

k-EN YAKIN KOMŞULUK ALGORİTMASININ HİLE DENETİMİNDE KULLANIMI

Yrd. Doç. Dr. Nermin AKYEL*

Arş. Gör. Keziban SEÇKİN**

ÖZET

İşletmede üretilen finansal tabloların amacı, doğru, dürüst, güvenilir bilgiler aracılığıyla işletmenin içerisinde bulunduğu durumu olduğu gibi sunmak ve ilgili taraflara karar almada yardımcı olacak finansal bilgiyi sağlamaktır. Finansal tablolarda yer alan bilgilerin hata ve hileleri barındırıyor olma olasılığı, finansal tabloların güvenilirliğinin onaylanması aşamasında muhasebe denetimine olan ihtiyacı giderek arttırmıştır. Bu doğrultuda denetim faaliyetleri içerisinde hile denetimi kendisine ayrı bir yer edinmiştir. Günümüz koşullarında iş ve işlemlerin sayısındaki artış ve karmaşıklık sonucu incelenen veri sayısının oldukça büyük olması, karar alma süreçlerinin de karmaşılaşmasına yol açmış ve doğal olarak daha fazla sayısal verinin incelenmesi ihtiyacını ortaya çıkartmıştır.

Denetçi açısından verilen kararların doğruluğu, karar vericinin yeteneği, deneyimi ve bilgi birikiminin yanı sıra elde edilen veri setinin uygunluğu ve yeterliği ile de ilişkilidir. Veri depolarının büyüklüğünün manuel olarak kontrol edilemeyecek boyutlara ulaşmış olması uygulayıcıları ve araştırmacıları, denetim faaliyetlerinde bilgisayar destekli uygulamalardan yararlanma yollarını araştırmaya yöneltmiştir. Bu çalışmada, mevcut verilerin bilgiye dönüşebilmesi için gerekli işlemler bütünü olarak da tanımlayabileceğimiz veri madenciliği tanıtılmış; karar alma aşamasında veri madenciliğinin kullanılması suretiyle ilgililerin doğru karar alma süreçlerine yardımcı olunabileceği vurgulanmış ve veri madenciliği yöntemlerinden biri olan “k- En Yakın Komşuluk Algoritması”nın hileli uygulamaların tespitinde kullanılabilmesi üzerinde durulmuştur.

Hile denetiminde veri madenciliğinden yararlanılabileceğinin vurgulandığı çalışmada, özellikle çalışan hilelerinin önlenmesinde en yakın k- en yakın

* Sakarya Üniversitesi Geyve Meslek Yüksekokulu e-posta: nakyel@sakarya.edu.tr

** Yıldırım Beyazıt Üniversitesi seckin.kezban@gmail.com

komşuluk algoritmasının kullanılabilceği düşünçesine yer verilmektedir. Bu yöntemin kullanılmasıyla hile yapma eğiliminde bulunabilecek çalışanlar önceden kestirilebilecek, dolayısıyla hilenin oluşmasına fırsat verilmeyecektir. Çalışmada veri madenciliği ve k - en yakın komşuluk algoritması hakkında literatür taraması ile potansiyel hilebazın önceden tespitinde k - en yakın komşuluk algoritmasından yararlanılabileceğine ilişkin örneklere yer verilmiştir.

Anahtar Kelimeler: denetim, hile, çalışan hileleri, veri madenciliği, k - en yakın komşuluk algoritması

Jel Kodlar: M40, M42, M49

USE OF k -NEAREST NEIGHBOR ALGORITHM IN FRAUD AUDITING

Abstract

The financial statements produced at an establishment aim to present the current situation of the establishment as is by means of accurate, flawless and reliable information and to provide the parties concerned with the financial information to help them in decision making. The presence of fraudulent financial statements has brought forward the reliability and approval of financial statements. In line with this, accounting auditing has become a need and fraud auditing has gained further importance among the auditing activities. The quite large number of data to be examined as a result of the increase in the numbers of jobs and transactions and as a result of complexity under the present conditions has also caused the decision making processes to become complex and, naturally, led to the need to examine more numerical data.

The accuracy of the decisions made in terms of an auditor is associated with the suitability and efficiency of the data set obtained besides the ability, experience and accumulation of knowledge of the decision maker. The fact that the size of data repositories has reached sizes that cannot be manually controlled has directed implementers and researchers to search for ways to make use of computer-aided applications in auditing activities. This study introduces data mining, which we may also define as the whole of procedures required for the transformation of existing data into information; it emphasizes that by using data mining at the stage of decision making, assistance may be provided for the accurate decision making processes of those concerned; and it considers that “ k -Nearest Neighbor

Algorithm”, one of the methods of data mining, might be used to detect fraudulent applications.

The idea that k -nearest neighbor algorithm might be used particularly to prevent employee frauds is contained in the study where it is emphasized that data mining might be utilized in fraud auditing. By using this method, it will be possible to predict those employees who might tend to do fraud and therefore the occurrence of fraud will not be allowed. The study provides a literature review on data mining and k -nearest neighbor algorithm and examples of the possibility of making use of k -nearest neighbor algorithm to detect a potential fraud beforehand.

Key words: auditing, fraud, employee frauds, data mining, k -nearest neighbor algorithm

Jel Codes: M40, M42, M49

1. GİRİŞ

Yirminci yüzyılın özellikle ikinci yarısında yaşanan bilim ve teknoloji alanındaki hızlı değişimler, işletme faaliyetlerinin de çeşitliliğini arttırmış; dolayısıyla iş ve işlemlerin sayısı ve mahiyetinde de oldukça önemli artışlar olmuştur. İş çevreleri de bilim ve teknolojide yaşanan bu hızlı değişimden payını almış; işletmede gerçekleşen iş ve işlemlerin uygulanmasında farklı fonksiyonlar ve karmaşık tanımlamalar geliştirilmiştir. Yaşanan değişim sadece iş hayatı ve işlemlerle sınırlı kalmayıp, hayatın her aşamasında kendini hissettirmeye başlamıştır. Her alanda yaşanan hızlı ve köklü değişim, çevresel faktörler üzerinde etkili olmuş, bunların bir uzantısı olarak sosyal hayatta da önemli değişiklikler yaşanmıştır. Sosyal çevrede yaşanan değişikliklerden iş çevreleri ve kişisel çevre de payını almış; kişilerin, dolayısıyla kurumların beklenti ve davranışları değişmiş, insani bir takım özelliklerden olan kişisel hırs ve kurumsal beklentiler artmıştır. Kişisel hırs ve kurumsal beklentilerdeki bu artış, pek çok alanda doğruluk, dürüstlük gibi kavramların unutulması ikinci plana atılmasına yol açmıştır. Kişiler ve kurumlar kendilerini olduklarından farklı bir biçimde gösterme ihtiyacı ile bazı olumsuz davranış kalıpları geliştirerek kasıtlı bir biçimde, karşısındaki kişi ve kuruluşları yanıltıcı sonuçlar doğuran uygulamalara yönelmişlerdir.

Muhasebe, işletmede gerçekleşen olayları yansıtmayı nedeniyle, “işletmenin dili” olarak nitelendirilmektedir. Muhasebenin ürettiği bilgiler ise finansal kararlara esas teşkil etmekte olup, muhasebeye “finansal ka-

rarların dili” demek de mümkündür (Sevilengül, 2003: 11). İşletme sahipleri, ortakları, yöneticiler, kreditorler, devlet ve yatırımcılar başta olmak üzere pek çok kişi işletmenin içerisinde bulunduğu durum ve mali yapısı hakkında bilgi sahibi olmak için muhasebede üretilmiş bilgilere ihtiyaç duymaktadır. İlgili tarafları yanıltıcı amaçla gerçekleştirilen uygulamalar, işletmenin faaliyetlerinin sonucunun, dolayısıyla finansal tablolarda yer alan bilgilerin, olduğundan farklı gösterilmesine yol açmaktadır. Çalışanlardan bir kısmı, gerek kendi kişisel hırsları, gerekse işletme yöneticilerinden gelen talepler doğrultusunda, işletmenin içerisinde bulunduğu durumu olduğundan farklı olarak gösterme çabasıyla yanıltıcı bilgi üretmeye yönelik uygulamalar gerçekleştirmektedirler.

Genel kabul görmüş muhasebe ilkelerine ve standartlara uygun olarak düzenlenmiş finansal bilgileri içeren finansal tabloların hazırlanması, güvenilir finansal raporlama olarak tanımlanmaktadır (Doyrangöl, 2007: 53). Güvenirlikten uzaklaşmış, gerçekte var olmayan kayıtlar aracılığıyla ve yanıltıcı sonuçlara ulaştıran, hileli finansal tabloların varlığı ise işletmeler ve diğer finansal tablo kullanıcıları için pek çok açıdan önemli bir risk unsuru haline gelmektedir. Bu doğrultuda işletme yönetimi, yönetim fonksiyonunun gereği olarak; hata, yolsuzluk ve hilelerden koruyacak, önleyici tedbirleri alarak sistemlerini buna göre kurgulamak zorundadır. Bu çalışmada işletme yönetiminden gelen talep doğrultusunda değil, kişisel istek ve beklentileri doğrultusunda hile yapma eğiliminde bulunan mevcut çalışanların ve çalışan adaylarının tespit edilmesinde kullanılabilecek sayısal hesaplamalara dayanan bir tanıma yönteminin var olup olmayacağı sorusuna cevap aranmıştır.

2. HİLE VE HİLELİ UYGULAMALAR

Türk Dil Kurumunun yayınladığı Türkçe Sözlükte hile, “birini aldatmak, yanıltmak için yapılan düzen, dolap, oyun” olarak tanımlanmakta olup, düzenbazlık, yolsuzluk ve sahtekarlık kelimeleriyle eşanlamlı olarak kullanılmaktadır. Literatürde “işletmelerde olması gerekenden farklı hareket biçimleri sergileyerek değişik sonuçların ortaya çıkartılması” yolsuzluk (Kaval, 2005: 62) olarak tanımlanırken, “muhasebe bilimi açısından hile, yanlış olduğu bilinen ya da doğruluğuna inanılmayan birtakım uygunsuzlukların ve yasadışı hareketlerin, maksatlı kandırma veya yanlış sunma niyetiyle yapılması” (Emir, 2008: 12), “hile, başkası ya da başkalarından haksız kazanç sağlamak amacıyla yapılan sahtekar ve aldatıcı davranışları içerir (Graycar ve Smith, 2002: 5), “belirli çıkarlar için sahtecilik yaparak, dürüstlüğü aykırı bir biçimde işlem yapılması, kurallara ve yasalara bile-

rek aykırı davranılması” (Erdoğan, 2006:91), “işletme yönetimindekiler ve yönetimden sorumlu kişilerin, işletme çalışanlarının ya da üçüncü şahısların, kasıtlı olarak adil ya da yasal olmayan bir menfaat sağlamak amacıyla aldatma içeren davranışlarda bulunmaları” (SPK, Sermaye Piyasasında Bağımsız Denetim Standartları hakkında Tebliğ, 2006: 134) gibi tanımlamalar yer almaktadır. Bu tanımlamalardan çıkartılabilecek ortak bir ifade ile “kişilerin, üyesi oldukları işletmenin varlıklarını ve kaynaklarını, kasıtlı olarak uygun olmayan bir biçimde kullanmak ya da ele geçirmek suretiyle haksız kazanç elde etmek amacıyla gerçekleştirdikleri aldatıcı davranışlar”, hile olarak adlandırılabilir (Akyel, 2009: 174). Hileli finansal raporlamayı inceleme amaçlı kurulan Amerikan Komisyonu, hileyi “finansal tabloları önemli ölçüde yanlış gösteren herhangi bir eylem” şeklinde tanımlamakta olup; hile, belge ve dokümanları değiştirmek, kayıtlardan işlemleri silmek, hayali işlemleri kaydetmek ya da önemli bilgileri yok etmek şeklinde yapılan uygulamalar sonucu ortaya çıkmaktadır (Demir ve Bahadır, 2007: 115; Stolowy ve Breton’dan alıntı). 1997 yılında Amerikan Serbest Muhasebeci ve Mali Müşavirler Enstitüsü (AICPA) tarafından yayınlanan Denetim Standartlarından (Statement on Auditing Standards) SAS 82’de “kasıtlı olarak doğru olmayan bilgilere yer verilen finansal raporlar” hileli finansal raporlar olarak tanımlanmış ve hilenin, hileli finansal raporlama ya da varlıkların kötüye kullanımından kaynaklandığı ifade edilmiştir (Özşahin, 2000: 24).

İşletme yönetimi, ilgili taraflara gerekli bilgilerin sağlanmasından ve güvenilirliğinden sorumludur. Muhasebeleştirilecek işlemlerin sayısının ve karmaşıklığının artmış olması bilginin güvenilirliğini etkileyecektir. Söz konusu bilgiye ihtiyaç duyan ilgili tarafların, bilgi istenen konu ve muhasebeleştirilmesi hakkında ayrıntılı bilgi sahibi olmamaları, muhtemel hatalı ve hileli uygulamaların gözden kaçmasına sebep olabilecektir. Hile içeren finansal tablolardan elde edilen finansal bilginin karar alma süreçlerinde kullanılması durumunda ise ilgili taraflar açısından oldukça ağır ve olumsuz sonuçlar ile karşılaşılacaktır.

3. HİLENİN KAVRAMSAL TEMELLERİ

Finansal tablolarda yer alan bilgilerin yanıltıcı niteliğe sahip olmasında hata ve hile olmak üzere iki temel unsur yer almaktadır. Hata, bazı işlemleri gerçekleştirirken, bilgisizlik, tecrübesizlik, dikkatsizlik ya da yorum farkından doğan yanlış uygulamaları içerirken, hile yanlışlık ya da uygun olmayan işlemlerin bilerek ve isteyerek gerçekleştirilmesi durumudur. Hilenin temelinde çıkar sağlama amacı ve kasit unsurunun varlığı yer almaktadır.

İşletmenin çıkarlarının ikinci plana atılarak kişisel çıkarlar doğrultusunda gerçekleştirilen hileli işlemlerle mücadele edebilmek için, hilenin temel yapısının doğru olarak tespit edilmesi gerekir. 1950'li yıllarda Cressey tarafından ileri sürülen ve daha sonra geniş çevrelerce kabul gören hile üçgeni hipotezinde, suç işlemiş ve işlemeye meyilli kişiler incelenmiş ve bu kişilerin olumsuzluk içeren davranış kalıplarında *baskı* (incentive), *fırsat* (opportunity) ve *haklı gösterme* (rationalization) olmak üzere üç temel unsurun etkili olduğu ifade edilmiştir (Kandemir, 2010: 15). Bu üç risk kategorisi, SAS (Statements of Auditing Standards) 99'da hile üçgeni olarak tanımlanmış ve IAS (International Auditing Standards) 240 da, bu üç faktörün, hilenin varlığı durumunda aranması gereken temel anahtar unsurlar olduğunu ifade etmiştir. Buna göre, yöneticiler şirketin finansal sonuçlarının olduğundan daha iyi gözükmesini istedikleri için, çalışanlar ise çeşitli kişisel sebeplerle (genellikle para ihtiyaçlarının baskısı ile) hile yapmaya yönelmektedirler (Çubukçu, 2009: 118).

Çalışanları hile yapmaya iten unsurlar hile üçgeni bağlamında sınıflandırılacak olursa aşağıdaki örneklere rastlamak mümkündür:

- Bazı tutku ve alışkanlıkları dolayısıyla yüksek tutarlı borçlanmış ve borçlarını ödemekte zorlanmakta olup, çıkış yolu aramaktadır (baskı unsuru),
- Daha rahat bir yaşantı sürebilmek amacıyla, işletmede sağlıklı bir iç kontrol yapısı ve denetim mekanizması olmadığını tespit ederek yakalanma olasılığının bulunmadığına inanmaktadır (fırsat unsuru),
- Hilenin tespiti durumunda, yaptığı işin ahlaki nitelik taşımadığını bilse de kaybına yol açtığı tutarı ya da varlığı daha sonra yerine koymayı düşündüğünü ve bu sebepten suçlanabileceğini düşünmediği gibi bahaneler ileri sürmektedir (haklı gösterme unsuru).

Baskı, fırsat ve haklı gösterme olarak tanımlanan üç unsurun varlığı ya da bir araya gelmeleri halinde hileli eylem oluşur. Kesin olarak ifade edilmese de, bu üç unsurdan birinin ya da birkaçının bulunmaması, hile olasılığını azaltabilecek ya da tamamen ortadan kaldıracaktır. İşletmede karşılaşılabilecek hileler, hile üçgeninde yer alan üç unsurun birden bünyesinde barındırıldığında risk oranı daha yüksek olacaktır (Hillison ve Diğerleri, 1999: 353).

PriceWaterhouse Coopers tarafından yapılan 2009 Ekonomik Suçlar raporunda 54 farklı ülkeden 3000 üst düzey yönetici için eğilim yoklamasına yer verilmiş, katılımcılardan %68'i baskı (güdü), %18'i fırsat (ortam), %14'ü haklı gösterme (tutum) faktörünün hile oluşumunda birinci derecede etkili

olduğunu düşündüklerini ifade etmişlerdir (Kandemir, 2010, 15).

4. ÇALIŞAN HİLELERİ

Belge ve dokümanları değiştirmek, kayıtlardan bazı işlemleri silmek, hayali işlemleri kaydetmek ya da önemli bilgileri yok etmek şeklinde gerçekleştirilen işlemler sonucu ortaya çıkan (Demir ve Bahadır, 2007: 115; Stolowy ve Breton'dan alıntı) hileli uygulamalar sonucunda, işletme, yöneticiler ya da çalışanlar lehine olumlu bir takım transferlerin varlığı söz konusudur. 2008 ACFE Raporuna göre, işletmede gerçekleştirilen hilelerin yaklaşık %40'ı düz çalışanlardan, %37'si yöneticilerden, %23'ü ise ortak ve yöneticiler tarafından gerçekleştirilmektedir (Bozkurt, 2009: 55).

Taşıdığı farklı niteliklere göre farklı hile sınıflandırmaları yapılmakta olup, bunlardan bir tanesi nakdi ve nakdi olmayan hileler şeklinde ifade edilmektedir. ACFE'nin 2008 yılı raporunda nakdi hilelerin nakdi olmayan hilelere oranla çok yüksek seviyede olduğu (%85) ifade edilmektedir (Bozkurt, 2009: 48). Nakit menfaat sağlama amacıyla gerçekleştirilen ve önemli bir boyut kapsayan nakit hileleri daha çok gelen nakit, çek ya da senedin kayıttan önce ya da kayda alındıktan sonra çalınması; ödemeler aracılığıyla kaynakların çalınması ve doğrudan nakit hırsızlığı olmak üzere üç şekilde uygulanmaktadır (Bozkurt, 2009: 49).

5. HİLEBAZIN NİTELİKLERİ

Hilenin ortaya çıkartılması için dikkat edilmesi gereken hususlardan en önemlisi; hile araştırmacısının, hile yapan kişinin (hilebazın) kişisel nitelikleri ve karakteristik yapıları hakkında bilgi sahibi olmasıdır. ACFE tarafından farklı yıllarda hazırlanan hile raporlarında hile yapan ya da yapmaya meyilli olan çalışanlara ait bazı istatistiki bilgilere yer verilmektedir.

ACFE 2008 Hile Raporunda, hile yapanların % 59'1'inin erkek % 40,9'unun kadın olduğu, evli olanların sayısının, bekar olanlara göre daha yüksek olduğu; düşük eğitim seviyelerinde daha sık hile yapma eğiliminin bulunduğu; ancak eğitim seviyesi yükseldikçe verilen zarar tutarında ciddi ölçüde artışlar olduğu (lise seviyesinde % 33,9 ve 100.000 USD, kolej seviyesinde % 20,8, ve 196.000 USD, üniversite seviyesinde % 34,4 ve 210.000 USD, lisansüstü seviyede ise % 10,9 ve toplam 550.000 USD tutarında) ifade edilmiştir. Hilelerin yarısından çoğu (% 52'si) 35-50 yaş arasındaki kişiler tarafından gerçekleştirilmektedir (Bozkurt, 2009: 78-85).

Yine aynı raporda zeka seviyesi yüksek olanlarda hile yapma isteğinin daha yüksek olduğu, hile yapma eğilimindeki kişilerin önemli bir kısmının diğer-

lerine göre daha egoist oldukları, kendilerini önemseyerek, diğerlerini küçük gördükleri ve daha yüksek yerlerde olmayı hakettiklerini düşündükleri, risk alma yeteneklerinin yüksek olduğu, başarısızlıktan korkmadıkları ve yakalanmayacakları düşüncesiyle hile yapmak suretiyle şanslarını denemek istedikleri ifade edilmiştir. Herhangi bir sebeple sorun yaşayan, stres altındaki kişiler (parasal sıkıntısı olan, ev ve aile yaşantıları kötü olanlar, uyuşturucu, alkol, kumar gibi bağımlılıkları bulunanlar), rahat bir yaşantı sürmek isteyenler ile pahalı zevk ve hobileri olanlardan bir kısmı, işletmede çalıştıkları pozisyonun hile yapmaya uygun olması durumunda hileli uygulamalara meyil göstermektedirler. İşine karşı meraklı ve yeni uygulamalar yaratmak için hevesli kişilerde, kurallara uymada problem yaşayanlarda ve mesai saatlerine uyum konusunda titiz davranarak işe erken gelip geç giden, hatta hiç izin kullanmayan kişilerde hile yapmış olma olasılığının daha yüksek olduğu görülmektedir. Yapılan iş karşılığı ödül, prim gibi beklentilerin olması; ya da tepe yönetim tarafından ulaşılmaması güç hedeflerin belirlenmesi ve hedefi tutturamama durumunda işi kaybetme riskinin olması (ödül ya da cezai uygulamalar) da çalışanları hile yapmaya yönelten işletme içi faktörler sayılmaktadır (Bozkurt, 2009: 85-94).

6. VERİ MADENCİLİĞİ TEKİNİĞİ

Teknolojik gelişmeler sonucu dijital veri sayısında da oldukça büyük artışlar yaşanmış, çok miktarda ve çok boyutlu karmaşık haldeki veri yığınlarını ayrıştırarak analiz yapmak için yeni yöntem ve sistemler geliştirilmesi ihtiyacı çeşitli modellemeler yapılmasını gündeme getirmiştir. Ana kütle hakkında anlamlı bilgi edinebilmek amacıyla geliştirilen uygulama ve kavramlardan biri de veri madenciliğidir. İstatistik biliminin bilgisayar teknolojisiyle bütünleşmesinin bir ürünü olan veri madenciliği 1990'lı yıllarda ortaya çıkan ve son yıllarda uygulamada yaygınlık gören kavramlardan biri olup, veri madenciliğinde büyük hacimli veri yığınlarından oluşan veritabanı sistemleri içerisinde gizli kalmış bilgilerin elde edilmesi ve bu bilgilerin karar almada kullanılması sözkonusudur.

Basit bir tanımlama yapılacak olursa veri madenciliği, büyük ölçekli veri yığınları arasından “değeri olan” bir bilgiyi elde etme işlemi olarak tanımlanabilir ve veri madenciliği sayesinde veriler arasındaki anlamlı ilişkiler ortaya koyulmak suretiyle gerektiğinde ileriye dönük çıkarımlarda bulunulabilir (Özkan, 2008: 38).

Veri madenciliğinde kullanılan modeller “Tanımlayıcı Modeller” ve “Tahmin Edici Modeller” olmak üzere ikiye ayrılır.

Tanımlayıcı modellerde amaç, karar vermeye yardımcı olabilecek mevcut veri setinde yer alan veriler arasındaki ilişkileri, bağıntıları ve davranışları belirlemektir. Tanımlayıcı modeller; “İlişki/Bağıntı-Birliktelik Analizi” ve “Kümeleme” olmak üzere iki grupta incelenmektedir.

- i. *İlişki / Bağıntı-Birliktelik Analizi*: Birliktelik, bir veri seti içerisinde öğeler arasındaki ilişkiyi ifade eder. Öğe ya da öğelerin bir veri setinde beraber bulunma durumuna ilişkin yapılan analizlerdir (Seçkin, 2011: 16).
- ii. *Kümeleme Analizi*: Kategorisi verilmemiş verilerin ön bir bilgi olmadan anlamlı gruplara atanmasıdır. Kümeleme yöntemi, birbiri ile alakalı ve benzer verileri diğer alakasız verilerden ayırarak benzerliğe dayalı olarak geliştirilen etkin bir yöntemdir (Seçkin, 2011: 16).

Tahmin edici modellerde amaç ise, geçmiş verilerden yararlanarak, gelecekle ilgili bir sonucu tahmin etmektir. Tahmin edici modeller; “İstatistiksel Tahmin Modelleri” ve Sınıflandırma” olmak üzere iki ana başlık altında incelenmektedir.

- i. *İstatistiksel Tahmin Modelleri*: Veri madenciliği çalışması esas olarak bir istatistiksel model uygulamasıdır. Veri madenciliği analizi için yaygın olarak kullanılan istatistiksel teknikler; “Regresyon Analizi”, “Diskriminant Analizi” ve “Lojistik Regresyonu” olarak adlandırılır.
- ii. *Sınıflandırma*: Verilerin kategorize edilmesinde kullanılan sınıflandırma, karmaşık veri setlerinin analiz edilmesinde kullanılan en yaygın yöntemlerden biridir. Sınıflandırma, önceden belirlenmiş kategorilere veri örneklerinin atanması şeklinde uygulanmaktadır (Seçkin, 2011: 15).

Veri madenciliğinin uygulanmasında sektör farkı gözetilmemekle birlikte geniş veri ambarlarının olduğu bankacılık, sigortacılık, sağlık, perakende satış gibi farklı alanlarda kullanılması daha yaygın ve uygundur (Silahtaroglu, 2008: 11). Veri madenciliği uygulamaları müşterilerin satın alma ürünlerinin belirlenmesi, demografik özellikleri arasındaki bağıntıların bulunması, posta kampanyalarına cevap verme oranlarının artırılması, mevcut müşterilerin elde tutulması/yeni müşteri kazanılması, pazar sepeti analizi ve risk yönetimi/dolandırıcılık saptama gibi konularda pazarlama yönetimi amaçlı kullanılmaktadır. Bunun yanı sıra veri madenciliği DNA analizi konusunda biyoloji alanında ve hastalıkların teşhisi konusunda ise tıp alanında kullanılabilir (Silahtaroglu, 2008: 11-14).

6.1. K- EN YAKIN KOMŞULUK ALGORİTMASI YÖNTEMİ

Verilerin sınıflandırılmasında kullanılan bu algoritmaya göre, sınıflandırma

sırasında dikkate alınan özelliklerden yola çıkılarak, sınıflandırılmak istenen yeni bireyin daha önce tanımlanmış olan bireylerden k tanesine yakınlığına bakılmaktadır.

k - En Yakın Komşuluk Algoritması (k - Nearest Neighbors Algorithm / k -NNA) yöntemi, sınıfları belirli olan bir örnek kümedeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu tespit etmek amacıyla kullanılır (Özkan, 2008: 117). Algoritmada sınıflandırılmak üzere alınan bir bilinmeyene en yakın k adet komşu belirlenir ve bu komşuların ait oldukları sınıflara bakarak, sınıflandırılmak istenen bilinmeyenin bu sınıflardan en yakın mesafede olduğu birine ya da birden fazlasına ataması yapılır. En sık kullanılan k değerleri veri kümesinin büyüklüğüne göre 3,5 ve 7'dir (Silahtaroglu, 2008: 65).

En yakın komşu algoritması yöntemini uygulayabilmek için belirlenmesi gereken uzaklık ölçüm yöntemi olarak genellikle "Öklid Uzaklığı" ya da "Kosinüs Benzerliği" ölçülerinden biri kullanılır¹ (Kesgin, 2007: 36). Kullanılan uzaklık ölçüsü aracılığıyla gözlem değerleri arasındaki mesafesinin hesaplanmasının ardından, sınıflamaya esas olacak en yakın k adedi dikkate alınarak değerlendirme yapmak suretiyle yargıya ulaşılır (Silahtaroglu, 2008: 65). Örneğin yeni bir verinin $k = 3$ için sınıflandırılmak isteniyor olması durumunda, daha önce tanımlanarak sınıflandırılmış komşularından en yakın 3 tanesi dikkate alınır. En yakın gözlem değerine sahip komşular hangi sınıfa dahil ise, bilinmeyen elemanın da o sınıfa dahil olması gerektiği kabul edilir.

Öklid Uzaklık Ölçütü

Sınıflandırılmak istenen örnek ile sınıflarda bulunan örnekler arasındaki yakınlık Öklid uzaklığı hesaplanarak bulunur. Sınıflarda bulunan örnekleri $X = (X_1, X_2, \dots, X_n)$ ve sınıflandırılmak istenen veriyi de $Y = (Y_1, Y_2, \dots, Y_n)$ temsil ettiğini düşünülerek iki vektör arasındaki Öklid uzaklığı aşağıdaki formül yardımı ile hesaplanır.

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

¹ Çalışmada veriler arasındaki mesafenin hesaplanmasında Öklid uzaklık ölçütü kullanılmıştır.

7. ÇALIŞAN HİLELERİNİN ÖNLENMESİNDE K-EN YAKIN KOMŞU ALGORİT-MASININ KULLANILMASI

Hile insanlık tarihinin başlangıcı kadar eski olup, hileli uygulamalarla farklı zamanlarda farklı şekillerde karşılaşılmaktadır. Bilim ve teknolojiye yaşanan gelişmeler, hileli uygulamaların sayısının ve tutarının da artmasına sebep olmuş, bu artışlar hile türlerinde ve hile kontrol yöntemlerinde de farklılaşmalara yol açmıştır. Hile ile mücadelede en önemli handikap, hilenin oluşumundan önce tespitidir. Hileyi önceden tespit etmek çok zor, hatta çoğu zaman imkansız gibidir. Her geçen gün gelişen teknolojinin de yardımıyla yeni bir hile uygulama yöntemiyle karşı karşıya kalınan günümüz çalışma ortamında, çözüm yine mevcut veriler ve gelişen teknoloji yardımıyla olacaktır. ACFE Raporunda yer alan bilgilere göre, hile yapmaya meyilli bir işletme çalışanının taşıdığı özellikler; “erkek, evli, iyi eğitilmiş, zeki, egoist, meraklı, risk alabilen, kural tanımaz, sıkı çalışan, stres altında, iyi yaşamayı ve bol para harcamayı seven, bulunduğu iş pozisyonuna oranla lüks yaşantı içinde olan, parasal sıkıntılar yaşayan, kötü alışkanlıkları olan, yönetici konumunda çalışan, hırslı, kişisel kazanç beklentisi olan ve işletme dışındaki kişilerle yakın ve sıkı ilişkiler içerisinde bulunan” olarak sıralanmaktadır (Bozkurt, 2009: 94).

Gerçekte hileli uygulamalarda bulunan işletme çalışanları, yaptıkları hileli uygulamaların ortaya çıkacağı kaygısı ile işletmede olabildiğince fazla zaman geçirmek, işini ve masasını boş bırakmamak, yerine bir başkasının geçmesi/ oturması olasılığını yok etmek amacıyla mesai saatlerine erken başlayıp geç bitirmek, yıllık iznini kullanmamak, hastalık, doğum, ölüm gibi durumlarda dahi işe gelmek, kendisinin gelme zorunluluğunun olmadığı, ancak diğer çalışanların işletmeye geldiği günlerde dahi işe gelmek gibi titiz görünen bir çalışma disiplini sergilemektedir.

Bu çalışmada, çalışan personel ya da yeni bir işe başvuran adayların hileye yakınlıklarını (hileye başvurma eğilimlerini) ölçmek için, sınıflandırma işleminde en çok kullanılan makine öğrenmesi algoritmalarından, *k*-en yakın komşuluk algoritmasını kullanmak suretiyle önceden belirlenmiş bazı özellik ya da ölçütler dikkate alınarak (cinsiyet, yaş, eğitim, gelir vb.) kişinin profilinin hangi kategoriye daha yakın oldukları, ya da hangi sınıf içerisinde yer alacakları belirlenmeye çalışılmıştır. Öncelikle çalışanların her birine ait kişisel özellikler analiz edilerek, kişilerin profilleri ortaya çıkartılmalı ve kişilerin özelliklerini, tutum ve davranışlarını içerecek şekilde mevcut çalışanlar gruplandırılmalı, daha sonra da ölçüm yapılması amaçlanan personele ait kişisel özellikleri içeren veriler dikkate alınarak kendisinin hangi gruba

daha yakın olduğunu belirleyecek şekilde Öklid uzaklık ölçüsü kullanılmak suretiyle uzaklık hesaplamaları yapılmış ve bilinmeyen örnekler için sınıflama yapılmıştır.

Örnek Olay 1: A Şirketi, İnsan Kaynakları Departmanı mevcut çalışanlarının hile yapıp yapmadıklarını araştırıyor olsun: Geçmiş verilerden şirket çalışanlarından bazılarının normal çalışma hayatlarını hileye başvurmadan sürdürdükleri, bazılarının hile yaptıklarının tespit edilerek kanuni takibe alındıkları görülmektedir. İnsan kaynakları departmanı, tespit edilen normal çalışan ve hile yapan çalışanların geçmiş verilerinden hareketle, mevcut çalışanların özelliklerine bakarak kendilerinin hile yapıp yapmayacaklarını tahmin etmeye bir faaliyet planlanmıştır. Literatürdeki çalışmalardan, cinsiyet, medeni hal, eğitim gibi demografik özelliklerin hilebazın tanınmasında etkili unsurlar arasında yer aldığı anlaşılmaktadır. Örnekte veriler; yaş, yıllık fazla mesai saati, yıllık izin kullanım süresi, yıllık geliri, birikim, miras dışı yollarla sahip olduğu ev sayısı olarak tanımlanmaktadır. Hilebazın tanımlanması için oluşturulacak algoritmada çalışanlar, normal çalışan ve kanuni takibe düşmüş çalışan olmak üzere iki sınıfa ayrılır. Tablo 1’de üzerinde çalışılacak örnek veri seti verilmiştir.

Tablo 1. *k* - En Yakın Komşuluk Algoritması Yardımıyla Sınıflandırma İçin Kullanılacak Örnek Veri Seti

Hile Durumu	Yaş	Yıllık Fazla Mesai (Saat)	Yıllık İzin Kull. (gün)	Yıllık Gelir	Birikim	Ev Sayısı
Normal	28	100	20	15.000	3.000	1
Normal	25	150	15	10.000	1.500	1
Normal	30	175	25	18.000	5.000	0
Kanuni Takip	35	250	15	27.000	10.000	1
Kanuni Takip	45	275	10	35.000	18.000	2
Kanuni Takip	46	280	5	45.000	25.000	3

Çalışan X=(yaş= “34”, yıllık fazla mesai süresi= “280 (saat)”, yıllık izin kullanım süresi= “5 (gün)”, yıllık geliri= “28000”, birikim= “10000”, sahip olduğu ev sayısı= “2”) örneği sınıflandırmak istenilen bilinmeyen örnek olsun. Bu niteliklere sahip çalışanın hile yapıp yapmadığı/ hile yapmaya meyilli olup olmadığını tahmine yönelik çalışmada:

Bilinmeyen örnek X’in tüm sınıflardaki her elemanla arasındaki Öklid uzak-

liğı hesaplanır ve arařtırmacının inisiyatifine gre bir *k* deęeri seilir. Seilen bu deęere gre en yakın klid uzaklıkları tespit edilerek ait olduklarını sınıfa gre bilinmeyen (arařtırılan rneęin) hangi sınıfa atanacağı belirlenir. rnekte *k* deęeri 3 seilmiş olup, *X* bilinmeyen rneęi ile en kçük ilk 3 deęere sahip komřuları tespit edilerek bunlar arasında en yakın bulunduęu sınıfa ataması yapılmıřtır. Ancak, deęiřkenlerin birbiri ile kıyaslanabilmesi ve etki dzeylerinin deęerlendirilebilmesi iin Tablo 1’de yer alan deęiřkenler SPSS Paket Programı aracılıęı ile

$$Z = \frac{X - \mu}{\sigma}$$

forml uygulanmak suretiyle normalleřtirilmiř ve sınıflandırma iřlemi, bu normalleřtirilmiř veri seti zerinde gerekleřtirilmiřtir. Bu bilgilere gre normalleřtirilmiř veri seti zerinde hesaplanan klid uzaklıkları Tablo 2’de yer almaktadır.

Tablo 2. *k* - En Yakın Komřuluk Algoritması Yardımıyla alıřan *X* iin Normalleřtirilmiř Veri Seti ile Sınıflandırma

Hile Durumu	Yař	Yıllık Fazla Mesai	Yıllık İzin Kull.	Yıllık Gelir	Birikim	Ev	klid Uzaklıęı
Normal	-,90820	-,1,42575	,86667	-,72532	-,97704	-,39318	$2\sqrt{[(-0,90820-(-0,10422))^2+(-1,42575-0,83867)^2+(0,86667-(-0,93333))^2+(-0,72532-0,40618)^2+(-0,97704-(-0,192292))^2+(-0,39318-0,49147)^2]} = 2,907949$
Normal	-,1,31019	-,79674	,26667	-,1,16052	-,1,14507	-,39318	$2\sqrt{[(-1,31019-(-0,10422))^2+(-0,79674-0,83867)^2+(0,26667-(-0,93333))^2+(-1,16052-0,40618)^2+(-1,14507-(-0,192292))^2+(-0,39318-0,49147)^2]} = 2,952475$
Normal	-,64021	-,48224	1,46667	-,46421	-,75301	-,1,27783	$2\sqrt{[(-0,64021-(-0,10422))^2+(-0,48224-0,83867)^2+(1,46667-(-0,93333))^2+(-0,46421-0,40618)^2+(-0,75301-(-0,192292))^2+(-1,27783-0,49147)^2]} = 2,553085$
Kanuni Takip	,02978	,46127	,26667	,31914	-,19292	-,39318	$2\sqrt{[(0,02978-(-0,10422))^2+(0,46127-0,83867)^2+(0,26667-(-0,93333))^2+(0,31914-0,40618)^2+(-0,19292-(-0,192292))^2+(-0,39318-0,49147)^2]} = 1,181103$
Kanuni Takip	1,36974	,77577	-,33333	1,01545	,70322	,49147	$2\sqrt{[(1,36974-(-0,10422))^2+(0,77577-0,83867)^2+(-0,33333-(-0,93333))^2+(1,01545-0,40618)^2+(0,70322-(-0,192292))^2+(0,49147-0,49147)^2]} = 2,226032$
Kanuni Takip	1,50374	,83867	-,93333	1,88584	1,48734	1,37612	$2\sqrt{[(1,50374-(-0,10422))^2+(0,83867-0,83867)^2+(-0,93333-(-0,93333))^2+(1,88584-0,40618)^2+(1,48734-(-0,192292))^2+(1,37612-0,49147)^2]} = 3,444594$

$X=(\text{yaş}= "34", \text{yıllık fazla mesai süresi}= "280 (\text{saat})", \text{yıllık izin kullanım süresi}= "5 (\text{gün})", \text{yıllık geliri}= "28000", \text{birikim}= "10000", \text{sahip olduğu ev sayısı}= "2")$ olup, *değişkenler için normalleştirilmiş değerler* $X=(\text{yaş}= "-0,10422", \text{yıllık fazla mesai}= "0,83867", \text{yıllık izin kullanım}= "-0,93333", \text{yıllık gelir}= "0,40618", \text{birikim}= "-0,19292", \text{sahip olduğu ev}= "0,49147")$ şeklindedir.

Yapılan çalışma sonucu elde edilen ilk 3 değer sırasıyla 1,181103 (kanuni takip), 2,226032 (kanuni takip), 2,553085 (normal) değerleridir. Bu değerlere bakıldığında X bilinmeyen örneğinin kanuni takip sınıfına daha yakın olduğu görülmektedir. Bu çalışanın, k - en yakın komşuluk algoritması çıktısına göre hile yapma eğiliminde olabileceği anlaşılmakta olup, bu kişinin ve çalışmalarının incelemeye alınması uygun olacaktır.

Örnek Olay 2 ve 3 için A Şirketi, İnsan Kaynakları Departmanının işe yeni eleman alımı başvurularını değerlendirirken adayların hile yapma eğilimlerini araştırdıklarını varsayalım. Geçmiş çalışanlarının bazıları, normal çalışma hayatlarını hileye başvurmadan sürdürmüş bazılarıysa hile yapmış ve kanuni takibe alınmış olsun. Örneklerde, adayların başvuru formlarından kendi ifadeleriyle alınan bazı özellikleri ile işletmenin geçmiş tecrübelerinden elde edilen değerlendirme kriterlerini karşılaştırmak suretiyle yeni işe alınacak personelin hile yapanlar sınıfına dahil olup olmayacakları, ya da hile yapma eğiliminde olup olmadıkları yönünde tahminde bulunulacaktır. Veri seti, örnek olay 1 ile aynı olup aynı sınıflandırma algoritmasından yararlanılacaktır.

Örnek Olay 2: Başvurulardan Aday $Y = (\text{yaş}= "30", \text{yıllık fazla mesai süresi}= "110 (\text{saat})", \text{yıllık izin kullanım süresi}= "20 (\text{gün})", \text{yıllık geliri}= "12000", \text{birikim}= "8000", \text{sahip olduğu ev sayısı}= "0")$ özelliklerine sahiptir. Aday Z ise $(\text{yaş}= "40", \text{yıllık fazla mesai süresi}= "300 (\text{saat})", \text{yıllık izin kullanım süresi}= "0 (\text{gün})", \text{yıllık geliri}= "20000", \text{birikim}= "25000", \text{sahip olduğu ev sayısı}= "3")$ özelliklerini taşıdığını ifade etmiştir. Sınıflandırmak istenilen adaylar Y ve Z 'nin hile yapıp yapmadığı/ hile yapmaya meyilli olup olmadığını tahmine yönelik çalışmada; bilinmeyen örnekler olan Aday Y ve Aday Z 'in tüm sınıflardaki her elemanla aralarındaki Öklid uzaklıkları hesaplanır. Bir önceki örnekte olduğu gibi k değeri olarak yine 3 seçilmiş olup, Y ve Z bilinmeyen örnekleri ile en küçük ilk 3 değere sahip komşuları tespit edilecek ve adaylar bunlar arasında en yakın mesafenin bulunduğu sınıfa atanacaktır. Bu bilgilere göre Aday Y ve Z için, Tablo 3 ve Tablo 4'te yer alan Öklid uzaklıkları elde edilmiştir.

Tablo 3. k- En Yakın Komşuluk Algoritması Yardımıyla Aday Y için Normalleştirilmiş Veri Seti İle Sınıflandırma

Hile Durumu	Yaş	Yıllık Fazla Mesai	Yıllık İzin Kull.	Yıllık Gelir	Birikim	Ev	Öklid Uzaklığı
Normal	-90820	-1,42575	,86667	-72532	-97704	-39318	$\sqrt{((-0,90820-(-0,64021))^2+(-1,42575-(-1,29994))^2+(0,86667-0,86667)^2+(-0,72532-(-0,98644))^2+(-0,97704-(-0,41695))^2+(-0,39318-(-1,27783))^2)}$ = 1,118989
Normal	-1,31019	-79674	,26667	-1,16052	-1,14507	-39318	$\sqrt{((-1,31019-(-0,64021))^2+(-0,79674-(-1,29994))^2+(0,26667-0,86667)^2+(-1,16052-(-0,98644))^2+(-1,14507-(-0,41695))^2+(-0,39318-(-1,27783))^2)}$ = 1,550855
Normal	-64021	-48224	1,46667	-46421	-75301	-1,27783	$\sqrt{((-0,64021-(-0,64021))^2+(-0,48224-(-1,29994))^2+(1,46667-0,86667)^2+(-0,46421-(-0,98644))^2+(-0,75301-(-0,41695))^2+(-1,27783-(-1,27783))^2)}$ = 1,189241
Kanuni Takip	,02978	,46127	,26667	,31914	-19292	-39318	$\sqrt{((0,02978-(-0,64021))^2+(0,46127-(-1,29994))^2+(0,26667-0,86667)^2+(0,31914-(-0,98644))^2+(-0,19292-(-0,41695))^2+(-0,39318-(-1,27783))^2)}$ = 2,539307
Kanuni Takip	1,36974	,77577	-33333	1,01545	,70322	,49147	$\sqrt{((1,36974-(-0,64021))^2+(0,77577-(-1,29994))^2+(-0,33333-0,86667)^2+(1,01545-(-0,98644))^2+(0,70322-(-0,41695))^2+(0,49147-(-1,27783))^2)}$ = 4,263946
Kanuni Takip	1,50374	,83867	-93333	1,88584	1,48734	1,37612	$\sqrt{((1,50374-(-0,64021))^2+(0,83867-(-1,29994))^2+(-0,93333-0,86667)^2+(1,88584-(-0,98644))^2+(1,48734-(-0,41695))^2+(1,37612-(-1,27783))^2)}$ = 5,597315

$Y=(\text{yaş}="30", \text{yıllık fazla mesai süresi}="110 (\text{saat})", \text{yıllık izin kullanım süresi}="20 (\text{gün})", \text{yıllık gelir}="12000", \text{birikim}="8000", \text{sahip olduğu ev sayısı}="0")$ olup, değişkenlere ait veriler normalleştirildiğinde $Y=(\text{yaş}=" -0,64021", \text{yıllık fazla mesai}=" -1,29994", \text{yıllık izin kullanım}=" 0,86667", \text{yıllık gelir}=" -0,98644", \text{birikim}=" -0,41695", \text{sahip olduğu ev}=" -1,27783")$ değerlerine ulaşılmaktadır.

En küçük ilk 3 değere bakıldığında sırasıyla 1,118989 (normal), 1,189241(normal), 1,550855 (normal) şeklindedir. Bu değerlerin tümü normal sınıfta yer almaktadır. Dolayısıyla Y bilinmeyen örneği normal sınıfa atanır. Bu adayın, k - en yakın komşuluk algoritması çıktısına göre hile yapma eğiliminde olmadığı düşünülerek işe alınması uygun görülebilir.

Tablo 4. k- En Yakın Komşuluk Algoritması Yardımıyla Aday Z İçin Normalleştirilmiş Veri Seti İle Sınıflandırma

Hile Durumu	Yaş	Yıllık Fazla Mesai	Yıllık İzin Kull.	Yıllık Gelir	Birikim	Ev	Öklid Uzaklığı
Normal	-,90820	-1,42575	,86667	-,72532	-,97704	-,39318	$^2\sqrt{((-0,90820-0,69976)^2+(-1,42575-1,09028)^2+(0,86667-(-1,53333))^2+(-0,72532-(-0,29013))^2+(-0,97704-1,48734)^2+(-0,39318-1,37612)^2)}$ = 4,906009
Normal	-1,31019	-,79674	,26667	-1,16052	-1,14507	-,39318	$^2\sqrt{((-1,31019-0,69976)^2+(-0,79674-1,09028)^2+(0,26667-(-1,53333))^2+(-1,16052-(-0,29013))^2+(-1,14507-1,48734)^2+(-0,39318-1,37612)^2)}$ = 4,653851
Normal	-,64021	-,48224	1,46667	-,46421	-,75301	-1,27783	$^2\sqrt{((-0,64021-0,69976)^2+(-0,48224-1,09028)^2+(1,46667-(-1,53333))^2+(-0,46421-(-0,29013))^2+(-0,75301-1,48734)^2+(-1,27783-(-1,27783))^2)}$ = 5,035997
Kanuni Takip	,02978	,46127	,26667	,31914	-,19292	-,39318	$^2\sqrt{((0,02978-0,69976)^2+(0,46127-1,09028)^2+(0,26667-(-1,53333))^2+(0,31914-(-0,29013))^2+(-0,19292-1,48734)^2+(-0,39318-1,37612)^2)}$ = 3,226365
Kanuni Takip	1,36974	,77577	-,33333	1,01545	,70322	,49147	$^2\sqrt{((1,36974-0,69976)^2+(0,77577-1,09028)^2+(-0,33333-(-1,53333))^2+(1,01545-(-0,29013))^2+(0,70322-1,48734)^2+(0,49147-1,37612)^2)}$ = 2,256054
Kanuni Takip	1,50374	,83867	-,93333	1,88584	1,48734	1,37612	$^2\sqrt{((1,50374-0,69976)^2+(0,83867-1,09028)^2+(-0,93333-(-1,53333))^2+(1,88584-(-0,29013))^2+(1,48734-1,48734)^2+(1,37612-1,37612)^2)}$ = 2,409261

$Z=(\text{yaş}= "40", \text{yıllık fazla mesai süresi}= "300 (\text{saat})", \text{yıllık izin kullanım süresi}= "0 (\text{gün})", \text{yıllık geliri}= "20000", \text{birikim}= "25000", \text{sahip olduğu ev sayısı}= "3")$ olup değişkenlere ait veriler normalleştirildiğinde $Z=(\text{yaş}= "0,69976", \text{yıllık fazla mesai}= "1,09028", \text{yıllık izin kullanım}= "-1,53333", \text{yıllık gelir}= "-0,29013", \text{birikim}= "1,48734", \text{sahip olduğu ev}= "1,37612")$ değerlerine ulaşılmaktadır.

En küçük ilk 3 değere bakıldığında sırasıyla 2,256054 (kanuni takip), 2,409261 (kanuni takip), 3,226365 (kanuni takip) şeklindedir. Bu değerlerin biri normal sınıfta diğerleri kanuni takip sınıfında yer almaktadır. Dolayısıyla Z bilinmeyen örneği kanuni takip sınıfına atanır. Bu adayın, k - en yakın komşuluk algoritması çıktısına göre hile yapma eğiliminde olduğu düşünülerek işe alınmaması yönünde karar verilebilir.

8. SONUÇ

Teknolojik gelişmeler ve küreselleşme muhasebe uygulamalarında da pek çok değişikliklerin ortaya çıkmasına sebep olurken, bu gelişmeler hata ve hilelerin yapısında ve içeriğinde de önemli değişiklikleri gündeme getirmiştir. Son yıllarda yapılan araştırmalarda bu gelişmelere paralel olarak, hata ve hileli uygulamalar için ödenen bedellerde de oldukça yüksek artışlar olduğuna dikkat çekilmektedir. Bu gelişmeler göz önüne alındığında finansal tablo kullanıcıları için finansal tablolarda yer alan bilgilerin gerçekliği ve güvenilirliği daha da önem kazanmakta olup muhasebede üretilecek finansal bilgilerin doğru, dürüst ve güvenilir bilgiler olabilmesi için dürüst ve güvenilir kişilerin bulunduğu çalışma ortamlarında üretilmiş olmaları hususu ayrı bir gerçekliktir. Bu açıdan ele alındığında finansal tablo kullanıcılarında, işletme çalışanlarının hileli uygulamalarda bulunup/bulunmadıkları (bulunup/bulunmayacakları) yönünde bazı soru işaretleri oluşmaktadır.

Bu çalışmada sayısal bir takım verilerden hareketle davranışsal/sosyal bazı olayların tanımlanabileceği, açıklanabileceği ve çıkan sonuçlar doğrultusunda davranış kalıpları sergilenebileceği düşüncesinden hareket edilmiştir. Bu düşünceden yola çıkılarak çalışan hilelerinin tanımlanması ve engellenmesinde veri madenciliği tekniğinin uygulama yöntemlerinden biri olan k -en yakın komşuluk algoritmasından yararlanılabileceği ileri sürülmektedir.

İşletmelerin geçmiş yıllardaki tecrübelerden hareketle mevcut çalışan ya da çalışan adayları için veri madenciliği tekniklerinden biri olan k -en yakın komşuluk algoritması yönteminin uygulanması sonucu çalışanlarının ya da çalışan adaylarının kişisel menfaatleri için hileli uygulamalar gerçekleştirmeye yakın olup olmadıkları yönünde tahminde bulunulabileceği düşünülmektedir. Bu düşünceden yola çıkarak çalışanların/adayların belli nitelikleri dikkate alınmak suretiyle, k -en yakın komşuluk algoritmasını kullanarak daha önce hile yaptığı tespit edilmiş kişilere yakın olup olmadıkları, dolayısıyla kendi verilerinin hile içeren göstergelere yakın olup olmadıkları araştırılarak kendileri hakkında bir değer yargısına ulaşılabileceği varsayılmaktadır. Bu varsayımdan hareketle geliştirilen örneklerde aynı göstergeler/değişkenler kullanılmak suretiyle biri mevcut çalışan, diğer ikisi çalışan adayı üç farklı kişi için değerlendirme yapmak suretiyle bu kişilerin hilebaz sınıfına dahil olup olmayacakları konusunda tahminde bulunulmuştur.

Çalışmada; yaş, yıllık fazla mesai saati, yıllık izin kullanım süresi, yıllık ge-

lir, birikim, miras dışı yollarla sahip olunan ev sayısı gibi gözlem değerleri dikkate alınarak hesaplamalar yapılmış ve araştırılan kişiler hakkında tahmini yargılara ulaşılmıştır. Değişkenler arasında, yaşam standartlarına ilişkin sinema, tiyatro, konser gibi ücret mukabili sosyal etkinliklere katılım sıklığı, sahip olunan araç sayısı ve modeli, varsa çocuklarının öğrenim gördükleri okul türü, gibi harcama niteliklerin alınması suretiyle kişilerin gelir-harcama ilişkilerinden hareketle hile yapma olasılıkları değerlendirilerek daha ayrıntılı sınıflandırma yapılmasına imkan sağlanacağı düşünülmektedir.

KAYNAKÇA

AKYEL, Nermin (2009), "Adli Muhasebeciliğin Altyapı Bileşenleri Ve Türkiye Penceresinden İncelenmesi" Yayımlanmamış Doktora Tezi, Sakarya Üniversitesi, Sosyal Bilimler Enstitüsü, Sakarya.

BOZKURT, Nejat (2009), İşletmelerin Kara Deliği Hile Çalışan Hileleri, Alfa Basım Bayım Dağıtım Ltd.Şti., İstanbul.

ÇUBUKÇU, Sezen (2009), "Muhasebe Hilelerini Ortaya Çıkartmada Benford Modeli'nin İlk İki Basamak Yaklaşımı İle Kullanılması" Muhasebe Bilim Dünyası Dergisi, Muhasebe Öğretim Üyeleri Bilim ve Dayanışma Vakfı, Sayı:2009/3, 113-142, Ankara.

DEMİR, Volkan ve Oğuzhan BAHADIR (2007), "Muhasebe Manipülasyonu-Yöntemler ve Teknikler", Mali Çözüm Dergisi, İSMMMO Yayını, Yıl:17, Sayı: 84, 103-119, İstanbul.

DOYRANGÖL, Nuran, (2007) İşletmelerde İç Kontrol ve İç Denetim, İstanbul, Deloitte Academy.

EMİR, Murat (2008) Hile Denetimi, Mali Çözüm, İstanbul Serbest Muhasebeci Mali Müşavirler Odası Yayınları, Sayı:86, 109-121, İstanbul.

ERDOĞAN, Melih (2006), Denetim, Kavramsal ve Teknolojik Yapı, 3. Bası, Ankara

GRAYCAR, Adam ve Russel SMİTH (2002), Identifying and Responding to Corporate Fraud in the 21 Century, Australian Institute of Management Sydney 20 March, Australian Institute of Criminology.

HİLLISON, W., C. PACİNİ ve D. SİNASON (1999), " The Internal Auditor as Fraud Buster", Managerial Auditing Journal, Vol:14-7.

KARABINAR, Selahattin ve Nermin AKYEL (2011), "Denetimde Standart Dışı Davranışların Ortaya Çıkarılmasında Veri Madenciliğinin Rolü Ve Önemi", 3. Balkanlarda Sosyal Bilimler Kongresi, Saraybosna.

- KAVAL, Hasan (2005), Muhasebe Denetimi, Gazi Kitabevi, Ankara.
- KESGİN, Fatih (2007), “Türkçe Metinler İçin Konu Belirleme Sistemi”, İstanbul Teknik Üniversitesi, Basılmamış Yüksek Lisans Bitirme Projesi, İstanbul.
- KOYUNCUGİL, Serhan ve Nermin ÖZGÜLBAŞ (2009), Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, Bilişim Teknolojileri Dergisi, Cilt:2, Sayı:2, ss:21-32.
- ÖZKAN, Yalçın (2008), Veri Madenciliği Yöntemleri, Papatya Yayıncılık Eğitim, İstanbul.
- ÖZŞAHİN, Gülfer (2000), “Bağımsız Denetçinin Hata ve Hilelere İlişkin Sorumluluğu” SPK Muhasebe Standartları Dairesi Yeterlilik Etüdü, Ankara.
- SEÇKİN, Keziban. (2011), “Metin Madenciliğinde Kullanılan Yöntemlerin Karşılaştırılması: Siyasi Parti Liderlerinin Grup Genel Toplantı Konuşmaları İle Uygulama”, Sakarya Üniversitesi Sosyal Bilimler Enstitüsü Yayınlanmamış Yüksek Lisans Tezi, Sakarya.
- SEVİLENGÜL, Orhan (2003), Genel Muhasebe, Ankara, Gazi Kitabevi, 11. Bası
- SİLAHTAROĞLU, Gökhan (2008), Kavram ve Algoritmalarıyla Temel Veri Madenciliği, Papatya Yayıncılık Eğitim, İstanbul.
- SPK, Sermaye Piyasasında Bağımsız Denetim Standartları hakkında Tebliğ, Seri X, No: 22, Kısım:6, Madde: 4, Resmi Gazete 26241; 12.06.2006.

