



Veri Madenciliğinde Karar Ağacı Algoritmaları İle Demir Çelik Endüstrisinde İş Kazaları Üzerine Bir Uygulama

Neslihan Köse*, Filiz Ersöz²

¹ Karabük Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Karabük, Türkiye (ORCID: 0000-0002-0503-0247)

² Karabük Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Karabük, Türkiye (ORCID: 0000-0000-0000-0000)

(International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2020 – 22-24 Ekim 2020)

(DOI: 10.31590/ejosat.843387)

ATIF/REFERENCE: Köse, N. & Ersöz, F. (2020). Veri Madenciliğinde Karar Ağacı Algoritmaları İle Demir Çelik Endüstrisinde İş Kazaları Üzerine Bir Uygulama. *Avrupa Bilim ve Teknoloji Dergisi*, (Özel Sayı), 397-407.

Öz

Çelik endüstrisinde çalışanlar tesislerin yapısı, üretim akışı, fazla ve tekrarlayan iş öğeleri, üretim proseslerinin doğası gereği yüksek sıcaklık ve gürültülü iş ortamları sebebiyle sık sık iş kazası risk ile karşıya kalmaktadırlar. "İş kazası" kavram olarak; çalışma ortamında çalışanın karşılaştığı istenmeyen, beklenmeyen, ihmalkârlık, kusur, dikkatsizlik, kasıt ve şanssızlık sonucu meydana gelen olayları kapsamaktadır. Yaşanan iş kazaları, hem sıklığı hem de sonuçları nedeniyle işletmeler ve toplum nazarında çok önemli bir sorundur. İş kazalarının başlıca nedeninin personelin güvenli olmayan hareketleri olduğu konusunda yaygın bir görüş vardır. Ancak birçok çalışma, kazaların çoğunun çalışanların kişilik özellikleri ve uygun olmayan ortam koşulları ile ilgili olduğunu göstermiştir. Bu çalışmada, bir demir çelik fabrikasında yaşanan iş kazalarına ilişkin, belirli alt gruplara özgü olan ilişkilerin tanımlanması, vakaların yüksek, orta, düşük risk grupları gibi kategorilendirilmesi ve gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması amaçlanmaktadır. Bu amaçla, bir demir çelik işletmesine ait 205 iş kazası verisi, veri madenciliği sınıflayıcı yöntemler ile araştırılmıştır. Yapılan modellemelerde; kazalının yaşı, cinsiyeti, medeni durumu, eğitim durumu, iş tecrübesi, kadrolu veya taşeron olarak çalışması, çalıştığı alan, geçirdiği kazanın şiddeti (hafif, orta, yüksek) bilgileri kullanılmıştır. Model çözümlemesi için Chaid, C5.0 ve CRT algoritmaları tekniklerinden faydalanılmış ve model sonuçları karşılaştırılmıştır. Veri analizi IBM SPSS modeller paket programı aracılığı ile yapılmış ve veri madenciliği tüm aşamaları ortaya konulmuştur. Veri madenciliği sınıflayıcı teknikler arasında en yüksek doğruluk oranına karar ağacı tekniklerinden CRT algoritması ile ulaşılmıştır. Karar ağaçları yöntemleri, demir çelik endüstrisindeki iş kazalarının sonucunu tahmin etmek için kullanılmasıyla önleyici tedbirler ve eğitim ihtiyaçları konusunda için tahminler kullanılarak kazalanma oranları azaltılabilir.

Anahtar Kelimeler: Demir-çelik endüstrisi, iş kazaları, veri madenciliği, karar ağacı algoritmaları

An Application on Decision Tree Algorithms in Data Mining and Occupational Accidents in the Iron and Steel Industry

Abstract

Workers in the steel industry frequently face the risk of occupational accidents due to the structure of the facilities, production flow, excessive and repetitive work items, high temperature due to the nature of the production processes, and the noisy work environments. "Occupational accident" as a concept; covers unwanted, unexpected, negligence, fault, carelessness, willfulness, and unfortunate events encountered by the worker in the work environment. Occupational accidents are a very important problem for businesses and society due to their frequency and consequences. There is a widespread opinion that the main cause of occupational accidents is the unsafe movements of the personnel. However, many studies have shown that most accidents are related to employee's personality traits and unsuitable environmental conditions. In this study, it is aimed to define the relationships specific to certain subgroups regarding occupational accidents in an iron and steel factory, to categorize the cases as high, medium and low-risk groups, and to establish rules for predicting future events. For this purpose, data of 205 occupational accidents belonging to an iron and steel enterprise were investigated using data mining classification methods. In the models made; Information on the age, gender, marital status, educational status, work experience, employment as a permanent or subcontractor, field of work, severity of the accident (mild, moderate, high) of the accident were used. Chaid, C5.0, and CRT algorithms techniques were used for model analysis and model results were compared. Data analysis was carried out through the IBM SPSS Modeler package program and all stages of data mining are revealed. Among the data mining classification techniques, the highest accuracy rate was achieved with the CRT algorithm, one of the decision tree techniques. By using decision trees methods to predict the outcome of occupational accidents in the iron and steel industry, accident rates can be reduced by using predictions for preventive measures and training needs.

Keywords: Iron and steel industry, occupational accidents, data mining, decision tree algorithms

* Sorumlu Yazar: Karabük Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Karabük, Türkiye

1. Giriş

Demir ve karbon alaşımı olan çelik, modern dünyanın en önemli mühendislik ve yapı malzemesidir. Hayatımızın her alanında (otomobillerde ve inşaat ürünlerinde, ulaşım ve alt yapı elemanlarında, beyaz eşyalarda, kargo gemilerinde ve cerrahi araç gereçlerde) kullanılmaktadır. Aynı zamanda sağladığı katma değerlerle ülkeler için stratejik öneme sahiptir. Dünya ham çelik üretimi 2019 yılında 1.868,8 milyon tona (Mt) ulaşmıştır. Ülkemiz ise 2019 yılında gerçekleştirdiği 33.7 milyon ton çelik üretimi ile Dünyadaki en büyük 8. çelik üreticisidir (World Steel Association, 2020).

Çelik, iki ana yöntem ile üretilir. Bunlarda ilki yüksek fırın-bazık oksijen fırın (BF-BOF) üretim teknolojisi iken, diğeri ise elektrikli ark ocağı (EAF) yöntemidir. Üretim yollarının varyasyonları ve kombinasyonları mevcut olmakla birlikte temel farklılık kullanılan hammaddedir. BF-BOF üretim teknolojisinde ana hammaddeler; demir cevheri, kömür ve hurda çeliktir, EAF ise ağırlıklı olarak hurda ve elektrik enerjisi kullanılmaktadır. Her iki üretim akışında da çalışanlar, tesislerin yapısı, üretim akışı, fazla ve tekrarlayan iş ögeleri, üretim proseslerinin doğası gereği yüksek sıcaklık ve gürültülü iş ortamları sebebiyle sık sık iş kazası riski karşı karşıya kalmaktadırlar.

Çelik üretimi, kitlesel olarak çeliğin üretilmeye başlandığı ilk dönemlerde daha tehlikeli bir süreçti ve bu süreçte kazalar kaçınılmazdı. Günümüzde ise birçok şirket, modern ve teknik olarak gelişmişliğin sonucu olarak sıfır kaza ilkesiyle üretimlerine devam etmektedirler. Buna rağmen, Sosyal Güvenlik Kurumu'nun verilerine göre 2019 yılında ülkemiz ana metal sanayiinde 16.419 adet iş kazası raporlanmıştır.

“İş kazası” kavramı olarak; çalışma ortamında çalışanın karşılaştığı istenmeyen, beklenmeyen, ihmalkârlık, kusur, dikkatsizlik, kasıt ve şanssızlık sonucu meydana gelen olayları kapsamaktadır (Demir ve Öz, 2018). Demir ve çelik endüstrisindeki en yaygın yaralanma ve hastalık nedenleri (International Labour Organization, 2005); kayma, takılma ve düşme (düz zeminde), yüksekte düşme, korumasız makineler, düşen nesnelere, yutulma/kapılma, kapalı alanlarda çalışmak, makinelerin taşınması, forkliftler ve vinçler, kontrollü ve kontrolsüz enerji kaynaklarına maruz kalma, asbeste maruz kalma, mineral yünlere ve liflere maruz kalma, solunabilir etmenler (gazlar, buharlar, tozlar ve dumanlar), kimyasallar (tahriş edici maddeler (asitler, alkaliler), çözücüler ve hassaslaştırıcılar) cilt teması, sıcak metal ile temas, ateş ve patlama, aşırı sıcaklıklar, radyasyon (iyonlaştırıcı olmayan, iyonlaştırıcı), gürültü ve titreşim, elektrik yanıkları ve elektrik çarpması, manuel kullanım ve tekrarlayan çalışma, patojenlere maruz kalma, otomasyondan kaynaklanan arızalar, ergonomi, İSG eğitimi eksikliği, zayıf iş organizasyonu, yetersiz kaza önleme ve denetim, yetersiz acil durum ilk yardım ve kurtarma tesisleri, tıbbi tesislerin ve sosyal korumanın eksikliğidir.

Dünya Çelik Örgütü (World Steel Association)'nın üyeleri için hazırladığı, çelik endüstrisi için altı güvenlik ve sağlık ilkesi bulunmaktadır; (World Steel Association, 2020) Tüm yaralanmalar ve işle ilgili hastalıklar önlenemez ve önlenmelidir. Yönetim, güvenlik ve sağlık performansından sorumludur. Çalışan katılımı ve eğitimi esastır. Güvenli çalışmak bir istihdam koşuludur. İş sağlığı ve güvenliğinde mükemmellik, mükemmel iş sonuçlarını destekler. İş sağlığı ve güvenliği, tüm iş yönetimi süreçlerine entegre edilmelidir.

İş kazalarının başlıca nedeninin personelin güvenli olmayan hareketleri olduğu konusunda yaygın bir görüş vardır. Ancak birçok çalışma, kazaların çoğunun çalışanların kişilik özellikleri ve uygun olmayan ortam koşulları ile ilgili olduğunu göstermiştir.

Bu çalışmada, bir demir çelik işletmesinde kazalanan 205 çalışana ait veri tabanı üzerinden iş kazalarına ilişkin, belirli alt gruplara özgü olan ilişkilerin tanımlanması, vakaların yüksek, orta, düşük risk grupları gibi kategorilendirilmesi ve gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması amaçlanmaktadır. Ayrıca benzer konularda yapılan çalışmalar incelenerek bir literatür taraması yapılmıştır.

2. Literatür

İş kazaları ile ilgili mevzuatlarında zorunlu kıldığı üzere yaşanan her iş kazası vakası kayıt altına alınmaktadır. Derlenen iş kazası verileri ileriye yönelik tahmin yapmak için çoğu zaman yetersizdir. Bu sebeple kullanılan veri madenciliği yöntemleri, eldeki verilerden faydalı bilgiler oluşturarak gelecekte oluşabilecek kazaları ve/veya şiddeti azaltacak kuralları ortaya koymaktadır.

Günümüze kadar yapılan pek çok çalışmada veri madenciliği yöntemleri kullanılarak iş kazalarının analiz edilmiş, kuralları dizisi oluşturulmuş ve tahminleme modellerinin kurgulanmıştır.

Chi ve arkadaşları (2003) 1989, 1990 ve 1992 yıllarından itibaren 1230 vaka yeniden analiz edilmiştir. Her bir ölüm, mağduru yaş, cinsiyeti, endüstrisi ve kaza türü açısından değerlendirilmiştir. Yaş en az kritik değişken olarak belirlenmiştir. Kazalananların yaşını, cinsiyetini ve endüstrilerini ilişkilendiren beş farklı etkileşimli kalıp belirlenmiştir. ANOVA, SMR ve CHAID arasında yapılan bir karşılaştırma, CHAID'in diğer yöntemlerden daha iyi performans gösterdiği ortaya konulmuştur.

Madencilik ve inşaat sektörlerinde yapılmış olan bir diğer çalışmada (Rivas vd., 2011), araştırmacılar kazadan hemen sonra yapılan görüşmelerde elde edilen kaza ve olay verilerini modellemek için veri madenciliği tekniklerini (karar ağaçları, Bayes ağları, destek vektör makineleri ve sınıflandırma ağaçları) kullanılmıştır. Sonuçlar, lojistik regresyon ile analiz edilerek, kazaların altında yatan faktörlerin öngörülmesine ve belirlenmesine fayda sağlamıştır.

Verma ve ark. (2014) yaptıkları çalışmada çelik endüstrisinde mesleki yaralanmalar ve iş kazalarına ait 843 olay 3 temel kriter altında incelenmiştir. Sonuç olarak yapılan güvensiz davranışlar sonucunda oluşan yaralanmaların en yüksek frekansa sahip olduğu gözlenmektedir.

Eser (2015) tarafından gerçekleştirilen çalışmada resmi kayıtlar kullanılarak son yıllarda yaşanmış iş kazaları için zaman dağılımlarından faydalanılmış, kaza zaman yoğunlukları aylık ve günlük olarak belirlenmiştir. Aynı şekilde iş kazası geçiren çalışanların iş deneyimleri ile iş kazası arasındaki ilişki yorumlanmıştır. Toplam 17 adet faaliyet grubu gruplandırmaya dahil edilmiştir. Yapılan bu çalışmalar neticesinde kaza sıklıkları ve şiddetlerinin nedenleri irdelenmiş, tecrübe, eğitim gibi çevresel faktörlerin kazalanmadaki etkileri ortaya koyulmuştur.

Ghousi, (2015 yaptığı çalışmasında 20-25 yaş arasında ve iş tecrübesi 2 yıl ve daha az olan çalışanlarda yüksek oranda ve şiddette kazaların meydana geldiği ortaya koyulmuştur. Cheng ve arkadaşları, 2013 yılında yayınladıkları çalışmada 447 vakayı CART algoritması kullanmışlar ve Ghousi R.'nin sonuçlarına yakın kurallar oluşturmuşlardır. Yaş ve tecrübe kazalanmada önemli bir kriter olarak belirlenmiştir.

Sarkar ve arkadaşlarının (2016) yaptıkları çalışma ile yalnızca kazaları tahmin etmekle kalmayıp aynı zamanda ramak kala, mal hasarı veya yaralanma vakaları gibi kaza senaryolarını açıklamak için kurallar sağlayan bir öngörücü model oluşturmayı amaçlamışlardır. CART algoritmasının kullanıldığı modelde kullanılan ekipmanın tehlike sınıfının kaza sınıfın belirlenmesinde en önemli etken olduğu ortaya konmuştur.

2018 yılında yapılan (Mutlu, Alp ve Altundaş, 2018) çalışmada ülkemizdeki otuz şehirde tekstil işletmelerinde yaşanmış iş kazaları ele alınmıştır. 522 adet kaza değerlendirilerek toplam 31 iş kaza faktörü göz önüne bulundurulmuştur. Paket program olarak WEKA (Waikato Environment for Knowledge Analysis) kullanılmıştır. Değişkenler arasındaki ilişkileri belirlemek için veri madenciliği yöntemlerinden biri olan birliktelik kurallarından faydalanılmıştır. Analiz edilen çalışma ortamlarından %100'ü faaliyetlerini sürdürmüştür. İş Sağlığı ve Güvenliği eğitimine katılmış ve çalışırken iş kazası geçiren personele sahip işletmelerin tamamı faaliyetine devam etmiştir.

Altunkaynak'ın (2018) yaptığı çalışmada Ki-Kare testi kullanılarak, bu tür kazalara neden olan faktörlerini belirlemek ve Kayıp İş Günleri (LWD), Kaza Sonrası Durumu (PAS) ve Kaza Türü (TA) arasındaki ilişki değerlendirilmiştir. Çalışmanın istatistiksel analizinde Weka R-project yazılımları kullanılmıştır. Analizler, en sık karşılaşılan kazalanma şeklinin "vücuda veya uzuvlara iki nesne arasında baskı yapması" olduğunu ve kazalanmada en önemli değişkenlerin sektör, boyut ve yaş, deneyim olduğunu göstermiştir

Bir diğer çalışmada ise (Yapıcı ve Çetinkaya, 2018) Eurostat (Avrupa İstatistik Kurumu) ve SGK (Türkiye Sosyal Güvenlik Kurumu) verileri kullanılarak Avrupa ülkeleri ile Türkiye kıyaslanmış ve Türkiye'nin iş sağlığı ve güvenliği konusunda Avrupa'daki konumunun belirlenmesi hedeflenmiştir. Çalışmada, 2008-2015 tarihleri arasındaki veriler değerlendirilerek ölümlü iş kazalarının değişimleri incelenmiştir. Değişimin istatistiksel bir ölçüsü olarak değişkenlik katsayısı kullanılmıştır. Avrupa ülkeleri için 2008-2015 yılları arasında gerçekleşen toplam iş kazaları içerisinde ölümlü iş kazalarının oranları verilerine kümeleme analizi uygulanmıştır. Ayrıca, aynı verilere çok boyutlu ölçekleme analizi uygulanmıştır. Araştırma sonucunda iş kazası sonucu ölüm oranlarına göre en tehlikeli sektörler; madencilik, inşaat, metal-makine sektörleri olduğu belirlenmiştir. İş kazası sayılarında ise madencilik, metal/makine, ulaşım (havayolu taşımacılığı) sektörlerinin ilk sıralarda olduğu belirlenmiştir.

2018 yılında Shirali ve arkadaşlarının İran çelik sektöründe gerçekleştirdiği çalışmada, 2001'den 2014'e kadar 2127 mesleki yaralanma vakası için 12 değişkenli verileri toplanmıştır. IBM SPSS Modeler programı ile CART ve CHAID algoritmaları karar ağaçları ve tahminler oluşturmak için kullanılmış ve her yöntem için iş kazası sonucu için beş tahmin oluşturulmuştur. CART yöntemi için en önemli prediktör değişkenler sırasıyla yaş, kaza nedeni ve eğitim düzeyini içermektedir. CHAID yöntemi için yaş, kaza yeri ve eğitim düzeyi sırasıyla en önemli değişkenler olarak tespit edilmiştir.

Sanmiquel vd. (2018) İspanya'da 2005 ve 2015 yılları arasında madencilik sektöründe yaşanan iş kazalarını veri madenciliği ile algoritmaları (birliktelik analizi) analiz etmişler ve sonuç olarak önemli sebepler olarak, fiziksel efor veya aşırı eforla vücut hareketi olduğunu belirlemişlerdir. Ayrıca kazalanın yaşının kaza türü üzerindeki etkisini çok az olduğunu ifade etmişlerdir.

Ayanoğlu ve Kurt (2019) yaptıkları çalışmada Türk metal sektöründe 165 adet işyerinde, 2012-2015 yılları arasında meydana gelen iş kazalarının değerlendirilmesi sonucunda, 44 değişkenli 193 kazaya dair bir iş kazası veri kümesi ortaya çıkarmışlardır. Veri kümesinde çok değişkenli veri analizi yöntemleri kullanılarak, çeşitli çıkarımlar yapmışlar ve veri madenciliği programı kullanılarak veri setinden en iyi tahmin modeli üreten algoritmanın yapay sinir ağları olduğu belirlemişlerdir. Çift katmanlı ileri beslemeli yapay sinir ağı ile bir tahmin modellemesi oluşturmuş, işyerlerindeki olası kazalanma risklerini değerlendirilmişlerdir. Çalışma sonucunda; Bayes Ağları ve Karar Ağacı gibi diğer yöntemlerin de yapay sinir ağlarına yakın başarı gösterdiğini savunmuşlardır.

Bir başka çalışmada Türkiye Sosyal Güvenlik Kurumu'ndan alınan 2013-2016 yılları arasında imalat sanayiinde 242.537 adet iş kullanılmıştır. Veri analizi yöntemleri olarak karar ağacı analizi ve birliktelik kuralları uygulanmıştır (Mutlu, Altundaş, 2019). Analiz sonucunda imalat sistemlerindeki kaza türü üzerinde beş faktörün (malzeme (kaza sırasında kullanılan ekipman), özel aktivite, genel aktivite, yer (kazanın meydana geldiği bölüm) ve kazalanın mesleği) etkili olduğu görülmüştür.

Erdoğan ve arkadaşlarının (2019) yapmış olduğu çalışmada Türkiye Taş Kömür İşletmeleri (TTK) ele alınmıştır. TTK'da 15 yılda (2000-2014,39.738 kaza) meydana gelen kazalar öncelikle kazaların sayısı, tipi ve yeri, yaş, deneyim, eğitim düzeyi ve kazazedelerin asıl görevi ile bu kazalardan kaynaklanan yaralanmalar açısından istatistiksel olarak analiz edilmiştir. Tehlikelerin alt ve üst sınırları belirlenmiş ve her tehlike sınıfı için olay ağaçları hazırlanmıştır.

Ersöz (2019) çalışmasında, Demir-Çelik sektöründe meydana gelen iş kazaları sonucu oluşan iş gücü kayıplarını en çok etkileyen bağımsız değişkenlerin belirlenerek iş gücü kayıplarının en aza indirilmesini amaçlayarak, Demir-Çelik tesislerinin yoğun olduğu bir ilde meydana gelen 2553 çalışana ait iş kazası verisini, çeşitli istatistiksel yöntemler ile araştırmıştır. CRT Karar ağacı yöntemi kullanılarak yapılan araştırma modeli sonucunda bağımlı değişken olan "iş kazası saatine" etki eden en önemli bağımsız değişkenler araştırılmış ve bağımsız değişkenler olarak "yaş", "tecrübe" ve "işgücü kayıp günü" alınmıştır. Araştırmada CRT yöntemi sonuçlarına

göre iş kaza saatini belirleyen en önemli değişken çalışanın yaşı çıkmıştır. C5.0 karar ağacı algoritma sonucuna göre iş kaza saatlerini etkileyen en önemli değişken “çalışanın görevi” değişkeni bulunmuştur. Bunu sırasıyla “tecrübe”, “yaralanan bölge” ve “yaş” değişkeni olduğu tespit edilmiştir. İş kazası oluş şekline göre en yüksek “Burkulma” ve daha sonra “Demir çarpması, düşmesi, parça düşmesi” olduğu görülmüştür. Ayrıca araştırma sonucuna göre işgücü kaybı ile tecrübe arasında ters orantılı bir ilişki ve önemli bir ilişki tespit edilmiştir. Tecrübe arttığında işgücü kaybı azalmaktadır.

3. Metodoloji

3.1 Veri Madenciliği

Veri Madenciliği, büyük veri setleri içerisinde saklı, değerli ve faydalı bilgilerin açığa çıkartılması olarak tanımlanabilir. Veri madenciliğindeki temel amaç, derlenmiş ve hazırlanmış verilerin bir takım istatistik yöntemlerle analiz edildikten sonra karar destek sistemlerinde kullanılacak kararlar üretmektir. Veri madenciliği süreci; problemin tanımlanması, verileri tanınması, verilerin hazırlanması, modelleme, değerlendirme ve uygulama aşmalarından oluşmaktadır.

3.2 Sınıflandırıcı Modeller ve Karar Ağaçları

Veri madenciliğinde modeller, işlevlerine göre üç ana grupta toplanır. Bunlar; Sınıflama (Classification), Kümeleme (Clustering) ve Birlikte Kuralları (Association Rules)’dır.

Sınıflayıcı modeller, veri madenciliğinde yer alan en tanınmış ve kullanışlı modellerdir. Sınıflama ve regresyon modelleri tahmin edici modellerdir ve söz konusu modeller öngörülse (Prediktif) modeller olarak da adlandırılmaktadır. Prediktif modeller, bir dizi girdi değeri baz alınarak bir sonuç değerinin tahmin edilmesi amaçlanmaktadır. Prediktif modeller içerisinde en önemli teknikler regresyon ve karar ağaçlarıdır (Ersöz, 2015).

Veri madenciliğinde kullanılmakta olan karar ağaçları, sınıfları bilinmekte olan ve sınıflandırılabilir veri setleri üzerinden tümevarım yöntemleriyle kullanılmakta olan bir karar yapısı çeşididir (Sun ve ark., 2008). Karar ağaçları, mega verileri, bir dizi karar verme aşamalarından geçirilerek, küçük gruplarına ayıran ve bu amaçla kullanılmakta olan yöntemlerdir. Yapılmakta olan her bir ayırım işlemi, nihai gruplarının üyelerini, diğerleriyle daha benzer hale getirmektedirler (Berry, 2004). Karar ağacı metotları, diğer istatistik yöntemlere kıyasla detaylı bilgi üretilmesi konusunda daha başarılıdır. Akademik çalışmalar için çok sık başvurulan yöntemler olmamasına karşın, sonuçların gruplandırılması ve kriterler arasında en önemli kriterin belirlenmesinde oldukça etkin çözümleme araçlarıdır (Ersöz, 2015). Yöntemin sık tercih edilmesi, ağaç yapılarını oluşturan kurallar dizinin kolay, sade ve araştırmacı tarafından anlaşılabilir olmasından kaynaklanmaktadır. Karar ağacı bir sınıfın olası üyelerinin belirlenmesinde, vakaların yüksek, orta, düşük risk grupları gibi sınıflara ayrılmasında, parametrik modellerin kurularak çok sayıda değişkenden en önemlilerinin seçilmesinde, gelecekteki olayların tahminlenmesi adına kurallar oluşturulmasında, belirli alt gruplara özgün ilişkilerin tanımlanmasında, kategorilerin birleştirilmesiyle sürekli değişkenlerden kesikli değişkenlere dönüştürülmesinde yaygın olarak kullanılmaktadır [Akpınar, 2000].

Veri madenciliğinde karar ağaçları oluşturmak amacıyla çok sayıda algoritma geliştirilmiştir. Bu algoritmalar, CHAID (Chi-Squared Automatic Interaction Detector), Exhaustive CHAID, CRT (Classification and Regression Trees), ID3, C4.5, MARS (Multivariate Adaptive Regression Splines), QUEST (Quick, Unbiased, Efficient Statistical Tree), C5.0, SLIQ (Supervised Learning in Quest), SPRINT (Scalable Paralleizable Induction of Decision Trees) şeklinde sıralanabilir (Akpınar, 2000).

CHAID Algoritması

CHAID yöntemi 1980 tarihinde araştırmacı Kaas tarafından en iyi bölmeyi hesaplamak için ortaya atılmıştır. En uygun bölümleri seçebilmek adına entropy/gini’den değil Ki-Kare analizlerinden faydalanılmaktadır (Albayrak ve Yılmaz, 2009). Değişkenlerin bölünmeye uygunluğuna Bonferroni düzeltilmiş p değeri kullanılarak karar verilmektedir. Bonferroni yaklaşımı, her bir grubun ortalama vektörlerinin genel ortalama vektöründen farklarının sıfır olup olmadığını araştırmaktadır. Genel ortalamalar vektörü \bar{x} ve her grubun i . değişkene göre ortalama vektörleri \bar{x}_i eşitlik 1’deki gibi hesaplanmaktadır.

$$\begin{matrix} \bar{x}_1 & \bar{x}_{11} & \bar{x}_{1g} \\ \bar{x}_2 & \bar{x}_{21} & \bar{x}_{2g} \\ \bar{x}_3 & \bar{x}_{31} & \bar{x}_{3g} \\ \vdots & \vdots & \vdots \\ \bar{x}_p & \bar{x}_{p1} & \bar{x}_{pg} \end{matrix} \quad (1)$$

Her grubun ortalama vektörünün, genel ortalama vektöründen farkları,

$$d_i = \bar{x}_i - \bar{x} = \begin{matrix} \bar{x}_{i1} - \bar{x}_1 & \bar{x}_{i2} - \bar{x}_2 \\ \bar{x}_{i3} - \bar{x}_3 & \vdots \\ \bar{x}_{ip} - \bar{x}_p & \bar{x}_{ig} - \bar{x}_g \end{matrix} \quad (2)$$

şeklinde hesaplanır. k . grup ile j . grup i . değişken ortalamaları arasındaki ortalama farklarına ait $1-\alpha$ güven aralığı,

$$d_{ki}-d_{ij}=(\bar{x}_{ki}-\bar{x}_{ji})\pm t(\alpha/pg(g-1))\sqrt{\left(\frac{1}{n_k}+\frac{1}{n_j}\right)*\frac{w_{ij}}{N-g}} \quad (3)$$

eşitliği ile hesaplanmaktadır. Burada, $N=n_1+n_2+\dots+n_g$, p değişken sayısı, g grup sayısı ve w_{ij} ise W matrisinin köşegen elemanlarını ifade etmektedir. W matrisi gruplar için değişimi göstermekte olup,

$$W=\sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) * (x_{ij} - \bar{x}_i)' \quad (4)$$

denklemleri ile hesaplanmaktadır. Grupların sayısı “ g ” ile i . gruptaki birim sayısı ise “ n_i ” ile gösterilmektedir. Her bir değişken için gruplar ikiye bölünecek şekilde değerlendirilir ve elde edilen aralığın sıfır değerini içerip içermediği belirlenir. Eğer aralık içerisinde sıfır değeri var ise gruplar arası istatistiksel olarak anlamlı bir farkın olmadığı, sıfır değeri yer almıyorsa ise gruplar arasındaki farklılığın istatistiksel olarak anlamlı olduğu sonucuna ulaşılır (Akpınar, 2000).

C5.0 ve CRT Algoritması

C5.0 algoritması kategorik hedef değişkenleri tahmin etmek amacıyla kullanılmaktadır. Hafızayı etkin kullanan algoritma daha küçük karar ağaçları oluşturmaya imkân tanımakta, verinin ayrıştırılması ve iyileştirilmesi için daha uygun çözümler sunmaktadır. Öğrenme zamanlarının da kısa olmasıyla modellerinin anlaşılması ve yorumlanması da kolaydır.

CRT algoritması son veya uç olmayan her bir düğümde iki adet dal bulunmaktadır. Budama işlemi ağacın karmaşıklık ölçüsüne dayanır. Sınıflandırma ve regresyonu destekleyici bir yapıdadır. Sürekli hedef değişkenleri ile çalışır. Verinin hazırlanmasına gereksinim duyar (Ersöz, 2015).

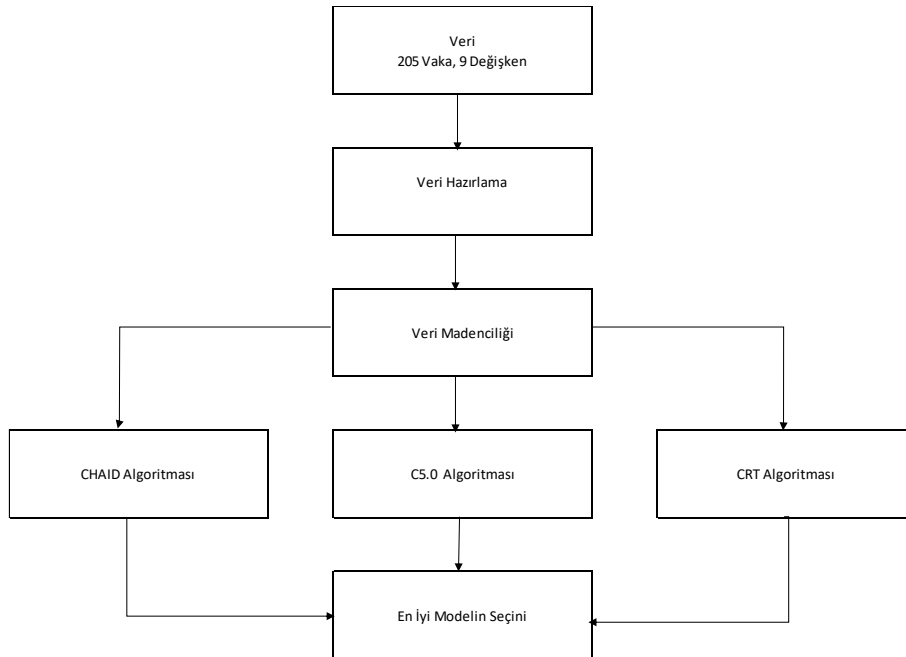
4. Veri Madenciliği Uygulaması

4.1 Problem tanımı ve amacı

Bu çalışmada, bir demir çelik fabrikası çalışanlarına ait veri tabanı üzerinden iş kazalarına ilişkin, belirli alt gruplara özgü olan ilişkilerin tanımlanması, vakaların yüksek, orta, düşük risk grupları gibi kategorileştirilmesi ve gelecekteki olayların tahmin edilebilmesi için veri madenciliği ile kurallar oluşturulması amaçlanmıştır. Çalışmada görsel, anlaşılır, basit yorumlanabilir ve kural çıkarımına imkân tanınması nedenleriyle veri madenciliği yöntemlerinden karar ağaçları tekniklerinden yararlanılmıştır.

4.2 Veri Toplama ve Verinin Modele Hazırlanması

Çalışmada kullanılan veri seti, bir demir çelik fabrikasında 2018 yılında iş kazasına uğramış çalışanların bilgilerinden oluşturulmuştur. Veri seti oluşturulurken; kazalının yaşı, cinsiyeti, medeni durumu, eğitim durumu, iş tecrübesi, kadrolu veya taşeron olarak çalışması, çalıştığı alan, geçirdiği kazanın şiddeti bilgileri kullanılmıştır (Tablo 1). Uygulama adımları Şekil 1.’deki modelde gösterilmiştir.



Şekil 1. Uygulama Adımları Akış Şeması

Veri girişleri IBM SPSS Statistics paket programı aracılığı ile yapılmış olup, modelleme öncesinde veriler içindeki aykırı, uç değerler temizlenmiştir. Araştırmada kullanılan nitel veriler aşağıda Tablo 1’de nicel veriler haline kodlanarak dönüştürülmüştür.

Tablo 1. Değişkenler ve Nitel Verilerin Kodlanması

Değişkenler	Kodlama	Değişken Alt Kategorisi
Cinsiyet	1	Kadın
	2	Erkek
Yaş	1	18-22 yaş
	2	23-30 yaş
	3	30-40 yaş
	4	40 yaş ve üstü
Eğitim Durumu	1	İlkokul
	2	Ortaokul
	3	Lise
	4	Lisans ve Lisans Üstü
Tecrübe	1	0-1 yıl
	2	1-2 yıl
	3	2-5 yıl
	4	5 yıl ve üzeri
Medeni Hal	1	Evli
	2	Bekâr
Çalışma Alanı	1	Saha
	2	Ofis
Kadro Durumu	1	Taşeron
	2	Taşeron Değil
Kaza Nedeni	1	Uygunsuz Ortam
	2	Uygunsuz Davranış
Kaza Şiddeti	1	Hafif
	3	Orta
	4	Şiddetli

Araştırmada yer alan 205 kazalının; %88’i erkek (181 kişi), %12’si kadın, %34’ü 18-22 yaş arasında, %43’ü ilkokul veya ortaokul mezunu, %45’i 0-2 yıl iş tecrübesine sahip, %84’ü saha da iş görmekte, %58’i taşeron personel olduğu ve %57’sinin uygunsuz davranıştan kazalandığı tespit edilmiştir.

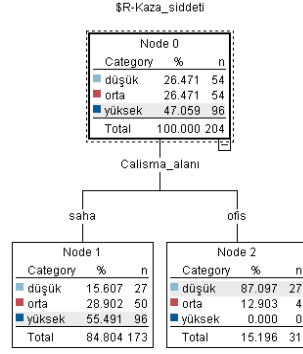
5. Bulgular

Araştırmada veri madenciliği sınıflayıcı modellerden; Chaid, CRT ve C.5.0 algoritmaları kullanılmıştır. Veri madenciliği araştırması sonucunda, kullanılan modellere ilişkin başarı oranları Tablo 2’de gösterilmiştir. Karar ağacı algoritmalarındaki başarı oranları farklılık göstermiştir (%68,29 %87,80 %78,05).

Tablo 2. Karar Ağacı Model Sonuçları

CHAID Algoritması	%68,29
CRT Algoritması	%87,80
C5.0 Algoritması	%78,05

Araştırmada iş kazası şiddetini etkileyen değişkenler araştırılmış ve Chaid karar ağacı model sonucu aşağıda Şekil 2’de verilmiştir.



Şekil 2. Kaza Şiddetine Ait CHAID Analizi Tablosu

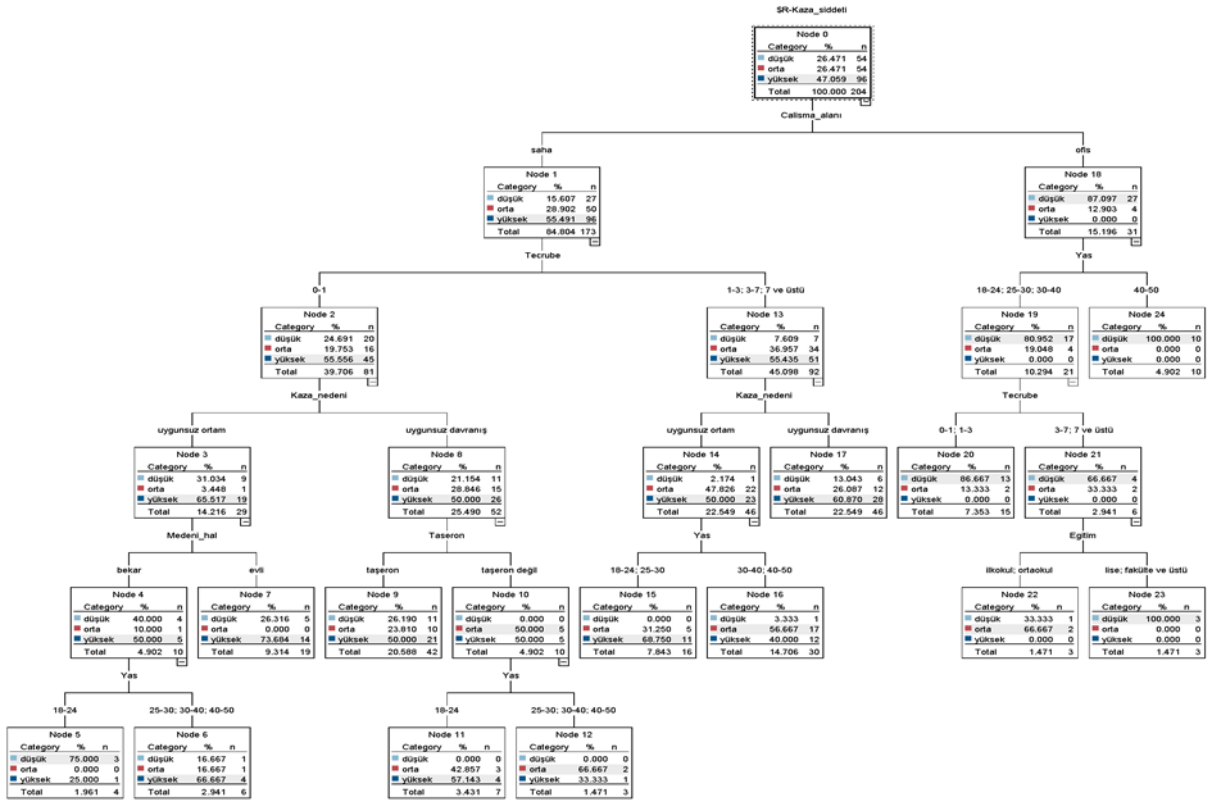
CHAID algoritmasının çözümünde sadece iki düğüm oluşmuştur. Ağaç sadece çalışma alanı değişkenine göre dallanmıştır. Yani “çalışma alanının” kazalanma şiddeti için en etkili faktör olduğu belirlenmiştir.

Tablo 3. Kaza Şiddetinin CHAID Algoritması ile Tahmini

Düğüm	Kurallar	Kaza Şiddeti	Oran	Popülasyon
1	Eğer (çalışma ortamı=saha)	Yüksek	55,491%	96
2	Eğer (çalışma ortamı=ofis)	Düşük	87,097%	27

CHAID analizine göre sahada çalışan personelin yaşadığı kazalar yüksek %55,491 oranında yüksek şiddetli olurken, ofis çalışanlarının geçirdiği kazaların %87,097’si düşük şiddette olduğu görülmüştür (Tablo 3).

İş kazası şiddetini etkileyen değişkenler C&RT algoritması ile de araştırılmış ve model sonucu aşağıda Şekil 3’de verilmiştir.



Şekil 3. Kaza Şiddetine Ait C&RT Analizi Tablosu

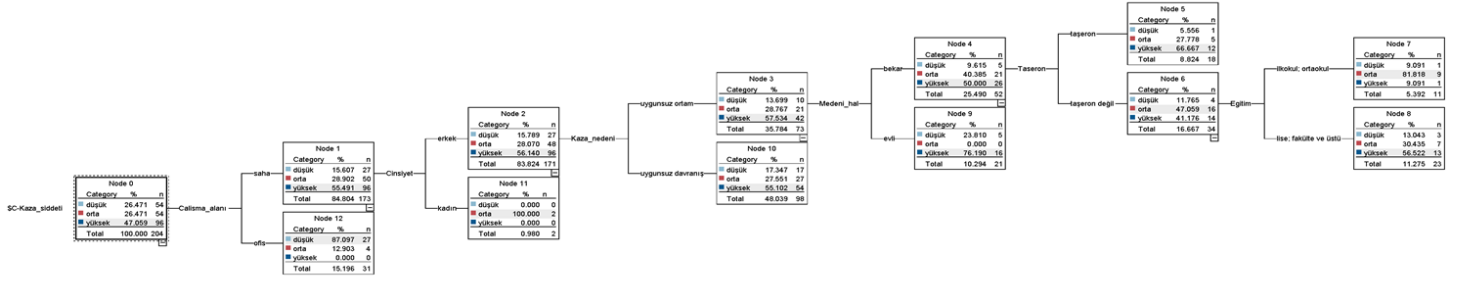
C&RT algoritmasının sonucunda 23 düğüm olmuştur. İlk dallanma Chaid algoritmasına benzer şekilde “çalışma alanı” olarak gözlemlenmiştir. Ofis çalışanlarında 40-50 yaş arasındaki personelin düşük riskle çalıştığı; 40 yaş altı personelin ise %80,952 düşük şiddetinde ve %19,048’i ise orta şiddette kazalandığı tespit edilmiştir.

Tablo 4. Kaza Şiddetinin C&RT Algoritması ile Tahmini

Düğüm	Kurallar	Kaza Şiddeti	Oran	Popülasyon
5	Eğer (çalışma ortamı=saha ise) ve tecrübe = 0-1 ve kaza nedeni=uygunsuz ortam ve medeni hali= bekar ve yaş= 18 ile 24 arasında	Düşük	75,000%	3
6	Eğer (çalışma ortamı=saha ise) ve tecrübe = 0-1 ve kaza nedeni=uygunsuz ortam ve medeni hali= bekar ve yaş= 25-30;30-40;40-50	Yüksek	66,667%	4
9	Eğer (çalışma ortamı=saha ise) ve tecrübe = 0-1 ve kaza nedeni=uygunsuz davranış ve taşeron	Yüksek	50,000%	21
11	Eğer (çalışma ortamı=saha ise) ve tecrübe = 0-1 ve kaza nedeni=uygunsuz davranış ve taşeron değil ve yaş=18-24	Yüksek	57,143%	4
12	Eğer (çalışma ortamı=saha ise) ve tecrübe = 0-1 ve kaza nedeni=uygunsuz davranış ve taşeron değil ve yaş=25-30;30-40;40-50	Orta	66,668%	2
15	Eğer (çalışma ortamı=saha ise) ve tecrübe = 1-3;3-7;7+ ve kaza nedeni=uygunsuz ortam ve yaş= 18-24;25-30	Yüksek	68,750%	11
16	Eğer (çalışma ortamı=saha ise) ve tecrübe = 1-3;3-7;7+ ve kaza nedeni=uygunsuz ortam ve yaş= 30-40;40-50	Orta	56,667%	17
17	Eğer (çalışma ortamı=saha ise) ve tecrübe = 1-3;3-7;7+ ve kaza nedeni=uygunsuz davranış	Yüksek	60,870%	28
20	Eğer (çalışma ortamı=ofis ise) ve yaş=18-40 ve tecrübe= 0-3	Düşük	86,667%	13
22	Eğer (çalışma ortamı=ofis ise) ve yaş=18-40 ve tecrübe= 7+ ve eğitim=ilkokul;ortaokul	Orta	66,667%	2
23	Eğer (çalışma ortamı=ofis ise) ve yaş=18-40 ve tecrübe= 7+ ve eğitim=lise,fakülte ve üstü	Düşük	100,000%	3
24	Eğer (çalışma ortamı=ofis ise) ve yaş=40-50	Düşük	100,000%	10

Sahada çalışan personelde ise kaza şiddeti en çok etkileyen ikinci faktör “tecrübe” ve “yaş” faktörleri olarak karşımıza çıkmıştır. Bir seneden tecrübesiz olarak çalışanların %55,556’sının iş kazası şiddetinin “yüksek” olduğu, uygunsuz ortamda çalışanların iş kazası şiddetinin (%65,517) ve uygunsuz davranıştan kazalanan çalışanların (%50) iş kazası şiddetinin “yüksek” olduğu tespit edilmiştir. Kaza nedeni “uygunsuz ortam” ve çalışanın medeni durumu “evli” ise, kaza şiddetinin “düşük” (%26,190) ve “yüksek” (73,604) şiddette kazalandığı; çalışanın yaşı 18-24 yaşında ise “düşük” (%75) ve “yüksek” (%25) şiddette kazalandığı görülmüştür. İşyerinde “uygunsuz davranış” gösteren taşeron çalışanların %50’si “yüksek” şiddette; %25,190 “düşük” ve %23,810 “orta” şiddette kazalandığı görülmüştür. Eğer çalışan taşeron değilse “düşük” şiddette iş kazası geçirmediikleri, %50 “orta” ve “yüksek” şiddette kazalandığı, yaş 25+ yaşında çalışanların %66,667 “orta” şiddette ve %33,333’ü ise “yüksek” şiddette iş kazası geçirdiği tespit edilmiştir (Tablo 4.).

İş kazası şiddetini etkileyen faktörler C5.0 algoritması ile de araştırılmış ve model sonucu aşağıda Şekil 4’te verilmiştir.



Şekil 4. Kaza Şiddetine Ait C5.0 Analizi Tablosu

C5.0 Algoritmasının sonucunda 12 düğüm olmuştur. İlk dallanma CHAID ve CRT algoritmasına benzer şekilde en önemli faktör “çalışma alanı” olduğu gözlemlenmiştir. İş kazası şiddetini etkileyen ikinci faktörün “cinsiyet” gözlemlenmiş olsa da sahada çalışan kadın sayısı ve yaptıkları işlerin farkları düşünüldüğünde karar vermek için yeterli olmadığı değerlendirilmiştir.

Tablo 4. Kaza Şiddetinin C5.0 Algoritması ile Tahmini

Düğüm	Kurallar	Kaza Şiddeti	Oran	Popülasyon
5	Eğer (çalışma ortamı=saha ise) ve cinsiyet=erkek ve kaza nedeni=uygunsuz ortam ve medeni durumu = bekar ve taşeron	Yüksek	66,667%	12
7	Eğer (çalışma ortamı=saha ise) ve cinsiyet=erkek ve kaza nedeni=uygunsuz ortam ve medeni durumu = bekar ve taşeron değil ve eğitim=ilkokul;ortaokul	Orta	81,818%	9
8	Eğer (çalışma ortamı=saha ise) ve cinsiyet=erkek ve kaza nedeni=uygunsuz ortam ve medeni durumu = bekar ve taşeron değil ve eğitim=Fakülte ve üstü	Yüksek	56,522%	13
9	Eğer (çalışma ortamı=saha ise) ve cinsiyet=erkek ve kaza nedeni=uygunsuz ortam ve medeni durumu = evli	Yüksek	76,190%	16
10	Eğer (çalışma ortamı=saha ise) ve cinsiyet=erkek ve kaza nedeni=uygunsuz davranış	Yüksek	55,102%	54
11	Eğer (çalışma ortamı=saha ise) ve cinsiyet=kadın	Orta	100,000%	2
12	Eğer (çalışma ortamı=ofis ise)	Düşük	87,097%	27

Çalışma alanı “saha” olan, cinsiyeti “erkek”, kaza nedeni “uygunsuz ortam” olan, medeni hal “bekar olan” ve taşeron çalışanların çalışanların %66,667’si “yüksek şiddette” kazalanmıştır. Çalışma alanı “saha” olan, cinsiyeti “erkek”, kaza nedeni “uygunsuz ortam” olan, medeni hal “bekar olan” ve taşeron olmayan çalışanların eğitim durumu “ilkokul ve ortaokul” olan çalışanların %81,818’i “orta şiddette”, eğer fakülte ve üstü mezuniyeti var ise %56,522 oranında “yüksek şiddetle” kazalanmıştır. Çalışma ortamı “saha” olan kadınların tamamı “düşük şiddetle” kazalanmıştır. Eğer personel ofiste çalışıyor ise %87,097 oranla “düşük şiddetli” kazalanmaktadır.

6. Sonuç ve Tartışma

Bu çalışmada bir demir çelik endüstrisinde yer alan işletmenin iş kazası verilerinden yararlanılarak, iş kazasının şiddetini etkileyen faktörler veri madenciliği karar ağacı teknikleri ile araştırılmıştır. Veri madenciliği karar ağacı algoritmaları uygulanmış ve oluşan kurallar değerlendirilmiştir.

En yüksek başarı oranına sahip (CRT Algoritmasına göre; bir seneden tecrübesiz olarak çalışanların %55,556'sinin iş kazası şiddetinin "yüksek" olduğu, uygunsuz ortamda çalışanların iş kazası şiddetinin (%65,517) ve uygunsuz davranıştan kazalanan çalışanların (%50) iş kazası şiddetinin "yüksek" olduğu tespit edilmiştir. Kaza nedeni "uygunsuz ortam" ve çalışanın medeni durumu "evli" ise, kaza şiddetinin "düşük" (%26,190) ve "yüksek" (73,604) şiddette kazalandığı; çalışanın yaşı 18-24 yaşında ise "düşük" (%75) ve "yüksek" (%25) şiddette kazalandığı görülmüştür. İşyerinde "uygunsuz davranış" gösteren taşeron çalışanların %50'si "yüksek" şiddette iş kazası geçirdiği tespit edilmiştir.

Chaid, CRT ve C5.0 algoritma sonuçları incelendiğinde; işyerindeki iş kazalarının şiddetini etkileyen en önemli değişken "çalışma alanı" olarak ortaya çıkmıştır. Daha sonra bunu "kaza nedeni" ve "tecrübe" faktörleri izlemiştir. Yaşın ise kazalanma için çok önemli bir kriter olduğu sonucuna varılmıştır. Buna rağmen; Ersöz (2019), Erdoğan ve arkadaşları (2019), Verma ve arkadaşları (2014) ve Eser (2015)'in yaptıkları çalışmalarda tecrübenin yanı sıra yaşın da çok önemli faktör olduğunu ortaya koymuşlardır. Ghousi (2015) 20-25 yaş arasında ve iş tecrübesi iki yıl ve daha az olan çalışanlarda yüksek oranda ve şiddette kazaların meydana geldiği ortaya koyulmuştur. Cheng ve arkadaşları, yaşın ve tecrübenin kazalanmada önemli bir kriter olduğunu sunmuşlardır. 2018 yılında Shirali ve arkadaşları (2018) ise sırasıyla yaş, kaza nedeni ve eğitim düzeyinin kazalanmadaki önemli değişkenler olduğunu sunmuşlardır.

İşletmelerdeki iş kazalarının nedenlerinin araştırılması, yönetimin önleyici önleyici tedbirler alınması açısından oldukça önemlidir. Söz konusu işletme önleyici faaliyet olarak İş Sağlığı ve Güvenliği eğitimleri ile mentörlük kavramlarının önemli unsurları olduğu değerlendirilmeli tecrübe ile kazanılacak deneyimin bir kısmını eğitimlerle tamamlamalıdır. Ayrıca taşeron personelin daha sık ve şiddetli kazalanması alt yüklenice getirilen şartların gözden geçirilmesi ve denetimlerin sıklaştırılması gerekliliğini göstermektedir. Taşeron personelin aldığı eğitimler de kontrol edilmelidir, hatta eğitimler işletme bünyesinde verilmeye başlanabilir.

Karar ağacı algoritmaları, demir çelik sektöründeki iş kazası şiddeti tahmini için uygun modeller oluşturmuştur. İşyeri yetkilileri kazalanmaya daha açık personel profilini belirleyemeye yönelik tahminler yaparak kaza oranını ve şiddetini azaltabilir, kişilere özgü eğitim ihtiyaçları planlayabilir ve özel tedbirler geliştirebilirler. Demir çelik endüstrisinde yaşanan kazaların sıklıkları göz önünde bulundurularak seçilen veri seti genişletilebilir.

Son olarak, gelecekteki çalışmalar için çelik endüstrilerindeki kazalanma şiddetini tahmin etmek için destek vektör makinesi ve Bayes ağları gibi diğer veri madenciliği yöntemlerinin kullanılması ve sonuçlarının karar ağacı algoritmaları ile karşılaştırılması önerilir.

Kaynakça

- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29 (1): 1-22.
- Albayrak, A., Yılmaz, Ş. (2009). Veri Madenciliği: Karar Ağacı Algoritmaları ve Imkb Verileri Üzerine Bir Uygulama", Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, C.14, pp.31-52,
- Altunkaynak, B. (2018). A statistical study of occupational accidents in the manufacturing industry in Turkey, · International Journal of Industrial Ergonomics 66, 0169-8141.
- Berry, Michael J. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, John Wiley and Sons Incorporated, USA, 194.
- Ayanoğlu C., Kurt M. (2019). Metal Sektöründe Veri Madenciliği Yöntemleri ile Bir İş Kazası Tahmin Modeli Önerisi, Vol. 2, pp. 78-87.
- Chi C., Chen C. (2003). Reanalyzing Occupational Fatality Injuries In Taiwan With A Model Free Approach, Safety Science, Volume 41, Issue 8, Pages 681-700.
- Demir, A., & Öz, A. (2018). Teolojik Açından İş Kazalarının İncelenmesi. Avrupa Bilim ve Teknoloji Dergisi, Sayı 14, S.189-197.
- Erdogan, H., Duzgun, H., Kestel, S. (2019). Quantitative Hazard Assessment For Zonguldak Coal Basin Underground Mines. International Journal Of Mining Science And Technology, Vol.29, 453-467.
- Ersöz F. (2015). "Veri Madenciliği Teknikleri ve Uygulamaları", Dijital Basımevi, 72. Tasarım, Ankara.
- Ersöz, T (2019). Demir-Çelik Sektöründe İş Kazalarının Analizi. Ekin Yayınevi, Bursa, Sayfa Sayısı 124.
- Ersöz T., Merdin, D., Ersöz F. (2015). Veri Madenciliği Yöntemi ile Memnuniyet Algısının Araştırılması, ISITES2015 Valencia -Spain
- Eser, A. (2015). Güncel İş Kazaları Verilerinin İstatistiksel Analizi, Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, Cilt 30, Sayı 2, Sayfalar 227 - 242.
- Ghousi, R. (2015). Applying a decision support system for accident analysis by using datamining approach: A case study on one of the Iranian manufactures, Journal of Industrial and Systems Engineering, Vol. 8, No. 3, pp 59-76.
- Hand, D. (2001). Manila, H. & Smyth, P., Principles of Data Mining. USA: MIT Press.
- International Labour Organization. (2005). Code of practice on safety and health in the iron and steel industry, Geneva, Sayfa Sayısı 5.
- Mutlu, N. G., Alp, S. & Altuntaş, S. (2018). Tekstil Sektöründe Yaşanan İş Kazalarının Verimadenciliği Uygulaması ile Analizi. 3 International Congress On Occupational Safety And Security, İstanbul.
- Mutlu, N. G., & Altuntaş, S. (2019). Assessment of Occupational Risks in Turkish Manufacturing Systems with Data-Driven Models. Journal Of Manufacturing Systems, vol.53, 169-182.
- Rivas, T., Paz, M., Martín J., Matías, J., García J. (2011). Explaining and predicting workplace accidents using data-mining techniques, Reliability Engineering & System Safety, 96- 7, Pages 739-747.
- Sarkar S., Patel A., Madaan S., Maiti, J. (2016). Prediction of Occupational Accidents Using Decision Tree Approach, Conference: INDICON-2016 (IEEE), Bangalore.

- Sanmiquel, L., Bascompta M., Rossell J., Anticoi, H. (2018). Analysis of Occupational Accidents in Underground and Surface Mining in Spain Using Data-Mining Techniques, *International Journal of Environmental Research and Public Health*, 15, 462.
- Sarkar S., Pateshwari V., Maiti J., (2017). Predictive Model for Incident Occurrences in Steel Plant in India, 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT).
- Shirali, G., Noroozi, M., Malehi, A. (2018). Predicting the outcome of occupational accidents by CART and CHAID methods at a steel factory in Iran, *J Public Health Res*, 7(2): 1361.
- Sun, Jie ve Hui Li, (2008). Data Mining Method for Listed Companies, Financial Distress Prediction, *Knowledge-Based Systems*, 21 (1): 2.
- Verma A., Das Khan S., Maiti, J., (2014). Identifying Patterns of Safety Related Incidents in A Steel Plant Using Association Rule Mining of Incident Investigation Reports, *Safety Science*, 70, pp. 89-98.
- Yapıcı N., Çetinkaya C.(2018). Comparison of Fatal Occupational Accidents Statistics in Turkey with Some European Countries, *International Journal of Scientific and Technological Research* ISSN 2422-8702 (Online) Vol 4, No.6.
- World Steel Association, (2020). *Safety And Health Principles And Definitions Guidance Book*, Brüksel, S. 4-5.