**Bağımsız Denetim Firma Tercihinin Makine Öğrenmesi Yöntemleriyle Tahmini**

**Abstract**

*Araştırmanın amacı, Türkiye’ de Borsa İstanbul Yıldız Pazar’da yer alan işletmelerin finansal oranlar ve makine öğrenme algoritmaları kullanarak bağımsız denetim firma tercihlerini tahmin etmektir. Bu kapsamda, 2019-2021 döneminde Borsa İstanbul’da işlem gören 159 işletmeye ait 13 finansal veri girdi değişkeni olarak kullanılmıştır. İlk olarak, Mutual Information (MI) yöntemiyle girdi değişkenlerinin önem değerleri bulunmuştur. Ardından, veriyi temsil eden en doğru alt kümeyi seçmek için girdi değişkenleri önem sırasına göre ardışık olarak gruplandırılmıştır. Grup seçimi amacıyla makine öğrenmesi algoritmalarından SVM (Support Vector Machine-Destek Vektör Makineleri), DT (Decision Tree-Karar Ağaçları), RF (Random Forest-Rastgele Orman), NB (Naive Bayes), KNN (K-Nearest Neighbors-K-En Yakın Komşu) ve XGBoost algoritması yöntemleri kullanılmıştır. Yöntemlerin başlangıç parametrelerini optimize etmek için GridSearchCV (GridSearch+Cross-Validation) tekniği uygulanmıştır. Yapılan deneyler sonucunda, bağımsız denetim firma tercihinde XGBoost algoritması %88.4 tahmin doğruluk (accuracy) değeri ile en başarılı yöntem olmuştur. Yöntemin, 13 veri arasından seçilen 8 özniteliği kullanması yeterli olmuştur. Öte yandan, Aktif Karlılık verisi en önemli öznitelik olarak belirlenmiştir. Çalışma, hem öznitelik seçiminde MI yöntemini kullanmış hem de daha az önemli değişkenleri modelden çıkararak yüksek başarı ile bağımsız denetim firma tercihini tahmin etmiştir*.

***Keywords:*** *Independent Audit, Financial Ratios, Machine Learning,Classification*

**1.Introduction**

Türkiye’ de halka açık şirketlerin zorunlu rotasyon dışında bağımsız denetim firma tercihlerinde hangi faktörlerden etkilendiklerine dair kamuoyunu bilgilendirme zorunluluğu bulunmamaktadır. Bu durum, şirket performansı, denetim kalitesi gibi hususlar hakkında fikir sahibi olmak isteyen kreditörler, yatırımcılar ve aracı kurumlar gibi bilgi kullanıcıları için önemli bir eksiklik oluşturmaktadır. Literatürde, denetim firma tercihine etki eden birçok faktör olmasına karşın çalışmaların ağırlıklı noktasının finansal performans faktörleri olduğu görülmüştür. Tablo 1 ‘de denetim firma değişikliğini finansal performans ile açıklayan çalışmalar ve kullandıkları yöntemler listelenmiştir.

**Tablo 1.** Summary of studies on relationship between audit firm change and financial performance

|  |  |  |
| --- | --- | --- |
| **Author(s)** | **Methods used** | **Result(s)** |
| Dhaliwal (2007) | Logistic regression | Zayıf mali performans sergileyen işletme yönetiminin, işletmenin mali durumunu daha iyi gösteren bir denetim firmasını tercih edeceğini öne sürmüşlerdir. |
| Chadegani et al. (2011) | Logistic regression | Bağımsız denetim firma değişikliğini etkileyen 6 faktörden birinin finansal başarısızlık olduğunu tespit etmişlerdir. |
| Kwak et al. (2011) | Bayesian network | Denetim firmasını değiştiren şirketlerde; Net Çalışma Sermayesi/Toplam Aktif, Kısa Vadeli Borçlar/Dönen Varlıklar, Kar Payı Ödemesi, Faiz ve Vergi Öncesi Kar/Toplam Aktif değişkenlerinin denetim firmasını değiştirmeyen işletmelere göre anlamlı şekilde farklı olduğu görülmüştür. |
| Nazri et al. (2012) | Survey, Logistic regression | Değişkenler; yönetimdeki değişiklik, olumsuz denetim raporu, işletme büyüklüğü, denetim kalitesi, finansal yapı, denetim ücreti gibi unsurlardan oluşmaktadır. Çalışmanın sonucunda, borçluluk seviyesindeki artışın müşteri işletmenin denetim firmasını değiştirme olasılığını artırdığı bulunmuştur. |
| Huang ve Scholz (2012) | Panel data regression | Finansal verilerde sahtecilik (karı zarara çevirme vs.) yapan işletmelerin denetçi istifasına neden olduğunu ve sonrasında daha büyük denetim firmaları ile anlaşmaya çalıştıklarını tespit etmişlerdir. |
| Eldridge et al. (2012) | Logistic discriminant | Denetim firması değişikliği 13 finansal başarısızlık değişkenine bağlı olarak tahminlemeye çalışılmıştır. Denetim firması değişikliklerinin tahmininde etkili olan değişkenler; kar payı, zarar, dağıtılmamış kar/toplam aktif, faiz ve vergi öncesi kar/ toplam aktif olarak bulunmuştur. |
| Suyono et al. (2013) | Questionnaire survey | İşletmenin finansal performansı ve denetim firması değişikliği arasında anlamlı ve pozitif şekilde ilişki ortaya çıkmıştır. En önemli göstergeler; finansal performans, denetim firmaları arasındaki rekabet ve denetim kalitesi olmuştur. |
| Black et al. (2013) | Logistic regression, logistic discriminant | Denetim firma değişikliğini açıklamada 4 finansal oranın (Net İşletme Sermayesi/Toplam Aktif, Aktif Getirisi, Özkaynak/Toplam Borç, Satışlar/Toplam Aktif) etkili olduğu tespit edilmiştir. Ayrıca, lojistik regresyon yönteminin denetim firma değişikliğini doğru sınıflandırma başarısı (%92,3) iken diskriminant analizinin (%47,6) olduğu tespit edilmiştir. |
| Aroh et al. (2017) | Logistic regression | Finansal başarısızlık değişkeni ile denetim firma değişikliği arasında pozitif bir ilişki tespit edememiştir. |
| Kusuma ve Farida (2019) | Discriminant, logistic regression | Finansal başarısızlık, karlılık ve yönetim değişikliğinin denetim firması değişikliğinde önemli bir etkiye sahip olduğunu bulmuşlardır. |
| Dharmasari ve Suardana (2021) | Logistic regression | Endonezya borsasına kayıtlı 25 üretim işletmesinin mali tabloları analiz edilmiş ve finansal zayıflık ile denetim firma değişikliği arasında pozitif bir ilişki bulunmuştur. |

Buradan hareketle, Türkiye’ de halka açık şirketlerin denetim firma tercihini finansal oranlar yardımıyla tahminlemenin etkili olacağı düşünülmüştür. Literatürdeki çalışmalar genel olarak değerlendirildiğinde; anket, soru kağıdı, lojistik regresyon ve diskriminant analizi gibi istatistiki yöntemlerin çoğunlukla kullanıldığı tespit edilmiştir. Makine öğrenme algoritmalarından yararlanarak bu ilişkiyi açıklayan çalışmalara rastlanılmaması sebebiyle, bağımsız denetim firma tercihi ve finansal oranlar arasındaki ilişkinin SVM (Support Vector Machine-Destek Vektör Makineleri), DT (Decision Tree-Karar Ağaçları), RF (Random Forest-Rastgele Orman), NB (Naive Bayes), KNN (K-Nearest Neighbors-K-En Yakın Komşu), XGBoost gibi makine öğrenmesi algoritması ile tahminlemenin literatürdeki eksikliği dolduracağı düşünülmektedir.

**2.** **Methodology**

Öğrenme işlemi sınıflandırma, regresyon, kümeleme ve nitelik çıkarımı gibi işlemleri kapsamaktadır. Veriden öğrenme ise, sınırlı sayıda eğitim verisinden yola çıkılarak tahmin fonksiyonunun belirlenmesini sağlayan öğrenme makinesinin veya algoritmanın oluşturulması olarak tanımlanır. Bu karakteristiklerinden dolayı pazarlama, finans, yönetim, mühendislik, üretim, endüstri, bankacılık, sağlık, turizm ve tıp gibi birçok alanda veriye dayalı yöntemlerle (veri madenciliği) öğrenme kullanılmaktadır (Chunhong ve Licheng, 2004, s.1869). Elde edilen bu veri setlerinin ortak özelliklerine göre belirli sınıflara ayrılması aşamasında çeşitli sınıflandırma yöntemlerine ihtiyaç duyulmaktadır. Bu amaçla çeşitli algoritmalar geliştirilmiştir. Başlıca algoritmalar, entropi tabanlı sınıflandırma (C4.5 algoritması, C5.0 algoritması), Regresyon ve Karar Ağaçları (Gini algoritması, Twoing algoritması), Bellek Tabanlı Algoritmalar (k-en yakın komşu algoritması), Bayes Sınıflandırıcılar, Regresyon Ağaçları, Destek Vektör Makineleri, Rastgele Orman şeklinde sıralanabilir (Oleg, 2010, s.11).

Bu algoritmalar içerisinde, denetim firma tercihinin etkin bir tahmininin yapılıp yapılmayacağını ortaya koymak ve belirlenen veri setlerinin (finansal oranların) önem derecelerini belirlemek için, bağımlı değişken (denetim firma tercihi) ile girdi değişkenler (finansal oranlar) arasındaki ilişki Destek Vektör Makineleri, Karar Ağaçları, Rastgele Orman, Naive Bayes, K-en Yakın Komşu ve XGBoost sınıflandırma algoritmaları kullanılarak açıklanmıştır. Burada amaç, bağımlı ve girdi değişkenler arasındaki ilişkiyi en iyi uyuma sahip olacak biçimde tahminleyen (prediction) bir model kurmaktır. Araştırmada kullanılan makine öğrenmesi yöntemlerine ilişkin açıklayıcı bilgilere sırasıyla yer verilmektedir.

**2.1.Destek Vektör Makineleri (Support Vector Machine-SVM):** 1990’larda geliştirildiklerinde son derece popüler olmuş ve yüksek performanslı bir algoritma için en iyi yöntemlerden birisi haline gelmiştir (Pradhan, 2012, s.82). Destek vektör makineleri, Vapnik-Chervonenkis teorisine dayanan, güçlü temelleri olan denetimli bir makine öğrenme algoritmasıdır. SVM; sinir ağları ve radyal tabanlı yapay sinir ağlarına benzerlik gösterse de genellikle bu algoritmalardan daha iyi performans gösterdiği belirtilmektedir. Bu öğrenme algoritması, Polinom, Radyal Temel Fonksiyon ve Çok Katmanlı Perceptron (Yapay sinir ağı) sınıflayıcıları için alternatif bir eğitim tekniği olarak görülmektedir. Tekniğin arkasındaki ana fikir, verilerden oluşan sınıflar arasındaki marjı maksimize eden optimal bir hiper düzlem meydana getirmektir (Pradhan, 2012, s.83). SVM, basit bir yapısı olması ve pratik uygulamalarda yüksek performans göstermesi bakımından oldukça kullanışlıdır. SVM’lerde kullanılacak örnek sayısı önemli değildir. SVM, eğitim esnasında görülmemiş verileri de sorunsuz olarak sınıflandırdığı için genelleştirebilme yeteneğine de sahiptir. Genelleştirebilme özelliği SVM’yi diğer tekniklere göre iyi bir alternatif yapmaktadır. SVM’ler, sınıflandırma ve regresyon problemlerinde çalıştırılabilir. Bu nedenle, borsa endeks hareketlerinin tahmin edilmesi, finansal bilgi manipülasyonları öngörüsü ya da finansal başarısızlıklara karşı erken uyarı sistemlerinin oluşturulması gibi birçok alanda istatistiki yöntemlere göre daha başarılı oldukları gözlemlenmiştir (Cao and Tay, 2003, p.1506-1507).

SVM metodundaki temel fikir, eldeki eğitim verilerinin karakterini mümkün olduğunca gerçeğe yakın bir şekilde yansıtan ve istatistiksel öğrenme teorisine uyan doğrusal ayırıcı fonksiyonun bulunmasıdır. Doğrusal olmayan durumların işlenebilmesi, boyut değişikliğini sağlayan haritalama fonksiyonunun bilinmemesi ve yüksek boyutlarda işlem yapmanın zor olması nedeniyle çekirdek düzenlemesi (kernel trick) denilen düzenlemeler yapılır (Hoffman, 2006, s.4). Böylelikle, dönüştürülmüş uzaydaki haritalama fonksiyonu yerine, doğrudan girdi uzayındaki verilerin kullanımı sağlayan çekirdek fonksiyonlar işleme dahil edilmiş olur. Destek Vektör Makineleri (SVM) algoritmasında yaygın olarak kullanılan dört çekirdek fonksiyon vardır. Bunlar (Nanda vd., 2018, p.6):

* Lineer Fonksiyon,
* Radyal Tabanlı Fonksiyon,
* Polinomiyal Fonksiyon,
* Sigmoid Fonksiyon,

Çekirdek fonksiyonları kullanılarak bütün değerlerin tekrar tekrar çarpım değerlerinin hesaplanarak bulunması yerine doğrudan çekirdek fonksiyonunda değerin yerine koyularak nitelik uzayındaki değerinin bulunması sağlanır. Bu sayede, son derece yüksek boyutlu bir nitelik uzayı ile uğraşma olasılığı kalmaz.

**2.2.Karar ağaçları (Decision Tree-DT):** Genellikle sınıflandırma ve tahmin için kullanılan bir veri madenciliği yaklaşımıdır (Song ve Lu, 2015, s.130). Yapay sinir ağı gibi diğer yöntemlerde sınıflandırma için kullanılmalarına rağmen, doğrulama yapabilmek için kendi kararlarını onların etki alanı bilgisiyle karşılaştırarak kolay yorumlama ve anlaşılır olma gibi birçok avantajı bünyesinden barındırmaktadır (Nassif vd., 2013, s.220). Buna ek olarak varsayımlar gerektirmeden çeşitli verilerin analizine de imkân tanımaktadır. Bir karar ağacının oluşturulmasında atılacak ilk adım, ağaçtaki dallanmanın hangi kritere göre yapılacağı ya da hangi öznitelik değerlerine göre ağaç yapısının oluşturulacağını belirlemektir (Seethapathy ve Babu, 2021, s.13). Bunun için literatürde çok çeşitli algoritmalar geliştirilmiştir. Bunlara; C4.5, C5.0, CHAID, CART ve ID3 örnek olarak verilebilir (Patel ve Prajapati, 2018, s.75; Seethapathy ve Babu, 2021, s.13-14). Karar ağaçları, veritabanları ile bütünleşik olarak kolaylıkla uygulanabilmesi ve güvenilir olması sebebiyle yaygın bir kullanım alanına sahiptir (Nassif vd., 2013, s.220). Yöntem, en güçlü ve en iyi tahmini elde edebilmek için bağımlı ve bağımsız değişkenler arasında olası tüm ilişkilerin araştırılmasına dayanmaktadır (Seethapathy ve Babu, 2021, s.14-15). En kuvvetli ilişkiye sahip bağımsız değişken bulunduğunda, veri kümesi bu bağımsız değişkenin değerlerine göre ikiye ayrılmaktadır. Söz konusu süreç olası bölünmeler tamamlanıncaya kadar devam ettirilmektedir (Gupta vd., 2017, s.16). Karar ağaçlarının sunduğu mantıksal modelin yansıttığı karar kuralları, insanlar tarafından kolayca anlaşılabilecek kadar açıktır. Yüksek sınıflandırma doğruluk oranı ve üretilen basit kurallar gibi özelliklere sahip olduğundan dolayı bu yöntem geniş bir uygulama yelpazesine sahiptir (Barros vd., 2012, s.1237).

**2.3.Rastgele Orman algoritması (Random Forest-RF):** Breiman (2001) tarafından geliştirilmiş karar ağaçları ve düğümlerden oluşan bir yöntemdir. Bu yöntem ile regresyon ve sınıflandırma analizleri yapılabilmektedir (Liaw ve Wiener, 2022, s.18). Yönteme göre düğümleri dallara ayırmak için düğümlerdeki rastgele değerlerden en iyi olanı seçilmekte ve oluşturulan karar ağaçlarına belirli ağırlıklar verilmektedir (Siva vd., 2012, s.133). Bu ağırlıklar karar ağaçlarının iç hatalarına göre belirlenmekte ve en düşük hataya sahip olan karar ağacına en yüksek, en yüksek hataya sahip karar ağacına en düşük ağırlık verilmektedir. Verilen bu ağırlıklar sınıf tahmininde oy verme işlemi için kullanılmaktadır. Daha sonra bu oylar toplanarak nihai karar verilmektedir (Chen vd., 2009, s.244).

**2.4.Naive Bayes algoritması (Naive Bayes-NB):** Gausssan Naif Bayes sınıflandırıcısı veya genellikle sadece Naif Bayes olarak adlandırılan sınıflandırma algoritması, Bayes teoremine dayalı basit bir olasılık tabanlı sınıflandırma yöntemidir. Naif Bayes yöntemi tahmine dayalı modelleme yapmak için basit ama güçlü bir algoritmadır. Bu yüzden özellikle sinyal ve görüntü işleme alanlarında en çok kullanılan sınıflandırma, tahmin algoritmalarında biridir. Bu algoritmada bir sınıftaki belirli özelliklerin varlığının diğer herhangi bir özellik ile ilgisi olmadığı varsayılır (Srinivasa vd., 2020, s.7594).

**2.5.XGBoost algoritması (XGBoost):** XGBoost algoritması Tianqi Chen ve Carlos Guestrin tarafından 2016 yılında Washington Üniversitesi'nde bir araştırma projesi olarak geliştirildi (Chen vd., 2016, s.789). XGBoost algoritması gradyan arttırma çerçevesini kullanan karar ağacı tabanlı topluluk öğrenme temelli makine öğrenme algoritmasıdır. Çok sayıda Kaggle yarışmasını kazanan algoritma olmasıyla ilgiyi üzerinde toplamıştır. XGboost, regresyon ve sınıflandırma problemlerinde yapılandırılmış veya tablo halindeki veri setlerinde kullanılır (Zhou vd., 2021, s.28). XGBoost algoritması, ağaçları oluşturmak için CART (Classification and Regression Tree) algoritmasını kullanır. XGBoost, sistem optimizasyonu ve algoritmik geliştirmeler yoluyla temel Gradient Boosting Machines çerçevesinin geliştirilmiş bir uygulamasıdır (Chen vd., 2016, s.789).

**2.6.K-en Yakın Komşu algoritması (K-Nearest Neighbors-KNN):** Sınıflandırmada temel amaç, nesnelerin sahip olduğu özelliklere bakılarak nesnelerin hangi sınıfa ait olduğunun belirlenme işlemidir. Çok farklı sınıflandırma türleri ve algoritmaları bulunmaktadır. Karar ağaçları, en yakın komşu, bayes, yapay sinir ağları bunlardan bazılarıdır. KNN algoritması ya da diğer adıyla K-En yakın komşu algoritması makine öğrenim algoritmaları içerisinde en çok bilinen ve kullanılan algoritmalardan biridir (Hu vd., 2016, s.7). Seçilen bir özelliğin kendine en yakın olan özellikle arasındaki yakınlığı kullanarak sınıflandırma yapılır. Burada bulunan K değeri örnek olarak 3 veya 5 gibi bir sayı ile ifade edilir. Çalışma şekline baktığımızda, tanımlanan verilere göre yeni bir tanımlanması gereken nesne geldiğinde öncelikle K değerine bakılır. Burada eşitlik olmaması için genellikle K sayısı tek sayı olarak seçilir. Yeni gelen veri ile diğer veriler arasındaki mesafeler hesaplanırken Kosinüs, Öklid ya da Manhattan uzaklığı gibi yöntemler kullanılır (Han vd., 2011, s.45).

**3.Dataset**

Yatırımcılar, borç verenler ve düzenleyici kuruluşlar gibi bilgi kullanıcıları açısından büyük önem arz eden mali şeffaflık ilkesinin piyasalarda tam anlamıyla tatbikinin tartışıldığı bu ortamda işletmelerin bağımsız denetim firma tercihlerini hangi faktörlere göre yapıldığına dair herhangi bir bilgi paylaşımı söz konusu değildir. Bu eksiklik nedeniyle, işletmelerin bağımsız denetim firma tercihlerinde hangi faktörlerden etkilendiğine dair literatür araştırması yapılmış ve birçok neden tespit edilmesine rağmen çalışmaların ağırlıklı noktasının denetim ücreti ve finansal performans faktörleri olduğu görülmüştür (Örn: Nasser vd., 2006; Wan Mohamed vd., 2007; Ettredge vd., 2007; Ismail vd., 2008; Calderon ve Ofobike, 2008; Chen vd., 2008; Kwak vd., 2011; Eldridge vd., 2012; Huang ve Scholz, 2012; Suyono vd., 2013; Black vd., 2013). Ancak, Türkiye’de denetim ücretlerine ilişkin veriler, denetim firmalarının sadece o yılın toplam denetiminden elde ettiği geliri göstermesi ve müşteri işletme bazında denetim ücretinin yansıtılmaması nedeniyle, araştırma, finansal performans değişkeni üzerinde yoğunlaştırılmıştır. Bu anlamda, çalışma ile amaçlanan, Türkiye Borsa İstanbul (BIST) Yıldız Pazar’da yer alan işletmelerin bağımsız denetim firma tercihlerini etkileyen faktörleri Makine Öğrenme algoritmaları kullanarak tahmin etmektir. Makine öğrenme algoritmalarında kullanılmak üzere oluşturulan 1 bağımlı, 13 girdi değişkenlerin hesaplanmasına ilişkin bilgiler Tablo 2’de sunulmuştur.

**Tablo 2.** Modelde Kullanılan Değişkenler

|  |  |  |
| --- | --- | --- |
| **Bağımlı Değişken** | **Girdi Değişkenler** | **Hesaplanışı** |
|  | Cari Oran (CO) | Dönen varlıklar / K.V. finansal borçlar |
|  | Likidite Oranı (LO) | (Dönen varlıklar-Stoklar) / K. V. finansal borçlar |
|  | Nakit Oran (NO) | (Dönen varlıklar-Stoklar-Alacaklar) /K.V. finansal borçlar |
|  | Finansal Borç Oranı (FBO) | (K.V. Finansal Borçlar + U.V. Finansal Borçlar) / Toplam Varlıklar |
| Denetim Firma Tercihi (DFT) | Kaldıraç Oranı (KO) | Toplam Borçlar / Toplam Varlıklar |
|  | Aktif Devir Hızı (ADH) | Net satışlar / Aktif toplamı |
|  | Stok Devir Hızı (SDH)  Aktif Karlılık (AK)  Brüt Kar Marjı (BKM)  Favök Marjı (FM)  Net Kar Marjı (NKM)  Özkaynak Karlılığı (OK)  Yatırılan Sermayenin Getirisi (ROIC) | Satışların maliyeti / Ortalama stok tutarı  Net kâr / Aktif Toplamı  Brüt satış kârı / Net satışlar  FAVÖK / Toplam Satışlar  Net kâr / Net satışlar  Net kâr / Özkaynaklar  Vergi Sonrası Net Faaliyet Karı / Yatırılan Sermaye |

**Kaynak:** Saalem ve Rehman, 2011, s.97; Adjirackor vd., 2017, s.6

Çalışmada kullanılan bağımlı değişken Denetim Firma Tercihi (DFT) olarak tanımlanırken, girdi değişkenler ise Aktif Karlılık (AK), Finansal Borç Oranı (FBO), Net Kar Marjı (NKM), Yatırılan Sermayenin Getirisi (ROIC), Favök Marjı (FM), Kaldıraç Oranı (KO), Stok Devir Hızı (SDH), Cari Oran (CO), Brüt Kar Marjı (BKM), Nakit Oran (NO), Likidite Oranı (LO), Aktif Devir Hızı (ADH), Özkaynak Karlılığı (OK) olarak belirlenmiştir.

İşletmelerin finansal kabiliyetlerini en iyi temsil ettiği düşünülen bu girdi değişkenlere ilişkin veriler 2019-2021 yılları arasını kapsayacak şekilde Fintables Bilişim Teknolojileri A.Ş.’ nin işletmelere ait finansal veri analizlerinin yer aldığı web sitesi üzerinden alınmıştır. Finansal oranlar ile bağımsız denetim firma tercihi arasındaki ilişkiyi ortaya koyabilmek amacıyla belirlenen bu girdi değişkenler pek çok araştırmada (Örn: Cheng vd., 2005; Lei ve Liu, 2021; Yim ve Mitcbell, 2005; Lin, 2009) finansal performansı doğrudan etkileyebilen göstergeler olarak alınması sebebiyle bu çalışmada da kullanılması uygun görülmüştür.

Araştırmanın bağımlı değişkeni sınıflandırılırken, söz konusu işletmelerin denetim firma tercihi (DFT) dört büyük denetim firmasından biri (Ernst and Young-Güney Bağımsız Denetim, PWC-Başaran Nas Bağımsız Denetim, Deloitte-DRT Bağımsız Denetim, KPMG-Akis Bağımsız Denetim) ise= 1, dört büyükler dışındaki denetim firması ise= 0 değeri verilerek oluşturulmuştur.

BIST Yıldız Pazar’da işlem gören ambalaj, kimya ve plastik, metal eşya ve makine, tekstil giyim ve deri, ulaştırma, imalat, ana metal, savunma, bilişim ve yazılım, kağıt ürünleri, taş toprak çimento, mobilya, toptan ve perakende ticaret gibi sektöreler araştırmaya dahil edilmiştir. Bankalar, faktöring kuruluşları, finansal kiralama şirketleri, yatırım ortaklıkları gibi şirketlerin faaliyet konuları ve finansal tablo formatlarının farklılık arz etmesi, ayrıca, varlık yapılarının büyük ölçüde nakit ve nakit benzeri değerlerden oluşması nedeniyle çalışma dışı tutulmuştur. 2019-2021 yılları arasında faaliyetleri kesintiye uğrayanlar, iflas edenler vb. durumlardan dolayı varlığı sonlanan işletmeler de veri setinden çıkarılmıştır.

Klasik istatistiksel yöntemler, tahminleme aşamasında kısmen başarılı olabildikleri için son yıllarda örneklerden öğrenme yapabilen, daha az veri seti ile daha yüksek doğruluk oranı verebilen, genelleme yeteneğine sahip ve istatistiksel yöntemlere göre daha az varsayımları olan makine öğrenmesi yöntemlerinden SVM, RF, DT, KNN, NB ve XGBoost algoritmalarının kullanılması tercih edilmiştir.

**4.** **Experiment and Results**

Çalışma, Python 3.6 scikit-learn kütüphanesi kullanılarak kodlanmıştır. Araştırmanın veri setini, 2019-2021 dönemleri arasındaki 477 (159 şirket x 3 yıl) adet veri oluşturmaktadır. Değişken grubuna ait verilerin %80’i eğitim, %20’si ise test verisi olarak kullanılmıştır. Araştırmaya ilişkin gerçekleştirilen işlem adımları ve bu işlemlere ait bilgiler Şekil 1’de gösterilmiştir.

metin, ekran görüntüsü, diyagram, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Figure 1.** The Process of the Experiment

Şekil 1’ de görüldüğü üzere toplanan veriler ilk olarak, sınıflandırmada kullanılmak üzere asgari-azami normalleştirme (min-max normalization) yöntemi ile normalize edilmiştir. Bu yöntemde, bir grup veri içerisinde yer alan en küçük (0.01) ve en büyük (0.99) değerler alınıp, diğer bütün veriler bu değer aralığında (0.01-0.99) normalleştirilir (Wu vd., 2016, s.219). Bu durum matematiksel olarak (Arora vd., 2021, s.1337),

Input parameter value : Xi,t

New value : Xt

şeklinde formüle edilir. Tablo 3’ de Veri setine ait tanımlayıcı istatistikler verilmiştir.

**Tablo 3.** Başlangıç Veri Seti Tanımlayıcı İstatistikler

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CO** | **LO** | **NO** | **FBO** | **KO** | **ADH** | | **SDH** | **AK** | **BKM** | **FM** | **NKM** | **OK** | **ROIC** |
| **Adet** | 474 | 474 | 474 | 474 | 474 | | 474 | 474 | 474 | 474 | 474 | 474 | 474 | 474 |
| **Ort.** | 1.9 | 1.4 | 0.7 | 26.7 | 53.8 | | 0.9 | 9.9 | 8.9 | 25.5 | 16.0 | 11.2 | 19.8 | 22.5 |
| **Std.** | 1.9 | 1.8 | 1.6 | 20.4 | 25.2 | | 0.7 | 15.6 | 12.5 | 20 | 19.2 | 25.4 | 29.3 | 36.6 |
| **Min** | 0.1 | 0.09 | 0 | 0 | 0.9 | | 0 | -0.2 | -38.1 | -37.9 | -84.5 | -116 | -118 | -103.2 |
| **25%** | 1 | 0.6 | 0.09 | 9.03 | 35.8 | | 0.4 | 3.1 | 1.7 | 13.8 | 7.9 | 1.7 | 4.2 | 4.5 |
| **50%** | 1.3 | 0.9 | 0.2 | 25.5 | 57.0 | | 0.8 | 5.3 | 6.3 | 22.2 | 15.2 | 7.6 | 18.5 | 21 |
| **75%** | 2.1 | 1.5 | 0.6 | 39.6 | 72.7 | | 1.2 | 9.8 | 14.4 | 32.5 | 22.3 | 16.9 | 32.2 | 42 |
| **Max** | 14.6 | 14.5 | 17.3 | 158.8 | 260.5 | | 4.3 | 124.6 | 75 | 118.4 | 94.9 | 177.1 | 339.5 | 199 |

Veri seti içerisinde, önem derecesi yüksek özniteliklerin bulunması için karşılıklı bilgi öznitelik değerlendirme yöntemi (Mutual Information) kullanılmıştır. Mutual Information (MI) bir rastgele değişkenin diğer rastgele değişken hakkında elde edebileceği bilgi miktarını ölçmektedir (Witten, 2016, s.310). Matematiksel olarak karşılıklı bilgi, iki soyut rassal değişken olan x ve y için şu şekilde ifade edilir (Peng vd., 2005, s.1226):

Burada, p(x,y) bağıl yoğunluk fonksiyonu,

x ve y için, p(x) ve p(y) marjinal olasılık yoğunluk fonksiyonu.

MI kullanılma amacı, herhangi bir sınıflandırıcıdan bağımsız oluşu ve herhangi bir parametre ayarı gerektirmemesidir. Dolayısıyla, uygulaması kolay ve farklı veriler için iyi bir genelleştirme yeteneğine sahiptir. Öznitelik değerlendirme yöntemi olan MI’ya göre Tablo 4’de değişkenlerin önem sırası gösterilmiştir. Değişkenler, Tablo 4’ deki önem sırasına göre sırasıyla ilk 3 değişken olan (AK, NKM, FBO) Ranking (1)-(3), (AK, NKM, FBO, ROIC) Ranking (1)-(4), … Ranking (1)-(13) şeklinde 11 gruptan oluşmaktadır. Her bir grup, bağımlı değişken olarak belirtilen denetim firma tercihini tahmin etmek için makine öğrenmesi sınıflandırma algoritmaları olan SVM, RF, DT, KNN, NB ve XGBoost yöntemlerine girdi değişken (input variable group) olarak verilmiştir.

**Tablo 4.** Çalışmada Kullanılan Değişkenlerin Mutual Information Önem Sırası

|  |  |  |  |
| --- | --- | --- | --- |
| **Input Variable** | **Ranking** | **Mutual Information Value** | **Input Variable Group** |
| AK | 1 | 0.0580 |  |
| NKM | 2 | 0.0362 |  |
| FBO | 3 | 0.0300 | Ranking (1)-(3) |
| ROIC | 4 | 0.0284 | Ranking (1)-(4) |
| FM | 5 | 0.0256 | Ranking (1)-(5) |
| KO | 6 | 0.0211 | Ranking (1)-(6) |
| NO | 7 | 0.0210 | Ranking (1)-(7) |
| SDH | 8 | 0.0202 | Ranking (1)-(8) |
| ADH | 9 | 0.0063 | Ranking (1)-(9) |
| BKM | 10 | 0.0053 | Ranking (1)-(10) |
| OK | 11 | 0.0023 | Ranking (1)-(11) |
| CO | 12 | 0.0000 | Ranking (1)-(12) |
| LO | 13 | 0.0000 | Ranking (1)-(13) |

|  |
| --- |
|  |
| **Figure 2.** Importance Value of Features |

Makine öğrenme algoritması için onlarca hiper parametre ve bu hiper parametrelerin alabileceği onlarca değer olduğu düşünüldüğünde, her değişken grubunda, SVM, RF, DT, KNN, NB ve XGBoost yöntemlerinde kullanılacak en iyi parametre kombinasyonunu seçmek gerekmektedir. Bu anlamda, özellikle küçük veri setleri ile çalışıldığında, en iyi performansı gösteren hiperparametre setini belirlemede başarılı sonuçlar veren GridSearchCV (GridSearch+Cross-Validation) hiper parametre optimizasyon yöntemi tercih edilmiştir (Shuai vd., 2018, s.451). Bu yöntemde, denenmesi istenen hiper parametreler ve değerleri için bütün kombinasyonlar ile ayrı ayrı model kurulur ve belirlenen metriğe göre en başarılı hiper parametre seti belirlenir. Araştırmada, GridSearchCV yönteminde kullanılan parametreler ve olası değerleri Tablo 5’de gösterilmiştir.

**Tablo 5.** Parametre Optimizasyonunda Kullanılan Değerler

|  |  |  |
| --- | --- | --- |
| **Models** | **Parameter** | **Parameter’s Value** |
| SVM | C  kernel  degree | 0.1,1,5,10,15,20,100,1000  rbf,linear,poly,sigmoid  3,8 |
| RF | n estimators  max depth  min samples leaf  min samples split  criterion  max features | 1,5,10,50,100  4,5,6,7,8,9,10  2,10,20  10,15,20  gini,entropy  auto,sqrt |
| DT | max leaf nodes  min samples leaf  min sample split  criterion  max depth  max features | 2-100  1-5  2,3,4  gini,entropy  2,3,4,5,6,7,8,9,10,11,12  auto,sqrt,log2 |
| KNN | n\_neighbors  leaf\_size  p  weights  algorithm  metric | range(1,30)  20,40,1  1,2  'uniform', 'distance'  'auto', 'ball\_tree', 'kd\_tree', 'brute'  'minkowski', 'chebyshev' |
| NB | Priors  var\_smoothing | [None, [0.1,]\* len(n\_classes),]  [1e-9, 1e-6, 1e-12] |
| XGBoost | Objective  learning\_rate  max\_depth  min\_child\_weight  subsample  colsample\_bytree  n\_estimators | ['binary:logistic']  [0.001, 0.01, 0.1, 0.20, 0.25, 0.30]  [3,4,5,6,8,10,12,15]  [1,5,10,11]  [0.8]  [0.7]  [5,100,500,1000] |

Modeller için gerekli parametreler GridSearchCV yöntemiyle bulunduktan sonra SVM, RF, DT, KNN, NB ve XGBoost makine öğrenmesi yöntemlerinin her biri ile 11 farklı değişken grubu için 66 adet tahmin modeli kurulmuştur. Tahmin modellerinin başarısını ölçmek için sınıflandırmada kullanılan performans metriklerinden accuracy (doğruluk), precision (kesinlik), recall (duyarlılık), f1-score (f1 puanı) ve Cohen’s Kappa değerleri hesaplanarak Tablo 6’ da sunulmuştur.

**Tablo 6.** Prediction Result of Models According to the Variants of Input Variable

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Test** | | | | |
| **Model** | **Input Variable Group** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **Kappa** |
| SVM | Ranking (1)-(3) | 0.621 | 0.521 | 0.621 | 0.501 | -0.006 |
| Ranking (1)-(4) | 0.642 | 0.616 | 0.642 | 0.568 | 0.09 |
| Ranking (1)-(5) | 0.747 | 0.75 | 0.747 | 0.729 | 0.408 |
| Ranking (1)-(6) | 0.716 | 0.717 | 0.716 | 0.688 | 0.321 |
| Ranking (1)-(7) | 0.747 | 0.742 | 0.747 | 0.739 | 0.43 |
| Ranking (1)-(8) | 0.705 | 0.696 | 0.705 | 0.695 | 0.335 |
| Ranking (1)-(9) | 0.737 | 0.768 | 0.737 | 0.7 | 0.354 |
| Ranking (1)-(10) | 0.695 | 0.684 | 0.695 | 0.683 | 0.307 |
| Ranking (1)-(11) | 0.684 | 0.672 | 0.684 | 0.67 | 0.278 |
| Ranking (1)-(12) | 0.684 | 0.673 | 0.684 | 0.656 | 0.25 |
| Ranking (1)-(13) | 0.695 | 0.689 | 0.695 | 0.665 | 0.27 |
| RF | Ranking (1)-(3) | 0.653 | 0.632 | 0.653 | 0.625 | 0.18 |
| Ranking (1)-(4) | 0.726 | 0.735 | 0.726 | 0.697 | 0.341 |
| Ranking (1)-(5) | 0.716 | 0.711 | 0.716 | 0.713 | 0.378 |
| Ranking (1)-(6) | 0.758 | 0.753 | 0.758 | 0.751 | 0.457 |
| Ranking (1)-(7) | 0.758 | 0.757 | 0.758 | 0.745 | 0.443 |
| Ranking (1)-(8) | 0.768 | 0.765 | 0.768 | 0.761 | 0.478 |
| Ranking (1)-(9) | 0.789 | 0.791 | 0.789 | 0.78 | 0.519 |
| Ranking (1)-(10) | 0.789 | 0.787 | 0.789 | 0.786 | 0.537 |
| Ranking (1)-(11) | 0.705 | 0.698 | 0.705 | 0.698 | 0.343 |
| Ranking (1)-(12) | 0.811 | 0.808 | 0.811 | 0.808 | 0.583 |
| Ranking (1)-(13) | 0.779 | 0.776 | 0.779 | 0.777 | 0.516 |
| DT | Ranking (1)-(3) | 0.632 | 0.608 | 0.632 | 0.608 | 0.142 |
| Ranking (1)-(4) | 0.653 | 0.647 | 0.653 | 0.649 | 0.24 |
| Ranking (1)-(5) | 0.653 | 0.689 | 0.653 | 0.659 | 0.307 |
| Ranking (1)-(6) | 0.705 | 0.696 | 0.705 | 0.695 | 0.335 |
| Ranking (1)-(7) | 0.684 | 0.698 | 0.684 | 0.688 | 0.345 |
| Ranking (1)-(8) | 0.663 | 0.659 | 0.663 | 0.661 | 0.267 |
| Ranking (1)-(9) | 0.768 | 0.766 | 0.768 | 0.767 | 0.496 |
| Ranking (1)-(10) | 0.674 | 0.71 | 0.674 | 0.679 | 0.349 |
| Ranking (1)-(11) | 0.737 | 0.759 | 0.737 | 0.741 | 0.463 |
| Ranking (1)-(12) | 0.737 | 0.752 | 0.737 | 0.741 | 0.457 |
| Ranking (1)-(13) | 0.705 | 0.714 | 0.705 | 0.708 | 0.381 |
| NB | Ranking (1)-(3) | 0.642 | 0.614 | 0.642 | 0.588 | 0.115 |
| Ranking (1)-(4) | 0.663 | 0.647 | 0.663 | 0.62 | 0.178 |
| Ranking (1)-(5) | 0.684 | 0.673 | 0.684 | 0.656 | 0.25 |
| Ranking (1)-(6) | 0.737 | 0.733 | 0.737 | 0.734 | 0.424 |
| Ranking (1)-(7) | 0.579 | 0.619 | 0.579 | 0.586 | 0.165 |
| Ranking (1)-(8) | 0.579 | 0.671 | 0.579 | 0.578 | 0.216 |
| Ranking (1)-(9) | 0.568 | 0.653 | 0.568 | 0.568 | 0.193 |
| Ranking (1)-(10) | 0.579 | 0.66 | 0.579 | 0.58 | 0.208 |
| Ranking (1)-(11) | 0.568 | 0.664 | 0.568 | 0.566 | 0.201 |
| Ranking (1)-(12) | 0.558 | 0.657 | 0.558 | 0.554 | 0.186 |
| Ranking (1)-(13) | 0.547 | 0.65 | 0.547 | 0.542 | 0.171 |
| KNN | Ranking (1)-(3) | 0.663 | 0.653 | 0.663 | 0.655 | 0.249 |
| Ranking (1)-(4) | 0.716 | 0.708 | 0.716 | 0.708 | 0.363 |
| Ranking (1)-(5) | 0.726 | 0.721 | 0.726 | 0.722 | 0.398 |
| Ranking (1)-(6) | 0.747 | 0.743 | 0.747 | 0.744 | 0.444 |
| Ranking (1)-(7) | 0.726 | 0.721 | 0.726 | 0.722 | 0.398 |
| Ranking (1)-(8) | 0.653 | 0.664 | 0.653 | 0.657 | 0.275 |
| Ranking (1)-(9) | 0.789 | 0.788 | 0.789 | 0.782 | 0.525 |
| Ranking (1)-(10) | 0.811 | 0.811 | 0.811 | 0.804 | 0.572 |
| Ranking (1)-(11) | 0.779 | 0.775 | 0.779 | 0.775 | 0.51 |
| Ranking (1)-(12) | 0.779 | 0.776 | 0.779 | 0.773 | 0.504 |
| Ranking (1)-(13) | 0.768 | 0.764 | 0.768 | 0.763 | 0.484 |
| XGBoost | Ranking (1)-(3) | 0.684 | 0.684 | 0.684 | 0.684 | 0.321 |
| Ranking (1)-(4) | 0.716 | 0.748 | 0.716 | 0.669 | 0.292 |
| Ranking (1)-(5) | 0.737 | 0.735 | 0.737 | 0.72 | 0.387 |
| Ranking (1)-(6) | 0.768 | 0.771 | 0.768 | 0.755 | 0.464 |
| Ranking (1)-(7) | 0.758 | 0.753 | 0.758 | 0.751 | 0.457 |
| Ranking (1)-(8) | 0.884 | 0.884 | 0.884 | 0.884 | 0.75 |
| Ranking (1)-(9) | 0.874 | 0.874 | 0.874 | 0.872 | 0.722 |
| Ranking (1)-(10) | 0.853 | 0.853 | 0.853 | 0.853 | 0.683 |
| Ranking (1)-(11) | 0.821 | 0.819 | 0.821 | 0.819 | 0.608 |
| Ranking (1)-(12) | 0.863 | 0.863 | 0.863 | 0.863 | 0.704 |
| Ranking (1)-(13) | 0.811 | 0.808 | 0.811 | 0.808 | 0.583 |

Tablo 6’ya ait algoritmalarının performans metriklerinin hesaplanmasına ilişkin formüller aşağıda yer almaktadır (Briliani, 2019, s.102-103).

Accuracy (doğruluk) değerinin hesaplanabilmesi için aşağıdaki formül kullanılmıştır.

*Accuracy=*

Precision (kesinlik) değerinin hesaplanabilmesi için aşağıdaki formül kullanılmıştır.

*Precision=*

Recall (duyarlılık) değerinin hesaplanabilmesi için aşağıdaki formül kullanılmıştır.

*Recall=*

F1-score (F1 değeri) değerinin hesaplanabilmesi için aşağıdaki formül kullanılmıştır.

*F1-score=2\**

Cohen’s Kappa

Relative observed agreement among accuracy.

Formüllerde kullanılan, TP gerçek pozitif sayısını, TN gerçek negatif sayısını, FP yanlış pozitif sayısını, FN ise yanlış negatif sayısını ifade etmektedir. Makine Öğrenmesi algoritmalarının test sonuçlarına göre performans metriklerinin karşılaştırıldığı Şekil 3’ de XGBoost algoritmasının diğer makine öğrenmesi tekniklerinden daha başarılı tahmin sonucu verdiği görülmüştür.



**Figure 3.** ML Algoritmalarının Performans Metriklerine Göre Karşılaştırılması

Tablo 7’de modellerin en başarılı değişken gruplarına göre karşılaştırılması verilmiştir.

**Tablo 7.** Modellerin En Başarılı Değişken Gruplarına Göre Karşılaştırılması

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **Cohen’s Kappa** | **Ranking Group** |
| XGBoost | 0.884 | 0.884 | 0.884 | 0.884 | 0.75 | Ranking (1)-(8) |
| RF | 0.811 | 0.808 | 0.811 | 0.808 | 0.583 | Ranking (1)-(12) |
| KNN | 0.811 | 0.811 | 0.811 | 0.804 | 0.572 | Ranking (1)-(10) |
| SVM | 0.747 | 0.75 | 0.747 | 0.729 | 0.408 | Ranking (1)-(5) |
| DT | 0.737 | 0.759 | 0.737 | 0.741 | 0.463 | Ranking (1)-(11) |
| NB | 0.737 | 0.733 | 0.737 | 0.734 | 0.424 | Ranking (1)-(6) |

Performans metriklerine göre bir sıralama yapıldığında en iyi performansı, tahmin doğruluk (accuracy) değeri %88.4 (0.884) ile 8 değişkenli (AK, NKM, FBO, ROIC, FM, KO, NO, SDH) XGBoost algoritması göstermiştir. Bunu sırasıyla 12 değişkenli (AK, NKM, FBO, ROIC, FM, KO, NO, SDH, ADH, BKM, OK, CO) RF algoritması ve 10 değişkenli (AK, NKM, FBO, ROIC, FM, KO, NO, SDH, ADH, BKM) KNN algoritması takip etmiştir. RF ve KNN algoritmaları %81.1 tahmin doğruluk değeriyle aynı performansı göstermiştir. SVM, DT ve NB algoritmaları ise sırasıyla %74.7, %73.7, %73.7 tahmin doğruluk değerleri ile birbirine yakın sonuçlar vermiştir. SVM, DT ve NB algoritmaları ile en iyi doğruluk değerleri sırasıyla; 5, 11 ve 6 değişkenli tahmin modellerinde elde edilmiştir. DT ve NB en düşük performans gösteren makine öğrenme algoritmaları olmuştur.

Performans metriklerinden precision, recall, F1-Score ve Cohen’s Kappa değerlerine göre de en başarılı makine öğrenmesi yöntemi XGBoost algoritması olmuştur.

**5. Conclusion**

Bu çalışmada, Türkiye Borsa İstanbul Yıldız Pazar’da ambalaj, kimya ve plastik, metal eşya ve makine, tekstil giyim ve deri, ulaştırma, imalat, ana metal, savunma, bilişim ve yazılım, kağıt ürünleri, taş toprak çimento, mobilya, toptan ve perakende ticaret gibi sektörlerde yer alan işletmelerin denetim firma tercihleri makine öğrenmesi algoritmalarıyla tahmin edilmiştir.

Bu tahminlemede SVM, RF, DT, KNN, NB, XGBoost gibi makine öğrenmesi algoritmaları kullanılmıştır. Burada amaç, bağımlı ve girdi değişkenler arasındaki ilişkiyi en iyi uyuma sahip olacak biçimde tahminleyen (prediction) bir model kurmaktır. Modellere ait veri setlerinde bağımlı değişken denetim firma tercihiyken, girdi değişkenler finansal oranlar (AK, NKM, FBO, ROIC, FM, KO, NO, SDH, ADH, BKM, OK, CO, LO) olarak belirlenmiştir. Araştırmada kullanılan toplam 13 öznitelik içerisinden sırasıyla ilk 8 öznitelik kullanılarak en yüksek tahmin değerine ulaşılmıştır. Öznitelik seçiminde Mutual Information (MI) yöntemi uygulanarak önem değerleri bulunmuştur. Özniteliklerin önem sırasına göre sıralaması: AK, NKM, FBO, ROIC, FM, KO, NO, SDH şeklindedir. AK (Aktif Karlılık) verisinin en önemli öznitelik olduğu tespit edilmiştir.

SVM, RF, DT, KNN, NB, XGBoost makine öğrenmesi yöntemlerinin her biri ile 11 farklı değişken grubu için 66 adet tahmin modeli kurulmuştur. Tahmin modellerinin başarısını ölçmek için sınıflandırmada kullanılan performans metriklerinden accuracy (doğruluk), precision (kesinlik), recall (duyarlılık), f1-score (f1 puanı) ve Cohen’s Kappa değerleri hesaplanmıştır. Performans metriklerine göre yapılan sıralamada en iyi performansı, tahmin doğruluk (accuracy) değeri %88.4 ile 8 değişkenli (AK, NKM, FBO, ROIC, FM, KO, NO, SDH) XGBoost algoritması göstermiştir.

Bu ve benzeri yapay zeka destekli tahminleme çalışmalarıyla, halka açık şirketlerin denetim firma tercihlerinde isabetli öngörüler oluşturacağı düşünülmektedir. Gelecekteki çalışmalarda, daha büyük veri setleri oluşturarak farklı makine öğrenme yöntemleriyle tahminleme çalışmalarının gerçekleştirilmesi hedeflenmektedir.