

Atf İçin: Karakaya Özkan, E. and Ulaş, H. B. (2023). Türkiye Metal Sektöründe Yaşanan İş Kazalarının Random Forest Modeliyle Tahminlenmesi. *İğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(3), 1983-1997.

To Cite: Karakaya Özkan, E. & Ulaş, H. B. (2023). Estimation of Occupational Accidents in the Turkish Metal Industry with Random Forest Model. *Journal of the Institute of Science and Technology*, 13(3), 1983-1997.

Türkiye Metal Sektöründe Yaşanan İş Kazalarının Rassal Orman Algoritmasıyla Tahminlenmesi

Ekin KARAKAYA ÖZKAN^{1*}, Hasan Basri ULAŞ²

Öne Çıkanlar:

- Metal sektörü iş kazası verileri analiz edilmiştir
- Makine öğrenim algoritması olan Rassal Orman uygulanarak iş kazası nedenleri analiz edilmiştir
- K-katlı çapraz doğrulama metodu kullanılmıştır

Anahtar Kelimeler:

- İş sağlığı ve güvenliği
- Metal sektörü
- Makine öğrenimi
- Rassal orman
- K-katlı çapraz doğrulama

ÖZET:

Bu çalışmanın amacı, Çalışma ve Sosyal Güvenlik Bakanlığı (ÇSGB) tarafından kayıt altına alınan, 2013-2018 yılları arasında metal sektöründe gerçekleşen, ölümlü ve uzuv kayıplı ulusal iş kazası verilerini kullanarak makine öğrenimi (ML) yöntemiyle bir tahmin algoritması geliştirmektir. İş kazası nedenlerinin detaylı bir şekilde sınıflandırılması ve tahmin edilmesi kazaları azaltmak için gereklidir. Literatürde; iş kazalarını azaltma amacıyla kaza ile ilgili faktörleri araştırmak ve etkili tahmin modelleri oluşturmak için çeşitli ML algoritmaları kullanılmıştır. Bu çalışmada, iş kazası nedenlerini ve sonuçlarını tahmin etmek amacıyla ML yöntemlerinden birisi olan Rassal Orman (RF) algoritması kullanılmıştır. Modelin doğrulaması için 10 katlı çapraz doğrulama modeli kullanılmış ve modelin doğruluk değeri %4.7 oranında artırılmıştır. RF algoritmasının doğruluk değeri 0.9172 olarak bulunmuştur. Metal sektöründe iş kazası nedenlerini etkileyen önemli faktörlerin analizinde özyinelemeli olarak özellik seçme (Recursive Feature Elimination - RFE) metodu kullanılmış ve en önemli özellikler kazanın ikincil tehlike kaynağı, iş günü kaybı ve kaza sebebi sapma kodu olarak bulunmuştur.

Estimation of Occupational Accidents in the Turkish Metal Industry with Random Forest Algorithm

Highlights:

- Occupational accident data of metal sector data were analyzed
- The causes of occupational accidents were analyzed by applying random forest, which is a machine learning algorithm
- K-fold cross validation method was used

Keywords:

- Occupational health and safety
- Metal sector
- Machine learning
- Random forest
- K-fold cross validation

ABSTRACT:

The aim of this study is to develop a predictive model using machine learning (ML) to identify the causes of fatalities and amputations in the metal sector based on occupational accident data collected by the Turkish Ministry of Labor and Social Security (MLSS) from 2013 to 2018. It is necessary to classify and predict occupational accident reasons in detail to prevent occupational accident. Researchers have used ML algorithm to investigate correlated factors and create effective prediction models in an effort to lower occupational accidents. In this study, we used random forest (RF) which is one of the ML algorithm to predict occupational accident reasons and consequences. 10- fold cross validation model is used for model validation and it increased %4.7 of accuracy of algorithm. Accuracy of RF is found as 0.9172. We extracted important factors that affect the occupational accident reasons at metal sector using Recursive Feature Elimination (RFE) and it is found that most important factors are secondary reason of the accident, days lost and deviation.

¹ Ekin KARAKAYA ÖZKAN ([Orcid ID: 0000-0002-3277-7119](https://orcid.org/0000-0002-3277-7119)), Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Kazaların Çevresel ve Teknik Araştırması Anabilim Dalı, Ankara, Türkiye

² Hasan Basri ULAŞ ([Orcid ID: 0000-0002-9754-6055](https://orcid.org/0000-0002-9754-6055)), Gazi Üniversitesi, Teknoloji Fakültesi, İmalat Mühendisliği Bölümü, Ankara, Türkiye

***Sorumlu Yazar/Corresponding Author:** Ekin KARAKAYA ÖZKAN, e-mail: ekinkarakaya@gmail.com.tr

Bu çalışma Ekin KARAKAYA ÖZKAN'ın devam etmekte olan Doktora tez çalışmalarından üretilmiştir.

GİRİŞ

Günümüzde sanayisinin ve üretim sistemlerinin en önemli sorunlarından birisi yaşanan iş kazalarıdır (Kifle ve ark., 2014; Nazaripour ve ark., 2018; Fuentes-Bargues ve ark., 2022). Uluslararası Çalışma Örgütü (ILO) tarafından yapılan araştırmaya göre; her yıl yaklaşık iki milyon kişi iş kazalarından dolayı hayatını kaybediyor. Ayrıca her yıl ölümlerle sonuçlanmayan ancak birçoğu geçici iş göremezliğe neden olan 374 milyon iş kazası ve meslek hastalığı meydana gelmektedir (ILO, 2023). Bu kazaların önlenmesi için iş sağlığı ve güvenliği alanında kontrol önlemlerinin alınması önem teşkil etmektedir.

Türkiye’de metal sektörü istatistikleri incelendiğinde sanayi sektörlerindeki toplam istihdamın %2’ sine, Gayri Safi Milli Hasıla’nın %3’ üne, toplam ihracatın ise %12’ sine sahip olduğu görülmektedir. Metal sektörü ihracat yapan sektörler arasında 3. sırada yer almaktadır. 2016 yılında Türkiye’de her 100.000 çalışandan yaklaşık 10’ u iş kazası sebebiyle hayatını kaybederken, metal sektöründe de bu oran aynıdır (SGK, 2017; Güllüoğlu ve Güllüoğlu, 2019). İş kazaları sebebiyle yaşanan bu hayat kayıplarını önlemek için; sektörde yaşanan iş kazalarının kök nedenlerinin belirlenerek önlenmesi büyük önem taşımaktadır. İş kazaları benzer kök nedenlerden gerçekleşirler. Bu sebeple yaşanmış bir iş kazasının kök nedeninin belirlenmesi, gelecekte yaşanması muhtemel iş kazalarını önlemek için uygun kontrol tedbirlerinin alınması aşamasında çok önemlidir (Lindberg ve ark., 2010; Salguero-Caparros ve ark., 2015; Shao ve ark., 2019; Anyfantis ve ark., 2021).

Yağimli ve ark. (Yağimli ve İzci, 2017) yaptıkları çalışmada, metal sektörünün yoğun üretim yapması, tehlikeli makineler kullanılması, tam otomasyon ile vardiyalı şekilde çalışması ve yoğun uzun saatler süren çalışma nedeniyle en fazla iş kazası yaşanan sektörlerden biri olduğunu belirtmişlerdir. Metal sektörü çalışanlarıyla yapılan başka bir çalışmada ise, iş kazası nedenlerinin dağılımı; %44 kişisel koruyucu donanımın yetersiz kullanılması, %38 dikkatsizlik ve kişisel nedenler, %18 makine ve tezgahlarda güvenlik önlemlerinin alınmaması şeklinde olduğu tespit edilmiştir. Ayrıca, iş kazalarının çoğunlukla kişisel koruyucu donanımların kullanılmaması ve çalışanların mesleki eğitiminin yetersiz olmasından kaynaklandığı ortaya konulmuştur (Gülhan ve ark., 2012).

Fuentes-Bargues ve ark. (Fuentes-Bargues ve ark., 2022) İspanya’da yaptıkları bir çalışmada, metal sektöründe; yer, çalışan karakteristiği, sapma, işe gelinmeyen günler, günün saati ve yaralanmanın kazanın ciddiyeti ile %95 güven düzeyinde ilişkili olduğunu belirtmişlerdir. İş kazalarının sebepleri incelendiklerinde ise makinelerin kontrolünün kaybedilmesi ile kayma ve düşme kazalarının daha ciddi sonuçları olduğunu tespit etmişlerdir.

ML, kendilerine beslenen verilere dayalı olarak öğrenebilen, ayarlayabilen ve iyileştirebilen sistemler tasarlama yöntemidir. ML teknolojisinin amacı, verilerin yapısını anlamak ve bu verileri daha sonra anlaşılabilen ve insanlar tarafından yaşam boyunca çeşitli uygulamalar için kullanılabilen tahminler oluşturarak belirli modellere sığdırmaktır (Mining, 2020; Sahay,2021). ML uygulamaları hayatın birçok alanında sıklıkla kullanılmakta ve birçok soruna karşı yüksek oranda başarı sunmaktadır. ML algoritmaları kaza araştırmaları, medikal görüntü işleme, ziraat sektörü, savunma sanayi, yapı sektörü, iş güvenliği ve birçok alanda kullanılmaktadır (Li ve ark., 2003; Yan ve ark., 2005; Goh ve Ubeynarayana, 2017; Park ve ark., 2018; Veziroğlu ve ark., 2023; Paçal 2023).

Goh and Ubeynarayana (Goh ve Ubeynarayana, 2017) gerçekleştirdikleri çalışmada geçmiş kazaların gelecekteki kazaları önlemek için bir temel teşkil edebileceğini belirtmişlerdir. Bu nedenle, iş kazalarını azaltmak ve tamamen önlemek için iş kazalarının nedenlerini ve türlerini sınıflandırarak tahmin etmenin önemini vurgulamışlardır. Tahmin metotlarıyla yaşanmış kaza verileri kullanılarak yapılan son çalışmalarda, kazaların tesadüfi bir şekilde gerçekleşmediği ortaya konulmuştur. Çalışmalar ayrıca kazaların önlenmesi için nicel araştırmaların yapılmasının gerekliliğini de

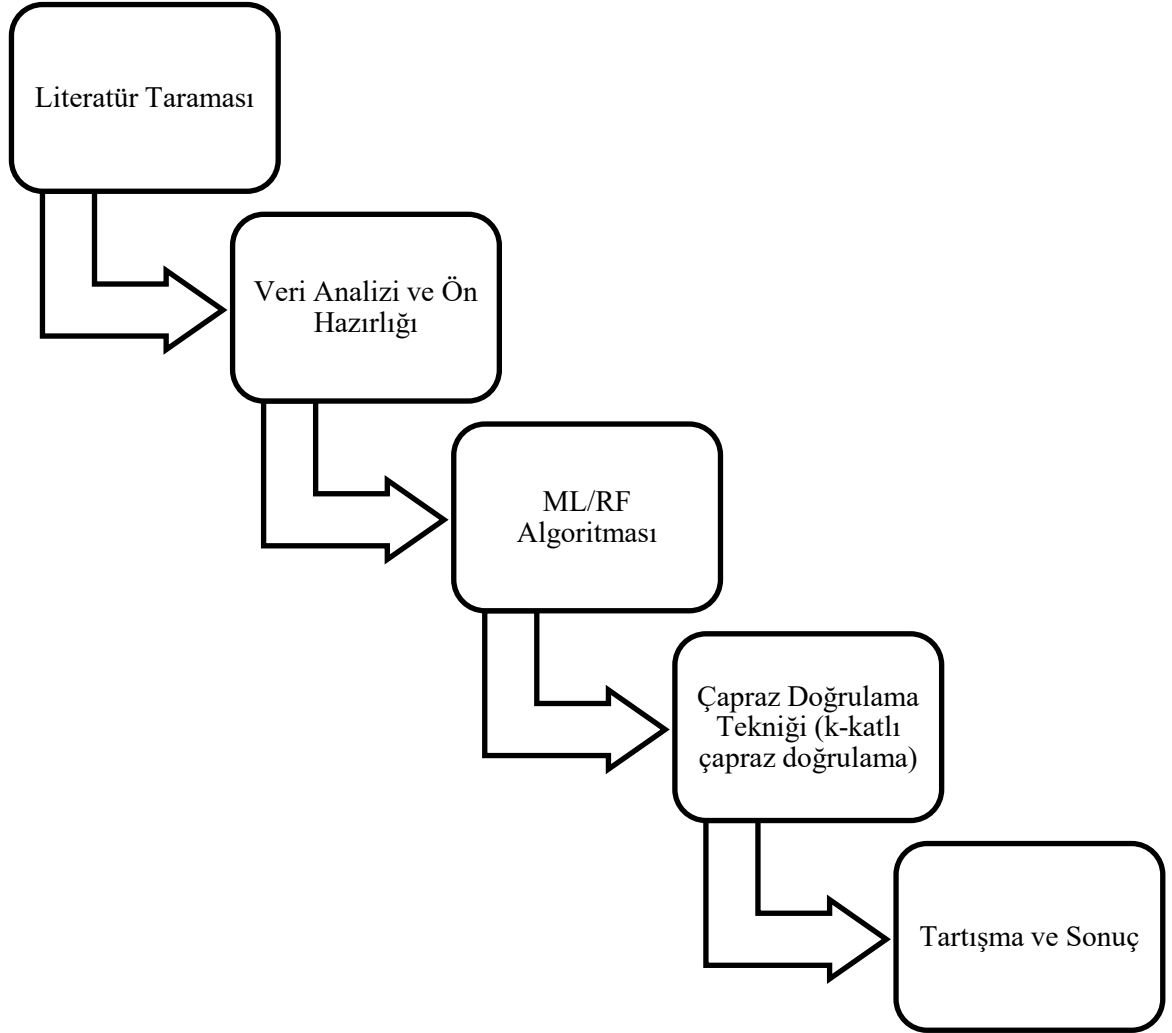
göstermiştir (Tixier ve ark., 2016). Bu sebeple, bazı araştırmalarda kazaları önlemek için ML algoritmaları, kaza veri setleri kullanılarak iş kazası parametrelerini tahmin etmek için kullanılmıştır. ML algoritmaları, farklı sektörlerde farklı amaçlar için kullanılmıştır. Örneğin, trafik kazası analizinde ML algoritmaları, potansiyel riskleri tahmin etmek için yaygın olarak kullanılmıştır (Yan ve ark., 2005; Yannis ve ark., 2010; Meng ve Weng, 2011; Weng ve ark., 2012; Gregoriades ve Mouskos, 2013; Karacasu ve ark., 2014; Usman ve ark., 2016; Nishimoto ve ark., 2017; Wang ve ark., 2019). Madencilik (Palei ve Das, 2009; He ve ark., 2010; Rivas ve ark., 2011; Sanmiquel ve ark., 2015), inşaat (Persona ve ark., 2006; Matías ve ark., 2008; Leu ve Chang, 2013; Alizadeh ve ark., 2015; Mıstıkoğlu ve ark., 2015; Amiri ve ark., 2016; Tixier ve ark., 2016; Yi ve ark., 2016; Goh ve Ubeynarayana, 2017; Chiang ve ark., 2018; Ghodrati ve ark., 2018; Park ve ark., 2018; Ayhan ve Tokdemir, 2019; Jahangiri ve ark., 2019; Jana ve ark., 2019; Sakhakarmi ve ark., 2019), havacılık (Li ve ark., 2003; Bazargan ve Guzhva, 2011; Yeoum ve Lee, 2013), sağlık (Veziroğlu ve ark., 2023; Paçal 2023) gibi sektörlerde aynı yaklaşım üzerine yapılmış çok sayıda çalışma olmasına rağmen, metal sektöründe (Yağimli ve İzci, 2017) ML algoritmalarının ne şekilde uygulanacağına dair literatürde hala yeterli sayıda araştırma bulunmamaktadır. Literatür incelendiğinde; ML kullanımı için en fazla kullanılan algoritmalar Destek Vektör Makineleri (support vector machine) (Matías ve ark., 2008; Rivas ve ark., 2011; Sánchez ve ark., 2011; Goh ve Ubeynarayana, 2017; Zhang ve ark., 2018; Aci ve Ozden, 2018; Sakhakarmi ve ark., 2019), karar ağaçları (decision tree) (Karlaftis ve Golias, 2002; Matías ve ark., 2008; Rivas ve ark., 2011; Weng ve ark., 2012; Sanmiquel ve ark., 2015; Aci ve Ozden, 2018; Jeong ve ark., 2018; Mafi ve ark., 2018; Zhang ve ark., 2018), Bayes Ağı (Bayesian Network) (Matías ve ark., 2008; Rivas ve ark., 2011; Leu ve Chang, 2013; Sanmiquel ve ark., 2015), K-En Yakın Komşuluk (k-nearest neighbour) (Goh ve Ubeynarayana, 2017; Aci ve Ozden, 2018; Mafi ve ark., 2018; Zhang ve ark., 2018), sınıflama ve regresyon ağacı (classification and regression tree) (Persona ve ark., 2006; Bevilacqua ve ark., 2008; Cheng ve ark., 2012; Amiri ve ark., 2016) ve RF (Siddiqui ve ark., 2012; Shanthi ve Ramani, 2012; Wang ve ark., 2015; Brown, 2016; Tixier ve ark., 2016; Mafi ve ark., 2018; Zhang ve ark., 2018; Jiang ve ark., 2019; Kang ve Ryu, 2019; Tang ve ark., 2019; Umer ve ark., 2020) şeklindedir. Bazı çalışmalarda birden fazla ML algoritması kullanılarak birbirleriyle başarı ölçütleri kıyaslanmıştır (Matías ve ark., 2008; Rivas ve ark., 2011; Sánchez ve ark., 2011; Tixier ve ark., 2016; Goh ve Ubeynarayana, 2017; Aci ve Ozden, 2018; Mafi ve ark., 2018; Jeong ve ark., 2018; Zhang ve ark., 2018; Wang ve ark., 2019; Jiang ve ark., 2019; Tang ve ark., 2019; Umer ve ark., 2020). Santos ve ark. (Santos ve ark., 2022) trafik kazaları verileri kullanarak gerçekleştirilen çalışmalar üzerine yaralanma şiddeti modellenmesinde kullanılan ML yöntemlerinin literatür araştırmasını yapmışlardır. RF algoritmasının, literatür araştırması yapılan bu çalışmaların içerisinde %29' unda kullanıldığı ve RF kullananların %70' inde RF'nin en yüksek performans gösterdiğini tespit etmişlerdir.

Zhang ve ark. (Zhang ve ark., 2018), Florida'da 3 yıllık bir süre içinde meydana gelen ve 51 tanesi ölümle sonuçlanan 5538 trafik kazasını incelemiştir. Trafik kazasının yaralanma ciddiyetini tahmin etmek için altı farklı algoritma (RF, K-En Yakın Komşuluk, Destek Vektör Makinesi, Sıralı Probit Model ve Multinomial Logit Modeli) kullanmışlar ve bu algoritmalar arasında RF en yüksek doğruluk oranı değerini göstermiştir.

İş kazalarının sınıflandırılması ve nedenlerinin tahmin edilmesi için bu çalışmada; metal sektöründe yaşanan ulusal iş kazası verileri kullanılarak kaza nedeni tahminleme çalışmalarında ML algoritması olan RF kullanılmıştır. Yukarıda bahsedilen tüm çalışmalar göz önüne alındığında metal sektörüne özel ML algoritmaları üzerine bir çalışma olmadığı için metal sektörü özelinde bu çalışma öncü olacaktır. İş kazalarına etki eden faktörler geleceğe yönelik olarak oluşturulacak iş kazası önleme politikalarına ve bu alanda verilecek kararlara veri sağlayacaktır.

MATERYAL VE METOT

Çalışmamızda öncelikle literatür taraması yapılmış, örnek çalışmalar incelenerek yapılacak çalışmanın amacı ve yöntemi belirlenmiştir. Daha sonra ÇSGB tarafından toplanan iş kazası verilerinin hepsi ayrıntılarıyla değerlendirilmiştir. Verilerin ön hazırlıkları ve temizleme işlemi yapılmıştır. Ardından ölümlü ve uzuv kayıplı iş kazaları için RF algoritması kullanılmıştır. Performans doğrulaması için ise çapraz doğrulama tekniği kullanılmıştır. Sonuç olarak, iş sağlığı ve güvenliği alanında çalışanların ve kamu görevlilerinin, tahmin edici bir algoritma kullanarak ciddi kazaların sebep ve sonuçlarını sınıflandırabilecekleri bir süreç önerilmiştir. Araştırma süresince izlenen yol Şekil 1' de gösterilmiştir.



Şekil 1. Araştırma Akış Şeması

Veri seti olarak İş Sağlığı ve Güvenliği Bilgi Yönetim Sistemi aracılığıyla ÇSGB tarafından kayıt altına alınan, 2013-2018 yılları arasında gerçekleşen metal sektörü iş kazası verileri (1542 adet) kullanılmıştır (ÇSGB, "İş Sağlığı ve Güvenliği Bilgi Yönetim Sistemi, "). Bu iş kazalarının 397 tanesi ölüme, 1145 tanesi ise uzuv kaybına sebep olmuştur. ÇSGB iş kazası bildirim formunu Avrupa İş Kazaları İstatistikleri metodolojisi (ESAW) ile elektronik ortamda kayıt altına almaktadır. ESAW metodolojisi iş kazası bildiriminde hangi parametrelerin ne şekilde bildirilmesi gerektiğini belirlemektedir. Bildirici her başlığın altında uygun olan seçeneği seçerek kaza bildirimini tamamlamaktadır. Bu çalışmada uzuv kaybına veya ölüme neden olan iş kazalarının bildircileri tarafından kazanın nasıl gerçekleştiğinin anlatıldığı açıklamaları ayrıntılı şekilde incelenmiş ve kazaların tehlike kaynakları birincil ve ikincil olarak belirlenmiştir. Ayrıca kaza sırasında makine

kullanılmışsa ve bu durum kaza açıklamasında belirtilmişse, makine bilgisinin kontrolü yapılmış ve bu veride eksiklik varsa kullanılan makine bilgisi hassasiyetle incelenerek eklenmiştir. ESAW parametrelerine ek olarak, tehlike kaynağı parametreleri de ML çalışmalarında da kullanılmıştır. Bu çalışmada, ML için 20 girdi değişkeni kullanılmış ve bu değişkenler Çizelge 1'de gösterilmiştir.

Çizelge 1. İş Kazası İstatistikleri İçin Değişkenler

Değişken	Açıklama
Kaza durumu	Uzun kayıplı veya ölümlü
Kaza tehlike kaynağı	Çizelge 2'de gösterilmiştir
Kaza tehlike kaynağı (2)	Çizelge 2'de gösterilmiştir (eğer varsa)
Kazanın gerçekleştiği yıl	
Kazanın gerçekleştiği ay	Kaza tarihi
Kazanın gerçekleştiği gün	
Makine	Makine ismi
Kaza yeri	ESAW değişkenleriyle belirlenir (Çalışılan ortam belirtilmemiş, sürekli olarak çalıştığı sabit işyeri, sabit olmayan geçici işyeri, diğer çalışılan ortam değişkenleri seçilir)
Kaza sebebi sapma kodu	ESAW değişkenleriyle belirlenir ("normalden farklı olan ve kazaya yol açan son olay."(Commission, 2012))
Yaralanma türü	ESAW değişkenleriyle belirlenir (Kırıklar, burkulmalar, incinmeler ve kimyasal yanıklar gibi fiziksel sonuçlar)
Yaralanmanın vücuttaki yeri	ESAW değişkenleriyle belirlenir (en fazla yaralanan yer)
Yaralanma sebebi	ESAW değişkenleriyle belirlenir
İş günü kaybı	İşe dönüş tarihi baz alınarak belirlenir
İSG eğitim durumu	Eğitim durumu "Evet" veya "Hayır" olarak girilir
Mesleki eğitim durumu	Eğitim durumu "Evet" veya "Hayır" olarak girilir
Kazazedenin cinsiyeti	Cinsiyet bilgisi (Erkek/Kadın)
Kazazedenin medeni durumu	Medeni durum bilgisi (Evli, bekar, dul, boşanmış ve bilinmiyor)
Çalışanın eğitim durumu	8 sınıflandırma bulunmaktadır (İlköğretim mezunu, Ortaokul mezunu, Lise mezunu, Üniversite mezunu, yüksek lisans, doktora, okuma-yazma bilen, okuma-yazma bilmeyen)
Kazazedenin yaşı	Kaza yaşandığında çalışanın yaşı
Kazazedenin iş hayatındaki toplam tecrübesi	Çalışanın iş hayatına başladığı günden kaza tarihine kadar geçen süre yıllık olarak gösterilmektedir
Kazazedenin kaza yaşadığı işyerindeki tecrübesi	Çalışanın kazanın olduğu işyerine geldiği günden kazanın olduğu tarihe kadar geçen süre yıllık olarak gösterilir

İş kazası bildircileri sisteme kazanın nasıl gerçekleştiği ile ilgili açıklamalar yazmaktadır. Bu açıklamalar ayrıntılı olarak incelenerek, Çizelge 2' de gösterilen sınıflandırmaya göre kaza nedenleri tehlike kaynaklarına göre sınıflandırılmıştır. İş kazası tehlike kaynakları (1) ve kazazedenin durumu çıktı değişkeni, diğerleri ise girdi değişkeni olarak alınmıştır.

Çizelge 2. Tehlike Kaynakları Sınıflandırması

Tehlike Kaynağı	Örnekler
Kimyasal Faktörler	Kimyasalların depolanması ve kullanımı vb.
Biyolojik Faktörler	Bulaşıcı hastalıklar, kişisel hijyen, mutfak hijyeni, hastalıklar vb.
Mekanik Faktörler	Makine, tezgahların ve el aletlerinin kullanımı, vb.
Elektrik Faktörleri	Prizlerin, kabloların, sigorta kutularının ve elektrik panolarının kullanımı, tesisatın topraklanması, elektrikli ekipmanlarla çalışma vb.
İnsan Kaynaklı Faktörler	Trafik kurallarına uymamak, KKD kullanmamak, İş ekipmanlarını güvensiz şekilde kullanmak, çalışanın izinsiz olduğu işlerde çalışması vb.
Ergonomik Faktörler	Yük kaldırma-taşıma, itme-çekme hareketleri, uygun olmayan vücut duruşları, tekrarlayan hareketler vb.
Organizasyonel Faktörler	İSG hizmetleri ve eğitimleri, sağlık gözetimi, çalışan seçimi, çalışma saatleri ve

Nakliye Kaynaklı Faktörler	molalar, personel uyumsuzlukları Çalışanların ve malzemelerin fabrikaya veya başka bir yere taşınması, malzeme taşıma, yükleme işlemlerini içeren tüm işler
Çalışma İzin Sistemi Faktörleri	Yüksekte veya tek başına çalışma, bakım onarım işleri, etiketleme-kilitleme sistemleri vb.
Genel Faktörler	Zeminler, yollar, merdivenler, korkuluklar, düzen ve temizlik, ikaz ve ikaz levhaları vb.
Diğer Faktörler	Yukarıda yer almayan kaynaklar

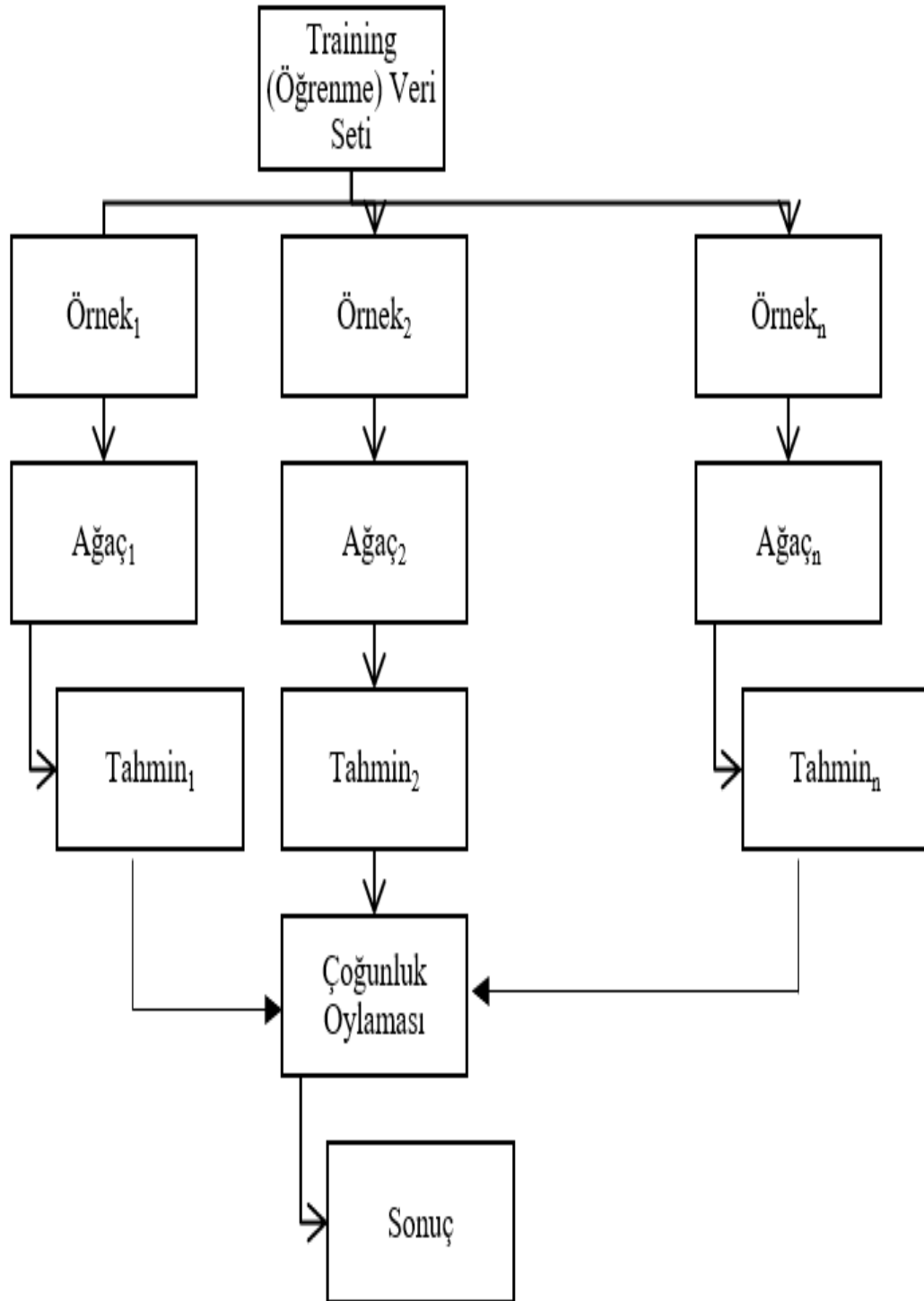
Makine Öğrenimi

ML, kendilerine beslenen verilere dayalı olarak öğrenebilen, ayarlayabilen ve iyileştirebilen sistemler tasarlama yöntemidir. ML, genellikle büyük veri olarak adlandırılan büyük miktarda veriyi analiz eden ve bunlardan modeller oluşturan bir yapay zekâ uygulamasıdır. ML yapay zekanın bir alt dalıdır ve çoğu zaman yapay zekâ teknolojisinin temelidir. ML yöntemleri, verileri anlamaya ve verilerden tahminler yapmaya yardımcı olan karmaşık modeller ve algoritmalar geliştirmek için kullanılmaktadır. ML uygulamaları, geçmiş verilerdeki eğilimlerden ve ilişkilerden öğrenerek araştırmacıların tahminler yapmasına olanak sağlamaktadır. Algoritmalar, sistemden daha fazla veri akarken öğrenmek ve geliştirmek için tasarlanmıştır. ML öncelikle karmaşık verileri anlamlandırmakla ilgilenir. Seçim sonuçlarını tahmin etmek, e-postadan gelen istenmeyen mesajları belirlemek ve filtrelemek, suç oluşturacak faaliyetleri öngörmek, trafik sinyallerini yol koşullarına göre otomatikleştirmek, fırtınalar ve doğal afetler için finansal tahminler üretmek, kazaların sebeplerini ve oluşma ihtimallerini belirlemek, müşteri kayıplarını incelemek, otomatik pilotlu uçaklar tasarlamak ML uygulamalarına bazı örneklerdir (Lantz, 2013; Mining, 2020; Sahay,2021).

RF, ML' de en çok kullanılan ve başarı oranı en yüksek olan algoritmalarındandır (Siddiqui ve ark., 2012; Shanthi ve Ramani, 2012; Wang ve ark., 2015; Brown, 2016; Tixier ve ark., 2016; Mafi ve ark., 2018; Zhang ve ark., 2018; Jiang ve ark. 2019; Kang ve Ryu, 2019; Tang ve ark., 2019; Umer ve ark., 2020).

RF algoritması

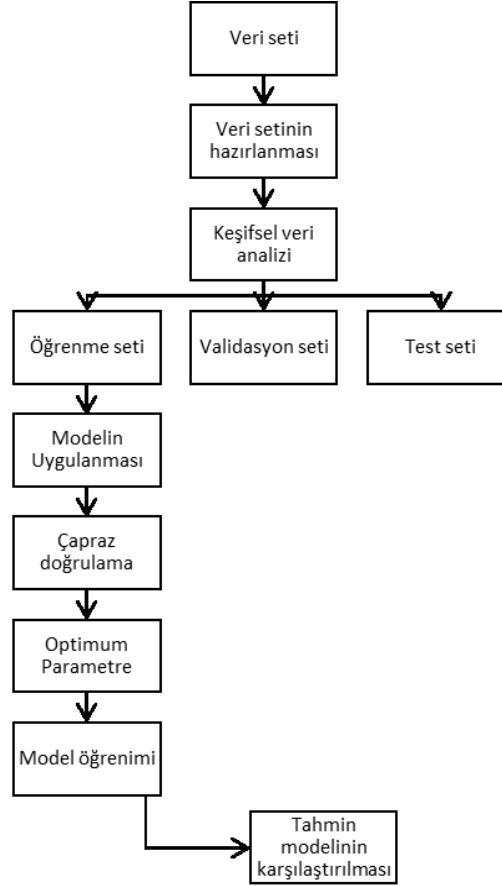
RF, sınıflandırma için birçok karar ağacı kullanan ve Breiman tarafından oluşturulmuş bir algoritmadır. RF, ağaç tahmincilerini, her ağaç bağımsız olarak örneklenmiş bir rastgele vektörün değerlerine bağlı olacak ve ormandaki tüm ağaçlar aynı dağılıma sahip olacak şekilde birleştirir. RF algoritması, doğruluk değerinde genellikle en iyi performansı göstermesi ve aşırı öğrenmeyi (overfitting) en aza indirmesi sebebiyle, ML alanında en çok kullanılan algoritmalarından birisidir (Breiman, 2001; Shanthi ve Ramani, 2012; Siddiqui ve ark., 2012; Wang ve ark., 2015; Brown, 2016; Tixier ve ark., 2016; Mafi ve ark., 2018; Zhang ve ark., 2018; Jiang ve ark., 2019; Kang ve Ryu, 2019; Tang ve ark., 2019; Umer ve ark., 2020; Zhen ve ark., 2023). RF karmaşık ara bağlantılara sahip büyük veri kümelerini etkili bir şekilde yönetir ve güvenilirdir (Strobl ve ark., 2008; Andriyas ve McKee, 2013). RF algoritması Şekil 2' de özetlenmiştir.



Şekil 2. RF Algoritması

RF, birden çok farklı karar ağacı kullanarak birden çok sonuç çıkarabilmekte ve bunları, verilen kararların etrafındaki diğer kararlar tarafından korunmasını sağlayacak şekilde kullanabilmektedir. Başka bir deyişle, karar verilmesine yardımcı olabilecek ne kadar çok karar ağacı varsa, bulguların daha doğru olması için bir şeylerin yanlış olma riskini azaltmada o kadar başarılı olmaktadır. Ağaçlar, sürekli olarak aynı yönde hareket etmedikleri veya girdilerde çok fazla örtüşmedikleri sürece, temelde birbirlerini bireysel hatalarından korumaktadırlar. Bu işlem, RF algoritmasının başarı performansını artırmaktadır (Lantz, 2013; Mining, 2020; Sahay,2021).

RF algoritması ML alanında en sık kullanılan ve en başarılı yöntem olduğu için gerçekleştireceğimiz iş kazası tahminleme çalışmalarında kullanılmış ve ML algoritmasının adımları Şekil 3'te verilmiştir.



Şekil 3. ML Şeması

Parametre tahmini ve doğrulama için kullanılan kaza veri setindeki verilerin %80' ini öğrenme aşaması için rastgele seçilmiş ve veri setinin kalan %20' si ise doğrulama (validasyon) için kullanılmıştır. Yukarıda belirtilen aşamalar 50 kere tekrar edilmiştir.

K-katlı çapraz doğrulama

ML algoritmaların da çapraz doğrulama, performans değerlendirmesi için kullanılan önemli bir yöntemdir. k-katlı çapraz doğrulamanın amacı; örneği k adet gruba bölerek ve her grubu bir doğrulama örneği (veya test veri seti) olarak ele alarak algoritmayı değerlendirmektir (Li ve ark., 2023). Son dönemlerde yapılan çalışmalar k-katlı çapraz doğrulama yönteminin, en çok kullanılan doğrulama yöntemlerinden biri olduğunu göstermektedir (Azadi ve Karimi-Jashni, 2016; Vu ve ark., 2022; Li ve ark., 2023). Örneğin, Garre ve arkadaşları, aşırı öğrenmeyi en aza indirmek için 10-katlı çapraz doğrulama yöntemini kullanmıştır. Sonuçları, ML algoritmasının 10-katlı çapraz doğrulama kullanıldığında daha iyi performans göstermiştir (Garre ve ark., 2020).

İş kazası verilerinin analizi için RF algoritmasının doğruluk değeri performansını artırmak için 10-katlı katlı çapraz doğrulama yöntemi kullanılmıştır.

RF algoritmasının performans ölçütü

Doğruluk değeri, dört istatistik kullanılarak hesaplanmasıyla performans ölçütü olarak değerlendirilir. Bu dört istatistik gerçekten gözlemlenen pozitif olarak tahmin edilen nesnelere sayısını temsil eden doğru pozitifler (TP), pozitif olarak tahmin edilen nesnelere sayısını temsil eden yanlış negatifler (FN), yanlış tahmin edilen negatif nesnelere sayısını temsil eden yanlış pozitifler

(FP) ve gerçekte gözlemlenen pozitif olarak tahmin edilen nesnelere sayısını temsil eden doğru negatifler (TN) şeklindedir (Gu ve ark., 2009). Doğruluk değeri aşağıda verilen denklem 1 ile hesaplanır:

$$\text{Doğruluk} = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

Doğruluk; ML' de performans karşılaştırılması için en çok kullanılan ölçüttür (Gu ve ark., 2009). Veri setinde çok fazla değişken olması ve iki farklı değişken için tahminleme yapılması sebebiyle ROC (Alıcı işlem karakteristikleri, Receiver Operating Characteristic) eğrileriyle değerlendirmek yerine, bu çalışmada, ML algoritmasının performans ölçütü olarak doğruluk değeri kullanılmıştır.

Değişken önceliklendirilmesi

Algoritmayı değerlendirmek ve iş kazası nedenlerini etkileyen bileşenleri analiz etmek için değişken önceliklendirilmesi yapılmalıdır. RF algoritması, birkaç karar ağacını kullanarak öngörülebilirliği artıran bir torbalama tekniği kullanır, fakat algoritmayı anlamak zordur. Bununla birlikte, değişken önceliklendirmesi, RF algoritmasını yorumlamak içinde kullanılır. Bu çalışmada özyinelemeli olarak özellik seçme (RFE) metodu uygulanmıştır. RFE değişken önemini değerlendirmek ve özellik seçimini gerçekleştirmek için kullanılan önemli bir metottur (Guyon ve ark., 2002). RFE, en uygun özellik alt kümelerini seçer ve belirli bir ML yöntemine dayalı olarak çok sayıda özelliği sıralayarak sınıflandırma performansını artırmaktadır (Li ve ark., 2022). RFE ilgisiz girdileri kaldırarak algoritma karmaşıklığını azaltmak için etkili ve verimli bir tekniktir (Guyon ve ark., 2002). Birçok ML algoritmalarında RFE tekniği sayesinde, özellik seçimi yapılmış ve performans değerini artırdığı görülmüştür (Freund ve Schapire, 1996; Friedman, 2000; Breiman, 2001).

BULGULAR VE TARTIŞMA

2013-2018 yılları arasında metal sektöründe yaşanmış 1542 adet uzuv kayıplı veya ölümlü iş kazası verileri tehlike kaynağı tahminlemesi yapılması amacıyla RF algoritmasıyla değerlendirilmiştir. Çalışmamızın bir diğer amacı da ulusal iş kazası verileri kullanılarak gelecekte meydana gelebilecek kazaların neden ve sonuçlarının tahmin edilmesi ve bu sonuçlara göre oluşturulacak politikalara veri sağlanmasıdır. Uygulanan ML yöntemini şirketler kendi kaza verileri için de kullanarak tahminlerde bulunabilir ve çıkan sonuçlara göre önleyici tedbirler alabilirler. RF algoritmasıyla elde edilen iş kazalarının tehlike kaynaklarına yönelik veriler, hazırlanacak ulusal politikalarla iş kazalarının önlenmesi sağlanacaktır.

İş Kazalarının Dağılımı

İş kazalarının nedenlerinin tehlike kaynağı dağılımı Çizelge 3'te gösterilmiştir.

Çizelge 3. İş Kazalarının Dağılımı

Tehlike Kaynakları	Uzuv kayıplı kazalar	Ölümlü kazalar	Toplam
Kimyasal Faktörler	13	16	29
Biyolojik Faktörler	2	55	57
Mekanik Faktörler	976	99	1075
Elektrik Faktörleri	3	26	29
İnsan Kaynaklı Faktörler	22	9	31
Ergonomik Faktörler	2	1	3
Organizasyonel Faktörler	2	4	6
Nakliye Kaynaklı Faktörler	49	74	123
Çalışma İzin Sistemi Faktörleri	48	83	131
Genel Faktörler	24	20	44
Diğer Faktörler	4	10	14

Metal sektöründe yaşanan iş kazalarının tehlike kaynaklarının en fazla mekanik faktörler kaynaklı olması bu sektörde makinelerin ve el aletlerinin çokça kullanılmasından dolayıdır. Çıkan sonuçlar incelendiğinde mekanik faktörlü kazaların metal sektöründe çok fazla iş kazasına sebep olduğu ve bu alanda ivedilikle çalışmalar yapılmasının gerekliliği görülmüştür.

RF algoritmasının performans ölçüt sonucu

Çalışmamızda ilk olarak 10-katlı çapraz doğrulama yöntemi uygulanmadan, tüm değişkenlerle RF algoritması uygulanmış ve programın verdiği doğruluk değeri sonuçları Çizelge 4'te verilmiştir.

Çizelge 4. RF Algoritmasının Doğruluk Değeri

mtry	Doğruluk	Kappa
2	0.6409997	0.04680341
118	0.8809192	0.79166057
234	0.8760542	0.78634286

Bu durumda RF algoritmasının doğruluk değeri 0.8760 olarak bulunmuştur. Çalışmanın ikinci aşamasında, RF algoritması için 10-katlı çapraz doğrulama yöntemi uygulanmıştır. Hesaplanan yeni doğruluk değeri 0.9172 olarak bulunmuştur. 10-katlı çapraz doğrulama uygulanarak RF algoritmasının doğruluk değerinde %4.7 oranında artış sağlanmıştır. Veri seti bu yöntem ile 10 eşit parçaya bölünerek, her bir parçanın hem öğrenme hem de test seti için kullanılması sağlanmıştır. Bu şekilde de dağılım ve parçalanmadan kaynaklanan sapma ve hatalar minimize edilerek doğruluk değeri artırılmıştır. Bu oran ile çok başarılı bir doğruluk oranı elde etmiştir.

Değişken önceliklendirilmesi

Değişken önceliklendirmesi çalışması için RFE yöntemi uygulanmış ve 20 değişken başlığı içerisinde öncelikli 10 değişken ve her bir değişken performans değeri üzerindeki etkisi yüzde olarak Çizelge 5' de verilmiştir.

Çizelge 5. Öncelikli değişkenler

	Özellik	Yüzde etkisi
1	Kaza tehlike kaynağı (2)	%27
2	İş günü kaybı	%16
3	Kaza sebebi sapma kodu	%12
4	Makine	%11
5	Yaralanmanın vücuttaki yeri	%9
6	Yaralanma türü	%5
7	Yaralanma sebebi	%4
8	Yaş	%1
9	Kazazedenin kaza yaşadığı işyerindeki tecrübesi	%0.8
10	Kazazedenin iş hayatındaki toplam tecrübesi	%0.74
11	Diğerleri	%13.46

20 tane değişken içerisinde sonuca etkisi %5' in üzerinde olan değişkenler (Kaza tehlike kaynağı (2), İş günü kaybı, Kaza sebebi sapma kodu, Makine, Yaralanmanın vücuttaki yeri) kullanılarak RF algoritması tekrar uygulanmıştır. Uygulanan RFE yönteminin RF algoritmasının doğruluk değerinde değişikliğe yol açmamıştır. Bunun sebebi ilk beş girdinin toplam değere etkisi %75'tir ve bu oldukça yüksek bir ortalamadır. Bu sebeple algoritma zaten gerekli girdileri kullanarak, ulaşabileceği en iyi doğruluk değerine ulaştığı için girdi seçilimi bir değişikliğe sebep olmamıştır. ÇSGB tarafından iş kazası verileri ESAW metodolojisi ile kayıt altına alındığı ve sistemin değiştirilmesi zor olması

sebebiyle bu sonuç ileride gerçekleştirilecek çalışmalar için güzel bir bulgudur. Uygulanmış olan ML algoritmasının, diğer sektörlerdeki iş kazası verilerine de uygulanabilir olduğunun göstergesidir.

Metal sektöründe gerçekleşen, iş kazalarının bildircileri tarafından yazılan açıklamaları ayrıntılı şekilde irdelenerek tehlike kaynak veya kaynakları belirlenmiştir. Tehlike kaynağı bilgisi ESAW metodolojisine yer almamaktadır. Fakat ESAW metodolojisinin istediği bir veri olmayan ikincil tehlike kaynağının tahminleme sürecinde en önemli etkenlerden biri olduğu görülmektedir. Bu bilgiler ışığında; ulusal bir veri setinde geleceğe yönelik tahminler yapılırken ESAW metodolojisinin geliştirilmesinin ya da ek olarak bazı ekstra bilgilerin de edinilmesinin gerekliliği görülmüştür. Kazaların kök nedenlerinin araştırılarak iş kazası veri setine eklenmesi, gelecekte yaşanma ihtimali olan kazalara yönelik bir tahmin yaparken kullanılacak önemli bulgular sağlayacaktır. Ortaya çıkan özellik önceliklendirilmelerinden, kaza tarihi ve çalışanların eğitim durumu değişkenlerinin iş kazası nedeni üzerinde çok az etkisinin olduğu veya hiç etkisinin olmadığı bulunmuştur.

SONUÇ

2013-2018 yılları arasında, metal sektöründe yaşanmış 1542 adet uzuv kayıplı ve ölümlü iş kazası; tehlike kaynağı tahminlemesi amacıyla RF algoritması ile değerlendirilmiştir. Araştırmanın ilk aşamasında RF algoritması, tüm verilere herhangi bir doğrulama yapılmadan çalışılmış olup akabinde 10-katlı çapraz doğrulama yöntemi uygulanmıştır. 10-katlı çapraz doğrulama uygulanarak RF algoritmasının doğruluk değerinde %4.7 artış sağlanarak, RF algoritmasının doğruluk değeri 0.9172 olarak hesaplanmıştır. Hesaplanan doğruluk değeri oldukça başarılı ve yüksek sonuçtur. ESAW metodolojisindeki başlıklar özelinde iş kazası verisi toplanması sebebiyle, kazaların kök nedeni yerine kazaya sebep olan son sapma kodu verisi toplanmaktadır. Doğruluk değerinin performansını etkileyen en önemli parametrelerden birisi ise kaza açıklamaları incelenerek belirlenen kök sebeplerdir. Bu sebeple; çalışmanın en büyük kısıtı kök nedenin araştırmacı tarafından okunarak belirlenmesidir. Gelecekteki araştırmalarda, iş kazalarını azaltmak için iş kazası nedenlerini ve sonuçlarını gerçek zamanlı olarak tahmin etmek için veri toplayacak ve ulusal düzeyde kullanılacak yöntemler geliştireceğiz. Tüm bunlara ek olarak, farklı sektörlerde yaşanan iş kazası verilerini de kullanarak aynı çalışmanın uygulanmasını önermekteyiz.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

KAYNAKLAR

- Aci, C., & Ozden, C. (2018). Predicting the Severity of Motor Vehicle Accident Injuries in Adana-Turkey Using Machine Learning Methods and Detailed Meteorological Data. *International Journal of Intelligent Systems and Applications in Engineering*, 6(1), 72-79. doi:10.18201/ijisae.2018637934
- Alizadeh, S. S., Mortazavi, S. B., & Mehdi Sepehri, M. (2015). Assessment of accident severity in the construction industry using the Bayesian theorem. *International Journal of Occupational Safety and Ergonomics*, 21(4), 551-557. doi:10.1080/10803548.2015.1095546
- Amiri, M., Ardeshtir, A., Fazel Zarandi, M. H., ve Soltanaghaei, E. (2016). Pattern Extraction For High-Risk Accidents In The Construction Industry: A Data-Mining Approach. *International Journal Of Injury Control And Safety Promotion*, 23(3), 264-276. doi:10.1080/17457300.2015.1032979
- Andriyas, S., ve McKee, M. (2013). Recursive Partitioning Techniques For Modeling Irrigation Behavior. *Environmental Modelling & Software*, 47, 207-217. doi:https://doi.org/10.1016/j.envsoft.2013.05.011

- Anyfantis, I., Leka, S., Reniers, G., ve Boustras, G. (2021). Employers' Perceived Importance And The Use (Or Non-Use) Of Workplace Risk Assessment In Micro-Sized And Small Enterprises In Europe With Focus On Cyprus. *Safety Science*, 139, 105256. doi:10.1016/j.ssci.2021.105256
- Ayhan, B. U., ve Tokdemir, O. B. (2019). Predicting The Outcome of Construction Incidents. *Safety Science*, 113, 91-104. doi:https://doi.org/10.1016/j.ssci.2018.11.001
- Azadi, S., ve Karimi-Jashni, A. (2016). Verifying The Performance of Artificial Neural Network And Multiple Linear Regression In Predicting The Mean Seasonal Municipal Solid Waste Generation Rate: A Case Study Of Fars Province, Iran. *Waste Management*, 48, 14-23. doi:https://doi.org/10.1016/j.wasman.2015.09.034
- Bazargan, M., ve Guzha, V. S. (2011). Impact Of Gender, Age and Experience Of Pilots On General Aviation Accidents. *Accident Analysis & Prevention*, 43(3), 962-970. doi:https://doi.org/10.1016/j.aap.2010.11.023
- Bevilacqua, M., Ciarapica, F. E., ve Giacchetta, G. (2008). Industrial And Occupational Ergonomics in The Petrochemical Process Industry: A Regression Trees Approach. *Accident Analysis & Prevention*, 40(4), 1468-1479. doi:https://doi.org/10.1016/j.aap.2008.03.012
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. doi:10.1023/A:1010933404324
- Brown, D. E. (2016). Text Mining the Contributors to Rail Accidents. *IEEE Transactions on Intelligent Transportation Systems*, 17(2), 346-355. doi:10.1109/TITS.2015.2472580
- Cheng, C.-W., Leu, S.-S., Cheng, Y.-M., Wu, T.-C., ve Lin, C.-C. (2012). Applying Data Mining Techniques To Explore Factors Contributing To Occupational Injuries In Taiwan's Construction Industry. *Accident Analysis & Prevention*, 48, 214-222. doi:https://doi.org/10.1016/j.aap.2011.04.014
- Chiang, Y.-H., Wong, F., ve Liang, S. (2018). Fatal Construction Accidents in Hong Kong. *Journal of Construction Engineering and Management*, 144. doi:10.1061/(ASCE)CO.1943-7862.0001433
- Commission, E. (2012). European Statistics on Accidents at Work (ESAW) — Summary methodology. In E. Commission (Ed.). Luxembourg Publications Office of the European Union.
- Freund, Y., ve Schapire, R. E. (1996). Experiments With A New Boosting Algorithm. Paper presented at the icml.
- Friedman, J. (2000). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29. doi:10.1214/aos/1013203451
- Fuentes-Bargues, J. L., Sánchez-Lite, A., González-Gaya, C., Victor Fco, R.-P., ve Reniers, G. (2022). A study of situational circumstances related to Spain's occupational accident rates in the metal sector from 2009 to 2019. *Safety Science*, 150, 105700. doi:https://doi.org/10.1016/j.ssci.2022.105700
- Garre, A., Ruiz, M. C., ve Hontoria, E. (2020). Application Of Machine Learning To Support Production Planning Of A Food Industry In The Context Of Waste Generation Under Uncertainty. *Operations Research Perspectives*, 7, 100147. doi:https://doi.org/10.1016/j.orp.2020.100147
- Ghodrati, N., Yiu, T. W., Wilkinson, S., ve Shahbazzpour, M. (2018). A New Approach To Predict Safety Outcomes In The Construction Industry. *Safety Science*, 109, 86-94. doi:https://doi.org/10.1016/j.ssci.2018.05.016
- Goh, Y. M., ve Ubeynarayana, C. (2017). Construction Accident Narrative Classification: An Evaluation Of Text Mining Techniques. *Accident; Analysis and Prevention*, 108, 122-130. doi:10.1016/j.aap.2017.08.026
- Gregoriades, A., ve Mouskos, K. C. (2013). Black Spots Identification Through A Bayesian Networks Quantification Of Accident Risk Index. *Transportation Research Part C: Emerging Technologies*, 28, 28-43. doi:https://doi.org/10.1016/j.trc.2012.12.008
- Gu, Q., Zhu, L., ve Cai, Z. (2009, 2009//). Evaluation Measures of the Classification Performance of Imbalanced Data Sets. Paper presented at the *Computational Intelligence and Intelligent Systems*, Berlin, Heidelberg.
- Gulhan, B., Ilhan, M., ve Civil, E. (2012). Occupational Accidents And Affecting Factors Of Metal Industry In A Factory In Ankara. *Turkish Journal of Public Health*, 10.
- Güllüoğlu, E., ve Güllüoğlu, A. (2019). Türkiye'de Metal Sektöründe Meydana Gelen İş Kazalarının Analizi. *International Journal of Advances in Engineering and Pure Sciences*. doi:10.7240/jeps.486478
- Guyon, I., Weston, J., Barnhill, S., ve Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46(1), 389-422. doi:10.1023/A:1012487302797
- He, X., Chen, W., Nie, B., ve Zhang, M. (2010). Classification Technique For Danger Classes Of Coal And Gas Outburst In Deep Coal Mines. *Safety Science*, 48(2), 173-178. doi:https://doi.org/10.1016/j.ssci.2009.07.007
- ILO. (2023). Safety and health at work.
- İş Sağlığı ve Güvenliği Bilgi Yönetim Sistemi. Retrieved from <https://ibys.csgb.gov.tr/>
- Jahangiri, M., Solukloei, H. R. J., ve Kamalinia, M. (2019). A Neuro-Fuzzy Risk Prediction Methodology For Falling From Scaffold. *Safety Science*, 117, 88-99. doi:https://doi.org/10.1016/j.ssci.2019.04.009

- Jana, D. K., Pramanik, S., Sahoo, P., ve Mukherjee, A. (2019). Interval Type-2 Fuzzy Logic And Its Application To Occupational Safety Risk Performance In Industries. *Soft Computing*, 23(2), 557-567. doi:10.1007/s00500-017-2860-8
- Jeong, H., Jang, Y., Bowman, P. J., ve Masoud, N. (2018). Classification Of Motor Vehicle Crash Injury Severity: A Hybrid Approach For Imbalanced Data. *Accident Analysis & Prevention*, 120, 250-261. doi:https://doi.org/10.1016/j.aap.2018.08.025
- Jiang, L., Xie, Y., ve Ren, T. (2019). Modelling Highly Unbalanced Crash Injury Severity Data By Ensemble Methods And Global Sensitivity Analysis. Paper presented at the *Proceedings of the Transportation Research Board 98th Annual Meeting*, Washington, DC, USA.
- Kang, K., ve Ryu, H. (2019). Predicting Types Of Occupational Accidents At Construction Sites In Korea Using Random Forest Model. *Safety Science*, 120, 226-236. doi:https://doi.org/10.1016/j.ssci.2019.06.034
- Karacasu, M., Ergül, B., ve Altın Yavuz, A. (2014). Estimating The Causes of Traffic Accidents Using Logistic Regression And Discriminant Analysis. *International Journal of Injury Control And Safety Promotion*, 21(4), 305-313. doi:10.1080/17457300.2013.815632
- Karlaftis, M. G., ve Golias, I. (2002). Effects Of Road Geometry And Traffic Volumes On Rural Roadway Accident Rates. *Accident Analysis & Prevention*, 34(3), 357-365. doi:https://doi.org/10.1016/S0001-4575(01)00033-1
- Kifle, M., Engdaw, D., Alemu, K., Sharma, H. R., Amsalu, S., Feleke, A., ve Worku, W. (2014). Work Related Injuries And Associated Risk Factors Among Iron And Steel Industries Workers In Addis Ababa, Ethiopia. *Safety Science*, 63, 211-216. doi:https://doi.org/10.1016/j.ssci.2013.11.020
- Lantz, B., *Machine Learning with R*. 2013: Packt Publishing.
- Leu, S.-S., ve Chang, C.-M. (2013). Bayesian-Network-Based Safety Risk Assessment For Steel Construction Projects. *Accident Analysis & Prevention*, 54, 122-133. doi:https://doi.org/10.1016/j.aap.2013.02.019
- Li, G., Baker, S. P., Grabowski, J. G., Qiang, Y., McCarthy, M. L., ve Rebok, G. W. (2003). Age, Flight Experience, and Risk of Crash Involvement in a Cohort of Professional Pilots. *American Journal of Epidemiology*, 157(10), 874-880. doi:10.1093/aje/kwg071
- Li, J., Gao, F., Lin, S., Guo, M., Li, Y., Liu, H., Wen, Q. (2023). Quantum k-fold Cross-Validation for Nearest Neighbor Classification Algorithm. *Physica A: Statistical Mechanics and its Applications*, 611, 128435. doi:https://doi.org/10.1016/j.physa.2022.128435
- Li, L., Ching, W.-K., ve Liu, Z.-P. (2022). Robust Biomarker Screening From Gene Expression Data By Stable Machine Learning-Recursive Feature Elimination Methods. *Computational Biology and Chemistry*, 100, 107747. doi:10.1016/j.compbiolchem.2022.107747
- Lindberg, A.-K., Hansson, S. O., ve Rollenhagen, C. (2010). Learning from Accidents – What More Do We Need to Know?. *Safety Science*, 48, 714-721. doi:10.1016/j.ssci.2010.02.004
- Mafi, S., AbdelRazig, Y., ve Doczy, R. (2018). Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups. *Transportation Research Record*, 2672(38), 171-183. doi:10.1177/0361198118794292
- Matías, J. M., Rivas, T., Martín, J. E., ve Taboada, J. (2008). A Machine Learning Methodology For The Analysis Of Workplace Accidents. *International Journal of Computer Mathematics*, 85(3-4), 559-578. doi:10.1080/00207160701297346
- Meng, Q., & Weng, J. (2011). A Genetic Algorithm Approach To Assessing Work Zone Casualty Risk. *Safety Science*, 49(8), 1283-1288. doi:https://doi.org/10.1016/j.ssci.2011.05.001
- Müstikoğlu, G., Gerek, I. H., Erdis, E., Mumtaz Usmen, P. E., Cakan, H., ve Kazan, E. E. (2015). Decision tree analysis of construction fall accidents involving Roofers. *Expert Systems with Applications*, 42(4), 2256-2263. doi:https://doi.org/10.1016/j.eswa.2014.10.009
- Mining, E., *Machine Learning for Beginners: A Complete and Phased Beginner's Guide to Learning and Understanding Machine Learning and Artificial Intelligence*. 2020: Everooks Limited.
- Nazaripour, E., Halvani, G., Jahangiri, M., Fallahzadeh, H., ve Mohammadzadeh, M. (2018). Safety Performance Evaluation In A Steel Industry: A Short-Term Time Series Approach. *Safety Science*, 110, 285-290. doi:https://doi.org/10.1016/j.ssci.2018.08.028
- Nishimoto, T., Mukaigawa, K., Tominaga, S., Lubbe, N., Kiuchi, T., Motomura, T., ve Matsumoto, H. (2017). Serious Injury Prediction Algorithm Based On Large-Scale Data And Under-Triage Control. *Accident Analysis & Prevention*, 98, 266-276. doi:https://doi.org/10.1016/j.aap.2016.09.028
- Pacal, I. (2023). Göğüs Röntgeni Görüntülerinden Otomatik COVID-19 Teşhisi için Görü Transformatörüne Dayalı Bir Yaklaşım . *Journal of the Institute of Science and Technology*, 13 (2) , 778-791 . DOI: 10.21597/jist.1225156

- Palei, S. K., ve Das, S. K. (2009). Logistic Regression Model For Prediction Of Roof Fall Risks In Bord And Pillar Workings In Coal Mines: An Approach. *Safety Science*, 47(1), 88-96. doi:https://doi.org/10.1016/j.ssci.2008.01.002
- Park, J., Cho, C., Cho, Y., ve Kim, K. (2018). Data-Driven Monitoring System for Preventing the Collapse of Scaffolding Structures. *Journal of Construction Engineering and Management*, 144. doi:10.1061/(ASCE)CO.1943-7862.0001535
- Persona, A., Battini, D., Faccio, M., Bevilacqua, M., ve Ciarapica, F. E. (2006). Classification Of Occupational Injury Cases Using The Regression Tree Approach. *International Journal of Reliability, Quality and Safety Engineering*, 13(2), 171-191. doi:10.1142/S0218539306002197
- Rivas, T., Paz, M., Martín, J. E., Matías, J. M., García, J. F., ve Taboada, J. (2011). Explaining And Predicting Workplace Accidents Using Data-Mining Techniques. *Reliability Engineering & System Safety*, 96(7), 739-747. doi:https://doi.org/10.1016/j.ress.2011.03.006
- Sahay, A., Essentials of Data Science and Analytics: Statistical Tools, Machine Learning, and R-Statistical Software Overview. 2021: Business Expert Press.
- Sakhakarmi, S., Park, J., ve Cho, C. (2019). Enhanced Machine Learning Classification Accuracy for Scaffolding Safety Using Increased Features. *Journal of Construction Engineering and Management*, 145. doi:10.1061/(ASCE)CO.1943-7862.0001601
- Salguero-Caparros, F., Suarez-Cebador, M., ve Rubio-Romero, J. C. (2015). Analysis Of Investigation Reports On Occupational Accidents. *Safety Science*, 72, 329-336. doi:https://doi.org/10.1016/j.ssci.2014.10.005
- Sánchez, A., Riesgo Fernández, P., Sánchez-Lasheras, F., de Cos Juez, F., ve Garcia Nieto, P. J. (2011). Prediction Of Work-Related Accidents According To Working Conditions Using Support Vector Machines. *Applied Mathematics and Computation*, 218, 3539-3552. doi:10.1016/j.amc.2011.08.100
- Sanmiquel, L., Rossell, J. M., ve Vintro, C. (2015). Study Of Spanish Mining Accidents Using Data Mining Techniques. *Safety Science*, 75, 49-55. doi:https://doi.org/10.1016/j.ssci.2015.01.016
- Santos, K., Dias, J. P., ve Amado, C. (2022). A Literature Review Of Machine Learning Algorithms For Crash Injury Severity Prediction. *Journal of Safety Research*, 80, 254-269. doi:https://doi.org/10.1016/j.jsr.2021.12.007
- Shanthi, S., ve Ramani, R. G. (2012). Feature Relevance Analysis And Classification Of Road Traffic Accident Data Through Data Mining Techniques. *Proceedings of The World Congress on Engineering and Computer Science*, 1, 24-26.
- Shao, B., Hu, Z., Liu, Q., Chen, S., ve He, W. (2019). Fatal Accident Patterns Of Building Construction Activities In China. *Safety Science*, 111, 253-263. doi:https://doi.org/10.1016/j.ssci.2018.07.019
- Siddiqui, C., Abdel-Aty, M., ve Huang, H. (2012). Aggregate Nonparametric Safety Analysis Of Traffic Zones. *Accident Analysis & Prevention*, 45, 317-325. doi:https://doi.org/10.1016/j.aap.2011.07.019
- SGK, (2017). SGK İstatistik Yıllıkları.
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional Variable Importance For Random Forests. *BMC Bioinformatics*, 9. doi:10.1186/1471-2105-9-307
- Tang, J., Liang, J., Han, C., Li, Z., ve Huang, H. (2019). Crash Injury Severity Analysis Using A Two-Layer Stacking Framework. *Accident Analysis & Prevention*, 122, 226-238. doi:https://doi.org/10.1016/j.aap.2018.10.016
- Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., ve Bowman, D. (2016). Application Of Machine Learning To Construction Injury Prediction. *Automation in Construction*, 69, 102-114. doi:https://doi.org/10.1016/j.autcon.2016.05.016
- Umer, M., Sadiq, S., Ishaq, A., Ullah, D. S., Saher, N., ve Madni, H. (2020). Comparison Analysis of Tree Based and Ensembled Regression Algorithms for Traffic Accident Severity Prediction.
- Usman, T., Fu, L., ve Miranda-Moreno, L. F. (2016). Injury Severity Analysis: Comparison Of Multilevel Logistic Regression Models And Effects Of Collision Data Aggregation. *Journal of Modern Transportation*, 24(1), 73-87. doi:10.1007/s40534-016-0096-4
- Vu, L., Ng, K., Richter, A., ve An, C. (2022). Analysis Of Input Set Characteristics And Variances On K-Fold Cross Validation For A Recurrent Neural Network Model On Waste Disposal Rate Estimation. *Journal of Environmental Management*, 311, 114869. doi:10.1016/j.jenvman.2022.114869
- Wang, J., Liu, B., Fu, T., Liu, S., ve Stipancic, J. (2019). Modeling When And Where A Secondary Accident Occurs. *Accident Analysis & Prevention*, 130, 160-166. doi:https://doi.org/10.1016/j.aap.2018.01.024
- Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., ve Bai, X. (2015). Flood Hazard Risk Assessment Model Based On Random Forest. *Journal of Hydrology*, 527, 1130-1141. doi:https://doi.org/10.1016/j.jhydrol.2015.06.008

- Weng, J., Meng, Q., ve Wang, D. Z. W. (2012). Tree-Based Logistic Regression Approach for Work Zone Casualty Risk Assessment. *Risk analysis: an official publication of the Society for Risk Analysis*, 33. doi:10.1111/j.1539-6924.2012.01879.x
- Veziroğlu, E. , Pacal, I. & Coşkunçay, A. (2023). Derin Evrişimli Sinir Ağları Kullanılarak Pirinç Hastalıklarının Sınıflandırılması . *Journal of the Institute of Science and Technology* , 13 (2) , 792-814 . DOI: 10.21597/jist.1265769
- Yağimli, M., ve İzci, F. (2017). Türkiye’de Makine ve Teçhizatı Hariç Fabrikasyon Metal Ürünleri İmalatı Sektöründe Yaşanan İş Kazaları ve Ölümlü İş Kazası Sayılarının Tahmini. *Karaelmas İş Sağlığı ve Güvenliği Dergisi*, 1, 9-15. doi:10.33720/kisgd.322546
- Yan, X., Radwan, E., ve Abdel-Aty, M. (2005). Characteristics Of Rear-End Accidents At Signalized Intersections Using Multiple Logistic Regression Model. *Accident Analysis & Prevention*, 37(6), 983-995. doi:https://doi.org/10.1016/j.aap.2005.05.001
- Yannis, G., Papadimitriou, E., Dupont, E., ve Martensen, H. (2010). Estimation of Fatality and Injury Risk by Means of In-Depth Fatal Accident Investigation Data. *Traffic Injury Prevention*, 11, 492-502. doi:10.1080/15389588.2010.492536
- Yeoum, S., ve Lee, Y. (2013). A Study On Prediction Modeling Of Korea Military Aircraft Accident Occurrence. *The International Journal of Industrial Engineering: Theory, Applications and Practice*, 20, 562-573.
- Yi, W., Chan, A. P. C., Wang, X., ve Wang, J. (2016). Development Of An Early-Warning System For Site Work In Hot And Humid Environments: A Case Study. *Automation in Construction*, 62, 101-113. doi:https://doi.org/10.1016/j.autcon.2015.11.003
- Zhang, J., Li, Z., Pu, Z., ve Xu, C. (2018). Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods. *IEEE Access*, 6, 60079-60087. doi:10.1109/ACCESS.2018.2874979
- Zhen, X., Ning, Y., Du, W., ve Huang, Y. (2023). An interpretable and augmented machine-learning approach for causation analysis of major accident indicators in the offshore petroleum Industry. *Process Safety and Environmental Protection*. doi:https://doi.org/10.1016/j.psep.2023.03.063