

## Makine ve Ekipman İmalatı Sektöründe İzolasyon Ormanı ve Yeniden Örnekleme Yöntemleri Kullanılarak Finansal Başarısızlığın Tespit Edilmesi

Kardelen Şeyma ERDEM<sup>1</sup>, Mehmet Akif BAKIR<sup>2</sup>

### ÖZET

**Amaç:** İşletmeler zaman zaman finansal başarısızlık/iflas durumu ile karşı karşıya kalmaktadır. Finansal başarısızlığın öngörülebilmesi ile işletmelerin erken aksiyon almalarının sağlanabileceği değerlendirilmektedir. Literatürde bu amaç için yapılmış birçok çalışma bulunmaktadır. Bu çalışmada finansal başarısızlığın tespiti üzerine makine öğrenmesi tabanlı bir uygulama yapılmıştır.

**Yöntem:** Türkiye’de 2010-2018 yılları arasında Makine ve Ekipman İmalatı Sektöründe faaliyet gösteren işletmeler çalışma kapsamında tutulmuştur. İlk olarak, işletmeler finansal başarı durumlarına göre İzolasyon Ormanı Yöntemi kullanılarak etiketlenmiştir. Daha sonra, dengesiz veri problemini ortadan kaldırmak için yeniden örnekleme yöntemleri ile veri ön işleme adımı gerçekleştirilmiştir. Yeniden örnekleme yöntemi ile elde edilen üç farklı veri setine son yıllarda sıklıkla kullanılan XGBoost modeli uygulanmış ve model başarıları karşılaştırılmıştır.

**Bulgular:** Yeniden örnekleme yöntemlerinden biri olan SMOTE yöntemi ile oluşturulan veri seti üzerinde kurulan XGBoost modelinin %78 doğru sınıflandırma oranıyla en iyi tahmin gücüne sahip model olduğu belirlenmiştir.

**Özgünlük:** Çalışmada diğer çalışmaların aksine finansal başarısızlığa uğrayıp uğramadığı bilinmeyen işletmeler üzerinde çalışılmıştır. Öncelikle işletmelerin finansal başarısızlık durumlarına ilişkin tahminde bulunulmuş sonrasında ise yeniden örnekleme yöntemleri kullanılmıştır.

**Anahtar Kelimeler:** İflas Tahmini, Altman Z Skor, Simülasyon Modelleme, Makine Öğrenmesi.

**JEL Kodları:** G33, C45, C63.

## Detecting Financial Failure in the Machinery and Equipment Manufacturing Sector Using Isolation Forest and Resampling Methods

### ABSTRACT

**Purpose:** Businesses are faced with financial failure/bankruptcy from time to time. It is considered that by predicting financial failure, it is possible for businesses to take early action. There are many studies in the literature for this purpose. In this study, a machine learning-based application was made on the detection of financial failure.

**Method:** The enterprises operating in the Machinery and Equipment Manufacturing Sector in Turkey between the years 2010-2018 were included in the study. First, businesses were labeled according to their financial success using the isolation forest method. Then, the data preprocessing step was carried out with resampling methods to eliminate the unbalanced data problem. The XGBoost model, which has been used frequently in recent years, was applied to three different data sets obtained by the resampling method and the model successes were compared.

**Results:** It was determined that the XGBoost model built on the data set created by the SMOTE method, which is one of the resampling methods, has the best predictive power with a 78% correct classification rate.

**Originality:** Contrary to other studies, in the study, it was studied on businesses whose financial failure was unknown. First of all, the estimation of the financial failures of the enterprises was made and then resampling methods were used.

**Keywords:** Bankruptcy Forecast, Altman Z Score, Simulation Modeling, Machine Learning.

**JEL Codes:** G33, C45, C63.

<sup>1</sup> Yüksek Lisans Öğrencisi, Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye, kardelennerdem@gmail.com, ORCID:0000-0002-5197-939X (Sorumlu Yazar-Corresponding Author).

<sup>2</sup> Prof. Dr., Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye, mabakir@gazi.edu.tr, ORCID:0000-0003-0774-0338.

**EXTENDED ABSTRACT**

The prediction of financial failure has been the subject of research for many years. Predicting financial failure in advance allows enterprises to take measures at an early stage and overcome this process relatively easily. The most well-known studies developed for predicting financial failure in the literature are Ohlson O-Score, Altman Z-Score and Taffer model. These models are known as financial-based models. Ohlson, Altman and Taffer have shown that it is possible to predict the financial failure of enterprises. Similar studies conducted in the literature have confirmed this finding.

It is observed that almost all studies on the prediction of financial failure have kept the sample small. In addition, studies have used enterprises that are certain to experience financial failure and enterprises that are known not to experience financial failure. In this study, one of the main objectives is to ensure that the sample is as large as possible. With a large sample size, the machine learning algorithms used in this study are aimed to obtain more generalizable models. Another aim of the study is to label financially unsuccessful and financially successful ventures using the statistics-based isolation forest method. In addition, another issue that distinguishes this study from the literature is that the data labeled as financially successful-financially unsuccessful are subjected to resampling methods. The most successful model for predicting financial failure was obtained with different data sets obtained using resampling methods.

The study focuses on the prediction of financial failure of medium-high technology enterprises operating in the NACE Rev.2 28 sector of the manufacturing industry between 2010-2018. In order to predict financial failure, 27 financial ratios under the main headings of liquidity ratios, indebtedness ratios, turnover ratios and profitability ratios were calculated from the enterprise data. The data were then labeled as financial success and financial failure using the isolation forest method. In response to the unbalanced data set problem, three different data sets were obtained using resampling methods and predictive modeling was performed using XGBoost, one of the machine learning algorithms.

As a result of the study, it is determined that the resampling method with the highest predictive power is the SMOTE method with 78% correct classification success. According to the XGBoost model established with the SMOTE method, the three most important variables that were found to be effective in predicting financial failure current ratio, leverage ratio and short-term liabilities/assets.

The most important feature that distinguishes this study from the literature is the large sample size and the post-labeling of the data. In addition, most of the other studies on the prediction of financial failure do not offer any solution for the unbalanced data set problem. In this study, this problem is addressed and modeling is performed over different data sets. When the final model success is analyzed, the financial failure of the venture is predicted by 78%. For future studies, different machine learning models can be built by developing isolation forest and resampling methods.

## 1.GİRİŞ

Bir işletmenin amacı; ürettiği mal ve hizmetler aracılığıyla kâr elde etmek, toplum yararına hizmette bulunmak ve sosyal açıdan fayda sağlayabilmektir. Bu amaç doğrultusunda zarar etmeden uzun yıllar boyunca yaşam döngüsünü sağlamak ister. Ancak yaşanan finansal krizler nedeniyle bazı işletmeler için uzun yaşam döngüsü mümkün olmayabilir. İşletmeler faaliyetleri boyunca çeşitli problemlerle karşı karşıya kalabilir. İşletme yaşamının uzun vadede bu problemlerden etkilenmemesi veya asgari düzeyde etkilenmesi için kısa sürede problemler iyi tanımlanmalı ve stratejik planlar oluşturulmalıdır. İyi tanımlanmayan ve çözüme ulaştırılamayan problemler işletmeleri başarısızlığa götürebilir.

İşletme başarısızlığı, işletmenin yükümlü olduğu finansal sorumlulukları karşılayamamasıyla başlayan ve çözüm bulunmazsa iflasa kadar gidebilen uzun bir süreçtir. Yaşanan finansal başarısızlığın iflas ile sonuçlanması ise işletme ve işletme çalışanlarının yanı sıra bulunduğu sektörde de domino etkisine sebep olabilir. Bu etki ile işletme ile iş birliği yapan girişimler, işletme yatırımcıları ve kreditorler dolaylı yoldan etkilenir (Wilson ve Sharda, 1994). Burada altı çizilmesi gereken temel husus işletme başarısının tanımlanmasıdır. Başarısızlık, yönetsel başarısızlık, teknolojik başarısızlık, teknik başarısızlık, ekonomik başarısızlık ve finansal başarısızlık olarak detaylandırılabilir (Soba ve diğerleri, 2016).

Yönetsel başarısızlık, yöneticilerin yeterli eğitim düzeyine sahip olmamaları, tedbirsiz davranışlar ve doğru olmayan örgütlenme stratejilerinin sebep olduğu bir başarısızlık türüdür. Yöneticilerin yaşanan hatalardan gerekli dersi çıkarmamaları durumunda yönetsel başarısızlık ortaya çıkmaktadır (Türksoy, 2007).

Teknolojik başarısızlık, üretim yapan işletmeler için insan emeği ağırlıklı üretimden makine ağırlıklı üretime geçiş süreci ağır maliyetlere sebep olmaktadır. İşletmeler kendi teknolojilerini üretmiyorlarsa değişen ve gelişen teknoloji karşısında kullandıkları makine ve teknolojilerin çağın gerisinde kalması kaçınılmazdır. Küçük büyüklükteki işletmeler için ise Ar-Ge çalışmaları ağır bir yük haline gelir. İşletmenin bu yükün altına girememesi ve sık sık teknoloji güncelleyememesi rakipleri ile mücadele etme gücünü azaltarak başarısızlığa ilerlemesine neden olur (Karacan ve Savcı, 2011).

Teknik başarısızlık, en yalın haliyle işletmenin ödeme yükümlülüklerini zamanında yerine getirememesi, ödeme gücünü yitirmesi olarak ifade edilir (Gönenli, 1994; 647). İşletmenin ödemekle sorumlu olduğu borçları ödeyememesi, alacaklarını tahsil edememesi, sipariş üzerine üretilen malların satın alınmaması gibi sebeplerle gerçekleşebilir. Bu gibi durumlarda işletmenin öz kaynak toplamı borçlarının toplamından fazla olsa bile işletme teknik olarak başarısız sayılır.

Ekonomik başarısızlık, bir işletmenin sermaye maliyetlerinin yatırım karlılıklarının üstüne çıkması sonucunda karşılaşılan durumdur. Tam aksi durumda yani sermaye maliyetlerinin yatırım karlılıklarının altına düşmesi durumu ise ekonomik başarı olarak adlandırılır (Dağlı, 1994).

Finansal başarısızlık, işletmenin finansal yükümlülüklerini yerine getirememesi durumudur. Okka (2009: 928) yaptığı çalışmada finansal başarısızlığı belirlenen şirket politikalarındaki ve alınan finansal kararlardaki başarısızlıkların sonucu hedefe ulaşılabilen durumu olarak ifade etmiştir. Her işletme finansal zorluklarla karşılaşabilir önemli olan bu zorlukların erken aşamada tespit edilerek gerekli tedbir planlarının oluşturulmasıdır. Zamanında oluşturulmayan tedbir planı işletmeyi nihai sona getirerek iflas ile karşı karşıya kalmasına neden olabilir.

Finansal başarısızlık ve iflasın erken tespit edilebilmesi finans ekosistemini doğrudan ve dolaylı yollarla etkiler. Finansal başarısızlık yaşama ihtimali yüksek bir işletmenin erkenden tespiti ile işletme yatırımcıları, işletmenin içerisinde bulunduğu tedarik ağı ve işletme alacaklıları gibi yapılar bu başarısızlıktan en az hasar ile kurtulmuş olur. İşletmeler açısından finansal başarısızlığın erken tespit edilmesi ise girişimin kendini kurtarması ve yeni kaynaklar aramaya başlaması açısından önemlidir.

Finansal başarısızlığın erken dönemlerde tespit edilmesi verimlilik açısından da büyük öneme sahiptir. Girişimin finansal anlamda başarılı olması verimliliğin artmasını sağlar. Verimliliği yüksek olan girişimler diğerlerine göre daha fazla kâr elde ederler. Finansal analizin iyi yapılması kaynakların doğru yerlere yönlendirilmesini sağlar. Doğru kaynak yönetimi ile girişimler iş süreçlerini iyileştirme, üretimi artırma ve otomasyon sistemlerini kullanarak daha verimli hale gelebilirler.

Geçmişten günümüze finansal başarısızlık tahmini üzerinde pek çok çalışma yapılmıştır. Yapılan çalışmalarda muhasebe kalemlerinden oluşturulan finansal oranlar üzerinde sıklıkla durulmuştur. Bu oranlar; Likidite Oranları, Finansal Yapı Oranları, Devir Hızları ve Karlılık Oranları olarak dört ana başlık altında toplanır. Likidite oranları, işletmenin kısa vadeli borç yükümlülüklerini yerine getirip getiremeyeceğinin analiz edilebileceği bilanço kalemlerinden oluşur. Finansal yapı oranları, işletmenin varlıklarını finanse ederken kullanılan kaynakların dağılımı ile ilgili bilgi verir. Devir hızları, stok ve alacaklar

kullanılarak işletmenin etkinlik ve yoğunluğu ile ilgili bilgiler verir. Karlılık oranları ise işletmenin satışları ve yatırımları üzerinden elde edilen kârlılığı gösterir.

İflas ve finansal başarısızlık risklerinin tahmini için yapılan çalışmalarda yaygın olarak Ohlson O-skoru, Altman Z-skoru ve Taffer modeli kullanılmaktadır (Toprak Kesgin ve diğerleri, 2019). Bu modeller muhasebe tabanlı modellerdir ve işletmelerin son yayımlanan finansal raporlarını baz alırlar. Ancak, son finansal raporların geçmiş dönem faaliyetlerinden bağımsız olmadığına dikkate alınması gerekir. Geçmişe yönelik yayımlanan finansal raporları çalışma kapsamında göz ardı etmek tahmin sonucunun yanıltıcı olmasına sebep olabilir.

İflas tahminine yönelik Altman Z-skoru modeli Altman (1968) tarafından yayımlanmıştır. Altman Z-skoru modeli oluşturulurken 66 işletmenin bilanço kalemlerinde bulunan 5 finansal oran kullanılmıştır. Ohlson O-skoru modeli Ohlson (1980) tarafından Altman Z-skora alternatif bir model olarak geliştirilmiştir. Ohlson O-skoru modeli 2.000'den fazla işletmenin verileri kullanılarak oluşturulmuştur ve doğruluk oranı Altman Z-skorundan daha yüksektir. O-skorda bir işletmenin iflas olasılığını tahmin etmek için bilanço verilerinde yer alan 9 faktör kullanılır. Hesaplanan olasılığa göre son iki yılda net zarar varsa *iflas durumu=1* diğer durumlarda *iflas durumu=0* anlamına gelir (Toprak Kesgin ve diğerleri, 2019). Taffler (1983) ise 70 işletmeye ait finansal oranları kullanarak finansal başarısızlığa etki eden oranları ve finansal başarısızlığın tahmininde kullanılacak bir model oluşturmayı hedeflemiştir.

Bu çalışmada ise geçmiş yılların finansal raporları da kapsama alınmıştır. Altman Z-skoru'na ek olarak istatistik tabanlı İzolasyon Ormanı (Isolation Forest) Algoritması ve yeniden örnekleme yöntemleri ile makine öğrenmesi yöntemleri kullanılmıştır. Ayrıca çalışma kapsamında bir işletmenin 2 yıl üst üste bilanço vermemesi iflas olarak kabul edilmiştir.

Çalışmanın ikinci bölümünde literatür kapsamında yapılan çeşitli çalışmalar incelenmiştir. Üçüncü bölümde ise bu çalışma kapsamında kullanılan yöntemlerin teknik anlatımına yer verilmiştir. Son olarak dördüncü bölümde ise araştırma sonucu elde edilen bulgular belirtilerek yorumlanmış ve diğer çalışmalar ile karşılaştırılmıştır.

## 2.LİTERATÜR TARAMASI

Finansal başarısızlığın ve iflasın tahminine yönelik çok sayıda çalışma yapılmıştır. Bu alanda yapılan ilk çalışmalardan bir tanesi Beaver (1966)'ın yaptığı çalışmadır. Beaver (1966) çalışmasında finansal başarısızlığı beş yıl öncesine kadar tek değişkenli model olarak oran analiziyle tahmin edebildiğini belirtmiştir. Ancak tek değişkenli modellerin en büyük dezavantajı her bir oranı tek tek ele alarak tahmin sonucunda bulunmasıdır. Oranların tek tek ele alınması çelişkili tahmin sonuçlarına neden olabilir. Bu sorunu çözmek için Altman (1968) yaptığı çalışmada çok boyutlu ve çok değişkenli modeller için kullanılan diskriminant analizini kullanmıştır. Yaptığı bu çalışma sonucunda elde ettiği modelle işletmeleri finansal başarılı ve finansal başarısız olarak sınıflayarak ilk yıl için %95'lik bir tahmin başarıları göstermiştir.

Aktaş ve diğerleri (2003) ise işletmelerin finansal oranlarını kullanarak yaptıkları çalışmada yapay sinir ağlarının diskriminant analizi, logit modeli ve çoklu regresyon modelinden daha iyi tahmin gücünün olduğunu bildirmişlerdir.

Tang ve Chi (2005), finansal başarısızlığın tahminine yönelik yaptıkları çalışmada diskriminant analizi ve yapay sinir ağları yöntemlerinin performans karşılaştırmasını yapmışlardır. Finansal başarısızlık tanımını ise iflas etmiş olmak, düşük stok değerlendirmesi ve likidite durumu olarak belirtmişlerdir. Tahmin modelini geliştirmek için Shangay ve Shezen Borsası'nda 1995-2002 yıllarında işlem gören 128 başarılı ve başarısız işletmeyi ele almışlardır. Çalışma sonucunda ise yapay sinir ağlarının tahmin başarısının diskriminant analizine göre daha yüksek olduğu sonucuna ulaşmışlardır.

Nguyen (2005), yılında yaptığı çalışmada finansal başarısızlığı yapay sinir ağları ve lojistik regresyon modellerini kullanarak tahmin etmiştir. Çalışmasında diğer çalışmalara benzer şekilde finansal oran verilerini kullanmıştır. Çalışma sonucunda yapay sinir ağları modelinin diğer istatistiksel modellerden daha başarılı tahmin yaptığı sonucuna ulaşmıştır.

Akkoç (2007), çalışmasında finansal başarısızlığın tespiti için sinirsel bulanık ağ modeli, çok boyutlu ayırma analizi ve yapay sinir ağları modeli üzerinde çalışmıştır. Bu amaçla 1983 ve 2004 yılları arasında Sermaye Piyasası Kuruluna tabi ve/veya İstanbul Menkul Kıymetler Borsası'nda faaliyet gösteren 142 işletmenin finansal oranlarını kullanmıştır. Akkoç 142 verinin yarısının başarılı işletmeler diğer yarısının ise başarısız işletmelerden oluştuğunu belirtmiştir. Buna ek olarak veri setini eğitim, test ve kontrol olarak üç grupta incelemiştir. Uygulama sonucunda çok boyutlu ayırma analizinin kontrol grubu üzerindeki başarısını %83,3; sinirsel bulanık ağ modelinin kontrol grubu üzerindeki başarısını %80,56 ve yapay sinir ağları modelinin kontrol grubu üzerindeki başarısını %86,1 olarak bulmuştur.

Kurtaran Çelik (2010), Türk bankacılık sektöründe yer alan bankaların finansal başarısızlıklarını tahmin etmek amacıyla diskriminat analizi ve yapay sinir ağları üzerinde bir çalışma gerçekleştirmiştir. Çalışmada bankaların finansal başarısızlığından önceki iki yıllık verileri kullanmıştır. Finansal başarısızlıktan bir yıl öncesi için yapılan diskriminant analizinde; takipteki kredilerin toplam kredilere oranı, faiz gelirlerinin ortalama getirili aktiflere oranı, faiz giderlerinin toplam giderlere oranı ve faaliyet giderinin toplam aktiflere oranı, finansal başarısızlıktan iki yıl öncesi için yapılan diskriminant analizinde ise; takipteki kredilerin toplam kredilere oranı, faiz gelirlerinin ortalama getirili aktiflere oranı ve duran aktiflerin toplam aktiflere oranı kullanılmıştır. Çalışma sonucunda finansal başarısızlığın bir yıl öncesi için kurulan yapay sinir ağları modelinin başarılı ve zarar eden bankaları %100 doğru tahmin ettiği sonucuna ulaşılmıştır. Benzer şekilde finansal başarısızlığın iki yıl öncesinde kurulan modellerden başarılı bankaları %89 tahmin gücüyle diskriminant analizi, başarısız bankaları ise %100 başarı ile yapay sinir ağlarının daha iyi tahmin ettiği ifade edilmiştir.

Kılıç ve Seyrek (2012), çalışmalarında 2005-2010 yılları arasında İstanbul Menkul Kıymetler Borsası'nda işlem gören 137 imalat sanayisi işletmesinin finansal oranlarını kullanarak Yapay Sinir Ağları ve Karar Ağaçları aracılığıyla finansal başarısızlıklarını tahmin etmeye yönelik çalışma gerçekleştirmişlerdir. Çalışma sonucunda finansal başarısızlığı Yapay Sinir Ağları modelinin %94, Karar Ağaçlarının ise %82 oranında doğru tahmin ettiği sonucuna ulaşılmıştır.

Öztürkler (2012)'in yaptığı çalışmaya göre finansal başarısızlık istatistikî model olan lojistik regresyon kullanılarak tahmin edilmiştir. Çalışmada 2004, 2005 ve 2006 yıllarında faaliyet gösteren 1.764 firmanın finansal oranları kullanılarak iki farklı lojistik regresyon modeli kurulmuştur. Kurulan iki logit tahmin sonucuna göre, finansman/net satışlar oranı düşük, net satış kârı/net satışlar oranı yüksek, faaliyet giderleri/net satışlara oranı düşük, asit-test oranı yüksek, dönen varlıklar toplamı/aktif oranı yüksek ve düşük kaldıraçla çalışıp borçlanmada uzun vadeyi tercih eden işletmelerin finansal olarak başarılı olma olasılıklarının daha yüksek olabileceğini söylemiştir.

Ertan ve Ersan (2018), çalışmalarında imalat sektöründeki finansal başarısızlığı tahmin etmek için 2000 ve 2014 yıllarında Borsa İstanbul'da işlem gören 208 işletme verisini kullanmışlardır. Finansal başarısızlık tanımını ise işletmenin ulusal pazardan Yakın İzleme Pazarı'na geçiş yapması olarak belirlemişlerdir. Tahmin yöntemi olarak log-lojistik, panel probit, Cox orantılı riskler, logit, tamamlayıcı log-log ve rassal etkiler modellerini ele almışlardır. Çalışmada Cox orantılı riskler yöntemi performansı en iyi model olarak belirlenmiştir.

Aksoy ve Boztosun (2019), Borsa İstanbul'da 2010-2012 yılları arasında kaydı bulunan imalat sanayi sektöründeki 86 işletme üzerinde yaptıkları çalışmada işletmelerin finansal başarısızlıklarını makine öğrenmesi sınıflama yöntemleri ile tahmin etmiştir. Çalışmalarında işletmelerin finansal başarısızlıklarını tahmin etmek için makine öğrenmesi sınıflama yöntemlerini kullanmışlardır. Sınıflama algoritması olarak Yapay Sinir Ağları, CART (Classification and Regression Tree), SVM (Support Vector Machine) ve K-En Yakın Komşu kullanılan çalışmanın model eğitiminde 14 öznitelik kullanmışlardır. Araştırmacılar bu öznitelikleri hem finansal oranlardan hem de finansal oranlardan bağımsız özel durum göstergelerinden elde ettiklerini belirtmişlerdir. Çalışma sonucunda işletmelerin finansal başarısızlıklarını hem 1 yıl hem de 2 yıl öncesinde en iyi tahmin eden algoritmanın Yapay Sinir Ağları modeli olduğunu bildirmişlerdir.

Kuzu ve Yakut (2020) yaptıkları çalışmada SVM ve Karar Ağaçları algoritmalarını kullanarak imalat sanayisinde faaliyet gösteren farklı teknolojik düzeylere sahip işletmelerin finansal başarısızlıklarını tahmin etmişlerdir. Uygulamalarında 2012-2015 yılları arasında Borsa İstanbul'a kayıtlı 153 işletmenin 11 mali oranını ve teknolojik düzeylerini tanımlayan kategorik değişkeni kullanmışlardır. İşletmelerin finansal başarısızlığı kavramını 3 yıl üst üste zarar etme durumu ve 2 yıl üst üste zarar etme durumu olarak iki farklı şekilde ele almışlardır. Her iki başarısızlık tanımı için de düşük teknoloji sınıfında olan işletmeler için SVM algoritmasının doğru sınıflama oranını %93, Karar Ağaçlarının doğru sınıflama oranını sırasıyla %90 ve %86 olarak elde etmişlerdir. Orta-düşük teknoloji düzeyinde yaptıkları çalışmada ise SVM algoritmasının başarısını sırasıyla %88 ve %85, Karar Ağaçlarının başarısını %88 ve %90 olarak bulmuşlardır. Son olarak orta-yüksek teknoloji düzeyine sahip olan işletmelerin 3 yıl üst üste zarar etme ve 2 yıl üst üste zarar etme durumları hedef değişken olarak belirlendiğinde sırasıyla SVM algoritmasında %94 ve %90, Karar Ağaçlarında %94 ve %87 doğru sınıflama başarısını elde etmişlerdir.

Bu çalışmada ise borsaya kote olan ve olmayan tüm girişimler ele alınmıştır. Literatür çalışmaları incelendiğinde ele alınan girişimlerin finansal başarısızlık durumu ile ilgili önsel bir bilgi bulunmaması çalışmayı literatür çalışmalarından ayıran en önemli unsur olarak belirtilebilir. Bununla birlikte makine öğrenmesi yöntemlerinin başarılı ve tutarlı sonuçlar vermesi için veri seti boyutunun fazla olması gerektiği bilinmektedir. Literatür çalışmalarının neredeyse hepsinde kullanılan veri setlerinin boyutu minimum düzeydedir. Yapılan çalışmada ise veri seti boyutu oldukça yeterlidir. Çalışmayı literatürden ayrı tutan bir diğer konu ise verilerin etiketlenmesidir. Literatür çalışmalarında finansal başarılı-finansal başarısız

şeklindeki hazır etiketli veriler kullanılmıştır. Bu çalışma kapsamında ise İzolasyon Ormanı Yöntemi kullanarak veri seti etiketlenerek yeniden örnekleme yöntemine tabii tutulmuş ve sonrasında modellenmiştir.

### 3.YÖNTEM

Bu çalışmanın amacı imalat sanayisi NACE Rev.2 28-Başka Yerde Sınıflandırılmamış Makine ve Ekipman İmalatı sektöründe faaliyet gösteren orta-yüksek teknoloji düzeyine sahip işletmelerin finansal başarısızlığının makine öğrenmesi yöntemleri aracılığıyla tahmin edilebilmesidir. Bu amaçla 2010-2018 yılları arasında NACE Rev.2 28 sektöründe faaliyet gösteren girişimlere ait likidite oranları, borçluluk oranları, devir hızı oranları ve karlılık oranları ana başlığı altında 27 finansal oran kullanılmıştır. Modelleme aşamasında 2010-2016 yıllarını kapsayan 56.443 işletme verisi eğitim 2017 yılına ait 14.112 işletme verisi test ve 2018 yılına ait 14.768 işletme verisi doğrulama verisi olarak belirlenmiştir. 2010-2016 yılları arasındaki tüm verilerin alınması bir girişim sıra numarasının yıllar bazında tekrar etmesine neden olmuştur. Bu işletmeler sıra numarasından arındırılıp tekilleştirilerek çalışmaya dahil edilmiştir. Veri setinde tahmin edilmesi beklenen finansal başarısızlık/iflasa dair bilgi bulunmadığı için izolasyon ormanı yöntemi kullanılarak veri setinin karakteristik özelliğini yansıtmayan veriler tespit edilmiş ve finansal başarısızlık yaşayan işletme olarak etiketlenmiştir. Etiketleme işlemi sonucunda oluşturulan veri etiketlerinin dağılımı incelendiğinde finansal başarısızlık yaşayan ve finansal başarısızlık yaşamayan işletme sınıflarının eşit dağılmadığı belirlenmiştir. Tahmin edilmesi beklenen sınıfların dağılımının eşit olmamasının makine öğrenmesi modellerinde çoğunlukta olan sınıfı aşırı öğrenme eğiliminde olduğundan yeniden örnekleme yöntemleri kullanılmıştır. Veri setine yeniden örnekleme yöntemlerinden olan rasgele düşük örnekleme, rasgele aşırı örnekleme ve SMOTE yöntemleri uygulanarak üç farklı yeni veri seti elde edilmiştir. Elde edilen her veri seti için XGBoost modeli kurularak yeniden örnekleme yöntemlerinin performansları karşılaştırılmıştır.

Makine öğrenmesi eldeki sınırlı sayıda veriden yola çıkarak eğitim verisiyle algoritmalar oluşturulup bir tahmin fonksiyonu belirlenmesi işlemidir. Öğrenme işlemi veri setine göre denetimli öğrenme ve denetimsiz öğrenme olarak ikiye ayrılır. Denetimli öğrenme yöntemlerinde girdi verilerine karşılık gelen çıktı verileri bulunur. Bu durum etiketlenmiş veri olarak adlandırılır. Denetimli öğrenmede algoritmanın öğrenmesi istenilen çıktılar eğitim setinde makineye verilerek eğitim gerçekleştirilir. Denetimsiz öğrenme yönteminde ise eğitim setinde öğrenilebilecek herhangi bir etiketli veri bulunmaz. Bu yöntemde veri setinin iyice analiz edilerek yapısal bir örüntü elde edilmesi amaçlanır.

Çalışmada kullanılan tüm veriler Sanayi ve Teknoloji Bakanlığı Stratejik Araştırmalar ve Verimlilik Genel Müdürlüğü Girişimci Bilgi Sistemi (GBS)'nden temin edilmiştir. GBS, farklı kamu kurum ve kuruluşlarının idari kayıtlarında yer alan işletmelerin ekonomik faaliyetlerine ilişkin verilerinin belirli standartlar çerçevesinde toplandığı bir veri entegrasyonu projesidir. GBS'nin temel amacı, tüm kamu kurum ve kuruluşları başta olmak üzere üniversiteler, uzmanlar, araştırmacılar, araştırma enstitüleri ve sivil toplum kuruluşlarındaki yetkililer için iktisadi, sektörel ve bölgesel politikaların tasarlanıp uygulanması ve etkinlik ölçümü için kullanılması amacıyla doğru ve güvenilir verilerin sunulacağı bir ortam sağlamaktır.

#### 3.1 İzolasyon Ormanı

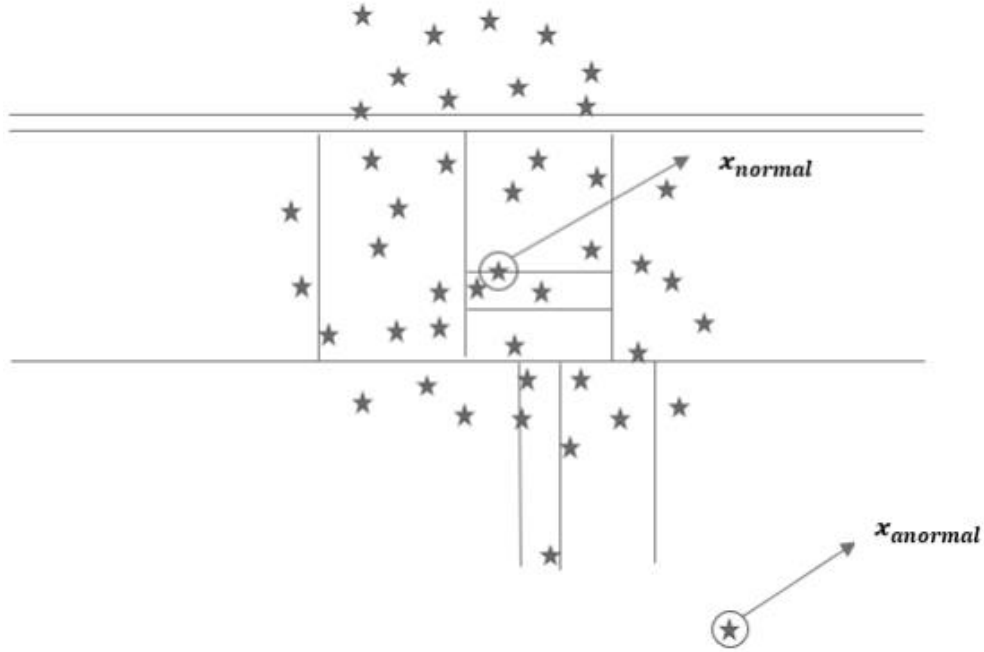
İzolasyon Ormanı, teoride karar ağacı algoritmasına dayanır ve mesafe tabanlı algılamaya sahiptir. Diğer anomali tespit yöntemlerinde algoritmalar normal veri profilini çıkartarak bu profilden uzak verileri "anomali" olarak etiketler. Bu algoritmaların temel amacı normal verilerin karakteristik yapısını tespit etmektir. Ancak İzolasyon Ormanı Yönteminde ise temel amaç anomali durumundaki noktaları tespit etmektir (Liu, 2008). Bu yöntem anomali noktalarının farklı olması ve veri setinde az bulunması üzerine kurulmuştur.

İzolasyon Ormanı Algoritmasında veri seti içerisinde rasgele seçilen bir özneliğin minimum ve maksimum değerleri arasından rasgele bir değer seçilir ve bu rasgele seçilen değer üzerinde veya altında olmasına göre dallara ayrılır. Anomali gösteren gözlem birimleri veri setinden en yakın dallara ayrılarak diğer birimlerden izole olurlar (Şekil 1).

Bu yöntemde gözlem birimleri yalnızca izole edilmekle kalmaz bunun yanında bir de gözlemlere ait anomali skoru elde edilir. Bu skor Eşitlik 1'deki gibi elde edilir.

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

Bir gözlem birimi için anomali skoru  $s(x, n)$  hesaplandıktan sonra elde edilen skor; 1'e yakınsa anormal, 0.5'ten küçükse normal olarak etiketlenir. Denklemde;  $h(x)$ ,  $x$  düğümüne olan uzaklık,  $E(h(x))$  izolasyon ağaçları üzerindeki düğümlerin  $h(x)$  değerleri ortalaması,  $c(n)$  ise  $x$  düğümüne ulaşmaya çalışırken başarısız aramaların ortalama yol uzunluğu olarak açıklanır.



Şekil 1. İzolasyon ormanı yöntemi

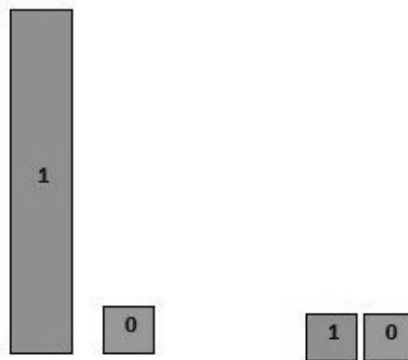
### 3.2 Yeniden Örnekleme Yöntemleri

Sınıflandırma problemleri makine öğrenmesinde denetimli öğrenme yöntemleri alt başlığında incelenir. Bu problemlerde çıktı değişkeni bir başka deyişle hedef vektör değerleri 0, 1, 2 gibi sınıflardan oluşur. Dengesizlik ise bir sınıfa ait örneklem büyüklüğünün diğer sınıf ya da sınıflardan yüksek olması durumudur. Bu durum fazla olan sınıf karakterinin aşırı öğrenilmesine, az olan sınıfın yeterince öğrenilememesine sebep olacağı için model performansını olumsuz etkiler. Bu sorunun çözümü için literatürde farklı teknikler geliştirilmiştir.

Eşit dağılmayan dengesiz veri setlerinde sınıfların dağılımlarını dengelemek için veri üzerinde çeşitli ön işleme adımları gerçekleştirilir. Sınıfların dağılımlarını eşitlerken yeniden örnekleme (resampling) yöntemleri kullanılmaktadır. Yeniden örnekleme, veri setini istatistiksel ya da istatistiksel olmayan tahmin yöntemleriyle yeniden yapılandırma işlemidir. Yeniden örnekleme teknikleri, Rasgele Düşük Örnekleme (random undersampling), Rasgele Aşırı Örnekleme (random oversampling) ve SMOTE olarak üç grupta incelenebilir (Galar, 2012).

#### 3.2.1. Rasgele Düşük Örnekleme (Random Undersampling)

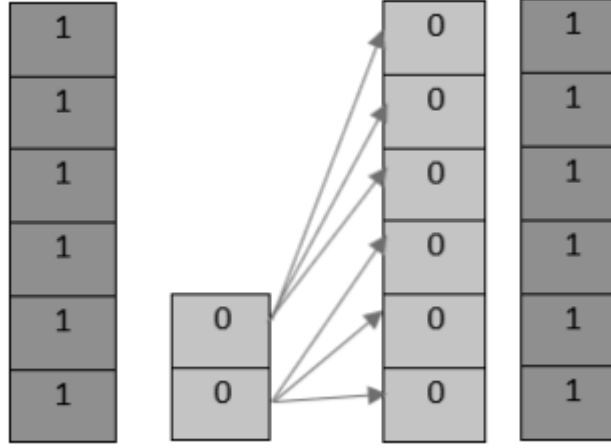
Bu yöntemde az olan sınıfın örneklemini korunur ve çok olan sınıf gözlemleri arasından rasgele seçimler yapılarak az olan sınıf sayısı kadar yeni örneklem oluşturulur (Han ve Kamber, 2006:384-385). Böylelikle az ve çok olan sınıflar eşitlenerek Şekil 2'deki gibi veri seti dengeli hale getirilir.



Şekil 2. Rasgele düşük örnekleme

### 3.2.2. Rasgele Aşırı Örnekleme (Random Oversampling)

Rasgele aşırı örnekleme, azınlık olarak bulunan sınıfın verilerinden örneklem çekerek veri setine eklemeye yöntemidir. (He ve Garcia, 2009). Bu yöntemde az olan sınıf birimleri rasgele seçilerek örnekleme eklendiği için aşırı öğrenme/aşırı uyum problemiyle karşılaşılabilir (Fernández ve diğerleri, 2018). Bu yöntemin oluşturacağı bir diğer problem ise birim sayısı arttığı için sınıflama yaparken sürenin artmasıdır. Şekil 3'te rasgele aşırı örnekleme yöntemi gözlenebilir.



Şekil 3. Rasgele aşırı örnekleme

### 3.2.3. Sentetik Azınlık Aşırı Örnekleme (SMOTE)

Dengesiz sınıf problemlerine çözüm olarak geliştirilen yöntemlerden biridir ve Chawla ve diğerleri (2002) tarafından geliştirilmiştir. Aşırı örnekleme yönteminde, gözlem birimlerinin bir kopyası oluşturulması sebebiyle meydana gelen aşırı öğrenme problemine karşı önerilir (Ay ve Yolaçan, 2022). SMOTE tekniği dengesiz veri seti probleminde kullanılan yaygın bir tekniktir (Patil ve diğerleri, 2020).

Bu yöntemde öncelikle azınlıkta bulunan sınıfın en yakın  $k$  komşusu belirlenir. Her bir yakın komşu ile seçilen gözlem arasındaki farklılıklar ölçülür ve elde edilen değer  $(0,1)$  arasından rasgele bir sayı seçilerek çarpılır. Son olarak, son elde edilen değer ilk olarak seçilen gözlem değerine eklenir. Böylelikle azınlık olan sınıftan sentetik olarak veri üretilmiş olur (Şekil 5). Yukarıda belirtilen adımlar Eşitlik 2'de formülize edilmiştir.

$$x_{yeni} = x_i + (x_j - x_i) * \alpha \quad (2)$$

### 3.3. XGBOOST Algoritması

XGBOOST (Extreme Gradient Boosting) Algoritması, makine öğrenmesinde sıklıkla kullanılan bir uygulamadır. XGBoost Algoritması, Gradient Boosting Machine (GBM) Algoritmasının optimize edilmesiyle ortaya çıkmıştır. XGBoost'un GBM'e göre en büyük avantajı boş gözlemleri işleyebilmesi, aşırı öğrenme durumuna karşın daha sağlam olması ve en önemlisi yüksek tahmin gücüne sahip olmasıdır. Algoritma ilk olarak Chen ve Guestrin (2016) tarafından geliştirilmiştir. Chen ve Guestrin çalışmalarında XGBoost algoritmasının diğer algoritmalara nazaran 10 kat daha hızlı çalıştığını söylemişlerdir (Yeşilyurt ve Dalkılıç, 2021).

Regresyon ve sınıflandırma problemlerinde boosting (arttırma) algoritmalarının daha yüksek performans sergilediği görülmektedir. XGBoost Algoritması diğer makine öğrenmesi yöntemlerine göre hesaplama kolaylığı sağlamaktadır. Bu hesaplama kolaylığı çok boyutlu verilerin analizinde kullanılmaktadır (Abar, 2020).

XGBoost Algoritması, güçlü ve sağlam bir sınıflandırıcı oluşturabilmek için çok fazla sayıda zayıf sınıflandırıcıyı boosting ile birleştirip yinelemeli olarak eğitimi gerçekleştirerek güçlü öğreniciyi elde etmeyi amaçlar. XGBoost algoritmasında eğitim seti için en iyi parametreyi bulmak ve model performansını ölçmek için aşağıdaki Eşitlik 3'te yer alan amaç fonksiyonu kullanılır.

$$Obj(\theta) = L(W) + \Omega(\theta) \quad (3)$$

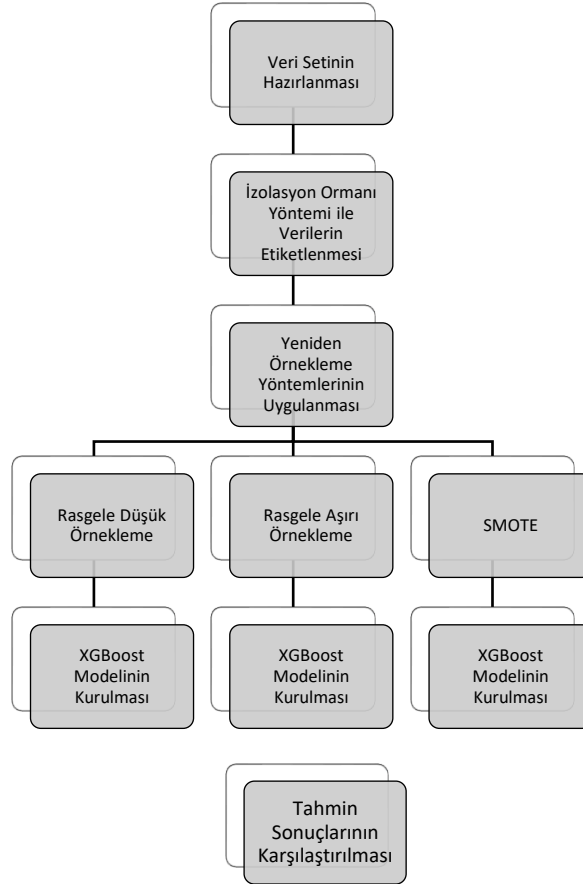
Eşitlikte  $L$ , eğitim kayıp fonksiyonudur ve eğitim verisini kullanarak modelin öngörüsünü ölçer (Dixit, 2017). Kayıp fonksiyonunun tanımlanarak en aza indirilmesi istenir.  $\Omega$ , düzeltme (regularization) terimini



ifade eder. Düzeltme, model karmaşıklığını kontrol ederek aşırı öğrenmeyi azaltmayı amaçlar (Yangın, 2019).

#### 4. BULGULAR

GBS'de bulunan ve NACE Rev.2 28 sektöründe faaliyet gösterdiği bilinen işletmelere ait bilanço verilerinden 27 finansal oran hesaplanmıştır. Daha sonra izolasyon ormanı yöntemi aracılığıyla veri setinde çoğunluk sınıfın karakteristik özelliklerine benzemeyen gözlem birimleri bulunarak etiketlenmiştir. Böylelikle yeniden örneklemeye yöntemlerinin uygulanacağı veri seti elde edilmiştir. Şekil 4'te tüm modelleme adımları gösterilmiştir. Çalışmada kullanılan 27 finansal oran ise Tablo 1'de verilmiştir.



Şekil 4. Modelleme adımları

Gelir tablosu ve bilanço verileri kullanılarak çok sayıda oran elde edilebilmektedir. Burada esas üzerinde durulması gereken konu problemi tanımlayacak oranları belirleyebilmektir. Yukarıdaki 24 oran 4 ana başlık altında toplanabilir. Bu başlıklar aşağıdaki gibi tanımlanmıştır.

- A) Likidite Oranları
- B) Finansal Yapı Oranları
- C) Devir Hızı Oranları
- D) Kârlılık Oranları

Likidite oranları işletmenin cari durumunun analizinde sıklıkla kullanılır. Likidite oranları kullanılarak işletmenin kısa vadeli borçlarını ödeyebilme gücü ve çalışma sermayesinin yeterli olup olmadığı belirlenebilir. Likidite oranları alt başlığında; Cari Oran, Asit-Test Oranı, Nakit Oranı, Stoklar/Dönen Varlıklar ve Stoklar/Aktif Toplam oranlarından oluşur.

**Tablo 1. Finansal oranlar**

<i>Finansal Oranlar</i>	<i>Finansal Oranlar</i>
<b>LİKİDİTE</b>	<b>DEVİR HIZI</b>
1. Cari Oran	Net Çalışma Serbest Devir Hızı
2 Asit Test Oranı	Aktif Devir Hızı
Nakit Oran	Net Kâr/ Özkaynak
Stoklar/Dönen Varlıklar	Net Kâr/ Aktif
Stoklar/Aktif	Faaliyet Kârı/Net Satışlar
<b>FİNANSAL YAPI</b>	<b>KARLILIK</b>
Kaldıraç Oranı	Brüt Satış Kârı/ Net Satışlar
Özkaynak/Aktif	Net Kâr/ Net Satışlar
Özkaynak/Yabancı Kaynak	Satışların Maliyeti/Net Satışlar
Kısa Vadeli Yabancı Kaynak/Aktif	Faaliyet Giderleri/Net Satışlar
Uzun Vadeli Yabancı Kaynak/Aktif	İşletme Sermayesi/Aktif
Uzun Vadeli Yabancı Kaynak/Devamlı Sermaye	
	<b>ALTMAN</b>
Stok Devir Hızı	Favök/Aktif Toplam
Alacak Devir Hızı	Net Satışlar/ Aktif Toplam
Çalışma Sermayesi Devir Hızı	Birikmiş Kârlar/Aktif Toplam

Finansal yapı oranları ise işletmenin borçluluk yapısı ve bu borçluluğun sürdürülebilirliğinin belirlenmesi için kullanılırlar. Borçluluk oranları alt başlığında; Kaldıraç Oranı, Özkaynak/Aktif Toplam, Özkaynak/Yabancı Kaynak, Kısa Vadeli Yabancı Kaynak/Aktif Toplam, Uzun Vadeli Yabancı Kaynak/Aktif Toplam ve Uzun Vadeli Yabancı Kaynak/Devamlı Sermaye oranları bulunur.

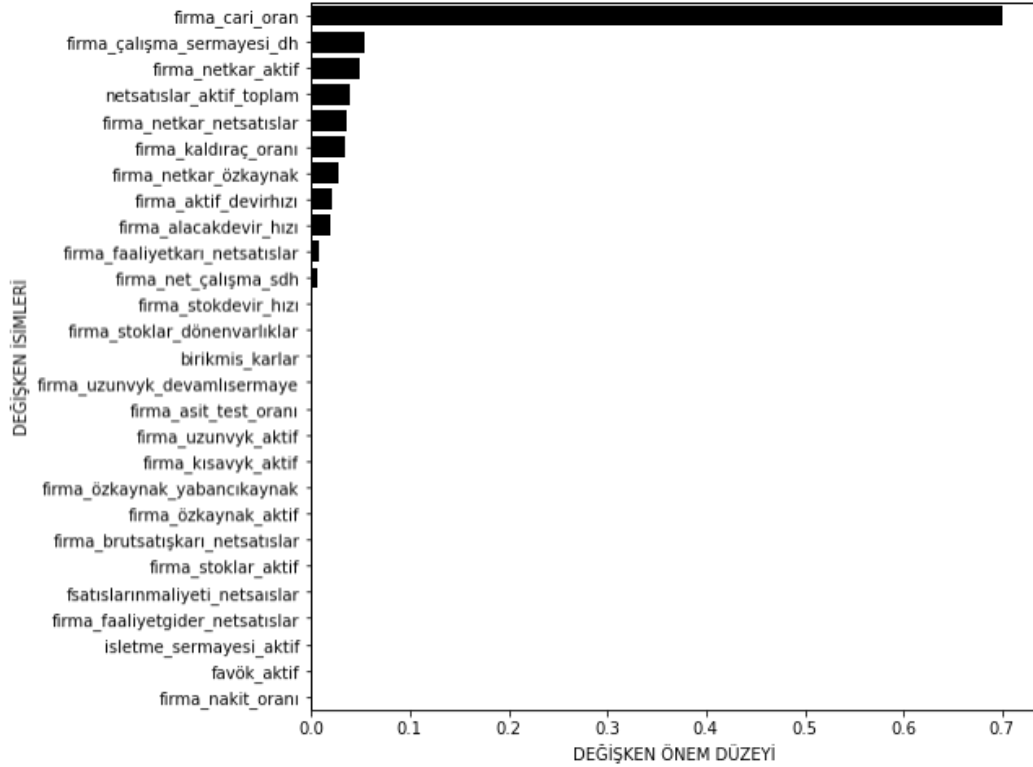
Devir hızı oranları kullanılarak işletme faaliyetlerinin etkinliği ve verimliliği ile ilgili bilgiler elde edilebilmektedir. Bu başlık altında temel olarak 5 farklı oran bulunur. Bu oranlar; Stok Devir Hızı Oranı, Alacak Devir Hızı Oranı, Çalışma Sermaye Devir Hızı, Net Çalışma Sermayesi Devir Hızı ve Aktif Devir Hızı Oranı olarak bilinmektedir.

Kârlılık oranları ise işletmenin faaliyetlerinde kârlı çalışılıp çalışılmadığı, özkaynaklarını ve yabancı kaynaklarını etkin kullanıp kullanmadığı hakkında bilgiler verir. En sık kullanılan kârlılık oranları Net Kâr/ Özkaynak, Net Kâr/Aktif Toplam, Faaliyet Kârı/Net Satışlar, Brüt Satış Kârı/Net Satışlar, Net Kâr/Net Satışlar, Satışların Maliyeti/Net Satışlar ve Faaliyet Giderleri/Net Satışlar olarak bilinmektedir.

Son olarak Altman'ın finansal başarısızlık tahmini için geliştirdiği Altman'ın Z Skoru yönteminde kullanılan Net İşletme Sermayesi/ Aktif Toplam, Birikmiş Karlar/ Aktif Toplam ve FAVÖK/ Aktif Toplam oranları modele dahil edilmiştir.

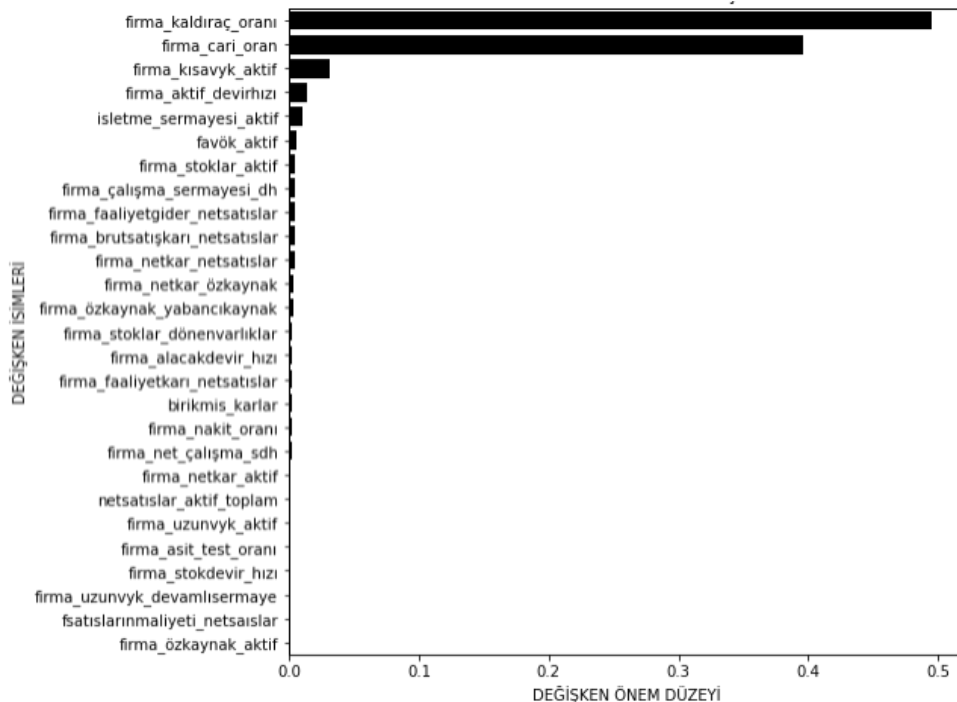
İlk olarak izolasyon oranı uygulamasıyla elde edilen veri setine rasgele düşük örnekleme yöntemi uygulanarak azınlıkta bulunan sınıfın örnekleme eşitlenmiştir. Daha sonra hazırlanan veri setine XGBoost modeli uygulanmıştır. Kurulan model kullanılarak 2017 yılına ait verilerden tahmin yapılmış ve 2018 yılında finansal başarısızlık yaşaması muhtemel işletmeler tahmin edilmiştir. Rasgele düşük örnekleme veri seti üzerine kurulan XGBoost modelinin tahminleri 2018 yılına ait veri seti ile doğrulandığında modelin %52 doğru sınıflama oranına sahip olduğu görülmüştür. Bu sonuca göre 2017'de finansal başarısızlık yaşayacağı tahmin edilen işletmelerin %52'sinin 2018'de gerçekten finansal başarısızlık yaşadığı belirlenmiştir.

Şekil 5, kurulan modelin belirlediği hedef değışkendeki değışimi en iyi açıklayan özelliklerin belirtildiği listedir. Grafik incelendiğinde kurulan modelin işletmelerin finansal başarısızlığının tahmin edilmesinde ayırt edici olarak belirlediği ilk üç değışken cari oran, çalışma sermayesi devir hızı ve net kâr/aktif oranlarıdır.



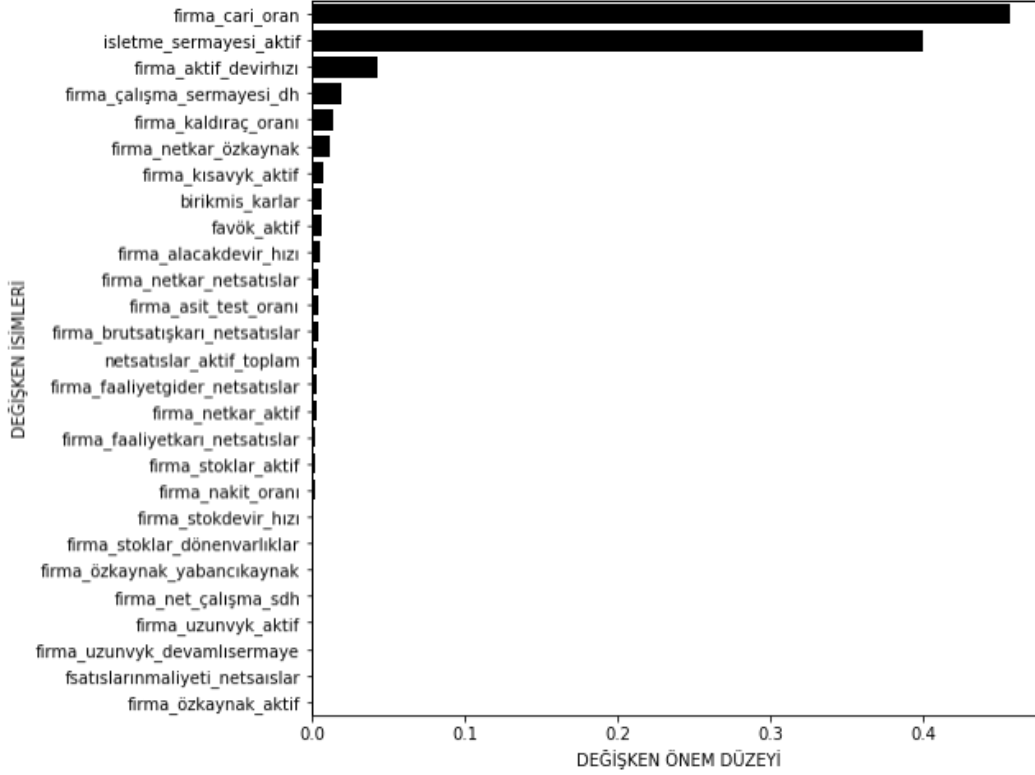
Şekil 5. rasgele düşük örnekleme verisi ile kurulan modelin önemli değişkenleri

İkinci olarak izolasyon ormanı uygulamasından elde edilen veri setine SMOTE yöntemi uygulanarak azınlıkta bulunan sınıf verilerinden sentetik veriler üretilmiş ve iki sınıfın örnekleme eşitlenmiştir. Hazırlanan veri seti üzerine XGBoost modeli kurularak 2017 yılına ait verilerden 2018 yılına ilişkin finansal başarısızlık tahmini yapılmıştır. Elde edilen tahmin sonuçları 2018 yılı verileri ile doğrulandığında finansal başarısızlık yaşayacağı tahmin edilen işletmelerin %78'inin doğru tahmin edildiği sonucuna ulaşılmıştır. Şekil 6'daki grafik incelendiğinde yeniden örnekleme yöntemlerinden SMOTE algoritması kullanılarak oluşturulan veri seti modellendiğinde işletmelerin finansal başarısızlığın tahmin edilmesinde ayırıcı değişkenlerin ilk üçünün kaldıraç oranı, cari oran ve kısa vadeli yabancı kaynak/aktifler olduğu gözlenmektedir.



Şekil 6. SMOTE verisi ile kurulan modelin önemli değişkenleri

Son olarak izolasyon ormanı uygulamasından elde edilen veri setine azınlık sınıfın örnekleminin eşitlenmesi için rasgele aşırı örnekleme yöntemi uygulanarak XGBoost modeli kurulmuştur. Kurulan model üzerinden diğer modellere benzer şekilde 2018 yılına ilişkin tahmin gerçekleştirilmiştir. Yapılan tahmin 2018 yılı verileri ile doğrulandığında finansal başarısızlık yaşayacağı tahmin edilen işletmelerin %75'inin doğru sınıflandırıldığı gözlemlenmiştir. Şekil 7'deki grafik incelendiğinde ise finansal başarısızlığın tahmini için kurulan modelin ayırt edici olarak belirlediği ilk üç değişken sırasıyla cari oran, işletme sermayesi/aktifler ve aktif devir hızı olarak belirlenmiştir.



Şekil 7. Rasgele aşırı örnekleme verisi ile kurulan modelin önemli değişkenleri

## 5. SONUÇ ve TARTIŞMA

Yapılan çalışmada, işletmelerin son yıl bilançosuna odaklanan literatürdeki yöntemlere ek olarak işletmelerin geçmiş yıllarına ait bilançoları da kullanılmıştır. Buna ek olarak, çalışmada Altman'ın Z-skoru hesaplamasında kullanılan değişkenler ve Z-skor sonuçları da veri setine eklenmiştir. Hazırlanan veri setine izolasyon ormanı yöntemi uygulanarak veri setinin karakteristik özelliği dışında davranan gözlem birimleri tespit edilerek veri seti finansal başarısızlık yaşayan ve finansal olarak başarılı olacak şekilde etiketlenmiştir. Etiketlenmiş ve modellenmeye hazır veri seti sınıflarının dağılımı incelendiğinde iki sınıfın dengesiz dağıldığı gözlemlenmiştir. Makine öğrenmesi modelleri dengesiz veri setleri ile çalışırken çoğunlukta olan sınıfı aşırı öğrenme eğiliminde olduğu için yeniden örnekleme yöntemleri kullanılarak düzenlemeye gidilmiştir. Daha sonra her veri setine denetimli sınıflama yöntemlerinden olan XGBoost modeli uygulanarak tahmin etme işlemi gerçekleştirilmiştir.

Çalışma sonucunda tahmin gücü en yüksek yeniden örnekleme yönteminin %78 ile SMOTE yöntemi olduğu belirlenmiştir. SMOTE yöntemi ile kurulan XGBoost modeline göre belirlenen en önemli üç değişken cari oran, kaldıraç oranı ve kısa vadeli yabancı kaynak/aktif olarak belirlenmiştir (Şekil 6). Tahmin gücü en yüksek ikinci yeniden örnekleme yöntemi ise %75 ile rasgele aşırı örnekleme modeli olarak belirlenmiştir. Rasgele aşırı örnekleme yönteminin uygulandığı veri seti üzerine kurulan XGBoost modelinin en önemli üç değişkeni ise cari oran, işletme sermayesi/aktif ve aktif devir hızı olarak belirlenmiştir (Şekil 7). Son olarak %52 ile tahmin gücü en düşük olan yeniden örnekleme yöntemi rasgele düşük örnekleme olarak bulunmuştur. Rasgele düşük örnekleme uygulanan veri seti ile kurulan XGBoost modelinin en önemli üç değişkeni ise cari oran, çalışma sermayesi devir hızı ve net kar/aktif olarak belirlenmiştir (Şekil 5). Her üç yöntem kullanılarak kurulan modellerin önemli değişkenleri incelendiğinde hepsinde de en önemli üç değişken arasında cari oran değişkeninin olduğu gözlemlenmiştir. Buradan yola çıkarak, makine ve ekipman imalatı sektörünün finansal başarısızlığının modellenmesinde en önemli finansal oranın likidite oranları alt başlığında bulunan cari oran kalemi olduğu söylenebilir. Cari oranın 1'den düşük olması dönen varlıklar

toplamının vadesi 1 yıldan kısa olan borçları karşılamadığı anlamına gelir. 2'nin üzerinde olması ise işletmenin nakit varlıklarını âtil kullandığı anlamına gelir.

Çalışmada, literatürdeki diğer çalışmalardan farklı olarak işletmelere ait 8 yıllık bilanço verisi kullanılmış ve işletmeler tekilleştirilerek çalışmaya dahil edilmiştir. Buna ek olarak, veri seti üzerinde anomali tespiti yapılmış ve veriler etiketlenmiştir. Anomali tespitini güçlendirmek için finansal başarısızlığın analizinde başarısı kanıtlanmış Altman Z-skoru hesaplanarak veri setine dahil edilmiştir. Literatürdeki diğer çalışmalar incelendiğinde yapılan tüm çalışmaların finansal başarısızlık yaşadığı ve finansal olarak başarılı olduğu kesinleşmiş işletmeler üzerinden modelleme yapıldığı görülmüştür. Bu çalışmada ise amaç finansal başarı durumu belli olmayan işletmelerin farklı bir yöntem kullanılarak tespit edilebileceği fikrini ortaya koymak ve daha sonraki çalışmalarda bu fikri geliştirebilmektir. Ayrıca, incelenen literatür çalışmalarında finansal olarak başarılı ve finansal olarak başarısız işletmelerden oluşan örneklem sayısının düşük olduğu görülmektedir. Örneklem sayısına bir bütün olarak bakıldığında yeterli gibi görünse de makine öğrenmesi modelinin eğitilmesi için eğitim ve test seti olarak bölündüğünde yetersiz kalabileceği öngörülmektedir.

Yapılan çalışmada ise eğitim seti olarak 2010-2016 yılları arasında faaliyet gösteren finansal olarak başarılı ve başarısız olduğu belirli olmayan 56.443 işletme verisi üzerinde çalışılmıştır. Bu belirsizliği gidermek için ise anomali tespit yöntemlerine odaklanılmıştır. İstatistiksel açıdan anomali değer kavramı incelendiğinde ortalama değerden  $\pm 3$  standart sapma dışında yer alan veriler anomali değer olarak adlandırılmaktadır. Ancak bu durum tek boyutlu başka bir deyişle tek değişkenli veri setlerindeki anomali gösteren veriyi belirlemek için kullanılabilir. Çok boyutlu veri setlerinde ise farklı istatistiksel uzaklık ölçüleri kullanılır. Bu ölçüler çoğunlukla doğru sonuç vermiş gibi görünse de birime bağımlı olduğu için yanıltıcı olabilir. Tüm bunlar göz önüne alınarak bu risklerden kaçınmak ve literatürde yaygın olarak çalışma yapılmamış bir yöntem kullanılması çalışmada izolasyon ormanı yöntemi kullanılması uygun görülmüştür. Böylelikle literatür çalışmalarından farklı olarak finansal başarısızlık yaşayan ve finansal başarısızlık yaşamayan işletmeler izolasyon ormanı yöntemi kullanılarak etiketlenmiştir. Veri setinin etiketlenmesi çalışmanın devamında denetimli öğrenme algoritmaları kullanmanın önünü açmıştır. Ayrıca izolasyon ormanı yönteminin 8 yıllık bilanço verisi üzerinde uygulanması örneklem sayısının yeterli kadar büyük olmasını sağlamıştır. Ancak anomali tespitinin doğası gereği etiketlenen veri setinde dengesiz sınıf dağılımı problemiyle karşılaşmıştır. Bu problem ise yeniden örnekleme yöntemleri olan rasgele aşırı örnekleme ve rasgele düşük örnekleme yöntemleri aracılığıyla ele alınmıştır. Bu doğrultuda çalışmanın ana amacı olan finansal başarısızlığın tespitinde izolasyon ormanı ve yeniden örnekleme yöntemlerinin başarısı ön plana alınmıştır. Yeniden örnekleme yöntemlerinin kullanılmasıyla birlikte her bir yöntem aracılığıyla farklı veri setleri elde edilmiştir. Elde edilen veri setleri ise ayrı ayrı modellenerek başarıları karşılaştırılmış ve nihai olarak en iyi veri seti ve model seçilmiştir. En başarılı model seçilirken test ve doğrulama veri setleri kullanılmıştır.

İzolasyon ormanı yöntemi test ve doğrulama veri setlerine uygulanmamıştır. Bunun nedeni model tahmini ve tahinlerin doğrulaması aşamasında herhangi bir yanlılığa sebebiyet vermemek olarak açıklanabilir. Çalışmadaki ana amacın izolasyon ormanı ve yeniden örnekleme yöntemleri kullanılarak finansal başarısızlığın tahmin edilmesi olduğu için farklı makine öğrenmesi yöntemlerinin karşılaştırılması yerine güncel olarak literatürde yaygın olarak kullanılan ve başarısını kanıtlamış XGBoost modeli üzerinde durulmuştur. Böylelikle modelin iyileştirilmesi ve yöntemlerin daha detaylı incelenmesinin önü açılmıştır.

Son olarak 2010-2016 yılları arasını kapsayan veriler eğitim verisi olarak kullandığında makine öğrenmesi yönteminin doğası gereği test ve eğitim olarak ayrılarak modelleme işlemi yapılmıştır. Bu sebeple bu aşamada verilecek olan model performans ölçütü olan karmaşıklık matrisini belirtmek yanıltıcı düşünceye sebep olabilir. Çalışma kapsamında izolasyon ormanı yönteminden etkilenmemesi için 2017 yılı test verisi, 2018 yılı verisi ise yeniden örnekleme ile oluşturulan modellerin doğrulanması için kullanılmıştır.

Yapılmış çalışmalar incelendiğinde araştırmaların hemen hemen hepsinde finansal başarısızlık yaşadığı kesin olan ve finansal başarısızlık yaşamayan işletmeler üzerinde çalışılmıştır. Bu çalışmada ise izolasyon ormanı yöntemi kullanılarak tüm veri seti incelenip verinin genel yapısını yansıtmayan birimler finansal başarısız işletme olarak etiketlenmiştir.

Makine öğrenmesi yöntemlerinde veri setinin yetersiz olması durumunda aşırı öğrenme ya da az öğrenme durumlarıyla karşılaşıldığı bilinmektedir. Veri setinin görece büyük olması ile ise daha genellenebilir bir model elde edilebilir. Literatürde konuyla ilgili yapılan diğer araştırmalar incelendiğinde, Kılıç ve Seyrek (2012) imalat sanayisinde faaliyet gösteren 137 işletme üzerinde yaptığı çalışmada yapay sinir ağları modelini kullanarak %94 doğru sınıflama oranına ulaştığını belirtmiştir. Benzer şekilde Kuzu ve Yakut (2020) çalışmalarında imalat sanayisinde faaliyet gösteren 153 işletme üzerinde çalışmışlardır. Farklı teknoloji düzeylerine sahip işletmelerin finansal başarısızlık tahminleri şu şekildedir; düşük teknoloji işletmeler için %93, orta-düşük teknoloji düzeyine sahip işletmeler için %90 ve orta-yüksek teknoloji düzeyine sahip işletmeler için %94. Gerçekleştirilen çalışmada ise en yüksek %78 doğru tahmin gücüne

sahip model elde edilmiştir. Ayrıca çalışma NACE Rev.2 28: Başka Yerde Sınıflandırılmamış Makine ve Ekipman İmalatı sektörü için gerçekleştirildiğinden tüm işletmelerin orta-yüksek teknoloji düzeyine sahip olduğu unutulmamalıdır. Literatür çalışmaları ve yapılan çalışma arasındaki tahmin gücündeki değişimi örnek sayısı ve kullanılan yöntemler açıklamaktadır.

Her çalışmada olduğu gibi bu çalışmada bazı kısıtlar altında gerçekleştirilmiştir. İleriki çalışmalarda finansal olarak başarılı ve başarısız işletmelerin etiketlenmesinde farklı yöntemler kullanılabilir ve yeniden örnekleme yöntemleri geliştirilerek farklı makine öğrenmesi modelleri uygulanarak model performansları karşılaştırılabileceği değerlendirilmektedir.

### **Bilgilendirme / Acknowledgements**

Çalışmada kullanılan tüm veriler Sanayi ve Teknoloji Bakanlığı Girişimci Bilgi Sistemi (GBS)'nden temin edilmiştir.

*All data used in the study were obtained from the Entrepreneur Information System (EIS) of the Ministry of Industry and Technology.*

### **Yazar Katkıları / Author Contributions**

*Kardelen Şeyma Erdem: Literatür Taraması, Kavramsallaştırma, Veri Derleme, Analiz, Modelleme Makale Yazımı-rijinal taslak Mehmet Akif Bakır: Metodoloji, Makale Yazımı-inceleme ve düzenleme Kardelen Şeyma Erdem: Literature Review, Conceptualization, Data Curation, Analysis, Modelling, Writing-original draft Mehmet Akif Bakır: Methodology, Writing-review and editing*

### **Çatışma Beyanı / Conflict of Interest**

Yazarlar tarafından herhangi bir potansiyel çıkar çatışması beyan edilmemiştir.

*No potential conflict of interest was declared by the authors.*

### **Fon Desteği / Funding**

Bu çalışmada herhangi bir resmi, ticari ya da kâr amacı gütmeyen organizasyondan fon desteği alınmamıştır.

*Any specific grant has not been received from funding agencies in the public, commercial, or not-for-profit sectors.*

### **Etik Standartlara Uygunluk / Compliance with Ethical Standards**

Yazarlar tarafından, çalışmada kullanılan araç ve yöntemlerin Etik Kurul izni gerektirmediği beyan edilmiştir.

*It was declared by the authors that the tools and methods used in the study do not require the permission of the Ethics Committee.*

### **Etik Beyanı / Ethical Statement**

Yazarlar tarafından bu çalışmada bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan edilmiştir.

*It was declared by the authors that scientific and ethical principles have been followed in this study and all the sources used have been properly cited.*



Yazarlar, Verimlilik Dergisi'nde yayımlanan çalışmalarının telif hakkına sahiptirler ve çalışmaları CC BY-NC 4.0 lisansı altında yayımlanmaktadır.

*The authors own the copyright of their works published in Journal of Productivity and their works are published under the CC BY-NC 4.0 license.*

## KAYNAKÇA

- Abar, H. (2020). "XGBoost ve Mars Yöntemleriyle Altın Fiyatlarının Kestirimi", *EKEV Akademi Dergisi*, 83, 427-446.
- Akkoç, S. (2007). "Finansal Başarısızlığın Öngörülmesinde Sinirsel Bulanık Ağ Modelinin Kullanımı ve Ampirik Bir Çalışma", Doktora Tezi, Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü, Kütahya.
- Aksoy, B. ve Boztosun, D. (2019). "İmalat İşletmelerinde Makine Öğrenmesi Yöntemleri Kullanılarak Finansal Başarısızlık Tahmini ve Sınıflandırma Performansının Karşılaştırılması: Borsa İstanbul Örneği", *2. Uluslararası Bankacılık Kongresi*, Çorum, 11-18.
- Aktaş, R., Doğanay, M. ve Yıldız, B. (2003). "Finansal Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırması", *Ankara Üniversitesi Siyasal Bilgiler Fakültesi Dergisi*, 58(4), 1-24.
- Altman, E.I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *The Journal of Finance*, 23, 589-609.
- Ay, A. ve Yolaçan E. (2022). "Yeniden Örnekleme Metotlarının Kredi Kartı Sahtecilik Tespiti için Topluluk Öğrenmesine Kapsamlı Analizi", *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 22(5), 1005-1015.
- Beaver, W.H. (1966). "Financial Ratios as Predictors of Failure", *Journal of Accounting Research, Empirical Research in Accounting, Selected Studies*, 4, 71-111.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. ve Kegelmeyer, W.P. (2002). "SMOTE: Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research*, 16(1), 321-357.
- Chen T. ve Guestrin C. (2016). "XGBOOST: A Scalable Tree Boosting System", *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794
- Dağlı, H. (1994). "İşletme Başarısızlıkları ve Alınması Gerekli Önlemler", *Verimlilik Dergisi*, 1, MPM Yayınları, Ankara.
- Dixit, A. (2017). "Ensemble Machine Learning", Birmingham, Packt Publishing Ltd., UK.
- Ertan, A.S. ve Ersan, Ö. (2018). "Finansal başarısızlığı Belirleyen Etkenler: Türkiye İmalat Sektörü Örneği", *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 40(2), 181-207.
- Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B. ve Herrera, F. (2018). "Learning from Imbalanced Data Sets", Springer Cham.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. ve Herrera, F. (2012). "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches", *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 42(4), 463-485.
- Gönenli, A. (1994). "İşletmelerde Finansal Yönetim", İstanbul Üniversitesi. İşletme Fakültesi, İstanbul.
- Han, J. ve Kamber, M. (2006). "Data Mining: Concepts and Techniques", Morgan Kaufmann, San Francisco, CA.
- He, H. ve Garcia, E.A. (2009). "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.
- Karacan, S. ve Savcı, M. (2011). "Kriz Dönemlerinde İşletmelerin Mali Başarısızlık Nedenleri", *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 21(1), 39-54.
- Kılıç, Y. ve Seyrek, İ.H. (2012). "Finansal Başarısızlık Tahmininde Yapay Sinir Ağlarının Kullanılması: İmalat Sektöründe Bir Uygulama", *1. International Symposium on Accounting and Finance*, Gaziantep, 677-689.
- Kurtaran Çelik, M. (2010). "Bankaların Finansal Başarısızlıklarının Geleneksel ve Yeni Yöntemlerle Öngörüsü", *Yönetim ve Ekonomi*, 17(2), 129-143.
- Kuzu, B. ve Yakut, S. (2020). "Destek Vektör Makineleri Yardımıyla İmalat Sanayisinde Mali Başarısızlık Tahminlerinin Teknoloji Yoğunluğuna Göre İncelenmesi", *Osmaniye Korkut ata Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(2), 36-54.
- Liu, F.T., Ting, K.M. ve Zhou, Z. (2008), "Isolation Forest," *2008 Eighth IEEE International Conference on Data Mining*, Pisa, 413-422.
- Nguyen, H.G. (2005). "Using Neutral Network in Predicting Corporate Failure", *Journal of Social Sciences*, 1(4), 199-202.
- Ohlson, J.A. (1980). "Financial Ratios and the Probabilistic Prediction of Bankruptcy", *Journal of Accounting Research*, 1980, 109-131.
- Okka, O. (2009). "Finansal Yönetim Örnek Olayları ve Örnek Çözümler", 2. Baskı, Nobel Yayın Dağıtım, İstanbul.
- Öztürkler, L. (2012). "Finansal Başarısızlık Tahmini", Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Patil A., Framewala A. ve Kazi F. (2020). "Explainability of SMOTE Based Oversampling for Imbalanced Dataset Problems", *3<sup>rd</sup> International Conference on Information and Computer Technologies (ICICT)*, San Jose, CA, USA.

- Soba, M., Akyüz, F. ve Uğurcan, Y. (2016). "Şirketlerin Finansal Performanslarının Altman Yöntemiyle Analizi: Borsa İstanbul Örneği", *Uşak Üniversitesi Sosyal Bilimler Dergisi*, 9(28/4), 65-87.
- Taffler, R.J. (1983). "The Assesment of Company Solvency and Performance Using a Statistical Model", *Accounting and Business Research*, 13(52), 295-307.
- Tang, T.C. ve Chi, L.C. (2005). "Neural Networks Analysis in Business Failure Prediction of Chinese Importers: A Between-Countries Approach", *Expert Systems with Applications*, 29(2), 244-255.
- Toprak Kesgin, H., Shakeri, S., Bulut, N., Yüzükl, S. ve Aktaş, M.S. (2019). "Makine Öğrenmesi Kullanılarak Şirket Bilanço Verilerine Dayalı İflas Riski Tahmini", *4<sup>th</sup> International Conference on Computer Science and Engineering (UBMK)*, Samsun, 195-200.
- Türksoy, A. (2007). "Konaklama İşletmelerinde Mali Başarısızlığa Yol Açan Etmenler". *Ege Academic Review*, 7(1), 99-115.
- Wilson, R.L. ve Sharda, R. (1994). "Bankruptcy Prediction Using Neural Networks", *Decision Support Systems*, 11(5), 545-557.
- Yangın, G. (2019). "XGBoost ve Karar Ağacı Tabanlı Algoritmaların Diyabet Veri Setleri Üzerine Uygulanması", Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Yeşilyurt, S. ve Dalkılıç, H. (2021). "XGBoost ve Gradient Boost Machine ile Günlük Nehir Akımı Tahmini", *3<sup>rd</sup> International Symposium of III Engineering Applications on Civil Engineering and Earth Sciences*, Karabük, Türkiye.