarak güvenli davranışların güvenlik ikliminde başlı başına etkili bir faktör olduğunu, dolayısıyla yaralanmaların azaltılmasında tek başına etkili olduğunu ortaya koymuşlardır.

Kang ve Ryu [54] çalışmalarında, 2008-2014 yılları arasında inşaat sektöründe gerçekleşen 9.796 kazanın verisini kullanarak çok sınıflı kaza tipini tahminlemek için Rassal Orman algoritmasını (RF) kullanmışlar ve %71.3 doğrulukla önemli faktörleri ortaya çıkarmışlardır. Sınıf dengesizliği problemini alt örneklem (RUS) yöntemiyle ele almışlardır. Benzer çalışmalardan farklı olarak hava durumu verisini de analize eklemişler ancak kaza tipi tahminlemede etkili olmadığını bulmuşlardır.

Zhang [59] çalışmasında OSHA web sitesinde yer alan 1983-2014 yılları arasında inşaat sektöründe gerçekleşmiş 16.323 kazanın raporunu 2 aşamalı bir yöntem ile incelemiştir. İlk aşamada önceden eğitilmiş bir metin madenciliği modeli olan Word2Vec uygulayarak kazayı açıklayan özet metinden kelime öbekleri oluşturmuş daha sonra bu oluşturduğu kelime öbeklerini kullanarak veride bulunan 11 farklı kaza sebebini Derin Sinir Ağları (DNN) yöntemiyle tahminlemişlerdir. Ayrıca çalışmasında Derin Sinir Ağları (DNN) Algoritmasının bu çalışmada Destek Vektör Makineleri (SVM), Doğrusal Regresyon (LR), K-yakın komşu (KNN), Karar Ağacı (DT) ve Naif Bayes (NB) algoritmaları ile karşılaştırıldığında daha iyi sonuç verdiğini ortaya koymuştur.

Sarkar [19] kayma takılma düşme kazalarının oluşumunda etkili faktörleri kullanarak, farklı makine öğrenmesi algoritmalarını ve parametre optimizasyon yöntemlerini kıyaslamış ve kazaların kayma takılma düşme kazası olup olmadığının tahminlenmesinde kullanılabilecek yeni bir metodoloji önermişlerdir. Kazaların oluşumunda etkili olan faktörlerin tanımlanması sırasında en çok bilginin saklı olduğunu düşündükleri yapılandırılmamış metin açıklamalarını Örtülü Diriklet

Yerayırması (LDA) yöntemini kullanarak kazayı tanımlayan sekiz kategorili bir niteliğe dönüştürmüşlerdir. Bu yeni nitelik ve diğer kaza niteliklerini kullanarak C5.0, CART ve RF karar ağacı algoritmalarının parametrelerini, PSO ve GA optimizasyon yöntemleriyle optimize ederek kombinasyonların tahmin doğruluğunu kıyaslamış ve PSO-RF modelinin en yüksek tahmin doğruluğunu verdiğini bulmuşlardır. Ayrıca en iyi model olarak bulunan PSO-RF yöntemi ile kayma düşme takılma kazalarının oluşumunun arkasındaki faktörleri açıklayan 20 adet yorumlanabilir güvenlik kuralı oluşturmuşlardır. Bulgular insan davranışlarının ve güvensiz davranışların kayma düşme ve takılma ile ilişkili yaralanmalarda en yaygın nedenler olduğunu göstermiştir. Ayrıca malzeme taşıma, engeller ve yoldaki gevşek parçacıklar ve zemine su dökülmesi gibi bir dizi faktör kayma takılma düşme kazalarının çoğunu açıklamaktadır.

Ayhan ve Tokdemir [27] 87 farklı inşaat projesinden 5 yılda gerçekleşen toplam 17.300 olayı incelemiş, olayların sonucunu (şiddet) riskli hareket, ramak kala, yaralanma, kayıp günlü, ölümlü olmak üzere 5 sınıfa ayırarak Yapay Sinir Ağları (ANN) algoritmasının tek başına ve bulanık set teorisi ile birlikte kullanımındaki tahmin performansını değerlendirmişlerdir. ANN algoritması tek başına bile % 84 doğruluk oranına ulaşırken, bulanık set teorisi kullanıldığında riskli hareket ve kayıp günlü iş kazalarının tahmin performansının arttığını bulmuşturlar.

Ayhan ve Tokdemir [26] diğer bir çalışmalarında, megaprojelerden toplanan 5.224 olay, 9 kategoriden toplam 60 nitelik üzerinden risk değerlendirmede kaza şiddeti tahmini için önce Örtük Sınıf Kümeleme (LCC) yöntemiyle homojen gruplar oluşturmuş, daha sonra bu grupların ramak kaladan ölümlü kazaya 6 şiddet seviyesi tahmini için Yapay Sinir Ağları (ANN) ve Vaka Tabanlı Akıl Yürütme (CBR) yöntemlerinin performanslarını kıyaslamışlardır. CBR algoritmasının %86 doğrulukla daha iyi performans gösterdiğini ortaya koymuşlardır. Ayhan ve Tokdemir [50] bir sonraki çalışmalarında farklı inşaat projelerinden toplanan 4.109 kazayı önce LCC yöntemiyle homojen gruplara ayırmış, sonra kaza sonucunun yüksek ve düşük şiddet olarak tahmini için dört farklı ANN modeli kullanılşmıştır. Analiz sonucunda yüksek şiddet ile sonuçlanan kazaların düşük şiddetle sonuçlanan kazalara kıyasla biraz daha yüksek doğrulukla tahmin edilebildiği ortaya konulmuştur.

Yedla vd. [60] çalışmalarında, maden sektöründeki iş kazalarının sonucunda ortaya çıkan iş günü kayıplarının tahminlenmesinde Karar Ağazı (DT), Rassal Orman (RF) ve Yapay Sinir Ağları (ANN) yöntemlerini hem yapılandırılmış hem de sözel/yapılandırılmamış veri üzerinde kullanarak performaslarını kıyaslamıştır. Algoritmaların sözel veri üzerindeki tahmin performansı daha yüksek bulunmuştur. Yedla vd. bu bulguyu sözel verinin yapılandırılmış kategorik/numerik veriden daha fazla ve doğru bilgi barındırması ile yorumlamışlardır. Algoritmalar arasında en

yüksek performansı ortaya koyan algoritman RF olmuştur.

Literatürde iş kazalarına gözetimli makine öğrenmesi yöntemleri uygulayan çalışmaların özeti Tablo 1.1'de verilmektedir. Tablo incelendiğinde en çok ağaç modellerinin kullanıldığı görülmektedir. Tahmin edilen bağımlı değişkenin ise çoğunlukla kaza şiddetini tanımlayan kayıp iş günü, yaralanma vb. değişkenler olduğu ortaya çıkmaktadır. Ağırlıklı sektör inşaat olmakla birlikte iş kazaları şiddetinin yüksek olduğu sektörlerin incelenmek için seçildiği görülmektedir. Tabloda yer alan çalışmalardan sadece ikisi birden fazla sektörü incelerken diğerleri tek bir sektöre odaklanmış, tüm sektörleri toplu olarak inceleyen bir çalışmaya rastlanmamıştır.

Tablo 1.1 Literatür taraması özet tablosu - Gözetimli Makine Öğrenmesi

Yayın	Veri	Sektör	Yöntem	Tahmin Edilen
Bevilacqua vd. (2008) [43]	Bir petrol rafinerisinde 10 yılda gerçekleşen 200 kaza	Petrokimya	DT, CART	Kaza tipi, kaza periyodu, risk kategorisi
Matías vd. (2008) [53]	4 yılda gerçekleşen 148 düşme kazası	Sanayi, Madencilik, İnşaat ve Hizmet	BN, DT, SVM, ELM	Yer seviyesinde düşme kazalarının tipleri
Cheng vd. (2012) [44]	Ulusal veri tabanında yer alan Taiwan inşaatlarında 10 yılda gerçekleşmiş 1.542 ciddi kaza	İnşaat	CART	Yaralanma tipi
Hajakbari (2014) [56]	İran ulusal kaza veri tabanında yer alan 19 yılda gerçekleşmiş 687 kaza	İnşaat, Sanayi	K-Means, DT	Risk seviyesi
Amiri vd. (2015) [57]	5 yıllık 21.864 kaza	İnşaat	Birliktelik Kuralları, Topluluk Öğrenmesi	Yüksek riskli kazalar
Mistikoglu vd. (2015) [22]	OSHA veri tabanından 594 ölümlü toplam 1.413 kaza	İnşaat	C5.0 CHAID	Ölümlü çatıdan düşme kazaları
Chen ve Luo (2016) [51]	OSHA veri tabanından 11 yıllık 1.161 kaza	İnşaat	k-mod, RBF, DT	15 farklı düşme sonucu yaralanma seviyesi
Abubakar vd. (2018) [25]	306 kaza	Metal	ANN	Yaralanma tipi

Tablo 1.1 Literatür taraması özet tablosu - Gözetimli Makine Öğrenmesi - (devamı)

Yayın	Veri	Sektör	Yöntem	Tahmin Edilen
Altunkaynak (2018) [58]	1 yıllık 39.327 kaza	Üretim	DT, Birliktelik kuralları	Kayıp gün ve kaza sonrası durum
Ayhan ve Tokdemir (2019a) [27]	87 farklı inşaat projesinden 5 yılda gerçekleşen 17.300 olay	İnşaat	ANN, Bulanık set teorisi	5 farklı şiddet seviyesi
Ayhan ve Tokdemir (2019a) [26]	Mega projelerden toplanan 5.224 olay	İnşaat	LCCA ANN CBR	6 farklı şiddet seviyesi
Kang ve Ryu (2019) [54]	Korea KOSHA veri tabanından 17 senelik 9.796 kaza	İnşaat	RF	9 farklı kaza tipi
Sarkar vd. (2019) [19]	3 senelik 1.964 yaralanmalı kaza	Metal	LDA, C5.0, CART ve RF	Kayma, takılma düşme kazaları
Zhang (2019) [59]	32 yıllık 16.323 kaza	İnşaat	SVM, LR, KNN, Dt, NB	11 farklı kaza tipi
Ayhan ve Tokdemir (2020) [50]	Farklı inşaat projelerinden toplanan 4.109 ramak kala uygunsuzluk ilkyardımlı ve diğer şiddetteki olaylar	İnşaat	LCCA, ANN	2 farklı şiddet seviyesi
Yedla vd. (2020) [60]	4 yıllık 16.394 kaza	Maden	DT, RF, ANN	İş günü kaybı

1.1.2.2 İş Kazaları Analizinde Gözetimsiz Makine Öğrenmesi Uygulamaları

Büyük verinin analizinde, veri bazen manuel olarak etiketlenemiyor olabilir. Ya da veri toplama esnasında veri etiketlenmemiş ise manuel olarak sonradan etiketlenmesinin bedeli de oldukça yüksek olabilmektedir[61]. Bu sebeple, bazı durumlarda gözetimli yerine gözetimsiz öğrenme yöntemleri tercih edilmektedir. Gözetimsiz öğrenmede, gözetimli öğrenmedeki gibi doğru veya yanlış cevaplar yoktur, yalnızca girdi veriler vardır ve bu girdi verilere karşılık gelen çıktı değişkenler bulunmaz [62]. Gözetimsiz öğrenmede algoritmalar, veriler hakkında daha fazla bilgi edinmek için verilerdeki ilginç yapıyı, temel yapıyı veya dağılımı modelleyerek ortaya koyarlar. Gözetimsiz makine öğrenmesi algoritmaları içinde en bilinenleri kümeleme (clustering), birlikteli kuralları (Association Rules) ve temel birleşen analizidir (Principal Component Analysis - PCA). Kümelemede ana amaç veriyi ortak özelliklerini ve benzerlikleri kullanarak farklı gruplara ayırmaktır. Başarılı şekilde sonuçlanan bir kümeleme sonrası aynı küme içinde benzerlikler fazla, kümeler arası benzerlikler azdır[63]. K-Ortalama ve Hiyerarşik Bölütleme yaygın olarak kullanılan kümeleme algoritmalarındandır.

Kaza vakalarının parametrelere göre dağılımının bilinmeyen doğası, vakaları benzerliklerine göre gruplandıran ve en yaygın olay zincirlerini tanımlamak için yol gösteren kümeleme yöntemleri ile ortaya çıkarılabilir [64]. Kazaların özellikleriyle birlikte sistematik analizi ve araştırılması kazaların tekrarını önlemede etkin bir yöntemdir. Son yıllarda bir yapay sinir ağları tekniği olan SOM gibi grafiksel yöntemler ve verileri hiyerarşik bir ilişki olmadan k sayıda kümeye ayıran Örtük sınıf analizi (LCA) gibi veri madenciliği yöntemleri de kümeleme için kullanılmaktadır [65]. İş kazaları literatüründe gözetimsiz makine öğrenmesi uygulamalarının kümeleme ve birliktelik kuralları alanında yoğunlaştığı görülmektedir.

Liao ve Perng [66] çalışmalarında, 1999-2004 yılları arasında Tayvan'da inşaat sektöründe gerçekleşen 309 ölümlü kazanın raporundan ilgili faktörleri çıkararak oluşturdukları veriyi kullanarak birliktelik kuralları ortaya koymuş ve işçinin yaşının, maaşının, haftanın gününün ve havanın yağmurlu olup olmayışının ölümlü kazaların üzerinde anlamlı bir öneme sahip olduğunu ortaya koymuşlardır.

Cheng vd. [67] Tayvan'da 2000-2007 yılları arasında gerçekleşen 1.347 iş kazası verisini bireysel faktörler ile iş kazaları arasındaki neden sonuç ilişkisini açıklamak üzere sitatistiksel olarak incelemiş ve daha sonra bina inşaat projelerinde 10, inşaat mühendisliği projelerinde 5 temel birliktelik kuralını ortaya çıkararak kazaların oluşumunda etkili tehlikeleri ortaya koymuşlardır. Analizlerin sonucunda hem çalışanların hem de yönetimin güvenlik sorunları ve potansiyel tehlikeler konusunda yetersiz farkındalığa sahip olduğunu göstermişlerdir. Ayrıca analiz sonucunda bu

tehlikelerin özellikle 10'dan az çalışanı olan işyerlerinde daha fazla ortaya çıktığı görülmektedir.

Palamara vd. [64] İtalyan ahşap endüstrisi ESAW kaza verilerini kullanarak 2 seviyeli bir kümeleme (SOM ve K-Means) yöntemi olarak SKM yöntemini önermiş ve bu yöntemi kümelerin sayısal risk değerlerinin bulunması ve kritik kaza gruplarının belirlenmesi amacıyla kullanmıştır. Ancak analizlerin tekrarlanması durumunda sonuçların farklı çıktığı belirtilmiştir. Comberti vd. [68] bu çalışmalarında önerdikleri bir duyarlılık analizi ile bu probleme çözüm getirmiştir.

Lombardi ve Rossi [69], ulusal veri tabanında bulunan 2002-2004 yılları arası inşaat sektöründe gerçekleşen 547 yüksekten düşme kazasını incelemiş, k-ortalama kümeleme yöntemi (K-Means) kullanarak %90d̂an fazla kazayı tanımlanabilir üç farklı kümede toplayabilmişlerdir.

Neonen [70], Finlandiya ulusal iş kazaları ve meslek hastalıkları istatistikleri veri tabanındaki 2 senelik 222.932 kayma takılma ve düşme kazalarını incelemiş, Karar ağaçları kullanarak bulduğu ilgili faktörlerden kurallar oluşturmuş, kazalının yaşının ve mesleğinin kayma, takılma düşme kazalarında etkili olduğunu ortaya koymuştur.

Ghousi [71] çalışmasında, İrandaki bir işletmenin 2003-2010 yılları arasında üretim biriminde gerçekleşen 1.954 kazayı K-ortalama algoritması ile 8 ayrı kümeye ayırmış ve her bir küme için apriori algoritması ile birliktelik kuralları çıkarmıştır.

Mure vd. [72], zorlu üretim ortamlarının iş kazası dinamikleri üzerindeki etkisini ortaya koymak için SKM ve Bulanık Mantık yaklaşımını birleştiren iki aşamalı bir yöntem önermişlerdir. Önerdikleri yöntemi metalurji endüstrisinde gerçekleşmiş 3.600 kazanın analizinde kullanmış ve tahmin edilen risk puanının doğruluk performansı üzerinden yöntemin etkinliğini ortaya koymuşlardır.

Moura vd. [73] çalışmalarında, bir önceki çalışmalarında ortaya koydukları çok nitelikli teknolojik kazalar veri kümesini (MATA_D) [74] SOM yöntemi ile 2 boyutlu risk haritalarına dönüştürerek büyük kazalar sırasında insan faktörleri, teknolojik yönler ve organizasyonlar arasındaki etkileşimlere dayalı bir risk değerlendirme doğrulama şeması önermişlerdir. Bir başka çalışmada, Moura vd. [75], SOM yöntemi ile, kazalara katkıda bulunan faktörleri dikkate alarak, kazaları benzerliğe göre gruplandırmış ve ortak kalıpları belirleyerek veri seti içindeki temel ilişkileri ortaya koymuşlardır.

Comberti vd. [76] çalışmalarında, SOM yöntemiyle iş kazaları veri tabanındaki veriyi benzer kazaların yer aldığı kümelere ayırmış, her bir kümenin frekansı ve iş göremezlik

gün sayısının çarpımı ile risk değerini oluşturmuş ve kaza kümelerinin en riskli 4 tanesi için tetikleyici değişkenlerini inceleyerek önleyici aksiyonlar önermişlerdir. Bir diğer çalışmada, Comberti vd. [77], 2006-2013 yılları arasında İtalya'nın yiyecek içecek endüstrisinde gerçekleşmiş 6.000'den fazla iş kazasını yürürlüğe giren ulusal güvenlik yönetmeliği öncesi ve sonrası olmak üzere iki farklı periyoda ayırarak incelemiş ve SKM yöntemiyle iki periyodta da birebir aynı kümelerin oluştuğunu ortaya koymuştur. Ayrıca, Comberti vd. [78] çalışmalarında, daha büyük bir veri setine SKM uygulamak amacıyla, 2006-2013 yılları arasında İtalya'nın ağaç endüstrisinde gerçekleşmiş 4.000'den fazla iş kazasını SKM yöntermiyle 4'ü kazanın oluşumunda etkili dinamikler olmak üzere yaranın vucüttaki yeri ve çalışan yaşı ile birlikte 6 değişken kullanarak 21 kaza kümesine ayırmıştırlar.

Chiang vd. [79] çalışmalarında, Hong Kong'da 2006-2015 yılları arasında, bakım onarım çalışmaları esnasında gerçekleşen 251 ölümlü kazayı medyada yayınlanan raporları kullanarak incelemişler ve gerçekleştirdikleri Temel Bileşenler Analizi (PCA) ve kümeleme analizi sonucunda dört küme ortaya çıkarmışlardır. Temel Bileşenler Analizi ile tutarlı olarak kümelemede, olayın gerçekleştiği ayın en güçlü tahminleyici olduğu bulunurken, bunu sırasıyla kazazedenin yaşı, olayın zamanı ve olayın gerçekleştiği haftanın günü değişkenlerinin takip ettiğini bulmuşlardır.

Shin vd. [80] Kore'de 2006-2010 yılları arasında inşaat sektöründe gerçekleşen 60.244 kazayı tanımlayan 55 nitelik ile birliktelik kuralları oluşturmuş ve bu kuralları inceleyerek içlerinden İSG yönetimi için belirgin düzeltici önleyici faaliyet tanımlayan 30 birliktelik kuralını ortaya koymuşlardır.

Literatürde iş kazalarına gözetimsiz makine öğrenmesi yöntemleri uygulayan çalışmaların özeti Tablo 1.2'de verilmektedir. Tablo incelendiğinde çalışmaların iş kazalarını ortaya çıkaran faktörlerin kümelendiği kombinasyonların ortaya çıkarılması olduğu görülmektedir.

Tablo 1.2 Literatür taraması özet tablosu - Gözetimsiz Makine Öğrenmesi

Yayın	Veri	Sektör	Yöntem	Amaç
Liao ve Perng (2008) [66]	309 ölümlü kaza	İnşaat	Birliktelik kuralları	Ölümlü iş kazalarına katkıda bulunan bireysel faktörlerin, görev faktörlerinin, yönetim faktörlerinin ve çevresel faktörlerin etkisinin incelenmesi.
Cheng vd. (2010) [67]	1.347 kaza	İnşaat	Birliktelik kuralları	Kazaların oluşumunda ortaya çıkan tehlike faktörleri kombinasyonlarının belirlenmesi
Palamara vd. (2011) [64]	1.207 olay	Üretim	SOM, K-Means	İtalyan ahşap endüstrisi ESAW verisinde en kritik kaza gruplarının belirlenmesi
Lombardi ve Rossi (2013) [69]	Ulusal veri tabanından 3 senelik 547 kaza	İnşaat	K-Means	Kazaların oluşumda etkili tekrar eden faktörlerin belirlenmesi için kümeleme
Nenonen (2013) [70]	Ulusal veri tabanından 2 senelik 222.932 kaza	İnşaat	DT	Kayma, takılma, düşme kazalarının incelenmesi ve oluşımındaki faktörlerin bulunması
Ghousi (2015) [71]	8 senelik 1.954 kaza	Üretim	K-Means, Birliktelik kuralları	Üretimde gerçekleşen iş kazalarının kümelenmesi ve kümelerin oluşum kurallarının çıkarılması

Tablo 1.2 Literatür taraması özet tablosu - Gözetimli Makine Öğrenmesi - (devamı)

Yayın	Veri	Sektör	Yöntem	Amaç
Moura vd. (2017a) [73]	238 kaza	Sanayi	SOM	Büyük kazalar sırasında insan faktörleri, teknolojik yönler ve organizasyonlar arasındaki etkileşimlerin ortaya çıkarılması
Moura vd. (2017b) [75]	238 kaza	Sanayi	SOM	Kazaları katkıda bulunan faktörlerin benzerliğine göre kümelendirme ve ortak kalıpları belirleyerek kaza veri seti içindeki temel ilişkileri ortaya koyma
Mure (2017) [72]	3.600 kaza	Metal	SOM, K-Means, Bulanık mantık	Kümeleme ile edinilen Zor üretim şartları girdilerini kullanarak, bulanık mantık yaklaşımı ile bir tesisteki iş kazalarının risk analizini gerçekleştirmek
Shin vd. (2017) [80]	60.244 kaza	İnşaat	Birliktelik Kuralları	Kazayı tanımlayan nitelikleri kullanarak birliktelik kuralları oluşturma ve uygun düzeltici önleyici faaliyet tanımlama
Chiang vd. (2018) [79]	Bakım onarım çalışmaları esnasında gerçekleşen 251 ölümlü kaza	Karışık	PCA ve Kümeleme	Ölümlü iş kazalarını ortaya çıkaran belirgin durumları incelemek için kümeleme
Comberti vd. (2018a) [77]	6.000 kaza	Gıda	SOM, K-Means, Duyarlılık Analizi	Yürürlüğe giren ulusal güvenlik yönetmeliği öncesi ve sonrası ortaya çıkan kazaların kümelenmesi ve kıyaslanması.

Tablo 1.2 Literatür taraması özet tablosu - Gözetimli Makine Öğrenmesi - (devamı)

Yayın	Veri	Sektör	Yöntem	Amaç
Comberti vd. (2018b) [76]	754 kaza	Karışık	SOM, K-Means, Duyarlılık Analizi	Kazaları katkıda bulunan faktörlerin benzerliğine göre kümelendirme, kümelerin frekansı ve iş göremezlik gün sayısının çarpımı ile risk değerlerinin oluşturulması
Comberti vd. (2018c) [78]	4.000 kaza	Ağaç	SOM, K-Means, Duyarlılık Analizi	Önceki çalışmalarında önerdikleri metodun iyileştirilmesi ve geliştirilmesi ve doğrulama amacıyla daha büyük bir veri tabanında uygulanması.

Tablo 1.3 Önerilen Yöntem

Veri	Sektör	Yöntem	Tahmin Edilen	Amaç
Ulusal veri tabanından 4.5 yıllık 187.910 kaza	Tüm sektörler	SGB, Birliktelik Kuralları	4 farklı ikili şiddet seviyesi	Kazaların oluşumda ve şiddetinde etkili faktörlerden oluşan kaza zincirlerini ortaya çıkarmak.

Yapılan literatür taramasında toplanan ulusal iş kazaları verisinin analizinde veri madenciliği kullanan yayınların sayıca sınırlı olduğu ve bu çalışmaların belirli bir kaza tipine veya belirli bir sektöre odaklandığı görülmüştür. İş kazası verisini ulusal boyutta bir bütün olarak analiz ederek, kazaların oluşumda ve şiddetinde etkili faktörleri karar alıcılara sunma amacı ile geliştirilen kapsamlı bir çalışmanın olmadığı tespit edilmiştir. Tablo 1.3'de gösterildiği üzere önerilen yöntem, hem gözetimli hem gözetimsiz makine öğrenmesi yöntemlerini kullanarak, ulusal iş kazası verisini, kazaların oluşumda ve şiddetinde etkili faktörlerden oluşan kaza zincirlerine dönüştüren ilk çalışma özelliğini taşımaktadır.

1.2 Tezin Amacı

Her gün meydana gelen iş kazaları sonucu çalışanlar yaralanmakta, çalışamaz hale gelmekte ve hatta yaşamlarını kaybetmektedirler. Artan iş kazalarının önlenmesi için ülkemizde mevzuat çıkarma atılımlarına ek olarak İSG alanında daha geniş araştırma çalışmalarına ihtiyaç bulunmaktadır. İş kazalarının önlenmesinde doğru kararlar alınabilmesi için, gerçekleşen iş kazalarının incelenmesi, oluşumundaki etkenlerin belirlenmesi ve karar alıcılara sunulması bu açıdan önemli bir rol oynamaktadır. Bu etkenlerin bilinmesi hem ulusal düzeyde faaliyetler belirlenirken hem de işyerlerinde gerçekleştirilen risk değerlendirme çalışmaları esnasında yol gösterici olacaktır.

Gelişen teknoloji ile toplanan ve her geçen gün artmaya devam eden ulusal kaza verisi, her bir kazayı tanımlayan çok sayıda parametreye de sahip olması ile birlikte, geleneksel istatistiksel yöntemler ile analizi zorlaşan geniş ve karmaşık bir veri tabanı oluşturmaktadır. Literatürde sunulan çalışmalar, veri madenciliğinin karmaşık verilerle başa çıkmada geleneksel istatistiksel yöntemlere kıyasla üstünlük oluşturduğunu ve iş kazaları verisinden sonuç tahminlemede, değerli çıkarımlar ortaya koymada etkinliğini göstermektedir. Bu göstergeler, veri madenciliği ile ortaya çıkarılan bulguların, iş kazalarının önlenmesine yönelik çalışmalara ışık tutmada etkili olduğunu göstermektedir.

Bu tez çalışmasının amacı, ulusal iş kazası veri tabanının karmaşık yapısını, ilgili devlet kurumları ve özel sektördeki karar alıcılara alınacak önlemler hususunda ışık tutan bir karar destek sistemi yapısına dönüştürecek bir analiz yöntemi geliştirmektir. Önerilen analiz yönteminde İSG alanında toplanan kaza verisinin bir bütün olarak ele alınması ve iş kazalarının oluşumunda ve şiddetinde etkili faktörlerden oluşan kaza zincirlerine dönüştürülmesi hedeflenmektedir. Ortaya çıkan kaza zincirlerinden oluşan karar destek sistemi, kamu ile paylaşılan özet istatistiklere güçlü çok daha etkili bir alternatif oluşturacak, iş kazalarının önlenmesinde hem ulusal düzeyde faaliyetler

hem de işyerlerinde alınacak önlemler belirlenirken yol gösterici olacaktır.

1.3 Orijinal Katkı

Literatür taramasında, kaza verisi ile veri madenciliği yöntemlerinin, sıklıkla ulaşımda gerçekleşen kazaların araştırılmasında kullanıldığı, İSG alanında toplanan kaza verilerini veri madenciliği ile analiz eden çalışmaların literatürün daha dar bir alanını oluşturduğu anlaşılmıştır. Dünyada İSG alanında veri madenciliği yöntemi kullanımının daha sınırlı olması bu alanda hala araştırmaya açık boşlukların bulunduğunu göstermektedir. Ülkemizde ise toplanan ulusal kaza verisinin sıklıkla betimsel istatistiksel çıkarımlar ile analiz edildiği, veri madenciliği yöntemini kullanan çok az çalışma olduğu bunlarında belli bir sektöre ve kaza çeşidine odaklandığı görülmektedir. Bu kapsamda literatürde ulusal iş kazaları verisinin, veriden öğrenmeyi sağlayan içgörüler edinmek amacıyla bütünsel analizine dair bir çalışmaya rastlanmamıştır. Bu açığın giderilmesi amacıyla bu çalışma kapsamında İSG kaza verisinin ilgililere karar desteği sunacak bir sisteme dönüştürülmesini sağlayacak bir analiz modelinin 2013-2017 yılları arasında toplanan ulusal kaza verisi üzerinde uygulanarak geliştirilmesi amaçlanmaktadır. Geliştirilen analiz modeli sürekli büyüyen iş kazaları verisinde kullanılarak yeni karar destek verileri oluşturulabilir ve yeni iç görüler ile karar alıcılara sunulabilir.

Özet olarak, bu çalışmanın özgün yönü;

(i) Toplanan ulusal iş kazası verisinin, geleneksel istatistiksel yöntemlerden farklı olarak, veri madenciliği gibi çıkarımsal yöntemlerle, sektörel, kaza tipi vb. parçalara ayırmadan, bir bütün olarak ele alınmasını sağlaması,

(ii) İş kazaları zincirlerini kümeleme yöntemi ile ortaya çıkarırken sadece iş kazası oluşumunda ve şiddetinde sınıflandırma algoritması tarafından etkili bulunmuş faktörleri kullanması,

(iii) Toplanan ham kaza verisinin, karar alıcılar için destek sistemine dönüştüren bir yöntem önermesidir.

2
MAKİNE ÖĞRENMESİ YÖNTEMİ

Veri madenciliği, veri toplama, temizleme, işleme, analiz etme ve verilerden yararlı içgörüler elde etme tekniklerini içeren bir şemsiye terimdir. Veri madenciliği, pazarlamadan sağlığa, üretimden coğrafi araştırmalara kadar çok geniş bir yelpazede yer alabilen veri setlerinden değerli bilgiler elde etmek amacıyla doğrusal regresyon, kümeleme gibi istatistiksel yöntemleri, sinir ağları ve karar ağaçları gibi yapay zeka teknikleri ile birleştirir [64]. Veri madenciliği uygulamaları genellikle dört konuda gruplanır: ilişki örüntüsü madenciliği, kümeleme, sınıflandırma ve aykırı değer tespiti [81].

Sınıflandırma problemlerinde kullanılan makine öğrenmesi metodolojileri ise temelde üç grupta incelenir; gözetimli ve gözetimsiz öğrenme yöntemleri ile pekiştirmeli öğrenme yöntemi. Gözetimli öğrenme yöntemlerde verinin ayrıldığı sınıflar bellidir ve öğrenme aşamasında verinin ait olduğu sınıf etiketi kullanılır. Bu yöntemde amaç öğrenme aşamasından sonra ait olduğu sınıf etiketi olmayan verilerin etiketini tahmin etmektir. Gözetimsiz yöntemlerde ise veri önceden her hangi bir grup etiketi içermez. Bu yöntemde amaç ise verinin içindeki ilişkileri kullanarak saklı grupların ortaya çıkarılmasıdır. Gruplar analizi gerçekleştiren kişi tarafından sonradan etiketlenebilir. Pekiştirmeli öğrenme yönteminde ise öğrenme dışardan gelen geribildirimler ile artarak devam etmektedir.

2.1 Veri Ön İşleme

Makine öğrenmesi yöntemlerinde en önemli aşamalardan biri veri ön işleme aşamasıdır. Özellikle çok boyutlu ve kirli verilerde öğrenme aşaması öncesi gerçekleştirilen bu ön işleme, öğrenmede ortaya çıkan modelin doğruluğunu arttırır. Veri ön işleme içerisinde eksik veri işleme, boyut indirgeme, çoklu doğrusallık analizi ve sıfır varyans analizi gibi aşamaları içermektedir.

Veri madenciliğinde eksik veri sorunu araştırmacıların ve uygulamacıların sıklıkla karşı karşıya kaldığı bir sorundur. Eksik veri işlemede eksik verilerin neden eksik

oldukları önemli bir bilgildir. Eksik verinin rastgele olup olmaması başka bir deyişle sistematik bir hata veya yanlılık sebebiyle eksik olması analist tarafından kesin olarak bilinemeyebilse de uygun istatistiksel yöntemlerin seçimi için yol göstericidir [82]. Veride bulunan bir gözlemin eksik olma sebebi gözlenen veriyle ilişkilendirilebildiğinde literatürde bu duruma Rastgele Eksik (Missing at Random – MAR) denmektedir ve eksik veri işleme açısından en iyi durumdur. Rastgele eksiğin özel bir durumu olan Tamamen Rastgele eksik (Missing Completely at Random - MCAR) ise eksik verinin verideki diğer değişkenlerin değerlerinden istatistiksel olarak bağımsız olması durumunda ortaya çıkar. Bu durumda eksik veriye sahip gözlemler, tüm gözlemlerin basit bir rastgele örneği gibidir [83]. Bu iki durum da literatürde görmezden gelinebilir (ignorable) eksiklik olarak da isimlendirilmektedir [84]. Eksik veri işlemede sorun çıkaran durum ise Rastgele Eksik Olmayan Veriler (NMAR) olarak adlandırılan, veride bulunan bir gözlemin eksik olma sebebinin kendisiyle ilişkili olması durumudur. Bu durumda eksik verinin tahmini için veride kullanılacak hiç bir ilişki bulunmuyor olur ve eksik veri işlemenin ötesine geçen başka yöntemlerle ele alınması gerekir[82].

Eksik veri sorununu ele almada bir çok farklı yöntemler kullanılmaktadır. Literatürde karşılaşılan geleneksel yöntemler arasında eksik veri içeren satırların veya bazen tüm bir sütunun silinmesi veya eksik verinin ortalama, regresyon gibi yöntemlere dayalı basit atamalar yoluyla doldurulması yöntemleri bulunmaktadır. Eksik verinin kategorik özellikte veri olması durumunda eksik veriyi NA kategorisi olarak ele almak da mümkündür. Geleneksel yöntemler en hızlı ve basit çözümü oluşturmakla birlikte yanlı sonuç oluşturma olasılığı da en yüksek yöntemlerdir [82]. Eksik verinin silinmesi ayrıca eksik verinin bulunduğu gözlemde yer alan diğer bilgilerin de siliniyor olması sebebiyle veriden öğrenmeyi azaltabilmektedir. Literatürde eksik verinin %10'u geçtiği durumlarda istatistiksel analizlerin yanlı sonuçlar çıkarabileceği de kabul edilmektedir [83]. Geleneksel yöntemlerde oluşan yanlı sonuç etkisinin azaltılması için çoklu atama yöntemleri iyi bir alternatif oluşturmaktadır. Bu tez çalışmasında, numerik değerlerde bulunan eksik verinin işlenmesinde, çoklu atama yöntemlerinden biri olan Zincirleme Denklemlerle Çok Değişkenli Atama (Multivariate Imputation by Chained Equations – MICE) yöntemi R programlama dilinde geliştirilen "mice" paketi kullanılarak uygulanmıştır [85]. Mice paketi standard olarak öngörücü ortalama eşleme (pmm - predictive mean matching) yöntemi ile atamaları gerçekleştirir ve genel koşullar altında iyi sonuçlar veren bir yöntemdir [85]. MICE algoritması ilk adım olarak gözlenen verilerden rassal örneklemler yaparak eksik veri kümesinin m adet kopyasını oluştur. Her bir veri kümesi için eksik veri içeren değişkenleri, oluşturduğu zincirleme (sıralı) denklemleri kullanarak değişkenden değişkene atamalar yaparak tamamlar. Bayesyen temellere dayanan bu sürecin sonunda birbirinden farklı

atamalara sahip m adet tamamlanmış veri kümesi elde edilir. Atama aşamasının ardından analizde kullanılacak veri kümesinin oluşturulması/seçilmesi için eksik verisi giderilmiş m adet veri kümesi farklı istatistiksel yöntemlerle incelenir [85].

Makine öğrenimi modellerinde değişken seçimi adımı da, öğrenmeye giren veri boyutunu azaltması ve öğrenme süresini kısaltması sebebiyle önemli adımlardan biridir. Modeli sadece amaç değişken üzerinde yüksek etkisi bulunan değişkenlere indirmek doğruluğunu arttırır (accuracy), anlaşılmasını kolaylaştırır (yorumlanabilirlik/interpretability) ve analizi hızlandırır [86]. Öznitelik seçim metotları temelde 3 ana kategoriye ayrılmaktadır; filtreleme, sarmalayıcı ve gömülü yöntemler.

Filtreleme yöntemlerinde Pearson Korelasyonu ve Ki-Kare en sık kullanılan iki yöntemdir [55]. Sarmalayıcı yöntemler bir alt özellik grubuna sahip modeller oluşturur ve model performanslarını ölçer. Bunun için iki tip arama stratejisi vardır. İlki boş bir öznitelik kümesi ile başlanan ve her seferinde gruptaki özniteliklerin tek tek eklenerek özniteliklerin kalitesinin test edildiği ileri arama stratejisidir. Diğeri ise tüm öznitelik kümesinden başlayan ve en iyi öznitelik alt kümesi bulunana kadar adım adım en kötü performans gösteren öznitelikleri eleyen geri arama stratejisidir. Gömülü yöntemlerde ise öznitelik seçimi bazı Lasso veya Rassal Orman gibi makine öğrenmesi modellerinin sağladığı iç görülerle elde edilir.

Bahsedilen metodların bazıları minimum öznitelik ile en yüksek performansı sağlamayı amaçlarken diğerleri tüm önemli öznitelikleri ortaya çıkarmayı amaçlar [87]. Bu nedenle, özellik seçme metodu, verinin özelliklerine ve elde edilmek istenen sonuca göre belirlenir. Bu çalışmada iş kazları zincirini oluşturan tüm önemli özelliklerin bulunması amaçlandığından, öznitelik seçimi metodu olarak R kütüphanesinde bulunan Boruta öznitelik seçim fonksiyonu kullanılmıştır [88]. Boruta, veri setine iteratif olarak eklediği gölge değişkenler ile veri setindeki önemli ve önemsiz nitelikler arasında istatistiksel olarak anlamlı bir ayrım elde etmek için, rastgele orman çalışması gerçekleştiren ve değişkenlerin tahmin edici değişkene olan yararlarını ölçen bir yöntemidir [88]. Boruta fonksiyonunun tercih edilmesindeki bir diğer özelliği ise büyük boyutlu verilerde iyi performans göstermesidir [87]. Boruta yönteminin ana fikri, istatistiksel testler ve birkaç rassal orman çalışması kullanarak gerçek tahmin değişkenlerinin önemini, rastgele sözde gölge değişkenlerinkiyle karşılaştırmaktır. Her çalıştırmada, tahmin değişkenleri seti, her bir değişkenin bir kopyası eklenerek ikiye katlanır. Bu gölge değişkenlerin değerleri, gözlemler arasında orijinal değerlere izin verilerek ve dolayısıyla sonuçla olan ilişki yok edilerek üretilir. Korelasyonları ortadan kaldırmak için eklenen değişkenler kendi içinde karıştırılır. Genişletilmiş veri seti üzerinde bir rassal orman eğitilir ve değişken önem değerleri

toplanır. Her gerçek değişken için, önemini tüm gölge değişkenlerin maksimum değeriyle karşılaştıran bir istatistiksel test yapılır. Önemli ölçüde daha büyük veya daha küçük önem değerlerine sahip değişkenler, sırasıyla önemli veya önemsiz olarak beyan edilir. Tüm önemsiz değişkenler ve gölge değişkenler kaldırılır ve önceki adımlar, tüm değişkenler sınıflandırılana veya önceden belirlenmiş sayıda çalıştırma gerçekleştirilene kadar tekrarlanır [88].

Makine öğrenmesinde kategorik tipteki verilerle çalışırken genel yaklaşım kategorik veriyi numerik veriye dönüştürme teknikleri uygulamaktır. En basit dönüştürme tekniği ise kategorik değişkenin seviyelerinin her birini, bulunması durumunda 1 değeri alan ve bulunmaması durumunda 0 değeri değeri alan yapay değişkenlere dönüştürülmektedir [63]. Dönüştürülen veri setinde yer alan yapay değişkenlerin modele yararlı bilgi sağlayıp sağlamadığı varyans analizi ile değerlendirilmektedir. ve bağımlı değişkenin tahmininde rol oynayamayacak kadar düşük varyansa sahip değişkenlerin veri setinden çıkartılması amaçlanmıştır. Bazı algoritmalar sıfır varyans özelliklerinden etkilenmez. Bununla birlikte, sıfıra yakın varyansa sahip özellikler de bir modele çok az bilgi sunar. Ayrıca, belirli bir örneğin o özellik için yalnızca tek bir benzersiz değer (baskın değer) içerme olasılığı yüksek olduğundan, yeniden örnekleme sırasında sorunlara neden olabilirler [89]. Bu nedenle çalışmada veri işleme aşamasında gerçekleştirilen yapay değişken oluşturma aşamasından sonra varyans analizi gerçekleştirilmesi gerekmektedir.

Veri ön işlemenin son aşamasında ise veri setindeki değişkenlerin kendi aralarında yüksek bir ilişkiye sahip olup olmadığını inceleyen çoklu korelasyon (Multicollinearity) analizi bulunmaktadır. Amaç değişken (target) üzerinde aynı açıklayıcılığı sahip birden fazla değişken varsa bunların tek bir değişkene indirgenmesi modelin sadeleştirilmesi açısından önemlidir. Analizin bu aşaması için tüm grupların veri setine spearman sıra korelasyon testi uygulanması ve aralarında %80'den daha fazla doğrusallık bulunan değişkenlerden Boruta analizinde daha az açıklayıcı değere sahip olanların veri setinden çıkartılması hedeflenmiştir.

2.2 Tahminleme

2.2.1 Stokastik Gradyan Artırma (SGB) Algoritması

Ağaçların Artırılması (Tree Boosting) algoritmaları, kayıp/maliyet fonksiyonu kullanılarak gerçekleştirilen ve parametrik olmayan bir optimizasyon algoritmaları olarak tanımlanmaktadır [90]. Gradyan arttırma ise maliyet fonksiyonunu minimize etmek amacıyla, öğrenmeyi sağlayacak fonksiyonların içinde, negatif gradyan yönüne işaret eden fonksiyonu, yinelemeli olarak seçerek, maliyet fonksiyonunu optimize

eder. Diğer arttırma yöntemleri gibi, gradyan arttırma da temel (zayıf) öğrenicileri (ağaçlar) yinelemeli bir şekilde tek bir güçlü öğrenicide birleştirir [91].

Gözetimli sınıflandırma problemlerinde, bir girdi değişkeni vektörü x ve bir çıktı değişkeni y bulunur. Amaç, girdi değişkenlerinin değerlerini kullandığında ortaya çıkardığı sonuç ile çıktı değişkenine en iyi yaklaşan fonksiyonu $\hat{F}(x)$ bulmaktır. Amaca bir kayıp fonksiyonunun $L(y, F(x))$, minimize edilmesiyle ulaşılır. Gradyan arttırma algoritmasının adımları ve açıklamaları aşağıda verilmektedir [92].

1. Gradyan artırma algoritması ilk adımında temel bir kayıp fonksiyonu belirler.

   $F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$

   Bu adımda gösterilen y_i gerçek değer, γ tahminlenen değerdir.

2. Daha sonra belirlenen ağaç sayısına (M) ulaşana kadar oluşturulan tüm iterasyonlar için aşağıdaki adımlar gerçekleştirilir:

   for $m = 1, \ldots, M$

   2.1 Sözde reziduallerin hesaplanması:

   $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ for $i = 1, \ldots, n$

   r_{im} değeri i gözleminin m ağacındaki hata değerini (rezidual) vermektedir.

   2.2 Girdi değişkenlerin (x_i) ve çıktı değişkenin hesaplanan sözde rezidullerin (r_{im}) olduğu veri seti ($\{(x_i, r_{im})\}_{i=1}^{n}$) ile karar ağacının oluşturulması.

   2.3 Model çarpanlarının (reziduallerin) hesaplanması.

   for $j = 1, \ldots, j_m$

   $\gamma_{jm} = \arg\min_{\gamma} \sum_{x_i \in R_{jm}} L\left(y_i, F_{m-1}(x_i) + \gamma\right)$

   2.4 Modelin güncellenmesi:

   $F_m(x) = F_{m-1}(x) + \alpha \sum_{j=1}^{j_m} \gamma_{jm}(x_i \in R_{jm})$

   Her gözlemin hata değeri tahmin için bir önceki ağaçtaki tahmin değeri $F_{m-1}(x)$, öğrenme hızı (α) ve hesaplanan rezidual (γ_{jm}) ile yeni tahmin hesaplanır.

   İkinci adımın tüm alt adımları belirlenen ağaç sayısı kadar (M) tekrarlanır.

3. Model fonksiyonunun oluşması:

   $F_M(x)$

Üçüncü adımda her gözlem için son ağaçta ortaya çıkan tahmin son tahmin olarak kullanılır. Son tahmin, Adım 1'de hesaplanan ortalamaya, ağaçların

tahmin ettiği tüm reziduallerin toplamıyla öğrenme oranıyla çarpımının eklenmesiyle bulunmuş olur.

Stokastik Gradyan Arttırma (Stochastic Gradient Boosting - SGB) algoritması ise Friedman tarafından geliştirilen, gradyan arttırma algoritmalarının iyileştirilmiş bir türevidir [93]. SGB algoritması, diğer gradyan arttırma algoritmalarından farklı olarak, her yinelemede, tam eğitim kümesi yerine rastgele ve yerine konulmadan alınan bir alt örnek eğitim verisi kullanır. Bu alt örnek bir sonraki ağacın oluşturulmasında kullanılır. Sonraki aşamalar yukarıda verilen gradyan arttırma adımları ile aynıdır. Algoritmaya eklenen bu rastgelelik sayesinde öğrenme aşamasında ortaya çıkan aşırı uyum gösterme (overfitting) azalır ve tahmin doğruluğu artar [92].

SGB algoritmasının kullandığı maliyet/kayıp fonksiyonları regresyonda hata karesi (squared error) iken sınıflandırma problemlerinde kayıp fonksiyonu olarak çok terimli sapma (multinomial deviance) kullanılmaktadır [93]. Sınıflandırmada her yinelemede K en küçük kareler ağacı oluşturulur. İkili sınıflandırmada (K=2) sadece tek bir ağaç kullanılır [92].

Stochastic Gradient Boosting (SGB) literatürde Boosted Regression Trees (BRT), ve Gradient Boosted Machine (GBM) gibi farklı isimlerle ifade edilebilmektedir. Bu çalışmada stokastik gradyan arttırma modelleri R kütüphanelerinde bulunan "gbm" paketi kullanılarak gerçekleştirilmiştir [94].

2.2.2 Tahmin Doğruluğu Yaklaşımı

Makine öğrenmesi yönteminde kullanılan algoritma öğrenme esnasında oluşturduğu modeli test veri setine uygular ve her bir veri satırının sınıflarını tahmin eder. İkili sınıflandırmalarda sınıflar pozitif ve negatif olarak adlandırılır. Sınıflandırma temelli makine öğreniminde algoritmanın performansı, tahminlenen ve gerçek değerlerin 4 farklı kombinasyonunu içeren karışıklık matrisi adı verilen bir matris ile gösterilmektedir (Tablo 2.1). Karışıklık matrisinde doğru pozitif tahminler DP ile, yanlış pozitif tahminler YP ile, doğru negatif tahminler DN ile ve yanlış negatif tahminler YN ile gösterilmektedir. Algoritmanın tahminleme performansında kullanılan tüm metrikler bu dört değere dayanmaktadır. Çalışmada kullanılan metrikler Denklem 2.2, 2.3, 2.4, 2.5, 2.6, 2.7 ve 2.8'te tanımlanmıştır.

Tablo 2.1 Karışıklık matrisi

	Gerçek Pozitif	Gerçek Negatif	
Pozitif Tahmin	DP	YP	Kesinlik
Negatif Tahmin	YN	DN	
	Duyarlılık DPO	Özgüllük YPO	

Doğruluk (Accuracy), hem pozitif sınıfta hem de negatif sınıfta doğru olarak sınıflandırılan örneklerin yüzdesini vermektedir.

$$Dogruluk = \frac{DP + DN}{(DP + YN + YP + DN)} \tag{2.1}$$

Ancak test veri setinde yer alan veri satırları iki sınıfa dengesiz dağıldığında sadece doğruluk metriği kulllanarak karar vermek yanıltıcı olabilir. Örneğin çok düşük gözlem sayısına sahip bir pozitif sınıf tahmininde tüm tahminlerin negatif olarak belirlenmesi ve hiç bir veri satırının pozitif olarak tahminlenmemesi durumunda bile DN değeri DP değerine göre çok yüksek olacağı için ortaya çok yüksek bir doğruluk çıkacaktır ancak hedef pozitif sınıfların tahmini olduğunda bu metrik algoritmanın performansı konusunda yanıltıcı olacaktır.

Dengesiz sınıf dağılımlarında algoritmanın tüm test veri setini baskın gruba sınıflandırmaması için farklı yöntemler kullanılmaktadır. SMOTE ve Down Balancing gibi yöntemlerin yanı sıra bazı makine öğrenmesi algoritmaların sunduğu olasılık değerlerindeki eşik değer üzerinden de sınıf dengesizliği problemi çözülebilmektedir. Birçok makine öğrenmesi algoritmasının oluşturduğu modeller 0 veya 1 gibi bir sınıf etiketini her bir örnek için bir olasılık değeri oluşturarak belirler. X'in sınıfı, Y'nin bir değeri belirttiği $Y = p(C|X)$ notasyonu Y'nin C sınıfına ait olma olasılığını belirtir. Algoritma $p(C|X)$ için 0-1 arasında bir olasılık puanı oluşturur ve genelde eşik değeri 0.5 belirleyerek bu değerin altında kalanları bir sınıfa, üstünde kalanları diğer sınıfa ayırır.

Sınıf dengesizliği gibi durumlar bu eşik değerin daha düşük veya daha yüksek belirlenmesiyle de ele alınabilmektedir. Eşik değer belirlenirken modelin tüm eşik değerlerdeki performanın eğrisi incelenir. Böylelikle 0.5 değeri yerine

tüm eşik değerler incelenir ve sınıf dengesizliğini en iyi şekilde ele alan bir eşik seçilebilir. Sınıflama dengesizliklerinde tahminleme 0-1 aralığındaki olasılık değerleri kullanılarak yapıldığında, hangi olasılık aralığının hangi sınıfa ait olduğunu belirlenebilmesi mümkün olmaktadır. Böylelikle çok az sayıda pozitif örneğin olduğu bir veride 0.5 eşik değeri neredeyse tüm örnekleri negatif sınıflarken eşik değeri düştükçe pozitif sınıflama oranı da artmaya başlar.

Dengesiz sınıf dağılımlarında, pozitif sınıfın doğru tahmin edilmesinin önemi büyükse, pozitif sınıfa (azınlık sınıfa) odaklanan kesinlik ve duyarlılık metrikleri ile bu metriklerin ortalamasını alan dengeli doğruluk (Balanced Accuracy) metriği 2.4 ve bu metriklerinin harmonik ortalaması olan F skoru tercih edilir [52]. Öğrenme sırasında ise ROC değeri kullanılır[95].

Duyarlılık/Hassasiyet (Sensitivity or Recall) doğru tahmin edilen pozitif sınıfların gerçek pozitif sınıfa oranını gösterir. Başka bir deyişle “Gerçek pozitiflerin yüzde kaçı doğru tahmin edildi?” sorunun cevabıdır.

$$Duyarlilik = \frac{DP}{(DP + YN)} \tag{2.2}$$

Kesinlik (precision) ise doğru tahminlenen pozitif örneklerin, pozitif tahmin edilen örneklere oranıdır. Başka bir deyişle "Pozitif tahminlenen örneklerin yüzde kaçı doğrudur?" sorusunun cevabıdır.

$$Kesinlik = \frac{DP}{(DP + YP)} \tag{2.3}$$

Kesinlik positif sınıf tahmininden emin olmayı hedeflerken, duyarlılık pozitif sınıfdaki verilerin hiç birini kaçırmamayı hedefler.

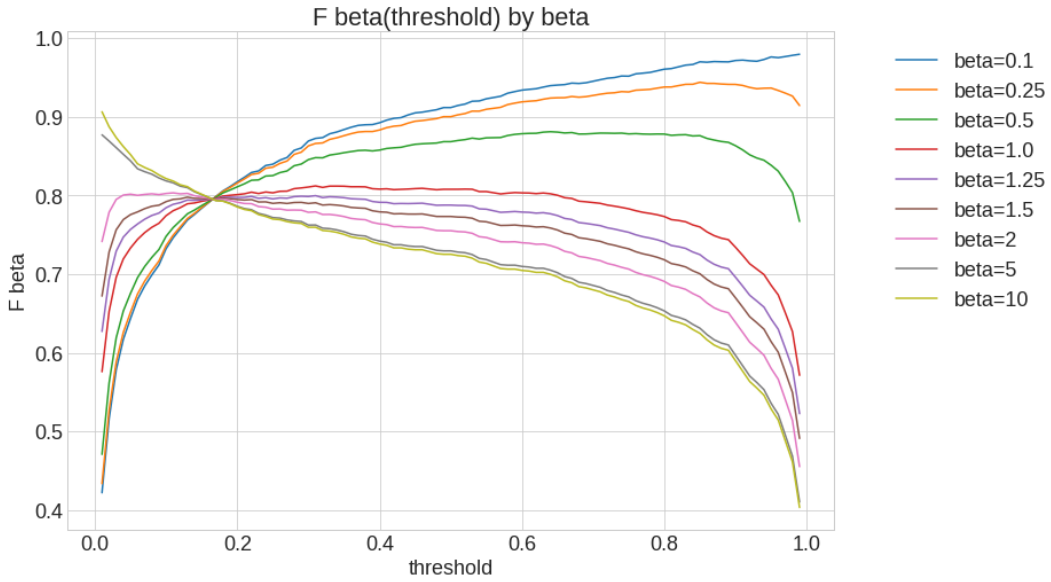
$$Dengeli\ Dogruluk\ = \frac{Duyarlilik + Kesinlik}{2} \tag{2.4}$$

Duyarlılık ve kesinlik arasında yapılan analize uygun bir denge oluşturmak gerekir. Duyarlılığın arttırılması pozitif sınıfta yer alan örneklerin doğru tahmin edilmesini yükseltir (DP artar) ancak bununla birlikte negatif sınıfta yer alan örneklerde çok daha fazla pozitif olarak tahmin edilmeye başlanır (YP de artar). Bu durumda duyarlılık yükseltmek kesinlik değerini azaltacaktır. Duyarlılık ve kesinliğin eşit dengede bulunması durumunda performans ölçümü için dengeli doğruluk metriği tercih edilebilinir 2.4. Ancak yapılan analizde pozitif değerlerin doğru tahminlenmesi,

yanlış tahminlenen pozitif örneklerden çok daha kritikse duyarlılığın daha yüksek olduğu bir denge daha uygun olacaktır. Bunun için kesinlik ve duyarlılığın harmonik ortalaması olan F_β skoru kullanılmaktadır .

$$F_\beta = (1+\beta)^2 * \frac{Kesinlik * Duyarlilik}{\beta^2 * Kesinlik + Duyarlilik} \tag{2.5}$$

F_β skoru, kesinlik ve duyarlılığın harmonik ortalamasıdır. Bu skor dengeli doğruluktan farklı olarak duyarlılık ve kesinlik değerleri arasındaki fark yükseldikçe daha düşük bir sonuç ortaya koyar. Skorun hesaplanmasında beta değeri duyarlılık ve kesinlik değerleri arasındaki önceliğe göre belirlenir. Örneğin F_1 skorunda duyarlılık ve kesinlik arasında bir öncelik farkı yokken, F_2 skoru duyarlılığı iki kat daha değerli kılmaktadır. Benzer şekilde $F_{0.5}$ ise kesinliği iki kat değerli kılar. Farklı eşik değerleri ile değişen F_β skorları Şekil 2.1'da gösterilmektedir [95].

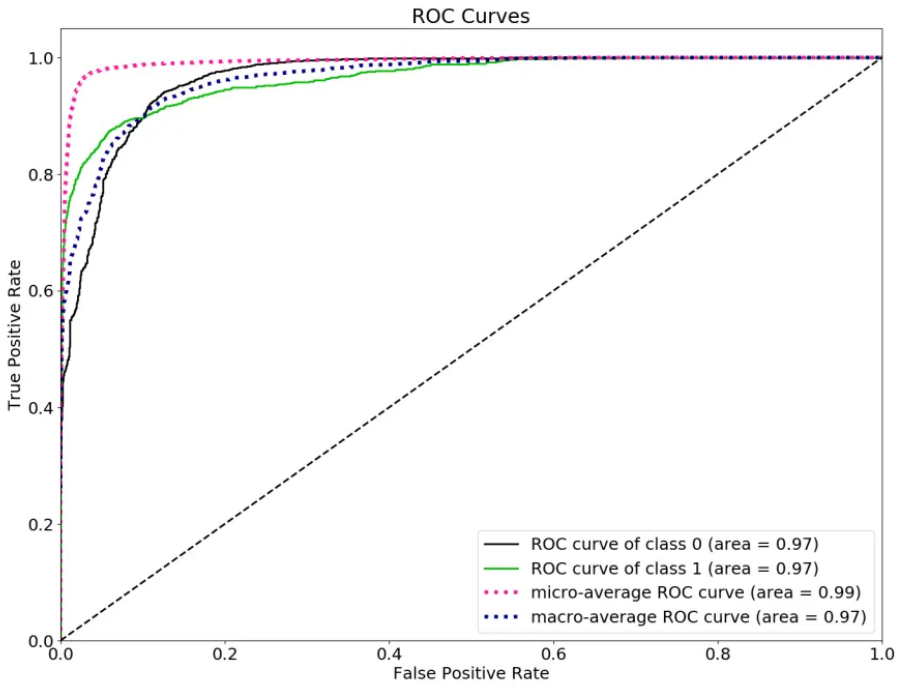


Şekil 2.1 Farklı eşik değerlere göre F_β skoru

Duyarlılık **DPO** (Doğru Pozitif Oran) olarak da adlandırılmaktadır. **YPO** (Yanlış Pozitif Oran) ise Kesinlik metriğinden farklı olarak, doğru pozitiflerin tüm pozitif tahminlere oranına bakmak yerine, yanlış pozitif tahminlerin gerçek negatiflere oranına bakar. YPO aynı zamanda **Özgüllük** (Specificity) olarak adlandırılır. Bu iki oran da aynı duyarlılık ve kesinlik gibi bir denge ilişkisi içerisindedir. Doğru Pozitif Oran arttıkça Yanlış Pozitif Oran da artar.

Bu iki oranın alabileceği olasılık değerlerinden oluşan eğriye **ROC** (receiver operation characteristics) eğrisi adı verilir (Şekil 2.2) [95]. ROC eğrisinde DPO x-eksenini, YPO ise y-eksenini oluşturur. İdeal durumda doğru pozitif sınıf tahminlerinin oranının 1

ve yanlış negatif sınıf tahminlerinin oranının 0 olması gerektiğinden, (0,1) noktası mükemmel performansı gösteren noktadır.

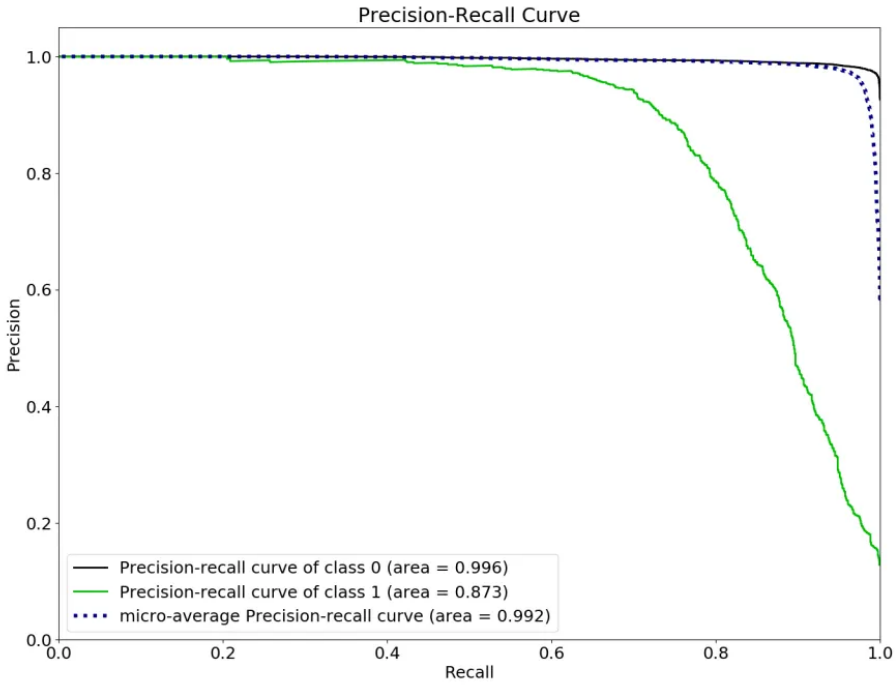


Şekil 2.2 ROC Eğrisi

ROC eğrisi tüm eşik değerler için farklı bir çok değer aldığından ve karşılaştırılması zor olduğundan eğrinin altında kalan alan hesaplanılarak performans değerlendirme metriği olarak kullanılır. ROC eğrisin altında kalan alan **AUC** olarak adlandırılır. Modelleri sınıflar arasında ne kadar ayırt edebildiğini anlatır. AUC yükseldikçe, model tahmin etmede daha iyi demektir. AUC ayrılabilirliğin derecesini veya ölçüsünü temsil eder ve 0 ile mükemmel performansı gösteren 1 değeri arasında bir değer alır. AUC rastgele sınıflandırma performansında 0.5 değeri alır. Şekil 2.2'de sağ alt köşede yer alan alan değerleri her eğrinin AUC değerini göstermektedir.

Dengesiz sınıf dağılımlarında görülebilen çok az sayıda doğru veya çok az sayıda yanlış tahmin AUC skorunda büyük değişiklikler oluşturabilir ve AUC azınlık sınıfının sadece bir kaç örnekten oluşması durumunda yanıltıcı olabilir. Bu nedenle böyle durumlarda DPO-YPO yerine Kesinlik-Duyarlılık eğrisi tercih edilir. Kesinliğin y-eksenini, Duyarlılığın ise x-eksenini oluşturduğu düzlemde farklı olasılık eşiklerine göre alınan değerleri gösteren eğriye **Kesinlik-Duyarlılık eğrisi** (PR Curve) denir (Şekil 2.3) [95]. Mükemmel performans gösteren model (1,1) koordinatlarındaki nokta ile gösterilir. İyi performans gösteren modeller bu noktaya doğru eğilen bir eğri ile temsil edilirkenü performans düştükçe model veri kümesindeki pozitif örneklerin

sayısıyla orantılı bir hassasiyetle grafik üzerinde yatay bir çizgiye dönüşmektedir. Dengeli bir veri kümesi için bu 0,5 olacaktır. PR eğrisinin azınlık sınıfına odaklanması, onu dengesiz ikili sınıflandırma modelleri için etkili bir performans değerlendirme metriği haline getirir .



Şekil 2.3 Kesinlik-Duyarlılık Eğrisi

Aynı ROC eğrisinde olduğu gibi Kesinlik-Duyarlılık Eğrisininde altında kalan alan da performans metriği olarak kullanılmaktadır ve literatürde PR-ROC olarak isimlendirilmektedir.

Çapraz entropi kaybı (Cross-entropy loss) veya log kaybı (LogLoss), çıktısı 0 ile 1 arasında bir olasılık değeri olan bir sınıflandırma modelinin performansını ölçer. Tahmin edilen olasılık gerçek değerden uzaklaştıkça LogLoss artar. Mükemmel bir modelin LogLoss değeri 0'dır. LogLoss değerinin hesaplandığı formül 2.6'de verilmiştir. Benzer şekilde Ortalama Kare Hata (The mean squared error- MSE) veya Brier skoru olarak adlandırılan metrik de tahmin edilen olasılık değerinin gerçek değerden farkını ele alan başka bir metriktir ve formülü 2.7'de verilmiştir. İki formülde de y_i gerçek değeri p_i sınıf tahmin değerini ve n gözlem sayısını vermektedir [96].

$$LogLoss = -\frac{1}{n}\sum_{i=1}^{n} y_i log_2 p_i + (1-y_i) log_2 (1-p_i) \tag{2.6}$$

$$Brier(MSE) = -\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2 \tag{2.7}$$

Eşik değer belirlemede kullanılan bir başka performans metriği de Youden J İstatistiği Metriğidir ve formülü 2.8'de verilmiştir. Formüldeki t değeri Youden J İstatistiği değerini maksimize eden eşik değeri belirtmektedir [96].

$$J_{max} = max_t\{Duyarlilik(t) + Ozgulluk(t) - 1\} \tag{2.8}$$

2.2.3 Sınıf Dengesizliği Yaklaşımı

Ölümlü kaza içeren sınıflandırmalarda ölümlü kaza sayısının diğer gruba/gruplara oranla çok az olması veride dengesizlik oluşturmaktadır. Bu tip sınıf dengesizliklerinin yarattığı problemler ve çözümler makine öğrenmesi literatüründeki önemli araştırma konularından birisi olmuştur. Sınıf dengesizliğinin oluşturduğu problemin çözümü sınıfları oluşturan gözlem sayılarını birbirine yakınlaştırmaktır. Veri setinde daha fazla gözlem sayısına ait sınıf çoğunluk sınıf, daha az gözlem sayısına ait sınıf ise azınlık sınıf olarak adlandırılmaktadır. Sınıf dağılımı dengelemesi literatürde genellikle 3 farklı yöntemle ele alınmaktadır. Bu yöntemler yeniden örnekleme yöntemleri kullanılarak aşırı örnekleme (oversampling) olarak bilinen azınlık sınıfın örneklerinin çoğaltılması, alt örnekleme (undersampling) olarak bilinen çoğunluk sınıfın örneklerininin çıkartılması ve SMOTE (Sentetik/Yapay Azınlık Aşırı Örnekleme) yöntemi olarak bilinen sentetik veri üretme yöntemleridir.

Aşırı örnekleme, mevcut veri noktalarını basitçe çoğaltırken, SMOTE, istatistiksel yöntemler ile yeni veriler oluşturarak aşırı örnekleme yapar [97]. Azınlık sınıfı örnekleri, k azınlık sınıfı en yakın komşularına katılan doğru parçalarından örnekler eklenerek oluşturulur (SMOTE, k=5 kullanır). Bu sayede aşırı örneklemede mevcut verilerin tam olarak kopyalanmasından kaynaklanabilecek olası aşırı öğrenme (overfitting) asorunun önüne geçilebilmektedir.

Alt Örneklem (RUS) yönteminde eğitimde kullanılan verinin sayısında azalma oluşturduğu için tez çalışmasında ölümlü olmayan kazaların rastgele silinmesi ile gerçekleştirilen alt örneklem yönteminin yanı sıra, ölümlü kazaların denge oluşuncaya kadar sentetik olarak veriye eklenmesi (SMOTE) yöntemi ve ROC eğrisi üzerinden Youden indexini maksimize eden eşik değer belirleme yöntemleri de uygulanmış ve en iyi performansı gösteren yöntem ile analize devam edilmesi amaçlanmıştır.

3
BİRLİKTELİK KURALI ÇIKARIMI

3.1 Apriori Algoritması

Birliktelik kuralı madenciliğinde amaç bir verideki sık görülen öğelerin bulunması ve güçlü birliktelik kurallarının üretilmesidir. Veri tabanlarında sık geçen öğe kümelerinin keşfedilmesinde en sık kullanılan algoritma Rakesh Agrawal ve Ramakrishnan Srikant tarafından 1994 [98] yılında tanıtılan ve temelinde iteratif (tekrarlayan) bir niteliğe sahip olan Apriori algoritmasıdır [99]. Apriori Algoritmasının ismi, bilgileri bir önceki adımdan elde ettiği için ingilizcede "prior (önceki)" kelimesinden gelmektedir. Bu algoritmada özellikler ve veri, var olma/olmama durumlarının gösterildiği "boolean" ilişki kuralları ile değerlendirilir. Bu çalışmada tüm kategorik özellikler yapay değişken olarak tanımlandığı için boolean gösterime uygundur.

Birliktelik kurallarında, öncül (LHS) ve sonuç (RHS) öğeleri araştırmacı tarafından belirlenir ve bulunacak birliktelik kurallarının kısıtı olarak destek ve güven alt sınırları belirlenir. Birliktelik kuralları üretilirken öncelikle araştırmacı tarafından belirlenen destek (Support) değerinin üstündeki öğeler bulunur ve kalan öğeler arasından belirlenen güven (Confidence) değeri üstündeki kurallar üretilir. Son aşamada üretilen kurallara ait kaldıraç (Lift) değeri kullanılarak araştırmacı için ilginç kurallar seçilebilir. Birliktelik kuralları oluşturulan kullanılan metrikler ve açıklamaları Tablo 3.1'de verilmektedir.

Tablo 3.1 Birliktelik kuralları metrikleri

Metrik	Gösterim	Formül
Destek	Destek(X -> Y) = Destek(Y -> X)	= P(X -> Y) = Frekans(X, Y) / N
Güven	Güven(X -> Y) Güven(Y -> X)	= P(Y|X) = P(X ve Y)/P(X) = Destek(X -> Y)/P(X) = Frekans(X, Y) / Frekans(X) = P(X|Y) = P(X ve Y)/P(Y) = Destek(Y -> X)/P(Y) = Frekans(X, Y) / Frekans(Y)
Kaldıraç	Kaldıraç(X -> Y) = Kaldıraç(Y -> X)	= P(X ve Y)/P(X)P(Y) = P(X|Y)/P(X) = P(Y|X)/P(Y) = Güven(Y -> X)/P(X) = Güven(X -> X)/P(Y)

Öncül öğeleri içeren durumların tüm durumlara oranı P(X), sonuç öğesini içeren durumların tüm durumlara oranı P(Y) biçiminde gösterilir. P(X -> Y) X öncül öğe kümesinin Y sonuç öğesi ile birlikte görülme olasılığını belirtir. P(X -> Y), P(X ve Y) olarak da gösterilmektedir.

Destek Değeri, X öncül öğe kümesinin Y öğe kümesi ile birlikte görülme olasılığını ifade eder [98]. X öncül öğelerin Y sonuç öğesi ile birlikte görülme frekansının toplam durum sayısına bölünmesiyle bulunur. Dolayısı ile sınırları 0–1 arasındadır. Aşağıdaki gibi gösterilir;
Destek(X -> Y) = P(X -> Y) = Frekans(X, Y) / N

Kaza analizi özelinde destek değeri, bir kaza zincirinin (X) belli bir şiddette (Y) görülmesinin toplam kazalar içinde görülme sıklığını ifade eder. Başka bir deyişle kaza şiddetinin belli bir kaza zinciri ile birlikte ortaya çıkış sıklığının tüm kazalara oranı olarak ifade edilebilir. Bu değerin bir dezavantajı veri setinde çok seyrek olarak meydana gelen ve bu sebeple de bu değer için belirlenen eşiğin altında kalan ancak yine de ilginç ve potansiyel olarak değerli kurallar üretebilecek öğelerin elenmesidir. Bu nedenle çalışmada eşik değer çok düşük tutularak eleme kaldıraç değeri üzerinden gerçekleştirilmiştir. Bu değerin başlıca özelliği eşik değer üzerindeki sıklıkla görülen bir öğe kümesinin tüm alt kümelerinin de sık olduğu anlamına gelen aşağı doğru

kapanma (down-ward closure) özelliğine sahip olmasıdır.

Güven Değeri, X öncül öğe kümesinin oluşması durumunda oluşabilecek sonuç öğeleri içinde Y sonuç öğesinin oluşma olasılığıdır[98]. X öncül öğe kümesinin Y sonuç öğesi ile birlikte görülme frekansının, X öncül öğe kümesinin görülme frekansına bölünmesiyle bulunur;
Güven(X -> Y) = P(Y|X) = P(X ve Y)/P(X) = Destek(X -> Y)/P(X) = Frekans(X, Y) / Frekans(X)
Güven(X -> Y) değeri ile Güven(Y -> X) değeri eşit değildir.
Güven(Y -> X) değeri Y sonuç öğresinin oluşmuş olması durumunda ortaya çıkmış öncül öğe kümelerinden X öncül öğe kümesinin oluşma olasılığını belirtir ve aşağıdaki gibi gösterilir;
Güven(Y -> X) = P(X|Y) = P(X ve Y)/P(Y) = Destek(Y -> X)/P(Y) = Frekans(X, Y) / Frekans(Y)

Kaza analizi özelinde güven değeri, Y kaza şiddetiyle ortaya çıkan bir X öğe kümesine sahip bir kaza zincirinin ortaya çıkış sıklığının, X öğe kümesinin ortaya çıkış sıklığına oranını belirtir. Başka bir deyişle Y kaza şiddeti ile sonuçlanan bir X kaza zincirinin, X kaza zincirlerinde hangi oranda tekrarlandığını belirtir. Bu değer aşağı doğru kapanma özelliğine sahip değildir. Bir kaza zinciri çok sık görülmemiş ise güven değeri Y sonuç öğesinin frekansına duyarlı olduğu için aldatıcı olabilir. Bu sebeple doğası gereği düşük frekansa sahip yüksek şiddet sonuçları araştırılırken güven değeri yerine kaldıraç değerini kullanmak daha iyi sonuç vermektedir.

Kaldıraç değeri, X öncül öğe kümesinin, Y sonuç öğesi ile birlikte görülme olasılığının, İstatistiksel olarak bağımsız olmaları durumunda ayrı ayrı görülme olasılıklarına oranını ifade eder[100];
Kaldıraç(X -> Y) = P(X ve Y)/P(X)P(Y) = P(Y|X)/P(Y) = Güven(X -> Y)/P(Y)
Kaldıraç(X -> Y) ve Kaldıraç(Y -> X) değerleri P(X ve Y)/P(X)P(Y) formülü sebebiyle eşit değerlere sahiptirler.
Kaldıraç(Y -> X) = P(X ve Y)/P(X)P(Y) = P(X|Y)/P(X) = Güven(Y -> X)/P(X)

Kaldıraç değeri, X meydana geldiğinde Y olasılığının ne kadar artacağını veya azalacağını gösterir. Bu değer 1 olduğunda X öncül öğe kümesinin oluşma olasılığı ile Y sonuç öğesinin oluşma olasılığı istatistiksel olarak birbirinden bağımsızdır. Başka bir deyişle her hangi birinin ortaya çıkışı diğerini etkilememektedir. Kaldıraç değeri 1'in üzerinde olduğunda iki olay pozitif ilişkiye sahiptir; 1'den küçük olduğunda negatif ilişkiye sahiptirler.

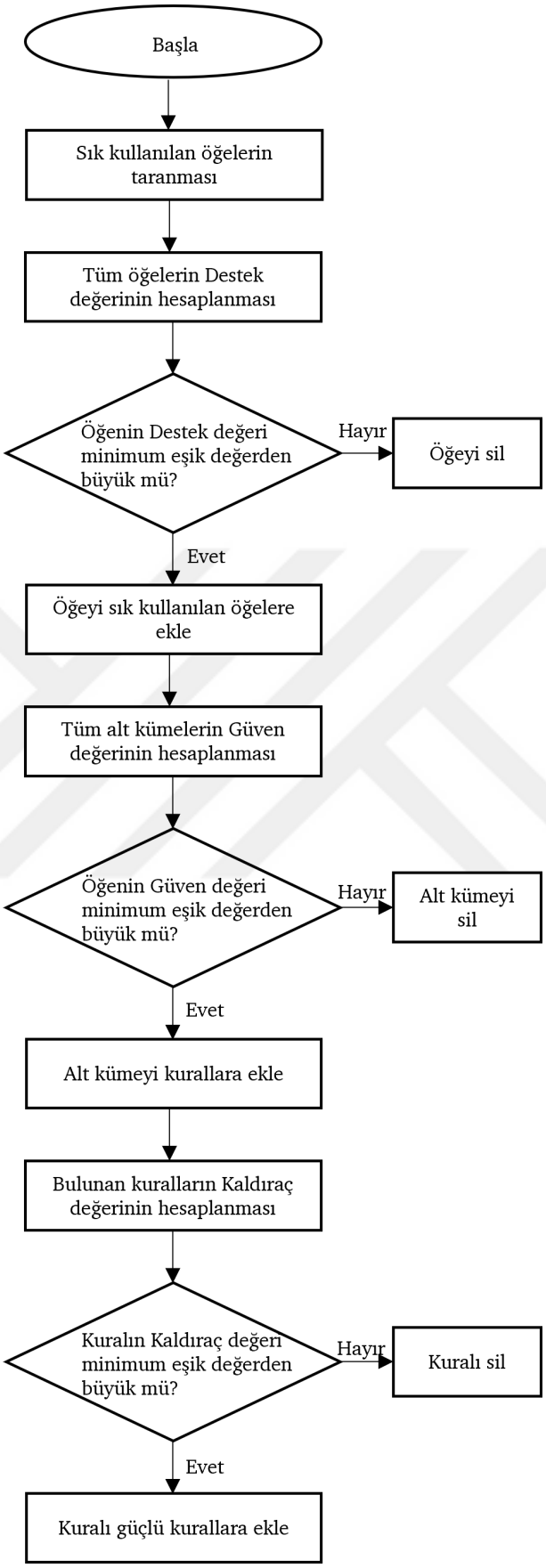
Kaza analizi özelinde Kaldıraç değeri bize, bir X öncül kaza zincirinin oluşması durumunda Y sonuç öğesinin ortaya çıkış olasılığının ne kadar yüksek olduğunu söyler.

Kaldıraç değerinin 1'den büyük veya küçük olması bir X öncül kaza zinciri ile bir Y sonuç öğesinin bağımlı ilişkisini ortaya koyar. Kaza zincirindeki faktörlerin Y sonuç öğesi ile ilişkisini anlamak için pozitif ilişki yani 1'den büyük kaldıraç değerlerini değerlendirmek daha uygundur. Bu değer aşağı doğru kapanma özelliğine sahip değildir ve frekanstan etkilenmez.

Kaza analizinde nadir ortaya çıksa da ilginç ve önemli kaza zincirleri için destek değeri düşük olabilir ve ölümcül kazalar gibi daha nadir olaylar söz konusu olduğunda güven değeri yanıltıcı olabilir. Bu nedenle, bir birliktelik kuralının gücünü belirlemek için destek ve güven değerlerinin yanı sıra kaldıraç değerinin değerlendirilmesi gerekir. Algoritmanın akışı gereği kurallar minimum destek ve güven değeri eşiğine göre keşfedildmesi kuralların kaldıraç değerlerine göre elenmeden önce gerçekleştiği için optimum destek ve güven eşiklerini belirlemek çok önemlidir.

Bu üç önemli değer için bir örnek vermek gerekirse. X öncül kaza zinciri öğeleri ve Y kaza sonucu öğesi için Destek değerinin %20'i, Güven değerinin %70 ve Kaldıraç değerinin 3.5 olduğunu varsayarsak; Analiz edilen tüm kazaların %20'sinde X öncül kaza zinciri öğelerinin Y kaza sonucu öğesi ile birlikte ortaya çıktığını, X öncül kaza zinciri öğelerinden oluşan kazaların %70'inin Y şiddetiyle oluştuğunu, X öncül kaza zinciri öğelerinin ortaya çıkması durumunda Y kaza sonucu öğesinin ortaya çıkma olasılığının 3.5 kat arttığını gösterir.

Apriori algoritması, Şekil 3.1'de gösterilen akış şemasında açıklanmış ve kural oluşturma adımları aşağıda verilmiştir:
(1) Veri tabanının taranması ve sık kullanılan tüm öğelerin bulunması.
(2) Sık görülen öğe kümelerinin Destek değerlerinin oluşturulması ve minimum eşik değere göre elenmesi.
(3) Alt öğe kümelerinin Güven değerlerinin oluşturulması ve minimum eşik değere göre elenmesi.
(4) Kalan öğe kümelerinin Kaldıraç değerlerinin oluşturulması ve minimum eşik değerden küçük/büyük değerlere sahip kuralların elenmesi.
(5) Kalan öğe kümelerinin güçlü birliktelik kuralları olarak kullanılması.



Şekil 3.1 Apriori algoritması akış şeması

4
ÖN ARAŞTIRMA

Bu tez çalışmanın ilk aşamasında inşaat sektörüne ait NACE kodlarının filtrelendiği bir veri seti ile ön araştırma uygulama gerçekleştirilmiştir. 20 bağımsız ve 1 bağımlı değişken ile toplam 25.577 kaza verisi kullanılarak 5 farklı makine öğrenmesi algoritması; Çok terimli Lojistik Regresyon (MLR), Destek Vektör Makineleri (SVM), Tek C5.0 Ağacı (C5), Stokastik Gradyan Artırma (SGB) ve Sinir Ağı (NN), iki farklı iş kazası şiddeti sınıflaması üzerinde denenmiştir. Makine öğrenmesi algoritmalarının modellerinin performansı dengeli doğruluk metriği üzerinden değerlendirilmiş ve ikili sınıflamada SVM ve SGB eşit performans gösterirken çoklu sınıflamada SGB en iyi performansı göstermiştir.

Modelin tahmin gücünü gösteren önem puanlarına göre en yüksek puana sahip değişkenler le SVM ve SGB algoritmaları kullanılarak tahmin sonuçları yeniden üretilmiş ve sadece en yüksek önem puanlarına sahip değişkenleri kullanmanın ölümlü kaza sınıflandırmasında doğruluğunu artırtığı ortaya çıkarılmıştır. Elde edilen sonuçlar, düzeltici ve önleyici eylemlerin belirlenmesinde, makine öğrenmesinin ortaya çıkardığı kaza şiddettine katkıda bulunan faktörlerin etkililiğine dair bir çıkarım ortaya koymuştur; Kaza veritabanlarının karmaşık yapılarının içinde kaza şiddettini en belirleyici birkaç faktörler ile ilgili önlem almak bile ciddi kazaların çoğunu önleyebilir. Ön çalışmanın sonucunda inşaat sektörü ile ilgili ortaya çıkan faktörlerin önereceği düzeltici önlemlere düşmeyi önleme önlemlerinin alınması, işyeri denetimlerine çalışan sayısına göre öncelik verilmesi ve çalışanın yaşı ve deneyimine göre ek güvenlik önlemlerinin alınması örnek verilebilir.

4.1 Yöntem

Bu ön araştırma çalışmasında 5. Bölüm içerisinde açıklanan ham veri işverenin ekonomik faaliyetini gösteren (NACE) kodlarından inşaat sektörü ile ilgili olan 41, 42 ve 43 NACE kodlarına göre filtrelenmiştir. Ön araştırma çalışmasında eksik veri tamamlama uygulanmadığı için "diğer" ve "bilgi yok" seçeneklerine sahip girişler

verilerden çıkarılmış, 20 bağımsız ve 1 bağımlı değişken ile toplam 25.577 kaza verisi ikili (Ölümlü/Ölümlü Olmayan) ve üçlü (Ölümlü/Majör/Minör) sınıflama tahmininde beş farklı makine öğrenmesi algoritması kullanılarak analizleri gerçekleştirilmiştir. Üçlü sınıflama tahmininde majör 15 iş gününden fazla kayıpla sonuçlanan kazaları ifade etmektedir.

Ön araştırma çalışmasında makine öğrenmesi algoritmaları seçilirken, sınıflandırıcıların temsil çeşitliliği ve öğrenme stili göz önünde bulundurulmuştur. Karşılaştırma amacıyla, hem doğrusal (MLR) hem de doğrusal olmayan modeller (SVM, C5.0, SGB, NN), ağaç tabanlı (C5.0, SGB) ve ağaç tabanlı olmayan modeller (MLR, SVM, NN) ve hem tekli (MLR, SVM, C5.0, NN) hem de toplu öğrenme modelleri (SGB) kullanılmaktadır. Doğrusal modellerde, DT(C5.0), NN ve SVMnin, iş kazaları sonuç tahminlerinde en çok kullanılan algoritmalar olduğu Bölüm 1.1'de verilen literatür taramasında tespit edildiğinden, performanslarının literatürdeki çalışmalarla karşılaştırılabilmesi amacıyla seçilmiştir. Doğrusal model için, hem ikili hem de çoklu sınıfda kullanılabildiği için MLR seçilmiştir. Toplu öğrenme yönteminde ise Tixier vd.'nin [20] SGB'nin daha iyi performans gösterdiğini ancak RF kadar yaygın kullanılmadığını ifade etmesi üzere performansını değerlendirme amacıyla popüler RF yerine SGB seçilmiştir.

Analiz edilen veriye ait nitelikler Tablo 4.1'de açıklanmıştır.

Tablo 4.1 Ön araştırma veri seti değişken sayıları

	İkili Sınıflama / Ölümlü / Ölümlü Olmayan		Üçlü Sınıflama / Majör / Minör	
	Ölümlü Kategorik	Numerik	Ölümlü Kategorik	Numerik
Başlangıç	15	5	15	5
Yapay Değişken	121		121	
Boruta Sonrası	39	2	53	5
Çoklu korelasyon	30	1	44	4
Toplam		31		48

Makine öğrenmesi algoritmalarına hazırlık amacıyla numerik ve yapay değişkenlerden hedef değişkenle maksimum alaka düzeyine sahip olanların seçimi için Boruta analizi gerçekleştirilmiş ve yeteri kadar ilgili bulunmayan değişkenler veri setinden silinmiştir. Daha sonra çoklu korelasyon testine göre yüksek korelasyona sahip değişkenlerden boruta analizinde daha az ilgi değerine sahip olanlar veri setinden silinmiştir. İkili sınıflamada kalan 31 ve üçlü sınıflamada kalan 48 değişkenin oluşturduğu verisetleri öğrenme ve test için ikiye ayrılmıştır. 2013-2016 yılları arasında gerçekleşen

kazalar öğrenmede kullanılmış ve test verisi olarak seçilen 2017 yılının ilk yarısında gerçekleşen kazalar tahmin edilmeye çalışılmıştır. 25.570 kazanın 5.524'ü test setine seçilerek bölünme oranı 78:22 olmuştur. Eğitim ve test setinde bulunan kaza sayıları Tablo 4.2'de verilmiştir.

Tablo 4.2 Ön araştırma eğitim/test veri seti kaza sayıları

		İkili Sınıflama		Üçlü Sınıflama	
	Total	Ölümlü	Ölümlü Olmayan	Ölümlü	Majör	Minör
Eğitim Seti	20.053	190	19.863	190	1.461	18.402
Test Seti	5.524	35	5.489	35	274	5.215

Makine öğrenmesi algoritmalarının eğitim aşamasında eğitim seti, 5 kat 10 tekrar CV konfigürasyonu ile eğitim ve doğrulama setine bölünmüştür. Konfigürasyonlar arasından ikili sınıflamada ROC metriğine, üçlü sınıflamada LogLoss metriğine göre optimum modeller seçilmiş ve test verileri bu konfigürasyondaki modeller kullanılarak tahmin edilmiştir. İkili sınıflandırmada en yüksek ROC değeri en iyi performansı verirken üçlü sınıflandırmada en düşük LogLoss değeri en yüksek performansı vermektedir. Eğitim veri setinde kazaların sayıca görece düşük olması nedeniyle makine öğrenmesi algoritma modeli tarafından tüm kazaların ölümlü olmayan bir kaza olarak sınıflandırılmasını önlemek amacıyla sınıf dağılımı dengesi için ön işlem uygulanmıştır. Veri dengeleme yöntemleri arasından veri setinin birçok kategorik değişken içermesi ve aşırı örnekleme tekniklerinin, sınıf dengesizliği sorununun çözümünde açık bir avantaj sağlamamakla birlikte hesaplama maliyetini artırması nedeniyle Rastgele Alt Örnekleme (RUS) seçilmiştir [101, 102]. Bu yöntemde, göreceli bir sınıf dengesi sağlanana kadar veri setindeki ölümlü olmayan kazalar rastgele elimine edilmektedir. Tablo 4.3'de Model Performansı bölümünde ikili sınıflama ve üçlü sınıflama için optimum model için kullanılan değerler gösterilmekteyken Tahmin Performansı bölümünde ise modellerin sınıflandırma performansını gösteren dengeli doğruluk (DD) değerleri verilmiştir.

Tablo 4.3 Ön araştırma optimum modeller

	İkili S. ROC	Üçlü S. LogLoss	İkili S. DD	Üçlü S. Ort. DD	Ölümlü DD	Majör DD	Minör DD
MLR	0,8021	1,0650	0,7276	0,6239	**0,7747**	0,5146	0,5824
SVM	**0,8062**	1,0050	**0,7415**	0,6159	0,7088	**0,5388**	0,6001
C5	0,7152	1,4058	0,7123	0,5582	0,6735	0,4678	0,5332
SGB	**0,8041**	**1,0027**	**0,7414**	**0,6276**	0,7340	0,5443	**0,6044**
NN	0,7743	1,0473	0,6808	0,5015	0,5045	0,4994	0,5007

4.2 Bulgular

Tablo 4.3 incelendiğinde ikili sınıflandırmada SVM ve SGB, üçlü sınıflandırmada ise SGB ve MLR benzer performans sergilemektedirler. Benzer dengeli doğruluk değerlerine sahip sınıflandırıcı modellerin performanslarındaki anlamlı farklılıkları test etmek için marjinal homojenlik testleri kullanılmıştır. İkili sınıflandırmada için parametrik olmayan bir test olan McNemar testi, üçlü sınıflandırmada ise genelleştirilmiş McNemar testi olarak anılan Stuart-Maxwell testi uygulanmış ve ikili sınıflandırmada SVM ile SGB arasında anlamlı bir performans farkı bulunmazken (p-değeri $= 1$) üçlü sınıflandırmada MLR ve SGB arasında anlamlı bir performans farkı bulunmuş (p-değeri < 0.0001) ve SGB'nin daha yüksek dengeli doğruluk değerine sahip olması sebebiyle daha iyi performans gösterdiği bulunmuştur.

Makine öğrenmesi algoritma modellerinin tahminleme performanslarını karşılaştırdıktan sonra, en iyi performansı gösteren algoritma modelinde bulunan değişkenlerin ölümlü kazaların tahmin edilmesine yaptığı katkıyı gösteren Değişken Önemi metriğine dayalı bir sıralama yapılmıştır. Bu sıralamada en yüksek önem puanlarına sahip değişkenler, en iyi tahmin veren ve modele en çok katkı sağlayan değişkenlerdir.

Analizin son aşamasında, en yüksek önem puanlarına sahip değişkenleri kullanmanın ölümlü kaza sınıflandırmasında doğruluğunu artırıp artırmadığını görmek için sadece bu değişkenler ile en iyi performans gösteren algoritmalar kullanılarak tahmin sonuçları yeniden üretilmiştir. Üçlü sınıflandırmada SGB en iyi performans gösteren algoritma olarak seçildiği için, İkili sınıflandırmada, SVM ve SGB aynı performansı göstermesine rağmen, farklılık oluşturması açısından SVM algoritması tercih edilmiştir. İkili sınıflandırmada en önemli ilk 31 değişken ile Dengeli Doğruluk 0.7415'ten 0.7964'e yükselirken üçlü sınıflandırmada ilk 48 değişken ile Dengeli Doğruluğun 0.7340'tan 0.7528'e yükseldiği bulunmuştur.

Bekleneceği üzere, şantiyelerdeki en büyük tehlikenin düşme olması nedeniyle, kayma ve zemin seviyesinden düşme ile ilgili sapma, ölümlü kazaların gerçekleşmesinde en önemli faktör olarak bulunmuştur. Sırasıyla, şirketteki toplam işçi sayısı, işçinin yaşı ve son işyerindeki iş deneyimi, kaza şiddeti sonuçları için en önemli belirleyiciler olarak bulunmuştur. Makine öğrenmesi algoritmalarının her iki sınıflandırma grubunda da ölümlü kazaları daha iyi bir doğrulukla tahmin etmiştir. Ancak üçlü sınıflandırmada majör ve minör kazaların ayırt edilememesi ve ölümlü kazaların tahminlenmesindeki doğruluğunda düşmesi sebebiyle ikili ve üçlü sınıflandırma karşılaştırıldığında, ikili sınıflandırmaların daha iyi doğrulukla sonuçlandığı bulunmuştur.

İlk denemede önemli bulunan değişkenlerle tekrarlanan analizlerde doğruluğun

artmasıyla, sadece en önemli faktörlerle ilgili önlem almanın bile ciddi kazaların büyük çoğunluğunu engelleyebileceği sonucuna varılabilmektedir. Bulunan en önemli değişkenlere özgü verilebilecek önleyici tedbirlere örnek olarak yaş ve deneyim gibi özellikleri sebebiyle yüksek risk grubunda bulunan çalışanlara eğitim ve benzeri ek önleyici tedbirlerin alınması, müfettişlerin kaza önleme için yaptığı denetimlerde çalışan sayısına göre şantiyeleri önceliklendirmeleri verilebilir. Ayrıca bu analiz sonucunda ortaya çıkan bulgular düşme önleyici eylemlerin öneminin altını çizmekte ve şantiyelerde ölümleri önlemek için düşmeyi önleyen toplu korunma yöntemlerini barındıran kesin bir uygulama önermektedir.

4.3 Tez Çalışmasına Katkısı

Ön araştırma çalışmanın neticesinde kaza şiddetini belirleyen faktörlerin ikili sınıflandırmada daha iyi performans oluşturduğunu ortaya koyması sebebiyle, tez çalışmasındaki şiddet sınıflandırmasında çoklu sınıflandırma yerine dört farklı ikili sınıflandırma kullanılmıştır. Şiddet gruplarının sınıflandırma tahminlerinde ön araştırma çalışmasında SVM ve SGB performansları aynı olmasına rağmen SVM algoritmasının hesaplama yükünün fazla olması ve bu sebeple de eğitim süresinin çok uzun olması sebebiyle tez çalışmasının tahminleme aşaması için SGB algoritması tercih edilmiştir. SGB algoritması ön araştırma çalışmasından farklı olarak 5 kat 10 tekrar CV konfigürasyonu yerine litertürde önerilen 10 kat 5 tekrar CV konfigürasyonu kullanılarak en iyi parametreye sahip model belirlenmiştir [103]. Ön araştırma çalışmasında veri dengesizliğinin çözümü için kullanılan Rastgele Alt Örneklem (RUS) yönteminde eğitimde kullanılan verinin sayısında azalma oluşturduğu için tez çalışmasında ölümlü olmayan kazaların rastgele silinmesi yanısıra ölümlü kazaların denge oluşuncaya kadar sentetik olarak veriye eklenmesi (SMOTE) yöntemi ve ROC eğirisinde daha dengeli bir eşik değer belirleme yöntemleri uygulanmış ve alt örneklem yöntemi ile kıyaslanarak en iyi performansı gösteren yöntem belirlenmiştir.

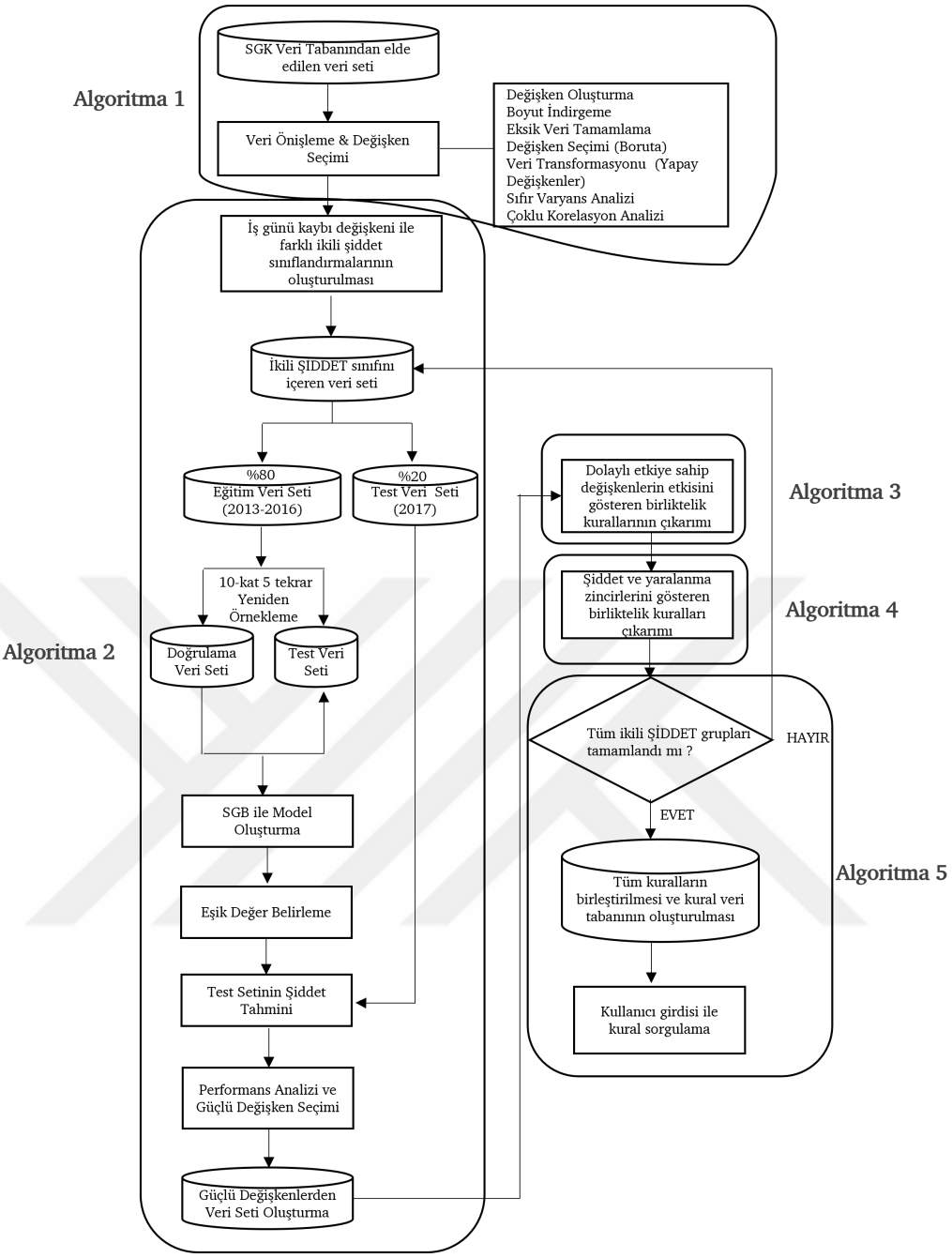
Yapılan ön araştırma çalışması tahmin gücü en yüksek faktörlerin ortaya çıkarılma metodunda yol gösterici olmuş ve bu sayede her bir şiddet grubu için en iyi tahminleyici değişkenler ve yaranın vucuttaki yeri, yaralanma türü değişkenleri ile birliktelik kuralları oluşturulmuştur. Birliktelik kurallarında daha fazla bilgiye sahip olabilmek amacıyla tez çalışmasında ek değişkenler kullanılmıştır. Bu sayede İSG faaliyetleri gerçekleştiren işyerleri ve denetleme mekanizmaları için karar destek verisi ortaya konmuştur.

5
ÖNERİLEN YÖNTEM

Bu bölümde önceki bölümlerde sunulan yöntemler Sosyal Güvenlik Kurumundan elde edilen ulusal kaza verisinin bütününe uygulanmış ve ulusal iş kazaları verisinin büyük veri kapsamında analizi için bir yöntem önerisi sunulmuştur.

Kazaların oluşumunda etkili faktörlerin belirlenmesi için öncelikle ulusal kaza verisini makine öğrenmesi için uygun hale getirecek yöntemler uygulanmış ve veri eğitim ile test setine ayrılmıştır. Bölüm 4'de verilen ön araştırma sonucu çoklu sınıflandırma yerine dört faklı ikili sınıflandırma kullanılarak farklı şiddetlerde oluşan kazalara etki eden faktörlerin belirlenmesi hedeflenmiştir. Eğitim seti 10 kat 5 tekrar CV konfigürasyonu ile SGB algoritması kullanılarak en iyi parametreye sahip model belirlenmiştir. Ayrıca sınıf dengesizliği üç farklı yöntemle ele alınmış, her bir şiddet kategorisi için algoritmanın test veri seti üzerindeki performansı karşılaştırılmıştır. Her bir şiddet grubu için, belirlenen doğruluk oranı ile sınıflandıran faktörler, başka bir deyişle kaza sonucunun ortaya çıkmasında etkili faktörler tespit edilmiş, yaranın vucuttaki yeri, yaralanma türü değişkenleri de eklenerek birliktelik kuralları çıkartılmıştır. Ortaya çıkan kaza zincirleri incelenerek bir karar destek verisi oluşturulmuştur. Çalışmada benimsenen metodolojinin akış şeması Şekil 5.1'de verilmektedir.

Analizler AMD Ryzen 5 2600X 3.6Ghz CPU, Nvidia GTX3070 8GB GPU, 32GB RAM'e sahip Windows 10 masaüstü ve AMD EPYC 7282 16 çekirdek 2.79 Ghz İşlemci ve 30 GB RAM'e sahip bir Window 2019 server sanal makine olmak üzere iki farklı bilgisayar üzerinde RStudio 4.0.5 kullanılarak tamamlanmıştır. Analiz kodları GitHub websitesinde bulunmaktadır (https://github.com/frecal/doktora_tezi).



Şekil 5.1 Metodoloji akış şeması

5.1 Veri Ön İşleme

Bu çalışmanın yürütülmesi için gerekli olan ve kamu erişimine açık olmayan iş kazası kayıtları bilimsel bir araştırmada kullanılmak üzere, 11.05.2017 tarih ve 3037382 sayılı yazı ile Sosyal Güvenlik Kurumu'ndan talep edilmiş ve 26/06/2017-03/07/2017 tarih aralığındaki veriler veri tabanından temin edilmiştir.

Elde edilen veri setine ait değişkenler işyeri, çalışma ortamı, çalışma koşulları, olaylar zinciri, kazazede ve kaza ile ilgili olmak üzere toplam 5 ana gruba

ayrılmıştır. Değişkenlerin ayrıldığı gruplar, değişken isimleri ve açıklamaları Tablo 5.1'de verilmektedir. Numerik değişkenler "(n)" işaretiyle belirtilmiştir.

Tablo 5.1 SGK veri tabanında yer alan değişkenler

Grup	Değişken	Açıklama
İşyeri	ISYERI_NACE	İşyeri faaliyet kodu
	SIGORTALI_SAYISI (n)	İşyerinde çalışan sigortalı sayısı
	KAZA_IL	Kazanın gerçekleştiği il
Çalışma Ortamı	SG_KAZAGERORTAM	Kazanın gerçekleştiği anda kazalının çalışıyor, dinleniyor veya işyeri dışında olma durumu
	CC	Kazanin gerçekleştiği çalışma ortamı
	CO	Çalışma istasyonunun sabit veya sürekli olması durumu
Olaylar Zinciri	GENELFAA	Kazadan az önceki zamanda kazalının yürütmekte olduğu genel faaliyet
	OZELFAA	Kazadan az önceki zamanda kazalının yürütmekte olduğu özel faaliyet
	OZELFAA_MAG	Özel faaliyet sırasında kullanılan araç / gereç
	KAZA_KS	Kazaya sebep olan olay
	KAZA_MAG	Kazaya sebep olan materyal/araç/gereç
	YS	Yaralanmaya neden olan olay
	YARAMAG	Yaralanmaya sebep olan materyal/araç/gereç
Kazazede	SG_MESLEK_ISCO08	Kazalı meslek kodu
	SG_CINSIYET	Kazalı cinsiyet
	SG_MEDENIHAL	Kazalı medeni durum
	OGR_DURUM	Kazalı öğrenim durumu
	SG_SONIS_TEC (n)	Kazalının son iş yerindeki tecrübesi
	SG_TOP_TEC (n)	Kazalının toplam tecrübesi
	SG_YAS (n)	Kazalı yaşı
	SG_UCRETLIZINGUN (n)	Kazalının birikmiş izin gün sayısı
	CALIS_SURE (n)	İşbaşından kazaya kadar geçen süre
Kaza	SAAT_ARALIK	Kazanın gerçekleştiği saat aralığı
	KAZA_AY	Kazanın gerçekleştiği ay
	KAZA_YIL	Kazanın gerçekleştiği yıl
	KAZA_MEVSİM	Kazanın gerçekleştiği mevsim
	KAZA_HGUN	Kazanın gerçekleştiği hafta günü

Tablo 5.1 SGK veri tabanında yer alan değişkenler - (devamı)

Grup	Değişken	Açıklama
	YT	Yaranın türü
	YVY	Yaranın vücuttaki yeri
	SG_ISGUNUKAYBI (n)	Kaza sonrası iş günü kaybı

5.1.1 Veri Temizleme

Ham veride bulunan bazı değişkenler yeni özellikler üretmek için kullanılmış ve daha sonra veri setinden çıkarılmıştır. Kazalı son iş tecrübesi ve toplam iş tecrübesi değişkenleri, ilk ve son işe başlama tarihleri ile kaza tarihi değişkenleri kullanılarak elde edilmiştir. Kazalı doğum tarihi ve kaza tarihi kullanılarak veriye kazalının yaşı eklenmiştir. Kaza günü işe başlama saati ve kaza saati değişkenleri kullanılarak kazaya kadar geçen süre hesaplanmıştır. Kaza gün ve saati kullanılarak kazanın gerçekleştiği saat aralığı, ay, yıl, mevsim ve hafta gün özellikleri üretilmiştir. Kullanılan değişkenlerdeki verinin tutarsızlığı sebebiyle sıfırdan küçük bulunan ilk ve son iş tecrübe süreleri eksik veri olarak düzeltilmiştir. Ölümlü kazaların iş günü kaybı kaza sonrası iş göremezlik durumu değişkeni kullanılarak ESAW metodolojisinde belirtildiği üzere 7500 olarak düzeltilmiştir [7].

İş kazası bildirim ekranından bildirimi yapan kişi bir çok değişkenin seçimini yaparken "Diğer" ve "Bilinmiyor" seçeneklerini kullanabilmektedir. Kazanın betimlenmesinde etkinliği olmayan bu seçenekler NA (eksik veri) olarak yeniden kodlanmıştır. Veri boyutunu düşürebilmek amacıyla kazayı betimleyen değişkenlerde ana kategoriler kullanılmıştır. Değişkenlerde, Avrupa ülkelerinde olduğu gibi ülkemizde de Avrupa İstatistik Ofisi tarafından belirlenen kategoriler kullanılmaktadır [7].

Elde edilen ham veri incelendiğinde değişkenlerin 22'sinin hem kod (K) hem de açıklama (KA) olarak iki farkli kolonda tanımlandığı gözlenmiştir. Ancak satırların bazılarında K-KA eşleşmelerinde tutarsızlıklar bulunmuştur. Tutarsızlığın kök sebebi bilinmemektedir. K ve KA değişkenleri farklı satırlarda eksik verilere sahip olduğundan bu iki kolon eşleştirilerek eksik veri sayısının azaltılması amaçlanmıştır.

Bir kazanın iş kazası olarak kabul edileceği durumlar Sosyal Sigortalar ve Genel Sağlık Sigortası Yasası'nın 13. maddesinde çalışmadan sayılar haller üzerinden belirlenmiştir [104]. Belirlenen bu durumların bilgisi iş kazası verisini giren kişi tarafından seçilmektedir ve veri ambarında SG_KAZAGERORTAM_K ve SG_KAZAGERORTAM_KA olarak iki ayrı değişkenle belirtilmektedir. İki değişkenin eksik verilerinin eşleştirilmesi sonrası ortaya çıkan kaza sayıları Tablo 5.2'de verilmektedir.

Tablo 5.2 Çalışmadan sayılan hallere ait iş kazası sayıları

	İş kazası sayıları
Ara Dinlenmesinde	31.628
Çalisirken	787.900
Süt Izni Esnasinda	64
Isten Eve Kendi Arabasiyla Gelirken	990
Evden Ise Kendi Arabasiyla Giderken	1.189
Isten Eve Servisle Gelirken	7.987
Evden Ise Servisle Giderken	11.893
Bilinmeyen	50.901

Bu tez çalışmasının amacı iş yerlerindeki faktörlerden dolayı oluşabilecek iş kazalarının oluşumunda etkili faktörlerin belirlenmesi olduğu için giriş yapılan kazalardan sadece "ara dinlenmesinde", "çalışırken" ve "bilinmeyen" olarak seçilmiş kazalar filtrelenmiş toplam 870.429 satırdan oluşan iş kazası ile analize devam edilmiştir.

Kaza verisi çoğunluğu kategorik özellikte değişkenlerin oluşturduğu karışık tipte bir veridir. Başlangıç verisindeki eksik veri yüzdeleri Tablo 5.3'de verilmektedir. Numerik değişkenlerin içinde eksik veri bulunduran toplam tecrübe, son iş tecrübesi ve çalışma süresi değişkenlerinin eksik verisi Zincirli Denklemlerle Çok Değişkenli Doldurma Algoritması olan MICE algoritması kullanılarak tamamlanmıştır [85]. Algoritma tarafından rassal oluşturulan iki veri seti kopyasına (m=2) eksik verilerle ilişkili olabilecek toplam ve son iş tecrübesi, çalışma süresi, çalışan yaşı ve kazanın gerçekleştiği saat aralığı değişkenleri kullanılarak atamalar gerçekleştirilmiş ve iki veri setinin tahmin değerlerinin ortalaması kullanılmıştır.

Kategorik özellikteki değişkenlerde bulunan eksik veriler ise birkaç adımda ele alınmıştır. Öncelikle ham veride kod ve açıklama olarak iki ayrı kolonla verilen değerler biribirleriyle karşılaştırılmış ve bir diğerinde var olan bilgi diğerinde tamamlanmıştır. Daha sonra kaza oluşumunda direk etkili değişkenlerde eksik veriye sahip satırlar silinmiş diğerlerindeki eksik veriler yeni bir kategori olarak belirlenmiştir. Böylelikle eksik veri sebebiyle kaza sayısında daha az azalma olmuştur.

Tablo 5.3 Başlangıç veri setinde bulunan eksik veri yüzdeleri

Grup	Kodu	Eksik Veri Yüzdesi
İşyeri	FAALGRUP	% 0
	ISYERI_NACE	% 0
	SIGORTALI_SAYISI	% 0
	KAZA_IL	% 0
Çalışma Ortamı	CC	% 20,88

Tablo 5.3 Başlangıç veri setinde bulunan eksik veri yüzdeleri - (devamı)

Grup	Kodu	Eksik Veri Yüzdesi
	SG_KAZAGERORTAM	% 5,85
	CO	% 2,18
Olaylar Zinciri	GENELFAA	% 36,60
	OZELFAA	% 37,85
	OZELFAA_MAG	% 0
	KAZA_MAG	% 37,44
	KAZA_KS	% 0
	YS	% 47,08
	YARAMAG	% 0
Kazazede	SG_MESLEK_ISCO08_K	% 0,13
	SG_CINSIYET	% 0
	SG_MEDENIHAL	% 0,33
	OGR_DURUM	% 0,05
	SG_SONIS_TEC	% 1,95
	SG_TOP_TEC	% 0,03
	SG_YAS	% 0
	SG_UCRETLIZINGUN	% 0
	CALIS_SURE	% 6,79
Kaza	SAAT_ARALIK	% 0
	KAZA_AY	% 0
	KAZA_YIL	% 0
	KAZA_MEVSİM	% 0
	KAZA_HGUN	% 0
	YT	% 25,59
	YVY	% 24,58
	SG_ISGUNUKAYBI	% 0

Eşleştirme analizi aynı değişkene ait kod ve açıklama sutunlarının yanı sıra aynı özelliği ortaya koyan iki değişkende daha uygulanmıştır. Böylelikle hem gene eksik veri sayısı azaltılmış hem de değişken sayısı azaltılarak boyut indirgenmiştir. Çalışma ortamını niteleyen değişkenlerde ESAW metodolojisinde çalışma ortamı olarak açıklanan değişken SGK tarafından kazayı bildirenden iki farklı değişken ile istenmektedir [105];

İş Kazasının Gerçekleştiği Yer / Bölüm (KAZA_GERYERBOLUM): Kazanın gerçekleştiği yeri belirtir.

Çalışılan Çevre (CC_KOD): Kazadan hemen önce kazalının çalışmakta olduğu veya bulunduğu işyeri türü, çalışma alanı veya yeridir.

Bu iki değişkenin alt kategorileri aynıdır ve kullanım kılavuzunda da farkı detaylı olarak belirtilmemiştir [105]. İş kazasının bilgilerini giren tarafından çok farklı seçimler yapılmamış olduğu ön kabulu ile daha az eksik veriye sahip olması nedeniyle CC değişkeni analiz için seçilmiş ve bu değişkendeki eksik veri KAZA_GERYERBOLUM değişkeninden alınan veriler ile tamamlanmıştır.

Benzer şekilde ESAW metodolojisinde tek bir değişkenle ifade edilen kazalı mesleği ham veride iki farklı değişken ile verilmektedir [105];

Esas İşi / Mesleği: Kazalının eğitim-öğretim mezuniyetine göre sahip olduğu meslektir.

Görevi: Kazalının o işyerindeki çalıştığı resmi pozisyondur. Örneğin; tekniker olarak eğitim almış nitelikli bir personel, görev itibariyle yönetici olarak çalışıyor olabilir.

Bu iki değişkenin de alt kategorileri aynıdır. Bu sebeple daha az eksik veriye sahip SG_MESLEK_ISCO08 değişkeni analiz için seçilmiş ve bu değişkendeki eksik veri SG_GOREVI_K değişkeninden alınan veriler ile tamamlanmıştır.

Tablo 5.4 CC ve SG_MESLEK_ISCO08 değişkenleri eksik veri sayıları

Değişken	Eksik Veri Sayısı	Eksik Veri Yüzdesi
CC_K	181.750	0,21
CC_KA	181.750	0,21
SG_MESLEK_ISCO08_K	1.051	0,0013
SG_MESLEK_ISCO08_KA	1.051	0,0013

Eksik verinin tamamlanmasında uygulanan stratejilerden biri Kod (K) ve Açıklama(KA) değişkenlerindeki verinin eşleştirilmesi ve eksik verinin diğer değişkende bulunan veriyle tamamlanması olmuştur. Eşleştirme ile eksik verileri tamamlanan değişkenlerin eksik veri sayıları ve eşleştirme sonrası geldikleri durum Tablo 5.5' de verilmektedir.

Tablo 5.5 Değişkenlerin eksik veri sayıları

	Değişken	Başlangıç	Eşleşme sonrası	Yüzde
1	FAALGRUP_K	344.530	318.556	0,37
2	FAALGRUP_KA	318.576	318.556	0,37
3	KAZA_GERYERBOLUM_K	227.572	212.801	0,24
4	KAZA_GERYERBOLUM_KA	212.891	212.801	0,24
5	KAZA_GERYER_K	0	0	0,00
6	KAZA_GERYER_KA	50.713	0	0,00
7	CO_K	19.022	19.022	0,02
8	CO_KA	19.022	19.022	0,02
9	CC_K	216.801	203.811	0,23
10	CC_KA	203.904	203.811	0,23
11	NACE_K	0	0	0,00
12	NACE_KA	0	0	0,00
13	KAZA_IL_K	0	0	0,00
14	KAZA_IL_KA	2	0	0,00
15	SG_IGOR_K	0	0	0,00
16	SG_IGOR_KA	4.357	0	0,00
17	SG_IGORDURUM_K	452.896	435.275	0,50
18	SG_IGORDURUM_KA	435.275	435.275	0,50
19	YT_K	248.710	222.770	0,26
20	YT_KA	225.593	222.770	0,26
21	YVY_K	215.202	213.936	0,25
22	YVY_KA	220.131	213936	0,25
23	SG_MESLEK_ISCO08_K	1.170	1.170	0,00
24	SG_MESLEK_ISCO08_KA	58.866	1.170	0,00
25	SG_GOREVI_K	1.171	1.171	0,00
26	SG_GOREVI_KA	58.866	1.171	0,00
27	MEDENIHAL_K	2.887	2.887	0,00
28	MEDENIHAL_KA	2.887	2.887	0,00
29	OGR_DURUM_K	433	433	0,00
30	OGR_DURUM_KA	433	433	0,00
31	GENELFAA_K	344.530	318.556	0,37
32	GENELFAA_KA	318.576	318.556	0,37
33	KAZA_MAG_K	402.944	379.273	0,44
34	KAZA_MAG_KA	386.752	379.273	0,44
35	OZELFAA_K	361.546	329.421	0,38
36	OZELFAA_KA	329.649	329.421	0,38

Tablo 5.5 Değişkenlerin eksik veri sayıları - (devamı)

	Değişken	Başlangıç	Eşleşme sonrası	Yüzde
37	OZELFAA_MAG_K	472.169	451.344	0,52
38	OZELFAA_MAG_KA	462.553	451.344	0,52
39	YARAMAG_K	431.334	407.521	0,47
40	YARAMAG_KA	415.092	407.521	0,47
41	SG_KAZAGERORTAM_K	56.293	50.901	0,06
42	SG_KAZAGERORTAM_KA	50.906	50.901	0,06
43	YS_K	450.938	409.839	0,47
44	YS_KA	410.017	409.839	0,47

Bu eşleştirme analizi ile birlikte eksiksiz satır yüzdesi %8 den %16.2'ye çıkmış, eksik veri %16.6'dan %11.8'e düşmüştür (Tablo 5.6).

Tablo 5.6 Eşleştirme öncesi ve sonrası eksik veri sayıları

	Başlangıç	Eşleşme Sonrası
Satır sayısı	870.429	870.429
Eksiksiz satır sayısı	69.719 (% 8)	142.940 (%16.4)
Toplam veri	53.096.169	28.724.157
Toplam eksik veri	8.834.712 (%16.6)	3.393.599 (%11.8)

Son olarak kazanın oluşumda en etkili değişkenlerden önce "CC_K", "GENELFAA_K", "OZELFAA_K", "KAZA_KS_K", "YS_K" değişkenlerinden herhangi birinde eksik veri bulunan satırlar silinerek satır sayısı 870.429'dan 259.689'a inmiş, daha sonra "OZELFAA_MAG_K", "KAZA_MAG_K", "YARAMAG_K" değişkenleri incelenmiştir.

Olaylar zincirini niteleyen değişkenlerde kazayı bildirenden kaza ile ilgili üç farklı materyal/ekipman bilgisi istenmektedir. İlki kazalının kaza sırasında gerçekleştirdiği faaliyet esnasında kullandığı (OZELFAA_MAG), ikincisi kaza sırasında beklenmedik durum oluşturan (KAZA_MAG), üçüncüsü ise yaralanmaya sebep olan ekipman/materyal (YARAMAG). ESAW terminolojisinde belirtildiği üzere bu üç ekipman bazı kalarda tamamen farklı olabilirken bazı kazalarda ise aynı olabilmektedir. Veri incelendiğinde kaza veri girişi esnasında bu üçünün sıklıkla veri girişini yapan kişi tarafından aynı veya fark oluşturmayan benzerliklerde seçildiği görülmüştür.

Tablo 5.7üç değişkene de aynı araç gereçin seçildiği kaza yüzdesi %68.67'dır. Bu durum kazayı bildirenlerin bu konuda bilgi sahibi olmadığını ve bilinçli ayrım yapmadan kazayı bildirmiş olabileceğini göstermektedir. Bu sebeple bu

üç materyal/ekipmandan herhangi birinde eksik veriye sahip satırların hepsini silmektense sadece üçününde boş olduğu satırlar silinmiştir. Bu satırların da silinmesi sonucu toplam satır sayısı başka bir deyişle kaza sayısı 187.910'a inmiştir.

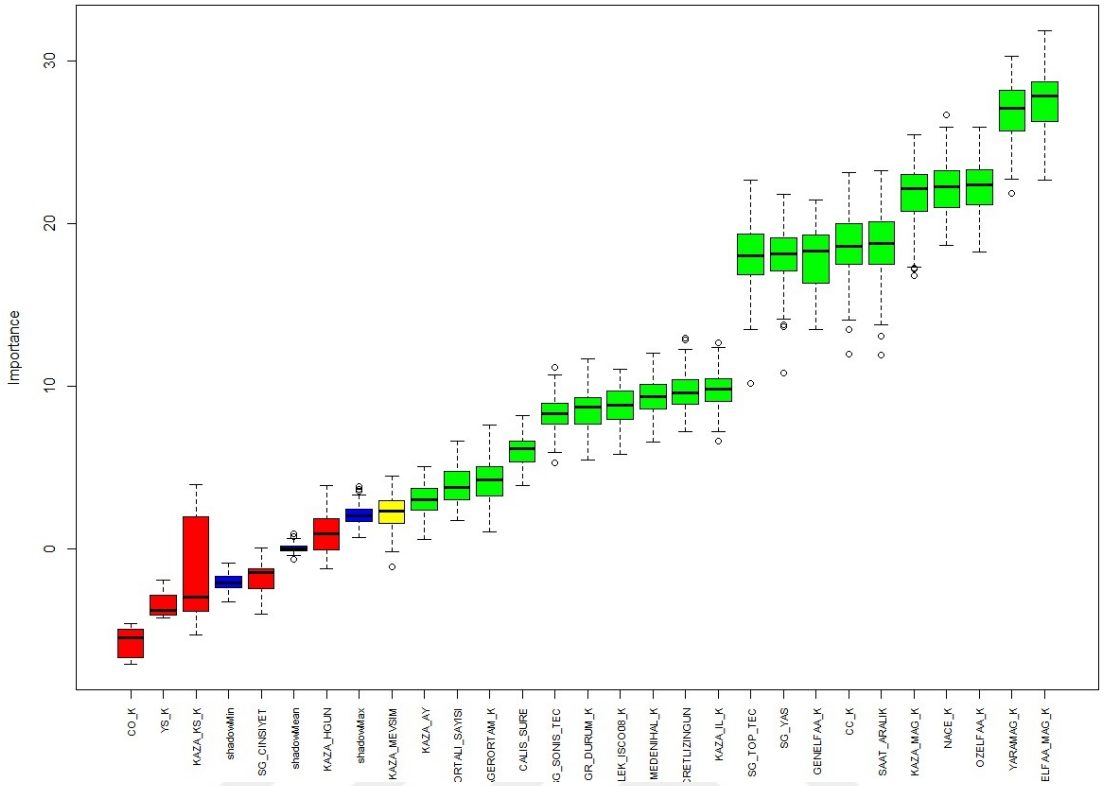
Tablo 5.7 Materyal/ekipman değişkenlerinin aynı seçilme yüzdeleri

	Veri Sayısı	Yüzde
KAZA_MAG = OZELFAA_MAG	137819	73.34
OZELFAA_MAG = YARAMAG	150959	80.33
KAZA_MAG = YARAMAG	149917	79.78
KAZA_MAG = OZELFAA_MAG = YARAMAG	129042	68.67

5.1.2 Boyut İndirgeme

Analizin bu bölümünde hazırlanan verideki değişkenler sırasıyla tahmin gücü (boruta) analizi, yapay değişken oluşturma, sıfır varyans testi ve çoklu korelasyon analizine tabi tutulmuştur. Öznitelik seçiminde değişkenlerin belirli bir tahmin edici değişkene olan yararlarına göre alt bölümlerini ölçen bir sarma algoritmasını içeren R kütüphanesinde bulunan Boruta öznitelik seçim fonksiyonu ile önemli değişkenler seçilmiştir [88]. Sonuçlara göre iş kazalarının sonucunu önemli ölçüde etkilemeyen değişkenler veri setinden çıkartılarak veri boyutu azaltılmıştır. İş kazalarının oluşumunda etkili faktörler şiddet gruplarındaki farklı amaç değişkenlere göre değişebileceği için öznitelik seçimi analizi, grup oluşturmadan önce, amaç değişken olarak ISGUNUKAYBI değişkeni kullanılarak gerçekleştirilmiştir.

187.910 kaza boruta yöntemi kullanılarak incelenmiş ve analiz sonucunda çalışma ortamının sabit veya sürekli oluşunu belirten CO, Kazaya sebep olan olayı belirten KAZA_KS_K, yaralanmaya sebep olan olayı belirten YS_K, kazalının cinsiyetini belirten SG_CINSIYET_K ve kazanın haftanın hangi gününde olduğunu belirten KAZA_HGUN_K değişkenlerinin, işgünü kaybını gün olarak gösteren SG_ISGUNUKAYBI bağımlı değişken üzerinde etkili olmadığı belirlenmiştir. KAZA_MEVSIM_K değişkenin amaç değişken üzerindeki etkisi belirsiz olarak bulunmuştur ancak bilgi kaybı oluşmaması açısından veri setinden çıkarılmamıştır. Şekil 5.2 analiz edilen verideki değişkenlerin, amaç değişken üzerindeki etkilerini önem sırasına göre göstermektedir. Bağımlı değişken sayısı bu analiz sonucu 26'dan 21'e düşmüştür. Kalan 21 bağımlı değişkenin kategorik özellikte olanların tüm kategorileri 1 ve 0 değerlerine sahip yapay değişkenler olarak tanımlanmış ve 15 kategorik değişken 372 yapay değişkene dönüştürülmüştür. Bazı alt kategorilerin varyasının analizi için yapay değişkenler sıfır varyans testine tabi tutulmuş ve 10 yapay değişkenin varyansının sıfır olduğu bulunmuştur (Tablo 5.8).



Şekil 5.2 Değişkenlerin tahmin gücü

Tablo 5.8 Analiz sonuçlarına göre veriden silinen bağımlı değişkenler

Grup	Boruta Analizi	Yapay Değişken Sıfır Varyans Analizi
İşyeri		
Çalışma Ortamı	CO	CC_K.999
Olaylar Zinciri	KAZA_KS YS	OZELFAA_K.999 OZELFAA_MAG_K.999 GENELFAA_K.29 GENELFAA_K.30 GENELFAA_K.31 GENELFAA_K.32 GENELFAA_K.999 YARAMAG K.999
Kazazede	SG_CINSIYET	
Kaza	KAZA_HGUN	

Son olarak gerçekleştirilen çoklu korelasyon analizinde sadece medeni hallerden %95 üzeri korelasyon gösteren evli ve bekar olma durumu ele alınmıştır. Sıfır varyansa

sahip yapay değişkenler ve medeni hal değişkeninin biri veriden çıkarıldığından toplam 361'i yapay değişken olmak üzere toplam 368 değişken ile makine öğrenmesi analizi aşamasına geçilmiştir (Tablo 5.9).

Tablo 5.9 Bağımlı değişken sayılarındaki değişimler

	Numerik	Kategorik	Toplam
Başlangıç	7	20	27
Boruta Sonrası	7	15	22
Yapay Değişken Sonrası	7	372	379
Sıfır Varyans Analizi Sonrası	7	362	369
Çoklu Korelasyon Analizi Sonrası	7	361	368

Veri Temizleme ve Boyut İndirgeme işlemleri, önerilen yöntemin ilk aşaması olan veri Ön işleme aşamasını oluşturmaktadır ve sözde kodu Algoritma 1'de verilmektedir.

Algorithm 1 Veri ön işleme

Input: Ulusal iş kazaları ham veri seti

Output: Ulusal iş kazaları ön işlenmiş veri seti

Güncel veriyi SGK veri tabanından çek
Yeni değişkenleri oluştur
if Hatalı veri var ise **then**
    Hatalı verileri NA olarak değiştir
end if
if eksik veri var ise **then**
    if Eksik veri kategorik ise **then**
        Kod ve Açıklama sutunlarını eşleştir
        Kalan eksik veriyi Diğer olarak tanımla
    elseEksik veri numerik ise
        Zincirli Denklemlerle çok değişkenli doldurma algoritması uygula
    end if
end if
if Eksik veri kazada direk etkili herhangi bir değişkende ise **then**
    Satır sil
end if
if Eksik veri kullanılan ekipmanı belirten değişkenlerin hepsinde ise **then**
    Satır sil
end if
Önemli bağımlı değişkenleri belirle
if İlgili bulunmayan değişken var **then**
    Değişken sil
end if
Sıfır varyans analizi yap
if Varyansı sıfır olan değişken var ise **then**
    Değişken sil
end if
Kategorik değişkenleri yapay değişkene dönüştür
Çoklu Korelasyon Analizi yap
if Herhangi iki değişkenin arasındaki korelasyon %95'den fazla ise **then**
    Önem ağırlığı daha az olan değişkeni sil
end if

5.2 Tahminleme

Analizin bu bölümünde, Algoritma 1'de verilen işlemlerin sonucu ortaya çıkan veri, şiddet gruplarına ayrılmıştır. Her bir şiddet grubunun veri seti, kaza tarihine göre sıralanarak %80'i eğitim için ve %20'si test için kullanılmak üzere eğitim ve test veri setlerine ayrılmıştır. Böylelikle, önerilen yöntem geçmiş kaza verisi üzerinden gelecekte oluşabilecek kazaların tahminlenmesine uyumlu hala getirilmiştir.

5.2.1 Şiddet Gruplarının Oluşturulması

Elde edilen veri seti SG_ISGUNUKAYBI bağımlı değişkenindeki kayıp gün sayısı kullanılarak 4 farklı 2'li şiddet sınıfına ayrılmıştır. Şiddet grubu numaralandırılması, açıklamaları ve eğitim/test veri seti sayıları Tablo 5.10'de verilmektedir. Makine öğrenmesi literatüründe ikili sınıflandırmada sınıflardan biri beklenen/araştırılan unsur olarak kabul edilir ve "pozitif" sınıf olarak adlandırılır. Kaza şiddeti sınıflandırmalarında kullanılan pozitif ve negatif sınıflar Tablo 5.10'de yer alan Şiddet Sınıfları kolonunda parantez içerisinde verilmiştir.

Tablo 5.10 Şiddet gruplamasına göre kaza sayıları

Şiddet Grubu	Şiddet Sınıflandırma	İş Günü Kaybı	Kaza Sayıları	Eğitim Seti (%80)	Test Seti (%20)
S1	Raporsuz (N)	<3	173.846	138.729	35.117
	Raporlu (P)	>3	14.064	11.599	2.465
S2	Minor (N)	<15	179.101	143.019	36.082
	Major (P)	>15	8.809	7.309	1.500
S3	KayıpYok (N)	=0	59.358	45.371	13.987
	KayıpVar (P)	>0	128.552	104.957	23.595
S4	OlumYok (N)	<7.500	187.276	149.800	37.476
	OlumVar (P)	>7.500	634	528	106

5.2.2 Makine Öğrenmesi ile Model Oluşturma ve Sınıf Tahminleme

Bir kazanın meydana gelmesinde etkili olabilecek bir çok faktör vardır. Bu çalışmada kullanılan veri, bu faktörlerin tümünü barındırmadığından, bu veride bulunan kaza zincirlerinin de her zaman aynı şiddette bir kaza ile sonuçlanmayabileceği kabul edilmelidir. Aynı kaza zinciri, veride bulunmayan bazı faktörlerin etkisiyle, bazen ramak kala durumlar oluşturabilecek iken bazen daha yüksek veya daha düşük şiddetle sonuçlanan kazalar ortaya çıkarabilir. Bu farklı şiddetteki kaza sonuçları makine öğrenmesi algoritmalarının öğrenme sürecini etkileyeceğinden, kazaların sonuçlarını yüksek doğruluk oranında tahminlemesi beklenmemektedir. Ancak şans faktöründen daha iyi bir sonuç ortaya koyabilmek bile kazaların belli bir örüntü

izlediğini ve bu örüntüdeki faktörlerin ne kadar çok analize katılabilinirse o kadar yüksek doğrulukta sonuçlar elde edilebileceğini gösterebilir.

Makine öğrenmesi sınıflandırmalarında bazı durumlarda doğru pozitif sınıf tahmini doğru negatif sınıf tahmininden daha fazla önem taşımaktadır. Örneğin kanser hücrelerinin tahmini analizinde kanser hücresini sağlıklı hücre olarak sınıflandırmak (Yanlış pozitif), sağlıklı bir hücreyi kanser hücresi olarak tahminlemekten çok daha risklidir. Bu gibi durumların makine öğrenmesinin öğrenme aşamasında ele alınması, tahminleme aşamasında daha fazla doğru pozitif elde edilmesine imkan sağlar. Elbette daha fazla doğru pozitif aynı zamanda daha fazla yanlış negatif demek de olacaktır. İş kazaları açısından bakıldığında durum kanser hücresi tahmininden çok farklı değildir.

İş kazalarının tahminlenmesindeki ana amaç yüksek şiddette oluşan kaza örüntülerindeki etkenleri ortaya çıkarmak ve bu etkenler ile ilgili önlemler alarak aynı örüntülerin tekrarlanmasını engellemek olduğundan, yüksek şiddetle sonuçlanan (pozitif sınıf) kazaların doğru tahmin edilmesi, düşük şiddetle sonuçlanan kazaların yanlış tahmin edilmesinden çok daha kritiktir. Örneğin, gerçekte ölümlü bir iş kazası olarak ortaya çıkmış kaza zincirlerini yüksek yüzeyde doğru tahmin eden bir algoritmanın, ölümlü olmayan bazı kazaları da ölümlü gibi tahmin etmiş olması nihai amaca engel oluşturmamaktadır. Bu sebeple sınıf tahmini performansının ölçülmesinde literatürdeki farklı metrikler incelenmiş ve yüksek şiddet oluşturan kazaların diğer gruptan ayrılmasında duyarlılık ve kesinlik metriklerine ve bu metriklerden çıkartılan metriklere odaklanılmıştır.

Makine öğrenmesi yöntemiyle tahmin performansı analizlerinde sıklıkla 4'ü temel 6 metrik kullanılmaktadır. Bu metriklerin iş kazalarının tahminlenmesi analizinde kullanılma şekli karışıklık matrisi üzerinden Tablo 5.11'de tanımlanmıştır.

Makine Öğrenmesi, gerçekteki sınıfların etiketlerinin bulunduğu eğitim verileri üzerinden çıkarılan modelin, test verilerinin sınıf etiketi tahmini için kullanılmasına dayanır. Bu çalışmada, tahmin için kullanılacak modelin oluşturulmasında Stokastik Gradient Boosting (SGB) algoritması kullanılmıştır. Analizler R programlama yazılımında bulunan Caret kütüphanesi kullanılarak gerçekleştirilmiştir [106, 107].

Veri ön işleme ile elde edilen veri kaza tarihine göre sıralanmış ve en yeni tarihli %20 kaza, kalan %80 kaza verisi kullanılarak tahminlenmiştir. Şiddet gruplarına ait eğitim ve test veri seti sayıları Tablo 5.10'de verilmektedir.

Her bir şiddet grubuna ait eğitim seti, 5 kat 10 tekrar CV konfigürasyonu ile eğitim ve doğrulama setlerine bölünmüştür. İkili sınıflandırmada kullanılan ROC metriğine göre optimum modeller seçilmiştir. Bu modeller test verilerinin sınıflarını

Tablo 5.11 Kaza karışıklık matrisi

	Gerçek Pozitif	**Gerçek Negatif**	
Pozitif Tahmin	**DP:** pozitif sınıfta yer alan kazaların pozitif sınıf olarak tahminlenmesi	**YP:**negatif sınıfta yer alan kazaların pozitif sınıf olarak tahminlenmesi	**Kesinlik:** pozitif olarak tahminlenenen kazaların gerçekte pozitif sınıfta olma oranı
Negatif Tahmin	**YN:** pozitif sınıfta yer alan kazaların negatif sınıf olarak tahminlenmesi	**DN:** negatif sınıfta yer alan kazaların negatif sınıf olarak tahminlenmesi	
	Duyarlılık/DPO: pozitif sınıfta yer alan kazaların pozitif sınıf olarak tahminlenme oranı	**Özgüllük/YPO:** negatif sınıfta yer alan kazaların negatif sınıf olarak tahminlenme oranı	

tahmin etmek için kullanılmıştır. Her bir şiddet grubununun farklı dengesiz veri ele alım yaklaşımlarına göre yapılmış analizlerinin tahmin performans ölçümleri Tablo 5.12'de verilmektedir. Dengesiz veri ele alım yaklaşımlarından alt örnekleme, smote örnekleme ile eşik değer belirleme yöntemlerinin performanslarında istatistiksel olarak belirgin farklar bulunmuştur (t-test p-değeri < 0.05). Ancak, Brier ve LogLoss metrikleri incelendiğinde, örnekleme yapmadan sadece eşik değer ayarı ile ele alınan yöntemin performansı hata oranları açısından üstünlük oluşturmaktadır. Bu sebeple, sınıflandırma tahmininde etkili önemli değişken seçimi için kullanılmak üzere örneklemin yapılmadığı, eşik değerin kullanıldığı temel model seçilmiştir.

Tablo 5.12 Şiddet grupları tahmin performansları

SG	DengelemeYöntemi	DengDoğ	AUC	Duyarlılık	Kesinlik	F1	F2	Brier	LogLoss	DP	DN
SG1	Eşik_Değer	0,6211	0,6646	0,6081	0,1045	0,1783	0,3096	**0,0600**	**0,2326**	1499	22269
	Alt_Örnekleme	0,6214	0,6633	0,5513	0,1115	0,1855	0,3082	0,2333	0,6574	1359	24286
	Smote	0,5796	0,6060	0,7063	0,0831	0,1487	0,2825	0,1240	0,4167	1741	15907
SG2	Eşik_Değer	0,6411	0,696	0,636	0,0695	0,1254	0,2419	**0,0375**	**0,1591**	954	23319
	Alt_Örnekleme	0,6469	0,6992	0,5993	0,0754	0,1339	0,2508	0,2246	0,6399	899	25058
	Smote	0,8307	0,6199	0,4753	0,063	0,1113	0,2059	0,1032	0,3419	713	25480
SG3	Eşik_Değer	0,6192	0,6615	0,6635	0,7247	0,6928	0,6749	0,2186	0,6274	15656	8040
SG4	Eşik_Değer	0,7780	0,8362	0,783	0,0097	0,0191	0,046	**0,0028**	**0,0171**	83	28967
	Alt_Örnekleme	0,7868	0,8458	0,8585	0,0085	0,0167	0,0407	0,1656	0,5089	91	26800
	Smote	0,7682	0,8451	0,8302	0,0079	0,0157	0,0382	0,0833	0,2827	88	26468

5.2.3 Önemli Değişkenlerin Belirlenmesi

Tez çalışmasının ana amaçlarından biri şiddet sonucunun tahmininde güçlü etkiye sahip değişkenlerin bulunması olduğundan analizin bu aşamasında her bir ikili sınıflandırmada kullanılan modelin önemli bulduğu değişkenler çıkarılmıştır.

Modellerdeki her bir tahminleyici için bir dizi eşik değerin Duyarlılık ve Özgüllük hesaplamalarıyla oluşturulan ROC eğrisi oluşturulmuş ve eğri altında kalan değer değişkenin tahminleyici gücü/önemi olarak kullanılmıştır. Aynı modeldeki tüm değişkenler modeller arası kıyaslama yapılabilmesi amacıyla yüz üzerinden ölçeklendirilmiştir.

Analizde kullanılan değişkenlerin ikili sınıflandırma modellerindeki önemi (tahmin gücü) yüzde olarak Tablo 5.13'de verilmektedir. Herhangi bir değerin bulunmaması değişkenin o modelde etkisi olmadığı anlamına gelmektedir. Bazı değişkenler sadece tek bir modelde önemli bulunurken, bazı değişkenler tüm ikili sınıflandırmalarda önemli bulunmuştur. Tabloda en çok dikkati çeken değişken SIGORTALI_SAYISI değişkenidir. Ölümlü kazaların tahminlenmesinde çok yüksek bir önemi bulunmaz iken diğer grupların tahminlenmesinde oldukça etkili olduğu görülmektedir. Kaza sonucu iş günü kaybının 3 günden fazla olduğu iş kazalarının tahminlenmesinde 18 değişken önem ağırlığının %70'den fazlasını oluştururken iş günü kaybının 15 günden fazla olup olmamasının tahminlenmesinde bu sayı 22'ye çıkmaktadır. SIGORTALI_SAYISI ve SG_YAS değişkenleri, iş günü kaybının 3 günden fazla olmasının tahminlenmesinde ve 15 günden fazla olmasının tahminlenmesinde en önemli iki değişken olarak bulunmuştur. Bu iki grubun önemli değişkenleri incelendiğinde, değişkenlerin önem ağırlığı değişse bile bir kaç farklılık ile hemen hemen aynı değişkenlerin önemli bulunduğu gözlenmektedir. İş günü kaybı olan ve olmayan kazaların belirlenmesinde ise sadece 11 değişken önem ağırlığının %70'den fazlasını oluşturmaktadır (SIGORTALI_SAYISI, CC_K.1, NACE_K.52, NACE_K.24, KAZA_IL_K.27, KAZA_IL_K.41, KAZA_IL_K.45, CC_K.5, CC_K.6, NACE_K.25, NACE_K.27). Gene bu grupta da SIGORTALI_SAYISI en önemli değişken olarak bulunmuştur.

Ölümlü kazaların tahminlenmesinde SIGORTALI_SAYISI diğer ikili sınıflandırmalardan farklı olarak sadece %1.28 ağırlığında öneme sahip bulunmuştur. Ölümlü kazaların tahminlenmesinde önem ağırlığının %70'den fazlasını 12 değişken oluşturmaktadır: KAZA_MAG_K.2, KAZA_IL_K.73, SG_TOP_TEC, KAZA_MAG_K.12, SG_YAS, SG_SONIS_TEC, CC_K.1, OZELFAA_K.3, SG_KAZAGERORTAM_K.2, KAZA_IL_K.13, OZELFAA_MAG_K.2, CC_K.2.

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
CC_K.1	0,20		11,03	5,37	16,61
CC_K.10	0,64		0,19		0,83
CC_K.2	4,51	4,09	0,33	2,80	11,74
CC_K.3	0,14	0,73	0,04	2,37	3,27
CC_K.5	0,19		3,36		3,56
CC_K.6			2,13		2,13
CC_K.9		0,17			0,17
SG_KAZAGERORTAM_K.1				0,64	0,64
SG_KAZAGERORTAM_K.2	0,82	2,29		3,79	6,90
SG_KAZAGERORTAM_K.999	1,20	1,27		0,65	3,12
GENELFAA_K.1	0,34	0,58	0,11	0,38	1,42
GENELFAA_K.12		0,13			0,13
GENELFAA_K.13		0,43			0,43
GENELFAA_K.14		0,30			0,30
GENELFAA_K.16	0,09		0,20		0,29
GENELFAA_K.25	0,67	1,03		2,44	4,14
GENELFAA_K.3		0,13			0,13
GENELFAA_K.4	0,22	0,27		1,74	2,23
GENELFAA_K.5		0,11			0,11
GENELFAA_K.6		0,10	0,11	0,30	0,51
GENELFAA_K.7	0,20	0,10			0,30

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
GENELFAA_K.8	0,09			0,34	0,43
GENELFAA_K.9	0,29	0,39			0,68
KAZA_MAG_K.1		0,11	0,05		0,16
KAZA_MAG_K.10	5,29	3,50	0,21		9,00
KAZA_MAG_K.11		0,09			0,09
KAZA_MAG_K.12	0,45	0,84		7,22	8,50
KAZA_MAG_K.14	1,56	1,41			2,97
KAZA_MAG_K.15	0,10				0,10
KAZA_MAG_K.17	0,08				0,08
KAZA_MAG_K.19	0,37	0,20			0,58
KAZA_MAG_K.2	0,64	1,30		11,29	13,23
KAZA_MAG_K.3		0,09		0,37	0,46
KAZA_MAG_K.6	2,70	1,80	1,46		5,95
KAZA_MAG_K.9		0,27		0,23	0,51
OZELFAA_K.1	2,93	1,76			4,68
OZELFAA_K.2		0,19	0,15		0,34
OZELFAA_K.3				4,47	4,47
OZELFAA_K.4		0,11	0,22		0,33
OZELFAA_K.5		0,20			0,20
OZELFAA_K.6	0,07				0,07
OZELFAA_K.7				0,23	0,23

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
OZELFAA_MAG_K.1		0,09	0,54		0,63
OZELFAA_MAG_K.10		0,21	1,68		1,88
OZELFAA_MAG_K.11			0,05		0,05
OZELFAA_MAG_K.12		0,23			0,23
OZELFAA_MAG_K.14	1,92	0,70	0,17		2,78
OZELFAA_MAG_K.17			0,09		0,09
OZELFAA_MAG_K.2	0,12	0,40		2,90	3,42
OZELFAA_MAG_K.20			0,06		0,06
OZELFAA_MAG_K.5		0,11			0,11
OZELFAA_MAG_K.6			0,06		0,06
OZELFAA_MAG_K.7		0,12			0,12
OZELFAA_MAG_K.8		0,12			0,12
OZELFAA_MAG_K.9	0,94	0,59			1,53
YARAMAG_K.10	1,40	2,96	0,34		4,70
YARAMAG_K.11		0,10			0,10
YARAMAG_K.12	0,10	0,42		2,31	2,84
YARAMAG_K.13				0,22	0,22
YARAMAG_K.14	2,09	1,49			3,58
YARAMAG_K.15	0,20				0,20
YARAMAG_K.2	0,40	0,81		0,45	1,65
YARAMAG_K.6	1,57	2,06	0,42		4,05

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
YARAMAG_K.7	0,21	0,21	0,07		0,49
KAZA_IL_K.1		0,15			0,15
KAZA_IL_K.10	0,10	0,09			0,19
KAZA_IL_K.11	0,18		0,24		0,41
KAZA_IL_K.12		0,10		0,69	0,79
KAZA_IL_K.13	0,31	0,33		3,20	3,83
KAZA_IL_K.14		0,20	0,04		0,24
KAZA_IL_K.15				0,87	0,87
KAZA_IL_K.16	0,68	0,84	0,78		2,31
KAZA_IL_K.17			0,10		0,10
KAZA_IL_K.19		0,09			0,09
KAZA_IL_K.2				0,78	0,78
KAZA_IL_K.20	0,56		0,22		0,78
KAZA_IL_K.21			0,34		0,34
KAZA_IL_K.26	0,29	0,11	1,41		1,81
KAZA_IL_K.27		0,11	4,10		4,21
KAZA_IL_K.28	0,52				0,52
KAZA_IL_K.3	0,12	0,31		0,40	0,83
KAZA_IL_K.30	0,45	0,96			1,41
KAZA_IL_K.31	0,83	0,75			1,58
KAZA_IL_K.32	0,19		0,09	0,29	0,57

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
KAZA_IL_K.33		0,32			0,32
KAZA_IL_K.34	0,34	0,95	0,84		2,13
KAZA_IL_K.35	0,59	1,63	0,26		2,48
KAZA_IL_K.36	0,18	0,26		0,32	0,76
KAZA_IL_K.37	0,74	0,73		1,19	2,65
KAZA_IL_K.38	0,43				0,43
KAZA_IL_K.4	0,14	0,30		1,43	1,88
KAZA_IL_K.41			3,81		3,81
KAZA_IL_K.42	2,57	5,02	1,20		8,78
KAZA_IL_K.44		0,10			0,10
KAZA_IL_K.45	0,11		3,72		3,82
KAZA_IL_K.46	0,80	1,51	0,15		2,46
KAZA_IL_K.47		0,30		1,17	1,47
KAZA_IL_K.48	0,34	0,47			0,80
KAZA_IL_K.49	0,18				0,18
KAZA_IL_K.52	0,42	0,17	0,04		0,64
KAZA_IL_K.54	0,41		0,06		0,47
KAZA_IL_K.55	0,26	0,09			0,35
KAZA_IL_K.57	0,29	0,47		0,57	1,33
KAZA_IL_K.58			0,12	0,62	0,74
KAZA_IL_K.59		0,09			0,09

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
KAZA_IL_K.6	0,41	0,13	0,41		0,95
KAZA_IL_K.60				0,38	0,38
KAZA_IL_K.61	0,55	0,50			1,05
KAZA_IL_K.62				0,21	0,21
KAZA_IL_K.63		0,21			0,21
KAZA_IL_K.64	0,64	0,77			1,41
KAZA_IL_K.65		0,09		1,36	1,46
KAZA_IL_K.67	1,10				1,10
KAZA_IL_K.7		0,43			0,43
KAZA_IL_K.70	0,40	0,07			0,47
KAZA_IL_K.72	0,29	0,22			0,51
KAZA_IL_K.73		0,31		9,52	9,83
KAZA_IL_K.74			0,54		0,54
KAZA_IL_K.75	0,22	0,11			0,33
KAZA_IL_K.77			0,28		0,28
KAZA_IL_K.78	0,15		0,20		0,35
KAZA_IL_K.8			0,11		0,11
KAZA_IL_K.80	0,94	0,17			1,12
KAZA_IL_K.81			0,11		0,11
KAZA_IL_K.9		0,58	0,11		0,69
NACE_K.1			0,45		0,45

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
NACE_K.10		0,13	0,47		0,60
NACE_K.12			0,48		0,48
NACE_K.13	1,22	0,52	0,76		2,50
NACE_K.14			0,26		0,26
NACE_K.16	0,58	0,23			0,81
NACE_K.19	0,26	0,21			0,47
NACE_K.2	0,11				0,11
NACE_K.22			0,16		0,16
NACE_K.24			4,84		4,84
NACE_K.25	0,17	0,17	2,12		2,46
NACE_K.26			1,74		1,74
NACE_K.27			1,80		1,80
NACE_K.28	0,35	0,31	0,48		1,15
NACE_K.29	0,09				0,09
NACE_K.35		0,20			0,20
NACE_K.37	0,17	0,25			0,41
NACE_K.38	0,10				0,10
NACE_K.41	0,86	1,04		1,88	3,78
NACE_K.42		0,35	0,16		0,51
NACE_K.46	0,62	0,22			0,84
NACE_K.49				2,02	2,02

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
NACE_K.5			0,40		0,40
NACE_K.52	1,04	0,28	10,30		11,62
NACE_K.56	0,45	0,24	0,58		1,27
NACE_K.71	0,25	0,51			0,76
NACE_K.79		0,10			0,10
NACE_K.8				0,41	0,41
NACE_K.81		0,15			0,15
NACE_K.82			0,05		0,05
NACE_K.85	0,11	0,14	0,09		0,34
NACE_K.86			0,11		0,11
NACE_K.95			0,21		0,21
NACE_K.99			0,17		0,17
KAZA_MEVSIM.1	0,18	0,15			0,33
KAZA_MEVSIM.2		0,15			0,15
KAZA_MEVSIM.3		0,15			0,15
KAZA_MEVSIM.4	0,14	0,07	0,10		0,32
KAZA_AY.2	0,09				0,09
KAZA_AY.3	0,07				0,07
KAZA_AY.7		0,40			0,40
KAZA_AY.8			0,04		0,04
SAAT_ARALIK.12	0,15		0,32		0,47

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
SAAT_ARALIK.2		0,16	0,05		0,21
SAAT_ARALIK.5	0,42	0,44			0,85
SAAT_ARALIK.8		0,12			0,12
SAAT_ARALIK.9		0,11			0,11
MEDENIHAL_K.1			0,33		0,33
MEDENIHAL_K.999		0,39	0,05		0,43
OGR_DURUM_K.2	0,17	0,33			0,51
OGR_DURUM_K.3		0,22			0,22
OGR_DURUM_K.4	0,68	0,24			0,92
OGR_DURUM_K.6			0,18		0,18
SG_MESLEK_ISCO08_K.12				0,25	0,25
SG_MESLEK_ISCO08_K.14	0,18		0,04		0,22
SG_MESLEK_ISCO08_K.15			0,47		0,47
SG_MESLEK_ISCO08_K.19			0,05		0,05
SG_MESLEK_ISCO08_K.23			0,59		0,59
SG_MESLEK_ISCO08_K.28			0,22		0,22
SG_MESLEK_ISCO08_K.30	0,23	0,08	0,17		0,48
SG_MESLEK_ISCO08_K.31	0,39	0,12			0,50
SG_MESLEK_ISCO08_K.33	0,07	0,28			0,35
SG_MESLEK_ISCO08_K.35			0,20		0,20
SG_MESLEK_ISCO08_K.36	0,09	0,10			0,20

Tablo 5.13 Şiddet gruplarının tahminlenmesinde önemli değişkenler - (devamı)

Satır Etiketleri	SG1 Raporlu	SG2 Majör	SG3 KayıpVar	SG4 ÖlümVar	Ortalama
SG_MESLEK_ISCO08_K.39		0,18	0,38		0,56
SG_MESLEK_ISCO08_K.4		0,41			0,41
SG_MESLEK_ISCO08_K.40			0,49		0,49
SG_MESLEK_ISCO08_K.42	0,31	0,14	0,48		0,93
SG_MESLEK_ISCO08_K.6	0,07				0,07
SG_MESLEK_ISCO08_K.8			0,06		0,06
SG_MESLEK_ISCO08_K.9			0,38		0,38
CALIS_SURE	0,74	1,71			2,45
SG_SONIS_TEC	2,43	4,02	1,04	6,23	13,72
SG_TOP_TEC	3,11	3,56	1,39	8,15	16,22
SG_UCRETLIZINGUN	0,68	0,99	0,68		2,35
SG_YAS	10,91	10,76	0,18	6,28	28,13
SIGORTALI_SAYISI	23,02	15,03	24,13	1,28	63,46

Analizin ikinci bölümünü oluşturan şiddet gruplarını belirleme, model oluşturma, sınıf tahminleme ve önemli değişkenlerin seçimi adımlarının sözde kodu Algoritma 2'de verilmektedir.

Algorithm 2 Tahminleyici değişkenlerin belirlenmesi

Input: Ulusal iş kazaları ön işlenmiş veri seti

Output: Her bir şiddet grubu için tahminleme performans değerleri ve tahminleyici değişkenlerden oluşan veri seti

1: İşgünü kaybı değişkenini kullanarak ikili şiddet grupları oluştur
2: **for** Bütün şiddet grupları için **do**
3: Veriyi kaza tarihine göre sırala
4: Verinin ilk %80'ini eğitim seti, son %20'sini test seti olarak belirle.
5: Eğitim setini kullanarak SGB algoritması ile model oluştur
6: Modeli kullanarak test setinde bulunan kazaların sınıflarını olasılık değeri olarak tahminle
7: Sınıf ayrışmasında olasılık eşik değerini youden indeksine göre belirle
8: Belirlenen eşik değere göre sınıf etiketlerini oluştur
9: Modelin tahmin gücünü gösteren performans değerlerini oluştur
10: Modelin güçlü bulduğu tahminleyici değişkenlerden oluşan ve tüm kazaları içeren veri setini oluştur.
11: **end for**

5.3 Kural Oluşturma

İkili şiddet gruplarına ait tahminleyici gücü yüksek değişkenler Algoritma 2'de verilen adımlar ile belirlendikten sonra bu değişkenlerin oluşturduğu kaza zincirleri apriori algoritması kullanılarak bulunabilir hale gelmektedir. Kural oluşturmada iki farklı yaklaşım benimsenmiştir;

1. Kazaların oluşumunda dolaylı etkisi bulunan değişkenleri içeren kaza zincirlerinin ortaya çıkarılması.

2. Şiddet ve yaralanma sonucu içeren kaza zincirlerinin ortaya çıkarılması.

Kaza zincirlerinin ortaya çıkarılması esnasında farklı eşik değerleri denenmiş ve incelenmesi zor olmayacak kadar çok kural çıkaracak yükseklikte, ancak ilginç kuralları da elemeyecek kadar düşüklükte eşik değerler belirlenmiştir.

5.3.1 Dolaylı Değişken Etkilerini İçeren Kaza Zincirlerinin Çıkarılması

Kazaların oluşumunda etkili faktörlerin bazıları direk yapılan iş ve ve ortaya çıkan risk ile direk ilgiliyken, bazı faktörler kaza oluşumnda dolaylı bir etki gösterebilmektedir. Örneğin çalışılan çevre, yapılan faaliyet, kullanılan ekipman ve görev ile ilgili olan değişkenler (CC_K, NACE_K, SG_KAZAGERORTAM_K, GENELFAA_K, OZELFAA_K, OZELFAA_MAG_K, KAZA_MAG_K, YARAMAG_K, SG_MESLEK_ISCO08_K) kaza oluşumunda direk etkili olabilecek iken, kazanın gerçekleştiği il, kazanın gerçekleştiği zaman ve kazalı ile ilgili değişkenler (KAZA_IL_K, OGR_DURUM_K, SAAT_ARALIK, KAZA_AY, KAZA_MEVSIM, MEDENIHAL_K, SIG_SAY, SG_SONIS_TEC, SG_TOP_TEC, CALIS_SURE, SG_YAS, SG_UCRETLIZINGUN) kazanın oluşumunda ancak dolaylı etkileri olabilecek değişkenlerdir. Bu dolaylı etkisi olabilecek değişkenlerin kaza zincirlerini nasıl ve ne şekilde etkilediğini ortaya koymak amacıyla bu aşamada dolaylı değişkenleri içeren kurallar oluşturulmuştur.

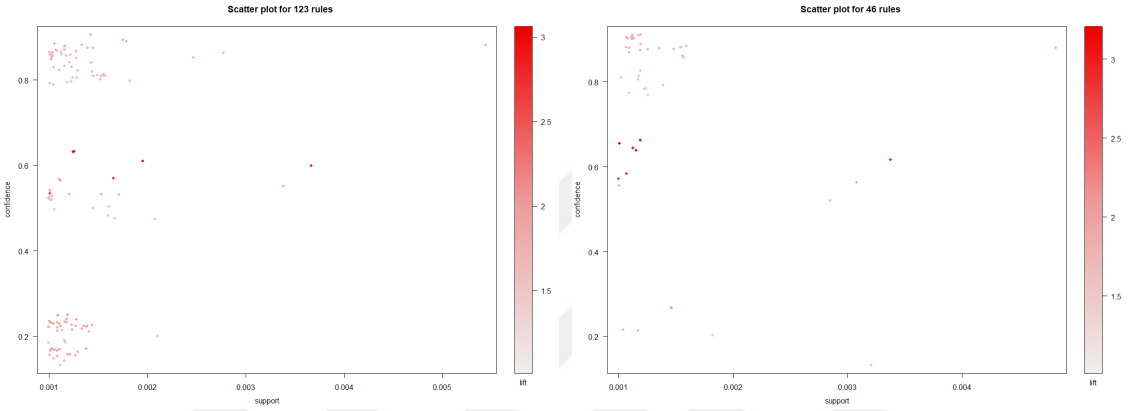
Dolaylı değişkenlerin kaza oluşumundaki etkisi, bu değişkenlerin kaza zincirlerinde bağımlı bir değişken olarak bulunup bulunmadığı üzerinden incelenmiştir. Birliktelik kuralları ölçütlerinden kaldıraç değeri, X ve Y'nin bağımsız olup olmadığını gösteren oransal bir değerdir. Bu sebeple dolaylı etkili değişkenlerin etkisi incelenirken kaldıraç değeri kullanılmıştır. Direk etkili değişkenlerden oluşan bir kaza zincirinin olasılığı, herhangi bir dolaylı değişkenin eklenmesi durumundaki olasılığından farklı ise, bu durum dolaylı değişkenin kaza zincirinde etkili olduğunu göstermektedir. Kaldıraç değerinin 1'den büyük olması, dolaylı değişkenin direk değişkenlerden oluşan kaza zincirinin oluşumunun olasılığını arttırdığını gösterirken 1'den küçük olması ise kaza zincirinin oluşumunun olasılığını azalttığını göstermektedir. Analizin bu aşamasında seçilen eşik değerler şu şekildedir;

> Ölümlü kazalar için (Şiddet Grubu 4) destek değeri en az 0.0001, diğer şiddet grupları için en az 0.001
>
> Tüm şiddet grupları için güven değeri en az 0.1

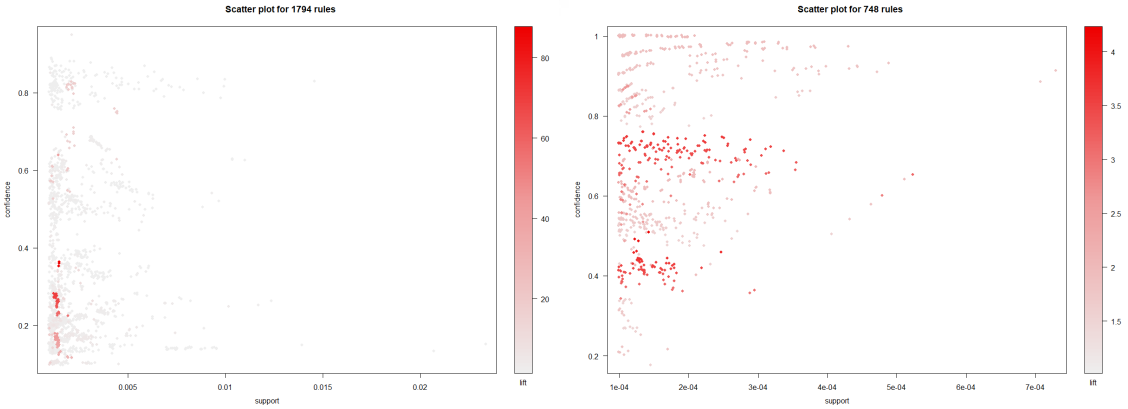
Bu ayarlar sonucu ortaya çıkan kural sayıları ve kaldıraç değeri en az 1.5 (K.d.>15) olarak seçildiğinde ortaya çıkan kural sayıları ve bu elemenin yüzdesi Tablo 5.14'te verilmektedir. Her bir şiddet grubunda ortaya çıkan kuralların kaldıraç değerlerinin dağılım grafikleri Şekil 5.3'de verilmektedir. Kaza zincirinin kaldıraç değeri arttıkça şekilde gözüken noktaların rengi koyulaşmaktadır.

Tablo 5.14 Dolaylı değişkenli kural sayıları

Şiddet Grup	Toplam Kural	K.d. > 1.5	Yüzde
1-Raporlu	5910	123	2%
2-Major	1792	46	2.5%
3-KayipVar	35959	1794	5%
4-OlumVar	1713	748	44%
Toplam		2711	



(a) Raporlu **(b)** Major



(c) KayipVar **(d)** OlumVar

Şekil 5.3 Dolaylı etkili faktörlerin etkilediği kazaların kaldıraç değeri dağılımları

Analiz sonucunda dolaylı değişkenlerin seçilen eşik değerler doğrultusunda etkili olduğu toplam 2.711 kaza zinciri bulunmuştur. Analiz sonucunda bulunan 2.711 kuralın tamamı https://github.com/frecal/doktora_tezi web adresinde bulunmaktadır. Kurallarda yer alan kısaltmaların açıklamaları Ek-A'da verilmektedir. Dolaylı değişkenlerin tümünün her bir şiddet grubunda ortaya çıkan kaza zincirlerindeki ortalama kaldıraç değerleri Ek-B.1'deki tabloda verilmektedir.

Ek-B.1'deki tabloda görülmektedir ki seçilen eşik değerler doğrultusunda bazı dolaylı değişkenler her hangi bir kaza zinciri üzerinde yeterince etkili bulunmaz iken, bazı dolaylı değişkenlerin kaldıraç değeri ise 1'den küçük bulunmuştur. Bu durum dolaylı değişkenin kaza zincirinin oluşum olasılığı üzerinde azaltıcı etkisi olduğu anlamına gelmektedir.

5.3.1.1 Dolaylı Etkili Değişkenlerin Genel Olarak İncelenmesi

Ek B.1'de verilen tablo incelendiğinde aşağıdaki çıkarımlar yapılabilmektedir:

- Aylardan ağutos, eylül, ekim, kasım, aralık ve ocak aylarının hiç bir şiddet grupuna ait kaza zincirleri üzerine etkisi bulunmamıştır. Mart ayı ise tüm şiddet gruplarına ait kaza zincirlerinde etkisi bulunan tek ay olarak görülmektedir.
- Tüm şiddet gruplarına ait kaza zincirlerinde etkili tek şehir İstanbul olarak belirmiştir. Ayrıca ölümlü kaza zincirlerinde etkili de tek şehirdir. Ancak kaldıraç değeri 0.88 olarak bulunduğu için bu etkinin negatif olduğu yorumlanmaktadır.
- İzmir ise sadece ölümlü kaza zincirlerinde etkili bulunmayarak diğer üç grupta da etkili tek şehir olarak gözükmektedir.
- Raporlu ve majör şiddet grubunda, başka bir deyişle kayıp iş günü 3 ile 15 günden fazla iş günü kayıplı iş kazalarında etkili olarak İstanbul ve İzmir dışında sadece Bursa ve Kocaeli illeri bulunmaktadır.
- Kayıp iş günü ile sonuçlanan kaza zincirlerinde etkili toplam 15 şehir bulunmaktadır. Edirne, Osmaniye ve Karabük illeri kayıp günlü iş kazalarının bazılarında 78, 52, 16 kaldıraç değerleriyle çok yüksek etkili olarak bulunmuşlardır.
- Boşanmış veya dul olmanın hiç bir kaza zincirinde etkisi bulunmazken, evli olmanın bazı kaza zincirlerinin oluşum olasılığını arttırdığı görülmektedir.
- İlkokul ve ortaokul mezunu olmak bazı kazı zincirlerinin oluşum olasılığını arttırıken lise mezunu olmanın sadece kayıp günlü iş kazaları kuralları incelendiğinde pozitif etkisi olmasına rağmen, 3 günden ve 15 günden fazla iş günü kayıplı ve ölümlü kazaların oluşumunun olasılığını düşürdüğü gözlemlenmektedir.
- Saat aralıklarının etkisi incelendiğinde sabah 08:00-10:00 arası saat aralığının tüm şiddet gruplarında etkili olduğu ve bazı kaza zincirlerinin oluşum olasılığını arttırdığı gözlemlenmektedir. Yemek öncesi 10:00-12:00 arası saat aralığı

incelendiğinde etkili bulunduğu tüm ölümlü kaza zincirlerinde etkisinin olasılığı arttırıcı yönde olduğu bulunmuştur. Diğer şiddet gruplarında ise hem olasılığı arttırdığı hem de azalttığı kaza zincirleri olduğu görülmektedir. Yemek saati olarak kabul edilebilecek 12:00-14:00 saat aralığı ölümlü kazaların oluşumunda hiç etkili bulunmazken, yemekten sonraki 14:00-16:00 ve 16:00-18:00 saat aralıklarının etkili oldukları ölümlü kazaların hepsinde olasılığı arttırıcı yönde etkisi olduğu görülmektedir.

- Son iş tecrübesinin 5 yıldan fazla olmasının ölümlü kaza zincirlerinin üzerinde hiç etkisi bulunmazken, son iş tecrübesinin 0-2 ay arasında olması etkili olduğu tüm kaza zincirlerinin oluşum olasılığını arttırdığı görülmüştür. Toplam tecrübede ise ilk 1 yılın hiç bir kaza zincirinde etkisi bulunamamıştır ancak 1-5 yıl arasında olmasının etkili olduğu tüm raporlu, majör ve ölümlü kaza zincirlerinin üzerinde oluşumunu azaltıcı etkisi olduğu görülmüştür.
- Ücretli izin günleri son iş tecrübesinde geçen sürelerle ve çalışanın bu izinleri kullanıp kullanmamasıyla belirlenmekedir. İzin kullanamamış olmanın kaza zincirleri üzerinde etkisi olup olmadığı incelendiğinde son iş tecrübesi etkisi ile arasında belirgin bir fark gözlemlenememiştir.
- En etkili dolaylı faktörlerden bir diğeri ise 50 yaş üzeri olunması olarak görülmektedir. 50 yaş üzeri olmanın etkili olduğu tüm raporlu, majör ve ölümlü kaza zincirlerinin üzerinde oluşumunu sırasıyla ortalama 1,2 1,3 ve 3,4 kat arttırıcı etkisi olduğu görülmüştür.
- 13-18 yaş arası olmanın ise sadece kayıp iş günlü kazalarla ilgili zincirlerde etkisi bulunmuştur ancak bu etki 40 kat arttırıcı olarak 13-18 yaş aralığı faktörünü en etkili dolaylı faktörlerden biri haline getirmektedir.
- 18-25 yaş arası olmanın ise etkili olduğu kaza zincirlerinin oluşumu üzerine azaltıcı etkisi olduğu görülmektedir. Yaş faktörü genel olarak incelendiğinde, yaşın artmasının etkili olduğu kaza zincirlerinin oluşumu üzerinde gittikçe olasılığını arttırıcı bir etki oluşturmaya başladığı yorumu çıkartılabilir.
- Benzer şekilde işyeri büyüklükleri incelendiğinde ise çalışan sayısı arttıkça etkili olduğu kaza zincirlerinin oluşumu üzerinde olasılığı azaltan bir etki görülmektedir.

Dolaylı etkili değişkenlerden en etkili olanların incelenmesi amacıyla tüm şiddet grupları için dolaylı değişkenlerin etkili bulunduğu kaza zincirlerinden kaldıraç değeri 1.5'dan büyük olan seçilmiş daha sonra dolaylı değişkenlerden sadece en yüksek lift değerine sahip olanların seçilmesi ile aşağıdaki tablo oluşturulmuştur.

Ek-B.1'deki tablonun özeti Tablo 5.15 verilmektedir. Tabloda sadece en az bir şiddet grubunda ortaya çıkan kaza zincirlerinin kaldıraç değeri ortalaması 1.5'dan büyük olanlar gösterilmektedir.

Tablo 5.15 Etkili dolaylı değişkenler ve ortalama kaldıraç değerleri

Dolaylı Değişken	1-Raporlu	2-Major	3-KayipVar	4-OlumVar
KAZA_AY=2			2,02	
KAZA_IL_K=6			3,14	
KAZA_IL_K=10			15,08	
KAZA_IL_K=16	1,26		1,82	
KAZA_IL_K=20			4,47	
KAZA_IL_K=22			78,23	
KAZA_IL_K=26			10,13	
KAZA_IL_K=35	1,54	1,26	1,44	
KAZA_IL_K=38			10,32	
KAZA_IL_K=41	1,74		1,72	
KAZA_IL_K=45			12,96	
KAZA_IL_K=54			6,61	
KAZA_IL_K=59			4,69	
KAZA_IL_K=78			16,26	
KAZA_IL_K=80			51,74	
OGR_DURUM_K=2	1,13	1,11	1,02	1,72
OGR_DURUM_K=5			5,40	
SAAT_ARALIK=5	1,23	1,26	1,07	1,72
SAAT_ARALIK=8	0,94	0,95	0,91	1,50
SAAT_ARALIK=11			1,80	
SAAT_ARALIK=12			1,96	
SG_SONIS_TEC=0-2ay	0,92	1,31	0,92	3,28
SG_UCRETLIZINGUN=0	1,08	1,15	0,94	1,75
SG_UCRETLIZINGUN=15-30	1,17	1,16	1,26	1,53
SG_YAS=13-18			40,52	
SG_YAS=>50	1,22	1,28	1,03	3,44
SIG_SAY=Küçük(0-6)	1,45	1,42	1,09	1,49
SIG_SAY=Orta(6-50)	1,26	1,28	1,03	1,78
SIG_SAY=Dev(>2000)			3,25	

Dolaylı değişkenlerin etkili olduğu 2.711 kaza zincirinin hepsini incelemek mümkün olmadığı için her bir şiddet grubu için her bir dolaylı değişkenin en yüksek kaldıraç

değeriyle etkilediği kaza zincirlerini gösteren Tablo 5.16 oluşturulmuştur. Tabloda sadece dolaylı değişkenlerin etkilediği kaza zincirinin en yüksek kaldıraç değeri 1.5'dan büyükse gösterilmektedir. Bu noktada altının çizilmesi gereken bir husus Tablo 5.16'den incelenen dolaylı değişkenin en az tabloda verilen kadar veya çok yakın oranda etkili olduğu başka kaza zincirlerinin de bulunuyor olabileceğidir. Bir dolaylı değişkenin etkili olduğu tüm kaza zincirlerinin sorgulanması için Bölüm 5.3.3'de verilen yöntem uygulanmalıdır.

Tablo 5.16'de bazı kaza zincirleri üzerinde birden fazla dolaylı değişkenin eşik değerler ölçüsünde etkili olduğu görülmektedir. Bu kaza zincirleri Tablo 5.17'de verilmektedir.

Tablo 5.16 Dolaylı değişkenlerin en etkili olduğu kaza zincirleri

RHS	LHS-SIDDET	LHS	Kaza zincirleri	Destek	Güven	Kaldıraç	Sıklık
1	KAZA_IL_K=10	KayipVar	NACE_K.10=1 , CC_K.1=1 , OZELFAA_K.4=1	0,0010	0,19	18,25	189
2	KAZA_IL_K=16	KayipVar	NACE_K.24=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.11=1 , SG_MESLEK_ISCO08_K.39=1	0,0010	0,52	5,87	189
3	KAZA_IL_K=20	KayipVar	NACE_K.13=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.10=1 , SG_MESLEK_ISCO08_K.35=1	0,0014	0,11	4,59	254
4	KAZA_IL_K=22	KayipVar	NACE_K.14=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_K.2=1	0,0015	0,36	87,34	277
5	KAZA_IL_K=26	KayipVar	NACE_K.27=1 , CC_K.1=1 , GENELFAA_K.1=1 , SG_MESLEK_ISCO08_K.35=1	0,0017	0,27	11,85	318
6	KAZA_IL_K=34	KayipVar	NACE_K.56=1 , CC_K.1=1 , OZELFAA_MAG_K.11=1	0,0010	0,85	4,02	189
7	KAZA_IL_K=35	KayipVar	NACE_K.10=1 , GENELFAA_K.1=1 , OZELFAA_K.4=1	0,0012	0,24	2,77	220
8	KAZA_IL_K=35	Raporlu	CC_K.1=1 , GENELFAA_K.1=1 , YARAMAG_K.14=1	0,0012	0,16	1,80	224
9	KAZA_IL_K=38	KayipVar	NACE_K.25=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_K.4=1 , SG_MESLEK_ISCO08_K.42=1	0,0011	0,61	19,49	200
10	KAZA_IL_K=41	KayipVar	NACE_K.24=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.14=1 , SG_MESLEK_ISCO08_K.35=1	0,0015	0,30	3,45	285
11	KAZA_IL_K=41	Raporlu	CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.14=1	0,0010	0,17	1,98	192
12	KAZA_IL_K=45	KayipVar	NACE_K.5=1 , CC_K.10=1 , GENELFAA_K.1=1 , OZELFAA_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0022	0,83	17,95	409
13	KAZA_IL_K=59	KayipVar	NACE_K.13=1 , CC_K.1=1 , OZELFAA_MAG_K.11=1	0,0015	0,25	6,53	278
14	KAZA_IL_K=6	KayipVar	NACE_K.24=1 , CC_K.1=1 , GENELFAA_K.1=1 , SG_MESLEK_ISCO08_K.42=1	0,0011	0,35	5,19	212
15	KAZA_IL_K=80	KayipVar	NACE_K.24=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.10=1 , KAZA_MAG_K.10=1 , YARAMAG_K.10=1 , SG_MESLEK_ISCO08_K.35=1	0,0013	0,28	70,06	237
16	KAZA_MEVSIM=2	OlumVar	CC_K.2=1 , OZELFAA_K.3=1 , KAZA_MAG_K.12=1	0,0001	0,63	2,30	19

Tablo 5.16 Dolaylı değişkenlerin en etkili olduğu kaza zincirleri - (devamı)

RHS	LHS-SIDDET	LHS	Kaza zincirleri	Destek	Güven	Kaldıraç	Sıklık
17	KAZA_MEVSIM=4	KayipVar	NACE_K.5=1 , CC_K.10=1 , OZELFAA_K.2=1	0,0012	0,37	1,65	221
18	SAAT_ARALIK=5	KayipVar	OZELFAA_MAG_K.1=1 , KAZA_MAG_K.1=1 , SG_MESLEK_ISCO08_K.39=1	0,0010	0,18	1,54	194
19	SAAT_ARALIK=5	OlumVar	NACE_K.41=1 , CC_K.2=1 , SG_KAZAGERORTAM_K.2=1	0,0002	0,21	1,80	32
20	SAAT_ARALIK=6	KayipVar	CC_K.2=1 , OZELFAA_K.4=1 , OZELFAA_MAG_K.6=1	0,0012	0,29	1,53	233
21	SAAT_ARALIK=8	OlumVar	NACE_K.41=1 , CC_K.2=1 , SG_KAZAGERORTAM_K.2=1 , YARAMAG_K.2=1	0,0001	0,32	1,77	19
22	SAAT_ARALIK=11	KayipVar	NACE_K.24=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.14=1	0,0015	0,10	1,81	286
23	SAAT_ARALIK=12	KayipVar	NACE_K.5=1 , CC_K.10=1 , SG_MESLEK_ISCO08_K.39=1	0,0010	0,15	2,68	196
24	CALIS_SURE=6-8	OlumVar	CC_K.2=1 , OZELFAA_MAG_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0001	0,34	1,54	19
25	SG_TOP_TEC=1-5yil	KayipVar	NACE_K.10=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_K.2=1	0,0012	0,24	1,51	221
26	SG_SONIS_TEC=0-2ay	Raporlu	GENELFAA_K.5=1 , OZELFAA_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0013	0,63	3,05	238
27	SG_SONIS_TEC=0-2ay	Major	CC_K.2=1 , SG_KAZAGERORTAM_K.2=1 , OZELFAA_MAG_K.2=1	0,0012	0,66	3,20	224
28	SG_SONIS_TEC=0-2ay	KayipVar	CC_K.2=1 , OZELFAA_MAG_K.1=1 , KAZA_MAG_K.1=1 , SG_MESLEK_ISCO08_K.30=1	0,0011	0,63	3,05	216
29	SG_SONIS_TEC=0-2ay	OlumVar	NACE_K.41=1 , CC_K.2=1 , SG_KAZAGERORTAM_K.2=1 , OZELFAA_MAG_K.2=1 , KAZA_MAG_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0001	0,76	3,65	25
30	SG_SONIS_TEC=2ay-1yil	KayipVar	NACE_K.5=1 , CC_K.10=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.14=1 , SG_MESLEK_ISCO08_K.39=1	0,0013	0,58	1,81	253
31	SG_SONIS_TEC=2ay-1yil	OlumVar	KAZA_MAG_K.14=1 , YARAMAG_K.14=1 , SG_MESLEK_ISCO08_K.39=1	0,0001	0,54	1,69	19
32	SG_SONIS_TEC=>5yil	KayipVar	NACE_K.24=1 , CC_K.1=1 , OZELFAA_K.4=1	0,0033	0,31	2,29	617
33	SG_UCRETLIZINGUN=0	Raporlu	CC_K.2=1 , GENELFAA_K.5=1 , OZELFAA_K.5=1	0,0014	0,90	1,76	270

Tablo 5.16 Dolaylı değişkenlerin en etkili olduğu kaza zincirleri - (devamı)

RHS	LHS-SIDDET	LHS	Kaza zincirleri	Destek	Güven	Kaldıraç	Sıklık
34	SG_UCRETLIZINGUN=0	Major	NACE_K.41=1 , CC_K.2=1 , KAZA_MAG_K.2=1	0,0012	0,91	1,78	225
35	SG_UCRETLIZINGUN=0	KayipVar	CC_K.2=1 , OZELFAA_K.4=1 , KAZA_MAG_K.1=1 , SG_MESLEK_ISCO08_K.39=1	0,0012	0,89	1,73	222
36	SG_UCRETLIZINGUN=0	OlumVar	NACE_K.41=1 , KAZA_MAG_K.2=1 , SG_MESLEK_ISCO08_K.30=1	0,0001	1,00	1,95	27
37	SG_UCRETLIZINGUN=1-15	KayipVar	NACE_K.24=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.11=1 , SG_MESLEK_ISCO08_K.39=1	0,0013	0,65	2,29	239
38	SG_UCRETLIZINGUN=15-30	KayipVar	NACE_K.24=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_K.4=1 , SG_MESLEK_ISCO08_K.35=1	0,0011	0,40	2,39	204
39	SG_YAS=>50	Raporlu	CC_K.2=1 , GENELFAA_K.5=1 , SG_MESLEK_ISCO08_K.39=1	0,0014	0,23	1,88	262
40	SG_YAS=>50	KayipVar	CC_K.2=1 , OZELFAA_K.2=1 , SG_MESLEK_ISCO08_K.30=1	0,0011	0,23	1,92	203
41	SG_YAS=>50	OlumVar	SG_KAZAGERORTAM_K.999=1 , OZELFAA_K.3=1 , SG_MESLEK_ISCO08_K.36=1	0,0001	0,51	4,22	27
42	MEDENIHAL_K=2	OlumVar	NACE_K.41=1 , CC_K.2=1 , YARAMAG_K.2=1 , SG_MESLEK_ISCO08_K.30=1	0,0001	1,00	1,54	19
43	OGR_DURUM_K=2	Raporlu	GENELFAA_K.5=1 , OZELFAA_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0011	0,57	1,79	213
44	OGR_DURUM_K=2	Major	NACE_K.41=1 , CC_K.2=1 , SG_KAZAGERORTAM_K.2=1	0,0031	0,56	1,78	579
45	OGR_DURUM_K=2	KayipVar	NACE_K.5=1 , CC_K.10=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.14=1 , SG_MESLEK_ISCO08_K.39=1	0,0015	0,65	2,06	283
46	OGR_DURUM_K=2	OlumVar	NACE_K.41=1 , CC_K.2=1 , SG_KAZAGERORTAM_K.2=1 , OZELFAA_MAG_K.2=1 , KAZA_MAG_K.2=1 , YARAMAG_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0001	0,63	2,00	19
47	OGR_DURUM_K=4	KayipVar	NACE_K.24=1 , CC_K.1=1 , GENELFAA_K.1=1 , OZELFAA_MAG_K.10=1 , KAZA_MAG_K.10=1 , YARAMAG_K.10=1 , SG_MESLEK_ISCO08_K.35=1	0,0031	0,68	1,99	576

Tablo 5.16 Dolaylı değişkenlerin en etkili olduğu kaza zincirleri - (devamı)

RHS	LHS-SIDDET	LHS	Kaza zincirleri	Destek	Güven	Kaldıraç	Sıklık
48	SIG_SAY=Küçük(0-6)	Raporlu	GENELFAA_K.1=1 , OZELFAA_K.2=1 , KAZA_MAG_K.7=1	0,0011	0,25	1,99	206
49	SIG_SAY=Küçük(0-6)	KayipVar	NACE_K.25=1 , OZELFAA_K.2=1 , YARAMAG_K.7=1	0,0017	0,22	1,75	326
50	SIG_SAY=Küçük(0-6)	OlumVar	CC_K.1=1 , SG_KAZAGERORTAM_K.2=1 , GENELFAA_K.1=1	0,0001	0,22	1,81	20
51	SIG_SAY=Orta(6-50)	Raporlu	CC_K.2=1 , OZELFAA_K.2=1 , YARAMAG_K.7=1	0,0010	0,53	1,59	190
52	SIG_SAY=Orta(6-50)	KayipVar	NACE_K.5=1 , CC_K.10=1 , GENELFAA_K.1=1 , OZELFAA_K.2=1 , SG_MESLEK_ISCO08_K.39=1	0,0016	0,61	1,85	304
53	SIG_SAY=Orta(6-50)	OlumVar	CC_K.2=1 , OZELFAA_MAG_K.2=1 , KAZA_MAG_K.2=1 , SG_MESLEK_ISCO08_K.30=1	0,0001	0,88	2,64	21
54	SIG_SAY=Büyük(50-2000)	KayipVar	NACE_K.56=1 , CC_K.1=1 , OZELFAA_MAG_K.11=1	0,0010	0,89	1,76	197

Tablo 5.17 Birden fazla dolaylı değişkenden etkilenen kaza zincirleri

	Kaza zincirleri	1-Raporlu	2-Major	3-KayipVar	4-OlumVar
1	**GENELFAA_K.5=1, OZELFAA_K.2=1, SG_MESLEK_ISCO08_K.39=1**				
	OGR_DURUM_K=2	1,79			
	SG_SONIS_TEC=0-2ay	3,05			
2	**NACE_K.24=1, CC_K.1=1, GENELFAA_K.1=1, OZELFAA_MAG_K.10=1, KAZA_MAG_K.10=1, YARAMAG_K.10=1, SG_MESLEK_ISCO08_K.35=1**				
	KAZA_IL_K=80			70,06	
	OGR_DURUM_K=4			1,99	
3	**NACE_K.41=1, CC_K.2=1, SG_KAZAGERORTAM_K.2=1**				
	OGR_DURUM_K=2		1,78		
	SAAT_ARALIK=5				1,80
4	**NACE_K.5=1, CC_K.10=1, GENELFAA_K.1=1, OZELFAA_K.2=1, SG_MESLEK_ISCO08_K.39=1**				
	KAZA_IL_K=45			17,95	
	SIG_SAY=Orta(6-50)			1,85	
5	**NACE_K.5=1, CC_K.10=1, GENELFAA_K.1=1, OZELFAA_MAG_K.14=1, SG_MESLEK_ISCO08_K.39=1**				
	OGR_DURUM_K=2			2,06	
	SG_SONIS_TEC=2ay-1yil			1,81	
6	**NACE_K.56=1, CC_K.1=1, OZELFAA_MAG_K.11=1**				
	KAZA_IL_K=34			4,02	
	SIG_SAY=Büyük(50-2000)			1,76	

Tablo 5.16 ve Tablo 5.17 incelendiğinde en fazla kaldıraç değeriyle en yüksek etkiye sahip dolaylı değişkenlerin iller olduğu görülmektedir. Edirne (KAZA_IL_K=22) ve Osmaniye (KAZA_IL_K=80) illeri kaldıraç değeri sırasıyla 87 ve 70 olan iki kaza zinciri ile göze çarpmaktadır.
Bu iki kaza zincirinin açıklaması şöyle verilebilir;

- Giyim eşyalarının imalatı (NACE_K.14) faaliyeti gösteren işyerlerinde, üretim alanı, fabrika, atölyede (CC_K.1), üretim imalat işleme işlerinde (GENELFAA_K.1), el makineleriyle çalışırken (OZELFAA_K.5) gerçekleşen iş günü kayıplı kazaların Edirne ilinde (KAZA_IL_K=22) gerçekleşmesi durumunda ortaya çıkma olasılığı 87 kat artmaktadır.
- Ana metal sanayi sektöründe (NACE_K.24) faaliyet gösteren işyerlerinde, üretim alanı, fabrika, atölyede (CC_K.1), üretim imalat işleme işlerinde (GENELFAA_K.1), sabit makine ve ekipmanlar ile çalışırken (OZELFAA_MAG_K.10, KAZA_MAG_K.10, YARAMAG_K.10), sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35) görevi esnasında gerçekleşen iş günü kayıplı kazaların Osmaniye ilinde (KAZA_IL_K=80) gerçekleşmesi durumunda ortaya çıkma olasılığı 70 kat artmaktadır.

Kaldıraç değeri 12 ile 19 arasında değişen dört kaza zincirinde etkili iller ise sırasıyla Kayseri, Balıkesir, Manisa, Eskişehir olmuştur;

- Bu illerin etkilediği kaza zincirleri sırasıyla Makine ve teçhizat hariç fabrikasyon metal ürünleri imalatı (NACE_K.25), Gıda ürünlerinin imalatı (NACE_K.10), Kömür ve Linyit Çıkartılması (NACE_K.5), Elektrikli teçhizat imalatı (NACE_K.27) NACE kodlarını içermektedir.
- Manisa hariç tüm illerin etkili olduğu kaza zincirleri üretim faaliyetleri içerirken (CC_K.1), Manisa, inşaat mevkiinde (CC_K.10) gerçekleşen kaza zincirinde etkili bulunmuştur.

Kaldıraç değeri 5 ile 7 arasında değişen dört kaza zincirinde etkili iller ise sırasıyla Tekirdağ, Bursa, Ankara ve Denizli olmuştur;

- Tekirdağ ve Denizli illerin etkilediği kaza zincirleri tekstil ürünlerinin imalatı (NACE_K.13) NACE kodu içerirken, Bursa ve Ankara ana metal sanayi (NACE_K.24) NACE kodu içermektedir.

- Tekirdağ ve Bursa illerinin etkili olduğu kaza zincirleri aktarma, taşıma ve depolama sistemleri (OZELFAA_MAG_K.11) özel faaliyet kodunu içerirken, Denizli ilinin etkili olduğu kaza zincirleri sabit makine ve ekipmanlar (OZELFAA_MAG_K.10) özel faaliyet kodunu içermektedir.

Bu illerin etkili olduğu kaza zincirlerinde meslek/görev faktörü incelendiğinde ise meslek/görevlerin çoğunlukla sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35), montajcılar (SG_MESLEK_ISCO08_K.36), tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar (SG_MESLEK_ISCO08_K.39) olduğu görülmektedir. Ancak Kayseri ve Ankara'da ilginç bir bulgu olarak meslek/görev, cadde ve sokaklarda satış ve hizmet işlerinde çalışanlar (SG_MESLEK_ISCO08_K.42) olarak ortaya çıkmıştır. Üretim ortamına uygun olmayan bu görevin iki il için de 200 farklı kazada ortaya çıkması incelenmesi gereken bir durum oluşturduğu şeklinde yorumlanabilir.

Tablo 5.16'da iller dışında diğer dolaylı faktörler incelendiğinde etkilediği kaza zinciri üzerindeki etkisi yüksek olan (kaldıraç değeri > 3) faktörler şu şekildedir;

- Kazanın gerçekleştiği ortamın bilinmeyen olarak seçildiği, özel faaliyetin Sürücülük/taşıt aracında bulunmak veya donanım kullanımı (OZELFAA_K.3=1) ve meslek/görevin montajcılar (SG_MESLEK_ISCO08_K.36) olması durumunda 50 yaşından büyük olma faktörü (SG_YAS=>50) ölümlü kaza olasılığını 4,22 kat arttırmaktadır.

- NACE faaliyet kodu yiyecek ve içecek hizmeti faaliyetleri (NACE_K.56) olduğunda, üretim alani, fabrika, atölyede (CC_K.1), özel faaliyetin aktarma, taşıma ve depolama sistemleri (OZELFAA_MAG_K.11) olması durumunda kazanın gerçekleştiği ilin İstanbul olması faktörü (KAZA_IL_K=34) iş günü kayıplı kaza olasılığını 4,02 kat, SIG_SAY=Büyük(50-2000) olması faktörü ise iş günü kayıplı kaza olasılığını 1,76 kat arttırmaktadır (Tablo 5.17).

- NACE faaliyet kodu ana metal sanayii (NACE_K.24) olduğunda, üretim, imalat, isleme işlerinde (GENELFAA_K.1), özel faaliyetin malzeme, nesne, ürün, makine aksamı, enkaz, toz (OZELFAA_MAG_K.14) olması durumunda, meslek/görevin sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35) kazanın gerçekleştiği ilin Kocaeli olması faktörü (KAZA_IL_K=41) iş günü kayıplı kaza olasılığını 3,45 kat arttırmaktadır.

- NACE faaliyet kodu bina inşaatı (NACE_K.41), çalışılan çevre bakim alani, onarim atölyesi (CC_K.2) ve dinlenme/çalışma durumundan çalışma durumu

olduğunda (SG_KAZAGERORTAM_K.2) kazalının ilkokul mezunu olması (OGR_DURUM_K.2) durumunda kazanın 15 günden fazla iş günü kaybıyla sonuçlanma olasılığı 1,78 kat artarken, saatin sabah 08:00-10:00 arası olması faktörü (SAAT_ARALIK=5) kazanın ölümle sonuçlanma olasılığını 1,80 kat arttırmaktadır (Tablo 5.17).

- NACE faaliyet kodu gene bina inşaatı faaliyetlerinde (NACE_K.41) bakim alani, onarim atölyesinde (CC_K.2) dinlenme/çalışma durumundan çalışma durumu iken (SG_KAZAGERORTAM_K.2) kaza zincirine malzeme/araç/gereç yer düzeyinin üstünde kapalı veya açık binalar, yapılar, yüzeyler (OZELFAA_MAG_K.2) eklendiğinde gerçekleşen ölümlü kaza zincirinin oluşumunu kazalının son iş tecrübesinin 0-2 ay olması faktörü (SG_SONIS_TEC=0-2ay) 3,65 kat arttırmaktadır.

- NACE faaliyet kodu Kömür ve Linyit Çıkartılması (NACE_K.5), genel faaliyetin üretim, imalat, işleme işleri (GENELFAA_K.1) ve meslek/görev tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar (SG_MESLEK_ISCO08_K.39) faktörlerini içeren iki kaza zinciri 4 farklı dolaylı değişkenden etkilenmektedir. Bu kaza zincirine el makineleriyle çalisma durumunda (OZELFAA_K.2) eklendiğinde iş günü kayıplı kaza oluşumu, kazanın gerçekleştiği ilin Manisa olması faktörü (KAZA_IL_K=45) ile 17,95 kat, işyerindeki sigortalı sayısının 6-50 arası olması faktörü (SIG_SAY=Orta(6-50)) ile ise 1,85 kat artmaktadır. Aynı kaza zincirine özel faaliyetin malzeme, nesne, ürün, makine aksamı, enkaz, toz (OZELFAA_MAG_K.14) eklendiğinde ise iş günü kayıplı kaza oluşumu, kazalının ilkokul mezunu olması (OGR_DURUM_K.2) ile 2,06 kat, kazalının son iş tecrübesinin 0-2 ay olması faktörü (SG_SONIS_TEC=0-2ay) ile ise 1,81 kat artmaktadır (Tablo 5.17).

- Kazalının son iş tecrübesinin 0-2 ay olması faktörü (SG_SONIS_TEC=0-2ay), bakim alani, onarim atölyesinde (CC_K.2) (SG_KAZAGERORTAM_K.2) malzeme/araç/gereç yer düzeyinin üstünde kapalı veya açık binalar, yapılar, yüzeyler olduğunda (OZELFAA_MAG_K.2) gerçekleşen 15 günden fazla iş günü kayıplı kazaların oluşumunu 3,20 kat arttırmaktadır.

- Kazalının son iş tecrübesinin 0-2 ay olması faktörü (SG_SONIS_TEC=0-2ay), yeni bina inşaatı esnasında (GENELFAA_K.5), el makineleriyle çalisma durumunda (OZELFAA_K.2), meslek/görev tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar (SG_MESLEK_ISCO08_K.39) olduğu durumda gerçekleşen 3 günden fazla iş günü kayıplı kaza oluşumunu 3,05 kat arttırmaktadır.

- Kazalının son iş tecrübesinin 0-2 ay olması faktörü (SG_SONIS_TEC=0-2ay), bakim alani, onarim atölyesinde (CC_K.2) malzeme/araç/gereç yer düzeyindeki her çeşit binalar, yapılar, yüzeyler olduğunda (OZELFAA_MAG_K.1 KAZA_MAG_K.1) meslek/görev, inşaat ve ilgili işlerde çalışan sanatkarlar (elektrikçiler hariç) (SG_MESLEK_ISCO08_K.30) olduğunda gerçekleşen iş günü kayıplı kaza oluşumunu 3,05 kat arttırmaktadır.

- (GENELFAA_K.5), el makineleriyle çalisma durumunda (OZELFAA_K.2), meslek/görev tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar (SG_MESLEK_ISCO08_K.39) olduğu durumda gerçekleşen 3 günden fazla iş günü kayıplı kazaların oluşumunu k,azalının son iş tecrübesinin 0-2 ay olması faktörü (SG_SONIS_TEC=0-2ay) 3,05 kat arttırırken, kazalının ilkokul mezunu olması (OGR_DURUM_K.2) 1,79 kat arttırmaktadır (Tablo 5.17).

Tablo 5.16 incelendiğinde ana metal sanayi NACE kodu (NACE_K.24) göze çarpmaktadır. Sadece bu faktörü içeren kazalar tabloda incelendiğinde aşağıdaki yorumlar ortaya çıkmaktadır;

- Dolaylı değişkenlerin etkili olduğu kaza zincirlerinden ana metal sanayi sektöründe (NACE_K.24) faaliyet gösteren işyerlerini içeren kaza zincirlerinin tümü sadece iş günü kayıplı kazalarda, çalışılan çevrenin üretim alanı, fabrika, atölye olması durumunda (CC_K.1) ve/veya genel faaliyetin üretim, imalat, işleme işleri oması durumunda (GENELFAA_K.1) ortaya çıkmıştır.

- İl bazında iş günü kayıplı kazaların oluşumunu kaldıraç değeri 70 ile en çok arttıran il Osmaniye (KAZA_IL_K=80) olmuştur. Dİğer iller etki sıralarıyla Bursa (KAZA_IL_K=16), Ankara (KAZA_IL_K=6) ve Kocaeli (KAZA_IL_K=41) olarak gözükmektedir.

- Özel faaliyet sabit makine ve ekipmanlar ile çalışırken (OZELFAA_MAG_K.10, KAZA_MAG_K.10, YARAMAG_K.10) ve meslek/görev sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35) iken kazanın Osmaniye ilinde (KAZA_IL_K=80) gerçekleşmesi faktörü iş günü kayıplı kazanın oluşum olasılığını 70 kat arttırken aynı zincirin oluşum olasılığını öğrenim durumunun lise olması da (OGR_DURUM_K=4) 1,99 kat arttırmaktadır (Tablo 5.17).

- Osmaniye ilinin etkili olduğu kaza zincirinde özel faaliyet sabit makine ve ekipmanlar ile çalışırken (OZELFAA_MAG_K.10, KAZA_MAG_K.10, YARAMAG_K.10) iken, Bursa'nın etkili olduğu kaza zincirinde özel faaliyet aktarma, taşıma ve depolama sistemleri (OZELFAA_MAG_K.11)

ve Kocaeli'de özel faaliyet malzeme, nesne, ürün, makine aksamı, enkaz, toz (OZELFAA_MAG_K.14) çalışmaları olarak belirmiştir.

- Osmaniye ve Kocaeli ilinin etkili olduğu kaza zincirinde meslek/görev sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35) iken, Bursa'nın etkili olduğu kaza zincirinde meslek/görev tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar (SG_MESLEK_ISCO08_K.39), Ankara'nın etkili olduğu kaza zincirinde meslek/görev cadde ve sokaklarda satış ve hizmet işlerinde çalışanlar (SG_MESLEK_ISCO08_K.42) olarak belirmiştir.

- Gerçekleştirilen özel faaliyetin malzeme, nesne, ürün, makine aksamı, enkaz, toz (OZELFAA_MAG_K.14) çalışmaları olması durumunda çalışılan saat aralığının 22:00-24:00 arası olması (SAAT_ARALIK=11) iş günü kayıplı kaza olasılığını 1,81 kat arttırmaktadır.

- Gerçekleştirilen özel faaliyetin nesnelerin kullanımı (OZELFAA_K.4) olması durumunda, son iş tecrübesinin 5 yıldan fazla olması (SG_SONIS_TEC=>5yil), iş günü kayıplı kaza olasılığını 2,29 kat arttırmaktadır. Bu özel faaliyetle birlikte meslek/görev sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35) olduğunda ise ücretli izin günün 15-30 gün arası olması (SG_UCRETLIZINGUN=15-30), iş günü kayıplı kaza olasılığını 12,39 kat arttırmaktadır.

- Ücretli izin gününün 1-15 gün arası olması (SG_UCRETLIZINGUN=1-15), özel faaliyet olarak aktarma, taşıma ve depolama sistemleri (OZELFAA_MAG_K.11) içeren, tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar (SG_MESLEK_ISCO08_K.39) meslek/görev tanımı içeren iş günü kayıplı kazaların olasılığını 2,29 kat artırırken, ücretli izin günün 15-30 gün arası olması (SG_UCRETLIZINGUN=15-30), özel faaliyet olarak nesnelerin kullanımı (OZELFAA_K.4) ve meslek/görev olarak sabit makine ve tesis operatörleri (SG_MESLEK_ISCO08_K.35) içeren iş günü kayıplı kazaların olasılığını 12,39 kat arttırmaktadır.

  Tekrar altını çizmek gerekirse; burada verilen yorumlar Tablo 5.16 ile sınırlıdır. NACE kodu olarak ana metal sanayi (NACE_K.24) içeren tüm kaza zincirlerinin sorgusu Bölüm 5.3.3'de verilmektedir.

Analizin kaza zincirlerinin ortaya çıkarılması bölümünün ilk adımını oluşturan dolaylı değişken etkilerini içeren kuralların oluşturulması adımlarının sözde kodu Algoritma 3'de verilmektedir.

Algorithm 3 Dolaylı değişken etkilerini içeren kuralların oluşturulması

Input: Her bir şiddet grubu için tahminleme performans değerleri ve tahminleyici değişkenlerden oluşan veri seti

Output: Her bir şiddet grubu için dolaylı etkili değişkenlerin etkisini gösteren birliktelik kuralları

1: **for** Bütün şiddet grupları için **do**
2: Kuralın soluna kazanın oluşumunda direk etkili değişkenleri ve şiddet değişkenini koy
3: Kuralın sağına kazanın oluşumunda dolaylı etkili değişkenleri koy
4: 10.000'de 1 destek, 100'de 10 güven aralığıiçinde en az 5 değişken ile oluşan tüm kuralları bul
5: **if** Kaldıraç değeri 1.5 altı olan kural var ise **then**
6: Kural sil
7: **end if**
8: Kurallara şiddet tahmin performans göstergelerini ekleyerek veri tabanını oluştur
9: **end for**

5.3.2 Şiddet ve Yaralanma Sonucu İçeren Kaza Zincirlerinin Ortaya Çıkarılması

Kuralların oluşturulmasındaki diğer yaklaşım ise şiddet ve yaralanma üzerinden kaza sonucunu ortaya koyan kaza zincirlerinin ortaya çıkarılmasıdır. Bu analizin amacı, karar alıcıların belirli faktörlerle ilgili kaza zincirlerine ulaşmasını sağlayarak, karar alıcılara karar destek verisi sunmaktır. Karar destek için gerekli veriyi oluşturmak için önce makine öğrenmesi ile tahmin analizi yapılmış ikili şiddet sınıflandırmalarına ait tahminleyici değişkenlerin bulunduğu verilerden kurallar ayrı ayrı oluşturulur ve daha sonra tüm bulunan kurallar birleştirilir. Kural oluştururken her bir ikili sınıf için 3 ayrı adımda kurallar oluşturulur;

1. X öncül öğeleri için şiddet sınıflandırmasına özel tahminleyici değişkenler, Y sonuç öğesi için ise birinci adımda ikili sınıflandırmanın pozitif kabul edilen sınıf seçilirek kurallar bulunur,
2. İkili sınıflandırmanın pozitif kabul edilen sınıfı X öncül değerlere eklenir, Y sonuç öğesi için yaralanma türü (YT) değişkenleri seçilirerek kurallar bulunur,
3. İkili sınıflandırmanın pozitif kabul edilen sınıfı X öncül değerlere eklenir, Y sonuç öğesi için yaralanmanın vücuttaki yeri (YVY) değişkeni seçilerek kurallar bulunur.

Tüm bu 3 adımda bulunan kurallar birleştirilerek ikili sınıflandırma grubuna ait kurallar belirlenmiş olur. Verilen bu üç adım tüm ikili şiddet grupları için tekrar edilir. Tüm kurallar birleştirildiğinde karar vericiler için karar destek veri tabanı oluşturulmuş olur.

Kural oluştururken sık geçen öğelerin bulunmasında kullanılan eşik değerler analist tarafından belirlenmektedir. Bölüm 3.1'de açıklandığı gibi kaza analizi özelinde yüksek kaldıraç değerlerine sahip kuralların bulunması önemli olduğundan analistlerin destek ve güven değerlerini hesaplama kapasite kaynağının izin verdiği ölçüde düşük tutabilmeleri önerilir. Tez çalışmasının bu aşamasındaki analizde kullanılan eşik değerler şu şekildedir;

Tüm şiddet grupları için destek değeri en az 0,0001

Tüm şiddet grupları için güven değeri en az 0,001

Bu eşik değerleri ile ortaya çıkan kural sayıları, kuralların minimum ve maksimum kaldıraç değerleri ve kaldıraç değeri 2'nin üstünde olan (K.d > 2) kaza sayıları Tablo 5.18'de verilmektedir.

Tablo 5.18 Kural sayıları

Şiddet Grup	Sonuç (RHS)	Destek>0.0001, Güven >0.001	Min. K.d.	Maks. K.d	K.d > 2
1	SIDDET=Raporlu	446.847	0,14	5,08	41.345
	YT	350.292	0,21	27,51	115.084
	YVY	372.884	0,08	6,24	3.014
2	SIDDET=Major	187.712	0,14	6,54	35.835
	YT	135.603	0,12	67,25	65.750
	YVY	140.254	0,09	19,11	4.136
3	SIDDET=KayipVar	1.197.952	0,03	1,46	0
	YT	1.365.011	0,20	17,46	37.786
	YVY	1.357.474	0,13	27,73	3.354
4	SIDDET=OlumVar	5.348	0,30	30,24	4.762
	YT	1.321	0,14	118,22	1.314
	YVY	1.498	0,49	31,92	1.466
Toplam					313.846

Görüldüğü üzere düşük eşik değerler yüksek sayıda kural ortaya koymaktadır. Bu kuralların kaldıraç değerleri ilgi çekici kuralların seçilmesi için kullanılabilmektedir. Tez çalışmasının uygulama aşamasında minimum kaldıraç eşik değeri 2 olarak seçilmiştir. Başka bir deyişle, Y sonuç öğesinin ortaya çıkış olasılığını en az 2 kat arttıran kaza zincirlerine ait kurallar ilginç kurallar olarak belirlenmiştir.

Önerilen yaklaşım ile yapılan analiz sonucunda karar alıcılar için 313.846 kural oluşturulmuştur. Gelecekte oluşan kazalar ile yeni veriler oluştukça önerilen yaklaşımdaki analiz tekrarlanabilir ve yeni, güncel kurallar oluşturulabilir.

Analizin kaza zincirlerinin ortaya çıkarılması bölümünün ikinci adımını oluşturan şiddet ve yaralanma sonucunu içeren kuralların oluşturulması adımlarının sözde kodu Algoritma 4'de verilmektedir.

Algorithm 4 Şiddet ve yaralanma sonucunu içeren kuralların oluşturulması

Input: Her bir şiddet grubu için tahminleme performans değerleri ve tahminleyici değişkenlerden oluşan veri seti

Output: Her bir şiddet grubu için şiddet ve yaralanma zincirlerini gösteren birliktelik kuralları

for Bütün şiddet grupları için **do**
    Kuralın soluna tahminleyici değişkenleri koy
    Kuralın sağına şiddet sınıflandırmasında pozitif kabul edilen sınıf değişkenini koy.
    10.000'de 1 destek ve 1.000'de 1 güven aralığı içindeki, kural içerisinde en az 5 değişkenin bulunduğu tüm kuralları bul
    if Kaldıraç değeri 2 altı olan kural var ise **then**
        Kural sil
    end if
    Kuralları veri tabanına ekle
    Kuralın soluna tahminleyici değişkenleri ve şiddet sınıflandırmasında pozitif kabul edilen sınıf değişkenini koy
    Kuralın sağına yaralanma türü değişkenini koy.
    10.000'de 1 destek ve 1.000'de 1 güven aralığı içindeki, kural içerisinde en az 5 değişkenin bulunduğu tüm kuralları bul
    if Kaldıraç değeri 2 altı olan kural var ise **then**
        Kural sil
    end if
    Kuralları veri tabanına ekle
    Kuralın soluna tahminleyici değişkenleri ve şiddet sınıflandırmasında pozitif kabul edilen sınıf değişkenini koy
    Kuralın sağına yaranın vucuttaki yeri değişkenini koy.
    10.000'de 1 destek ve 1.000'de güven aralığı içindeki, kural içerisinde en az 5 değişkenin bulunduğu tüm kuralları bul
    if Kaldıraç değeri 2 altı olan kural var ise **then**
        Kural sil
    end if
    Kuralları veri tabanına ekle
end for

5.3.3 Kaza Zinciri Sorgulama

Tez çalışmasında önerilen analiz yöntemi sonucu ilk aşamada ortaya çıkan 2.711 kural ile ikinci aşamada ortaya çıkan ortaya çıkan 313.846 kural birleştirildiğinde

ortaya 316.557 kaza zinciri çıkmıştır. Kullanım kolaylığı açısından doğru yaklaşım ortaya çıkan veri tabanının sorgu yapabilecekleri bir çevirim içi arayüz ile karar vericilere sunulmasıdır. Karar vericiler var olan koşulları bu arayüzden seçebilir ve bu koşulların oluşturabileceği kaza zincirlerini listeleyebilir. Mevcut aşamada kara vericiler https://github.com/frecal/doktora_tezi web adresinde bulunan kaza zincirleri dosyasını kullanarak aşağıda örneği bulunan kodlama ile sorgulama yapabilirler.

Bölüm 5.3.1'de dolaylı değişkenlerin etkilediği kaza zincirleri bulunurken dolaylı değişkenlerin en çok etkilediği ilk kaza zincirlerinin seçilerek oluşturulduğu Tablo 5.16'de ve Tablo 5.17'de sık ortaya çıkan ana metal sanayi sektörü (NACE_K.24), üretim alanı, fabrika, atölye (CC_K.1) ve/veya genel faaliyetin üretim, imalat, işleme işleri oması durumunda (GENELFAA_K.1) faktörlerini içeren kaza zinciri incelenmişti.

Tüm veri tabanında ana metal sanayi NACE kodu (NACE_K.24) sorgulandığında veri tabanında kaza zinciri faktörleri arasında bu NACE kodunu bulunduran 4.311 kaza zinciri olduğu görülmektedir.

Örnek olarak bir karar alıcı, ana metal sanayi (NACE_K.24) alanında faaliyet gösteren, 6-50 arası sayıda çalışanı olan bir işyerinde (SIG_SAY=Buyuk(50-2000)) işe yeni başlayan bir kişi için (SG_SONIS_TEC=0-2ay) aşağıdaki R kodunu kullanarak sorgu yaptığında, aşağıdaki 2 kaza zincirini görüntülemektedir.

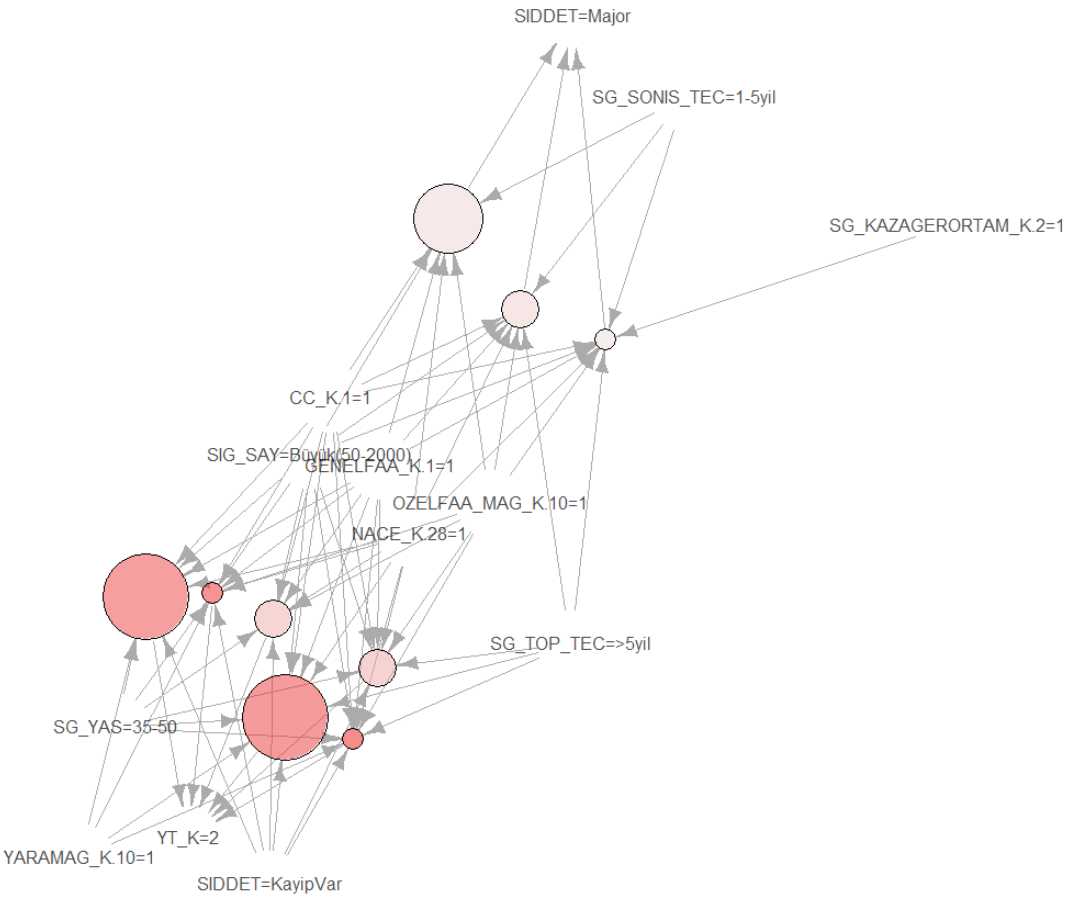
```
> load(paste0("RObjects/kaza_zincirleri.rdata"))
> query_lhs <- c("NACE_K.24=1",
                 "SIG_SAY=Buyuk(50-2000)",
                 "SG_SONIS_TEC=0-2ay")
> found_rules <- subset(rules, subset = lhs %ain% query_lhs)
> inspect(found_rules)
    lhs                          rhs            support   confidence  coverage    lift count
[1] {CC_K.1=1,
     GENELFAA_K.1=1,
     OZELFAA_K.2=1,
     SG_SONIS_TEC=0-2ay,
     SG_UCRETLIZINGUN=0,
     SIG_SAY=Buyuk(50-2000),
     NACE_K.24=1,
     SIDDET=KayipVar}        => {YVY_K=1}       0,000117  0,2136    0,00059     2,00    22
[2] {CC_K,1=1,
     OZELFAA_K,2=1,
     SG_SONIS_TEC=0-2ay,
     SG_TOP_TEC=>5yil,
     SG_UCRETLIZINGUN=0,
     SIG_SAY=Buyuk(50-2000),
     NACE_K,24=1,
     SIDDET=KayipVar}        => {YVY_K=1}       0,000101  0,2235    0,00045     2,09    19
```

Elbette her sorgunun bir kural listelemesi mümkün olmayacaktır. Bu durumda karar

alıcılardan bazı koşulların değiştirilmesi istenebilir veya bir kaç koşulun farklı olması durumunda ortaya çıkan kurallar önerilebilir.

Başka bir örnek bir karar alıcının belli bir işyeri özelinde risk değerlendirmesi için alacağı kararlar amacıyla kural sorgulaması yapması için verilebilir. Örneğin karar alıcı "Başka yerde sınıflandırılmamış makine ve ekipman imalatı" (NACE_K.28) faaliyeti gösteren, 50-2000 arası çalışanı bulunan (SIG_SAY=Büyük(50-2000)) bir işyerinde risk değerlendirme çalışması esnasında "Üretim alani, fabrika, atölye" (CC_K.1) alanında, "Üretim, imalat, isleme" (GENELFAA_K.1) süreçleri gerçekleştirilen ve kullanılan sabit makine ve ekipmanların (OZELFAA_MAG_K.2 veya YARAMAG_K.2 veya KAZA_MAG_K.2) kullanımı esnasında gerçekleşebilecek kazalara ait kuralların sorgulanması için aşağıdaki R kodunu kullanarak sorgu yaptığında Şekil 5.4'de gösterilen grafiğe ve Tablo 5.19'da verilen 9 kaza zinciri tablosuna ulaşmaktadır.

```
> load(paste0("RObjects/kaza_zincirleri.rdata"))
> query_lhs1 <- c("NACE_K.28=1", "CC_K.1=1", "GENELFAA_K.1=1", "SIG_SAY=Buyuk(50-2000)")
> query_lhs2 <- c("OZELFAA_MAG_K.10=1", "YARAMAG_K.10=1", "KAZA_MAG_K.10=1")
> found_rules <- subset(rules, subset = lhs %ain% query_lhs1) %>%
+                 subset(rules, subset = lhs %in% query_lhs2)
> plot(found_rules, method = "graph")
> found_rules_df <- DATAFRAME(found_rules, setStart='', setEnd='', separate = TRUE)
> row.names(found_rules_df) <- NULL
> write.table(found_rules_df, file=paste0("RObjects/kaza_sorgu.xls"),
+                 quote = FALSE, sep = "\t", dec = ",",
+                 row.names = FALSE, col.names = TRUE, append = F)
```



Şekil 5.4 İşyeri özelinde kural sorgusu grafiği

Tablo 5.19 İşyeri özelinde kural sorgusu sonuç tablosu

Kaza Zinciri					Destek	Güven	Kaldıraç	Sıklık
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1,	SG_SONIS_TEC=1-5yil,		SIDDET=Major	0,00012	0,0961	2,05	22
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1,	SG_SONIS_TEC=1-5yil, SG_TOP_TEC=>5yil,		SIDDET=Major	0,00011	0,0976	2,08	20
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1,	SG_SONIS_TEC=1-5yil, SG_TOP_TEC=>5yil, SG_KAZAGERORTAM_K.2=1		SIDDET=Major	0,00010	0,0941	2,01	19
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1,	SG_YAS=35-50,	SIDDET=KayipVar	YT_K=2	0,00011	0,2041	2,26	20
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1,	SG_TOP_TEC=>5yil, SG_YAS=35-50,	SIDDET=KayipVar	YT_K=2	0,00011	0,2062	2,28	20
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1, YARAMAG_K.10=1	SG_YAS=35-50,	SIDDET=KayipVar	YT_K=2	0,00010	0,2405	2,66	19
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	OZELFAA_MAG_K.10=1, YARAMAG_K.10=1	SG_TOP_TEC=>5yil, SG_YAS=35-50,	SIDDET=KayipVar	YT_K=2	0,00010	0,2436	2,70	19
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	YARAMAG_K.10=1	SG_TOP_TEC=>5yil, SG_YAS=35-50,	SIDDET=KayipVar	YT_K=2	0,00012	0,2371	2,63	23

Tablo 5.19 İşyeri özelinde kural sorgusu sonuç tablosu - (devamı)

Kaza Zinciri					Destek	Güven	Kaldıraç	Sıklık
NACE_K.28=1, CC_K.1=1, GENELFAA_K.1=1, SIG_SAY=Büyük(50-2000)	YARAMAG_K.10=1,	SG_YAS=35-50,	SIDDET=KayipVar	YT_K=2	0,00012	0,2347	2,60	23

Karar alıcının belirlediği koşullar ve bu koşullar ile birlikte ortaya çıkan diğer koşullar sonucunda oluşan kaza zincirleri incelendiğinde;

- Kaza şiddeti olarak 15 günden fazla iş günü kaybıyla sonuçlanan kazalar (SIDDET=Major) ve yaralanma türü olarak kemik kırıkları (YT_K=2) ile sonuçlanan kazalar görülmektedir.
- Karar alıcının belirlediği koşullar ile birlikte çalışanın son iş tecrübesi 1-5 yıl arası olduğunda 15 günden fazla iş günü kayıplı kaza olasılığı 2,05 kat artmaktadır.
- Aynı çalışanın toplam tecrübesi 5 yıldan fazla ise olasılık 2,08 kata çıkmaktadır. Bu bulgu tecrübeli hisseden çalışanların yeni bir işyerinde en az 1 yıl çalıştıktan sonra majör şiddette kaza yapma olasılıklarının arttığı şeklinde yorumlanabilir. Karar alıcı bu bulgu ışığında bu özelliği taşıyan çalışanlara hatırlatma eğitimleri aldırma, eğitim içeriğine tecrübeye güvenmek yerine kuralları uygulamanın önemine dair içerik ekleme gibi ek önleyici önlemler alabilir.
- Ayrıca karar alıcının belirlediği koşulların oluşturduğu kayıp günlü kazalarda kemik kırığı sonucu olasılığı en yüksek kuralın 2,70 kaldıraç değeriyle toplam tecrübesi 5 yıldan büyük olup yaş aralığı 35-50 arası çalışanların maruz kaldığı kazalar olduğu görülmektedir. Karar alıcılar bu bulgu ışığında özellikle bu niteliklere sahip çalışanları için kırıklara sebep olabilecek makine ve ekipmanlarda ek önlemler almak isteyebilirler.

Analizin son aşamasını oluşturan kaza zincirlerinin sorgulanması adımlarının sözde kodu Algoritma 5'de verilmektedir. Önerilen yönteme ait 5 aşamalı algoritma ulusal iş kazaları veri tabanına yeni veri eklendikçe tekrarlanabilir ve yeni veriler ışığında karar vericiler için yeni iş kazaları zincir kuralları oluşturulabilir.

Algorithm 5 Kaza zincirleri veri tabanının oluşturulması ve sorgulanması

Input: Her bir şiddet grubu için şiddet ve yaralanma zincirlerini gösteren birliktelik kuralları

Output: Sorgulanan kaza zincirleri

1: Tüm şiddet gruplarının birliktelik kurallarını birleştirerek kurallar veri tabanını oluştur.
2: Sorgu ekranından ilgili değişkenleri gir
3: Sorgu ekranından girilen değişkenleri veri tabanından sorgula
4: **if** Girdi değişkenleri barındıran kural var ise **then**
5: İlgili kuralları etkili model performans kısıtlarını belirterek listele
6: **end if**

6
SONUÇ VE ÖNERİLER

Ülkelerin iş kazası sayıları ülkelerin gelişmişlik seviyelerinin göstergelerinde önemli bir rol oynamaktadır [10]. Kaza verisi sadece nicel olarak değil nitel olarak da ülkelerin sistemlerinde risk oluşturan ve sağlıklı bir şekilde çalışmayan yönlerini tespit etmede kritik öneme sahiptir. Bu sebeple günümüzde ülkeler, gerçekleşen kazaları kayıt altına alarak veri tabanı oluşturmaktadır. Kaza veri tabanının sistematik bir şekilde analiz edilerek sistemlerin aksayan yönlerinin ortaya çıkarılması ve İSG konusunda ulusal düzeyde karar alan, mevzuat oluşturan kişi veya kurumlara ve işyerlerindeki karar alıcılara sunulması etkin kararlar alınabilmesi için gereklidir. Literatür incelemesi göstermiştir ki iş kazalarının analizinde makine öğrenimi, kural çıkarımı gibi veri madenciliği yöntemlerinin kullanımı son yıllarda artmaya başlamış ve etkinliği ortaya çıkan çalışmalarda kanıtlanmıştır.

Bu çalışma, karmaşık ulusal iş kazası verisini, kümeleme ve sınıflandırma veri madenciliği yöntemleriyle yorumlanabilir bir karar destek veri tabanına dönüştüren bir analiz yöntemi geliştirmeyi amaçlaması bakımından ve ulusal iş kazası verisini bir bütün olarak analiz eden ilk çalışma olması bakımından özgündür. Bu tez çalışmasında, iş kazalarının şiddetini, yaralanma sonuçlarını belirleyen kazanın oluşumda etkili faktörlerin makine öğrenmesi yöntemi ile belirlenmesinde ve bu ortaya çıkan faktörlerle kaza zincirleri kurallarının ortaya çıkarılmasında kullanılabilecek yöntem geliştirilirken uygulama aşamasında Türkiye'de 2013-2017 yılları arasında gerçekleşmiş kazalardan oluşan ulusal kaza veri seti kullanılmıştır.

Önerilen analiz yönteminin uygulamasında iş günü kaybı 3 günden fazla olan "Raporlu", iş günü kaybı 15 günden fazla olan "Major", iş günü kayıplı "KayipVar" ve ölümlü kazalar "OlumVar" olmak üzere 4 farklı ikili şiddet grubu seçilmiştir. Seçilen şiddet grubuna göre kazaların oluşumunda etkin faktörler belirlenirken kategorik tipteki değişkenler 0/1 değeri alan ikili değişkenlere dönüştürülmüş ve 7'si numerik olmak üzere toplam 368 değişken makine öğrenmesi algoritmaları kullanılarak değerlendirilmiştir. Seçilen ikili şiddet gruplarının tahminlenmesinde

daha önce gerçekleştirilen analizlerde iyi performansa gösteren SGB - Stokastik Gradyan Arttırma algoritması kullanılmış ve bu algoritma ile model belirlenirken sınıf dengezisliği problemi için farklı yöntemler incelenmiş, en verimli şekilde yüksek doğruluk ortaya koyan modelin eşik değer belirleme yöntemi ile ortaya çıktığı görülmüştür. Ortaya çıkan modellerin dengeli doğruluk değerleri Raporlu, Major, KayipVar ve Olumvar şiddet gruplarında sırasıyla 0.6211, 0.6411, 0.6192 ve 0.778 olarak bulunmuştur. Makine öğrenmesi literatüründe farklı veri setlerinde tahmin performansı %100'e yaklaşan modeller bulunuyor olsada, iş kazalarının oluşumunda etkili bütün faktörlerin veride bulunması mümkün olmadığından ortaya çıkan doğruluklar bu alandaki çalışmaların sürdürülmesi yönünde yinede motivasyon oluşturmaktadır. Her bir şiddet grubu için belirtilen doğruluk oranları ile geliştirilmiş modeller sırasıyla 105, 128, 93, 45 faktörü kaza oluşumunda etkili faktörler olarak belirlemiş ve kaza zincirleri bu faktörler ile oluşturulmuştur. Raporlu ve Major şiddet gruplarında tüm numerik değişkenler kaza oluşumunda etkili bulunurken, KayipVar durumundaki kazalarda çalışma süresi, ölümlü kazalarda hem çalışma süresi hemde ücretli izin gün sayıları etkili faktörler olarak bulunmamıştır. Etkili bulunan faktörler daha sonra birliktelik kuralları oluşturmak için kullanılmıştır. Birliktelik kuralları oluşturulurken iki ayrı yaklaşım belirlenmiştir. İlk yaklaşımda, belirlenen şiddet ile gerçekleşen kazalara, dolaylı değişkenlerin etkisini ortaya koymak amacı belirlenmiştir. İkinci yaklaşımda ise makine öğrenmesinde belirlenen faktörlerin, yaralanma türü (YT), yaranın vücuttaki yeri (YVY) ve belirlenen şiddet grubu ile birlikte ortaya çıktığı kazaları belirlemek amaçlanmıştır ve uygulamada 316.557 kaza zinciri ortaya çıkarılmıştır. İlk yaklaşım uygulamasında ortaya çıkan 2711 kaza zinciri ile birleştirilerek karar alıcıların sorgulama yapabilmesi için bir veri seti oluşturulmuştur. Önerilen yöntem uygulamada belirlenmiş dört ikili şiddet grubu dışında başka şiddet grupları için de tekrarlanabilir ve tekrarda ortaya çıkan kaza zincirleri karar destek veri tabanına eklenebilir. Ayrıca toplanan ham iş kazası verilerine yeni kazalar eklendikçe analiz tekrarlanabilir ve daha güncel iş kazası zincirleri ortaya çıkarılabilir.

Ortaya çıkan kaza zincirleri veri seti, makine öğrenmesi aşamasında ortaya çıkan tahmin doğruluğu ile sınırlıdır. Bu sebeple, iş kazalarının oluşumda belirleyici olan faktörlerin ortaya çıkarılmasında doğru ve güvenilir kaza verisinin önemi büyüktür. Bildirilen kazalara ilişkin iş kazası raporundaki tüm bilgiler tam ve eksiksiz bir şekilde doldurulmalıdır. Kullanılan iş kazası verilerine bakıldığında, doldurulması gereken pek çok kısım doldurulmadığı için "BELİRSİZ", "DİĞER" veya "BİLİNMEYEN" ifadelerine sıklıkla karşılaşılmaktadır. Bu durum iş kazası verisinde bilgi kaybına yol açarak, kök nedenlerin bulunmasını sınırlamaktadır. Jacinto ve Soares [108] çalışmalarında belirttikleri gibi ölümlü kazalara ilişkin bilinmeyen/belirtilmemiş

bilgilerin yüksek düzeyde olması bilgi edinilmesi en elzem konuda bilgi kaybına yol açmaktadır. İş kazası verisinin toplandığı çevirimiçi platformda bu gibi belirsiz seçenekleri seçen kullanıcılardan ek açıklama girmesini istemek bu sorunun çözümüne önerilebilir. Bu öneri hem kullanıcının kolaya kaçmasını engeller hem de girilen açıklama metin madenciliği gibi yöntemlerle analize katılabilir. Bu yöntemler kayıp bilginin azalmasını sağlayacaktır.

Analizin 2013-2021 yılları arasında gerçekleşen iş kazalarından oluşan daha büyük bir veri ile tekrarı önerilir. SGK veri tabanından yeni veriyi çekmeden, bölüm 5.1.1'de açıklanan Kod (K) ve Açıklama (KA) tablolarındaki uyumsuzluğun nedeni öncesinde araştırılmalı ve mümkünse uyumsuzluğun olmadığı bir veri seti oluşturulmalıdır. İş kazaları verisinde yer alan serbest girişli bölüm, iş kazası ile ilgili önemli bilgiler barındırabileceğinden, burada girilen metinin veri madenciliği ile işlenerek önerilen yönteme değişken olarak katılması öğrenme doğruluğunu arttırabilir [60].

Önerilen yöntemin uygulamasında ortaya çıkarılan karar destek veri seti R programlama dili ile sorgulanarak incelenmiştir. Bir sonraki geliştirmede çevirimiçi bir platform üzerinden karar vericilere sunulması için bir arayüzün geliştirilmesi önerilir. Arayüz, karar vericilerin kodlama yapmak zorunda kalmadan, açılır listelerden seçim yapmak yöntemiyle çeşitli sorgulamalar yaparak kaza zincirlerini görüntüleyebilmeleri ve açıklamalar sayesinde yorumlayabilmeleri için gereklidir. Sorgulanan kaza zincirleride yer alan faktörler, ilgili ulusal düzeyde faaliyetler belirlemede veya bu faktörlerle ilgili alınması gereken önlemler yönünden yol gösterici olacaktır. Ayrıca işyerlerinde gerçekleştirilen risk değerlendirme çalışmaları yönünden de karar alıcılar kendi sektörleri, faaliyet alanları özelinde kaza zincirlerini listeleyerek ve [80] çalışmasında önerildiği gibi ayrıntılı inceleyerek düzeltici önleyici faaliyetlerini belirleyebilirler.

Veri madenciliği tekniklerindeki gelişmeler, büyük veritabanlarının analizi için artan kapasite ve analiz için ihtiyaç duyulan algoritmaların yazılım kütüphanelerinin de artmasıyla bu tez çalışması ile önerilen yöntemin geliştirilmesi veya benzer yeni yöntemlerin oluşturulması ile ilerlemenin devam etmesi beklenmektedir.

KAYNAKÇA

[1] ILO, *Safety and Health at Work: A Vision for Sustainable Prevention*. Geneva, 2014, ISBN: 9789221289098.

[2] ——, *Improvement of national reporting, data collection and analysis of occupational accidents and diseases*. 2012, p. 12, ISBN: 978-92-2-126817-8. [Online]. Available: http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---safework/documents/publication/wcms_207414.pdf.

[3] A. K. Nurdoğan, "Uluslararası Çalışma Örgütünün(UÇÖ-ILO) Yüzüncü Yıl Dönümü ve Türkiye İlişkileri," *Bitlis Eren Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Akademik İzdüşüm Dergisi*, vol. 3, no. 4, pp. 78–95, 2018.

[4] ILO, "Technical and ethical guidelines for workers' health surveillance," Geneva, Tech. Rep., 1998.

[5] T. Cumhuriyeti, *6331 Sayılı İş Sağlığı ve Güvenliği Kanunu*, 2012. [Online]. Available: https://www.mevzuat.gov.tr/mevzuatmetin/1.5.6331.pdf.

[6] ILOSTAT, *ILO Sağlık ve Güvenlik İstatistikleri*. [Online]. Available: https://ilostat.ilo.org/topics/safety-and-health-at-work/.

[7] EUROSTAT, *European Statistics on Accidents at Work (ESAW) - Summary methodology*. Luxembourg, 2013, pp. 1–61, ISBN: 978-92-79-28419-9. DOI: 10.2785/40882.

[8] SGK, *SGK İstatistik Yıllıkları*. [Online]. Available: http://www.sgk.gov.tr/wps/portal/sgk/tr/kurumsal/istatistik/sgk_istatistik_yilliklari.

[9] M. Şen, S. Dursun, G. Murat, "Türkiye'de İş Kazaları: Avrupa Birliği Ülkeleri Bağlamında Bir Değerlendirme," *OPUS Uluslararası Toplum Araştırmaları Dergisi*, no. November, Dec. 2018, ISSN: 2528-9527. DOI: 10.26466/opus.463211.

[10] M. Baram, "Globalization and workplace hazards in developing nations," *Safety Science*, vol. 47, no. 6, pp. 756–766, Jul. 2009, ISSN: 09257535. DOI: 10.1016/j.ssci.2008.01.008.

[11] G. E. Gürcanli, U. Müngen, "An occupational safety risk analysis method at construction sites using fuzzy sets," *International Journal of Industrial Ergonomics*, vol. 39, no. 2, pp. 371–387, 2009, ISSN: 01698141. DOI: 10.1016/j.ergon.2008.10.006.

[12] G. E. Gürcanli, U. Müngen, "Analysis of construction accidents in Turkey and responsible parties," *Industrial Health*, vol. 51, no. 6, pp. 581–595, 2013, ISSN: 00198366. DOI: 10.2486/indhealth.2012-0139.

[13] G. E. Gurcanli, S. Bilir, M. Sevim, "Activity based risk assessment and safety cost estimation for residential building construction projects," *Safety Science*, vol. 80, no. 112, pp. 1–12, 2015, ISSN: 18791042. DOI: 10.1016/j.ssci.2015.07.002. [Online]. Available: http://dx.doi.org/10.1016/j.ssci.2015.07.002.

[14] S. Bilir, G. E. Gürcanlı, "A Method For Determination of Accident Probability in Construction Industry," *Teknik Dergi*, no. 511, pp. 8537–8562, 2018, ISSN: 1300-3453. DOI: 10.18400/tekderg.363613.

[15] Ö. Akboğa Kale, S. Baradan, "Identifying Factors that Contribute to Severity of Construction Injuries using Logistic Regression Model," *Teknik Dergi*, vol. 31, no. 2, pp. 9919–9940, 2020. DOI: https://dx.doi.org/10.18400/tekderg.470633.

[16] I. M. Uzun, S. Cebi, "A novel approach for classification of occupational health and safety measures based on their effectiveness by using fuzzy kano model," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 1, pp. 589–600, 2020, ISSN: 10641246. DOI: 10.3233/jifs-179432.

[17] M. Sari, A. S. Selcuk, C. Karpuz, H. S. B. Duzgun, "Stochastic modeling of accident risks associated with an underground coal mine in Turkey," *Safety Science*, vol. 47, no. 1, pp. 78–87, 2009, ISSN: 09257535. DOI: 10.1016/j.ssci.2007.12.004. [Online]. Available: http://dx.doi.org/10.1016/j.ssci.2007.12.004.

[18] O. Karadeniz, "Dünya'da ve Türkiye'de İş Kazaları ve Meslek Hastalıkları ve Sosyal Koruma Yetersizliği," *Çalışma ve Toplum Dergisi*, vol. 0, no. 34, pp. 15–75, 2012.

[19] S. Sarkar, R. Raj, S. Vinay, J. Maiti, D. K. Pratihar, "An optimization-based decision tree approach for predicting slip-trip-fall accidents at work," *Safety Science*, vol. 118, pp. 57–69, Oct. 2019, ISSN: 09257535. DOI: 10.1016/j.ssci.2019.05.009.

[20] A. J. Tixier, M. R. Hallowell, B. Rajagopalan, D. Bowman, "Application of machine learning to construction injury prediction," *Automation in Construction*, vol. 69, pp. 102–114, Sep. 2016, ISSN: 09265805. DOI: 10.1016/j.autcon.2016.05.016.

[21] L. Huang, C. Wu, B. Wang, Q. Ouyang, "A New Paradigm for Accident Investigation and Analysis in the Era of Big Data," *Process Safety Progress*, vol. 37, no. 1, pp. 42–48, 2017, ISSN: 10668527. DOI: 10.1002/prs.

[22] G. Mistikoglu, I. H. Gerek, E. Erdis, P. E. Mumtaz Usmen, H. Cakan, E. E. Kazan, "Decision tree analysis of construction fall accidents involving roofers," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2256–2263, 2015, ISSN: 09574174. DOI: 10.1016/j.eswa.2014.10.009. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2014.10.009.

[23] M. Gul, A. F. Guneri, F. Yılmaz, "Analysis of the relation between the characteristics of workers and occupational accidents using data mining," *The Turkish Journal of Pediatrics Environmental Medicine and Safety*, vol. 1, no. 4, pp. 102–118, 2016. [Online]. Available: http://www.turkishjournalpediatrics.org/?fullTextId=418&lang=eng#r33.

[24] G. TUNA, "TÜRKİYEDE MEYDANA GELEN İŞ KAZALARININ VERİ MADENCİLİĞİ KÜMELEME MODELİNDE GRAF TEORİ İLE ANALİZİ," Ph.D. dissertation, Gazi Üniversitesi, 2017, pp. 148–162.

[25] A. M. Abubakar, H. Karadal, S. W. Bayighomog, E. Merdan, "Workplace injuries, safety climate and behaviors: application of an artificial neural network," *International Journal of Occupational Safety and Ergonomics*, vol. 26, no. 4, pp. 651–661, 2018, ISSN: 10803548. DOI: 10.1080/10803548.2018.1454635. [Online]. Available: https://doi.org/10.1080/10803548.2018.1454635.

[26] B. U. Ayhan, O. B. Tokdemir, "Safety assessment in megaprojects using artificial intelligence," *Safety Science*, vol. 118, pp. 273–287, Oct. 2019, ISSN: 09257535. DOI: 10.1016/j.ssci.2019.05.027.

[27] ——, "Predicting the outcome of construction incidents," *Safety Science*, vol. 113, pp. 91–104, Mar. 2019, ISSN: 09257535. DOI: 10.1016/j.ssci.2018.11.001.

[28] N. Köse, F. Ersöz, "Veri Madenciliğinde Karar Ağacı Algoritmaları İle Demir Çelik Endüstrisinde İş Kazaları Üzerine Bir Uygulama An Application on Decision Tree Algorithms in Data Mining and Occupational Accidents in the Iron and Steel Industry," no. November, pp. 397–407, 2020. DOI: 10.31590/ejosat.843387.

[29] A. O. AKAY, M. AKGÜL, A. İ. ESİN, M. DEMİR, N. ŞENTÜRK, T. ÖZTÜRK, "Evaluation of occupational accidents in forestry in Europe and Turkey by k-means clustering analysis," *Turkish Journal of Agriculture and Forestry*, vol. 45, no. 4, pp. 495–509, 2021, ISSN: 13036173. DOI: 10.3906/tar-2010-55.

[30] K. Koc, Ö. Ekmekcioğlu, A. P. Gurgun, "Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers," *Automation in Construction*, vol. 131, no. March, 2021, ISSN: 09265805. DOI: 10.1016/j.autcon.2021.103896.

[31] B. Depaire, G. Wets, K. Vanhoof, "Traffic accident segmentation by means of latent class clustering," *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1257–1266, Jul. 2008, ISSN: 00014575. DOI: 10.1016/j.aap.2008.01.007.

[32] M. Alikhani, A. Nedaie, A. Ahmadvand, "Presentation of clustering-classification heuristic method for improvement accuracy in classification of severity of road accidents in Iran," *Safety Science*, vol. 60, pp. 142–150, Dec. 2013, ISSN: 09257535. DOI: 10.1016/j.ssci.2013.06.008.

[33] S. Kaplan, C. G. Prato, "Cyclist–Motorist Crash Patterns in Denmark: A Latent Class Clustering Approach," *Traffic Injury Prevention*, vol. 14, no. 7, pp. 725–733, Oct. 2013, ISSN: 1538-9588. DOI: 10.1080/15389588.2012.759654.

[34] Q. Zeng, H. Huang, "A stable and optimized neural network model for crash injury severity prediction," *Accident Analysis and Prevention*, vol. 73, pp. 351–358, 2014, ISSN: 00014575. DOI: 10.1016/j.aap.2014.09.006. [Online]. Available: http://dx.doi.org/10.1016/j.aap.2014.09.006.

[35] K. Kusano, H. C. Gabler, "Comparison and Validation of Injury Risk Classifiers for Advanced Automated Crash Notification Systems," *Traffic Injury Prevention*, vol. 15, no. February 2015, S126–S133, 2014, ISSN: 1538957X. DOI: 10.1080/15389588.2014.927577.

[36] A. Iranitalab, A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Analysis and Prevention*, vol. 108, pp. 27–36, 2017, ISSN: 0001-4575. DOI: 10.1016/j.aap.2017.08.008.

[37] Y. Kang, A. Khattak, "Cluster-Based Approach to Analyzing Crash Injury Severity at Highway–Rail Grade Crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2608, no. 1, pp. 58–69, Jan. 2017, ISSN: 0361-1981. DOI: 10.3141/2608-07.

[38] S. Kumar, D. Toshniwal, M. Parida, "A comparative analysis of heterogeneity in road accident data using data mining techniques," *Evolving Systems*, vol. 8, no. 2, pp. 147–155, Jun. 2017, ISSN: 1868-6478. DOI: 10.1007/s12530-016-9165-5.

[39] G. Prati, M. De Angelis, V. Marín Puchades, F. Fraboni, L. Pietrantoni, "Characteristics of cyclist crashes in Italy using latent class analysis and association rule mining," *PLOS ONE*, vol. 12, no. 2, J. Xu, Ed., e0171484, Feb. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0171484.

[40] M. Sun, X. Sun, D. Shan, "Pedestrian crash analysis with latent class clustering method," *Accident Analysis & Prevention*, vol. 124, no. November 2018, pp. 50–57, Mar. 2019, ISSN: 00014575. DOI: 10.1016/j.aap.2018.12.016.

[41] H. Alawad, S. Kaewunruen, M. An, "Learning from Accidents: Machine Learning for Safety at Railway Stations," *IEEE Access*, vol. 8, pp. 633–648, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2962072.

[42] A. Maheshwari, N. Davendralingam, D. A. Delaurentis, "A comparative study of machine learning techniques for aviation applications," in *Aviation Technology, Integration, and Operations Conference*, Atlanta, Georgia: American Institute of Aeronautics and Astronautics, Inc., 2018, ISBN: 9781624105562. DOI: 10.2514/6.2018-3980.

[43] M. Bevilacqua, F. Ciarapica, G. Giacchetta, "Industrial and occupational ergonomics in the petrochemical process industry: A regression trees approach," *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1468–1479, Jul. 2008, ISSN: 00014575. DOI: 10.1016/j.aap.2008.03.012.

[44] C.-W. Cheng, S.-S. Leu, Y.-M. Cheng, T.-C. Wu, C.-C. Lin, "Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry," *Accident Analysis & Prevention*, vol. 48, pp. 214–222, Sep. 2012, ISSN: 00014575. DOI: 10.1016/j.aap.2011.04.014.

[45] A. Chokor, H. Naganathan, W. K. Chong, M. E. Asmar, "Analyzing Arizona OSHA Injury Reports Using Unsupervised Machine Learning," *Procedia Engineering*, vol. 145, pp. 1588–1593, 2016, ISSN: 18777058. DOI: 10.1016/j.proeng.2016.04.200.

[46] T. Rivas, M. Paz, J. Martín, J. Matías, J. García, J. Taboada, "Explaining and predicting workplace accidents using data-mining techniques," *Reliability Engineering & System Safety*, vol. 96, no. 7, pp. 739–747, Jul. 2011, ISSN: 09518320. DOI: 10.1016/j.ress.2011.03.006.

[47] K. Vallmuur, "Machine learning approaches to analysing textual injury surveillance data: A systematic review," *Accident Analysis and Prevention*, vol. 79, pp. 41–49, 2015, ISSN: 00014575. DOI: 10.1016/j.aap.2015.03.018. [Online]. Available: http://dx.doi.org/10.1016/j.aap.2015.03.018.

[48] F. Davoudi Kakhki, S. A. Freeman, G. A. Mosher, "Evaluating machine learning performance in predicting injury severity in agribusiness industries," *Safety Science*, vol. 117, pp. 257–262, Aug. 2019, ISSN: 09257535. DOI: 10.1016/j.ssci.2019.04.026.

[49] S. Sarkar, V. Pateshwari, J. Maiti, "Predictive model for incident occurrences in steel plant in India," in *8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017*, Institute of Electrical and Electronics Engineers Inc., 2017, ISBN: 9781509030385. DOI: 10.1109/ICCCNT.2017.8204077.

[50] B. U. Ayhan, O. B. Tokdemir, "Accident Analysis for Construction Safety Using Latent Class Clustering and Artificial Neural Networks," *Journal of Construction Engineering and Management*, vol. 146, no. 3, p. 04 019 114, Mar. 2020, ISSN: 0733-9364. DOI: 10.1061/(ASCE)CO.1943-7862.0001762.

[51] H. Chen, X. Luo, "Severity Prediction Models of Falling Risk for Workers at Height," *Procedia Engineering*, vol. 164, pp. 439–445, 2016, ISSN: 1877-7058. DOI: 10.1016/j.proeng.2016.11.642.

[52] J. Choi, B. Gu, S. Chin, J.-S. Lee, "Machine learning predictive model based on national data for fatal accidents of construction workers," *Automation in Construction*, vol. 110, no. 3, p. 102 974, Feb. 2020, ISSN: 09265805. DOI: 10.1016/j.autcon.2019.102974.

[53] J. M. Matías, T. Rivas, J. E. Martín, J. Taboada, "A machine learning methodology for the analysis of workplace accidents," *International Journal of Computer Mathematics*, vol. 85, no. 3-4, pp. 559–578, 2008, ISSN: 00207160. DOI: 10.1080/00207160701297346.

[54] K. Kang, H. Ryu, "Predicting types of occupational accidents at construction sites in Korea using random forest model," *Safety Science*, vol. 120, pp. 226–236, 2019. DOI: 10.1016/j.ssci.2019.06.034.

[55] S. Sarkar, S. Vinay, R. Raj, J. Maiti, P. Mitra, "Application of optimized machine learning techniques for prediction of occupational accidents," *Computers and Operations Research*, vol. 106, pp. 210–224, 2019, ISSN: 03050548. DOI: 10.1016/j.cor.2018.02.021.

[56] M. S. Hajakbari, B. Minaei-Bidgoli, "A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data," *Journal of Loss Prevention in the Process Industries*, vol. 32, pp. 443–453, Nov. 2014, ISSN: 09504230. DOI: 10.1016/j.jlp.2014.10.013.

[57] M. Amiri, A. Ardeshir, M. H. Fazel Zarandi, E. Soltanaghaei, "Pattern extraction for high-risk accidents in the construction industry: a data-mining approach," *International Journal of Injury Control and Safety Promotion*, vol. 23, no. 3, pp. 264–276, Jul. 2016, ISSN: 1745-7300. DOI: 10.1080/17457300.2015.1032979.

[58] B. Altunkaynak, "A statistical study of occupational accidents in the manufacturing industry in Turkey," *International Journal of Industrial Ergonomics*, vol. 66, pp. 101–109, Jul. 2018, ISSN: 01698141. DOI: 10.1016/j.ergon.2018.02.012.

[59] F. Zhang, "A hybrid structured deep neural network with Word2Vec for construction accident causes classification," *International Journal of Construction Management*, vol. 0, no. 0, pp. 1–21, 2019, ISSN: 15623599. DOI: 10.1080/15623599.2019.1683692. [Online]. Available: https://doi.org/10.1080/15623599.2019.1683692.

[60] A. Yedla, F. D. Kakhki, A. Jannesari, "Predictive modeling for occupational safety outcomes and days away from work analysis in mining operations," *International Journal of Environmental Research and Public Health*, vol. 17, no. 19, pp. 1–17, 2020, ISSN: 16604601. DOI: 10.3390/ijerph17197054.

[61] D. Binu, B. R. Rajakumar, *Artificial Intelligence in Data Mining: Theories and Applications*. Academic Press, 2021, ISBN: 978-0-12-820601-0.

[62] J. Brownlee, "Master Machine Learning Algorithms," 2017.

[63] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2013, ISBN: 978-1-4614-7138-7. DOI: 10.1007/978-1-4614-7138-7.

[64] F. Palamara, F. Piglione, N. Piccinini, "Self-Organizing Map and clustering algorithms for the analysis of occupational accident databases," *Safety Science*, vol. 49, no. 8-9, pp. 1215–1230, Oct. 2011, ISSN: 09257535. DOI: 10.1016/j.ssci.2011.04.003.

[65] E. Farina, S. Bianco, A. Bena, O. Pasqualini, "Finding causation in occupational fatalities: A latent class analysis," *American Journal of Industrial Medicine*, vol. 62, no. 2, pp. 123–130, 2019, ISSN: 10970274. DOI: 10.1002/ajim.22936.

[66] C. W. Liao, Y. H. Perng, "Data mining for occupational injuries in the Taiwan construction industry," *Safety Science*, vol. 46, no. 7, pp. 1091–1102, 2008, ISSN: 09257535. DOI: 10.1016/j.ssci.2007.04.007.

[67] C. W. Cheng, C. C. Lin, S. S. Leu, "Use of association rules to explore cause-effect relationships in occupational accidents in the Taiwan construction industry," *Safety Science*, vol. 48, no. 4, pp. 436–444, 2010, ISSN: 09257535. DOI: 10.1016/j.ssci.2009.12.005. [Online]. Available: http://dx.doi.org/10.1016/j.ssci.2009.12.005.

[68] L. Comberti, G. Baldissone, M. Demichela, "Workplace accidents analysis with a coupled clustering methods: S.O.M. and K-means algorithms," *Chemical Engineering Transactions*, vol. 43, pp. 1261–1266, 2015, ISSN: 22839216. DOI: 10.3303/CET1543211.

[69] M. Lombardi, G. Rossi, "Cluster analysis of fatal accidents series in the Infor.Mo database: Analysis, evidence and research perspectives," *International Journal of Safety and Security Engineering*, vol. 3, no. 4, pp. 318–332, 2013, ISSN: 2041904X. DOI: 10.2495/SAFE-V3-N4-318-332.

[70] N. Nenonen, "Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database," *Applied Ergonomics*, vol. 44, no. 2, pp. 215–224, 2013. DOI: 10.1016/j.apergo.2012.07.001. [Online]. Available: http://dx.doi.org/10.1016/j.apergo.2012.07.001.

[71] R. Ghousi, "Applying a decision support system for accident analysis by using data mining approach: A case study on one of the Iranian manufactures," *Journal of Industrial and Systems Engineering*, vol. 8, no. 3, pp. 60–76, 2015.

[72] S. Murè, L. Comberti, M. Demichela, "How harsh work environments affect the occupational accident phenomenology? Risk assessment and decision making optimisation," *Safety Science*, vol. 95, pp. 159–170, 2017, ISSN: 18791042. DOI: 10.1016/j.ssci.2017.01.004.

[73] R. Moura, M. Beer, E. Patelli, J. Lewis, F. Knoll, "Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies," *Safety Science*, vol. 99, pp. 196–214, 2017, ISSN: 18791042. DOI: 10.1016/j.ssci.2017.05.001. [Online]. Available: http://dx.doi.org/10.1016/j.ssci.2017.05.001.

[74] ——, "Learning from major accidents to improve system design," *Safety Science*, vol. 84, pp. 37–45, 2016, ISSN: 18791042. DOI: 10.1016/j.ssci.2015.11.022. [Online]. Available: http://dx.doi.org/10.1016/j.ssci.2015.11.022.

[75] R. Moura, M. Beer, E. Patelli, J. Lewis, "Learning from major accidents: Graphical representation and analysis of multi-attribute events to enhance risk communication," *Safety Science*, vol. 99, pp. 58–70, 2017, ISSN: 18791042. DOI: 10.1016/j.ssci.2017.03.005. [Online]. Available: http://dx.doi.org/10.1016/j.ssci.2017.03.005.

[76] L. Comberti, M. Demichela, G. Baldissone, "A combined approach for the analysis of large occupational accident databases to support accident-prevention decision making," *Safety Science*, vol. 106, pp. 191–202, Jul. 2018, ISSN: 09257535. DOI: 10.1016/j.ssci.2018.03.014.

[77] L. Comberti, M. Demichela, G. Baldissone, R. Luzzi, G. Fois, "The impact of regulation on occupational safety - A regional study for Italian food industry," *Chemical Engineering Transactions*, vol. 67, no. 1989, pp. 397–402, 2018, ISSN: 22839216. DOI: 10.3303/CET1867067.

[78] L. Comberti, M. Demichela, G. Baldissone, G. Fois, R. Luzzi, "Large occupational accidents data analysis with a coupled unsupervised algorithm: The S.O.M. k-means method. An application to the wood industry," *Safety*, vol. 4, no. 4, 2018, ISSN: 2313576X. DOI: 10.3390/safety4040051.

[79] Y. H. Chiang, F. K. W. Wong, S. Liang, "Fatal Construction Accidents in Hong Kong," *Journal of Construction Engineering and Management*, vol. 144, no. 3, 2018, ISSN: 07339364. DOI: 10.1061/(ASCE)CO.1943-7862.0001433.

[80] D. P. Shin, Y. J. Park, J. Seo, D. E. Lee, "Association Rules Mined from Construction Accident Data," *KSCE Journal of Civil Engineering*, vol. 22, no. 4, pp. 1027–1039, 2017, ISSN: 19763808. DOI: 10.1007/s12205-017-0537-6.

[81] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015, ISBN: 978-3-319-14141-1. DOI: 10.1007/978-3-319-14142-8.

[82] E. R. Buhi, P. Goodson, T. B. Neilands, "Out of sight, not out of mind: Strategies for handling missing data," *American Journal of Health Behavior*, vol. 32, no. 1, pp. 83–92, 2008, ISSN: 10873244. DOI: 10.5993/AJHB.32.1.8.

[83] D. A. Bennett, "How can I deal with missing data in my study?" *Australian and New Zealand Journal of Public Health*, vol. 25, no. 5, pp. 464–469, 2001, ISSN: 13260200. DOI: 10.1111/j.1467-842X.2001.tb00294.x.

[84] A. Agresti, *An introduction to categorical Analysis*. John Wiley & Sons, Inc., 2019, p. 375, ISBN: 9781119405276.

[85] S. van Buuren, K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–67, 2011, ISSN: 15487660. DOI: 10.18637/jss.v045.i03.

[86] I. Guyon, A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, no. 3, pp. 1157–1182, 2003. DOI: 10.1162/153244303322753616.

[87] F. Degenhardt, S. Seifert, S. Szymczak, "Evaluation of variable selection methods for random forests and omics data sets," *Briefings in Bioinformatics*, vol. 20, no. 2, pp. 492–503, 2019, ISSN: 14774054. DOI: 10.1093/bib/bbx124.

[88] M. B. Kursa, W. R. Rudnicki, "Feature Selection with the Boruta Package," *Journal of Statistical Software*, vol. 36, no. 11, pp. 1–13, 2010, ISSN: 1548-7660. DOI: 10.18637/jss.v036.i11.

[89] B. Boehmke, B. Greenwell, *The R Series Hands-on Machine Learning with R*. Taylor & Francis, 2020, ISBN: 9781138495685.

[90] L. Breiman, "Arcing the edge," *Statistics*, vol. 4, pp. 1–14, 1997.

[91] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[92] J. Friedman, T. Hastie, R. Tibshirani, *The elements of statistical learning Data mining, inference, and prediction*. Springer, 2009, ISBN: 0387848576,9780387848570,9780387848587.

[93] J. H. Friedman, "Stochastic Gradient Boosting," 1999.

[94] B. Greenwell, B. Boehmke, J. Cunningham, "Package 'gbm' - Generalized Boosted Regression Models," *CRAN Repository*, p. 39, 2020. [Online]. Available: https://cran.r-project.org/web/packages/gbm/gbm.pdf%0Ahttps://github.com/gbm-developers/gbm.

[95] J. Czakon, *F1 Score vs ROC AUC vs Accuracy vs PR AUC: Which Evaluation Metric Should You Choose?* 2021. [Online]. Available: https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc.

[96] D. Berrar, "Performance measures for binary classification," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1-3, no. 1, pp. 546–560, 2018. DOI: 10.1016/B978-0-12-809633-8.20351-8.

[97] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, no. Sept. 28, pp. 321–357, Jun. 2002, ISSN: 1076-9757. DOI: 10.1613/jair.953.

[98] R. Agrawal, T. Imielinski, A. Swami, "Mining Association in Large Databases," *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, pp. 207–216, 1993, ISSN: 01635808.

[99] B. Lantz, *Machine learning with R, 2nd edition*. 2015, ISBN: 9781784393908.

[100] S. Brin, R. Motwani, J. D. Ullman, S. Tsur, "Dynamic Itemset Counting and Implication Rules for Market Basket Data," *SIGMOD Record (ACM Special Interest Group on Management of Data)*, vol. 26, no. 2, pp. 255–264, 1997, ISSN: 01635808. DOI: 10.1145/253262.253325.

[101] N. Japkowicz, "The Class Imbalance Problem: Significance and Strategies," *Proceedings of the 2000 International Conference on Artificial Intelligence*, pp. 111–117, 2000.

[102] D. J. Dittman, T. M. Khoshgoftaar, R. Wald, A. Napolitano, "Comparison of data sampling approaches for imbalanced bioinformatics data," *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014*, pp. 268–271, 2014.

[103] J. H. Kim, "Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap," *Computational Statistics and Data Analysis*, vol. 53, no. 11, pp. 3735–3745, 2009, ISSN: 01679473. DOI: 10.1016/j.csda.2009.04.009. [Online]. Available: http://dx.doi.org/10.1016/j.csda.2009.04.009.

[104] T. Cumhuriyeti, *5510 Sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu*, 2006. [Online]. Available: https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=5510&MevzuatTur=1&MevzuatTertip=5.

[105] S. G. Kurumu, *İş Kazasi ve Meslek Hastaliği Bildirim Formu Kullanım Kılavuzu*, 2012. [Online]. Available: http://e.sgk.gov.tr/wps/wcm/connect/13cbe3f9-f41c-4728-bb17-b1116b00e2f4/Kullanim+Kilavuzu.pdf?MOD=AJPERES&CACHEID=13cbe3f9-f41c-4728-bb17-b1116b00e2f4.

[106] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008, ISSN: 15487660. DOI: 10.18637/jss.v028.i05.

[107] M. Kuhn *et al.*, "Caret: Classification and Regression Training," 2021.

[108] C. Jacinto, C. G. Soares, "The added value of the new ESAW/Eurostat variables in accident analysis in the mining and quarrying industry," *Journal of Safety Research*, vol. 39, no. 6, pp. 631–644, 2008, ISSN: 00224375. DOI: 10.1016/j.jsr.2008.10.009. [Online]. Available: http://dx.doi.org/10.1016/j.jsr.2008.10.009.

A
DEĞİŞKEN KATEGORİLERİ

SGK veritabanından elde edilen kaza verisinde yer alan kategorik değişkenlerin seviyelerinin kodları ve karşılık gelen açıklamaları bu bölümdeki yer alan tablolarda verilmektedir.

Tablo A.1 NACE_K

Kod	Açıklama
1	01-Bitkisel ve hayvansal üretim ile avcılık ve ilgili hizmet faaliyetleri
2	02-Ormancılık ve tomrukçuluk
3	03-Balıkçılık ve su ürünleri yetiştiriciliği
5	05-Kömür ve Linyit Çıkartılması
6	06-Ham Petrol ve Doğalgaz çıkarımı
7	07-Metal Cevheri Madenciliği
8	08-Diğer Madencilik ve Taşocakçılığı
9	09-Madenciliği destekleyici hizmet faaliyetleri
10	10-Gıda ürünlerinin imalatı
11	11-İçeceklerin imalatı
12	12-Tütün ürünleri imalatı
13	13-Tekstil ürünlerinin imalatı
14	14-Giyim eşyalarının imalatı
15	15-Deri ve ilgili ürünlerin imalatı
16	16-Ağaç. ağaç ürünleri ve mantar ürünleri imalatı (mobilya hariç), saz. saman ve benzeri malzemelerden örülerek eşyaların imalatı
17	17-Kağıt ve kağıt ürünlerinin imalatı
18	18-Kayıtlı medyanın basılması ve çoğaltılması
19	19-Kok kömürü ve rafine edilmiş petrol ürünleri imalatı
20	20-Kimyasalların ve kimyasal ürünlerin imalatı
21	21-Temel eczacılık ürünlerinin ve eczacılığa ilişkin malzemelerin imalatı
22	22-Kauçuk ve plastik ürünlerin imalatı
23	23-Diğer metalik olmayan mineral ürünlerin imalatı
24	24-Ana metal sanayii
25	25-Makine ve teçhizat hariç. fabrikasyon metal ürünleri imalatı
26	26-Bilgisayarların. elektronik ve optik ürünlerin imalatı
27	27-Elektrikli teçhizat imalatı
28	28-Başka yerde sınıflandırılmamış makine ve ekipman imalatı

Tablo A.1 NACE_K - (devamı)

Kod	Açıklama
29	29-Motorlu kara taşıtı. treyler (römork) ve yarı treyler (yarı römork) imalatı
30	30-Diğer ulaşım araçlarının imalatı
31	31-Mobilya imalatı
32	32-Diğer imalatlar
33	33-Makine ve ekipmanların kurulumu ve onarımı
35	35-Elektrik. gaz. buhar ve havalandırma sistemi üretim ve dağıtımı
36	36-Suyun toplanması. arıtılması ve dağıtılması
37	37-Kanalizasyon
38	38-Atığın toplanması. ıslahı ve bertarafı faaliyetleri, maddelerin geri kazanımı
39	39-İyileştirme faaliyetleri ve diğer atık yönetimi hizmetleri
41	41-Bina inşaatı
42	42-Bina dışı yapıların inşaatı
43	43-Özel inşaat faaliyetleri
45	45-Toptan ve perakende ticaret ve motorlu kara taşıtlarının ve motosikletlerin onarımı
46	46-Toptan ticaret. motorlu kara taşıtları ve motosikletler hariç
47	47-Perakende ticaret (Motorlu kara taşıtları ve motosikletler hariç)
49	49-Kara taşımacılığı ve boru hattı taşımacılığı
50	50-Su yolu taşımacılığı
51	51-Havayolu taşımacılığı
52	52-Taşımacılık için depolama ve destekleyici faaliyetler
53	53-Posta ve kurye faaliyetleri
55	55-Konaklama
56	56-Yiyecek ve içecek hizmeti faaliyetleri
58	58-Yayımcılık faaliyetleri
59	59-Sinema filmi. video ve televizyon programları yapımcılığı. ses kaydı ve müzik yayımlama faaliyetleri
60	60-Programcılık ve yayıncılık faaliyetleri
61	61-Telekomünikasyon
62	62-Bilgisayar programlama. danışmanlık ve ilgili faaliyetler
63	63-Bilgi hizmet faaliyetleri
64	64-Sigorta ve emeklilik fonları hariç finansal hizmet faaliyetleri
65	65-Zorunlu sosyal güvenlik hariç. sigorta. reasürans ve emeklilik fonları
66	66-Finansal hizmetler ile sigorta faaliyetleri için yardımcı faaliyetler
68	68-Gayrimenkul faaliyetleri
69	69-Hukuki ve muhasebe faaliyetleri
70	70-İdare merkezi faaliyetleri, idari danışmanlık faaliyetleri
71	71-Mimarlık ve mühendislik faaliyetleri, teknik muayene ve analiz
72	72-Bilimsel araştırma ve geliştirme faaliyetleri
73	73-Reklamcılık ve pazar araştırması
74	74-Diğer mesleki. bilimsel ve teknik faaliyetler
75	75-Veterinerlik hizmetleri
77	77-Kiralama ve leasing faaliyetleri
78	78-İstihdam faaliyetleri

Tablo A.1 NACE_K - (devamı)

Kod	Açıklama
79	79-Seyahat acentesi. tur operatörü ve diğer rezervasyon hizmetleri ile ilgili faaliyetler
80	80-Güvenlik ve soruşturma faaliyetleri
81	81-Binalar ve çevre düzenlemesi faaliyetleri
82	82-Büro yönetimi. büro desteği ve iş destek faaliyetleri
84	84-Kamu yönetimi ve savunma; zorunlu sosyal güvenlik
85	85-Eğitim
86	86-İnsan sağlığı hizmetleri
87	87-Yatılı bakım faaliyetleri
88	88-Barınacak yer sağlanmaksızın verilen sosyal hizmetler
90	90-Yaratıcı sanatlar. gösteri sanatları ve eğlence faaliyetleri
91	91-Kütüphaneler. arşivler. müzeler ve diğer kültürel faaliyetler
92	92-Kumar ve müşterek bahis faaliyetleri
93	93-Spor faaliyetleri. eğlence ve dinlence faaliyetleri
94	94-Üye olunan kuruluşların faaliyetleri
95	95-Bilgisayarların ve kişisel ve ev eşyalarının onarımı
96	96-Diğer hizmet faaliyetleri
97	97-Ev içi çalışan personelin işverenleri olarak hane halklarının faaliyetleri
98	98-Hane halkları tarafından kendi kullanımlarına yönelik olarak üretilen ayrım yapılmamış mal ve hizmetler
99	99-Uluslar arası örgütler ve temsilciliklerinin faaliyetleri

Tablo A.2 KAZA_IL_K

Kod	Açıklama
1	ADANA
2	ADIYAMAN
3	AFYON
4	AGRI
5	AMASYA
6	ANKARA
7	ANTALYA
8	ARTVIN
9	AYDIN
10	BALIKESIR
11	BILECIK
12	BINGÖL
13	BITLIS
14	BOLU
15	BURDUR
16	BURSA
17	ÇANAKKALE
18	ÇANKIRI
19	ÇORUM
20	DENIZLI
21	DIYARBAKIR

Tablo A.2 KAZA_IL_K - (devamı)

Kod	Açıklama
22	EDIRNE
23	ELAZIG
24	ERZINCAN
25	ERZURUM
26	ESKISEHIR
27	GAZIANTEP
28	GIRESUN
29	GÜMÜSHANE
30	HAKKARI
31	HATAY
32	ISPARTA
33	MERSIN
34	ISTANBUL
35	IZMIR
36	KARS
37	KASTAMONU
38	KAYSERI
39	KIRKLARELI
40	KIRSEHIR
41	KOCAELI
42	KONYA
43	KÜTAHYA
44	MALATYA
45	MANISA
46	KAHRAMANMARAS
47	MARDIN
48	MUGLA
49	MUS
50	NEVSEHIR
51	NIGDE
52	ORDU
53	RIZE
54	SAKARYA
55	SAMSUN
56	SIIRT
57	SINOP
58	SIVAS
59	TEKIRDAG
60	TOKAT
61	TRABZON
62	TUNCELI
63	SANLIURFA
64	USAK
65	VAN
66	YOZGAT
67	ZONGULDAK

Tablo A.2 KAZA_IL_K - (devamı)

Kod	Açıklama
68	AKSARAY
69	BAYBURT
70	KARAMAN
71	BATMAN
71	KIRIKKALE
73	SIRNAK
74	BARTIN
75	ARDAHAN
76	IGDIR
77	YALOVA
78	KARABüK
79	KILIS
80	OSMANIYE
81	DÜZCE

Tablo A.3 CC_K

Kod	Açıklama
1	11 Üretim alani, fabrika, atölye
2	12 Bakim alani, onarim atölyesi
3	13 Temelde depo, yükleme, bosaltma için kullanilan alan
4	19 Yukarida listelenmemis diger baska 010 tür çalisilan çevre
5	21 Insaat mevkii - yeni insa edilen bina
6	22 Insaat mevkii - yikilan, onarilan, bakimi yapilan bina
7	23 Açik hava tasocagi, açik hava madeni, kazi, çukur (açik hava madenciligi ve çalisan tas ocaklari dahil)
8	24 Insaat mevkii - yeralti
9	25 nsaat mevkii - su üstü veya üzerinde
10	26 Insaat mevkii - yüksek basinçli ortamda
11	29 Yukarida listelenmemis diger baska 020 tür çalisilan çevre
12	31 Yetistirme alani
13	32 Çiftlik alani - toprak ürünleri
14	33 Çiftlik alani - agaç veya çali ürünleri
15	34 Ormancilik alani
16	35 Balik çiftçiligi alani, balikçilik, deniz ürünleri (gemi veya tekne üzerinden olmayan)
17	36 Bahçe, park, botanik bahçe, hayvanat bahçesi
18	39 Yukarida listelenmemis diger baska 030 tür çalisilan çevre
19	41 Büro, toplanti salonu, kütüphane, vb.
20	42 Ögretim Kurumu, okul, lise, yüksek okul, üniversite, kres, anaokulu
21	43 Küçük veya büyük satis alani (sokak satislari dahil)
22	44 Lokanta, dinlenme alani, geçici konaklama (müze, oditoryum, stadyum, fuar, vb. dahil)
23	49 Yukarida listelenmemis diger baska 040 tür çalisilan çevre
24	51 Saglik Kurumu, özel hastane, hastane, bakim evi

Tablo A.3 CC_K - (devamı)

Kod	Açıklama
25	59 Yukarida listelenmemis diger baska 050 tür çalisilan çevre
26	61 Sürekli olarak kamu geçisine açik alan (karayollari, yan yollar, park alanlari, istasyon veya havaalani bekleme salonlari, vb.)
27	62 Ulasim araçlari - kara veya demiryolu ile- özel veya kamu (her tür: tren,otobüs, araba, vb.)
28	63 Kamu alanlarina bagli alanlar olmakla birlikte sadece yetkili kimselerin erisimine izin verilen alanlar: demiryolu hatti, havaalani pisti, karayolu banketi
29	69 Yukarida listelenmemis diger baska 060 tür-çalisilan çevre
30	71 Kisisel ev
31	72 Bir binanin ortak alani, ekleri, kisisel aile bahçesi
32	79 Yukarida listelenmemis diger baska 070 tür-çalisilan çevre
33	81 Kapali spor alani - spor salonu, jimnastik salonu, kapali yüzme havuzu
34	82 Açik spor alani - spor sahasi, açik yüzme havuzu, kayak parkuru
35	89 Yukarida listelenmemis diger baska 080 tür çalisilan çevre
36	91 Yükseltme esnasinda - sabit bir düzeyde (çati, teras, vb.)
37	92 Yükseltme esnasinda-direk, pilon, asili platform
38	93 Havada - uçakta
39	99 Yukarida listelenmemis baska 090 türü Çalisilan çevre, insaat santiyesi disinda
40	101 Yeraltinda - tünel (yol, ten, tüp)
41	102 Yeraltinda - maden
42	102 Yeraltinda - maden
43	109 Yukarida listelenmemis diger baska 100 tür çalisilan çevre, insaat mevkii disinda
44	111 Deniz veya okyanus - her tür tekne, platform, gemi, sandal, mavnada
45	112 Göl, nehir, liman - her tür tekne, platform, gemi, sandal, mavnada
46	119 Yukarida listelenmemis diger baska 110 tür çalisilan çevre, insaat mevkii disinda
47	121 Yüksek basinç ortamlarinda - sualti (örnegin, dalma)
48	122 Yüksek basinç ortamlarinda - kamara
49	129 Yukarida listelenmemis baska 120 türü çalisilan çevre, insaat santiyesi disinda
NA	0 Bilgi Yok
NA	999 Siniflandirmada listelenmemis baska çalisma ortamlari

Tablo A.4 SG_KAZAGERYER_K

Kod	Açıklama
1	100 Isyerinde
2	200 Isyeri Disinda
NA	0 Bilgi Yok

Tablo A.5 KAZA_KS_K

Kod	Açıklama
1	Elektrik sorunları, patlama, yangın nedeniyle sapma
	11 Elektrik arizasi nedeniyle elektrik sorunu - dogrudan temasa yol açan
	12 Elektrik sorunu - dogrudan temasa yol açan
	13 Patlama
	14 Yangin, tutusma
	19 Yukarida listelenmemis baska 10 çesit Sapma
, 2	Taşma, devrilme, sızma, buharlaşma, emisyon sapması
	21 Kati durumunda - tasma, devrilme
	22 Sivi durumunda - sizma, kaçirma, akma, siçrama, püskürme
	23 Gaz durumunda - buharlasma, aerosol olusum, gaz olusumu
	24 Toz halindeki madde - duman olusumu, havadaki/yayilmis toz ve zerrecikleri
	29 Yukarida listelenmemis baska 20 çesit Sapma
3	Maddi aracın kırılma, patlama, ayrılma, kayma, düşme, çokmesi
	31 Maddenin kirilmasi - eklemlerde, birlesme noktalarinda
	32 Kirilma, patlama - kiymik olusumu (tahta, cam, metal, tas, plastik, digerleri)
	33 Maddi aracin kayma, düsme, çökmesi - yukaridan (kazazedenin üstüne düserek)
	34 Maddi aracin kayma, düsme, çökmesi - asagidan (kazazedeyi asagi çekerek)
	35 Maddi aracin kayma, düsme, çökmesi - ayni düzeyde
	39 Yukarida listelenmemis baska 30 türü-çesit Sapma
4	Bir makinenin, taşıma aracının veya işleme ekipmanının, elle kullanılan alet, nesne, hayvanın denetimden çıkması
	41 Denetim kaybi (tam veya kismi) - makine (istenmeyen baslama da dahil olmak üzere) veya maddi araç
	42 Denetim kaybi (tam veya kismi) - tasit araci veya ekipman kullaniminin (motorlu olsun veya olmasin)
	43 Denetim kaybi (tam veya kismi) - el aleti (motorlu olsun veya olmasin) veya alet tarafindan kullanilan maddi aracin
	44 Denetim kaybi (tam veya kismi) - nesnenin (tasinan, oynatilan, kullanilan, vb.)
	45 Denetim kaybi (tam veya kismi) - hayvanin
	49 Yukarida listelenmemis baska 40 çesit Sapma
5	Kayma veya tökezleme - düşme, kişilerin düşmesi
	51 Kisinin düsmesi - alt düzeye
	52 Kayma - tökezleme ve düsme - Kisinin düsmesi - ayni düzeyde
	59 Yukarida listelenmemis baska 50 çesit Sapma
6	Fiziki baskı olmadan beden hareketi
	61 Keskin bir nesne üzerinde yürüme
	62 Diz çökme, oturma, yaslanma
	63 Kapilma veya götürülme - bir sey veya ivme tarafindan
	64 Isbirligi olmaksizin yapilan hareket, gereksiz veya zamansiz eylemler
	69 Yukarida listelenmemis baska 60 çesit Sapma

Tablo A.5 KAZA_KS_K - (devamı)

Kod	Açıklama
	Fiziki baskıyla veya fiziki baskı altında beden hareketi
	71 Kaldirma, tasima, ayakta durma
	72 Itme, çekme
	73 Asagi birakma, egilme
	74 Bükülme, dönme
	75 Çignenmek, bacak veya bilek bükülmesi, düsmeden kaymak
	79 Yukarida listelenmemis baska 70 çesit Sapma
	Şok, korku, vahşet, saldırganlık, tehdit, bulunma
	81 Sok, korku
	82 Vahset, saldiri, tehdit - isverenin otoritesine tabi olan sirket çalisanlari arasinda
	83 Vahset, saldiri, tehdit - islerini yapmakta olan kazazedeye yönelik sirket disindaki kisiler tarafindan (banka soygunu, otobüs soförleri, vb.)
	84 Saldiri, dürtülme - hayvan tarafindan
	85 Üçüncü bir kimsenin veya kisinin kendisinin, kendine ve baskasina da tehlike yaratmasi
	89 Yukarida listelenmemis baska 80 türü çesit Sapma
NA	Bilgi yok
NA	99 Bu siniflandirmada listelenmemis baska Sapma

Tablo A.6 GENELFAA_K

Kod	Açıklama
1	11 Üretim, imalat, isleme - tüm türler
2	12 Depolama - tüm türler
3	19 Yukarida listelenmemis diger baska 10 tür Kaza aninda kazazedenin yürütmekte oldugu genel faaliyet
4	21 Kazi
5	22 Yeni insaat - bina
6	23 Yeni insaat - insaat mühendisligi, altyapi, yol, köprü, baraj ve limanlar
7	24 Yeniden modelleme, onarim, genisletme, bina bakimi - her tür insaat
8	25 Yıkım - her tür inşaat
9	29 Yukarida listelenmemis diger baska 20 çesit kaza aninda kazazedenin yaptigi faaliyet
10	31 Tarimla ilgili meslek türü - topragin islenmesi
11	32 Tarimla ilgili meslek türü - sebzelerle, bahçecilik
12	33 Tarimla ilgili meslek türü - canli hayvanlarla
13	34 Ormancilikla ilgili meslek türü
14	35 Balik çiftçiligi, balikçilik
15	39 Yukarida listelenmemis diger baska 30 çesit kaza aninda kazazedenin yaptigi faaliyet
16	41 Kamu hizmeti, bakim, yardim, servis
17	42 Zihinsel çalisma - ögretmenlik, egitim, bilgi islem, büro isi, organizasyon, yönetim
18	43 Ticari faaliyet - alim, satim ve ilgili hizmetler

Tablo A.6 GENELFAA_K - (devamı)

Kod	Açıklama
19	49 Yukarida listelenmemis baska 40 çesit kaza aninda kazazedenin yaptigi faaliyet
20	51 Kurulum, hazirlama, montaj, yükleme, sökme, parçalama
21	52 Bakim, onarim, ayar, akort
22	53 Is alaninin temizlenmesi, makine sanayi veya elle
23	54 Atik yönetimi, kontrol, her tür atik isleme
24	55 Izleme, donanimlari olsun veya olmasin imalat süreçlerinin, çalisma alanlarinin, tasit araçlarinin izlenmesi, teftisi
25	59 Yukarida listelenmemis baska 50 çesit kaza aninda kazazedenin yaptigi faaliyet
26	61 Tasit araçlariyla olani da dahil olmak üzere hareket ;
27	62 Spor, sanatsal faaliyet
28	69 Yukarida listelenmemis baska 60 çesit kaza aninda kazazedenin yaptigi faaliyet
NA	
NA	0 Bilgi yok
NA	99 Siniflandirmada listelenmemis baska kaza aninda kazazedenin yaptigi faaliyet

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K

Kod	Açıklama
1	Binalar, yapılar, yüzeyler - yer düzeyinde (kapalı veya açık, sabit veya seyyae, geçici veya kalıcı) 01.01 Bina parçalari, yapisal parçalar- kapi, duvar, bölme, vb. ve kasitli engeller (pencere, vb.) 01.02 Yer düzeyindeki yüzeyler - yer ve tabanlar (kapali veya açik, çiftlik topragi, spor alani, kaygan zemin, daginik zemin, çivili kereste) 01.03 Yer düzeyindeki yüzeyler -hareketli 01.99 01 numarali grup içinde, yukarida listelenmemis baska bilinen bina, yapi ve yüzeyler- ayni düzeyde
2	Binalar, yapılar, yüzeyler - yer düzeyinin üstünde (kapalı veya açık) 02.01 Yer düzeyinden yukaridaki bina parçalari - sabit (çati, teras, kapi ve pencereler, merdiven, iskele) 02.02 Yer düzeyinden yukaridaki yapilar, yüzeyler - sabit (geçit, sabit merdiven, pilonlar da dahil olmak üzere) 02.03 Yer düzeyinden yukaridaki yapilar, yüzeyler - seyyar (yapi iskelesi, seyyar merdiven, besik, yükseltme platformlari da dahil olmak üzere) 02.04 Yer düzeyinden yukaridaki yapilar, yüzeyler - geçici (geçici yapi iskelesi, baglar, salincaklar da dahil olmak üzere) 02.05 Yer düzeyinden yukaridaki yapilar, yüzeyler - yüzer (arama platformlari, mavna iskeleleri de dahil olmak üzere) 02.99 02 numarali grup içinde yukarida listelenmemis baska bilinen yer düzeyinden yukaridaki bina, yapi, yüzeyler
3	Binalar, yapılar, yüzeyler - yer düzeyinin altında (kapalı veya açık)

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K - (devamı)

Kod	Açıklama
	03.01 Hafriyat, hendek, kuyu, ocak, egik yüzey, dehliz
	03.02 Yeralti alanlar, tüneller
	03.03 Sualti ortamlar
	03.99 03 numarali grup içinde yukarida listelenmemis baska bilinen yer düzeyinin altindaki bina, yapi, yüzeyler
4	Malzeme temini ve dağıtımı sistemleri boru hatları
	04.01 Gaz, hava, sivi, kati malzeme temini ve dagitimi sistemleri, boru aglari - sabit-aksakliklar da dahil olmak üzere
	04.02 Malzeme temini ve dagitimi sistemleri, boru aglari - seyyar
	04.03 Kanalizasyon, lagim
	04.99 04 numarali grup içinde yukarida listelenmemis baska bilinen malzeme temini ve dagitim sistemleri, boru aglari
	Motorlar, enerji iletimi ve depolama sistemleri
5	05.01 Motorlar, jeneratörler (termal, elektrik, radyasyon)
	05.02 Enerji iletim ve depolama sistemleri (mekanik, basinçli, hidrolik, elektrik, pil ve akümülatörler de dahil olmak üzere)
	05.99 05 numarali grup içinde yukarida listelenmemis baska bilinen motor, enerji iletimi ve depolama sistemleri
6	El aletleri, motorsuz
	06.01 El aletleri, motorsuz - biçkicilik için
	06.02 El aletleri, motorsuz - kesmek, ayirmak için (makas, budama makaslari da dahil olmak üzere)
	06.03 El aletleri, motorsuz - oymak, yiv açmak, yontmak, tiraslamak, yontmak, kesmek için
	06.04 El aletleri, motorsuz - kazimak, parlatmak, ovmak için
	06.05 El aletleri, motorsuz - delmek, çevirmek, vidalamak için
	06.06 El aletleri, motorsuz - çakmak, perçinlemek, zimbalamak için
	06.07 El aletleri, motorsuz - dikmek, örmek için
	06.08 El aletleri, motorsuz - kaynatmak, yapistirmak için
	06.09 El aletleri, motorsuz - topragi islemek ve malzeme çikarmak için (çiftçilik aletleri de dahil olmak üzere)
	06.10 El aletleri, motorsuz - cilalamak, yaglamak, yikamak, temizlemek için
	06.11 El aletleri, motorsuz - boyamak için
	06.12 El aletleri, motorsuz - yerinde tutmak, kavramak için
	06.13 El aletleri, motorsuz - mutfak isleri için (biçaklar disinda)
	06.14 El aletleri, motorsuz - ameliyat ve tibbi isler için - keskin, kesme
	06.15 El aletleri, motorsuz - ameliyat ve tibbi isler için - kesme olmadan, digerleri
	06.99 06 numarali grup içinde yukarida listelenmemis baska bilinen motorsuz el aletleri
7	El makineleri veya elle yönlendirilern aletler, makineli
	07.01 Makineli el aletleri - biçkicilik için
	07.02 Makineli el aletleri - kesmek, ayirmak için (makas, budama makaslari da dahil olmak üzere)

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K - (devamı)

Kod	Açıklama
	07.03 Makineli el aletleri - oymak, yiv açmak, yontmak, (çim biçmek, 09.02ye bakin) tiraslamak, yontmak, kesmek için
	07.04 Makineli el aletleri - kazimak, parlatmak, ovmak için
	07.05 Makineli el aletleri - delmek, çevirmek, vidalamak için
	07.06 Makineli el aletleri - çakmak, perçinlemek, zimbalamak için
	07.07 Makineli el aletleri - dikmek, örmek için
	07.08 Makineli el aletleri - kaynatmak, yapistirmak için
	07.09 Makineli el aletleri - topragi islemek ve malzeme çikarmak için (çiftçilik aletleri, beton kiricilar da dahil olmak üzere)
	07.10 Makineli el aletleri - cilalamak, nemlendirmek, yikamak, temizlemek için (yüksek basinçli elektrik süpürgeleri de dahil olmak üzere)
	07.11 Makineli el aletleri - boyamak için
	07.12 Makineli el aletleri - yerinde tutmak, kavramak için
	07.13 Makineli el aletleri - mutfak isleri için (biçaklar disinda)
	07.14 Makineli el aletleri - isitmak için (kurutucular, alev tabancalari, ütüler de dahil olmak üzere)
	07.15 Makineli el aletleri - ameliyat ve tibbi isler için -sivri; kesici
	07.16 Makineli el aletleri - ameliyat ve tibbi isler için -kesici olmadan, digerleri
	07.17 Basinçli tabancalar (alet belirtilmeden)
	07.99 07 numarali grup içinde, yukarida listelenmemis baska bilinen el makineleri veya elle yönlendirilen aletler
8	El aletleri - güç kaynağı belirtilmemiş
	08.01 Güç kaynagi belirtilmeyen el aletleri - biçkicilik için
	08.02 Güç kaynagi belirtilmeyen el aletleri - kesmek, ayirmak için (makas, budama makaslari da dahil olmak üzere)
	08.03 Güç kaynagi belirtilmeyen el aletleri - oymak, yiv açmak, yontmak, tiraslamak, yontmak, kesmek için
	08.04 Güç kaynagi belirtilmeyen el aletleri - kazimak, parlatmak, ovmak için
	08.05 Güç kaynagi belirtilmeyen el aletleri - delmek, çevirmek, vidalamak için
	08.06 Güç kaynagi belirtilmeyen el aletleri - çakmak, perçinlemek, zimbalamak için
	08.07 Güç kaynagi belirtilmeyen el aletleri - dikmek, örmek için
	08.08 Güç kaynagi belirtilmeyen el aletleri - kaynatmak, yapistirmak için
	08.09 Güç kaynagi belirtilmeyen el aletleri - topragi islemek ve malzeme çikarmak için (çiftçilik aletleri de dahil olmak üzere)
	08.10 Güç kaynagi belirtilmeyen el aletleri - cilalamak, yaglamak, yikamak, temizlemek için
	08.11 Güç kaynagi belirtilmeyen el aletleri - boyamak için
	08.12 Güç kaynagi belirtilmeyen el aletleri - yerinde tutmak, kavramak için
	08.13 Güç kaynagi belirtilmeyen el aletleri - mutfak isleri için (biçaklar disinda)
	08.14 Güç kaynagi belirtilmeyen el aletleri - ameliyat ve tibbi isler için - keskin, kesme

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K - (devamı)

Kod	Açıklama
	08.15 Güç kaynagi belirtilmeyen el aletleri - ameliyat ve tibbi isler için - kesme olmadan, digerleri
	08.99 08 numarali grup içinde, yukarida listelenmemis baska bilinen güç kaynagi belirtilmeyen el aletleri
9	Makine ve ekipmanlar - taşınabilir veya seyyar
	09.01 Tasinabilir veya seyyar makineler - topragi islemek ve malzeme çikarmak için - bina yapimi ve insaat mühendisligi isleri için maden, tasocagi ve fabrika
	09.02 Tasinabilir veya seyyar makineler-topragi islemek, çiftçilik için
	09.03 Tasinabilir veya seyyar makineler (topragin islenmesi için olmayan) - insaat santiyeleri için
	09.04 Seyyar yer temizleme makineleri
	09.99 09 numarali grup içinde, yukarida listelenmemis baska bilinen tasinabilir veya seyyar makine ve ekipmanlar
10	Makine ve ekipmanlar - sabit
	10.01 Topragi islemek ve malzeme çikarmak için sabit makineler
	10.02 Malzemeleri hazirlamak, ezmek, tozlastirmak, filtrelemek, ayirmak, karistirmak, harmanlamak için makineler
	10.03 Malzemeleri islemek için makineler - kimyasal islemler (reaktif, mayalama islemleri)
	10.04 Malzemeleri islemek için makineler- sicak islemler (firin, kurutucu, ocaklar)
	10.05 Malzemeleri islemek için makineler - soguk islemler (sogukluk üretilmesi)
	10.06 Malzemeleri islemek için makineler - diger islemler
	10.07 Biçimlendirme makineleri - basmak, ezmek yoluyla
	10.08 Biçimlendirme makineleri - perdahlamak, yuvarlamak, silindir baskilar yoluyla (kagit baskilari da dahil olmak üzere)
	10.09 Biçimlendirme makineleri - enjeksiyon, itip çikarma, üfleme, egirme, kaliplama, eritme, dökme yoluyla
	10.11 Metal isleme makineleri - biçkicilik için
	10.12 Metal isleme makineleri - kesmek, ayirmak, yontmak için (kalip kesici, biçki makineleri, kirpma makaslari, oksijen tabancali ekipmanlar dahil olmak üzere)
	10.13 Yüzey isleme makineleri - temizlik, yikama, kurulama, boyama, baski
	10.14 Yüzey isleme makineleri - galvaniz, elektrolit yüzey bakimi
	10.15 Montaj makineleri (kaynatma, yapistirma, çakma, vidalama, perçinleme, egirme, telleme, dikme, zimbalama)
	10.16 Paketleme makineleri, sarma makineleri (doldurma, etiketleme, kapatma...)
	10.17 Özel sanayiler için diger makineler (çesitli izleme ve sinama makineleri)
	10.18 Çiftçilikte kullanilan ve yukaridaki makineler arasina dahil olmayan özel makineler

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K - (devamı)

Kod	Açıklama
	10.99 10 numarali grup içinde, yukarida listelenmemis baska bilinen sabit ve makine ve ekipmanlar
	10.10 Metal isleme makineleri - rendelemek, ögütmek, yüzey bakimi, çekmek, cilalamak, tornalamak, delmek için
11	Aktarma, taşıma ve depolama sistemleri
	11.01 Sabit tasiyici kayis, sürekli isleme ekipman ve sistemleri - kayis, yürüyen merdiven, teleferik, tasiyici kayis, vb.)
	11.02 Asansörler- agir yük asansörü, vinç, kriko, vb.
	11.03 Sabit vinç, seyyar vinç, araç tarafindan tasinan vinç, üst raylarda çalisan vinç, asili yükü olan kaldirma aygitlari
	11.04 Seyyar isleme aygiti, isleme kamyonlari (motorlu veya motorsuz) - el arabasi, paletli kamyonlar, vb.
	11.05 Kaldirma ekipmani, sabitlestirme, kavrama ve çesitli isleme aygitlari (aski, çengel, halat, vb. dahil olmak üzere)
	11.06 Depolama sistemleri, paketleme ekipmani, kaplar (silo, tanklar) - sabit - tank, fiçi, kap, vb.
	11.07 Depolama sistemleri, paketleme ekipmani, kaplar - seyyar
	11.08 Depolama aksami, raf, paletli raf, paletler
	11.09 Çesitli paketleme, küçük ve orta boy, seyyar (dekovil, çesitli kap, sise, sandik, söndürücüler...)
	11.99 11 numarali grup içinde, yukarida listelenmemis baska bilinen aktarma, tasima ve depolama sistemleri
12	Kara taşıtları
	12.01 Tasitlar - agir: kamyon/otobüsler (yolcu tasimasi)
	12.02 Tasitlar - hafif: esya veya yolcu
	12.03 Tasitlar - iki veya üç tekerlekli, motorlu veya motorsuz
	12.04 Diger kara tasitlari: kayak, patenler
	12.99 12 numarali grup içinde, yukarida listelenmemis baska bilinen kara tasitlari
13	Diğer taşıt araçları
	13.01 Tasitlar - rayda, asili monoraylar dahil: esya
	13.02 Tasitlar - rayda, asili monoraylar dahil: yolcu
	13.03 Tasitlar - deniz: esya
	13.04 Tasitlar - deniz: yolcu
	13.05 Tasitlar - deniz: balikçilik
	13.06 Tasitlar - hava: esya
	13.07 Tasitlar-hava: yolcu
	13.99 13 numarali grup içinde, yukarida listelenmemis baska bilinen diger tasit araçlari
14	Malzeme, nesne, ürün, makine akşamı, enkaz, toz
	14.01 Yapi malzemeleri - büyük ve küçük: prefabrike kabuk, kalip, kiris, hatil, tugla, kiremit, vb.
	14.02 Makine parçalari, tasit parçalari: sasi, dirsekli mahfaza, manivela, tekerlek, vb.
	14.03 Makine parça veya bilesikleri, metal isleme makineleri (isbu maddi araçlardan kopan parça ve yongalar da dahil olmak üzere)

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K - (devamı)

Kod	Açıklama
	14.04 Birlestirme aygitlari: somun, civata, vida, çivi, vb.
	14.05 Zerrecik, toz, kiymik, parçacik, zifos, kirik çömlek parçalari, diger enkaz
	14.06 Çiftlik ürünleri (tohum, saman, diger çiftlik ürünleri de dahil olmak üzere)
	14.07 Çiftçilik ve yetistiricilikte kullanilan ürünler (gübre, yemler de dahil olmak üzere)
	14.08 Depolanmis ürünler - depolama alanindaki nesne ve paketler de dahil olmak üzere
	14.09 Depolanmis ürünler- rulo, makara olarak
	14.11 Yükler- bir kaldirma aygiti, vinçte asili olan
	14.12 Yükler - elle islenen
	14.99 14 numarali grup içinde, yukarida listelenmemis baska bilinen malzeme, nesne, ürün, makine parçalari
	14.10 Yükler - makineli isleme veya aktarma aygiti tarafindan tasinan
15	Kimyasal, patlayıcı, radyoaktif, biyolojik maddeler
	15.01 Maddeler - yakici, çürütücü (kati, sivi veya gaz)
	15.02 Maddeler - zararli, zehirli (kati, sivi veya gaz)
	15.03 Maddeler - alevlenebilir (kati, sivi veya gaz)
	15.04 Maddeler - patlayici, reaktif (kati, sivi veya gaz)
	15.05 Belirgin etkisi olmayan gaz, buharlar (yasam biçimleri için atil, havasiz birakan)
	15.06 Maddeler - radyoaktif
	15.07 Maddeler - biyolojik
	15.08 Maddeler, malzemeler - özel bir rizikosu olmayan (su, atil maddeler...)
	15.99 15 numarali grup içinde, yukarida listelenmemis baska bilinen kimyasal, patlayici, radyoaktif, biyolojik maddeler
16	Güvenlik aygıt ve ekipmanı
	16.01 Güvenlik aygitlari - makinelerde
	16.02 Koruyucu aygitlar - münferit
	16.03 Alarm aygit ve ekipmani
	16.99 16 numarali grup içinde, yukarida listelenmemis baska bilinen güvenlik aygit ve ekipmani
17	Büro ekipmanı, kişisel ekipman, spor ekipmanı, silahlar, ev eşyaları
	17.01 Mobilya
	17.02 Ekipman - bilgisayar, büro otomasyonu, röprodüksiyon, iletisim
	17.03 Ekipman - ögretim, yazma, çizme için - daktilo, otomatik pullama makinesi, agrandizör, zaman kaydediciler de dahil olmak üzere
	17.04 Spor ve oyunlar için esya ve ekipman
	17.05 Silahlar
	17.06 Kisisel esya, giysi
	17.07 Müzik aletleri
	17.08 Ev esyasi türü ekipman, alet, nesne, çamasir (mesleki kullanim)
	17.99 17 numarali grup içinde, yukarida listelenmemis baska bilinen büro ekipmani, kisisel ekipman, spor ekipmani, silahlar, ev esyalari

Tablo A.7 KAZAMAG_K / OZELFAA_MAG_K / YARAMAG_K - (devamı)

Kod	Açıklama
18	Yaşayan organizmalar ve insanlar
	18.01 Agaç, bitki, mahsul
	18.02 Hayvan - evcil ve yetistirme amaçli
	18.03 Hayvan - vahsi hayvanlar, böcek, yilanlar
	18.04 Mikroorganizmalar
	18.05 Bulasici mikroplu amiller
	18.06 Insanlar
	18.99 18 numarali grup içinde, yukarida listelenmemis baska bilinen canlilar ve insanlar
19	Kütle atık
	19.01 Kütle atik - hammadde, ürün madde, nesnelerden
	19.02 Kütle atik - kimyasallardan
	19.03 Kütle atik - biyolojik madde, bitki, hayvanlardan
	19.99 19 numarali grup içinde, yukarida listelenmemis baska bilinen kütle atik
20	Fiziki olgu ve doğal unsurlar
	20.01 Fiziki olgu - gürültü, dogal radyasyon, isik, isik huzmesi, basinç yükseltme, basinç düsürme, basinç
	20.02 Dogal ve atmosfer unsurlari (akarsu, çamur, yagmur, dolu, kar, buz, rüzgar, vb. de dahil olmak üzere)
	20.03 Dogal afetler (sel, yanardag infilaki, deprem, tsunami, yangin da dahil olmak üzere)
	20.99 20 numarali grup içinde, yukarida listelenmemis baska bilinen fiziki olgu ve unsurlar
NA	00.01 Materyal araç yok
NA	00.02 Bilgi yok
NA	00.99 00 grubunda, yukarida listelenmemis baska bilinen durum .
NA	-1
NA	99.00 Bu siniflandirmada listelenmemis baska maddi araç

Tablo A.8 OZELFAA_K

Kod	Açıklama
1	Makine işletimi
	11 Makinenin çalistirilmasi, makinenin durdurulmasi
	12 Makinenin beslenmesi, makinenin bosaltilmasi
	13 Makinenin izlenmesi, makinenin isletilmesi veya sürülmesi
	19 Yukarida listelenmemis baska 10 çesit kaza aninda kazazedenin yaptigi faaliyet
2	El makineleriyle çalisma
	21 El makineleriyle çalisma - elle
	22 El makineleriyle çalisma - motorlu
	29 Yukarida listelenmemis baska 20 çesit kaza aninda kazazedenin yaptigi faaliyet
3	Sürücülük/taşıt aracında bulunmak veya donanım kullanımı

Tablo A.8 OZELFAA_K - (devamı)

Kod	Açıklama
	31 Bir tasit aracini sürmek veya donanimi kullanmak - seyyar ve motorlu
	32 Bir tasit aracini sürmek veya donanimi kullanmak - seyyar ve motorsuz
	33 Bir tasit aracinin yolcusu olmak
	39 Yukarida listelenmemis baska 30 çesit kaza aninda kazazedenin yaptigi faaliyet
4	Nesnelerin Kullanımı
	41 Ele almak, tutmak, kavramak, yakalamak, yerlestirmek - yatay bir düzeyde
	42 Baglamak, yapistirmak, ayirmak, sökmek, sikmak, vidasini açmak, vidalamak,çevirmek
	43 Sikilamak, asmak, yükseltmek, koymak - düsey bir düzeyde
	44 Atmak, savurmak
	45 Açmak, kapatmak (kutu, paket, parsel)
	46 Akitmak, dökmek, doldurmak, sulamak, püskürtmek, bosaltmak
	49 Yukarida listelenmemis baska 40 çesit kaza aninda kazazedenin yaptigi faaliyet
	Elle Taşıma
	51 Düsey olarak tasima - bir nesneyi kaldirma, yükseltme, alçaltma
	52 Yatay olarak tasima - bir nesneyi çekme, iteme, yuvarlama
	53 Bir yükü tasima - bir kisi tarafindan tasinmasi
	59 Yukarida listelenmemis baska 50 çesit kaza aninda kazazedenin yaptigi faaliyet
6	Hareket
	61 Yürüme, kosma, çikma, inme, vb.
	62 Çikma veya girme
	63 Atlama, ziplama, vb.
	64 Sürünme, tirmanma, vb.
	65 Kalkma, oturma
	66 Yüzme, dalma
	67 Yerdeki hareketler
	69 Yukarida listelenmemis baska 60 çesit kaza aninda kazazedenin yaptigi faaliyet
7	70 Bulunma - Belirtilmemis
NA	
NA	0 Bilgi yok
NA	99 Bu siniflandirma listelenmemis baska kazadan az önceki zamanda kazazedenin yürüttügü özel faaliyet

Tablo A.9 SG_KAZAGERORTAM_K

Kod	Açıklama
1	110 Ara Dinlenmesinde
2	120 Çalisirken
3	201 Süt Izni Esnasinda
4	211 Isten Eve Kendi Arabasiyla Gelirken

5	212 Evden Ise Kendi Arabasiyla Giderken
6	221 Isten Eve Servisle Gelirken
7	222 Evden Ise Servisle Giderken
NA	202 BILINMEYEN
NA	

Tablo A.10 YS_KOD_K

Kod	Açıklama
1	Elektrik akımı, ısı, tehlikeli maddelerle temas
	11 Bir kaynak arki, kivilcim veya çakmasi (pasif)
	12 Elektrikle dogrudan temas, elektrik yüklenmenin bedene alinmasi
	13 Çiplak alev veya sicak veya yanan bir nesne veya ortam ile temas
	14 Soguk veya donmus bir nesne veya ortam ile temas
	19 Yukarida listelenmemis baska 10 çesit sapma
	15 Tehlikeli maddelerle temas - burun, agiz yoluyla teneffüs
	16 Tehlikeli maddelerle temas - cilt veya gözler yoluyla/üzerinden
	17 Tehlikeli maddelerle temas - yutma veya yeme yoluyla sindirim sisteminden
2	Boğulma, gömülme, sarılma
	21 Sivi içinde bogulma
	22 Kati madde altinda gömülme
	23 Gaz veya havadaki zerrecikler içinde veya tarafindan sarilma
	29 Yukarida listelenmemis baska 20 çesit sapma
3	Sabit bir nesneye yatay veya düşey darbe (kazazede hareket halindeyken)
	31 Bir seye dogru düsey hareket, çarpisma (düsüsten dolayi)
	32 Bir seye dogru yatay hareket, çarpisma
	39 Yukarida listelenmemis baska 30 çesit sapma
	83 Darbe, çifte, toz veya sikarak bogma
4	Hareket halindeki bir nesnenin çarpması, çarpışma
	41 Uçan bir nesnenin darbesi
	42 Düsen bir nesnenin darbesi
	43 Salinan bir nesnenin darbesi
	44 Araçlar da dahil olmak üzere dönen, hareket eden, tasinmakta olan bir nesnenin darbesi
	45 Araçlar da dahil olmak üzere bir nesne ile çarpisma - bir kisi ile çarpisma (kazazede hareket halindeyken)
	49 Yukarida listelenmemis baska 40 çesit sapma
5	Sivri, uçlu, sert veya kaba bir materyal araç ile temas
	51 Keskin bir materyal araç ile temas (biçak, keski, vb.)
	52 Sivriltilmis bir materyal araç ile temas (çivi, sivri alet, vb.)
	53 Sert veya kaba materyal araç ile temas
	Bilgi yok
6	Kasılmak, ezilmek
	61 Kisilmak, ezilmek - içeride
	62 Kisilmak, ezilmek - altinda
	63 Kisilmak, ezilmek - arasinda

Tablo A.10 YS_KOD_K - (devamı)

Kod	Açıklama
	64 Kol, el veya parmagin kopmasi, kesilmesi
	69 Yukarida listelenmemis baska 60 çesit sapma
7	Fiziksel veya ruhsal baskı
	71 Fiziksel baski - kas/iskelet sistemi üzerinde
	72 Fiziksel baski - radyasyon, gürültü, isik veya basinç nedeniyle
	73 Ruhsal stres veya sok
	79 Yukarida listelenmemis baska 70 çesit sapma
8	Isırılma, tekme, vb. (hayvan veya insan tarafından)
	81 Isirik
	82 Böcek veya balik sokmasi
	89 Yukarida listelenmemis baska 80 çesit sapma
NA	
NA	0 Bilgi yok
NA	99 Bu siniflandirmada listelenmemis yaralanmaya sebep olan hareket (olay)

Tablo A.11 SG_MESLEK_ISCO08SG_GOREVI_KOD

Kod	Açıklama
1	01 Subaylar
2	02 Subay olmayan silahlı kuvvetlerin daimi mensupları
3	03 Silahlı kuvvetlerde diğer rütbelerdeki meslekler
4	11 Başkanlar, üst düzey yöneticiler ve kanun yapıcılar
5	12 Ticari ve idari müdürler
6	13 Üretim ve uzmanlaşmış hizmet müdürleri
7	14 Ağırlama, perakende ve diğer hizmet müdürleri
8	21 Bilim ve mühendislik alanlarındaki profesyonel meslek mensupları
9	22 Sağlık profesyonelleri
10	23 Eğitim ile ilgili profesyonel meslek mensupları
11	24 İş ve yönetim ile ilgili profesyonel meslek mensupları
12	25 Bilgi ve iletişim teknolojisi ile ilgili profesyonel meslek mensupları
13	26 Hukuk, sosyal ve kültür ile ilgili profesyonel meslek mensupları
14	31 Bilim ve mühendislik ile ilgili yardımcı profesyonel meslek mensupları
15	32 Yardımcı sağlık profesyonelleri
16	33 İş ve idare ile ilgili yardımcı profesyonel meslek mensupları
17	34 Hukuk, sosyal, kültür ve benzeri alanlar ile ilgili yardımcı profesyonel meslek mensupları
18	35 Bilgi ve iletişim teknisyenleri
19	41 Genel büro elemanları ile klavye kullanan büro elemanları
20	42 Müşteri hizmetlerinde çalışan elemanlar
21	43 Sayısal işlemler yapan ve malzeme kayıtları tutan büro elemanları
22	44 Diğer büro hizmetlerinde çalışan elemanlar
23	51 Kişisel hizmetler veren elemanlar
24	52 Satış hizmetleri veren elemanlar
25	53 Kişisel bakım hizmetleri veren elemanlar
26	54 Koruma hizmetleri veren elemanlar

Tablo A.11 SG_MESLEK_ISCO08SG_GOREVI_KOD - (devamı)

Kod	Açıklama
27	61 Pazara yönelik nitelikli tarım çalışanları
28	62 Pazara yönelik nitelikli ormancılık, su ürünleri ve avcılık çalışanları
29	63 Kendi geçimine yönelik çiftçiler, balıkçılar, avcılar ve toplayıcılar
30	71 İnşaat ve ilgili işlerde çalışan sanatkarlar (elektrikçiler hariç)
31	72 Metal işleme, makine ve ilgili işlerde çalışan sanatkarlar
32	73 El sanatları ve basım ile ilgili işlerde çalışanlar
33	74 Elektrik ve elektronik işlerde çalışan sanatkarlar
34	75 Gıda işleme, ağaç işleri, giyim eşyası ve diğer sanatkarlar ve ilgili işlerde çalışanlar
35	81 Sabit tesis ve makine operatörleri
36	82 Montajcılar
37	83 Sürücüler ve hareketli makine ve teçhizat operatörleri
38	91 Temizlikçiler ve yardımcılar
39	92 Tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar
40	93 Madencilik, inşaat, imalat ve ulaştırma sektörlerinde nitelik gerektirmeyen işlerde çalışanlar
41	94 Yiyecek hazırlama yardımcıları
42	95 Cadde ve sokaklarda satış ve hizmet işlerinde çalışanlar
43	96 Çöpçüler, atık toplayıcılar ve diğer nitelik gerektirmeyen işlerde çalışanlar

Tablo A.12 SG_CINSIYET

Kod	Açıklama
1	Erkek
2	Kadın

Tablo A.13 SG_MEDENIHAL

Kod	Açıklama
1	Bekar
2	Evli
3	Boşanmış, Dul
NA	Bilinmiyor

Tablo A.14 OGR_DURUMKOD

Kod	Açıklama
1	11 OKUR YAZAR DEGIL
2	20 ILKOGRETIM MEZUNU
2	21 ILKOKUL
2	10 OKUR YAZAR
3	22 ORTAOKUL
4	30 LISE

Tablo A.14 OGR_DURUMKOD - (devamı)

Kod	Açıklama
4	31 MESLEK LISESI
5	40 YUKSEKOKUL
6	50 UNIVERSITE
7	52 DOKTORA
7	51 Master, yüksek mühendislik vb.

Tablo A.15 SAAT_ARALIK

Kod	Açıklama
1	[0,2)
2	[2,4)
3	[4,6)
4	[6,8)
5	[8,10)
6	[10,12)
7	[12,14)
8	[14,16)
9	[16,18)
10	[18,20)
11	[20,22)
12	[22,24]

Tablo A.16 KAZA_MEVSİM

Kod	Açıklama
1	Kış
2	İlkbahar
3	Yaz
4	Sonbahar

Tablo A.17 YT_KOD

Kod	Açıklama
1	Yaralar ve yüzeysel yaralanmalar
	11 Yüzeysel yaralanmalar
	12 Açik yaralar
	19 Diger tür yaralar ve yüzeysel yaralanmalar
2	Kemik kırıkları
	21 Kapali kiriklar
	22 Açik kiriklar
	29 Diger tür kemik kiriklari
3	Çıkıklar, burkulmalar ve incinmeler
	31 Çikiklar ve yari çikiklar
	32 Burkulmalar ve incinmeler

Tablo A.17 YT_KOD - (devamı)

Kod	Açıklama
	39 Diger tür çikik, burkulma ve incinmeler
4	40 Travma sonucu organ kaybi (bedenin bir parçasinin kaybi)
5	Beyin sarsıntısı ve iç yaralanmalar
	51 Beyin sarsintisi ve kafatasi yaralanmalari
	52 Iç yaralanmalar
	59 Diger tür beyin sarsintisi ve iç yaralanmalar
6	Yanıklar, kaynar su ile kavrlma ve donmalar
	61 Yaniklar ve kaynar su ile kavrulmalar (termal)
	62 Kimyasal yaniklar (korozyon)
	63 Donmalar
	69 Diger tür yaniklar, kaynar su ile kavrulma ve donmalar
7	Zehirlenme ve enfeksyonlar
	71 Akut zehirlenmeler
	72 Akut enfeksiyonlar
	79 Diger tür zehirlenme ve enfeksiyonlar
8	Suda boğulma ve nefesin kesilmesi
	81 Nefesin kesilmesi
	82 Suda bogulma ve ölümle sonuçlanmayan bogulma
	89 Diger tür bogulma ve nefesin kesilmesi
9	Ses titreşm ve basınç etkileri
	91 Akut isitme kaybi
	92 Basinç etkileri (yüksek basinca bagli)
	99 Diger tür ses, titresim ve basinç etkileri
10	Aşırı ısı, ışık ve radyasyon etkileri
	101 Sicaklik ve günes çarpmasi
	102 Radyasyon etkileri (termal olmayan)
	103 Indirgenmis isi etkileri
	109 Diger tür asiri isi, isik ve radyasyon etkileri
11	Şok
	111 Saldiri ve tehdit (korkutma) sonrasi soklar
	112 Travma soklari
	119 Diger tür soklar
12	120 Birden fazla sayida yaralanmalar
NA	-1
NA	0 Yaranin türü bilinmeyen veya belirtilmemis
NA	999 Diger basliklar altinda içerilmeyen diger belirtilmis yaralanmalar

Tablo A.18 YVY_KOD

Kod	Açıklama
1	Kafada meydana gelen belirtilmemiş alanlar
	11 Kafa, beyin ve beyine bagli sinir ve damarlari
	12 Yüz bölgesi
	13 Göz(ler)
	14 Kulak(lar)
	15 Disler

Tablo A.18 YVY_KOD - (devamı)

Kod	Açıklama
	18 Kafa; çesitli bölgeleri etkilenmis
	19 Kafa; yukarida belirtilmemis diger bölgeler
2	Boyun, boyundaki omurilik ve omur dahil
	21 Boyun; boyundaki omurilik ve omur dahil
	29 Boyun; yukarida belirtilmeyen diger bölgeleri
3	Sırt, sırttaki omurilik ve omur dahil
	31 Sirt; sirttaki omurilik ve omur dahil
	39 Sirt; yukarida belirtilmeyen diger bölgeleri
4	Gövde ve organlar
	41 Gögüs kafesi (kaburga), eklem ve kürek kemikleri dahil olmak üzere kaburgalar
	42 Organlariyla birlikte gögüs bölgesi
	43 Organlariyla birlikte karinla ilgili ve sirtin alt bölümü (kalça kismi)
	48 Gövde; çesitli bölgeleri etkilenmis
	49 Gövde; yukarida belirtilmeyen diger bölgeleri
5	Kollar
	51 Omuz ve omuz eklemleri
	52 Dirsek dahil kol
	53 El
	54 Parmak(lar)
	55 Bilek
	58 Kollar; çesitli bölgeleri etkilenmis
	59 Kollar; yukarida belirtilmemis diger bölgeleri
6	Bacaklar
	61 Kalça ve kalça eklemleri
	62 Diz dahil bacak
	63 Ayak bilegi
	64 Ayak
	65 Ayak parmak(lar)i
	68 Bacaklar; çesitli bölgeleri etkilenmis
	69 Bacaklar; yukarida belirtilmemis diger bölgeleri
7	Tüm beden ve çeşitli bölgeler
	71 Tüm beden (bütün vücuda tesir eden etkiler)
	78 Bedenin, etkilenmis çesitli bölgeleri
NA	-1
NA	0 Yaranin vücuttaki yeri belirtilmemis
NA	99 Vücudun yaralanmis diger bölgelerinden yukarida belirtilmemis alanlar

B
KURALLAR

Tablo B.1 Dolaylı etkili değişkenlerin ortalama kaldıraç değerleri

	Dolaylı değişkenler	1-Raporlu	2-Major	3-KayipVar	4-OlumVar
1	CALIS_SURE=0-1	1,11	1,13	1,35	1,35
2	CALIS_SURE=1-2	0,99	0,98	1,30	1,30
3	CALIS_SURE=2-4	1,13	1,11	0	0
4	CALIS_SURE=4-6	0,92	0,90	0,90	0,90
5	CALIS_SURE=6-8	0,99	1,01	1,33	1,33
6	CALIS_SURE=>8	1,01	0,98	1,22	1,22
7	KAZA_AY=1	0	0	0	0
8	KAZA_AY=2	0	0	2,02	0
9	KAZA_AY=3	1,11	1,12	1,11	1,43
10	KAZA_AY=4	0	0	1,12	0
11	KAZA_AY=5	1,03	0	1,05	0
12	KAZA_AY=6	0	0	1,07	0
13	KAZA_AY=7	0	0	1,04	0
14	KAZA_AY=8	0	0	0	0
15	KAZA_AY=9	0	0	0	0
16	KAZA_AY=10	0	0	0	0
17	KAZA_AY=11	0	0	0	0
18	KAZA_AY=12	0	0	0	0
19	KAZA_IL_K=1	0	0	0	0
20	KAZA_IL_K=2	0	0	0	0
21	KAZA_IL_K=3	0	0	0	0
22	KAZA_IL_K=4	0	0	0	0
23	KAZA_IL_K=5	0	0	0	0
24	KAZA_IL_K=6	0	0	3,14	0
25	KAZA_IL_K=7	0	0	0	0
26	KAZA_IL_K=8	0	0	0	0
27	KAZA_IL_K=9	0	0	0	0
28	KAZA_IL_K=10	0	0	15,08	0
29	KAZA_IL_K=11	0	0	0	0
30	KAZA_IL_K=12	0	0	0	0
31	KAZA_IL_K=13	0	0	0	0
32	KAZA_IL_K=14	0	0	0	0
33	KAZA_IL_K=15	0	0	0	0

Tablo B.1 Dolaylı etkili değişkenlerin ortalama kaldıraç değerleri - (devamı)

	Dolaylı değişkenler	1-Raporlu	2-Major	3-KayipVar	4-OlumVar
34	KAZA_IL_K=16	1,26	0	1,82	0
35	KAZA_IL_K=17	0	0	0	0
36	KAZA_IL_K=18	0	0	0	0
37	KAZA_IL_K=19	0	0	0	0
38	KAZA_IL_K=20	0	0	4,47	0
39	KAZA_IL_K=21	0	0	0	0
40	KAZA_IL_K=22	0	0	78,23	0
41	KAZA_IL_K=23	0	0	0	0
42	KAZA_IL_K=24	0	0	0	0
43	KAZA_IL_K=25	0	0	0	0
44	KAZA_IL_K=26	0	0	10,13	0
45	KAZA_IL_K=27	0	0	0	0
46	KAZA_IL_K=28	0	0	0	0
47	KAZA_IL_K=29	0	0	0	0
48	KAZA_IL_K=30	0	0	0	0
49	KAZA_IL_K=31	0	0	0	0
50	KAZA_IL_K=32	0	0	0	0
51	KAZA_IL_K=33	0	0	0	0
52	KAZA_IL_K=34	1,07	1,01	1,09	0,88
53	KAZA_IL_K=35	1,54	1,26	1,44	0
54	KAZA_IL_K=36	0	0	0	0
55	KAZA_IL_K=37	0	0	0	0
56	KAZA_IL_K=38	0	0	10,32	0
57	KAZA_IL_K=39	0	0	0	0
58	KAZA_IL_K=40	0	0	0	0
59	KAZA_IL_K=41	1,74	0	1,72	0
60	KAZA_IL_K=42	0	0	0	0
61	KAZA_IL_K=43	0	0	0	0
62	KAZA_IL_K=44	0	0	0	0
63	KAZA_IL_K=45	0	0	12,96	0
64	KAZA_IL_K=46	0	0	0	0
65	KAZA_IL_K=47	0	0	0	0
66	KAZA_IL_K=48	0	0	0	0
67	KAZA_IL_K=49	0	0	0	0
68	KAZA_IL_K=50	0	0	0	0
69	KAZA_IL_K=51	0	0	0	0
70	KAZA_IL_K=52	0	0	0	0
71	KAZA_IL_K=53	0	0	0	0
72	KAZA_IL_K=54	0	0	6,61	0
73	KAZA_IL_K=55	0	0	0	0
74	KAZA_IL_K=56	0	0	0	0
75	KAZA_IL_K=57	0	0	0	0
76	KAZA_IL_K=58	0	0	0	0
77	KAZA_IL_K=59	0	0	4,69	0
78	KAZA_IL_K=60	0	0	0	0
79	KAZA_IL_K=61	0	0	0	0

Tablo B.1 Dolaylı etkili değişkenlerin ortalama kaldıraç değerleri - (devamı)

	Dolaylı değişkenler	1-Raporlu	2-Major	3-KayipVar	4-OlumVar
80	KAZA_IL_K=62	0	0	0	0
81	KAZA_IL_K=63	0	0	0	0
82	KAZA_IL_K=64	0	0	0	0
83	KAZA_IL_K=65	0	0	0	0
84	KAZA_IL_K=66	0	0	0	0
85	KAZA_IL_K=67	0	0	0	0
86	KAZA_IL_K=68	0	0	0	0
87	KAZA_IL_K=69	0	0	0	0
88	KAZA_IL_K=70	0	0	0	0
89	KAZA_IL_K=71	0	0	0	0
90	KAZA_IL_K=72	0	0	0	0
91	KAZA_IL_K=73	0	0	0	0
92	KAZA_IL_K=74	0	0	0	0
93	KAZA_IL_K=75	0	0	0	0
94	KAZA_IL_K=76	0	0	0	0
95	KAZA_IL_K=77	0	0	0	0
96	KAZA_IL_K=78	0	0	16,26	0
97	KAZA_IL_K=79	0	0	0	0
98	KAZA_IL_K=80	0	0	51,74	0
99	KAZA_IL_K=81	0	0	0	0
100	KAZA_MEVSIM=1	1,02	1,03	1,02	1,17
101	KAZA_MEVSIM=2	0,91	0,90	0,97	1,21
102	KAZA_MEVSIM=3	1,02	1,02	1,01	1,27
103	KAZA_MEVSIM=4	1,09	1,07	1,03	0,85
104	MEDENIHAL_K=1	0,84	0,82	0,97	0,64
105	MEDENIHAL_K=2	1,08	1,12	1,03	1,23
106	MEDENIHAL_K=3	0	0	0	0
107	MEDENIHAL_K=999	0	0	0	0
108	OGR_DURUM_K=1	0	0	0	0
109	OGR_DURUM_K=2	1,13	1,11	1,02	1,72
110	OGR_DURUM_K=3	1,07	1,03	1,07	1,22
111	OGR_DURUM_K=4	0,98	0,97	1,15	0,64
112	OGR_DURUM_K=5	0	0	5,40	0
113	OGR_DURUM_K=6	0	0	0	0
114	OGR_DURUM_K=7	0	0	0	0
115	OGR_DURUM_K=999	0	0	0	0
116	SAAT_ARALIK=1	0	0	0	0
117	SAAT_ARALIK=2	0	0	0	0
118	SAAT_ARALIK=3	0	0	0	0
119	SAAT_ARALIK=4	0	0	0	0
120	SAAT_ARALIK=5	1,23	1,26	1,07	1,72
121	SAAT_ARALIK=6	0,98	0,96	0,94	1,27
122	SAAT_ARALIK=7	0,98	1,08	1,06	0
123	SAAT_ARALIK=8	0,94	0,95	0,91	1,50
124	SAAT_ARALIK=9	0,99	0,96	0,97	1,21
125	SAAT_ARALIK=10	0	0	0	0

Tablo B.1 Dolaylı etkili değişkenlerin ortalama kaldıraç değerleri - (devamı)

	Dolaylı değişkenler	1-Raporlu	2-Major	3-KayipVar	4-OlumVar
126	SAAT_ARALIK=11	0	0	1,80	0
127	SAAT_ARALIK=12	0	0	1,96	0
128	SG_SONIS_TEC=0-2ay	0,92	1,31	0,92	3,28
129	SG_SONIS_TEC=2ay-1yil	0,93	0,88	0,98	1,01
130	SG_SONIS_TEC=1-5yil	1,13	1,16	1,17	1,24
131	SG_SONIS_TEC=>5yil	1,37	1,31	1,32	0
132	SG_TOP_TEC=0-2ay	0	0	0	0
133	SG_TOP_TEC=2ay-1yil	0	0	0	0
134	SG_TOP_TEC=1-5yil	0,83	0,86	1,06	0,85
135	SG_TOP_TEC=>5yil	1,04	1,06	1,01	1,07
136	SG_UCRETLIZINGUN=0	1,08	1,15	0,94	1,75
137	SG_UCRETLIZINGUN=1-15	1,11	1,11	1,18	0
138	SG_UCRETLIZINGUN=15-30	1,17	1,16	1,26	1,53
139	SG_UCRETLIZINGUN=30-45	0	0	0	0
140	SG_UCRETLIZINGUN=>45	0	0	0	0
141	SG_YAS=13-18	0	0	40,52	0
142	SG_YAS=18-25	0,79	0,81	1,05	0,83
143	SG_YAS=25-35	0,92	0,90	1,06	0,69
144	SG_YAS=35-50	1,16	1,18	1,03	1,07
145	SG_YAS=>50	1,22	1,28	1,03	3,44
146	SIG_SAY=Küçük(0-6)	1,45	1,42	1,09	1,49
147	SIG_SAY=Orta(6-50)	1,26	1,28	1,03	1,78
148	SIG_SAY=Büyük(50-2000)	0,86	0,81	1,14	0,57
149	SIG_SAY=Dev(>2000)	0	0	3,25	0

TEZDEN ÜRETİLMİŞ YAYINLAR

Makale

1. F. Recal, T. Demirel, "Comparison of machine learning methods in predicting binary and multi-class occupational accident severity". Journal of Intelligent Fuzzy Systems, vol. 40, pp. 10981-10998, 2022. DOI:10.3233/JIFS-202099