

Öz nitelik Çıkarım Yöntemleri ve Makine Öğrenmesi Kullanarak Şirket Bilanço Verilerine Dayalı İflas Riski Tahmini

Bankruptcy Risk Forecasting Based on Company Balance Sheet Data Using Feature Extraction Methods and Machine Learning

Necip BULUT
İstanbul Üniversitesi İngilizce
İktisat Bölümü
necipbulut@ogr.iu.edu.tr
ORCID: 0000-0003-2532-3992

Saber SHAKERI
Finnet Elektronik, YTÜ
Teknopark, İstanbul
saber.shakeri@finnet.gen.tr
ORCID: 0000-0002-8563-8470

Seçil YÜZÜK
MGA Yazılım, YTÜ
Teknopark, İstanbul
secil.yuzuk@mgasoft.com.tr
ORCID: 0000-0001-9123-8396

Mehmet S. Aktaş
Bilgisayar Müh. Bölümü, YTÜ
aktas@yildiz.edu.tr
ORCID: 0000-0001-7908-5067

Öz

Bir şirketin başarısı hem firmanın iç muhatapları hem de yatırımcılar ve üçüncü kişilerce büyük önem taşımaktadır. Finansal olarak başarısızlık kimi zaman iflaslar ile sonuçlanabilmekte ve firmanın muhatapları üzerinde yıkıcı etkiler yaratabilmektedir. Yatırımcılar, finansörler, yöneticiler bazen de politika yapımcıları için firmaların iflas risklerini tahmin etmek oldukça önemlidir. Literatürde iflas riskinin tahmini için birçok yöntem geliştirilse de Ohlson O-skoru ve Altman Z-skoru iflas riskini tahmin için oldukça sık kullanılan iki yöntemdir. Bu iki modelin hem lineer model olmaları hem de firmaların yalnızca son bilançolarıyla ilgilenmeleri bazen hatalı tahminlere yol açabilmektedir. İflas olgusunun bir süreç olduğu düşünüldüğünde şirketin sadece son finansal raporlarının incelenmesi bir takım sakıncalar barındırmaktadır. Bu sebeple iflas risklerini doğru tahmin etmek için şirketlerin geçmiş finansal raporlarının da incelenmesi gerekmektedir. Literatürdeki bu iki iflas riski tahmin yöntemi şirketlerin sadece son finansal raporlarıyla ilgilenmektedir. Bu çalışmada literatürdeki klasik lineer modeller yerine, lineer olmayan makine öğrenmesi algoritmaları kullanılarak şirketlerin iflas

riskleri tahmin edilmeye çalışılmıştır. Deneysel değerlendirme sonucunda elde ettiğimiz sonuçlar, şirketlerin iflas riskini makine öğrenmesi algoritmalarıyla tahmin etmenin, lineer klasik finansal modellere göre daha başarılı sonuçlar verdiğini ortaya koymaktadır.

Anahtar Sözcükler: Ohlson O-skoru, Altman Z-skoru, Şirket İflas Riski Tahmini, Öz nitelik Çıkarımı, Makine Öğrenmesi.

Abstract

The success of a company has a significant issue for both the interlocutors of companies and other related persons. Financial failure sometimes ends up bankrupt and can have a critical effect on the company's interlocutors. Prediction of bankruptcy is significant for investors, backers, director and sometimes policy makers. Although there are a lot of models to predict bankruptcy in the financial literature, Ohlson O-score and Altman Z-score are models that are used quite often. The fact that the set of models are both linear models and companies are only interested in their latest balance sheets can sometimes lead to incorrect predictions. We argue that bankruptcy should be viewed as a process. Hence, modeling bankruptcy based on only the latest financial reports of the companies has some

Gönderme ve kabul tarihi: 22.10.2019 - 11.12.2019

Makale türü: Araştırma

drawbacks. For this reason, in addition to the latest balance sheets, previously published financial reports of companies should also consider to predict bankruptcy risk of the company. However, aforementioned two financial models are only interest in the latest financial reports of companies. In this study, we study the prediction of the bankruptcy risk of companies by using non-linear machine learning algorithms rather than classical linear models in the finance literature. Our experimental study results indicate that the prediction of the bankruptcy risk of companies by using non-linear machine learning algorithms is more successful compared to linear classical financial models.

Keywords: Ohlson O-score, Altman Z-score, Bankruptcy Risk Prediction, Feature Extraction, Machine Learning.

1. Giriş

Bir şirketin iflas riski tahminleri, şirketin yöneticilerinin yanı sıra o şirkete yatırım yapacak yatırımcılar için oldukça önemlidir. Yatırımcılar yatırım yapacakları şirketin iflas risklerini tahmin etmek için güvenilir yollar aramaktadır. Diğer taraftan firmaların finansal yöneticileri de bu güvenilir metodları aramaktadır. İflas riski tahmin yöntemlerinin başarısı, doğru zamanda firmanın muhataplarının gerekli önlemleri almasına olanak verebilecektir. Yatırımcılar ise elde edebilecekleri bulgularla iflastan doğabilecek yatırım risklerini en aza indirmeyi amaçlamaktadır.

Finans literatüründe iflas risklerini tahmin etmek için çeşitli yöntemler bulunur. Yaygın olarak kullanılan modeller Ohlson O-skoru modeli, Altman Z-skoru modeli ve Taffer modeli olarak sıralanabilir [1, 2, 3]. Ancak yaygın olarak kullanılan bu 3 model de finansal başarısı tahmin edilmek istenen şirketlerin, en son yayınlanan finansal raporunu esas almaktadır. Bu durum, bazı hatalara neden olabilmektedir. Modellerin en son yayınlanan finansal tabloya odaklanması konjonktürel etkileri yakalayabilmektedir. Böylece finansal olarak başarılı bir firma tahmin sonucunda başarısız olarak nitelendirilebilmektedir. Oysa şirketlerin güncel durumlarında, geçmiş faaliyetlerinin önemli bir etkisi bulunduğu görülmektedir. Bu sebeple, finansal başarı veya başarısızlık tahmini (iflas riski tahmini) yapılırken şirketlerin geçmiş finansal raporlarının da dikkate alınması oldukça önemli hale gelmektedir.

Bu çalışmada klasik iflas riski tahmini finansal modellerinden yola çıkılarak, lineer olmayan ve

makine öğrenmesi yöntemleri kullanılan, tahmin modelleri geliştirilmiştir. Yine bu çalışma kapsamında, makine öğrenmesi algoritmalarının iflas riski tahminlerindeki başarıları değerlendirilmiştir.

Borsada işlem gören şirketlerin Kamuyu Aydınlatma Platformu'nda yayınlanan son bilançolarından başlanarak geriye doğru ulaşılabilen tüm bilançoları incelenmiştir. Bilançolardaki tüm ana ve tali kalemler arasından öznitelik çıkarımları, Bilgi Kazanımı ve Temel Bileşenler Analizi olmak üzere 2 farklı öznitelik çıkarım metodu kullanılarak gerçekleştirilmiştir. Diğer bir deyişle, veri seti üzerinde makine öğrenmesi algoritması çalıştırılmadan önce şirketlerin iflas riski tahmininde önemli rol oynayabilecek kalemler, öznitelik çıkarımı yöntemleri kullanılarak belirlenmiştir.

Bu araştırma kapsamında, şirketlerin iflas riski tahmini için Türkiye Finansal Raporlama Standartları'nda yer alan tüm tali ve ana bilanço kalemleri öznitelik havuzuna dahil edilmiştir. Veri seti olarak borsada işlem gören 150 şirketin 2000-2018 yılları arasında bulunan çeyrek yıllık bilanço verileri kullanılmıştır. Şirketlerin iflas risk durumlarını etiketlemek için literatürde yer alan klasik lineer iflas riski tahmin yöntemleri olan Altman Z-skoru ve Ohlson O-skoru değerleri kullanılmıştır. Veri seti eğitim ve test veri seti olarak iki bölüme ayrılmıştır. Daha sonra, eğitim veri seti üzerinde, Rassel Orman, Karar Destek Vektörleri ve Lojistik Regresyon sınıflandırma algoritmaları kullanılarak öğrenme modelleri oluşturulmuş ve test veri seti üzerinde modellerin başarıları irdelenmiştir. Elde edilen modeller, iflas riskine sahip olan ve risk bulunmayan şirketleri başarılı bir şekilde sınıflandırabilmektedir.

Deneysel çalışmada, her bir öznitelik çıkarım yöntemi her bir makine öğrenmesi algoritması ile kombinlenerek sonuçlar elde edilmiştir. Elde edilen sonuçlar, klasik lineer yöntemlerle elde edilen sonuçlarla karşılaştırmalı olarak incelenmiştir. Ayrıca her bir sınıflandırma ve öğrenme kombinasyonu kendi aralarında karşılaştırmaya tabii tutulmuştur. Bu bağlamda iflas riski tahmini için birden fazla model alternatifini sunulmuş ve bu modellerin avantaj ve dezavantajları belirlenmeye çalışılmıştır.

Bu çalışmanın organizasyon yapısı ise şu şekildedir. Bölüm-2'de literatürde yer alan geçmişte yapılmış iflas riski tahmin modellerine yer verilmiştir. Bölüm-3 bu çalışmada önerilen metodolojiyi içermektedir. Bölüm-4'te ampirik bölüm yer almakta ve onun sonuçları karşılaştırmalı olarak incelenmektedir. Son olarak Bölüm-5'te bu çalışmanın sonuçları ve gelecekteki çalışmalar yer almaktadır.

2. Literatür Taraması ve Temel Bilgiler

Günümüzde işletmelerin finansal durumunu değerlendirmek için farklı finansal modeller kullanılmaktadır. Yatırımlarda karar almak için kullanılan finansal modeller, şirketlerin düzenli olarak yayınladıkları bilanço verilerini (mali tablo verileri) kullanmaktadır.

Literatür incelendiğinde şirket başarısızlık tahmininde oldukça yaygın olarak kullanılan iki model olduğu görülmüştür. Bu modeller Altman Z-skoru (1968) ve Ohlson O-skoru (1980)'dur [2]. İflas tahminine yönelik ilk model Altman Z-skoru formülü 1968 yılında Edward I. Altman tarafından yayımlanmıştır [2]. Formül oluşturulurken 66 şirket verileri, bir firmanın iki yıl içinde iflas etme ihtimalini tahmin etmek için kullanılmıştır. Altman Z-skoru, formülü şirketin bilanço verilerini esas almaktadır. Altman Z-skoru iki yıl önce iflasın ön görülmesinde %72 doğru sonuç vermiştir. Altman Z-skorumun formülü Denklem 1'de gösterilmiştir.

$$Z = 1.2T_1 + 1.4T_2 + 3.3T_3 + 0.6T_4 + 0.999T_5 \quad (1)$$

Burada;

T₁: Çalışma sermayesi / Toplam Varlıklar

T₂= Alınmış karlar / Toplam Varlıklar

T₃= Faiz ve vergi öncesi kar / Toplam Varlıkları

T₄= Toplam Piyasa Değeri / Toplam Yükümlülükler

T₅= Toplam Satışlar / Toplam Varlıklar

İflas tahminine yönelik diğer model, Ohlson O-skoru modeli 1980 yılında Dr. James Ohlson tarafından Altman Z-skorumun alternatif olarak geliştirilmiştir [1]. O-skoru, Altman Z-skorumun 66 şirketten geliştirilmiş olmasına karşı 2000'den fazla şirket kullanılarak oluşturulmuştur ve Altman Z-skorumdan daha yüksek bir doğruluk oranına sahiptir. O-skoru, bir mali tabloda yer alan bir dizi finansal orandan üretilen istatistiksel bir iflas riski göstergesi olarak tanımlanabilir. O-skoru modeli, bir firmanın iflas olasılığını tahmin etmek için bilanço verilerinden kolaylıkla elde edilebilecek 9 faktörü kullanmaktadır ve şirketin iflasına dair bir olasılık değeri üretmektedir. O-skoru hesaplanması Denklem 2'de verilmiştir.

$$T = -1.32 - 0.407 \log\left(\frac{TA_t}{GNP}\right) + 6.03 \frac{TL_t}{TA_t} - 1.43 \frac{WC_t}{TA_t} + 0.0757 \frac{CL_t}{CA_t}$$

$$-1.72X - 2.37 \frac{NI_t}{TA_t} - 1.83 \frac{FFO_t}{TL_t} + 0.285Y - 0.521 \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|} \quad (2)$$

Burada;

TA = Toplam Varlıklar

GNP = GNP fiyat endeksi

TL = Toplam Borçlar,

WC = İşletme Sermayesi,

CL = Cari Borçlar,

CA = Cari Varlıklar

NI = Net Kar

FFO = Faaliyet Geliri

Eğer TL > TA ise X=1 değilse X=0

Eğer son iki yılda net zarar varsa Y = 1 yoksa Y = 0

t: Bilanço dönemi

Elsa Imelda ve Clara Ignacia Alodia çalışmasında, Ohlson O-skoru modeli ile Altman Z-skoru modelini karşılaştırmıştır [4]. Bu çalışmada Altman modelinin iflas eşik değeri 2,99 olarak Ohlson O-skorumun ise 0,38 olarak belirlenmiştir. Altman modelinin sırasıyla 1, 2, 3 yıl önceden doğruluk oranı %63, %60, %60 olarak, O-skoru modelinin ise %65, %78, %75 olarak hesaplanmıştır. Bu makale kapsamındaki çalışmada, O-skoru modelinin daha doğru sonuçlar ürettiği gösterilmiştir.

Nanxi Wang ise yaptığı çalışmada, makine öğrenme algoritmalarından Destek Vektör Makinesi (Support Vector Machine), Yapay sinir ağı (Artificial Neural Network) ve kendiliğinden kodlayıcı (Auto Encoder) yöntemlerini kullanarak şirket iflas tahmini yapan bir algoritma geliştirmiştir [5]. Kullandığı bu 3 modelin daha önce kullanılan modellerden (robust logistic regression, inductive learning algorithms, genetic algorithms) daha hassas tahmin yaptığını ve performanslarının daha etkili olduğunu göstermiştir.

Altaş ve Giray çalışmalarında, finansal başarısızlık riski olan şirketleri belirlemek amacıyla Lojistik Regresyon algoritmasını kullanmıştır [6]. Çalışmada, tekstil sektöründe yer alan 2001 yılına ait bilançoları yardımıyla mali oranları hesaplanmış ve dönem sonu kâr-zarar durumuna göre finansal olarak başarısız ve

başarılı ayrımı yapılmıştır. Araştırmacıların geliştirdiği modelin doğru sınıflandırma başarısı %74 olarak hesaplanmıştır.

Muzır ve Çağlar yaygın olarak kullanılan sekiz adet finansal başarısızlık tahmin modelinin Türkiye’de test edilmesi ve başarısızlık öncesi ilk yıl için tahmin performanslarının karşılaştırılmasını hedeflemişlerdir [7]. Çalışmanın sonuçları doğrultusunda, uygulamada yer alan mevcut model önerilerinin %90 düzeyinde veya daha yüksek bir doğru sınıflandırma oranına sahip olamadığı görülmüştür. Ohlson tarafından önerilen O-Skoru modelinin, %81 doğru sınıflandırma oranı ile diğer modellere göre daha başarılı olduğu tespit edilmiştir.

Raut, Sneha, Milind Tiwari, ve Kuldeep Kumar çalışmalarını 2016-2018 yılları arasında Avusturya’daki emlak şirketleri üzerinde finansal başarısızlık tahmini yapmışlardır [8]. Çalışmalarında birçok istatistiksel yöntemi karşılaştırmışlardır. Bu yöntemler arasında başarı oranı en yüksek modelin Yapay Sinir Ağları, Lojistik Regresyon ve Olasılıksal Gradyan Artırma (Stochastic Gradient Boosting) birleştiren hybrid model olduğu gösterilmiştir.

Le, Hong Hanh, ve Jean-Laurent Viviani çalışmasında Amerika Birleşik Devletleri’nde yer alan 3000 bankanın verisi kullanılarak geleneksel metotlar ile makine öğrenmesi algoritmalarının iflas riski tahminleri karşılaştırılmıştır [9]. Bu çalışmadaki deneysel sonuçlar Yapay Sinir Ağları ve K-En Yakın Komşu (K-Nearest Neighbor) algoritmalarının daha doğru sonuçlar ürettiğini ortaya koymaktadır.

Lin, Wei-Chao, Yu-HsinLu ve Chih-FongTsay çalışmalarında şirket iflas modeli için öznelik seçimi aşamasında Bilgi Kazancı ve Genetik Algoritma ile öznelik alt kümesi belirleme yöntemlerini karşılaştırmışlardır [10]. Üç veri seti içinde en düşük hata oranı, Genetik Algoritmayı, Naive Bayes ve Destek Vektör Makinesi birleştiren model ile elde edilmiştir.

Vu, Loan Thi,ve diğerleri çalışmalarında finansal başarısızlık tahmini modeli oluştururken öznelik seçimi yöntemlerinden Faktör Analizini ve F-skoru seçimini kullanmışlardır[11]. Toplamda 12 model oluşturulan bu çalışmada Destek Vektör Makinesi modelinin F-skor seçim yöntemiyle birlikte en iyi performansı verdiği gözlenmiştir.

Onur Akpınar ve Gökçe Akpınar finansal başarısızlık riskinin belirleyicilerinin saptanması için bir çalışma yapmışlardır [12]. Çalışmada düzeltilmiş Altman Z-skoru finansal başarısızlık göstergesi olarak

kullanılmıştır. Elde edilen bulgulara göre; entelektüel sermayenin Z-skorunu pozitif etkilediği saptanmıştır.

Zeytinoglu ve Akarım çalışmalarında Altman’ın finansal olarak başarılı-başarısız ayrımına dayalı uygulamasını Borsa İstanbul örneğinde incelemişlerdir [13]. Elde ettikleri sonuçlara göre; uygulanan yöntemin işletmelerin mali başarısızlığının tahmininde %88’in üzerinde açıklayıcı olduğu tespit edilmiştir.

Bellovary, Jodi L., Don E. Giacomino, ve Michael D. Akers çalışmalarında 1930 yılından 2007 yılına kadar oluşturulan iflas tahmin yöntemlerini karşılaştırmışlardır [14]. Araştırmanın sonucu Çok Değişkenli Doğrusal Ayrıcılık Analizinin ve Yapay Sinir Ağları ile oluşturulan modellerin en başarılı sonuçları oluşturduğu gözlenmiştir. Ayrıca daha fazla değişken kullanan modellerin daha iyi tahmin yaptığı sonucuna ulaşılamamıştır. Çalışmaya göre 2 değişkenli bir algoritma 21 değişkenli bir algoritma kadar doğruluk oranına sahip olabilmektedir.

Acosta-González, Eduardo, Fernando Fernández-Rodríguez, and Hicham Ganga çalışmalarında finansal iflas tahmini için genetik algoritmaları kullanmışlardır [15]. Algoritmaları 1 yıl içinde başarılı olan şirketlerde%98,5, başarısız olan şirketlerde ise %82,5oranında doğruluk oranına sahiptir. Bu başarı oranını 2, 3, 4 yıl içinde korumuşlardır.

Chen, James Ming çalışmalarında klasik istatistiksel yöntemlerin yerini makine öğrenmesi algoritmalarının aldığını belirtmişlerdir [16]. Bu algoritmaların Destek Vektör Makine Öğrenmesi Algoritmalarının; Karar Ağacı Algoritmalarının ve Genetik Algoritmaların doğruluk yüzdelerini artırdığını tespit etmişlerdir [17].

Christopoulos, Apostolos G.,ve diğerleri çalışmalarında finansal iflas modellerinde öznelik seçimi için bir yöntem önermişlerdir [18]. Küme teorisi ile yaptıkları öznelik seçiminin temel varsayımları teyit ettiği gözlenmiştir.

Fernández, Manuel Ángel ve diğerleri çalışmalarında iflas tahmininde kullanılan odaklanmış ve odaklanmamış yöntemleri karşılaştırmışlardır [19]. Bu karşılaştırmada odaklanmamış modellerin daha başarılı sonuçlar ürettiği görülmüştür. Bu çalışma aynı zamanda farklı ekonomik sektörlerdeki iflası açıklayan değişkenlere ilişkin içgörü de sağlamıştır.

Aşağıda, bu çalışma kapsamında kullandığımız öznelik indirgeme (çıkarma) yöntemleriyle ilgili özet açıklamaları veriyoruz. Principle Component Analysis (Temel Bileşenler Analizi): Temel bileşen analizi

(PCA) (Jolliffe 1986) popüler bir veri işleme ve boyut indirgeme tekniğidir [22]. Öznitelik çıkarımı, öznitelik eleme / dönüştürme gibi alanlarda yaygın olarak kullanılan istatistiksel bir metottur. Bu projede ise boyut dönüştürme ve boyut indirgeme (gereksiz boyutlardan kurtulma veya bazı boyutları birleştirme) için kullanılmıştır. Lineer Diskriminant Analiz (Linear Discriminant Analysis - LDA): Linear Discriminant Analysis (LDA) makine öğrenme uygulamaları için ön işleme aşamasında boyut azaltma tekniği olarak kullanılır. LDA, genel olarak PCA'ya benzese de LDA'in çalışma mantığında sınıflar arasındaki uzaklığı maksimize etmek vardır. PCA'da sınıf kavramı yoktur. PCA sadece datapoint'lar arası mesafeyi maksimize ederek öznitelik boyutunu indirgemektedir. LDA sınıflar arasındaki farkı maksimize ederek veri setinde boyut indirgemektir. PCA kümeleme problemlerinde kullanılırken, LDA sınıflandırma problemlerinde kullanılmaktadır. LDA kullanılan eğitim örneğinde yer alan aynı sınıflardaki gözlemlerin geometrik düzlemdeki izdüşümlerinin uzaklıklarını birbirine yaklaştırır [20]. Bu açıdan, LDA eğitim örneğinde yer alan verilerin normalizasyonunu da yaparak daha doğru sınıflandırıcı modeller oluşturulmasında kullanılabilir. Olabildiğince az gözlem kaybına sebep olan Temel Bileşenler Analizi'nin aksine bu yöntem boyut indirgeme işlemi sonrası her bir gözlemin sınıfını daha belirgin hale getirmektedir. Bilgi Kazancı (Information Gain): Bilgi Kazancı özniteliklerin önem değerini belirtmek için kullanılan bir metriktir. Bir değişken kullanılarak veri ayrıştırıldığında, hedef değişkendeki belirsizliğin ne kadar değiştiğini görece olarak gösteren bir değerdir. Entropinin (rassal bir değişkenin belirsizliğinin ölçütü) tersidir ve verilen bir öznitelik için verilen sınıflandırma sonuçlarının ne kadar değer ile kazanılabileceğini gösterir. Bu çalışma kapsamında Bilgi Kazancı metriği kullanılarak, öznitelik çıkarımı (boyut indirgemesi) üzerinde çalışılmıştır. Tüm özniteliklerin Bilgi Kazancı değerleri hesaplanmış ve Bilgi Kazancı değeri belirli bir eşik değeri altında kalan öznitelikler, öznitelik kümesi içerisinde çıkarılmıştır. Bilgi Kazancı değeri eşik değeri üstünde kalan öznitelik kümesi ise makine öğrenmesi algoritmalarında kullanılmıştır.

3. Önerilen Yöntem

Önerilen yöntemin veri akış şeması Şekil-1'de adımları ise Şekil-2'de resmedilmektedir. Burada, ilk olarak, gelen bilanço verileri JSON objesine dönüştürülüp veri tabanına kaydedilmektedir. JSON objesinin bir kısmının örneği Şekil-1'de verilmiştir. Daha sonra her bir JSON objesi içinde saklanan

bilanço verileri üzerinde, Altman Z-skoru ve Ohlson O-skorumu hesaplanıp veri tabanına kaydedilmektedir.

Bir tahmin modeli oluşturmak için bilançoların başarılı ve başarısız olduğuna dair bir etiketle etiketlenmesi gerekmektedir. Bu etiketler finansal analistler tarafından oluşturulabilir. Ancak bu projede bilançolar Ohlson O-skoru ile belirlenen eşik değerine göre otomatik olarak başarılı ve başarısız olmak üzere iki sınıf için etiketlenmiştir. (Literatürdeki verilere dayalı olarak, bilançoların iflas riski yoktur olarak etiketlenmesi için Ohlson O-skorumun eşik değeri 0,38 olarak seçilmiştir.) Daha sonra makine öğrenmesi algoritmaları ile oluşturulacak öğrenme modelleri için veri ön işleme aşamasında öznitelik boyut indirme işlemi yapılmaktadır.

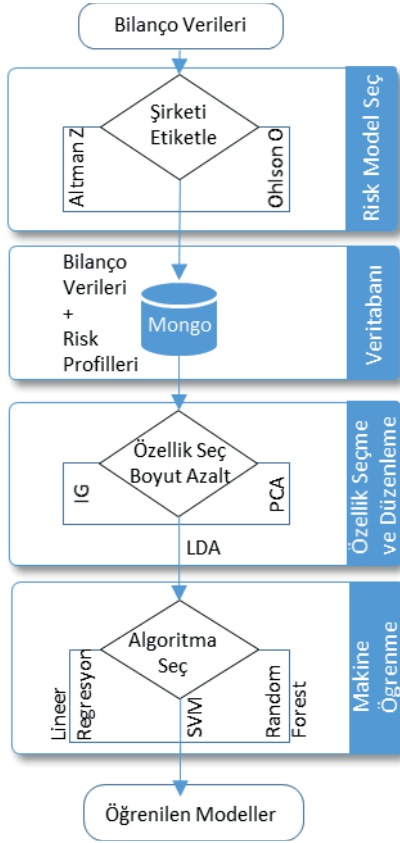
Temel Bileşenler Analizi (Principal Component Analysis-PCA) için veri standart normal forma getirilip boyut sayısı seçildikten sonra öznitelik istenilen boyut sayısına indirgenmiş olur.

Bilgi Kazanımı (Information Gain) için ise veri öncelikle kategorik hale getirilip tüm öznitelik için Bilgi Kazanımı hesaplanır. Bu öznitelikler büyükten küçüğe sıralanır ve belirli bir eşik değerinin üzerindeki bilgi kazanımına sahip öznitelikler seçilir.

Bu iki boyut indirme işleminden sonra, Lineer Diskriminant Analizi (Linear Discriminant Analysis-LDA) yöntemi, öznitelik çıkarımı amaçlı değil, sadece seçilmiş olan özniteliklerin normalizasyonu amaçlı olarak kullanılmıştır [21].

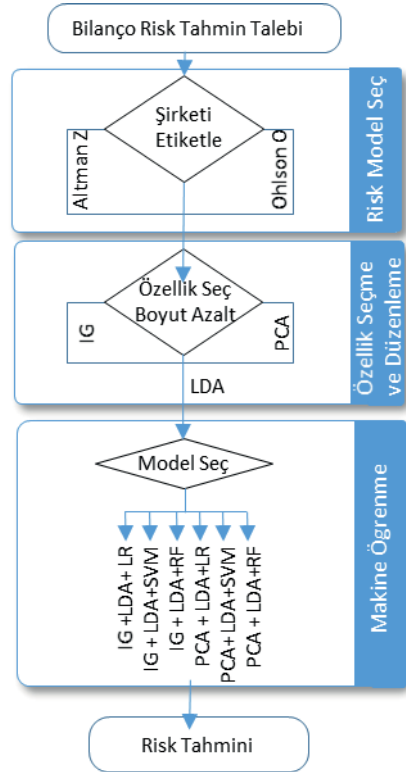
Kullanılan iki öznitelik çıkarım yöntemi (PCA ve Bilgi Çıkarımı tabanlı yöntem), iflas riskini belirleyecek olan 3 farklı sınıflandırma algoritması ile kombinlenerek toplamda 6 farklı iş süreci oluşturulmuştur.

Önerilen yöntem, şirketin mevcut başarı performansını sadece son bilançosuna göre değil geçmişten gelen bilançolarla birlikte değerlendirilmesini sağlamaktadır. Bilançoda yer alan özniteliklerden sabit olarak Ohlson O-skoru yöntemi 10 özelliği, Altman Z-skoru yöntemi 7 özelliği kullanılmaktadır. Önerilen yöntemde şirket etiketleme işlemi bu 2 yöntemle yapılarak daha önce üretilen 6 farklı iflas riski belirleme yöntemiyle tahmin edilir.



Şekil-1: Makine Öğrenmesi Süreci

Bilgi Kazanımı veri boyutu indirgeme yöntemi kullanılarak, öncelikle bilanço verilerinin önem sıralaması elde edilmektedir. Bilançoların iflas riski tahmininde, hangi özniteliklerin en önemli olduğu, önem sırasına göre oluşturulmuş artan boyuttaki öznitelik ile eğitilen modellerin doğruluk metriği açısından başarı değerleri ölçülerek elde edilmektedir. Model değerlendirmesi ile elde edilen değerlerden deneysel olarak optimum öznitelik boyut sayısı bulunmaktadır. Burada, başarısız (iflas riski olan) değerlendirilen bilançolarda, şirket bilançosunun hangi özniteliklerinden dolayı, bu değerlendirmenin yapıldığı görülebilmektedir.



Şekil-2: Tahmin ve Sonuçların Test Süreci

4. Değerlendirme

Bu çalışma kapsamında geliştirilen uygulama Visual Studio 2017 editörü kullanılarak, C# programlama dili kullanılarak gerçekleştirilmiştir. Makine öğrenmesi çalışmaları için "The Accord.NET ML Framework" (link: <http://accord-framework.net>) kullanılmıştır. Borsa İstanbul'da işlem gören 150 şirketin Kamuyu Aydınlatma Platformu'nda açıkladıkları 2000-2018 yıllarına ait çeyrek dönemlik bilanço verileri MongoDB'de JSON formatında saklanmıştır. Bu veriler Türkiye Muhasebe Standartları'nda yer alan 297 farklı bilanço kaleminden oluşmaktadır. Her bir şirket için bu 297 farklı kalem Türk Lirası cinsinden veri tabanında tutulmuştur. İki farklı test uygulanarak, makine öğrenmesi algoritmalarının, şirketlerin iflas riski tahminindeki başarı performansı, finans literatüründe yer alan geleneksel modeller (Altman Z-skoru ve Ohlson O-skoru) ile karşılaştırılmıştır.

Test-1: Bu test için, BIST İstanbul Hisse Senedi Piyasasında yer alan MGROS, EREGL, AKSA, TOFAS, ARCLK, ASELS, FROTO, THYAO

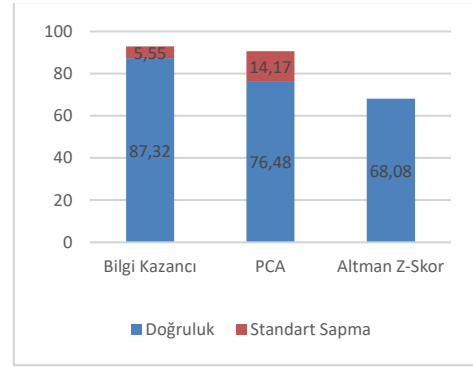
şirketlerine ait bilanço verileri kullanılmıştır. Bu şirketlerinin halka açık bilanço verilerinden toplamda 739 bilanço verisi kullanılmıştır. Şirketlerin bilanço verilerinin iflas riskleri, O-skoru finansal modeli kullanılarak işaretlenmiştir. Makine öğrenmesi algoritmaları kullanılırken, veri boyutu indirgemede, Temel Bileşenler Analizi ve Bilgi Kazanımı yöntemleri Lineer Diskriminant Analizi ile kombinlenerek kullanılmıştır. Makine öğrenmesi algoritmaları için en iyi sonucu veren boyut sayısı deneysel yöntemlerle belirlenmiştir. Uygulama modelin oluşturulması istenen veri setini, etiketler için eşik değerini, eğitim test veri oranını ve özniteliklerin çıkarımı için kullanılan modeli girdi olarak alıp oluşturduğu modelin test sonuçlarını çıktı olarak vermektedir.

Test-1 için, Lojistik Regresyon modelinin başarı değerlendirmesinde 5-katlı çapraz doğrulama yöntemi kullanılmış ve elde edilen doğruluk değerlerinin ortalamaya ve standart sapma değerleri bulunmuştur. Bu test kapsamında, öğrenme modelinin verdiği sonuçlarla, Altman Z-skoru finansal modelinin verdiği değerler karşılaştırmalı olarak incelenmiştir. Bilgi Kazanımı yöntemi kullanılarak, bilançodan elde edilen özniteliklerin önem sıralaması belirlenmiştir. İflas riski tahmininde, en önemli öznitelik gruplarının belirlenmesi için deneysel çalışma gerçekleştirilmiştir. Öznitelik vektör boyut sayıları artırılırken, bu vektörlerden öğrenen makine öğrenme algoritmalarının bilançoların iflas riski tahminindeki başarıları incelenmiştir. En önemli kalemlerin hangileri olduğu deneysel olarak tespit edilmeye çalışılmıştır. Bu test kapsamında; aynı zamanda, Bilgi Kazanımı yöntemi ile öznitelik çıkarımı ve Temel Bileşenler Analiziyle yapılan boyut indirgeme yöntemi ile yapılan öznitelik çıkarım yöntemlerinin, hangisinin daha iyi bir model oluşturduğu araştırılmıştır. Bu test kapsamında, farklı boyut sayıları için, Bilgi Kazanımı ve Temel Bileşenler Analizi Lineer Diskriminant Analizi Visual Studio 2017 - C# programı kullanılarak kombine edilmiş ve Doğruluk (Accuracy) değeri tabloları oluşturulmuştur. Çizelge-1’de modelin 5katman çapraz doğrulama yöntemiyle test sonuçları gösterilmiştir. Burada, her bir katman için yapılan test sonucu elde edilen değerlerin ortalama ve standart sapma değerleri verilmektedir.

Çizelge-1: Modelin 5-KatlıÇaprazDoğrulamaTest Başarı Oranları

	Öznitelik Çıkarımı /Boyut İndirgeme	Doğruluk	Standart Sapma
10 Boyut	Bilgi Kazancı	88,59	6,41
	PCA	59,61	6,49
20 Boyut	Bilgi Kazancı	87,32	5,55
	PCA	76,48	14,17
30 Boyut	Bilgi Kazancı	83,67	6,30
	PCA	71,63	9,14
40 Boyut	Bilgi Kazancı	83,44	6,16
	PCA	72,16	13,12

Modelin Altman Z-skoru ile karşılaştırılması Şekil - 3’te gösterilmiştir. Bu değerlendirmede Altman Z-skorunun gri alan olarak belirlediği değerler dikkate alınmamıştır. Model için ise boyut sayısı 20 olarak kullanılmıştır. Böylece Bilgi Kazancı boyut indirgeme yöntemi diğer yöntemlere nispeten iflas riski tahminlerinde daha iyi performans göstermiştir. Yapılan analizlerden elde edilen bilgiler doğrultusunda; diğer bir boyut indirgeme modeli olan Temel Bileşenler Analizi %76,48 ile ikinci en iyi performansı sergilemiştir.



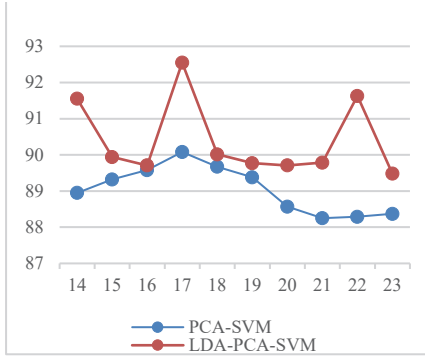
Şekil-3: Modelin Altman Z-Skoru İle Karşılaştırılması

Bilgi Kazancı Yöntemi ile oluşturulan model kullanılarak, şirketlerin iflas riskine dair bilgiler içeren en önemli kalemler tespit edilmiştir. Bu kalemler sırasıyla şu şekildedir:

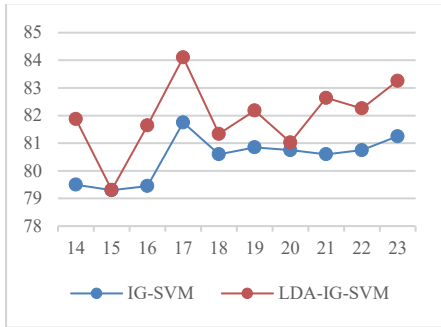
- 1) Uzun vadeli yükümlülükler (TL)
- 2) Ertelenen vergi varlıkları (TL)
- 3) Ana ortaklık dışı öz sermaye (TL)

- 4) Diğer faaliyetlerden zararlar ve giderler (-) (TL)
- 5) Net dönem karı/zararı (TL)
- 6) Finansal duran varlıklar (TL)
- 7) Vergi öncesi kar/zarar (TL)
- 8) Ödenecek vergi ve yasal yükümlülükler (TL)
- 9) Kısa vadeli ticari alacaklar (TL)
- 10) Ticari Borçlar (TL)

Test-2: Bu test kapsamında daha geniş bir bilanço veri seti üzerinde makine öğrenmesi algoritmaları uygulanarak, iflas riski tahminleri gerçekleştirilmiştir. Burada, 150 farklı şirket bilanço verisi kullanılmıştır. Şirket verilerinden çıkartılan öznelik sayısı ise 230 adettir. Test kapsamında kullanılan bilanço verisi sayısı 10800 olmakla birlikte, 18 yıl süresince yayınlanmış olan verilerden oluşmaktadır. Bu test kapsamında, etiketli verilerin %80' i eğitim ve %20' si test verisi olarak kullanılmıştır.



Şekil-4: PCA-SVM ve LDA-PCA-SVM Deneyleri İçin Doğruluk Oranları [21]



Şekil-5: Farklı Öznelik Boyutları İçin IG-SVM ve LDA-IG-SVM Deneyleri İçin Doğruluk Oranları [21]

Şekil-4 ve Şekil-5 bu araştırma kapsamında incelediğimiz makine öğrenmesi algoritmalarının iflas riski tahminindeki başarılarını tespit etmek için

gerçekleştirdiğimizi deneysel çalışmanın sonuçlarını içermektedir. Buradaki testlerde, veri boyutu indirgeme yöntemleri Temel Bileşenler Analizi (PCA) ve Bilgi Kazanımı (IG) yöntemleri Linear Diskriminant Analizi (LDA) ile kombine edilerek farklı öznelik boyutları için öğrenme algoritmalarının başarıları incelenmiştir. Şekil 4 ve Şekil-5'te 14'ten başlanarak 23'e kadar farklı öznelik sayılarıyla olan başarı gözlemlenmektedir. Her iki yöntemde de 17 öznelikle yapılan iflas riski tahminin en yüksek başarıya sahip olduğu tespit edilmiştir.

Şekil-4'te ise aynı SVM algoritması farklı öznelik çıkarım yöntemleriyle kombine edilerek küçük örneklem boyutuyla tahmin yapılmıştır. Bu sonuçlara göre PCA yöntemi LDA yöntemiyle kombine edilerek yapılan tahminlerin daha başarılı olduğu gözlemine ulaşılmıştır. Daha büyük örneklem boyutlarında ise Rassal Orman sınıflandırma yöntemi başarılı sonuç verse de LDA-PCA-SVM kombinasyonuna nispeten daha düşük başarı yüzdesine sahiptir.

Şekil-5, aynı SVM algoritmasının önce sadece Bilgi Kazanımı ile daha sonra Bilgi Kazanımı ve Linear Diskriminant Analizi ile kullanımını göstermektedir [21]. Görüldüğü üzere 15 öznelik ile yapılan tahmin yöntemi haricinde diğer tüm öznelik sayılarında LDA-IG-SVM yönteminin iflas riskini %80 üzerinde bir başarıyla tahmin ettiği tespit edilmiştir. Ek olarak Rassal Orman ve Lojistik Regresyonla üretilen iflas riski yöntemleri, Şekil-4 ve Şekil-5'te gösterilen 4 farklı yöntemden daha düşük bir başarı sergilemiştir.

5. Sonuç ve Gelecekteki Çalışmalar

Yapılan çalışmada şirketlerin sadece son bilançosuna odaklanan literatürdeki lineer yöntemlere ek olarak şirketlerin geçmiş dönem bilançolarını da analize katan 6 farklı melez model kullanılmıştır. Bu modeller Borsa İstanbul'da işlem gören 150 firmanın bilanço verileriyle analiz edilmiştir. Yapılan analiz sonucunda makine öğrenmesi yöntemleriyle yapılan lineer olmayan tahminlerin literatürdeki lineer yöntemlerden daha başarılı olduğu sonucuna varılmıştır. Ayrıca makine öğrenmesi algoritmaları çeşitli öznelik çıkarım yöntemleri ile kombine [KE1] edilerek kendi aralarında da bir değerlendirmeye tabii tutulmuştur. Bu öznelik çıkarım yöntemleri ile şirketlerin iflas risklerini açıklayabilecek en önemli bilanço kalemlerini belirleyerek daha başarılı sonuçlar elde edilmiştir. Veri etiketleme yöntemi olarak ise AltmanZ-skoru ve Ohlson O-skoru kullanılmıştır. Yapılan çalışmalar boyunca şirketlerin iflas riskini en düşük hata ile tahmin eden yöntemin Karar Destek Vektörleri algoritması olduğu tespit edilmiştir. Bu

algoritmaya veriler gönderilmeden önce Temel Bileşenler Analizi ve Linear Diskriminant Analizi kombinasyonu kullanılarak en önemli ana ve tali bilanço kalemlerinin seçilmesi daha yüksek başarıya sahip tahmin sonuçları doğurmuştur. Kısaca bu çalışmada üretilen PCA-LDA-SVM şirket iflas riski tahmin yöntemi diğer modellere göre oldukça yüksek başarı kaydetmiştir. Bu çalışma literatürde yeni iflas riski tahmin modellerinin geliştirilmesinin önünü açmaktadır. Şirket etiketleme yöntemi için piyasa ve saha araştırması yapılarak finansal makine öğrenmesi yöntemlerinin geliştirilmesi hem literatüre hem de finans piyasalarına oldukça önemli katkılar sunacaktır.

Mali tablolara dayalı finansal riskin tahmin edilmesinde en önemli aşama mali tablolardaki kalemlerin risk tahminine olan etkilerinin belirlenmesidir. Hangi kalemlerin iflas göstergesi olarak nitelendirilebileceğini tespit edebilmek tahmin işlemlerine geçmeden önceki en kritik aşamadır. Birçok sınıflama ve regresyon algoritmalarında tespit edilen öznitelikler kullanılarak hesaplama yapılmaktadır. Ayırt edici bilgi içermeyen özniteliklerin elenmesi algoritmaların doğruluğunu arttıracaktır. Gelecek çalışmalar arasında; çok sayıda kalem içeren bilançolardan en az sayıda bağımsız kalemlerin seçim için Genetik Algoritmaların kullanımını [planlıyoruz](#)[KE2].

Kaynakça

- [1] Ohlson, James A. "Financial ratios and the probabilistic prediction of bankruptcy." *Journal of accounting research*, 1980, pp. 109-131
- [2] Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The journal of finance* 23.4, 1968, pp. 589-609
- [3] Taffler, Richard J. "Empirical models for the monitoring of UK corporations." *Journal of Banking & Finance* 8.2, 1984, pp. 199-227
- [4] Imelda, Elsa, and Ignacia Alodia. "The Analysis of Altman Model and Ohlson Model in Predicting Financial Distress of Manufacturing Companies in the Indonesia Stock Exchange." *Indian-Pacific Journal of Accounting and Finance* 1.1, 2017, pp. 51-63
- [5] Wang, Nanxi. "Bankruptcy prediction using machine learning." *Journal of Mathematical Finance* 7.04, 2017, pp. 908
- [6] Altaş, Dilek, and Selay Giray. "Mali Başarısızlığın Çok Değişkenli İstatistiksel Yöntemlerle Belirlenmesi: Tekstil Sektörü Örneği.", 2005
- [7] Muzir, Erol and Caglar, Nazan (2009), "The Accuracy of Financial Distress Prediction Models in Turkey: A Comparative Investigation with Simple Model Proposals", *Anadolu University Journal of Social Sciences*, 9(2): 15-48
- [8] Raut, Sneha, Milind Tiwari, and Kuldeep Kumar. "Financial Distress Prediction Using Cutting-Edge Statistical Techniques: A Study of Australian Real Estate Sector." *Shodh-Amrit: JKLU Journal of Engineering & Management* 1.2, 2018, pp. 2-32
- [9] Viviani, Jean-Laurent, and Hong LE Hanh. "Why Do Banks Fail?-The Explanation from Text Analytics Technique.", *SSRN Electronic Journal*, 2018
- [10] Lin, Wei-Chao, Yu-Hsin Lu, and Chih-Fong Tsai. "Feature selection in single and ensemble learning-based bankruptcy prediction models." *Expert Systems* 36.1, 2019, e12335
- [11] Vu, Loan Thi, et al. "Feature selection methods and sampling techniques to financial distress prediction for Vietnamese listed companies." *Investment Management & Financial Innovations* 16.1, 2019, pp. 276
- [12] Akpınar, Onur, and Gökçe Akpınar. "Finansal Başarısızlık Riskinin Belirleyicileri: Borsa İstanbul'da Bir Uygulama.", *Journal of Business Research - Turk*, Volume 9, Issue 4 2017
- [13] Zeytinoglu, Emin, and Yasemin Deniz Akarim. "Financial failure prediction using financial ratios: An empirical application on Istanbul Stock Exchange." *Journal of Applied Finance and Banking* 3.3, 2013, pp. 107
- [14] Bellovary, Jodi L., Don E. Giacomino, and Michael D. Akers. "A review of bankruptcy prediction studies: 1930 to present." *Journal of Financial education*, 2007, pp. 1-42
- [15] Acosta-González, Eduardo, Fernando Fernández-Rodríguez, and Hicham Ganga. "Predicting corporate financial failure using macro economic variables and accounting data." *Computational Economics* 53.1, 2019, pp. 227-257
- [16] Chen, James Ming. "Models for Predicting Business Bankruptcies and Their Application to Banking and to Financial Regulation." Available at SSRN 3329147, 2019
- [17] Shakeri, S., & Ashouraei, M. (2016). Neur Ada: Combining artificial neural network and Adaboost for accurate object detection. *International Journal of Next-Generation Computing*, 7(2), pp. 155-163
- [18] Christopoulos, Apostolos G., et al. "An implementation of Soft Set Theory in the Variables Selection Process for Corporate Failure Prediction Models. Evidence from NASDAQ Listed Firms." *Bulletin of Applied Economics* 6.1, 2019, pp. 1-20
- [19] Fernández, Manuel Ángel, et al. "Focused vsunfoc used models for bankruptcy prediction: Empirical evidence for Spain." *Contaduría y Administración* 64.2, 2019, e96

- [20] Varatharajan, R., Manogaran, G., & Priyan, M. K. (2018). A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing. *Multimedia Tools and Applications*, 77(8), 10195-10215
- [21] Zengin T., Shakeri S., Bulut N., Yüzük S. and Aktaş M.S. (2019, September). Bankruptcy Risk Forecast Based on Company Balance Sheet Data Using Machine Learning. In 2019 4rd International Conference on Computer Science and Engineering (UBMK). IEEE. 2019
- [22] I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.