

PAPER DETAILS

TITLE: Mali Tablo Denetiminde Ön Analitik İnceleme Teknigi Olarak Veri Madenciliginin Kullanimi:
Borsa Istanbul Uygulaması

AUTHORS: Hilmi KIRLIOGLU,İsmail Fatih CEYHAN

PAGES: 13-36

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/39499>

Mali Tablo Denetiminde Ön Analitik İnceleme Tekniği Olarak Veri Madenciliğinin Kullanımı: Borsa İstanbul Uygulaması *

Hilmi KIRLIOĞLU*, İsmail Fatih CEYHAN**

ÖZET

Kaliteli bir denetim gerçekleştirmek adına denetçi mümkün olduğunca denetleyeceği firma hakkında varacağı yargıya kısa bir süre içinde ulaşmalıdır. Bu yargıya varmak için denetçi, firmanın mali tablolarından yararlanmaktadır. Uluslararası Finansal Raporlama Standartlarına uygun olarak düzenlenmiş bu mali tablolar, firma hakkında bilgi sağlayan önemli göstergeler olup, denetim programı aşamalarından olan analitik inceleme faaliyeti sırasında denetçinin yararlanacağı araçlardan biridir.

Çalışmamızda Borsa İstanbul'a kayıtlı 40 şirketin mali tablo verileri kullanılmış, bu şirketlerin veri madenciliği algoritmalarıyla finansal olarak başarılı ya da başarısız olarak sınıflandırılarak ön analitik inceleme faaliyetinin kısa sürede tamamlanması hedeflenmiştir. Sonuç olarak, k-en yakın komşu algoritması ve 10 kat çapraz doğrulama tekniği ile %95 gibi yüksek bir oranda doğru finansal sınıflama tahmini elde edilmiştir.

Anahtar Kelimeler: Mali Tablo Denetimi, Veri Madenciliği, Sınıflama Analizi, Weka

ABSTRACT

Auditors will have to be reached the judgment about the firm that will be supervised to perform a good quality audit as much as possible in a short time. To reach this judgment, the auditors benefit from the company's financial statements. Financial statements arranged in accordance with International Financial Reporting Standards are important indicators that provide information about companies. These statements is one of the instruments the auditor can use it during the analytical review activities in stage of the audit program.

In our study, we aimed to classify 40 companies listed in Borsa İstanbul as financially successful or financially unsuccessful with data mining algorithms using datas of the financial statement of these companies. As a result, correct classification estimation was achieved at a rate as high as 95% with the k-nearest neighbor algorithm and 10-fold cross-validation technique.

Keywords: Financial Statement Audit, Data Mining, Classification Analysis, Weka

* Bu çalışma Sakarya Üniversitesi S.B.E. Muhasebe-Finansman A.B.D.'da yapılmakta olan "Muhasebe Denetiminde Kalite ve Veri Madenciliği- Bir Uygulama" adlı doktora tezinden üretilmiştir.

* Prof. Dr., Sakarya Üniversitesi, İşletme Fakültesi, İşletme Bölümü, Muhasebe ve Finansman A.B.D., hilmik@sakarya.edu.tr

1. Giriş

Her gün sınırsız miktarda verinin toplandığı bir dünyada yaşamakta olduğumuz için bu verilerin analiz edilmesi önemli bir ihtiyaç haline gelmiştir. Veri zengini fakat bilgi fakiri olduğumuz bu ortamda veri analizi yapılabilmesi için güçlü veri analizi araçlarına ihtiyaç duyulmaktadır. Terabaytlarca veya petabaytlarca veri, internet aracılığıyla bilgisayar ağlarına ve çeşitli veri depolama cihazlarına her gün aktarılmakta ve toplanan verilerin hacmi de hızla artmaktadır. Çok yüksek miktardaki veriden otomatik olarak anlamlı bilgi ortaya çıkaracak ve tecrübeye dönüştürecek güçlü ve becerikli araçlara ihtiyaç duyulmaktadır. Bu nedenle veri madenciliği, veri çağından bilgi çağına geçişte önemli açılımlar sağlamaktadır. (Han vd., 2012:1-2).

Çalışmamızın amacı bağımsız denetçinin, denetim aşamalarından biri olan ön analitik inceleme sürecinde işletmelerin mali tablolarından elde edilen veriler kullanılarak veri madenciliği modellerinin uygulanmasında, “Naive Bayes” ve “K-En Yakın Komşu” algoritmalarının hangisinin daha iyi sınıflandırma sonucu verdiği ve veri madenciliği ile şirketin finansal performans tahminini hızlandırmanın mümkün olup olmadığı tespit edilmeye çalışılmıştır.

Çalışmanın birinci bölümünde, mali tablo oranları ile veri madenciliği uygulamalarının birlikte kullanımının yer aldığı çalışmalarla ilgili literatür taraması yapılmıştır. İkinci bölümde ise çalışmada yer alan veri madenciliği ve sınıflandırma kavramlarının tanımı, önemi, kullanım amaçları, özellikleri ve türleri ele alınmıştır. Çalışmanın uygulama aşaması olan üçüncü bölümde, Borsa İstanbul’a kayıtlı 40 firmanın mali tablo verileri kullanılarak farklı bir işletmenin mevcut pazar durumu hakkında tahmin yapılmıştır. Çalışmanın dördüncü bölümü olan sonuç ve değerlendirme kısmında ise ulaşılan bulguların analizi ve değerlendirilmesi yapılarak önerilerde bulunulmuştur.

2. Literatür

Mali tablo oranları ile veri madenciliği uygulamalarının birlikte kullanımının yer aldığı çalışmalarla ilgili yapılan taramada uygulanan yöntemlerle ilgili şu bilgilere ulaşılmıştır. Fanning, Cogger ve Srivastava (1995) yaptıkları çalışmada yönetim kaynaklı hile tespiti için yapay sinir ağları kullandıkları bir model geliştirmişlerdir. Green ve Choi (1997) çalışmalarında kurumsal hilenin tespitinde içsel mali verileri kullanarak bir sinir ağı hile

** Arş. Gör., Bartın Üniversitesi, İ.İ.B.F., İşletme Bölümü., Muhasebe ve Finansman A.B.D., ismailc@bartin.edu.tr

sınıflandırma modeli geliştirmişlerdir. Fanning ve Cogger (1998) ile Cerullo ve Cerullo (1999) araştırmalarında kurumsal hilenin tespiti için geliştirdikleri modelde sinir ağlarını kullanmışlardır. Koskivaara (2000) çalışmasında mali hesapların denetimi için sinir ağı tahmin yeteneği üzerinde çeşitli ön-işleme modellerinin etkisini incelemiştir. Feroz ve diğerleri (2000) yaptıkları araştırmada olası dolandırıcıları ve muhasebe hilelerini tahmin etmeye çalışmışlardır.

Lin, Hwang ve Becker (2003) çalışmalarında lojistik model ve bulanık sinir ağlarını kullanarak kurumsal dolandırıcılık tespiti için bir bütünlük bulanık sinir ağı modeli programını değerlendirmişlerdir. Kotsiantis ve diğerleri (2006) yaptıkları araştırmada Karar Ağaçları, Yapay Sinir Ağları, Bayes İnanç Ağı, K-En Yakın Komşu tekniklerini kullandıkları hibrid bir karar destek sistemi ile hileli finansal tabloları tespit etmede istif değışken yöntemi kullanmışlardır. Kirkos ve diğerleri (2007) çalışmalarında hileli finansal tabloların tespitinde sinir ağları, karar ağaçları ve Bayes İnanç Ağlarının etkinliğini araştırmış ve ilişkili faktörleri tespit etmişlerdir. Liou (2008) araştırmasında hileli finansal raporlamayı tespit için algılama / tahmin modelleri kurmuştur.

Kapardis ve diğerleri (2010) çalışmalarında hile tespiti aracı olarak yapay sinir ağlarının kullanılmasını test etmişlerdir. Jans ve diğerleri (2010) çalışmalarında tanımlayıcı veri madenciliği yaklaşımı ve çok değışkenli gözlemlenemeyen sınıf kümeleme tekniğinin kullanılmasının, şirket içi dolandırıcılık riskini azaltmada daha etkili olduğunu ortaya çıkarmışlardır. Ravisankar ve diğerleri (2011) ve Perols (2011) çalışmalarında finansal tablo hilesine başvuran şirketleri belirlemek için Sinir Ağı ve Destek Vektör Makinelerini kullanmışlardır. Zhou ve Kapoor (2011) yaptıkları çalışmada kendi uyarladıkları tepki yüzeyi modeline dayanan bir çerçeve araştırma ile finansal tablo hilesini tespit etmeyi hedeflemişlerdir. Glancy ve Yadav (2011) yönetimin kararları ve 10-K başvuru analizi kullanarak hileli raporlar ile hilesiz raporları ayırt etmişlerdir. Pandey ve Pal (2011) çalışmalarında kurumun satış oranının önemli bir seviyeye düşüşünü azaltmak ve kurumun performans düzeyini artırmak için Bayesian sınıflandırma modelini uygulamışlardır. Ngai ve diğerleri (2011) çalışmalarında, finansal dolandırıcılık tespiti için veri madenciliği tekniklerinin uygulanmasına ilişkin literatürü gözden geçirmişlerdir.

Kuzey ve diğerleri (2014) çalışmalarında, karar ağaçları ve yapay sinir ağları kullanarak firma değeri üzerindeki çok ulusluluk etkisini ve diğer on dört finansal göstergiyi 1997-2011 dönemi için araştırmışlardır. Kou ve diğerleri (2014) çalışmalarında, finansal risk analizi alanında kümeleme algoritmalarının kalitesini artırmak için Çok Ölçütlü Karar Verme

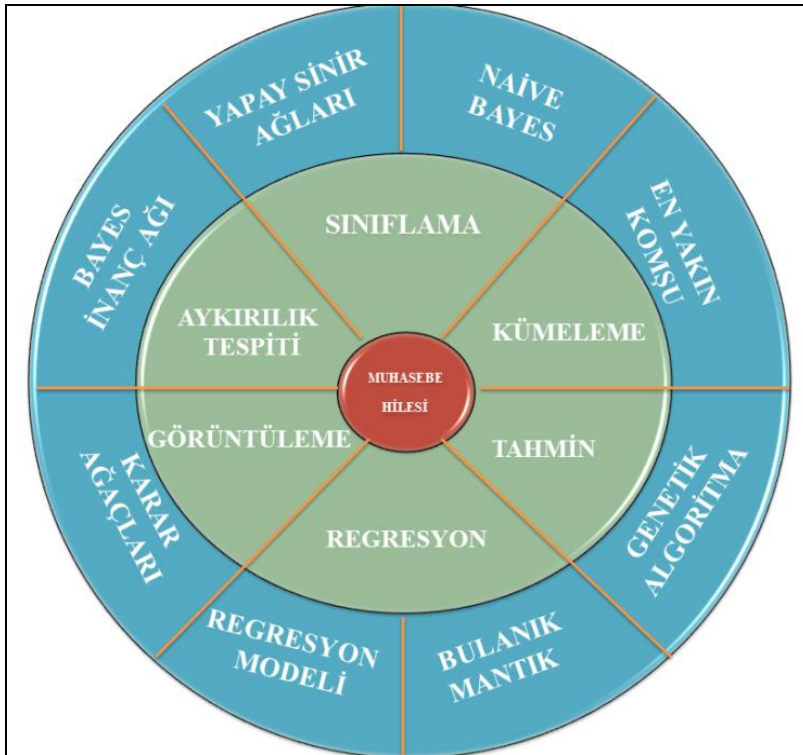
yöntemlerinin kullanıldığı yeni bir değerlendirme yaklaşımı önermişlerdir. Huang ve diğerleri (2014) yaptıkları araştırmada topolojik kalıpları ortaya çıkarma ve Büyüyen Hiyerarşik Kendinden Düzenlenen Haritalar yaklaşımı ile etkili finansal hile algılama ve özellik çıkarma üzerine çalışmışlardır.

Türkiye’ de yapılan çalışmalarda; Terzi (2012) yaptığı çalışmada muhasebe denetiminde kullanılan veri madenciliği yöntemlerini ve nasıl kullanılacağını incelemiştir. Albayrak ve Yılmaz (2009) çalışmalarında, İMKB’ de faaliyet gösteren 173 işletmenin yıllık finansal göstergelerine karar ağaçları tekniği uygulayarak firmaları ayıran en önemli değişkenleri saptamışlardır. Boran (2012) yaptığı araştırmada İMKB’ye tabi şirketleri veri madenciliği yöntemleri aracılığıyla finansal olarak değerlendirmiş ve uygun kümelere ayırmıştır. Akyel (2011) çalışmasında potansiyel hilebazın tespitinde k en yakın komşuluk algoritmasından yararlanılabileceğine ilişkin örneklerle yer vermiştir. Yücel (2013) çalışmasında SAS No.99’da belirtilen 42 kırmızı bayrağın hile ve yolsuzlukların tespitindeki etkinliğinin Türkiye’deki bağımsız denetçilerin bakış açılarıyla değerlendirmiştir.

Finansal tablolarda hile tespitinde veri madenciliği uygulamasının kavramsal çerçevesi Şekil 1’de sunulmuştur.

Şekil 1

Finansal Tablolarda Hile Tespitinde Veri Madenciliği Uygulamasının Kavramsal Çerçevesi



Kaynak: Sharma, 2012: 37-47

3. Veri Madenciliği Ve Analitik İnceleme

Bu bölümde veri madenciliği ve analitik inceleme hakkında özet bilgi verilmiştir.

3.1. Veri Madenciliği

Veri madenciliği, hangi potansiyel bilginin veriler arasında olduğunu bulmak ve problemi çözmek için yöntemlerin nasıl uygulanacağını keşfetmek için veriler ile uğraşmanın bir diğer adıdır. Veri madenciliği için kullanılan yazılım araçlarının birçoğu istatistik, yapay zeka ve makine öğrenmesi gibi konulardan faydalanılarak oluşturulmuştur. Bu araçlar sadece veri setinin bir parçasındaki veriler ile diğer parçasındaki verileri karşılaştırarak bir ilişki ortaya çıkarmaya çalışmaktadır. Buradan bilgiyi kişinin çıkarması gerekmektedir. Veri madenciliği projelerinde veri madencisinin zamanının çoğu veri seti hazırlama işinde geçer (Shmueli vd., 2010:275).

Veri madenciliği ve bilgi keşfi alanındaki gelişmeler çeşitli faktörlerden dolayı önemlidir (Larose, 2005:4);

- Toplanan verilerin hızla artması,
- Verinin veri ambarlarında toplanması, böylelikle tüm kuruluşların güvenilir bir veri tabanına ulaşması,
- İnternette dolaşma ve intranetler aracılığıyla veriye ulaşımın artması,
- Global ekonomide piyasa paylaşımının artması için rekabet baskısı,
- Satışa hazır ticari veri madenciliği yazılımlarının geliştirilmesi,
- Hesaplama gücü ve depolama kapasitesinde hızla artan büyüme.

Veri madenciliği uygulamalarına örnek olarak, yapılan bir çalışmada aynı cinsiyet, yaş ve tedavi geçmişine sahip, belirli tedavilere yanıt veren son 100 hastasının tanımlanması amacıyla, veri madenciliği teknikleri kullanıldığından bahsetmektedir. (Swartz, 2004:8). Ayrıca bankaların kredi kartı müşterisi kazanmasında (Weng vd., 2007:46–57), sigorta ve telekomünikasyon firmalarında hile tespiti (Rejesus vd., 2004:24–32), üretici firmalarda kalite kontrolünde (Cunha, 2006:4041–4054) ve daha birçok uygulamada veri madenciliği sıkça kullanılmaktadır.

3.1.1. Veri Madenciliği Uygulamasında Sınıflandırma

Sınıflandırma, veri madenciliğini de içeren çeşitli alanlarda yaygın olarak kullanılan bir tekniktir. Sınıflandırma, gözetimli öğrenme yöntemleri ile büyük veri setlerini öznitelik değerleri kullanılarak daha öncesinde tanımlanmış olan sınıflardan birine dahil etmektir (Mastrogiannis, 2009: 2829-2839).

Diğer bir deyişle sınıflandırma, hangi sınıfa ait olduğu bilinmeyen bir kayıt için bir sınıf belirleme sürecidir. Sınıflandırma yapılırken öğrenme algoritmalarından biri kullanılır. Eğitim işlemi tüm veriler kullanılarak değil, bu veriler içinden seçilen bir örnek veri seti üzerinde gerçekleştirilir. Eğitimin amacı, bir sınıflandırma modelinin oluşturulmasıdır. (Özkan, 2013:34).

Sınıflandırma işlemi dört temel bileşenden oluşmaktadır (Gorunescu, 2011:15):

- Sınıf: Sınıflandırılan objeye verilen etikettir.
- Öngörücüler: sınıflandırılacak verinin karakteristikleri. Bu karakteristiklere kan basıncı, satın alma sıklığı, medeni durumu, rüzgâr ve hız yönü, mevsim vb. örnekler verilebilir.
- Öğrenme veri seti: yukarıda belirtilen iki bileşen için veriler içeren veri setidir. Uygun sınıf belirlemek için öğrenme modeli oluşturmada kullanılır.
- Test veri seti: sınıflandırıcı model tarafından sınıflandırılacak yeni veriler içerir ve böylece sınıflandırıcı doğruluğu (model performansı) elde edilebilir.

3.1.1.1. Sınıflandırma Algoritmaları

Araştırmada, sınıflandırma algoritmalarından K-En Yakın Komşu Metodu ve Naive Bayes Metodu kullanılmıştır.

3.1.1.1.1. K-En Yakın Komşu Metodu

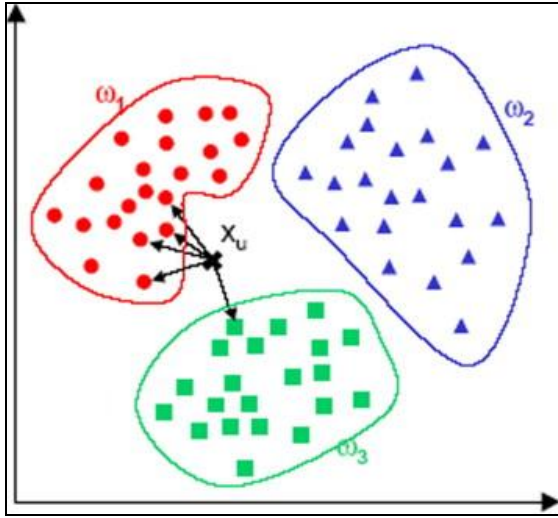
K-en yakın komşu metodu eğitici öğrenme tabanlı en eski ve en basit metotlardan biridir (Aci vd., 2011:12585-12591). K-en yakın komşu sınıflandırması, eğitim veri setindeki yeni bir vektörün kategorisini belirlemede en yakın uzaklığı kullanan bir algoritmadır. Eğitim aşamasında, özellik uzayı bölgelere ayrılır ve eğitim veri noktaları içeriklerinin benzerliklerine göre bu bölgeler içinde gösterilirler. Etiketlenmemiş giriş verisi noktaları, bu noktaların kategorilere olan uzaklığı hesaplanarak kategorilendirilir. K-en yakın komşu sınıflandırma algoritması sadece küçük bir eğitim veri setine ihtiyaç duyar. Bu özellik bu algoritmanın daha basit olmasını ve diğer algoritmalarından daha üstün olmasını sağlar. Şekil2’de w1, w2 ve w3 isimli üç kategoriden oluşan bir 5-en yakın sınıflandırıcıya örnek

gösterilmektedir. X_u ise test işlemi aşamasında sınıflandırılacak olan yeni etiketsiz giriş verisidir (Wan vd., 2012:11880-11888).

K-en yakın komşu metodunun en önemli avantajı çoklu kategorize edilmiş veri noktaları ile sınıflandırma işleminde başarılı çalışabilmesidir. En önemli dezavantajı ise yeni verinin sınıfını belirlemede uzaklık hesabı yaptığından dolayı eğitim setinin boyutu büyüdüğünde hesaplama zorlaşacaktır. Ayrıca sınıflandırmada harcanan zaman ile hafıza ve işlemci kullanımı artacaktır (Wan vd., 2012:11880-11888).

Şekil 2

Üç Boyutlu Bir 5-En Yakın Sınıflandırıcının Öznitelik Uzayı



Kaynak: Osuna, 2002: Introduction to pattern recognition, Wright State University, Dayton, Ohio, USA

K-en yakın komşu sınıflandırıcıda en sık kullanılan uzaklık fonksiyonu Öklid uzaklık formülüdür ve yeni bir etiketsiz veri noktası ile eğitim veri noktaları arasındaki uzaklığın hesaplanması için kullanılır. K-en yakın komşu algoritmasının sınıflandırma aşamasında temel adım yeni giriş veri noktasının en yakın komşularını belirlemek için uzaklığın hesaplanmasıdır (Wan vd., 2012:11880-11888).

3.1.1.1.2. Naive Bayes Metodu

Veri madenciliği uygulamalarında kullanılan Naive Bayes algoritması sınıflandırma problemlerinde sıkça kullanılan etkili ve basit bir yöntemdir (Soria vd., 2011:775–784). Hızlı öğreticili sınıflandırma tekniklerinden biri olan Naive Bayes algoritması, büyük ölçekli, karışık ve tamamlanmamış veri setlerinde sınıflandırma yapmak için uygun bir yöntemdir.

Naive Bayes sınıflandırma, belirlenen sınıflardaki değişkenlerin bağımsız olduğunu varsaymaktadır. Bu sınıflandırıcı yüksek miktardaki değişkenler ve büyük veri setleriyle ilgilenmektedir (Soria vd., 2011:775–784).

3.1.1.2. Eğitim-Test Yöntemleri

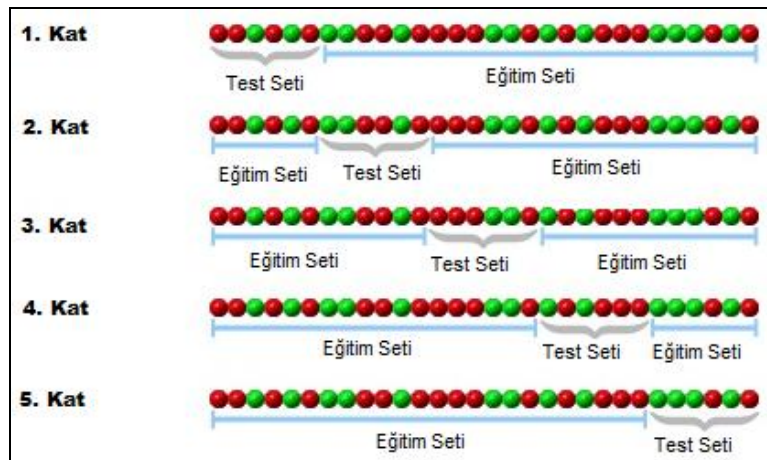
Araştırmada kullanılan eğitim-test yöntemlerinden, “K-Kat Çapraz Doğrulama Yöntemi” ve “Veri Setinin Eğitim Seti - Test Seti Olarak Ayrılması Yöntemi” hakkında özet bilgi aşağıda verilmiştir.

3.1.1.2.1. K-Kat Çapraz Doğrulama Yöntemi

Bir sınıflandırıcının performansı genellikle tahmin hatası ile ölçülmektedir. Çoğu problemlerde, hata tam olarak hesaplanamamaktadır ve bu tahmin edilmelidir. Dolayısıyla uygun bir hata tahmin edici seçilmesi önemlidir. K-kat çapraz doğrulamada veri seti k adet eşit parçaya ayrılmaktadır, bir sınıflandırıcı k-1 parça ile eğitilmektedir ve geri kalan parçada sınıflandırıcı test edilerek bir hata değeri hesaplanmaktadır. Şekil 3’de örnek olarak 5 kat çapraz doğrulama modeli gösterilmiştir. K-kat çapraz doğrulamada hata tahmini, her parçadaki elde edilen hatanın ortalama değeri olarak elde edilmektedir. Dolayısıyla k-kat çapraz doğrulama hata tahmin edici iki etkene bağlıdır. Bunlardan biri eğitim seti, diğeri bölünen parçalardır (Rodriguez vd., 2010:569-575).

Şekil 3

5 kat çapraz doğrulama modeli



Kaynak: http://ecs.victoria.ac.nz/Courses/COMP307_2014T1/Lect13-team1

Uygun modellerin performansını hesaplamak için kestirim hatasının tahmini gerekmektedir. Çapraz doğrulama kestirim hatasının tahmininde çoğunlukla kullanılmaktadır. Birçok araştırmada birini dışarıda bırakan çapraz doğrulama çalışılmasına rağmen, k-kat çapraz doğrulama, k değeri küçük olduğunda gerçek veri analizinde bir problem haline gelebilmektedir. Çapraz doğrulama model seçimi için kullanılmaktadır fakat daha iyi bir kestirim hatası tahmini, daha iyi bir model seçim kriterine götürmek zorunda değildir (Fushiki, 2011:137–146).

3.1.1.2.2. Veri Setinin Eğitim Seti - Test Seti Olarak Ayrılması Yöntemi

Belirli bir veri setinin, seçilen bir sınıflandırıcı ile sınıflandırılması sırasında kullanılabilecek diğer bir yöntem ise, veri setinin eğitim seti ve test setine bölünmesidir. Veri setinin bölünmesinde %66 eğitim seti ve %34 test seti gibi oranlarda setler oluşturulabilmektedir. Eğitim seti, sistemin eğitiminde, test seti ise eğitilen bu sistemin test setindeki değerlerle test edilerek sınıflandırma başarısının bulunmasında kullanılmaktadır. Test setindeki verilerin doğruluk oranı ile seçilen veri seti üzerinde sınıflandırıcının başarısı ölçülebilmektedir [Goyal, 2011:1-4, Almonacid, 2013:165-172, Togun, 2010:349–355).

3.2. Analitik İnceleme

Denetimin planlama, yürütme ve tamamlama aşamalarında bir risk değerlendirme yöntemi olarak kullanılan analitik inceleme tekniği, denetlenen işletmenin faaliyet koşullarının ve çevresiyle olan ilişkilerinin kavranmasına imkân tanımaktadır. Ayrıca kayıtlı değerler ile beklenen değerler arasında farklılıklar olup olmadığının araştırılmasında da uygulanan bir yöntemdir.

Analitik inceleme; cari dönem finansal bilgilerinin; önceki dönemlere ait karşılaştırılabilir nitelikteki bilgilerle, bütçe ve tahminler gibi işletmenin beklediği sonuçlarla ve sektörel göstergelerle basit olarak ve/veya ileri derecede istatistiki teknikler kullanılmak suretiyle karşılaştırılması şeklinde açıklanabilir.

Analitik inceleme prosedürlerinin tanımı AICPA (American Institute of Certified Public Accountant-Amerikan Sertifikalı Kamu Muhasebecileri Kurumu) tarafından 1988 yılında yayınlanan SAS No:56 “Analitik Prosedürler” yönetmeliğinde şu şekilde yapılmıştır; “çeşitli finansal ve finansal olmayan veriler ile işletmenin kayıtları arasındaki anlamlı ilişkilerin incelenmesiyle elde edilen bilgilerin, denetçi tarafından geliştirilen beklentilere

uyup uymadığına bakılması işlemlerine analitik inceleme prosedürleri adı verilmektedir” (SAS No:56, par. 2).

IFAC (International Federation of Accountants- Uluslararası Muhasebeciler Federasyonu) tarafından yayınlanan ISA 520 Analitik Prosedürler ’de ise; “finansal ve finansal olmayan veriler arasındaki ilişkilerin makullüğünü araştıran çalışmalardan elde edilen finansal verilerin değerlendirilmesi ve olağandışı dalgalanmaların, tutarsız ilişkilerin ve tahmin edilen tutarlardan sapmaların karşılaştırılmasıdır” şeklinde bir tanımlama yapılmıştır (ISA 520, par. 4).

BDKK (Bankacılık Düzenleme ve Denetleme Kurumu) tarafından yayınlanan Bankalarda Bağımsız Denetim Gerçekleştirecek Kuruluşların Yetkilendirilmesi ve Faaliyetleri Hakkında Yönetmelik’de, analitik inceleme “finansal tablolar ile banka hakkında edinilen bilgilerin tutarlı olup olmadığının değerlendirilmesinde, finansal tablolarda yer alan bilgi ve açıklamaların karşılaştırılması ve aralarındaki ilişkilerin araştırılarak gözden geçirilmesidir” şeklinde tanımlanmıştır.

SPK (Sermaye Piyasası Kurulu) tarafından yayınlanan Seri X, No:22 sayılı Sermaye Piyasasında Bağımsız Denetim Standartları Hakkında Tebliğ’de de analitik prosedürler; “finansal ve finansal olmayan bilgiler arasındaki inandırıcı ilişkilere dayalı çalışmaların ortaya çıkardığı finansal bilgilerin değerlendirilmesi” olarak tanımlanmıştır (Seri X, No:22 sayılı Tebliğ– Onsekizinci Kısım m.2).

Denetimin planlama, test etme ve nihai gözden geçirme aşamalarında analitik inceleme prosedürlerinin kullanılmasında en uygun denetim prosedürünün seçilmesi en önemli aşamalardan biridir. Literatürde ve denetim standartlarında aşağıda açıklanan beş çeşit analitik inceleme prosedüründen bahsedilmektedir (Arens vd., 2006:209);

- İşletme verileri ile önceki yıllara ait benzer verilerin karşılaştırılması,
- İşletme verileri ile işletme tarafından belirlenmiş bütçe ve tahminlerin karşılaştırılması,
- İşletme verileri ile denetçi tarafından belirlenmiş beklenen sonuçların karşılaştırılması,
- İşletme verileri ile sektör verilerinin karşılaştırılması,
- İşletme verileri ile finansal olmayan verilerin ortaya koyduğu beklenen sonuçların karşılaştırılması.

Analitik inceleme; denetimin planlanması, kanıt toplama ve denetimin tamamlanması ve raporlama olmak üzere denetim sürecinin her aşamasında kullanılmaktadır (Arens vd., 2000:190)

AICPA tarafından yayınlanan SAS No:56 Analitik Prosedürler Yönetmeliği ve ülkemizde SPK tarafından yayınlanan Seri X, No:22 sayılı Tebliğ de analitik inceleme prosedürlerinin denetimin planlama aşamasında kullanılmasını zorunlu tutmaktadır (SAS No:56, para.1; Seri X, No:22 sayılı Tebliğ – Onsekizinci Kısım- m.3).

Analitik inceleme prosedürlerinin denetimin planlama aşamasında kullanılması denetçiye, mali zorlukların ve başarısızlıkların önceden tespit edilmesi, birçok gelir tablosu hesaplarının uygunluğunun belirlenmesi, gelir tablosuyla doğrudan ilişkili bîlânço hesaplarının (stoklar ve alacaklar gibi) uygunluğunun belirlenmesi veya finansal zorlukların tespit edilmesi gibi yararlar sağlamaktadır.

Denetim planının hazırlanmasında, denetçinin tecrübesi, iç kontrol sisteminin güvenilirliği ve analitik inceleme prosedürlerinin sonuçları çok etkilidir. Denetçi tarafından hazırlanacak olan denetim planı aynı zamanda denetimin kalitesini de doğrudan etkileyeceği için denetçinin analitik inceleme prosedürlerini planlama aşamasında kesinlikle göz ardı etmemesi gerekir. Denetçi planlama aşamasında kullanacağı analitik prosedürlerin seçiminde serbesttir. Prosedürler basit oranlardan başlayıp karmaşık matematiksel modellere kadar değişebilmektedir (SAS No:56, par.2)

Analitik inceleme tekniklerinden biri olan oran analizi işletmelerin finansal durumu ve faaliyet sonuçlarını kendi içinde ve aynı veya farklı sektörden diğer işletmelerin sonuçlarıyla karşılaştırılarak incelemek amacıyla kullanılmaktadır. Bu oranlar sadece bir döneme ait olabileceği gibi, birden fazla döneme ait de olabilmektedir. Oranlar, finansal tabloda yer alan değerler arasındaki ilişkileri kurmaya yarar.

Oran analizi yapılırken; ilgili finansal tablolardan gerekli bilgilerin toplanarak oranların hesaplanması, hesaplanan oranların şirketin farklı dönemlerine ait oranlarla veya başka şirketlerin oranlarıyla karşılaştırılması, derleme ve çıkarımların ifade edilmesi adımları uygulanır (Bhat, 2008:291).

Oran analizinin avantajları aşağıda listelenmiştir (Singh, 2007:128).

- Finansal Açıdan Sağlıklı Olması: Oran analizi çeşitli ilgi grupları için şirket çapında sağlıklı analiz yapılmasına yardımcı olur.
- Planlama ve Tahmin Yapabilme: Belirli bir dönem üzerinde yapılan oran analizi yönetimin planlama yapmasına ve tahminlerde bulunmasına yardımcı olur.
- Kontrol Edilebilir Ölçütler: Oran analizi şirketin güçlü ve zayıf yönlerinin tanımlanmasını ve zayıf yönleri için önlem alınmasını sağlar.

- Karar Verme: Oran analizi organizasyonun faaliyet verimliliğinin artmasına katkıda bulunur.
- Kullanışlılık: Oran analizi, yatırımcılar tarafından uygulanan temel analizin en basit aracıdır. Finansal tabloların gösterdiği bilginin ilgililerce anlaşılır ve kullanılabilir olmasını sağlar.

4. Araştırma Yöntemi

4.1. Veri Toplama Yöntemi

Bu çalışma kapsamında, Borsa İstanbul'da gözültü pazarı ve ulusal pazarda yer alan firmaların 2013 yılı ilk 9 ayına ait bilanço ve gelir tabloları incelenmiş ve veri madenciliği tekniklerinden faydalanılarak karşılaştırma yapılmıştır. Bu çerçevede herhangi bir firmanın anlık finansal değerleri ile önümüzdeki dönemde gözültü pazarında yer alıp almayacağı tahmin edilebilmektedir. Böylelikle denetçi, firma hakkında ön bilgi sahibi olacaktır. Önerilen yöntemde Waikato Üniversitesi tarafından geliştirilen açık kaynak kodlu Weka veri madenciliği aracından yararlanılmıştır (<http://www.cs.waikato.ac.nz/ml/weka/>).

Bu amaçla öncelikle Kamuyu Aydınlatma Platformu'nun internet sitesinden gözültü pazarında yer alan tüm firmaların (20 firma) 2013 yılının ilk 9 ayına ait ve ulusal pazarda yer alan rastgele seçilmiş 20 firmanın aynı döneme ait finansal verileri elde edilmiştir. Bu firmaların listesi Ek 1' de sunulmuştur. Ulusal Pazar; Borsa İstanbul kotasyon koşullarını tümüyle karşılayan şirketlerin paylarının işlem gördüğü pazardır. Gözültü Pazarı ise; belirli koşulların ortaya çıkması sonucunda, şirketlerin izleme ve inceleme kapsamına alınması durumlarında sürekli gözetim, denetim ve izleme ortamında, yatırımcıların devamlı ve zamanında bilgilendirilmesini sağlayacak önlemlerle birlikte, hisse senetlerinin Borsa İstanbul bünyesinde işlem görebileceği pazardır (<http://www.kap.gov.tr/sirketler/islem-goren-sirketler/pazarlar>).

Bu verilerden faydalanılarak araştırmada anlamlı nitelikler oluşturulurken kullanılan mali tablo oranları aşağıda sunulmuştur.

Öznitelik Çıkarımında Uygulanan Mali Tablo Oranları

- Cari Oran = Dönen Varlıklar / Kısa Vadeli Yabancı Kaynak
- Asit_Test Oranı = (Dönen Varlıklar - Stoklar) / Kısa Vadeli Yabancı Kaynak
- Çabuk Oran = Ticari Alacak / Ticari Borç
- Nakit Oran = Hazır Değerler / Kısa Vadeli Yabancı Kaynak

- Stok Bağımlılık Oranı = (Kısa Vadeli Yabancı Kaynak - Hazır Değerler) / Stoklar
- Devamlı Sermaye Bağımlılık Oranı = (Stoklar + Ticari Alacak - Ticari Borç) / (Özkaynaklar + Uzun Vadeli Yabancı Kaynak)
- (Kısa Vadeli Yabancı Kaynak + Uzun Vadeli Yabancı Kaynak) / Varlıklar
- Özkaynaklar / Toplam Varlıklar
- Özkaynaklar / Toplam Borç
- (Özkaynaklar - Ödenmemiş Sermaye) / (Kısa Vadeli Yabancı Kaynak + Uzun Vadeli Yabancı Kaynak + Ödenmemiş Sermaye)
- (Kısa Vadeli Yabancı Kaynak + Uzun Vadeli Yabancı Kaynak) / (Özkaynaklar – Maddi Olmayan Duran Varlıklar)
- Maddi Duran Varlıklar / (Özkaynaklar + Uzun Vadeli Yabancı Kaynak)
- Maddi Duran Varlıklar / Özkaynaklar
- Maddi Duran Varlıklar / Toplam Varlıklar
- Dönen Varlıklar / Toplam Varlıklar
- Duran Varlıklar / Dönen Varlıklar.
- Satılan Mal Maliyeti / Ortalama Stok
- Ticari Alacak / Net Satışlar
- Net Satışlar / Ortalama Hazır Değerler
- Net Satışlar / (Dönen Varlıklar – Kısa Vadeli Yabancı Kaynak)
- Net Satışlar / Ortalama Özsermaye
- Net Satışlar / Ortalama Varlıklar
- Net Satışlar / Ortalama Duran Varlıklar
- Brüt Satış Kârı / Net Satışlar
- Faaliyet Kârı / Net Satışlar
- Olağan Kâr / Net Satışlar
- Dönem Kârı / Net Satışlar
- Net Kâr / Net Satışlar
- Faaliyet Giderleri / Net Satışlar
- Net Kâr / Özkaynak
- Net Kâr / Varlıklar
- Dönem Zararı / Özkaynaklar
- Altman Z Skoru (Altman :1968)

Her bir firma için yukarıda sayılan rasyolardan elde edilen nitelikler veri tabanına aktarılarak her bir firma için veri tabanında bir satır oluşturulmuştur. Bir satırdaki her bir sütun hesaplanan rasyolara aittir. Aynı zamanda bu sütunun kesiştiği her bir satır farklı bir firmanın bu değişkene ait verisini içermektedir.

Bazı firmaların sütunlarındaki bazı nitelikler hesaplanamamıştır. Bunun sebebi formüllerde kullanılan firma verilerinden pay veya paydada bir veya daha fazla verinin bulunmamasıdır. Bu nedenle oluşturulan veri setine veri temizleme işlemi uygulanarak sütunlarda eksik veri kalmaması sağlanmıştır. Veri temizleme işleminde, eksik nitelik değerleri o niteliğin ortalama değeri ile doldurulmuştur (<http://www.stat.gen.tr/index.php?istek=sinif&dersid=ist01&konuid=ver01&max=1>).

Böylelikle eksik veri kalmamış ve elde edilen veri seti Weka'da kullanılmaya hazır hale getirilmiştir. Elde edilen veri seti, Weka veri madenciliği programında K-en yakın komşu ile Naive Bayes algoritmaları ve 10-kat çapraz doğrulama ve veri setinin rastgele belirlenen %66 eğitim %34 test seti olacak şekilde eğitim ve test kümesi oluşturma yöntemleri kullanılarak test edilmiştir. Eğitim ve test kümeleri her bir algoritma için ayrı ayrı denenmiş ve sınıflandırıcıların başarıları doğru sınıflandırılan örneklerin oranları ile değerlendirilmiştir. Daha sonra sınıflandırıcılardan elde edilen bu başarı oranları karşılaştırılarak hangi tekniğin yapılan analizde daha doğru sınıflandırma yaptığı, dolayısıyla denetçiye firma seçilirken ya da denetim programı yürütülürken bu tekniklerden hangisinin daha fazla yardımcı olabileceği belirlenmiştir.

Buna ilaveten ayrıca 2013 yılında gözaltı pazarında yer alan firmaların verileri eğitim seti olarak kullanılarak, 2014 yılında gözaltı pazarına alınan bir firmanın verileri test seti olarak sınanmıştır. Bu test işleminde Naive Bayes algoritması ve 10-kat çapraz doğrulama tekniği kullanılmış, elde edilen sonuçlar değerlendirilmiştir.

4.2. Araştırmanın Kısıtları ve Karşılaşılan Zorluklar

Yapılan eğitim ve test işlemlerinden anlamlı sonuçlar elde edebilmek için veri setinde kullanılmak üzere belirlenen sınıfların eşit miktarda elemana sahip olması gerekmektedir. Bu çalışmada karşılaşılan kısıtlardan en önemlisi gözaltı pazarında yer alan firma sayısının az olması nedeniyle, veri setinin ulusal pazardaki ve gözaltı pazarındaki 20'şer şirket ile sınırlandırılmış olmasıdır. Ulusal pazardaki sadece 20 firmaya ait verilerin kullanılmasının sebebi, veri madenciliği teknikleri kullanılırken sınıflar arasında dengeli dağılım sağlanabilmesini için iki sınıfın da eşit dağıldığı bir veri setinin oluşturulması zorunluluğudur.

Ayrıca her bir şirket için mali tablo oranlarının hesaplanması, veri temizleme işlemi ve Weka programında kullanılan “arff” formatına dönüştürülmesi uzun zaman almıştır.

Diğer bir kısıt; analizde uygulanan şirket verilerinin sadece 2013 yılı 9 aylık mali tablolarından elde edilmesi ve sadece k-en yakın komşu algoritması ve Naive Bayes algoritmasının kullanılmasıdır.

5. Bulgular

Çalışmamızda, 2013 yılının ilk 9 ayına ait veri seti ile farklı sınıflandırma teknikleri kullanılarak sınıflandırma başarıları ölçülmüştür. Bu amaçla oluşturulan veri seti Weka veri madenciliği programına yüklenmiştir. Daha sonra sınıflandırıcı ve test işlemi alternatifleri seçilmiştir. Farklı başarı oranlarını karşılaştırmak amacıyla Naive Bayes ile K-en yakın komşu sınıflandırma algoritmaları ve 10-kat çapraz doğrulama ile %66 eğitim %34 test seti oluşturma teknikleri kullanılmıştır..

Weka programından elde edilen sonuç tablolarında ortaya çıkan “doğru pozitif”, “yanlış pozitif”, “doğru negatif” ve “yanlış negatif” değerlerinin yerleri Tablo 1’de gösterilmiştir. Çalışmadaki karışıklık matrislerinde de belirtildiği üzere “a” sınıfı ulusal pazardaki şirketleri, “b” sınıfı ise gözüaltı pazarındaki şirketleri temsil etmektedir. Yani ulusal pazarda sınıflandırılması gereken bir firma ulusal pazarda sınıflandırılmış ise “Doğru Pozitif”, fakat bu firma gözüaltı pazarında sınıflandırılmış ise “Yanlış Pozitif” sınıflandırılmıştır. Firma gözüaltı pazarında sınıflandırılması gerekirken gözüaltı pazarında sınıflandırılmış ise “Doğru Negatif”, fakat bu firma ulusal pazarda sınıflandırılmış ise “Yanlış Negatif” olarak sınıflandırılmış olur.

Tablo 1

Karışıklık Matrisi (Confusion Matrix)

		TAHMİN EDİLEN		
		Pozitif	Negatif	Sınıflama
GERÇEK	Pozitif	Doğru Pozitif	Yanlış Negatif	a = a
	Negatif	Yanlış Pozitif	Doğru Negatif	b = b

İlk olarak Naive Bayes algoritması ve 10-kat çapraz doğrulama tekniği kullanılmış, elde edilen sonuçlar Tablo 2’de sunulmuştur. Sonuç olarak karışıklık matrisinde (Confusion Matrix) veri setinin doğru pozitif sınıflandırılan örnek sayısı 19, doğru negatif sınıflandırılan örnek sayısı 16, yanlış pozitif sınıflandırılan örnek sayısı 4 ve yanlış negatif sınıflandırılan örnek sayısı ise 1 olarak bulunmuştur. Dolayısıyla doğru sınıflandırılan örnek sayısı 35 ve yanlış sınıflandırılan örnek sayısı ise 5 olarak elde edilmiştir. Oluşturulan modelin doğru sınıflandırma oranı %87,5 olarak hesaplanmıştır.

Tablo 2

Naive Bayes Algoritması Ve 10-Kat Çapraz Doğrulama Kullanılarak Sınıflandırma Tahmini Sonuçları

	NAİVE BAYES 10-Kat Ç.D.
Doğru Sınıflandırılan Örnek Sayısı	35
Yanlış Sınıflandırılan Örnek Sayısı	5
Doğru Pozitif	19
Yanlış Negatif	1
Yanlış Pozitif	4
Doğru Negatif	16
Doğru Sınıflandırma Oranı (%)	87,5
Yanlış Sınıflandırma Oranı (%)	12,5

Yine Naive Bayes algoritması kullanılarak bu defa sınıf verilerinin karışık olarak kaydedildiği veri setinin %66’sı eğitim seti, %34’ü ise test seti olarak ayrılarak sınıflandırıcının başarısı ölçülmüştür. Elde edilen sınıflama başarısı Tablo 3’de gösterilmiştir. Test setinde bulunan 14 firmanın sınıflandırıldıktan sonra elde edilen karışıklık matrisine göre doğru pozitif sınıflandırılan örnek sayısı 9, yanlış pozitif sınıflandırılan örnek sayısı 1 ve doğru negatif sınıflandırılan örnek sayısı 4 olarak bulunmuştur. Yanlış negatif sınıflandırılan firma ise olmamıştır. Toplamda doğru sınıflandırılan örnek sayısı 13 ve yanlış sınıflandırılan örnek sayısı ise 1 olarak elde edilmiştir. Oluşturulan modelin doğru sınıflandırma oranı %92,9 olarak hesaplanmıştır.

Tablo 3

Naive Bayes Algoritması Ve %66 Eğitim-%34 Test Seti Kullanılarak Sınıflandırma Tahmini Sonuçları

	NAİVE BAYES %66-%34
Doğru Sınıflandırılan Örnek Sayısı	13
Yanlış Sınıflandırılan Örnek Sayısı	1
Doğru Pozitif	9
Yanlış Negatif	0
Yanlış Pozitif	1
Doğru Negatif	4
Doğru Sınıflandırma Oranı (%)	92,86
Yanlış Sınıflandırma Oranı (%)	7,14

K-en yakın komşu algoritması ile 10-kat çapraz doğrulama tekniği kullanılarak oluşturulan modele ait sonuç Tablo 4’de verilmiştir. Burada k değeri “8” olarak modele dahil edilmiştir. Tablodaki karışıklık matrisinde veri setinde doğru pozitif sınıflandırılan örnek sayısı 20, yanlış pozitif sınıflandırılan örnek sayısı 2 ve doğru negatif sınıflandırılan örnek sayısı ise 18 olarak bulunmuştur. Yanlış negatif sınıflandırılan örnek ise olmamıştır. Doğru sınıflandırılmış örnek sayısı 38 ve yanlış sınıflandırılan örnek sayısı ise 2 olarak elde edilmiştir. Oluşturulan modelin doğru sınıflandırma oranı %95 olarak hesaplanmıştır.

Tablo 4

K-En Yakın Komşu Algoritması Ve 10-Kat Çapraz Doğrulama Kullanılarak Sınıflandırma Tahmini Sonuçları

	k-NN 10-Kat Ç.D.
Doğru Sınıflandırılan Örnek Sayısı	38
Yanlış Sınıflandırılan Örnek Sayısı	2
Doğru Pozitif	20
Yanlış Negatif	0
Yanlış Pozitif	2
Doğru Negatif	18

Doğru Sınıflandırma Oranı (%)	95
Yanlış Sınıflandırma Oranı (%)	5

Son olarak K-en yakın komşu algoritması ile %66 eğitim %34 test seti kullanılarak oluşturulan modele ait sonuç ise Tablo 5’de verilmiştir. K değeri yine “8” olarak alınmıştır. Karışıklık matrisinde, veri setinde doğru pozitif sınıflandırılan örnek sayısı 9, yanlış pozitif sınıflandırılan örnek sayısı 1 ve doğru negatif sınıflandırılan örnek sayısı ise 4 olarak bulunmuştur. Yanlış negatif sınıflandırılan örnek ise olmamıştır. Toplamda doğru sınıflandırılmış örnek sayısı 13 ve yanlış sınıflandırılan örnek sayısı ise 1 olarak elde edilmiştir. Oluşturulan modelin doğru sınıflandırma oranı %92,86 ve yanlış sınıflandırma oranı ise %7,14 olarak hesaplanmıştır.

Tablo 5

K-En Yakın Komşu Algoritması Ve %66 Eğitim-%34 Test Seti Kullanılarak Sınıflandırma Tahmini Sonuçları

	<i>k-NN</i> %66-%34
Doğru Sınıflandırılan Örnek Sayısı	13
Yanlış Sınıflandırılan Örnek Sayısı	1
Doğru Pozitif	9
Yanlış Negatif	0
Yanlış Pozitif	1
Doğru Negatif	4
Doğru Sınıflandırma Oranı (%)	92,86
Yanlış Sınıflandırma Oranı (%)	7,14

Çalışmanın daha sonraki aşamasında ilk uygulamadan farklı olarak, 2. Bölümde araştırma yönteminde belirtilen 2013 yılı 9 aylık mali tablo verileri kullanılarak oluşturulan veri seti, eğitim seti olarak kullanılarak 2014 yılında gözaltı pazarına alınmış bir firmanın gerçek zamanlı pazar tahmini yapılmıştır. Bu firmanın sınıflandırılmasında Naive Bayes algoritması kullanılmıştır. Elde edilen sonuç Tablo 6’da sunulmuştur. Sonuçlara göre 2014 yılında gözaltı pazarına alınan bu şirket, 2013 verilerine göre eğitilen modelde denenmiş ve

doğru negatif olarak sınıflandırılmıştır. Sonuç olarak başarılı bir şekilde gözetli pazarında olduğu tahmin edilmiştir.

Tablo 6

2014 Yılında Gözetli Pazarına Alınmış Olan Bir Firmanın Gerçek Zamanlı Pazar Sınıflandırma Tahmini Sonuçları

	<i>NAİVE BAYES</i> 10-Kat Ç.D.
Doğru Sınıflandırılan Örnek Sayısı	1
Yanlış Sınıflandırılan Örnek Sayısı	0
Doğru Pozitif	0
Yanlış Negatif	0
Yanlış Pozitif	0
Doğru Negatif	1
Doğru Sınıflandırma Oranı (%)	100
Yanlış Sınıflandırma Oranı (%)	0

Yaptığımız çalışmadan elde edilen sonuçlar Tablo 7’de toplu olarak ifade edilmiştir.

Tablo 7

2013 Yılı Verilerine Göre Sınıflandırma Algoritmalarının Başarı Durumu Karşılaştırması

	<i>BAYES</i>		<i>k-NN</i>	
	10-Kat Ç.D.	%66- %34	10-Kat Ç.D.	%66- %34
Doğru Sınıflandırılan Örnek Sayısı	35	13	38	13
Yanlış Sınıflandırılan Örnek Sayısı	5	1	2	1
Doğru Pozitif	19	9	20	9
Yanlış Negatif	1	0	0	0
Yanlış Pozitif	4	1	2	1
Doğru Negatif	16	4	18	4
Doğru Sınıflandırma Oranı (%)	87,5	92,86	95	92,86
Yanlış Sınıflandırma Oranı (%)	12,5	7,14	5	7,14

6. Sonuç ve Değerlendirme

Naive Bayes algoritması ve 10 kat çapraz doğrulama tekniği uygulandığında toplam doğru sınıflandırılan örnek sayısı 35, toplam yanlış sınıflandırılan örnek sayısı 5 olarak bulunmuştur. Naive Bayes algoritması ve %66-%34 tekniği uygulandığında toplam doğru sınıflandırılan örnek sayısı 35, toplam yanlış sınıflandırılan örnek sayısı 5 olarak bulunmuştur. k-en yakın komşu algoritması ve 10 kat çapraz doğrulama tekniği uygulandığında toplam doğru sınıflandırılan örnek sayısı 38, , toplam yanlış sınıflandırılan örnek sayısı 2 olarak bulunmuştur. k-en yakın komşu algoritması ve %66-%34 tekniği uygulandığında toplam doğru sınıflandırılan örnek sayısı 13, , toplam yanlış sınıflandırılan örnek sayısı 1 olarak bulunmuştur.

Bu sonuçlara göre Naive Bayes algoritması ve 10 kat çapraz doğrulama tekniği uygulandığında doğru sınıflandırma oranı %87,5 olarak, toplam yanlış sınıflandırma oranı ise %12,5 olarak bulunmuştur. Naive Bayes algoritması ve %66-%34 tekniği uygulandığında doğru sınıflandırma oranı %92,86, yanlış sınıflandırma oranı %7,14 olarak bulunmuştur. k-en yakın komşu algoritması ve 10 kat çapraz doğrulama tekniği uygulandığında doğru sınıflandırma oranı %95, yanlış sınıflandırma oranı ise %5 olarak bulunmuştur. k-en yakın komşu algoritması ve %66-%34 tekniği uygulandığında doğru sınıflandırma oranı %92,86 ve yanlış sınıflandırma oranı ise %7,14 olarak bulunmuştur.

Sonuç olarak Naive Bayes algoritması ve %66-%34 tekniği ile k-en yakın komşu algoritması ve %66-%34 tekniği aynı doğru sınıflandırma oranına ulaştırmıştır. En iyi doğru sınıflandırma oranına k-en yakın komşu algoritması ve 10 kat çapraz doğrulama tekniği ile ulaşılmış, en kötü doğru sınıflandırma oranına Naive Bayes algoritması ve 10 kat çapraz doğrulama tekniği ile ulaşılmıştır.

Çalışmamızda ortaya çıkan sonuçlara göre denetim firmaları denetlenecek firmalar ile sözleşme yapmadan önce veya denetçiler denetim sürecinin bir parçası olan işletmeyi tanımak için yaptıkları ön analitik inceleme sırasında veri madenciliği yöntemlerinden k-en yakın komşu algoritması ve 10 kat çapraz doğrulama tekniğini kullanmaları daha doğru olacak, bu sayede firma hakkında ön bilgi sahibi olunabilecektir.

Çalışmamızda iki önemli kısıt vardır. Birincisi, çalışmada analizde uygulanan şirket verilerinin sadece 2013 yılı 9 aylık mali tablolarından elde edilmesidir. Bu konuda çalışacak araştırmacılar veri setini genişletebilirler. İkincisi, çalışmada sadece k-en yakın komşu algoritması ve Naive Bayes algoritması kullanılmıştır. Araştırmacılar diğer veri madenciliği algoritmalarını da kullanarak çalışmayı genişletebilirler.

Kaynakça

- Aci, M. & Avci, M. (2011). K nearest neighbor reinforced expectation maximization method. *Expert Systems with Applications*, 38, 10, 15, 12585-12591.
- Akyel, N. & Seçkin K. (2011). K-en yakın komşuluk algoritmasının hile denetiminde kullanımı. *Muhasebe ve Vergi Uygulamaları Dergisi*, 51, 21-39.
- Albayrak, A.S. & Koltan Yılmaz, Ş. (2009). Veri madenciliği karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14, 1, 31-52.
- Almonacid F., Fernández, E.F., Rodrigo, P., Pérez-Higueras, P.J. & Rus-Casas, C. (2013). Estimating the maximum power of a high concentrator photovoltaic (HCPV) module using an Artificial Neural Network. *Energy*, 53, 165-172.
- Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, XXIII, 4, 589-609.
- Arens, A. & Loebbecke, J. (2000). *Auditing: An Integrated Approach*. New Jersey: Prentice Hall.
- Arens, A., Randall, J., & Beasley, M. (2006). *Auditing and Assurance Services: An Integrated Approach* (11. b.). New Jersey: Prentice Hall.
- Bankalarda Bağımsız Denetim Gerçekleştirecek Kuruluşların Yetkilendirilmesi ve Faaliyetleri Hakkında Yönetmelik, md.33/g
- Bhat, M.S. & Rau, A.V. (2008). *Managerial Economics and Financial Economics*. Hyderabad Sultan Bazar: BS Publications,.
- Boran, L. (2012). *Veri Madenciliğinin Türk İşletmelerin Finansal Tablolarına Uygulanması ve Uygulama Örneği*. Yayınlanmamış Doktora Tezi, Marmara Üniversitesi S.B.E., İstanbul.
- Cerullo, M. J. & Cerullo, V. (1999). Using neural networks to predict financial reporting fraud. *Computer Fraud & Security*, May/June, 14–17.
- Cunha C. D., Agard, B., & Kusiak, A. (2006). Data mining for improvement of product quality. *International Journal of Production Research*, 44:18/19, 4041–4054.
- Fanning, K., Cogger, K., & Srivastava, R. (1995). Detection of management fraud: a neural network approach. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 4, 2, 113– 26.
- Fanning, K. & Cogger, K. (1998). Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7, 1, 21- 24.
- Feroz, E. H., Kwon, T. M., Pastena, V. & Park, K. J. (2000). The efficacy of red flags in predicting the SEC's targets: an artificial neural networks Approach. *International Journal of Intelligent Systems in Accounting, Finance, and Management*, 9, 3, 145–157.
- Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Stat Comput*, 21, 137–146.
- Glancy, F. H. & Surya, B. Y. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, 50, 95–601.
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques* (12.b.). Romania: Intelligent Systems Reference Library.

Goyal, S. & Goyal, G. K. (2011). A New Scientific Approach of Intelligent Artificial Neural Network Engineering for Predicting Shelf Life of Milky White Dessert Jeweled with Pistachio. *International Journal of Scientific & Engineering Research*, 2, 9, 1-4.

Green, B. P. & Choi, J. H. (1997). Assessing the risk of management fraud through neural-network technology. *Auditing: A Journal of Practice and Theory*, 16(1), 14–28.

<http://www.cs.waikato.ac.nz/ml/weka/> Erişim Tarihi: 18.09.2013

http://ecs.victoria.ac.nz/Courses/COMP307_2014T1/Lect13-team1 Erişim Tarihi: 17.04.2014

<http://www.kap.gov.tr/sirketler/islem-goren-sirketler/pazarlar> Erişim Tarihi: 11.04.2013

<http://www.stat.gen.tr/index.php?istek=sinif&dersid=ist01&konuid=ver01&max=1> Erişim Tarihi: 17.04.2014

<http://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00329.pdf> Erişim Tarihi: 16.06.2013

<http://www.ifac.org/sites/default/files/publications/files/A027%202012%20IAASB%20Handbook%20ISA%20520.pdf> Erişim Tarihi: 18.06.2013

Han, J., Kamber, M. & Pei J. (2012). *Data Mining* (Third Edition). Boston: Morgan Kaufmann.

Huang, S.Y., Tsaih, R. H. & Yu, F. (2014). Topological pattern discovery and feature extraction for fraudulent financial reporting. *Expert Systems with Applications*, 41, 9, 4360–4372.

Jans M., Lybaert, N. & Vanhoof, K. (2010). Internal fraud risk reduction: Results of a data mining case study. *International Journal of Accounting Information Systems*, 11, 17–41.

Kapardis, M. K., Christodoulou, C. & Agathocleous, M. (2010). Neural networks: the panacea in fraud detection? *Managerial Auditing Journal*, 25, 659-678.

Kirkos, E., Spathis, C. & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32, 4, 995–1003.

Koskivaara, E. (2000). Different pre-processing models for financial accounts when using neural networks for auditing. *Proceedings of the 8th European Conference on Information Systems*, 1, (pp.326–3328).

Kotsiantis, S., Koumanakos, E., Tzelepis, D. & Tampakas, V. (2006). Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence*, 3, 2, 104–110.

Kou, G., Peng Y. & Guoxun W. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences*. <http://dx.doi.org/10.1016/j.ins.2014.02.137> Erişim Tarihi: 27.03.2014.

Kuzey, C., Uyar, A. Delen, & D. (2014). The impact of multinationality on firm value: A comparative analysis of machine learning techniques. *Decision Support Systems*, 59, 127–142.

Larose, D., T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining.*, Hoboken: John Wiley and Sons.

Lin, J. W., Hwang, M. I. & Becker, J. D. (2003). A Fuzzy Neural Network for Assessing the Risk of Fraudulent Financial Reporting. *Managerial Auditing Journal*, 18, 8, 657-665.

Liou, F. M. (2008). Fraudulent financial reporting detection and business failure prediction models: a comparison. *Managerial Auditing Journal*, 23, 7, 650-662.

- Mastrogiannis, N., Boutsinas, B. & Giannikos, I. (2009). A method for improving the accuracy of data mining classification algorithms. *Computers and Operations Research*, 36, 10, 2829-2839.
- Ngai, E.W.T., Yong, H., Wong, Y.H., Yijun, C. & Xin Sun (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50, 559-569.
- Osuna, R.G. (2002). *Lecture Notes CS 790: Introduction to pattern recognition*. Ohio: Wright State University.
- Özkan, Y. (2013). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim.
- Pandey, U.K. & Pal S. (2011). Data Mining: A prediction of performer or underperformer using classification. *International Journal of Computer Science and Information Technologies*, 2, 2, 686-690.
- Perols, J. (2011). Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *Auditing: A Journal of Practice and Theory*, 30, 2, 19-50.
- Ravisankar, P., Ravi, V., Rao, G.R. & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support System*, 50, 491-500.
- Rejesus R.M., Little, B.B. & Lovell, A.C. (2004). Using data mining to detect crop insurance fraud: Is there a role for social scientists? *Journal of Financial Crime*, 12, 1, 24-32.
- Rodriguez, J. D., Perez A. & Lozano J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *Transactions On Pattern Analysis And Machine Intelligence*, 32, 3, 569-575.
- Sermaye Piyasasında Bağımsız Denetim Standartları Hakkında Tebliğ- Seri X, No:22, Kısım 18, md.2
- Sharma, A. & Panigrahi, P. K. (2012). A Review of Financial Accounting Fraud Detection based on Data Mining Techniques. *International Journal of Computer Applications*, 39-1, 37-47.
- Shmueli, G., Nitin, R. P., Peter, C. B. (2010). *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. Hoboken: John Wiley and Sons.
- Singh, Y.P. (2007). *Accounting and Financial Management For IT Professionals*. New Delhi: New Age International Limited Publishers,.
- Soria, D., Garibaldi, J. M., Ambrogi, F., Biganzoli, E. M. & Ellis I. O. (2011). A 'non-parametric' version of the naive Bayes classifier. *Knowledge-Based Systems*, 24, 775-784.
- Swartz, N. (2004). IBM, Mayo clinic to mine medical data. *The Information Management Journal*, 38, 6.
- Terzi, S. (2012). Hile Ve Usulsüzlüklerin Tespitinde Veri Madenciliğinin Kullanımı. *Muhasebe ve Finansman Dergisi*, 54, 51-64.
- Togun, N. K. & Sedat Baysec (2010). Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks. *Applied Energy*, 87, 349-355.
- Wan, C. H., Lee, L. H., Rajkumar, R. & Dino, I. (2012). A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine. *Expert Systems with Applications*, 39, 15, 11880-11888.
- Weng, S.-S., Chiu, R.-K., Wang, B.-J. & Su S.-H. (2006/2007). The study and verification of mathematical modeling for customer purchasing behavior. *Journal of Computer Information Systems*, 47, 2, 46-57.
- Yücel, E. (2013). Hileli Finansal Raporlamanın Tespitinde Kırmızı Bayrakların Etkinliği: Türkiye Uygulaması. *Muhasebe ve Finansman Dergisi*, 60, 139-158.

Zhou, W. & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. Decision Support Systems, 50, 3, 570-575.

Ek :

2013 Yılı İlk 9 Aylık Mali Tabloları Kullanılan Firmalar

<i>Gözetli pazarındaki firmalar</i>	<i>Ulusal pazardaki firmalar</i>
1 Ansa Yatırım Holding A.Ş.	Adana Çimento Sanayii T.A.Ş.
2 Artı Yatırım Holding A.Ş.	Adel Kalemcilik Ticaret ve Sanayi A.Ş.
3 ÇBS Boya Kimya Sanayii ve Ticareti A.Ş.	Afyon Çimento Sanayi T.A.Ş.
4 ÇBS Printaş Oto Boya ve Gereçleri Sanayii A.Ş.	Akçansa Çimento Sanayi ve Ticaret A.Ş.
5 Dardanel Önentaş Gıda Sanayi A.Ş.	Akenerji Elektrik Üretim A.Ş.
6 Egeplast Ege Plastik Ticaret ve Sanayi A.Ş.	Akfen Holding A.Ş.
7 Ekiz Yağ Ve Sabun Sanayii A.Ş.	Akın Tekstil A.Ş.
8 Esem Spor Giyim Sanayi ve Ticaret A.Ş.	Aksa Akrilik Kimya Sanayii A.Ş.
9 Feniş Alüminyum Sanayi ve Ticaret A.Ş.	Aksa Enerji Üretim A.Ş.
10 Frigo-Pak Gıda Maddeleri Sanayi ve Ticaret A.Ş.	Aksu Enerji ve Ticaret A.Ş.
11 İCG Yatırım Holding A.Ş.	Akyürek Tüketim Ürünleri Pazarlama Dağıtım ve Ticaret A.Ş.
12 İdaş İstanbul Döşeme Sanayii A.Ş.	Alarko Carrier Sanayi ve Ticaret A.Ş.
13 Kapital Yatırım Holding A.Ş.	Alarko Holding A.Ş.
14 Mango Gıda Sanayi Ve Ticaret A.Ş.	Alcatel Lucent Teletaş Telekomünikasyon A.Ş.
15 Mazhar Zorlu Holding A.Ş.	Alkim Alkali Kimya A.Ş.
16 Metro Altın İşletmeciliği İnşaat Sanayi ve Ticaret A.Ş.	Alkim Kağıt Sanayi ve Ticaret A.Ş.
17 Selçuk Gıda Endüstri İhracat İthalat A.Ş.	Altınyaz Kombinaları A.Ş.
18 Şeker Piliç ve Yem Sanayi Ticaret A.Ş.	Altınyıldız Mensucat ve Konfeksiyon Fabrikaları A.Ş.
19 Transtürk Holding A.Ş.	Anadolu Cam Sanayii A.Ş.
20 Yeşil Yatırım Holding A.Ş.	Anadolu Efes Biracılık Ve Malt Sanayii A.Ş.
21* Ataç İnşaat ve Sanayi A.Ş.	

*2014 yılında Borsa İstanbul'da "Gözetli Pazarı"na alınan şirket