VERİ MADENCİLİĞİ YÖNTEMİNİN FİNANS SEKTÖRÜNE UYGULANMASINA YÖNELİK KAVRAMSAL BİR DEĞERLENDİRME

A Conceptual Evaluation on Implementation of Data Mining Method in the Finance Sector

Mesut DOĞAN

*Doç. Dr., Afyon Kocatepe Üniversitesi
mesutdogan07@gmail.com
https://orcid.org/0000-0001-6879-1361*

Mustafa KEVSER

*Dr. Öğr. Üyesi, Bandırma Onyedi Eylül Üniversitesi
mkevser@bandirma.edu.tr
https://orcid.org/0000-0003-0586-1662*

Makale Başvuru Tarihi / Received: 29.05.2021

Makale Kabul Tarihi / Accepted : 29.06.2021

Makale Türü / Article Type: Araştırma Makalesi

ÖZET

Anahtar Kelimeler:

*Veri Madenciliği,
Finans, Bankacılık*

Bu araştırmada veri madenciliğinin finans sektörüne uygulanmasına yönelik kavramsal bir tarama yapılmıştır. Başka bir ifade ile finansal veriler ile veri madenciliği yöntemi kullanan ampirik çalışmalar özetlenmiştir. Yapılan genel tarama sonucunda veri madenciliğinin bir çok alanda kullanıldığı ve finans alanında yapılan ampirik çalışmalarda yaygın bir şekilde yer bulduğu belirlenmiştir. Veri madenciliği yöntemleri kullanılarak yapılan finans çalışmaları incelendiğinde söz konusu çalışmaların önemli bir bölümünün risk yönetimi ve finansal başarısızlık alanında yapıldığı tespit edilmiştir. Çalışmada ayrıca finans alanında veri madenciliği yöntemlerinden en fazla yapay sinir ağlarının kullanıldığı ortaya konmuştur. Yapılan genel tarama sonucunda veri madenciliği yöntemleri ile kredi skorlama, anormal getiri tahmini, manipülasyon tespiti ve portföy yönetimi konularında araştırmaların yapılmasına ihtiyaç duyulduğu belirlenmiştir.

ABSTRACT

Keywords:

*Data Mining,
Finance, Banking*

In this research, a conceptual review was made for the application of data mining to the financial sector. In other words, it is summarized in empirical studies using financial data and data mining method. As a result of the general scanning, it was determined that data mining is used in many areas and it is widely used in empirical studies in the field of finance. When the financial studies using data mining methods are examined, it has been determined that a significant part of the studies in question are carried out in the field of risk management and financial failure. In the study, it was also revealed that artificial neural networks are the most used data mining methods in the field of finance. As a result of the general survey, it was determined that there is a need for research on credit scoring, abnormal return estimation, manipulation detection and portfolio management with data mining methods.

1. GİRİŞ

Teknolojinin gelişmesi ile ticaret hacmi tüm dünyada gelişmekte, bu gelişmeye paralel olarak üretilen veri miktarı da hızlı bir şekilde artmaktadır. Söz konusu veriler her geçen gün yaşamla daha fazla ilgili olarak vazgeçilmez unsurlarından biri olmaktadır. Yüksek veri miktarı, veri depolama ve işleme faaliyetlerinin ulaşılabilirliğinin artması birtakım fırsatları da beraberinde getirmiştir. İnsan doğası gereği, günlük yaşamın belirsizliği nedeniyle yaşamı boyunca riski minimuma indirmeye çalışmakta ve tüm fırsatlardan yararlanmak istemektedir. Kendine ayırmak istediği vakti mümkün olduğunca arttırmaya çalışır. Bu ve bunun gibi sebeplerden dolayı veri madenciliğinin gelecekte olabilecek olayları tahmin edebilmesi, gözle görülmeyen ya da görüle bile önemsenmeyen fırsatları görebilmesi, daha önce tespit edilememiş olan ilişkileri tespit edebilmesi ve bütün bunları çok hızlı bir biçimde yapabilmesi nedeniyle tercih sebebi olmaktadır (Gültekin, 2017).

Veri madenciliği yöntemi kapsamlı veriler içerisinde “değeri olan” bir bilgiyi temin etme işlemidir. Söz konusu veriler arasındaki ilişkiyi belirlemek ve gerektiğinde geleceğe yönelik tahminlerde bulunmak mümkündür. Başka bir ifade ile veri madenciliği yöntemi bir birimde üretilen bütün verilerin bir takım tekniklerden yararlanarak mevcut veya ileride ortaya çıkabilecek örtük bilgiyi gün ışığına çıkarma süreci olarak ifade edilebilir. Bu bağlamda veri madenciliğinin örgütlerin karar destek sistemleri için önemli bir yere sahiptir (Özkan, 2008).

Veri Madenciliğinin, finans alanında kullanımı dünyada giderek yaygın hale gelmekte ve sık kullanılan yöntemlerden birisi olma yolunda ilerlemektedir. Müşteri verisi, portföy verisi, şirket verisi vb. bilgiler veri tabanlarında depolanmaktadır. Veri madenciliği teknikleri kullanılarak, kredi riski, pazar riski, portföy yönetimi ve daha bir çok alanda değerlendirme ve önlem amaçlı kullanılmaktadır (Uzar, 2013). Özellikle risk yönetimi, portföy yönetimi, mali başarısızlık tahminleri, manipülasyon tespiti, döviz piyasası, dolandırıcılık tespiti, risk profillerinin belirlenmesi, olağandışı borsa getirilerinin tahmini, yatırımcıların korunması gibi finans alanlarına uygulanmaktadır.

Veri madenciliği yönteminin gizli örüntüleri ve faydalı bilgileri ortaya çıkarma durumu finans alanında kullanılmasını sağlamıştır. Özellikle kredi taleplerinin değerlendirilerek kredi geri ödemelerinin kontrol altında tutulmasında, müşteri gruplarının ve riskli müşteri tiplerinin belirlenmesinde, sadık müşteri portföyünün oluşturulmasında, müşterilere özgü satış politikalarının oluşturulmasında, kredi kartı müşterilerinin kaybedilme sebeplerinin belirlenmesinde, banka satış personellerinin performanslarının değerlendirilmesinde, banka yatırım fonu müşterisi iken yatırım fonu müşterisi olmaktan çıkmış müşterilerin işlem hareket detayının öğrenilerek hesabını kapatmaya meyilli müşterilerin tespit edilmesinde, bilanço ve gelir tablosuna bakarak işletmelerin başarılı olup olmadığının tahmin edilmesinde, kredi kartı dolandırıcılıklarının tespitinde, anormal getiri tespiti, hisse senedi fiyatı, döviz kuru tahmininde, kurve faiz oranı tahmininde, finansal risklerin anlaşılmasında ve yönetilmesinde, portföy yönetimi, finansal kriz ve sıkıntıların tahmininde yaygın bir şekilde kullanılmaktadır (Kaymaz, 2019).

Bu çalışmanın amacı veri madenciliğinin finans sektörüne uygulanmasına yönelik literatür taraması gerçekleştirmektedir. Başka bir ifade ile finansal veriler ile veri madenciliği yöntemi kullanan ampirik çalışmalar özetlenecektir. Çalışmada veri madenciliğinin finans alanında uygulama alanı örnekleri ortaya konarak literatürün gelişmesine ve bu yöntemin daha yaygın kullanılmasına yardımcı olacaktır.

2. VERİ MADENCİLİĞİ MODELLERİ VE ALGORİTMALARI

2.1.Yapay sinir ağları

Yapay sinir ağları değişkenler arasındaki karmaşık ilişkilerin ve etkileşimin bulunduğu ya da tek bir çözüm kümesinin bulunmadığı hallerde çok iyi sonuçlar tespit eden yapay zeka teknolojilerinden birisidir. Bu bağlamda yapay sinir ağı finansal başarısızlık tahmininde kullanıma uygun bir araçtır (Keskin, 2005). Yapay sinir ağları farklı alanlarda kullanılan ve söz konusu alanlarda problemlerin çözümünde yararlanılabilen çok güçlü ve esnek yöntemlerden birisidir. Yapay sinir ağlarının kullanılma amaçları arasında örüntü tanıma, sınıflandırma, örüntü tamamlama, eşleştirme, optimizasyon ve kontrol sayılabilir. Söz konusu alanlarda gerçekleştirilen güncel araştırmalar yapay sinir ağlarının güçlü örüntü sınıflandırma ve kestirim kapasitelerini kanıtlamışlardır. Bu ağların yapısı bağımlı değişken ve kestirimci değişkenler arasındaki çok karmaşık ilişkileri bulmaya yardımcı olur (Irmak, 2009).

2.2.K- En Yakın Komşu

K-En Yakın Komşu yöntemi; Büyük veri tabanlarında kullanılan sınıflandırma tekniklerinden birisidir. Sınıflandırılmak istenen nesnenin ait olduğu kümeyi, en yakınında yer alan K birim nesneden en fazla birime ait olanla aynı kümede sınıflandırması mantığına dayanmaktadır (Koyuncugil ve Özgülbaş, 2009: 27). Söz konusu yöntemde örneklemelerin tamamı bir örüntü uzayında saklanmaktadır. Algoritma, bilinmeyen bir örneklemin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örnekleme en yakın olan k örneklemini bulmaktadır. Ardından bilinmeyen örneklem, k en yakın komşu içinden en çok benzediği sınıfa atamaktadır. K-en yakın komşu algoritması, aynı zamanda, bilinmeyen örneklem için bir gerçek değerin tahmininde de kullanılmaktadır (Toktaş ve Demirhan, 2004).

2.3.Sınıflama Analizi

Sınıflandırma, veri sınıflarını tanımlamak veya ayırt etmek amacıyla bir model bulmaktadır. Model üzerinde çalışılan veri seti üzerinden oluşturulmakta ve amaç yeni bir nesnenin belirli sınıflar içinde herhangi sınıfa ait olup olmadığını belirleyecek bir sınıflayıcı oluşturmaktır. Benzer şekilde bir veri kümesi içinde ortak yönleri veya farklılıkları ortaya koyacak şekilde sınıflandırma yapılabilir (Tuffery, 2008). Sınıflandırma, her bir kaydı veri setinde yer alan çeşitli gruplara dağıtmaktadır. Sınıflandırmada yeni örneklemin özellikleri irdelenerek, var olan gruplardan hangisine ait olduğunu belirleme amacı bulunmaktadır. Sınıflandırma, hangi gruba ait olduğu bilinmeyen verilerin sınıflarının belirlenmesidir (Özkan, 2008).

2.4.Aykırlık Analizi

Aykırlık analizi, verilerin geriye kalan bölümünden karakteristik olarak önemli derecede farklı olan gözlemlerin tanımlanmasıdır. Bu tarz gözlemler aykırılık veya anomali olarak bilinmektedir. Aykırılık tespiti algoritmasında gerçek anomalilerin keşfedilmesine ve normal objelerin yanlışlıkla aykırı olarak etiketlenmesine engel olmaktır. Başka bir ifade ile iyi bir anomali bulucunun yüksek bir tespit oranına ve düşük bir yanlış alarm oranına sahip olması gerekmektedir. Anormal getiri bulma uygulamaları arasında hile araştırmaları, manipülasyon, network saldırıları ve ekosistem bozulmaları olarak gösterilebilir (Kantardzic, 2003).

2.5.Karar Ağacı Sınıflandırma

Bu yöntem basit tek faktörler ve karar ağacı yapısını gösterecek biçimde oluşturulmakta ve karar ağacı aynı zamanda kategorik ve sürekli veriyi gösterebilir. Bir karar ağacı üç ögeden oluşmaktadır. Birincisi karar düğümüdür ve test değişkenini belirtmektedir. İkincisi bir uç veya daldır ve olası değişken sonucuna uygundur. Sonuncusu ise bir yaprak veya cevap düğümüdür. Nesnenin ait olduğu sınıf bilgisini içerir (Amor vd., 2004). Düğüm gösterimi bütün veri seti kayıtlarını, alanları ve alan değerlerini gösterir. Kuralları oluşturan alanlar veya sütunlar giriş verileri olarak adlandırılırlar. Kurallara ayırma işlemi ardı ardına gerçekleştirilerek ters bir ağaç gibi dalların, alt dalları oluşturduğu hiyerarşik bir sistem oluşturur. Başka bir ifade ile karar ağaçları, sınıflandırma ve tahminlemede kullanılan yöntemlerden birisidir ve veri madenciliğinde çok sık kullanılmaktadır. Bu model düğümler ve dallardan oluşan ve anlaşılabilirliği kolay olan bir tekniktir. Karar ağacı modelinde olan her bir dalın belirli bir olasılığı da mevcuttur. Böylelikle son dallardan köke ya da istenilen yere ulaşana kadar olasılıkların hesaplanması da yapılabilir (Ertikin, 2017: 82). Karar ağaçlarının, anlaşılır ve basit olması çok kolay bir şekilde yorumlanabilmesi, uygun maliyetli olması, diğer veri tabanları ile ortak hareket edebilmesi, iyi sonuçlar vermeleri nedeniyle bir çok kişi veya kurum tarafından yaygın olarak kullanılan tahmin yöntemlerinden birisidir (Chien ve Chen, 2008).

2.6.Bayes Ağları

Bayes sınıflandırma tekniği Bayes Teoremi'nin üzerine kurulmuştur. Bayes sınıflandırıcıları istatistiksel sınıflandırıcılardan birisidir. Bayes ağları değişken veri setleri arasındaki olasılıksal ilişkileri gösteren etkili ve doğal bir yöntemdir. Bir bayes ağı yakın öncülleri verilen grafikte yönlendirilmiş her düğüm için bir olasılık dağılımı oluşturmaktadır ve. Bir bayes ağı sınıflandırıcısı kategorik veri içeren bir veri seti üzerindeki ortak olasılık dağılımını temsil etmektedir. Bu yöntem düğümlerden oluşan yönlendirilmiş G aksilik grafiği ve yaylar ve koşullu olasılık tablolarıdır. Düğümler; değerlerin direkt bağımlılıklarını içeren yayları temsil eder. Bayes ağındaki yayların yoğunluğu karmaşıklığı ölçmektedir. Seyrek bayes ağları ise basit olasılıksal modelleri yoğun bayes ağları ise yüksek karmaşıklığa sahip modelleri temsil eder. Bu bağlamda bayes ağları olasılıksal modellemede esnek bir yöntemdir (Gültekin, 2017). Bayes ağının iki bileşeni vardır: yönlendirilmiş bir asiklik grafik ve olasılık dağılımıdır. Yönlendirilmiş asiklik grafikteki düğümler, stokastik değişkenleri temsil etmekte ve yaylar koşullu olasılık dağılımları ile ölçülen değişkenler arasındaki yönlendirilmiş bağımlılıkları temsil etmektedir (Lior ve Oded, 2008).

2.7.Kümelenme

Kümelenme olayların, nesnelerin, durumların birbirine benzer olanlarının bir arada sınıflandırılmasıdır. Kümelenme analizi, sınıflandırma yöntemlerinden farklı olarak nesnelerin daha önceden bilinmeyen sayıda sınıfa ayrılmasına imkan sağlamaktadır. Değişkenler arası bir bağımlılık söz konusu olmadığından bu yöntemle denetimsiz öğrenme ya da danışmansız öğrenme gibi isimler de verilmektedir (Balaban ve Kartal, 2015). Kümelenme analizlerinde genel olarak nesneler arasındaki uzaklık kullanılarak nesnelerin ortak yönleri veya farklılıkları ortaya çıkarılmaya çalışılmaktadır. Ulusal ve uluslararası literatürde birçok kümeleme analizi bulunmakla beraber en çok kullanılan kümeleme analizleri K-ortalamlar ve hiyerarşik kümeleme analizleridir (Hastie vd., 2001).

2.8.Genetik Algoritmalar

Genetik algoritmalar, John Holland tarafından verilerin kodlanmış biçimlerini kullanan olasılık kuramlarına göre kullanılan sürekli bir amaç fonksiyonu olan teknik algoritmalarından birisidir. Örneklem sayısına göre verilerin tamamı üzerinde çalışma yapmak yerine belirli bir kısmı üzerinde çalışma yaparak zaman tasarrufu olanak sağlar. Genetik algoritmalar bilgisayar ara yüzlerini etkin olarak kullanılmakta ve tek bir yapı yerine, çözüm kümeleri meydana getirerek problemin çözümü için alternatif sonuçlar getirmektedirler. Söz konusu çözüm kümelerine nüfus adı verilmekte ve nüfus içerisindeki elemanlar gen, birey, kromozom adlarını almaktadır. (Gülçe, 2010). Genetik algoritmalar yöntemi olasılık kuramları çerçevesinde çalışmakta, çözüm olayının karmaşık ve geniş olduğu problem tipleri açısından olumlu sonuçlar vermektedir. Genetik algoritmalar, finans alanında ve bu alanda özellikle yatırım araçlarının performansını ölçmede, piyasaların getirileri ilgili tahminle yapma, işletmedeki mali kayıpların araştırılmasında ve tahmininde, pazarlama ve üretim alanında yaygın kullanıma sahiptir (Gül ve Çağatan, 2002).

2.9.Destek Vektör Makineleri

Vapnik ve Chervonenkis (1971) tarafından geliştirilen Destek Vektör Makineleri istatistiksel öğrenme kuramına dayanan bir makine öğrenmesi yöntemlerinden birisidir. Destek vektör yöntemi son yıllarda özellikle veri madenciliğinde değişkenler arasındaki ilişkilerin tahmin edilemediği durumlarda yaygın bir şekilde kullanılmaktadır. Bu yöntem, temelde iki grulu problemlerin tahminlemesinde doğrusal bir sınıflayıcı olarak düşünülse de, daha sonra doğrusal olarak ayrılamayan veya çok sınıflı sınıflama problemlerinin çözümüne de genelleştirilerek, bu problemlerin çözümünde de kullanılmıştır. (Noble, 2006). Başka bir ifade ile Destek Vektör Makineleri iki ayrı veri sınıfını temel olarak en iyi şekilde ayırt edebilmek için kullanılmaktadır. Bu amaç ile karar sınırları ya da aşırı düzlemler belirlenmektedir. Doğrusal olmayan veri kümesinde, Destek Vektör Makineleri doğrusal bir hiper düzlem çizememektedir (Demirci, 2019).

2.10. Gradyen Artırma Modeli

Sınıflandırma, gerileme, düğüm artırma ve problemleri için bir makine öğrenme yöntemlerinden birisidir ve zayıf tahminleme modelleri, çoğu zaman karar ağaçları biçiminde bir tahminleme yapmaktadır (Friedman, 2002). Yükseltme ise uygun bir maliyet fonksiyonu üzerinde bir optimizasyon algoritmasıdır. Gradyen artırma, zayıf öğrencileri tek bir güçlü öğrenciye tekrarlayan şekilde birleştirmektedir(Jain, 2015).

3.VERİ MADENCİLİĞİ İLE FİNANS

3.1. Finansal Başarısızlık Tahminleri

İflas/finansal sıkıntı tahmini, 1960'lardan itibaren kritik bir muhasebe ve finans araştırma alanı haline gelmiştir. İflas tarihi/finansal sıkıntı tahmin modelleri iki ana döneme ayrılabilir. 1960'ların sonundan 1980'lerin sonlarına kadar uzanan ilkinde, araştırma, en doğru modelleri oluşturma yöntemleri olarak büyük ölçüde diskriminant analizine ve lojistik regresyona dayanıyordu (Salehi vd. 2016).

Yakut (2012) Borsa İstanbul (BİST) açısından 120 sınai işletmesini finansal başarısızlıklarını veri madenciliği yöntemi ile tahmin etmiştir. Çalışmada C5.0 algoritması, NN ve SVM yöntemleri kullanarak söz konusu firmaların 2-3-4 yıl öncesine ait tahmin sonuçlarını ortaya koymuşlardır. Ayrıca Kaygın vd. (2016) 143 BİST'te faaliyet gösteren sınai işletmesi açısından ve finansal oranlar yardımıyla tahminleme yapmışlardır. Çalışmada veri madenciliği ve LA yöntemlerinden faydalanmıştır. Chen (2011) çalışmasında da finansal başarısızlık tahmini yapmıştır. Benzer araştırmalar Chen ve Du (2009); Torun (2007); Santos (2006); Barnes (1987); Shirata (1998) yazarlar tarafından gerçekleştirilmiştir.

Veri madencilik yönteminde yapay zekaya dayalı makine öğrenimi yöntemlerini kullanarak tek değişkenli modellerden çok değişkenli modellere doğru finansal başarısızlık tahminlerinde bulunmuştur. Wilson ve Sharda, (1994) ve Carlos, (1996) finansal başarısızlığı tahmin etmede veri madencilik yöntemlerin yapay sinir ağları en yaygın yöntemlerden birisidir. Diğer veri madenciliği teknikleri arasında Frydman, Altman ve Kao, (1985) karar ağaçları, Li ve Sun, (2009) vakaya dayalı akıl yürütme, Shin ve Lee, (2002) genetik algoritmalar, Cohen vd. (2012) simülasyon analizi yer almaktadır.

Yukarıda özetlenen çalışmaların önemli bir bölümü firmaları başarısız ve başarılı olarak sınıflayarak hangi finansal oranların finansal başarısızlıkta daha önemli olduğunu ortaya koymuştur. Diğer bir kısmı ise işletmeleri sınıflandırarak finansal başarısızlıktaki evreleri tespit etmiştir.

3.2. Risk Yönetimi

Finans sektöründe veri madenciliği yaygın olarak kullanılan yöntemlerden birisidir. Bankaların, gelişen teknoloji ile birlikte müşterilerini daha iyi tanıma, onlara daha fazla kredi kullanma yönündeki çabaları, verilen kredilerin geri ödenmemesi ya da geç ödenmesi gibi hususlar banka yöneticilerinin dikkatlerini özellikle veri madenciliği ile ilgilinen analistlere yönlendirmiştir. Bankalar açısından temerrüde düşen müşterileri belirlemesi ve bunlara yönelik önlemler alması son derece önemlidir. Veri madenciliğini diğer tahmin tekniklerinden ayıran en önemli özellik daha önceleri yapılan değerlendirmelerden tahminler yapabilme durumudur. Bu da kredi skorlama ya da risk analizine yardımcı olmaktadır (Yıldız, 2019).

Literatürde, finansal riski analiz yapmanın birçok yolu vardır. Farklı araştırmacılar, farklı amaçlar için kurumsal finansal riski kontrol etmenin birçok yolunu araştırmıştır. Literatürde bankalar açısından kredi riskini değerlendirmek amacıyla Lessmann vd. (2015) makine öğrenmesi, Angelini ve diğerleri (2008) yapay sinir ağları ve hibrit modeller ile veri madenciliği tekniklerini kullanmışlardır. Ayrıca kredi riskini tahmin etmede erken uyarı modelleri oluşturmak veya sorunlu kredileri yönetmek veya müşterilerin faiz oranlarındaki ayarlamalara nasıl tepki vereceğini tahmin etmek için veri madenciliği yöntemlerinden de yararlanılmıştır (Chitra ve Subashini, 2013; Kruppa vd., 2013; Bjorkegren ve Grissen, 2018). Bunun yanı sıra kredi riski değerlendirmelerinin güvenilirliğini artırmak ve güvenli olmayan müşterileri tahmin etmek amacıyla da bu yöntem kullanılmıştır Mercaldo vd., 2017; Martinelli vd. 2017).

Yukarıda özetlenen çalışmaların çok büyük bir bölümü veri madenciliği yönteminden yararlanarak kredi risk değerlendirmesi yaparak, geçmişte kredi kullanan bireylerin davranışlarından yola çıkarak modelleme yapmışlardır. Sonucunda ise hangi bireylerin kredi ödemesini nasıl yapacaklarını veya yapmayacaklarını tahmin etmeye çalışmışlardır.

3.3.Portföy Yönetimi

Portföy yönetimi, yatırım riskinin analizi gibi finans sektöründe veri madenciliği teknikleri kullanılmaktadır. Finansal bilgi sistemine söz konusu veri madenciliği tekniklerini dahil eden işletmelerin sağlayacağı faydalardan birisi de yatırım risklerini minimize etmektir. Finansal bilgi sistemine entegre edilen veri madenciliği yöntemleri kalifiye personelin de desteği ile örgütün her bir alanında benimsetilirse, yapılacak olan tahminlerle uygun portföy oluşturulur (Uzar, 2013).

Duan ve Miao (2016) toplam beklenen kayıpların dağılımını tahmin etmek için varlıkların temerrüt olasılıkları arasındaki korelasyonu dikkate alan bir model sunmuştur. Kredi portföyündeki risk profili değerlendirmesinin, ekonomik koşullardaki değişiklikler nedeniyle farklı anlarda büyük ölçüde değişebileceğini göstererek, bu bağımlılıkları modellemek ve sonuçları tahmin etmek için sistem dinamiklerinin kullanılmasını önermektedir.

Nario vd. (2016). kredi portföyü analizinin karmaşıklığı ve tek bir portföyü diğer iş alanlarıyla olan korelasyonlarını dikkate almadan verimli bir şekilde yönetmenin mümkün olamayacağını ifade etmişlerdir. Avrupa'daki 41 finans kurumuyla, onların gelecek beklentilerini, yönetim uygulamalarını ve kullandıkları araçları anlamak için bir araştırma yaptı. CPM'deki artışın özellikle yayılmanın azalması ve maliyetleri düşürme baskısı tarafından motive edildiği sonucuna varmışlardır.

3.4. Manipülasyon Tespiti

Manipülasyon suçu menfaat sağlamak amacıyla kasıtlı olarak finansal yatırımların fiyatları ile oynamaktır ve menkul kıymetlerin fiyat mekanizmasının bozulmasına neden olur (Uzar, 2013). İçerden öğrenenlerin ticareti ve manipülasyon suçlarının incelendiğinde, söz konusu suçların özünde miktar ve fiyat hareketlerinin

değişiklikleri olduğu görülmektedir. Anormal bir duruma sahip işlemlerin ve hareketlerin bu hiyerarşi içinde incelenmesi gereklidir. Bu yüzden içerden öğrenenlerin ve manipülasyon tespiti aslında hesaplama dayanan bir takım işlemler gerektiren matematiksel süreçtir. Bundan dolayı menkul kıymet piyasasının doğal akışı içinde izlemiş olduğu yatırımcılar arasındaki ilişkiyi görülmektedir. Erken uyarı sistemlerinin ve gözetim sistemlerini de, insan gücü ile yapılmış olan hesaplamalar halkası ve teknolojiye dayanarak gerçekleştirilmektedir. Böylelikle şüpheli görülen durumlarda muhtemel ilişkileri tespit ederek analizler yapılmasıdır (Koyuncu, 2007).

3.5. Olağandışı Borsa Getirilerinin Tahmini

İçeriden öğrenenlerin ticareti haksız bir kazanç ve olağandışı getirilere neden olmaktadır. Yapay sinir ağları hisse senedi getiri tahminlerinde kullanılan en sık kullanılan veri madenciliği yöntemidir (Zhang ve Zhou, 2004). Safer (2001) yapmış olduğu çalışmada ABD borsalarının verileri yardımıyla içeriden öğrenenlerin ticaretine ilişkin bir analiz gerçekleştirmiştir. Yapılan analiz sonucunda içeriden öğrenenlerin ticaretine anormal getirilere neden olduğunu ve haksız bir kazanç oluştuğunu tespit etmiştir.

3.6. Kredi Skorum

Finans sektöründe; şirketlerin, müşterilerinin kredi skorlarını analiz edebilmeleri için; müşterilerinin daha önceki eğilimlerini bilmeleri oldukça önemlidir. Bu bağlamda analizler yapılarak müşterilerin borçlarını geri ödeyip / ödemeyecekleri ortaya çıkartılarak şirketlerin zarar durumları minimize edilirken; kar oranlarını artırmak istenmektedir (Yıldız, 2019).

SONUÇ

Bu çalışmada veri madenciliğinin finans sektörüne uygulanmasına yönelik kavramsal bir tarama yapılmıştır. Başka bir ifade ile finansal veriler ile veri madenciliği yöntemi kullanan ampirik çalışmalar özetlenmiştir. Yapılan genel tarama sonucunda veri madenciliğinin bir çok alanda kullanıldığı ve finans alanında yapılan ampirik çalışmalarda yaygın bir şekilde kullanıldığı görülmüştür. Özellikle risk yönetimi, portföy yönetimi, mali başarısızlık tahminleri, manipülasyon tespiti, döviz piyasası, dolandırıcılık tespiti, risk profillerinin belirlenmesi, olağandışı borsa getirilerinin tahmini, yatırımcıların korunması gibi finans alanlarına uygulanmıştır.

Veri madenciliği yöntemleri kullanılarak yapılan finans çalışmaları incelendiğinde söz konusu çalışmaların önemli bir bölümü risk yönetimi ve finansal başarısızlık alanında yapıldığı anlaşılmaktadır. Özellikle ulusal literatürde firmalar; başarılı ve başarısız şeklinde yapılan kümeleme analizlerin yoğunlukta olduğu belirlenmiştir. Çalışmada ayrıca finans alanında veri madenciliği yöntemlerinden en fazla yapay sinir ağları kullanıldığı tespit edilmiştir.

Gelecek çalışmalarda uluslararası literatürde sıkça kullanılan ancak ulusal literatürde fazla kullanılmayan Gradyen Artırma Modeli, Makine Öğrenmesi, Genetik Algoritmalar gibi veri madenciliği yöntemleri kullanılması literatüre katkı sağlayacaktır. Ayrıca söz konusu veri madenciliği yöntemleri ile kredi skorum, anormal getiri tahmini, manipülasyon tespiti ve portföy yönetimi konularında çalışmaların yapılması yatırımcılara, kurumlara ve firmalara yardımcı olacaktır.

KAYNAKÇA

- AMOR, N. B., BENFERHAT S., ELOUEDI Z., (2004), “**Naive Bayes vs Decision Trees in Intrusion Detection Systems**”, Proceedings of the 2004 ACM symposium on Applied computing ACM Symposium on Applied Computing, 420 – 424.
- BJORKEGREN, D. and GRISSIN, D. (2018), “**Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment**”, Available at SSRN 2611775.
- CEREN, K. (2019), “**Veri Madenciliği Yöntemi İle Risklerin Yönetilmesi Ve Sigorta Sektörü Üzerine Bir Uygulama**”, Marmara Üniversitesi, Yüksek Lisans Tezi.
- CHIEN, C.F. ve CHEN, L.F. (2008), “**Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry**”, Expert Systems with Applications, vol. 34, pp.280-290.
- DEMİRCİ, M., (2019), “**Destek Vektör Makineleri ve M5 Karar Ağacı Yöntemleri Kullanılarak Yağış-Akış İlişkisinin Tahmini**”, Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, C.10, S.3, (1113-1124).

- DUAN, J. C. and MIAO, W. (2016), “**Default Correlations and Large-Portfolio Credit Analysis**”, J. Bus. Econ. Stat., vol. 34, no. 4, pp. 536-546.
- ERTİKİN, Kıvanç (2017), “**Hile Denetimi: Kırmızı Bayrakların Tespiti için Kullanılan Proaktif Yaklaşımlar**”, Muhasebe ve Finansman Dergisi, Temmuz 2017, s. 76-77.
- FRIEDMAN, J.H., (2002), “Stochastic gradient boosting. Computational statistics and data analysis”, 38 (4), 367-378.
- GÜL, Emel, Gökay ve ÇAĞATAN Taşkın (2002), “**Genetik Algoritmalar ve Uygulama Alanları**”, Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt: 21, Sayı: 1, 2002, s.141-142.
- GÜLÇE, G. (2010), “**Veri ambarı Ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma**”, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, 2010, 275300.
- GÜLTEKİN, S.U. (2017), “**Veri Madenciliği: Yapay Sinir Ağı Ve Doğrusal Regresyon Yöntemleri İle Fiyat Tahmini**”, Pamukkale Üniversitesi, Yüksek Lisans Tezi.
- HASTIE, T., R TIBSHIRANI, ve J. FRIEDMAN (2001), “**The Elements of Statistical Learning; Data Mining**”, Inference and Prediction, Springer Series in Statistics, USA.
- IRMAK, S. (2009), “**Veri Madenciliği Yöntemleri İle Sağlık Sektörü Veri Tabanlarında Bilgi Keşfi: Tanımlayıcı Ve Kestirimci Model Uygulamaları**”, (Yayınlanmamış Doktora Tezi). Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü.
- JAIN, A., MENON, M.N. ve CHANDRA, S. (2015), “**Sales Forecasting For Retail Chains**”, <https://pdfs.semanticscholar.org/76a2/44f4da1d29170a9f91d381a5e12dc7ad2c0f.pdf>, [Ziyaret tarihi: 5 Temmuz 2021].
- KANTARDZIC, Mehmed (2003), “**Data Mining: Concepts, Models, Methods and Algorithms**, USA”, The Institute of Electrical and Electronics Engineers, Inc, 2003
- KAYGIN, Ceyda YERDELEN, TAZEGÜL Alper ve YAZARKAN Hakan (2016), “**İşletmelerin Finansal Başarılı ve Başarısız Olma Durumlarının Veri Madenciliği ve Lojistik Regresyon Analizi ile Tahmin Edilebilirliği**”, Ege Akademik Bakış, Cilt 16, Sayı1, 147-159.
- KESKİN Benli, Y. (2005), “**Bankalarda Mali Başarısızlığın Öngörülmesi Lojistik Regresyon ve Yapay Sinir Ağı Karşılaştırması**”, Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi , 31-46.
- KOYUNCUGİL, Ali Serhan ve ÖZGÜLBAŞ, Nermin (2009), “**Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları**”, Bilişim Teknolojileri Dergisi, Cilt: 2, Sayı: 2, Mayıs 2009, 21-32, s.28.
- LESSMANN, S., BAESSENS, B., SEOW, H.-V. and THOMAS, L. C. (2015), “**Benchmarking State-of-the-Art Classification Algorithms For Credit Scoring**”, An update of research. European
- LİOR, Rokach ve ODED, Maimon (2008), “**Data Mining with Decision Trees: Thoery and Applications**”, Singapore: World Scientific Publishing Co. Pte. Ltd.,
- MARTINELLI, F., MERCALDO, F., and SARACINO, A. (2017), “**Bridemaid: An Hybrid Tool for Accurate Detection of Android Malware**” In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, pages 899–901. ACM.
- MERCALDO, F., NARDONE, V., and SANTONE, A. (2017), “**Diabetes Mellitus Affected Patients Classification and Diagnosis Through Machine Learning Techniques**”, Procedia computer science, 112:2519–2528.
- NARIO, L. PFİSTER, T. T. POPPENSIEKER, and U. STEGEMANN (2016), “**The Evolving Role of Credit**”, portfolio management.
- NOBLE ,WS (2006), “**What is a Support Vector Machine?**”, Nat Biotechnol, 24(12): 1564-67.
- ÖZKAN, Yalçın (2008), “**Veri Madenciliği Yöntemleri**”, 1. Basım, İstanbul: Papatya Yayıncılık Eğitim.
- SAFER, A. M. (2002), “**The Application of Neural Networks to Predict Abnormal Stock Returns Using Insider Trading Data**”, Applied Stochastic Models in Business and Industry, 18 (4): 381–389.

- SALEHI, Mahdi, MAHMOUD Mousavi, SHIRI Mohammad Bolandraftar Pasikhani , (2016), **“Predicting corporate financial distress using data mining techniques: an application in Tehran Stock Exchange”**, International Journal of Law and Management, Vol.58 Iss 2.
- TOKTAŞ, Peral ve DEMİRHAN Melek Basak (2004), **“Risk Analizinde Veri Madenciliği Uygulamalar,”** YA/EM Yöneylem Arastırması/Endüstri Mühendisligi- XXIV Ulusal Kongresi, 15- 18 Haziran, Adana.
- TORUN, Talip (2007), **“Finansal Başarısızlık Tahmininde Geleneksel İstatistikî Yöntemlerle Yapay Sinir Ağlarının Karşılaştırılması ve Sanayi İşletmeleri Üzerinde Uygulama”**, Erciyes Üniversitesi Sosyal Bilimler Enstitüsü, Kayseri 2007 (Yayınlanmamış Doktora Tezi).
- TUFFERY, Stephane (2008), **“Data Mining and Statistics for Decision Making”**, United Kingdom: John Wiley & Sons Ltd., 2008
- UZAR, C. (2013), **“Finansal Bilgi Sisteminde Veri Madenciliği Teknolojisinin Kullanılması: Borsa İstanbul Üzerine Bir Uygulama”**, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora Tezi.
- YAKUT, Emre (2012), **“Veri Madenciliği Tekniklerinden C5.0 Algoritması ve Destek Vektör Makineleri ile Yapay Sinir Ağlarının Sınıflandırma Başarılarının Karşılaştırılması: İmalat Sektöründe Bir Uygulama”**, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum 2012 (Yayınlanmamış Doktora Tezi).
- ZHANG, D. ve ZHOU, L. (2004), **“Discovering Golden Nuggets: Data Mining in Financial Application”**, Ieee Transactions on Systems, Man, and Cybernetics, 4(4):513-522.