



VERİ MADENCİLİĞİ YAZILIMLARININ POZİTİF TANIMLI MATRİS SORUNUNDA KESTİRİM UYUŞMAZLIKLARI

Ömer Utku ERZENGİN *

İstatistik Bölümü, Fen Edebiyat Fakültesi, Süleyman Demirel Üniversitesi, Isparta, Türkiye

ÖZET

Bu çalışmada örnek veri olarak 2012 yılı Borsa İstanbul'da işlem gören 270 işletmenin bilanço ve gelir tabloları (BvGT) kullanılmıştır. Veri madenciliği yazılımları da (VMY) kullandıkları yöntemleri gerçekleştirmek için başvuru kitabı (reference manual) kullanırlar ve elde edilen sonuçların tutarlı olduğunu ve analizlerin hangi bilimsel dayanağa göre yapıldığını kullanıcılarına beyan ederler. VMY'lerin başvuru kitaplarının kaynakları birbirinden farklı olabilir ve aynı analizler farklı VMY'leri için farklı sonuçlar ortaya çıkarabilir. Çalışmanın amacı aynı veri ve aynı analiz için farklı VMY' den elde edilen sonuçların uyumsuzluğunu göstermektir. Edward I. Altman ve arkadaşları BvGT'ye bağlı finansal oranlara göre mali başarısızlık-başarı Altman Z Skora (AZS) yöntemlerini belirlemiştir. Çalışmada BvGT'deki kalemlere bağlı 21 oran elde edilmiştir ve bu oranlar kendi doğalarından ötürü birbirleriyle istatistiksel açıdan ilişkili içindedirler. Ortaya çıkabilecek çoklu bağlantı sorunu temel bileşenler analizine (TBA) göre çözülmek istenmiştir ve boyut indirilmesi yapılmıştır. TBA'DA işleme alınan kovaryans matrisi (KM) pozitif tanımlı matris (PTM) olmadığı fark edilmiştir. Her VMY'nin PTM olmayan matris sorununu çözmek için kendi nümerik analiz yöntemlerini kullandığı anlaşılmıştır. İşletmelerin mali başarısızlık-başarıları AZS'ye göre kategorik olarak 0-1 şeklinde belirlenmiştir. TBA kullanılarak indirgenmiş sayıdaki değişkenler ikili lojistik regresyon (İLR) analizine sokulmuş ve mali başarı ve başarısızlık kestirilmiştir. Veri madenciliği yazılımlarının İLR'deki ayrıştırma performansı ROC eğrisine göre yapılmıştır. IBM Modeler (SPSS), Statistica, Stata, SAS, R, Weka, Orange yazılımlarında aynı işlemler tekrarlanmıştır. VMY'lerin PTM olmayan kovaryans matrisine bağlı sonuçlarının uyumsuzluğu tartışılmıştır. Çalışmada ücretli VMY'leri analizleri 258 gözlem üzerinden yapmış; hepsi 174 başarısız gözlemden 169 tanesini başarısız ve 84 başarılı gözlemden 75 tanesi başarılı bulmuştur. Özgür yazılımların VMY'leri 270 gözlem üzerinden yapılmış; 182 başarısız gözlemin 173 tanesi R tarafından başarısız, 176 tanesi Weka tarafından başarısız ve 173 tanesi Orange tarafından başarısız bulunmuştur. Özgür yazılımların VMY'leri 88 başarılı gözlemin R tarafından 82 tanesi, Weka tarafından 62 tanesi ve Orange tarafından 75 tanesi başarılı bulunmuştur.

Anahtar Kelimeler: Veri madenciliği, Pozitif tanımlı matris, Finansal başarısızlık, Uyuşmazlık

MISMATCH OF DATA MINING SOFTWARE PREDICTION UNDER POSITIVE DEFINITE MATRIX PROBLEM

ABSTRACT

In this study as a sample data, 270 balance sheet and income tables (BS&IT) of Borsa İstanbul companies were used in 2012. Data mining software (DMS) uses a guidebook (reference manual) to perform its methods and declares their results are consistent. DMS declares to the users that the results obtained are analyzed according to the scientific base. The sources of reference books of the DMS may be different and the same analyzes may produce different outcomes. The aim of the study is to show the mismatch of DMS results for the same analysis and data. Edward I. Altman et al. determined the financial distress procedure called Altman Z score (AZS) according to financial ratios of the balance sheet and income table (BS&IT). In this study according to items of BS&IT, 21 financial ratios were calculated. The financial ratios of BS&IT are related to each other because of BS&IT nature. The multicollinearity problem, that could occur, was wanted to solved with principal component analysis (PCA) and dimension was reduced. Covariance matrix that was analyzed with PCA, was found that it was not a positive definite matrix (PDM). It was understood each DMS use its own numerical analysis methods to solve non PDM problem. Financial distress companies were determined according to Altman Z score (AZS) as categorically labeled with 0-1. Using the PCA, reduced number of variables were analyzed by binary logistic regression (BLR) and the companies financial distress were predicted. BLR parsing performance of the data mining software was based on the to ROC curves. The same procedures were repeated in IBM Modeler (SPSS), Statistica, Stata, SAS, R, Weka, Orange software. The mismatch of DMS results was discussed. The shareware DMS completed analysis from 258 observation; all of them found 169 observations as unsuccessful among 174 unsuccessful observations and all of them found 75 as successful among 84 successful observations. The free DMS completed analysis from 270 observation; R found 173 observations as unsuccessful, Weka found 176 observations as unsuccessful, Orange found 173 observations as unsuccessful among 182 unsuccessful observations and R found 82 observations as successful, Weka found 62 observations as successful, Orange found 75 observations as successful among 84 successful observations.

Keywords: Data mining, Positive definite matrix, Financial distress, Mismatch

*Sorumlu Yazar: omererzengin@sdu.edu.tr

Geliş: .08.052018 Kabul: 14.02.2019

1. GİRİŞ

Ticari (commercial) analiz yazılımları ve ticari olmayan (nonprofit) veya açık kaynak kodlu özgür (GNU) analiz yazılımları belirli başvuru kitabına (reference manual) göre içerdikleri yöntemleri programlama dilleriyle gerçekleştirmek zorundadır. Zorunluluğun sebebi gerçekleştirilecek olan yazılımın bilimsel kaynağının ve araç olarak sunulacak analizlerin tutarlılığının gösterilmesidir. Yazılımlar ortaya çıkarken proje yönetiminin kullandığı (project management) çeşitli disiplinlere göre şekillenirler [1]. Bu disiplinler genel olarak yazılım geliştirme felsefesi (Software development philosophy) olarak adlandırılır [1, 2]. Yazılım geliştirme yöntemleri birbirlerine göre farklılık gösterse de bütün yazılım geliştirme felsefelerinde kullanılacak fonksiyonların başvuru kitabı her zaman ilk sıraya alır [1, 2, 3].

Bütün yazılımlarda olduğu gibi VMY geliştiricileri tarafından doğal kabul edilip raporlanmayan kısımlardan birisi programlama dillerinin kendine ait kütüphaneleridir. Yazılım kütüphanelerinde kod geliştiricilerin işini hızlandıracak hazır yöntem veya fonksiyonlar bulunur. Örneğin C/C++ dilinin matematiksel işlemlerinde kullanılacak `<math.h>` kütüphanesi dışında özgür yazılım GSL ve ücretli yazılım IMSL (Roguewave Software) kütüphanesi de mevcuttur. Programla dili kütüphanesi geliştirmek ayrı bir branştır. Her bir yazılım kütüphanesi aynı zamanda doğal olarak kendi analiz yaklaşımını getirmektedir. Hazır kullanıma sunulan bu kütüphanelerle geliştirilen analiz yazılımlarının dokümanite edilmesi imkansıza yakındır. Ayrıca programlama lisanslarının da sürümleri olduğu gibi hazır yazılım kütüphanelerinin de sürümleri mevcuttur. Bütün yazılımlar açısından başvuru kitabının mükemmel bir biçimde dokümanite edilmemesinin temel sebebi personel, zaman ve paradır.

VMY' ler kullanılıken unutulmaması gereken kısım, kurulum aşamasında kullanıcıya EULA (End User License Agreement: Son Kullanıcı Lisans Anlaşması) onayı sorulmaktadır [4]. Bu kısımda genelde LIMITED WARRANTY (Sınırlı Güvence) veya daha kötüsü NO WARRANTY (Güvence Yoktur) ibareleri bulunmaktadır. VMY'lerden elde edilen sonuçlara bağlı verilen kararlardan araştırmacılar daha fazla sorumluluk sahibi olmaktadır.

Yapılacak çalışmanın amacı, ticari olan ve ticari olmayan, veri madenciliğinde kullanılan yazılımların sonuç uyumsuzluklarını incelemektir. Uyuşmazlıklardaki farklılıkların temel sebebi dokümanite edilemeyen yazılım geliştirme dillerine dayanan kütüphane farklılıklarına dayanmaktadır. Yapılan çalışma, konu aşırı geniş olduğundan, sadece belli bir perspektifte incelenecektir.

Literatür incelenecek olursa analiz yazılımlarının sonuç farklılıkları tartışılmıştır. Bergmann ve arkadaşların 2000 yılında yaptığı çalışmada Wilcoxon-Mann-Whitney testinde bile farklı istatistiksel analiz yazılımlarında farklı sonuçların alındığını belirtmişlerdir [5]. Keeling ve arkadaşları 2007 yılında 9 tane istatistiksel analiz yazılımının güvenilirliği araştırmışlardır. Çalışmada Anova, doğrusal regresyon, doğrusal olmayan regresyon ve temel bileşenler analizinin yazılımlara göre sonuç farklılıkları tartışılmıştır [6]. Odeh ve arkadaşları 2010 yılında çeşitli analiz yazılımlarının sonuçlarının güvenilirliğini doğrusal regresyon üzerinde tartışmışlardır [7]. Bu ve benzer makalelerde bilinen veri setleri veya sorunlu olmayan veriler kullanılmıştır. Ayrıca yazılım güvenliğinin tayini (Software Reliability Assessment) ayrı bir konu olup bunun kendine ait analizleri mevcuttur [8, 9].

Halka açık Kamu Aydınlatma Platformundan elde edilen 2012 yılı Borsa İstanbul'da işlem gören 270 işletmenin BvGT'larına bağlı finansal oranlar elde edilmiştir [10]. Finansal oranlar bulunduktan sonra veri olarak kullanılıp her yazılım için ayrı ayrı, sırasıyla TBA ve İLR çalışmaları yapılmıştır. İLR analizine bağlı denklem katsayıları ve 2x2'lik sınıflandırma tabloları elde edilmiştir. Her bir yazılımdan elde edilen katsayılar ve sınıflandırma tabloları bir birleriyle karşılaştırılmıştır.

Yapılan çalışmada tescilli ürünlerin ilgili sitelerinden deneme sürümleri çekilmiş ve tescilli ürünlerin kurumları bilgilendirilerek çalışmada kullanılmıştır. Ücretli yazılımların firmalarının cevaplarına bağlı

ücretli analiz yazılımlarının ürünleri X_VMY1, X_VMY2, ... şeklinde kullanılmıştır. Kısaltmada X bilinmeyi temsil ederken VMY ise veri madenciliği yazılımını temsil etmektedir.

2. YÖNTEM VE ANALİZ

Yazılımlara tek tek müdahale yerine geçerli olan (default) özellikler kullanılmıştır. Eksik gözlemler için yazılımın varsayılan özelliği kullanılmıştır. Analizlerde yazılımları asgari mutlak bir zemine çekmektense İLR'ye bağlı 2x2'lik sınıflandırma tablosunda azami ayrıştımayı sağlayacak uçlara çekmek tercih edilmiştir.

2.1. Çalışmada Kullanılan Yazılımlar

R projesi açık kaynak kodlu (gnu.org) özgür bir yazılımdır [11]. R projesinin kendisine ait kaynak kitabı ve her paketin kendine ait kullanım kitapçığı vardır [12]. Ayrıca R'ye paket eklemek belirli kurallar dahilinde yapılmakta ve belirli makaleler istenmektedir [12]. R yazılımı TBA ve İLR için “stats” paketini kullanmaktadır [13]. Çalışmada 2018.04.23 tarihli R yazılımının 3.5.0 (Joy in Playing) sürümü kullanılmıştır.

Weka GNU tabanlı özgür yazılım projesidir [14]. Weka yazılımı kaynak kitap olarak Witten ve arkadaşlarının yazdığı “Data Mining: Practical Machine Learning Tools and Techniques” [15] ve 21.12.2017 tarihli “WEKA Manual for Version 3-8-2” [16] referans kitaplarına göre yapılmaktadır.

2018 tarihli deneme sürümleri IBM Modeler 18.1.1. ve IBM SPSS 25 yazılımı IBM firmasının tescilli ürünüdür. Algoritmalar için “IBM SPSS Statistics 25 Algorithms” [17] dokümanını kullanmaktadır. IBM diğer analizlerde kullandığı çeşitli kaynaklar da bulunmaktadır. TBA için “Factor Algorithms” bölümünü kullanmaktadır. İLR için “Logistic Regression Algorithms” bölümünü kullanmaktadır.

2017 tarihli deneme sürümü Stata 15, Stata firmasının tescilli ürünüdür. Stata'nın 15'te TBA için “Statamultivariate Statistics Referencemanual Release 15” [18] ve İLR için “Statabase Referencemanual Release 15” [19] kullanılmıştır.

Statistica 12.5'in deneme sürümü TIBCO firmasının tescilli ürünüdür. Statistica'nın web sitesinde detaylı bir şekilde kullanılan yöntemler açıklanmaktadır [20,21]. Statistica'da yapılan çalışmada İLR için STATISTICA Formula Guide: Logistic Regression [22] ve TBA için “Principal Component Analysis (PCA) and Partial Least Squares (PLS) Technical Notes” [23] kısımları kullanılmıştır.

Orange yazılımı açık kaynak kod komünüsüyle (open source community) ortaklaşa çalışan özgür yazılımdır. Orange yazılımı Slovenya'daki Ljubljana Üniversitesi Bioinformatics Laboratory tarafından yürütülmektedir. Sürüm olarak 2018 tarihli Orange3-3.13 kullanılmıştır. Orange kaynak kitap olarak kendi web sitesindeki ilgili bölümleri kullanmaktadır [24, 25].

SAS Enterprise Miner 14.2 ve SAS 9.4 Acedemik SAS Institute Inc firmasının tescilli yazılımıdır. Ve her iki yazılım aynı prosedürleri kullanıldığından “Base Sas 9.4 Procedures Guide Statistical Procedures”, “Base SAS 9.4 Procedures Guide” başvuru kitapları kullanılmıştır [26, 27].

2.2. Veri

Mali başarısızlık, işletmelerin borçlarını ödeyememesi ve ödeyecek durumunun olmaması durumu işletmenin faaliyetlerini tamamen durdurup iflasına kadar giden bir süreç olup T.C. Borçlar Kanuna [28] ve T.C. İcra ve İflas Kanuna [29] göre hukukta yerini almıştır. İşletmelerin başarısızlığı, finansal yükümlülüklerini yerine getirememekten başlayıp iflasa kadar uzanan durumları içeren geniş bir süreçtir [28, 29].

Finansal olarak veri sağlama ve işletmelerin riskini belirlenmesinin ilk örneklerinden bir tanesi kuruluş tarihi 20 Temmuz 1841 olan Dun ve Bradstreet (D&B) işletmesine göre yapılmıştır [30]. Tanımlanan mali başarısızlık daha sonraları Altman, Beaver ve bir çok araştırmacı tarafından mukayese edilebilecek bir skor haline dönüştürülmüştür [31, 32].

Literatürde yapılan önemli çalışmalardan biriside Altman'ın 1968 yılındaki yayımladığı makaledir [33]. Altman 1968'den sonra 1983 yılında yaptığı çalışmada [31, 33, 34] Z' skorunun 1.23 ile 2.99 arasında kalan aralığını gri (bulanık) bölge olarak tanımlamıştır. Altman Z' skoruna göre 1.23'ün altında kalan işletmeler başarısız ve 2.99 üzerine çıkan işletmeler başarılı olarak kabul edilmiştir.

Altman Z' skoruna göre bulanık bölgenin belli bir kesim noktasından altı başarısız kabul edilecektir. Gri (bulanık) bölgenin kesim noktasının bulunması birkaç bilim dalına bağlı analiz yöntemlerini aynı anda içermektedir. Ekonomik olarak piyasa şartlarına da göre değişen kesim noktasını Erzengin ve arkadaşları 2013 yılında istatistiksel olarak incelenmiştir [35]. Gri bölgenin kesim noktası, bu makale içinde incelenemeyecek kadar geniş bir konudur. Yapılan çalışmadaki bütün yazılımlarda Altman skoru 2.84 altı başarısız olarak kabul edilmiştir.

Altman'ın yenilediği Z' skor modeli:

$$Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5$$

Eşitlikte X_1 : Net İşletme Sermayesi/Toplam Varlıklar, X_2 : Geçmiş Yıl Karları /Toplam Varlıklar, X_3 : Faiz ve Vergiden Önceki Kâr/Toplam Varlıklar, X_4 : Özkaynaklar/ Toplam Borçlar, X_5 : Satış Gelirleri/ Toplam Varlıklar, Z': Kapsamlı endeks şeklindedir.

Altman'ın Z' skorunda olduğu gibi bir işletmenin mali başarısız/başarılı (MBB) durumu BvGT' deki tek bir kalemle belirlenemez. BvGT' deki kalemleri birlikte kullanmakta kalemler arasında mantıksal bağlantılar kurulamadığı için MBB durumunu yeterli bir şekilde açıklanamayabilir. İşletmin ticari durumunu açıklamak için finansal oranların kullanılması gerekmektedir. Yapılan çalışmada kullanılan finansal oranlar

Likidite oranları: X_1 : Cari Oran=Dönen Varlıklar/Kısa Vadeli Yükümlülükler, X_2 : Likidite Oranı=(Dönen Varlıklar-Stoklar-Diğer Dönen Varlıklar)/Kısa Vadeli Yükümlülükler, X_3 : Stokların Net Çalışma Sermayesine Oranı= Stoklar/(Dönen Varlıklar-Kısa Vadeli Yükümlülükler)

Finansal yapı oranları: X_4 : Kısa Vadeli Yükümlülükler/Özkaynaklar, X_5 : (Kısa Vadeli Yükümlülükler+Uzun Vadeli Yükümlülükler)/Özkaynaklar, X_6 : Duran Varlıklar/ Özkaynaklar, X_7 : Kısa Vadeli Yükümlülükler/ Toplam Varlıklar, X_8 : Uzun Vadeli Yükümlülükler/ Toplam Varlıklar, X_9 : (Kısa Vadeli Yükümlülükler+Uzun Vadeli Yükümlülükler)/ Toplam Varlıklar

Faaliyet Oranları: X_{10} : Alacak Devir Hızı= Satış Gelirleri/ Ticari Alacaklar, X_{11} : Ortalama Tahsilat Dönemi= 365/Alacak Devir Hızı, X_{12} : Hazır Değerler Devir Hızı= Satış Gelirleri/ Nakit ve Benzerleri, X_{13} : Stok Devir Hızı= SMM/ Stoklar, X_{14} : Dönen Varlıklar Devir Hızı= Satış Gelirleri/ Dönen Varlıklar, X_{15} : Maddi Duran Varlıklar Devir Hızı= Satış Gelirleri/ Maddi Duran Varlıklar, X_{16} : Özsermaye Dönüş Hızı= Satış Gelirleri/ Özkaynaklar, X_{17} : Aktif Dönüş Hızı= Satış Gelirleri/ Toplam Varlıklar

Karlılık Oranları: X_{18} : Brüt Kar Oranı= Brüt Karı-Zararı/ Satış Gelirleri, X_{19} : Faaliyet Kar Oranı= Faaliyet Karı-Zararı/ Satış Gelirleri, X_{20} : Dönem Kar Oranı= Dönem Karı-Zararı/ Satış Gelirleri, X_{21} : Özsermayenin Amortismanı Oranı= Dönem Karı-Zararı/ Özkaynaklar şeklindedir.

2.3. Yöntem

Çalışmada özgün veri (bilanço oranları) X, TBA ile indirgenmiş veri Y, İLR'deki bağımlı değişken Z ile gösterilmiştir.

2.3.1. Temel Bileşenler Analizi

Örneklem matrisi $X_{n \times p}$ n gözlemlili p değişkene sahiptir. Kovaryans Matrisi (KM) $S = \frac{1}{n-1} (X - \bar{X})(X - \bar{X})^T$ şeklinde gösterilir.

Temel Bileşenler Analizin (TBA) amaçlarından bir tanesinde bağımsız değişkenler X'in, genelleştirilmiş varyansa bağlı (|S|), bilgi kaybı olmadan daha az sayıda faktör değişkenler matrisine ($Y_{n \times k}$) dönüştürmektir. Girdi matrisi bağımsız değişkenler X'in TBA sonrası elde edilen indirgenmiş matrisin (Y) örneklem büyüklüğünün gözlem matrisiyle eşit olup değişken sayının daha az olması beklenir ($k \leq p$). TBA varsayımlarından bir tanesi gözlem matrisindeki değişkenlerin doğrusal ilişkiye sahip olması, indirgenmiş matrisin faktör değişkenlerinin arasında herhangi bir ilişki olmaması istenir [36, 37]. $|S - \lambda I| = 0$ eşitliğinin çözüme bağlı temel bileşenler bulunur. Değişken sayısı (p) kadar öz değer (λ eigen value) mevcuttur ve elde edilen eşitlik p dereceden bir polinomun kökleridir. Eşitlikte "I" birim matristir. TBA, varyans kovaryans matrisi S üzerinde hesaplanır.

$$S = \sum_{j=1}^k \lambda_j \beta_j \beta_j^T$$

şeklinde yazılabilir. Eşitlikte S varyans kovaryans matrisi, λ öz değer ve β ise $1 \times k$ 'lık öz vektördür (eigen vector). Buradan indirgenmiş faktör matrisi $Y = BTX$ şeklinde hesaplanır. Eşitlikte $Y_{n \times k}$ indirgenmiş veri, $X_{n \times p}$ gözlenen veri $B_{k \times k}$ özvektördür. Temel bileşenler bulunduktan sonra indirgenmiş matriste kaç tane indirgenmiş değişken olacağı

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \geq \frac{2}{3} \quad k=1,2,\dots,p$$

yoluyla hesaplanır [37]. İndirgenmiş matristeki değişken sayısı için farklı yöntemlerde mevcuttur.

2.3.2. Pozitif Tanımlı Matris (PTM)

Kovaryans matrisi gibi $p \times p$ boyunda $S = Cov(X)$ matrisi simetrik bir matris olsun. Sıfıra eşit olmayan a vektörünün devriği a^t olsun. $a^t S a > 0, \forall a \neq 0$ koşulu

- 1) S sadece pozitif özdeğerlere sahiptir,
 - 2) S alt matrisleri pozitif determinantlara sahiptir,
 - 3) S matrisi sadece pozitif pivotlar sahiptir
- şartlarında gerçekleşir [38].

Doğrusal cebir teorisi ve çok değişkenli istatistik teorisi açısından KM PTM sorunu değişkenlerin doğrusal kombinasyonlarının 0 varyansa eşit olması gibi sorunlarını göstermektedir. Negatif varyans ise ayrı bir sorun olup bu makalede tartışılmayacaktır. KM PTM olmaması sorunun basit ispatı:

Eğer $Cov(X)$ PTM tanımlı değilse $a \in R^p \setminus \{0\}$ beraber $Cov(X)a = 0$ öyleki

$$0 = a^t Cov(X)a = \sum_{i,j} a_j Cov(X_i, X_i) a_i = Var\left(\sum_i a_i X_i\right)$$

olduğu şeklindedir. Sonuç olarak bazı değişkenlerin doğrusal kombinasyonlarının 0 varyanslı olduğu ortaya çıkar. Eğer kovaryans PTM değilse matrisin tersinin alınmasında bazı sorunlar çıkabilir. PTM olmama sorunları (KM tersinin alınamaması, doğrusal kombinasyonların 0 veya negatif olması) bu makalenin konusu değildir fakat sonuçları tartışılacaktır.

2.3.3. İkili Lojistik Regresyon (İLR)

Lojistik regresyon fonksiyonu

$$f(z) = \frac{1}{1+e^{-z}}$$

0-1 aralığında olup bağımsız değişkenler TBA'ya bağlı k tanedir Z'nin açılımı

$$Z = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \beta_3 Y_3 + \dots + \beta_k Y_k = \beta_0 + \sum_{j=1}^k \beta_j Y_j$$

şeklinde gösterilir. Denklemden bütün Y'ler 0 olduğunda β_0 bağımlı değişken Z'nin beklenen değeridir. β_0 , sabit $\beta_1, \beta_2, \dots, \beta_k$ doğrusal toplamın bilinmeyen katsayılarıdır. Kestirimdeki artık (hata) terimi ε 'la gösterilmiştir [39]. Çeşitli çalışmalarda Altman Z skoruna göre başarısızlık ve başarı kategorik olarak 1-0 şeklinde belirlenmiştir [40].

Katsayıların önemliliği: İLR'de bulunan katsayılar (β_j) test edilirken Wald test istatistiği kullanılır. Wald testi katsayıların en çok olabilirlik kestirimlerinin standart hatasına oranlanmasıyla bulunur [30, 33]. Wald testinin hipotezi $H_0: \beta_j = 0$, $H_A: \beta_j \neq 0$ şeklindedir. Wald test istatistiği yokluk hipotezi altında normal dağılır ve

$$W = \frac{\hat{\beta}_j}{S.H(\hat{\beta}_j)} \sim Z$$

şeklinde gösterilir. Ayrıca β en çok olabilirlik kestirimlerinin karesi ile standart hatasının karesinin oranı Ki-kare dağılır ve

$$W = \frac{\hat{\beta}_j^2}{S.H(\hat{\beta}_j)^2} \sim \chi^2$$

şekline gösterilir. R, SPSS ve SAS yazılımları başvuru kitaplarına göre Ki-kare dağılımı kullanırken Stata normal dağılımı kullanmaktadır.

Hosmer & Lemeshow Uyum İyiliği Testi

Hosmer Lemeshow testi esasında bir Ki-kare testi olup İLR'nin uyum iyiliğini inceleyen yöntemlerden birisidir [30, 33]. İLR'den bağımlı değişkene için kestirilen olasılıklar küçükten büyüğe sıralanır ve genellikle 10 eşit alt gruba bölünür. Her alt grup için beklenen ve gözlenen değerler bulunduktan sonra Ki-Kare Testi uygulanır.

$$\varepsilon_i = \frac{G_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$$

eşitliğinde G_i gözlenen değer, $n_i\pi_i$ beklenen değer, π_i olasılık, n_i her bir alt gruba düşen gözlem sayısıdır. $\hat{C} = \sum_{i=1}^g \varepsilon_i^2$ eşitliğinde g grup sayısıdır. \hat{C} Hosmer-Lemeshow istatistiği olup χ_{g-2}^2 (Ki-Kare) dağılımı gösterir [36, 39].

Alıcı Eğrisi Altında Kalan Alan (ROC)

Sınıflandırma yapılırken belli değer altında kalan bölge başarısız diğer taraf ise başarılı olarak değerlendirilir. Altman Z skorunun belli bir değer altında kalan kısmı başarısız olarak değerlendirilmiştir. Yapılan çalışmada ayrıştırmanın en iyi kesim noktası olup olmadığı alıcı eğrisi (ROC) altında kalan alan (AUC) ile gösterilmiştir [39].

Açıklayıcılık Katsayısı

Bağımlı değişkendeki varyansın ne kadarının bağımsız değişken tarafından belirlendiği açıklayıcılık katsayısıyla (R^2) yapılır.

$$\text{Cox-Snell sanki } R_{CS}^2 = 1 - \left(\frac{LL_0}{LL_1}\right)^{\frac{2}{n}} = 1 - e^{-\frac{2}{n}(LL_1 - LL_0)}$$

$$\text{Nagelkerke } R_N^2 = \frac{1 - \left(\frac{LL_0}{LL_1}\right)^{\frac{2}{n}}}{1 - \left(\frac{LL_0}{LL_1}\right)^{\frac{2}{n}}} = \frac{R_{CS}^2}{1 - e^{-\frac{2}{n}(LL_1 - LL_0)}}$$

Eşitliklerde LL_1 modelin en çok olabilirlik doğal logaritma değeridir ve LL_0 modeldeki bütün değişkenleri içerir. LL_0 da sadece sabitin bulunduğu modelin en çok olabilirlik doğal logaritma değeridir [36, 39].

3. BULGULAR

TBA analizinde değişkenlerin hangi bileşenlere yüklendiği önemli olup, yazılımlar arası farklılıklar göstermektedir. Çalışmaya dahil edilen 7 yazılımın TBA sonuçları çok fazla yer tutacağından bu yazı kapsamında verilmemiştir. Erzengin ve arkadaşlarının yaptığı çalışmalarda [35] TBA analizi sonucu 8 adet faktörün uygun sonuç verdiği ön görülmüştür. 21 adet finansal oranın 8 değişkenlik faktör skorlarına dönüştürülmesiyle varyans bütün yazılımlarda yaklaşık % 73 - %73.5'i açıklanmaktadır.

TBA analizi yapılırken R yazılımı elde edilen matrisin PTM olmadığını belirtmiştir. WEKA ve Orange yazılımları böyle bir uyarıda bulunmamıştır. Ücretli yazılımlardan sadece iki tanesi çıktılarında KM'nin PTM olmadığını belirtmiştir. Diğer ücretli yazılımlarda ise belli bir uyarı sadece uyarı bildirisi (warning code, warning message) olarak verilmiştir. KM'nin PTM olmadığı uyarı mesajının ne anlama geldiği kolaylıkla internette ve yazılım başvuru kitabında bulunamayabilir.

Çalışmamızda BvGT bağlı finansal oranlar hesaplanırken bazı oranlar bazı işletmeler için hesaplanamamıştır. Hesaplanamayan oranların nedeni ise bazı işletmelerin kalemlerinin BvGT'da olmamasıdır. Hesaplanamayan oranlar TBA'da yazılımlara bağlı farklı sonuçlar ortaya koymaktadır. R, SPSS, SAS, Stata, Statistica, Minitab yazılımlarında eksik veri hesaplamalardan çıkarılmaktadır. Orange ve Weka yazılımında ise TBA gerçekleştirirken eksik veriler yerine bir şekilde veri üretme işlemi yapılmıştır. Bu veri üretmenin nasıl yapıldığı başvuru kitaplarından anlaşılamamıştır. VMY'de yapılan analizlerin sonunda Tablo 2 ve Tablo 3 oluşmuş fakat analizde kullanılan gözlem miktarı birbirinden farklı olmuştur.

Tablo 1. Yazılımlara bağlı İLR katsayı sonuçları

	X_VMY1	X_VMY2	X_VMY3	X_VMY4	R	Weka	Orange
β_0	-4.765	2.717	-4.765	-4.765	-3.132	4.206	-2.654
β_1	-12.925	3.280	-2.869	-5.670	-4.943	-4.293	3.056
β_2	4.178	2.413	0.364	0.907	2.818	2.684	1.535
β_3	-0.024	-0.433	3.407	3.648	2.183	3.409	1.679
β_4	-13.601	8.693	-9.587	-15.400	-9.971	-7.368	-4.514
β_5	0.824	5.185	5.361	9.371	8.512	-0.463	0.367
β_6	0.252	0.359	0.187	0.297	-1.295	1.049	0.059
β_7	1.932	-0.177	-0.566	2.244	<u>2.222</u>	-0.357	0.030
β_8	1.714	-1.644	2.687	0.606	0.414	1.551	0.175

Bütün yazılımlarda Hosmer-Lemeshow testi sonucu (R ve ücretli yazılımlarda) 0.99 ve üstünde çıkmıştır. Yazılımlarda Nagelkerke Sanki R^2 ise 0.880 ve Cox-Snell sanki R^2 ise 0.63 olarak bulunmuştur. TBA'ya bağlı faktörlerden elde edilen model tatminkar sonuçlar vermiştir. ROC eğrisi açısından ücretli yazılımlarda (X_VMY1, X_VMY2, X_VMY3, X_VMY4) 0.932, R'da 0.838, Weka' da 0.836, Orange'da 0.884 çıkmıştır.

İLR analizi sonucu çıkan 4 gözlü sınıflandırma tablolarının kestirim hesaplamaları İLR denklemine göre yapılmaktadır. İLR katsayıları aynı zamanda risk ölçüsü birimdir [36, 39]. Tablo 1'de görüleceği gibi VMY'lerde oluşturulan denklemlerin İLR katsayıları bir birindenden taban tabana farklıdır. Tablo 1 sütun bazında bütün olarak incelenirse, pozitif tanımlı matris sorunda VMY'lerin oluşturduğu denklemlere bağlı tahminlerin birbirinden tamamen farklı olduğu açıktır. İLR'deki katsayıların yorumlanması İLR'nin doğası gereği üstsel olarak yapılması gerektiğinden tahmin veya kesitirim için elde edilecek sonuçlar doğrusal sistemlerdekinden çok farklı olacaktır. İLR katsayıları başarılı başarısız (0-1) karşılaştırması üzerinden hesaplandığından sonuçların pozitif ve negatif çıkması bütün yorumları değiştirecektir.

Tablo 1'ye göre İLR analizine bağlı elde edilen sonuçlarda yazılımlara bağlı farklılıklar vardır. X_VMY1, X_VMY2 ve X_VMY3 yazılımları β_0 sabitte aynı sonuçları verirken diğer yazılımlar farklı sonuçlar vermiştir. Orange yazılımı TBA'da herhangi bir rotasyon özelliğine sahip değildir ve Orange yazılımından elde edilen sonuçlar italik olarak verilmiştir. Orange yazılımı katsayıların önemliliklerini vermemektedir. Orange'da β_6 , β_7 ve β_8 için bulunan katsayılar diğer yazılımlara göre sifira en yakın değerlerdir. β_1 açısındansa bulunan katsayı ikinci en büyük pozitif değerdir.

Tablo 1'ye göre X_VMY1 yazılımı β_3 katsayısı 0'a çok yakın bulunmuştur. Diğer yazılımlara göre β_1 katsayısı en küçük negatif değerdir. β_2 katsayısı ise en büyük pozitif değerdir. β_4 katsayısı ikinci en küçük negatif değerdir. β_8 katsayısı ikinci en büyük pozitif değerdir. X_VMY1 sabit terimi X_VMY3 ve X_VMY4 ile aynı bulunmuştur. X_VMY1 yazılımında β_0 , β_1 , β_2 , β_4 , β_7 ve β_8 katsayıları önemlidir. Diğer yazılımlara göre β_2 katsayısı en büyük değer olarak bulunmuştur. Diğer yazılımlara göre β_7 katsayısı üçüncü en büyük değerdir. β_8 katsayısı ikinci en büyük pozitif değerlerdir.

X_VMY2 WEKA'dan sonra en büyük sabit değeri β_0 bulunmuştur. Diğer yazılımlara göre β_4 pozitif olan tek değerdir. X_VMY2 β_0 , β_1 , β_2 , β_4 , β_5 ve β_8 katsayılarını önemli bulunmuştur.

X_VMY3 β_0 , β_1 , β_2 , β_3 , β_4 , β_5 , ve β_8 katsayılarını önemli bulunmuştur. β_8 katsayısı diğer yazılımlara göre bulunan en büyük değerdir. Daha önce belirtildiği gibi β_0 katsayısı X_VMY1 ve X_VMY2 ile aynı bulunmuştur. β_2 katsayısı ikinci en büyük bulunan değerdir.

X_VMY4 ise sabit terim β_0 katsayısını X_VMY1 ve X_VMY3 ile aynı bulunmuştur. Diğer faktörlerin önemlilikleri açısından ise X_VMY4 yazılımı β_0 , β_1 , β_3 , β_4 , β_5 ve β_7 katsayılarını önemli olarak bulunmuştur. β_4 katsayısı en küçük negatif değerdir. β_3 , β_5 ve β_7 ise en büyük pozitif değerler olarak karşımıza çıkmıştır.

R yazılımı $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ katsayılarını önemli bulmuştur. β_7 katsayısı ise 0.05002 olarak bulunmuştur. Buna bağlı olarak β_7 katsayısı altı çizili olarak belirtilmiştir. R yazılımının kestirdiği katsayılar genelde ortada yer almıştır. β_6 katsayısı tek negatif değerdir. β_5 ve β_7 ise ikinci büyük katsayılardır.

Weka yazılımı ise kendisine ait topluluk forumunda [41] geliştiricilerinden olan Mark Hall'ın belirttiği üzere sürekli olarak varimax rotasyonu yapmaktadır. Weka yazılımında katsayıların önemliliklerini belirtmemektedir. Weka yazılımının bulunduğu katsayıların önemlilikleri bilinmediğinden italik olarak gösterilmiştir. Weka'nın sabit için bulunduğu katsayı en büyük pozitif değerdir. β_3 için bulunan katsayılar ikinci en büyük pozitif değerdir. β_7 için bulunan katsayı en küçük ikinci değerdir. β_6 için bulunan katsayı en büyük pozitif değerdir. β_5 için bulunan katsayı tek negatif değerdir.

Tablo 2. Ücretli veri madenciliği yazılımlarının 4 gözlü sınıflandırma tablosu

	X_VMY1, X_VMY2, X_VMY3, X_VMY4	Gözlem		Toplam
		Başarısız (0)	Başarılı (1)	
Kestirim	Başarısız (0)	169	9	178
	Başarılı (1)	5	75	80
	Toplam	174	84	258

Tablo 2'de görüleceği gibi eşitlikleri birbirinden farklı olsada (ücretli) yazılımları aynı ayırıştırma gücüne sahiptir. Bazı finansal oranlar hesaplanamadığından eksik gözlemler oluşmuştur. Başarısız 174 gözlemde 169 (%97.12) tanesi başarısız olarak kestirilmiştir. Başarılı 84 gözlemde de 75 (%89.28) tanesi başarılı olarak kestirilmiştir. Yanlış pozitif oranı (YPO) ise 84 gözlemde 9 (%10.72) tanesidir. Yanlış negatif oranı ise (YNO) 174 gözlemde 5 (%2.88) tanesidir. Doğruluk oranı ise 258 gözlemde 243 (169+75) tanesi olup %94.57' dir.

Tablo 3. Özgür yazılımlar 4 gözlü sınıflandırma tablosu

	Özgür Yazılımlar	Gözlem		Toplam
		Başarısız (0)	Başarılı (1)	
R Kestirim	Başarısız (0)	173	6	179
	Başarılı (1)	9	82	91
Weka Kestirim	Başarısız (0)	176	26	202
	Başarılı (1)	6	62	68
Orange Kestirim	Başarısız (0)	173	13	186
	Başarılı (1)	9	75	84
Toplam*		182*	88*	270*

Tablo 3 en alt satırında görüleceği gibi toplam ek bir yıldızla belirtilmiştir. Başarısız 182 başarılı 88 işletme olduğunu ifade etmektedir. R, Weka ve Orange için ayrı ayrı toplamı ifade etmektedir. Tablo 3'te görüleceği gibi R, Weka ve Orange yazılımları başarılı başarısız kestirimini 270 gözlem üzerinden yapmıştır. TBA analizinde finansal oranlara bağlı bazı gözlemler eksik olduğundan faktör skorlarında eksik olarak ortaya çıkmıştır. R verisi içinde eksik gözlemler NA olarak ifade edilmektedir. Fakat İLR analizinde ise eksik gözlemler içinde tahminde bulunmuştur. R'ın İLR'de de kullanılan komutu glm fonksiyonunda eksik gözlemler için varsayılan (default) özelliği eksik gözlemlere tahmin yapmaktadır. Başarısız 182 gözlemde 173 (%95.06) tanesi başarısız olarak kestirilmiştir. Başarılı 88 gözlemde de 82 (%93.18) tanesi başarılı olarak kestirilmiştir. YPO ise 88 gözlemde 6 (%6.82) tanesidir. YNO ise 182 gözlemde 9 (%4.94) tanesidir. Doğruluk oranı ise 270 gözlemde 255 (173+82) tanesi olup %94.44' dür.

Tablo 3'te görüleceği gibi Weka yazılımı TBA analizinde finansal oranlara bağlı bazı gözlemler eksik olmasına rağmen faktör skorlarında eksik gözlemler için hesap yapılmıştır. Eksik gözlemlere rağmen faktör skorlarında değerler mevcuttur. Weka'nın eksik gözlemler için varsayılan özelliği eksik gözlemlere tahmin yapmaktadır. Başarısız 182 gözlemde 176 (%96.70) tanesi başarısız olarak kestirilmiştir. Başarılı 88 gözlemde de 62 (%70.46) tanesi başarılı olarak kestirilmiştir. YPO ise 88

gözlemden 26 (%29.54) tanesidir. YNO ise 182 gözlemden 6 (%3.30) tanesidir. Doğruluk oranı ise 270 gözlemden 238 (176+62) tanesi olup %88.14' dür.

Tablo 3'te görüleceği gibi Orange yazılımı TBA analizinde finansal oranlara bağlı bazı gözlemler eksik olmasına rağmen faktör skorlarında eksik gözlemler için hesap yapılmıştır. Eksik gözlemlere rağmen faktör skorlarında değerler mevcuttur. Orange yazılımının eksik gözlemler için varsayılan özelliği eksik gözlemlere tahmin yapmaktadır. Başarısız 182 gözlemden 173 (%95.06) tanesi başarısız olarak kestirilmiştir. Başarılı 88 gözlemde de 75 (%85.23) tanesi başarılı olarak kesitirilmiştir. YPO ise 88 gözlemde 13 (%14.77) tanesidir. YNO ise 182 gözlemde 9 (%4.94) tanesidir. Doğruluk oranı ise 270 gözlemde 248 (173+75) tanesi olup %91.85' dir.

4. TARTIŞMA VE SONUÇ

Tablo 1, 2, 3' de de görüldüğü gibi ücretli yazılımlarla ücretsiz yazılımlar arasında ciddi farklılıklar bulunmaktadır. Yazılımlar arasında ciddi anlamda uyumsuzluklar mevcuttur.

Tablo 1'de görüleceği gibi İLR katsayıları VMY'lere göre oldukça birbirinden farklıdır. İLR'de β katsayılarının pozitif ve negatiflik durumuna göre riskin yönünü tamamen değiştirmektedir. Örneğin β_1 katsayısı X_VMY1, X_VMY3, X_VMY4, R ve Orange'da negatifken X_VMY2 ve Weka'da pozitiftir. İLR açısından karar vermede bu ciddi bir sıkıntı çıkarmaktadır. İkinci en büyük dengesizlik ise β_2 katsayısında görülmektedir. β_2 katsayısı için X_VMY1, X_VMY3, X_VMY4, R ve Weka için pozitifken X_VMY2 ve Orange için pozitiftir.

Eksik gözlemler açısından ücretli yazılımlar bir bakıma bedellerini ödemektedir. Eksik gözlemlerin işleme alınması ücretli yazılımlarda (IBM Modeler (SPSS), Statistica, Stata, SAS) kullanıcıya dolaylı olarak bırakılmaktadır. Kullanıcı eğer isterse eksik gözlemlere müdahale edebilmektedir. Varsayılan özellik eksik gözlem varsa o gözlem satırının analize alınmaması durumudur.

Özgür yazılımlarda ise bir kaç durum mevcuttur. R eksik gözlem işlemlerini tamamen kullanıcıya bırakmıştır. Her aşamada kullanıcı ayrı dikkat göstermek zorundadır. TBA için eksik gözleme bağlı işlem yapılmazken İLR'de eksik gözlem için analiz yapılmaktadır. R yazılımdan TBA analizinde eksik gözlemler NA (Not Applicable) olarak geçmektedir. R, İLR analizinde NA gözlemleri için kestirimde bulunmaktadır. Weka ve Orange yazılımlarında ise yazılım varsayılan özellik eksik gözlemler için bir şekilde tahmin yapılmasıdır. Eksik gözlemler (missing value) üzerindeki işlem kaynaklarında açık bir şekilde belirtilmemektedir.

Araştırmacı yazılımları kullanırken yazılımın özelliklerini dikkatlice incelemek zorundadır. Yazılıma güvenmeden önce yazılımın başvuru kitabı dikkatle incelenmeden yorum yapılmamalıdır. Örneğin eksik gözlemlerin nasıl ele alındığı bilinmeden yazılımı kullanmak ciddi sorunlara sebebiye verecektir. Olgunun ilgili olduğu ortamdan toplanan veriler tamamen mükemmel değildir. Eksik gözlen haricinde yazılımlar bir şekilde sorunlarıda çözmek için yöntemler geliştirmiş olabilirler. Örneğin pozitif tanımlı olmayan matrislerin tersinin alınmasında özel yöntemler kullanılabilir. Yazılımlara bağlı farklılıklar araştırmacıyı hatalı kararlara götürebilir. Yazılımlarda elde edilen katsayılar bir birlerinde oldukça farklı olabilir. İLR gibi yarı parametrik yöntemlerin ortaya çıkaracağı sonuçlar yazılımlara göre oldukça büyük farklılıklar ortaya koyacaktır.

Araştırmacı başvuru kitabına göre elindeki verinin hangi yazılım tarafından daha sağlıklı incelendiği belirlemekle yükümlüdür. Yazılımların başvuru kitaplarındaki yakınsama ve yineleme teknikleri incelenmelidir. Yazılımlara bağlı elde edilecek sonuçlar ciddi farklılıklar doğurabilmektedir. Girişte belirtildiği gibi, EULA'ya bağlı veri madenciliği yazılımlarında üretilen bilgiye göre, araştırmacı verilen kararlarda daha fazla sorumluk sahibidir.

TEŞEKKÜR

Verinin kontrolü için yardım eden Zübeyde Karıcı'ya ve Nurzem Üzümcü' ye teşekkür ederim.

KAYNAKÇA

- [1] Chemuturi M, Cagley TM. Mastering Software Project Management: Best Practices, Tools and Techniques. USA: J. Ross Publishing Project Management Professional Series, 2010.
- [2] https://en.wikipedia.org/wiki/List_of_software_development_philosophies
- [3] Birrell ND, Ould MA. A Practical Handbook for Software Development. London, UK: Cambridge University Press, 1988.
- [4] https://www.wikiwand.com/en/End-user_license_agreement
- [5] Bergmann R, Ludbrook J, & Spooren Will PJM. Different Outcomes of the Wilcoxon—Mann—Whitney Test from Different Statistics Packages. The American Statistician; 2012; 54-1:72-77.
- [6] Keeling KB, Pavur RobertJ. A comparative study of the reliability of nine statistical software packages. Computational Statistics & Data Analysis. May 2007; 51-8: 3811-3831.
- [7] Odeh OO, Featherstone AM, Bergtold JS. Reliability of Statistical Software. American Journal of Agricultural Economics. October 2010; 92-5:1472-1489.
- [8] Cristescua MP, Stoicaa EA, Ciovică LV. The Comparison of Software Reliability Assessment Models. Procedia Economics and Finance. 2015; 27: 669 – 675.
- [9] Kapur PK, Pham H, Gupta A, Jha PC. Software reliability assessment with OR applications. London, UK: Springer, 2011
- [10] <https://www.kap.org.tr/>
- [11] <https://www.r-project.org/about.html>
- [12] <https://journal.r-project.org/index.html>
- [13] <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>
- [14] [https://en.wikipedia.org/wiki/Weka_\(machine_learning\)](https://en.wikipedia.org/wiki/Weka_(machine_learning))
- [15] Witten IH, Frank E, Hall MA. Data Mining Practical Machine Learning Tools and Techniques Third Edition. Newyork, USA: Elsevier Science, 2011.
- [16] Bouckaert RR, Frank E, Hall M, Kirkby R, Reuteman P, Seewald A, Scuse A. “WEKA Manual for Version 3-8-1”. New Zealand: GNU General Public License version 3, December 2016.
- [17] IBM. IBM SPSS Statistics 25 Algorithms. IBM Corporation: 2017.
- [18] StataCorp LP. Stata Multivariate Statistics Reference Manual Release 15. Texas: Stata Press Publication, 2017.
- [19] StataCorp LP. Stata Base Reference Manual Release 15. Texas: Stata Press Publication, 2017.

- [20] <http://documentation.statsoft.com>
- [21] <http://www.statsoft.com/Textbook>
- [22] Statsoft. STATISTICA Formula Guide: Logistic Regression. 2013.
- [23] <http://documentation.statsoft.com/STATISTICAHelp.aspx?path=MSPC/PCAandPLSTechnicalDetails>
- [24] <https://docs.orange.birolab.si/3/data-mining-library/#reference>
- [25] <https://docs.orange.birolab.si/3/data-mining-library/#tutorial>
- [26] SAS Institute Inc. Base SAS 9.4 Procedures Guide. Fifth Edition. NC, USA: SAS Institute Inc, September 2016.
- [27] SAS Institute Inc. SAS Enterprise Miner 14.2. Fourth Edition. NC, USA: SAS Institute Inc, November 2016
- [28] <http://www.mevzuat.gov.tr/MevzuatMetin/1.5.6098.pdf>
- [29] <http://www.mevzuat.gov.tr/MevzuatMetin/1.3.2004-20120702.pdf>
- [30] <http://www.sec.gov/Archives/edgar/data/1115222/000111522213000004/dnb-2012x12x31x10xk.htm>
- [31] Altman, EI. Corporate Financial Distress and Bankruptcy. 3rd edition. USA: John Wiley & Sons, Inc., 2006.
- [32] Beaver WH. Financial ratios as predictors of failure. Journal of Accounting Research; 1966: 71-111.
- [33] Altman EI. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance 1968; 23-4: 589-609.
- [34] Uzun E. İşletmelerde Finansal Başarısızlığın Teorik Olarak İrdelenmesi. Mufad Journal of Accounting and Finance Temmuz 2005; 27: 158-168.
- [35] Erzengin, ÖU, Üzümcü N. Specifying Fuzzy Region of Altman Z Score on Estimating Financial Distress of Borsa İstanbul Companies. In: XVIth International Symposium on Econometrics, Operations Research And Statistics; 07-12 December 2015; Edirne: sf. 484-485.
- [36] Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. 4. Baskı. Ankara: Detay Yayıncılık, 2013.
- [37] Hair JF. Multivariate data analysis. USA: Pearson Education, 2007.
- [38] Rosen KH. Handbook of Linear Algebra. Newyork: Chapman & Hall/CRC, 2007.
- [39] Hosmer DW, Lemeshow S. Introduction to the logistic regression model. Applied Logistic Regression. Second Edition. Newyork: A Wiley-Interscience publication, 2000.
- [40] Suntraruk P. A review of statistical methods in the financial distress literature. AU Journal of Management, 2010; 8:2.