

**VERİ MADENCİLİĞİ VE MUHASEBEDE KULLANIMI:
MUHASEBE DOLANDIRICILIĞININ TESPİTİNDE EN ÇOK KULLANILAN
YÖNTEMLER**

Fatih Ömür BİNİCİ¹

ÖZET

Günümüzde teknolojinin yaygınlaşması ve gelişmesiyle birlikte işletmeler, analiz edilmesi ve yorumlanması gereken muazzam miktarda finansal veri üretmektedir. Üretilen bu büyük miktardaki verilerin analizi, geleneksel yöntemlerle mümkün olmamaktadır. Bu nedenle işletmeler, geleneksel yöntemler yerine Veri Madenciliği (VM) yöntemlerine yönelmiştir. VM, büyük veri tabanlarında istatistiksel olarak güvenilir, daha önce bilinmeyen, eyleme dönüştürülebilir içgörüler ve ilginç desenler elde etmek için istatistik, matematik, yapay zeka ve makine öğrenmesi gibi teknikleri kullanan kapsamlı bir süreçtir. İşletmelerde VM'nin kullanım alanları çok çeşitli olmakla birlikte, yoğun olarak muhasebe ve finans alanında dolandırıcılık faaliyetlerinin tespitinde kullanılmaktadır. Bu yoğun kullanım nedeniyle, verilerin analizinde seçilecek yöntem ve uygulamaların belirlenmesi büyük önem taşımaktadır.

Bu doğrultuda çalışmada, muhasebedeki veri madenciliği uygulamalarına ilişkin alan yazını incelemek ve muhasebe dolandırıcılığının tespiti için en yaygın kullanılan veri madenciliği tekniklerine dair kavramsal bir çerçeve sunmak amaçlanmıştır. Belirlenen amaç doğrultusunda, alan yazınında yer alan sistematik literatür taramaları birleştirilerek birincil kaynaklardan toplanan veriler “Şemsiye İncelemesi (Umbrella Review)” yöntemiyle analiz edilmiştir. Yapılan analiz sonucunda, VM yöntemlerinden “Regresyon (Logit/Logistic/Probit)” yönteminin, muhasebe dolandırıcılığının tespitinde en çok kullanılan yöntem olduğu sonucuna ulaşılmıştır. En az kullanılan yöntemlerin ise Otokodlayıcı (Autoencoder, AE), Azaltılmış Hata Budaması (Reduced Error Pruning-REP) ve Stokastik Gradyan İnişi (Stochastic Gradient Descent, SGD) olduğu tespit edilmiştir.

Anahtar Kelimeler: Muhasebe, Veri Madenciliği, Muhasebe Dolandırıcılığı

¹Ağrı İbrahim Çeçen Üniversitesi, Eleşkirt MYO, fobinici@agri.edu.tr, 0000-0001-6147-1955

DATA MINING AND ITS USE IN ACCOUNTING: THE MOST COMMONLY USED METHODS IN DETECTING ACCOUNTING FRAUD

ABSTRACT

In today's world, businesses generate large amounts of financial data that must be analyzed and interpreted due to technological advances and proliferation. Traditional methods cannot analyze this massive data set. As a result, businesses have turned to data mining (DM) techniques. DM is an extensive process that employs methods such as statistics, mathematics, artificial intelligence, and machine learning to extract reliable, previously unknown insights and patterns from large databases, informing decision-making. While DM has various applications in business, it is especially prevalent in accounting and finance for fraud detection. Due to this widespread use, selecting the right data analysis methods and applications is crucial.

Against this backdrop, the present study aims to review the literature on data mining applications in accounting and to present a conceptual framework for the most commonly used data mining techniques for detecting accounting fraud. To achieve this, a systematic review of the literature was conducted, and data collected from primary sources were analyzed using the 'umbrella review' method. The analysis found that the 'Regression (Logit/Logistic/Probit)' method is the most frequently used for this purpose. The least common methods identified were Autoencoder (AE), Reduced Error Pruning (REP), and Stochastic Gradient Descent (SGD).

Keywords: Accounting, Data Mining, Accounting Fraud

GİRİŞ

İşletmelerin ham verileri stratejik bir varlığa dönüştürülmesini sağlayan günümüz akıllı iş analitiği ve karar destek sistemlerinin en önemli paradigmalarından biri olarak öne çıkan veri madenciliği (VM), kurumsal geleceği şekillendiren eğilimleri tahmin etmede, yöneticilerin daha isabetli kararlar almasını desteklemede ve nihayetinde işletmelerin rekabet gücünü artırmada kullanılan bir yönetim aracıdır (Amani ve Fadlalla, 2017: 32). Büyük verileri anlamlı, işlenebilir ve değerli bilgilere dönüştürmek için kullanılan bu temel teknoloji (Chen, Sakaguchi ve Frolick, 2000: 66), işletmelere yalnızca geçmişe dair içgörü sunmakla kalmaz, aynı zamanda onları geleceğe hazırlayan yenilikçi bir yönetim felsefesi olarak da tanımlanır (Karami, Baber ve Ojala, 2022: 2).

Veri madenciliği veya bir diğer adıyla bilgi keşfi, büyük veri tabanlarında istatistiksel olarak güvenilir, daha önce bilinmeyen, eyleme dönüştürülebilir içgörüler ve ilginç desenler elde etmek için istatistiksel, matematiksel, yapay zeka ve makine öğrenmesi gibi teknikleri kullanan kapsamlı bir süreçtir (Mraovic, 2008: 439; Sharma ve Panigrahi, 2013: 38; Liu ve Vasarhelyi, 2014: 10).

Günümüzde şirketler, VM sayesinde büyük verileri analiz ederek stratejik kararlar almakta, operasyonel süreçleri ve kampanyaları yönetmekte ve tüm operasyonlar için etkili ve verimli çözümler geliştirmektedir (Chen, Sakaguchi ve Frolick, 2000: 66). VM hem özel sektörde hem de kamu sektöründe çok sayıda pratik uygulamaya olanak tanıyarak verilere sentetik bir yaklaşım kazandırmakta ve muhasebe gibi geleneksel disiplinlerde yeni bir veri organizasyonu modelinin uygulanmasına imkan tanımaktadır (Mraović, 2008: 349). İşletmelerin analiz edilmesi ve yorumlanması gereken muazzam miktarda finansal veri üretmesi nedeniyle, VM muhasebe mesleğinde de giderek daha yaygın hale gelmektedir (Theodorakopoulos, Thanasis ve Halkiopoulos, 2024: 1201).

Muhasebede büyük veri kullanımı, muhasebecilerin finansal veriler hakkında derinlemesine fikir edinmeleri ve daha iyi kararlar almaları için yeni fırsatlar sunmaktadır (Farooq, 2023: 947). Gelişmiş analitik ve makine öğrenmesi araçlarının yardımıyla muhasebeciler artık büyük veri kümelerini gerçek zamanlı olarak analiz edebilmekte, kalıpları ve eğilimleri belirleyebilmekte ve gelecekteki finansal performans hakkında tahminlerde bulunabilmektedir (Lee ve Siau, 2001: 41). Bununla birlikte muhasebede harcama modellerinin izlemesi, anormalliklerin belirlenmesi, finansal düzenlemelere uyumu sağlama, mali hesap verebilirlik, finansal raporlama, iç denetim, maliyet yönetimi ve dolandırıcılık tespiti gibi alanlar da veri madenciliği uygulamalarının kullanımı yaygındır (Kumar vd., 2020; Ngai vd., 2011: 544). Veri madenciliği teknikleri, büyük veri hacimleri ve karmaşık finansal verilerle başa çıkmak konusunda büyük kolaylık sağlamaktadır (Sharma ve Panigrahi, 2013: 37). Özellikle mevcut ekonomik senaryoda yaşanan finansal muhasebe dolandırıcılığındaki artışla birlikte, finansal muhasebe dolandırıcılığı tespiti önemli bir çalışma alanı haline gelmiştir (Mraović, 2008: 450). Bu doğrultuda çalışmanın amacı, temel bir iş disiplini olan muhasebe ile etkili bilgi sistemleri teknolojilerinden biri olan veri madenciliğinin kesişim noktasındaki durumu değerlendirmektir. Çalışmada, muhasebede veri madenciliği uygulamalarına ilişkin alan yazımını incelemek ve finansal muhasebe dolandırıcılığının tespiti için en yaygın olarak kullanılan veri madenciliği tekniklerine dair kavramsal bir çerçeve sunulması amaçlanmıştır. Çalışmanın, alandaki boşluğa katkı sunması, gelecekteki araştırmalar için potansiyel fırsatları ortaya koyması ve hem akademisyenler hem de uygulayıcılar için çeşitli veri madenciliği teknikleri hakkında yararlı bir bilgiler sağlaması beklenmektedir. Ayrıca finansal muhasebe dolandırıcılığının tespitine uygulanabilir veri madenciliği tekniklerinin sistematik ve kapsamlı literatür incelemesi, bu alanda yapılacak gelecekteki araştırmalar için bir temel oluşturabilir.

1. VERİ MADENCİLİĞİ, TEKNİKLERİ VE UYGULAMALARI

1.1. Veri Madenciliği

Büyük veri, dağıtılmış ve merkezi olmayan, büyük hacimli, heterojen olma özellikleri taşır. Bu özellikler, büyük verilerden yararlı bilgi keşfetmeyi aşırı derecede zorlaştırır. Bu zorluk (Wu, Zhu, Wu ve Ding, 2013: 98) tarafından yapılan çalışmada, bir grup kör adamın bir fili ölçerek çizmeye çalışmasına benzetilmiştir.

Buradaki fil büyük veriyi temsil etmektedir. Her kör adamın amacı, süreç boyunca topladığı bilgi parçasına göre filin bir resmini çizmektir. Sonuç olarak, kör adamların hissettiği, kendi yerel bölgesiyle sınırlı olacağından, topladıkları bilgiler ile fili çizmeyi başaramayacaklardır. VM’de bu tür zorlukları aşmak için, kurumsal veri ambarlarında bulunan büyük verilerde gömülü bilgiyi bulur ve çıkarır. Bu yönüyle altın veya kömür madenciliğine benzetilir. VM yaklaşımı, istatistik, çevrimiçi analitik işleme, elektronik tablolar ve temel veri erişimi gibi diğer veri analizi tekniklerinin tamamlayıcısıdır. Başka bir ifade ile, VM verilerde anlam bulmanın başka bir yoludur (Rygielski, Wang ve Yen, 2002: 484).

VM, veri veya bilgi keşfi olarak da adlandırılır. Aynı zamanda belirli bir disiplindeki kararları iyileştirmek için büyük miktarda veriden bilgi çıkarma veya madencilik yapma süreci olarak da ifade edilir (Han ve Kamber, 2006: 2). VM, örtülü bilgi elde etmek için toplanan verilerden yararlanan ve değerli bilgiler elde etmek için tarihsel kayıtları analiz eden tüm aktivite ve teknikleri içerir (Seng ve Chen, 2010: 8043). VM, anlamlı desenler ve kurallar keşfetmek için büyük miktarda verinin otomatik veya yarı otomatik yollarla keşfedilmesi ve analiz edilmesi sürecidir (Jain ve Srivastava, 2013: 116). VM araçları gelecekteki eğilimleri ve davranışları tahmin eder, kuruluşların proaktif bilgi odaklı kararlar almasına yardımcı olur. Örneğin; dünyanın popüler arama motorları genellikle, her internet araması için tanımlayıcı bilgileri depolar. Bir Evrensel Kaynak Bulucu (Universal Resource Locator-URL) adresini ve kullanılan arama ifadesini, 18 aya kadar bir veri tabanında saklar (Ramzan ve Ahmad, 2014: 1).

Günümüzde hızla artan veri miktarı karşısında temel zorluk, görünüşte anlamsız büyük miktarda veriden yararlı bilgiler çıkartarak, nasıl rekabetçi bir istihbarat aracı haline getirileceğidir (Seng ve Chen, 2010: 8043). Çünkü büyük veri kümelerini analiz etme ve anlama, toplama ve saklama yeteneğini geliştirecek araçlar sınırlıdır (Fayyad, Piatetsky-Shapiro ve Smyth, 1996). İşletmeler rekabetçi kalabilmek için bilginin analizine giderek daha fazla ihtiyaç duyduğundan, bu zorluğun nasıl karşılanacağı kritik bir konudur (Seng ve Chen, 2010: 8043). Ayrıca veri miktarındaki hızlı büyüme, geleneksel veri analiz yöntemlerinin kullanılmasını zorlaştırmaktadır. Bu da çeşitli uygulama alanlarına uygun büyük veri analiz araçlarının geliştirilmesine neden olmuştur (Liu ve Vasarhelyi, 2014: 10). Veri madenciliği bu zorluğun üstesinden gelmek için kullanılan teknolojilerden biridir (Chen, Sakaguchi ve Frolick, 2000: 66).

1.2. Veri Madenciliğinin Evrimi ve Gelişimi

Veri madenciliğinin açıklama, tahmin ve öneri olmak üzere üç ana hedefi vardır. Açıklama, verileri tanımlayan insan tarafından yorumlanabilir desenleri bulmaya odaklanır. Tahmin, ilgi duyulan diğer değişkenlerin bilinmeyen veya gelecekteki değerlerini tahmin etmek için veri tabanındaki bazı değişkenleri veya alanları kullanmayı içerir. Öneri, tahminlerin değerli içgörüler ve kararlar için kullanıcıya sunulmasıdır (Fayyad vd., 1996: 27).

VM teknikleri uzun bir araştırma ve ürün geliştirme sürecinin sonucudur. Veri madenciliğinin kökeni, verilerin bilgisayarlarda ilk depolanmasıyla başlar, veri erişimindeki gelişmelerle devam eder ve günümüzde ise, teknoloji kullanıcıların veriler arasında gerçek zamanlı olarak gezinmesine olanak tanır (Rygielski vd., 2002: 486). Veri madenciliğinin evrimsel aşamaları Tablo 1’de sunulmuştur.

Tablo 1: VM Evrimsel Aşamaları

Sahne	İş sorusu	Etkinleştirici teknolojiler	Ürün sağlayıcıları	Özellikler
Veri Toplama (1960’lar)	Son beş yılda ortalama toplam gelirim ne kadardı?	Bilgisayarlar, bantlar, diskler	IBM, CDC	Geriye dönük, statik veri
Veri Erişimi (1980’ler)	Geçtiğimiz mart ayında X’de birim satışları ne kadardı?	İlişkisel Veri Tabanları (RDBMS), Yapılandırılmış Sorgu Dili (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Rekor düzeyde geriye dönük, dinamik veri
Veri Gezintisi (1990’lar)	Geçtiğimiz mart ayında X’de birim satışları ne kadardı? Boston’da ne kadardı?	Çevrimiçi Analitik İşleme (OLAP), çok boyutlu veri tabanları, veri ambarları	Pilot, IRI, Arbor, Redbrick, Evrimsel Teknolojiler	Çoklu düzeylerde geriye dönük, dinamik veri
Veri Madenciliği (2000 ve sonrası)	Boston biriminin önümüzdeki ay satış miktarı ne kadar olur? Neden?	Gelişmiş algoritmalar, çok işlemcili bilgisayarlar, devasa veri tabanları	Lockheed, IBM, SGI,	İleriye dönük, proaktif bilgi sunumu

Kaynak: (Rygielski vd., 2002: 486).

Tablo 1’de görüldüğü üzere, ilk zamanlar bilgisayarlar kullanılarak geriye dönük veriler elde edilirken, 2000’li yıllar ve sonrasında artık ileriye dönük ve proaktif bilgi sunumuna geçilmiştir. İlk aşama, toplama veya ortalama gibi basit hesaplamalar yapmak için kullanılan verilerin toplanmasıdır. Bu aşama da üretilen bilgiler, toplam gelir veya belirli bir zaman dilimindeki ortalama toplam gelir gibi veri toplama sitelerinden elde edilen rakamlarla ilgili iş sorularını cevaplamak için kullanılır. Günümüzde ise, artık gelişmiş algoritmalar, çok işlemcili bilgisayarlar, devasa veri tabanları ve veri yazılımları kullanılmaktadır (Rygielski vd., 2002: 484). VM süreci beş ana operasyonel aşamadan oluşur (Mraović, 2008: 439). Bunlar;

- İşlemsel verilerin veri ambarı sistemine çıkarılması, dönüştürülmesi ve yüklenmesi,
- Depolanması ve yönetilmesi
- Veri erişimine izin verilmesi
- Veri analizi
- Verilerin görsel bir biçimde sunulmasıdır

1.3. Veri Madenciliğinde Kullanılan Teknik ve Uygulamalar

VM'de kullanılan yöntemler yapay zeka, makine öğrenimi, istatistik, veri tabanı sistemleri ve iş zekasının kesişim noktasındadır (Jain ve Srivastava, 2013: 116). Veri madenciliği yöntemleri genel olarak iki kategoriye ayrılabilir. İlk kategori, istatistiksel modellerin kullanılmasıdır. Veri madenciliğine en büyük katkısı, hipotezleri değerlendirmek, sonuçları değerlendirmek ve sonuçları uygulamaktır. Kullanılan bazı popüler istatistiksel teknikler arasında olasılık dağılımları, korelasyon, regresyon, küme analizi ve ayırıcı analiz bulunur. Veri madenciliğinde uygulanan yöntemlerin ikinci kategorisi, makine öğrenimi adı verilen öncü yapay zekanın bir dalıdır. Veri madenciliği sisteminin öğrendiği ve modelleri için parametreleri bulduğu bir eğitim veri kümesinin kullanılmasını önerir. Bu tür bir yaklaşıma, veri tabanındaki çok sayıda örneği inceleyerek kurallar türetmeyi içeren tümevarımsal akıl yürütme de denir (Lei, Sakaguchi ve Mark, 2000: 67). Diğer bir ayırım ise, yönlendirilmiş ve yönlendirilmemiş yaklaşım yöntemleridir. Yönlendirilmiş veri madenciliğiyle, ilgi duyulan belirli bir hedef değişken belirlenir. Yönlendirilmemiş veri madenciliğiyle, belirli bir hedef değişken (bağımlı değişken) yoktur; bunun yerine amaç, bir veri popülasyonundaki herhangi bir değişken arasında herhangi bir ilişki bulmaktır. Bu iki geniş kategoriye karakterize etmenin bir başka yolu, yukarıdan aşağıya yaklaşımların belirli hipotezleri test ettiğini ve aşağıdan yukarıya yaklaşımın yeni hipotezler ürettiğini söylemektir (Gray ve Debrecey, 2014: 358).

Veri madenciliğinde verilerdeki ilgili ilişkileri çıkarmak için bir dizi algoritmik yaklaşımla desteklenir. Bu yaklaşımlar farklı problem sınıflarını ele alabilir. Veri madenciliği uygulamasında altı sınıf bulunur (Sharma ve Panigrahi, 2013: 38). Bunlar;

- **Sınıflandırma:** Bilinmeyen nesnelerin kategorik etiketlerini tahmin etmek ve farklı sınıflardaki nesneler arasında ayırım yapmak için (eğitim setinden) oluşturulur ve bir model (hedef sette) kullanılır. Yaygın sınıflandırma teknikleri arasında sinir ağları, Naïve Bayes tekniği, karar ağaçları ve destek vektör makineleri bulunur (Sharma ve Panigrahi, 2013: 38).
- **Kümeleme:** Kümeleme, nesneleri daha önce bilinmeyen kavramsal olarak anlamlı gruplara (yani kümelere) ayırmak için kullanılır, bir kümedeki nesneler birbirine benzer ancak diğer kümelerdeki nesnelerden çok farklıdır. En yaygın kümeleme teknikleri K-en yakın komşu, Naïve Bayes tekniği ve kendi kendini organize eden haritalardır (Gray ve Debrecey, 2014: 358).
- **Tahmin:** bir veri kümesinin desenlerine dayalı olarak sayısal ve sıralı gelecekteki değerleri tahmin eder. Tahmin için, tahmin edilen değerlerin kategorik (ayrık değerli ve sırasız) olmaktan ziyade sürekli değerli (sıralı) olduğu belirtilmektedir. Sinir ağları ve lojistik model tahmini en yaygın kullanılan tahmin teknikleridir (Gray ve Debrecey, 2014: 358).
- **Aykırı değer algılama:** kalan veri kümesinden büyük ölçüde farklı veya tutarsız olan nesneleri tespit etmek için veri nesneleri arasındaki mesafeyi ölçmek için kullanılır (Sharma ve Panigrahi, 2013: 38).

- **Regresyon:** bir veya daha fazla bağımsız değişken ile bağımlı değişken (yani sürekli değerli) arasındaki ilişkiyi ortaya çıkarmak için kullanılan istatistiksel bir metodolojidir. Regresyon tekniği genellikle lojistik regresyon ve doğrusal regresyon gibi matematiksel yöntemler kullanılarak gerçekleştirilir (Sharma ve Panigrahi, 2013: 39).
- **Görselleştirme:** verilerin kolayca anlaşılabilir sunumunu ve karmaşık veri özelliklerini kullanıcıların veri madenciliği sürecinde ortaya çıkarılan karmaşık örüntüleri veya ilişkileri görmelerini sağlamak için net örüntülere dönüştüren metodolojiyi ifade eder (Gray ve Debreceny, 2014: 358).

Tablo 2: Veri Madenciliği Yöntemleri, Güçlü Yönleri ve Sınırlılıkları

Yöntem	Güçlü yönler	Sınırlılıklar
Yapay sinir ağı	Sahtekârlık tespitinde köklü bir geçmişe sahiptir. Algoritmik olmayan ikili sınıflandırma problemlerinde de kanıtlanmış uygunluk.	Eğitim ve çalışma için yüksek hesaplama gücü ister; gerçek zamanlı kullanım için uygun değildir. Eğitim kümesi alanı iyi temsil etmezse aşırı öğrenme riski vardır; yeni dolandırıcılık yöntemlerine uyum için sık sık yeniden eğitim gerekir.
Lojistik model	Uygulaması basittir. Dolandırıcılık tespitinde iyi bilinen bir geçmişe sahiptir.	Sınıflandırma başarımı, diğer veri madenciliği yöntemlerine göre düşüktür; dolandırıcılık tespitinin karmaşıklığıyla baş etmek zor olabilir.
Destek vektör makinesi (SVM)	Dolandırıcılık tespiti gibi doğrusal olmayan sınıflandırma problemlerini çözebilir. Eğitim ve çalışma düşük hesaplama gücü gerektirir; gerçek zamanlı çalışmaya elverişlidir.	Girdi kümesinin dönüştürülmesi nedeniyle denetçilerin sonuçları işlemesi/yorumlaması zordur.
Karar ağaçları, ormanlar ve CART	Uygulaması ve anlaşılması kolaydır. Eğitim ve çalışma düşük hesaplama gücü gerektirir; gerçek zamanlı çalışmaya elverişlidir.	Eğitim kümesi alanı iyi temsil etmezse aşırı öğrenme görülebilir. İlk kurulumdaki optimizasyon yüksek hesaplama gücü gerektirir.
Genetik algoritma/programlama	Uygulaması nispeten basittir; uygunluk (fitness) çözümü olarak sınıflandırma doğruluğu kullanılabilir. Algoritmik olmayan ikili sınıflandırma problemlerinde kanıtlanmış uygunluğu vardır.	Eğitim ve çalışma için yüksek hesaplama gücü ister; gerçek zamanlı için uygun değildir. Yerel maksimum/minimum problemi nedeniyle yeni sahtekârlık yöntemlerine uyumda güçlüklerle karşılaşılır.
Metin madenciliği	Finansal tablo hilesi gibi büyük hacimli metinsel veriye sahip dolandırıcılık türleri için çok kullanışlıdır.	Gerçek dolandırıcılık tespiti için ayrıca bir sınıflandırma yöntemine ihtiyaç duyar. Metin verisi genellikle daha öznel olduğundan işlenmesi zordur.
Veriyi İşlemenin Grup Yöntemi (GMDH)	Uygulaması basittir. Elde edilebilir en iyi çözüme ulaşma garantisi sunar.	Pek çok dolandırıcılık türünde görülen gürültülü veriyi sınıflandırmada zorluk.
Tepki yüzeyi metodolojisi (RSM)	Dolandırıcılık tespiti gibi doğrusal olmayan sınıflandırma problemlerini çözebilir.	Sınıflandırma başarımı, diğer veri madenciliği yöntemlerinden genelde düşüktür; dolandırıcılığın karmaşıklığıyla baş etmek zor olabilir.

Yöntem	Güçlü yönler	Sınırlılıklar
Öz-örgütlenen harita (SOM)	Uygulaması basittir; sonuçların görsel doğası nedeniyle denetçilerin anlaması çok kolaydır. Algoritmik olmayan ikili sınıflandırma problemlerinde kanıtlanmış uygunluğu vardır.	Görselleştirme denetçi gözlemi gerektirir; tamamen otomatikleştirilmesi kolay değildir.
Bayesyen inanç ağı (BBN)	Yüksek hesaplama verimliliği sayesinde gerçek zamanlı çalışmaya elverişlidir.	İncelenen dolandırıcılık türüne özgü tipik ve atipik davranışların güçlü biçimde anlaşılmasını gerektirir.
Süreç madenciliği	Her yinelemeli adım için bilginin mevcut olduğu iç soruşturmalarda yararlıdır. Tekil öznitelikler yerine tüm süreç zincirine odaklanma olanağı sağlar.	İncelenen tür için tipik/atipik davranışın güçlü biçimde anlaşılmasını gerektirir. Pek çok dolandırıcılık türünde görülen gürültülü veriyi sınıflandırmak zordur.
Yapay bağışıklık sistemi	Dengesiz sınıf dağılımlarına sahip (fraud tespiti gibi) sınıflandırma problemlerine yüksek uyumludur.	Çalıştırma için yüksek hesaplama gücü ister; gerçek zamanlı kullanım için uygun değildir.
Hibrit yöntemler	Birden çok geleneksel tespit yönteminin güçlü yönlerini birleştirerek yeni dolandırıcılık tekniklerine uyum sağlayabilir.	Dolandırıcılık yüksek maliyetli bir sorun olduğundan, yeni ve yeterince incelenmemiş yöntemler yüksek risk taşır.

Kaynak: Westve Bhattacharya, 2016: 52

Veri madenciliğin de lojistik regresyon, ayırıcı analiz ve küme analizi gibi klasik istatistiksel prosedürleri, sinir ağları, karar ağaçları ve genetik algoritmalar gibi makine öğrenme tekniklerini kullanır (Özekes, 2003: 80). Genel olarak kullanılan VM yöntemleri güçlü yönleri ve sınırlılıkları Tablo 2’de sunulmuştur. Veri madenciliği modellerinden ve tekniklerinden amaca en uygun olanı seçilmelidir. Uygun olmayan bir teknik seçilerek yapılan veri madenciliği uygulaması başarılı olma ihtimali düşüktür (Özekes, 2003: 80). VM, organizasyonel amaçlara ulaşmada çok geniş kullanım alanına sahiptir. Pazarlama, finans, muhasebe, insan kaynakları, denetim, üretim ve birçok işletme de fiyat, planlama, personel becerileri gibi iç faktörler de, ayrıca ekonomik göstergeler, rekabet ve pazarın yapısı gibi dış faktörlerin analiz edilmesinde kullanılmaktadır (Tüzüntürk, 2010: 70).

Veri madenciliği, yaşam kalitesini artırmak, hayatı kolaylaştırmak ve kurumların davranışlarına daha fazla düzen ve sorumluluk getirme konusunda önemli faydalar sunan bir tekniktir. Ancak veri madenciliği tekniklerinin kullanımında gizlilik, yasallık ve etikle ilgili sorunlar bulunur. Ayrıca veri madenciliği uygulamalarının orijinal amaçlarının dışında kullanılması da söz konusudur. Veri’nin bilgiye dönüşmesi için, insanların onu yorumlayabilmesi, anlamlı eylem seçeneklerine dönüştürmesi ve istenen sonuçları alabilmesi için beklenen eylemleri gerçekleştirmesi gerekir. Seifert’in (2004) belirttiği gibi, veri madenciliği teknikleri örüntüleri ve ilişkileri ortaya çıkarmamıza yardımcı olabilir, ancak bize bu bulguların gerçek değerini veya önemini söyleyemez. Bu değerlendirmeler hâlâ insanın ve onun iradeli eyleminin ayrıcalığı olmaya devam etmektedir (Mraović, 2008: 450). VM ile ilgili yapılan açıklamalar sonrasında bir sonraki kısımda muhasebe de kullanımı konusuna değinilmiştir.

2. MUHASEBEDE VERİ MADENCİLİĞİ

İletişim teknolojilerinin ve bilgi erişilebilirliğinin hızla ilerlediği çağımızda şirketler, keşfedildiğinde iş karar alma süreçleri için kritik öneme sahip olabilecek değerli bilgilerin saklı olabileceği büyük hacimli verilerle karşı karşıyadır. Bu zorluğun üstesinden gelmek için veri madenciliği önemli bir çözüm sunar (Chahadah, Refae ve Qasim, 2018: 443). Veri madenciliği, kuruluşlara kurumsal gelişimin gelecekteki eğilimlerini etkili bir şekilde tahmin etme, yöneticilerin daha iyi kararlar almasına yardımcı olma ve bir işletmenin rekabet gücünü artırma dahil olmak üzere çok çeşitli avantajlar ve yetenekler sağlar (Aslantas, 2024: 292). VM, işletmenin temelini oluşturan ve kaydetme, iç ve dış raporlama, maliyetlendirme, tahmin, değerlendirme, analiz ve denetleme gibi çok çeşitli görevleri yerine getiren muhasebe bilgi sistemi (Amani ve Fadlalla, 2017: 33) tarafından üretilen yüksek hacimli verinin anlamlı bilgiye dönüştürülmesi sağlar (Alagöz, Öge ve Ortakarpuz, 2014: 18). Muhasebe bilgi sistemi tarafından sağlanan veriler miktar bağlamında çok olmasına rağmen, bilgi açısından yetersizdir. Bu durum finansal karar destek sistemini zayıflatır. Veri madenciliği teknikleri sayesinde büyük verilerden eğilimler tahmin edilerek yöneticilere karar alma desteği sağlar ve işletmelerin rekabet gücünü artırır (Xiao, Xiaoli ve Gaojin, 2010: 381). Ayrıca geleneksel analiz yöntemlerine göre VM, büyük veri yığınları arasından değerli bilgilerin yakalanıp çıkarılmasına daha fazla imkan tanır (Coman vd., 2023: 118). Bununla birlikte VM, tekniklerinin benimsenmesi, harcama modellerinin izlemesini, anormalliklerin belirlemesine ve hesap verilebilirliği iyileştirir (Kumar vd., 2020: 2108).

Muhasebede veri madenciliği uygulamalarının kullanımı, özellikle finansal raporlama, iç denetim, maliyet yönetimi ve dolandırıcılık tespiti gibi alanlar da yaygındır (Ngai vd., 2011: 544). Veri madenciliği araçları, kuruluşların gelişmiş bütçeleme uygulamalarına ve daha iyi kaynak tahsisi yapmalarını kolaylaştırır. Kuruluşların finansal işlemlerdeki olağandışı kalıpları tespit etmelerini sağlar (Baesens vd., 2003: 315). VM tekniklerin uygulanması, hesap verebilirliği ve raporlamada şeffaflığı artırarak paydaşlar arasında daha fazla güven oluşturabilir (Feng vd., 2018: 608). Öngörücü analizlerle gelecekteki harcamaları tahmin edebilir ve daha iyi bütçe planlaması ve kaynak yönetimine olanak tanır (Zhang, Li ve Chen, 2014: 90).

Muhasebede veri madenciliğinin uygulamalarını tanımlayan çok sayıda araştırma vardır. Bu araştırmaların çoğu literatür incelemesi sunsa da bu incelemeler genellikle belirli bir muhasebe alanına ve/veya veri madenciliği tekniğine odaklanmıştır (Amani ve Fadlalla, 2017: 32). Yapılan çalışmalar değerlendirildiğinde, büyük veri ve veri analizlerinin kullanılması, yönetim muhasebesi açısından kontrol ve bütçelemeye etkinliğin artmasına, finansal muhasebe alanında kullanılması bilgi kalitesi ve ihtiyaca uygun bilgi üretimin de şeffaflık ve güçlü kararlar alınmasına ve finansal tablolarda kullanılması dinamik, gerçek zamanlı raporlar sunulmasına yardımcı olmaktadır (Warren, Moffitt ve Byrnes, 2015: 397).

2.1. Muhasebede VM Kullanımı

Muhasebede veri madenciliği uygulamalarının konu odaklı bir analizi, bu uygulamaların neredeyse üçte ikisinin (%64) güvence ve uyumluluğa, dörtte birinin (%25) yönetim muhasebesine ve geri kalanının finansal muhasebe ve muhasebe bilgi sistemlerine odaklandığını göstermektedir (Amani ve Fadlalla, 2017: 32). Muhasebe denetiminde veri madenciliği kullanımı, genellikle çok büyük veri kümelerindeki gizli kalıpları keşfetmek ve finansal tablolar da yer alan hileli muhasebe işlemlerin tespitinde önemli bir rol oynamaktadır (Sharma ve Panigrahi, 2013: 39; Söylemez ve Türkmen, 2017: 278; Mercan vd., 2023:39-40; Kandemir ve Kardeş, 2025: 334). Bu önemli rol, önde gelen muhasebe meslek kuruluşları tarafından kabul edilmektedir. Amerikan Sertifikalı Muhasebeciler Enstitüsü (Association of International Certified Professional Accountants-AICPA), veri madenciliğini geleceğin en iyi on teknolojisinden biri olarak belirlemiş ve İç Denetçiler Enstitüsü (Institute of Internal Auditors-IIA) veri madenciliğini dört araştırma önceliğinden biri olarak listelemiştir (Koh ve Low, 2004: 465). Sertifikalı Küresel Yönetim Muhasebeciler (Chartered Global Management Accountants-CGMA), şirket liderlerinin %50'sinden fazlasının büyük veriyi ve veri madenciliğini, veri odaklı iş çağı için temel iş önceliği arasında sıraladığını bildirmiştir (CGMA, 2013). VM muhasebe alanından önemli roller üstlense de veri kalitesi problemleri, çalışanların değişim direnci, sınırlı eğitim ve gerekli kullanım becerilerinin çalışanlarda olmaması gibi çeşitli sorunlar ortaya çıkabilir (Alsharif ve Alvi, 2021:129; Chen ve Zhang, 2014: 89). Ayrıca VM kullanımı, gerekli olan yazılım temini, teknolojik altyapı oluşturulması, personel eğitimi gibi maliyetlere yol açabilir (Sağlar ve Kefe, 2021: 266). Muhasebe ve finans alanındaki veri madenciliği uygulamaları içinde en çok araştırılan konularından biri dolandırıcılık tespitidir (Chahadah, Refae ve Qasim, 2018: 443). Dolayısıyla dolandırıcılığın tespitinde kullanılan yöntemlerin ve uygulamaların iyi bilinmesi gerekir. Bu amaçla sonraki kısımda muhasebe alanında dolandırıcılık ayrıntılı olarak ele alınmıştır.

2.2. Muhasebe Dolandırıcılığı ve VM

Muhasebe dolandırıcılığı, bir şirketin sahte/çarpık finansal raporlar hazırlayıp, yayınlarak finansal tablo kullanıcılarını kasıtlı olarak aldatma veya yanıltma eylemidir. Bu eylemler muhasebe standartlarının tahrif edilmesi, değiştirilmesi, ihmal edilmesi veya seçici olarak açıklanması, kötüye kullanılması veya yanlış beyan gibi yöntemleri içerir (Rezaee, 2005: 279). Muhasebe dolandırıcılığı, özellikle yatırımcılar ve alacaklılar olmak üzere kamuya açık bilgilerin kullanıcılarını aldatmak için kasıtlı bir çabadır (Papík ve Papíková, 2022: 6). Bu dolandırıcılık faaliyetleri teknoloji gelişimi, internet ve mobil bilgi işlem gibi gelişmelerden dolayı artmakta (Laourou, 2025: 289) ve işletmelerde her yıl dünya çapında büyük zarara neden olmaktadır (Popat vd., 2018: 1120). Dolandırıcılığın ve özellikle muhasebe ve finansal tablo dolandırıcılığının tespiti, son on yılda finansal tablo denetiminin giderek daha önemli bir bileşeni haline gelmiştir (Debrecey ve Gray, 2010: 159).

Kurumsal sektöre zarar veren yüksek profilli dolandırıcılık vakalarının yarattığı olumsuzluklarla mücadele etmek için ise, veri madenciliği giderek daha önemli hale gelmektedir (Mraović, 2008: 450). Ayrıca dolandırıcılık tespiti aynı zamanda bir tanımlama ve sınıflandırma sorunu olarak da kabul edilir.

Büyük miktardaki finansal işlemler arasından gerçek ve sahte işlemleri ayırmak önemli bir sorundur. Yanlış sınıflandırma önemli maliyetlere neden olur (Duman ve Özelçi, 2011: 13058). Bu maliyetlerden kurtulmak veri madenciliği yöntemleri ile mümkündür (Yue vd., 2007: 3). Veri madenciliği, sınıflandırma ve tahmin kalıplarını araştırmak için farklı perspektiflerden büyük veri örnekleri üzerinde gerçekleştirilen bir veri analiz sürecidir (Ringsdorf ve Kajanová, 2017). Veri madenciliğinin uygulanması, veri örneği kategorizasyonunun farklı alanlarında, örneğin farklı tahmin modellerindeki uygulamalarda gerçekleşir. Bu yöntemlerle oluşturulan veri madenciliği modelleri, finansal bilgilerin dış kullanıcılarının, kamuya açık finansal veriler aracılığıyla şüpheli davranışları tespit etmesini sağlar (Papík ve Papíková, 2022: 6). Veri madenciliği yöntemleri muhasebe kayıtlarını, belirsiz gerçekleri, eğilimleri veya finansal tablolardaki tutarsızlıkları tespit eder ve yüksek riskli alanları belirlemek için şirket personeliyle yakın iş birliği içinde çalışır (Mraović, 2008: 450). Yine de finansal tablo dolandırıcılıklarının tespiti, bu alanda genel bir anlayış eksikliği ve aldatmacaları gizleyebilen sektördeki bilgili kişiler tarafından gerçekleştirilmesi nedeniyle zor olabilir (West ve Bhattacharya, 2016: 51).

Muhasebe dolandırıcılığının tespitinde VM yöntemlerini kullanan ilk dolandırıcılık tespiti çalışmaları, lojistik regresyon ve sinir ağları gibi VM yöntemlerine yoğun olarak odaklanmıştır (Quah ve Sriganesh, 2008: 1724). Zaman içinde alan yazını genişledikçe karar ağaçları, SVM, Bayesçi sınıflandırıcılar, metin madenciliği, ağ-tabanlı ve aykırı değer/anomali temelli algoritmalar gibi tekniklerin de kullanıldığı görülmüştür.

Muhasebe dolandırıcılığı yapmanın nedenleri arasında hisse senedi performansını iyileştirmek, vergi yükümlülüklerini azaltmak veya yönetim baskısı nedeniyle performansı abartma girişimi yer alır (Ravisankar vd., 2011: 498). Bununla birlikte dolandırıcılık nedenleri ve etkileyen faktörleri “Dolandırıcılık Üçgeni Teorisi” ile de açıklanmaya çalışılmıştır (Rezaee, 2005: 279). Teori “motivasyon, fırsat ve rasyonalizasyon” olmak üzere dolandırıcılık faaliyetlerini üç ana unsura bağlar. Motivasyon faktörleri arasında kurumsal genişleme ve iş büyüme oranı, finansman ihtiyaçları, öz sermaye teşvikleri, performans, listeleme statüsünün korunması ve hisse senedi fiyat manipülasyonu bulunur (Tan vd., 2021: 3). Rasyonelleştirme faktörleri esas olarak yöneticilerin kişisel etik, dolandırıcılığa karşı tutumlar ve kişisel deneyimler dahil olmak üzere kişisel faktörlerini içerir (Liu vd., 2023: 76). Fırsat faktörü bilgi asimetrisi, muhasebe dolandırıcılığını mümkün kılan, bilgi gizlemesinin tespit edilmesini zorlaştırır ve bu da kötü haberlerin gizlenmesine ve birikmesine yol açarak nihayetinde finansal dolandırıcılığa yol açar (Yin, Sun ve Kong, 2025: 3). Öte yandan büyük verinin gelişimiyle birlikte büyük veri platformlarının inşası, veri madenciliği yöntemleri ile veri analizi, şirketler ve paydaşlar arasındaki bilgi asimetrisini hafifleterek kurumsal finansal dolandırıcılığı azaltır (Li vd., 2024).

Ancak bu süreçte, veri kalitesi, algoritmalar ve modellerdeki sınırlamalar, eksik veri altyapısı, dijital denetimlerin uygulanmasının yüksek maliyetleri ve ilgili verilerin ve iş bilgilerinin gizliliği nedeniyle önemli sınırlılıklar barındırır (Yin, Sun ve Kong, 2025: 3).

Bu sınırlılıklara rağmen, borsa ve döviz kuru öngörüsü, banka iflaslarının tahmini, finansal riskin anlaşılması ve yönetimi, vadeli işlem ticareti, kredi derecelendirmesi ve kredi riski yönetimi ve kara para aklama analizleri; veri madenciliğinin görevleri arasında yer almaktadır (West ve Bhattacharya, 2016: 51). Bu görevler VM'nin finansal sistemlerde düzensizliklerinin erken saptanmasının yalnızca operasyonel verimlilik için değil, aynı zamanda düzenleyici uyum ve paydaş güveni açısından da kritik olduğunu göstermektedir.

Sonuç olarak, muhasebe dolandırıcılığının tespiti için sağlam stratejiler ve teknikler geliştirilmesinin önemi açıktır. Özellikle dolandırıcılıkla ilgili uygulanan anomali tespit yaklaşımlarına ilişkin yayımlanmış güncel araştırmaların kapsamlı ve sistematik bir görünümünü çıkarmak, alanın bütüncül olarak anlaşılmasına hizmet etmektedir (Hilal, Gadsden ve Yawney, 2022: 2). Bu çerçevede, çalışmanın sonraki kısmında araştırma amacına uygun olarak tasarlanan yöntem ve analiz adımlarına geçilmektedir.

3. ARAŞTIRMANIN METODOLOJİSİ

3.1. Araştırmanın Amacı ve Kapsamı

Bu araştırmanın amacı, muhasebe alanında muhasebe dolandırıcılık ve finansal tablo dolandırıcılığı vakalarının tespitinde kullanılan veri madenciliği yöntemlerini incelemektedir. Araştırmada, birden fazla sistematik literatür incelemesini tek bir çalışmada birleştiren, “Şemsiye İncelemesi (Umbrella Review)” olarak adlandırılan yöntemle elektronik veri tabanlarında yer alan sistematik inceleme yapan çalışmalardaki birincil ampirik çalışmaları birleştirerek, kullanılan veri madenciliği yöntemlerinin frekans dağılımı ve zaman içindeki kullanım eğilimlerini analiz etmektir. Araştırmanın amacı ve kapsamı doğrultusunda oluşturulan araştırma soruları aşağıdaki gibidir.

RQ1: Muhasebe dolandırıcılığının tespitinde literatürde en çok hangi veri madenciliği yöntemleri kullanılmıştır?

RQ2: Bu yöntemlerin zaman içinde kullanım eğilimleri nasıl değişmiştir?

3.2. Araştırmanın Yöntemi ve Uygulaması

Araştırmada, sistematik incelemelerden elde edilen bulguları bir araya getirilerek daha bütüncül ve sentezlenmiş bir sonuç oluşturmayı amaçlayan “Şemsiye İncelemesi (Umbrella Review)” kullanılmıştır. Literatürde yer alan sistematik inceleme çalışmalarının seçiminde PRISMA 2020 araştırma yöntemi kullanılmıştır. PRISMA 2020, araştırmaları belirleme, seçme, değerlendirme ve sentezleme yöntemlerindeki gelişmeleri yansıtan raporlama rehberliği içermektedir (Page vd., 2021). Literatür taramasına ilişkin anahtar kelimeler hem Türkçe hem de İngilizce olarak Tablo 3'te sunulmuştur.

Tablo 3: Taranan Veri Tabanları, Anahtar Kelimeler ve Araştırma Sonuçları

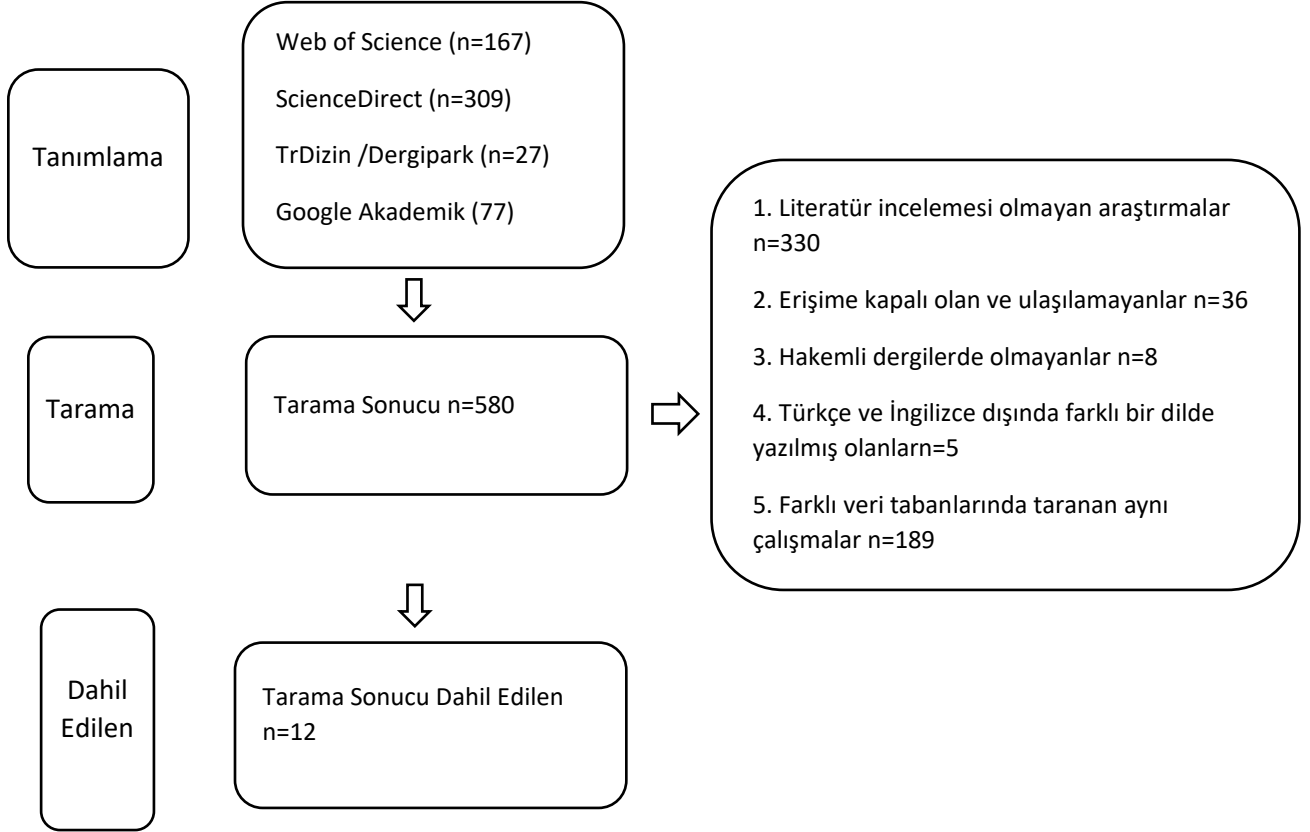
Veri Tabanı	Anahtar Kelime	Araştırma Sonucu
Web of Science	("financial fraud" OR "corporate fraud" OR "financial statement fraud" OR "accounting fraud" OR "Literature review" OR "systematic review") AND ("financial statement" OR "financial reporting" OR "accounting information system") AND ("data mining")	167
ScienceDirect	("financial fraud" OR "corporate fraud" OR "financial statement fraud" OR "accounting fraud" OR "Literature review" OR "systematic review") AND ("financial statement" OR "financial reporting" OR "accounting information system") AND ("data mining")	309
TrDizin/Dergipark	title /abstract: ("financial fraud"); ("corporate fraud"); ("financial statement fraud"); ("accounting fraud"); ("Literature review"); ("systematic review"); ("data mining") Başlık/Özet: ("finansal dolandırıcılık"); ("kurumsal dolandırıcılık"); ("finansal tablo dolandırıcılığı"); ("muhasabe dolandırıcılığı"); ("Literatür taraması"); ("sistematik inceleme"); ("veri madenciliği")	27
Google akademik	"financial fraud" and "corporate fraud" and "financial statement fraud" and "accounting fraud" and "data mining" and "Literature review" and "systematic review" "finansal dolandırıcılık" ve "kurumsal dolandırıcılık" ve "finansal tablo dolandırıcılığı" ve "muhasabe dolandırıcılığı" ve "veri madenciliği" ve "Literatür taraması" ve "sistematik inceleme"	77
Toplam		580

Tablo 3'te yer alan anahtar kelimeler aracılığıyla "Web of Science (n=167)", "ScienceDirect (n=309)", "TrDizin/DergiPark (n=27)", "Google akademik (n=77)" olmak üzere beş elektronik veri tabanı taranmıştır. Tarama sonucunda başlık, özet ve anahtar kelimeleri ile uyumlu İngilizce/Türkçe 580 çalışma Ağustos 2025 ayı içerisinde bulunmuştur. Bulunan çalışmalar Tablo 4'te hariç tutma kriterlerine göre ayırma işlemine tabi tutulmuştur.

Tablo 4: Çalışmaları Ayırtırmak İçin Kullanılan Hariç Tutma Kriterleri

Hariç Tutma Kriterleri
1. Literatür incelemesi veya sistematik inceleme olmayan araştırmalar n=330
2. Erişime kapalı olan ve ulaşılamayanlar n=36
3. Hakemli dergilerde olmayanlar n=8
4. Türkçe ve İngilizce dışında farklı bir dilde yazılmış olanlar=5
5. Farklı veri tabanlarında taranan aynı çalışmalar n=321

Ayırırda kullanılan yöntem ve nihai olarak araştırmaya dahil edilecek çalışmaların tespiti Prisma 2020 prosedürlerine göre yapılmıştır. Prisma yöntemi akış şeması Şekil 1'de sunulmuştur.



Şekil 1: Prisma Yöntemi Akış Diyagramı

Şekil 1’ de gösterilen şemada makalelerin hangi safhalardan geçirildiğini göstermektedir. Tarama sonucunda 580 çalışmaya ulaşılmış, hariç tutma kriterlerinin uygulaması sonucunda toplam 565 çalışma analiz dışında tutulmuştur. Analiz dışında tutulma nedenleri incelendiğinde, literatür incelemesi olmayan çalışmalar (n=330) ile farklı veri tabanlarında taranan aynı çalışmalar (n=189) yoğunluğu gözlemlenmiştir.

Yapılan hariç tutma işlemi sonunda araştırmanın amacıyla uyumlu 12 adet literatür veya sistematik inceleme örneği bulunmuştur. Bulunan çalışmalar Tablo 4’te sıralanmıştır. Alan yazını araştırmalarının kapsadığı yıllar incelendiğinde 1990 ile 2024 yılları arasındaki geniş bir dönemi kapsadığı görülmüştür. En geniş yılları kapsayan alan yazını araştırması 32 yıl ile Ashtiani ve Raahemi (2022) tarafından yapılan çalışma olmuştur. Yine Tablo 4’te sunulan çalışmalar içerisinde Rao ve Mandhala (2024) tarafından yapılan literatür taraması 187 çalışma ile en fazla araştırmanın dahil edildiği literatür taraması olmuştur. Çalışmada belirlenen araştırma sorularından ilki olan “RQ1: Muhasebe dolandırıcılığının tespitinde literatürde en çok hangi veri madenciliği yöntemleri kullanılmıştır?” sorusunun cevabını bulmak için literatür incelemelerinde bulunan birincil makalelerde kullanılan ve muhasebe dolandırıcılığı ile ilgili çalışmalarda kullanılan VM yöntemlerinin frekans dağılımlarının hesaplamasıdır. Birincil makalelerde muhasebe dolandırıcılığı haricinde; sigorta, kripto, kredi kartı, e-ticaret gibi dolandırıcılık türlerinde kullanılan VM yöntemlerine de yer verilmiştir.

Bu çalışmalarda yer alan finansal tablo ve muhasebe dolandırıcılığı (kurumsal) dikkate alınarak frekans dağılımları yapılmış, diğer dolandırıcılık türleri hariç bırakılarak Tablo 5’te sunulmuştur.

Tablo 4: Analize Dahil Edilen Literatür Çalışmaları

Makale Adı ve Yılı	Yazarlar	Kapsadığı Yıllar	Makale Sayısı
Financial Fraud Detection and Machine Learning Algorithm (Unsupervised Learning): Systematic Literature Review	Husnaningtyas ve Dewayanto (2023).	2010–2023	27
Financial fraud detection through the application of machine learning techniques: a literature review	Aros vd. (2024).	2012–2023	104
Unveiling financial fraud: A comprehensive review of machine learning and data mining techniques	Rao ve Mandhala (2024)	1995–2021	187
The application of machine learning to study fraud	Ramzan ve Lokanan (2025)	2011–2022	88
Intelligent Fraud Detection in Financial Statements	Ashtiani ve Raahemi (2022)	1990–2021	47
Anti-money laundering and financial fraud detection	Goecks vd. (2022)	2010–2021	45
A Systematic Literature Review of Machine Learning Techniques in Financial Fraud Prevention and Detection	Adekoya ve Tagbo (2023)	2010–2022	145
A Review of Financial Accounting Fraud Detection based on Data Mining Techniques	Sharma ve Panigrahi (2012)	2000–2012	25
Financial fraud detection applying datamining techniques: A comprehensive review from 2009 to 2019	Al-Hashedi ve Magalingam (2021)	2009–2019	75
The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature	Ngai vd. (2011)	1997–2008	49
Intelligent financial fraud detection: A comprehensive review	West ve Bhattacharya (2016)	2004–2014	14
Hileli Finansal Tabloların Tespitinde Veri Madenciliği Uygulamaları: Mevcut Araştırma Eğilimlerinin İncelenmesi	Kandemir ve Kardeş (2025)	2006-2024	41
Toplam			847

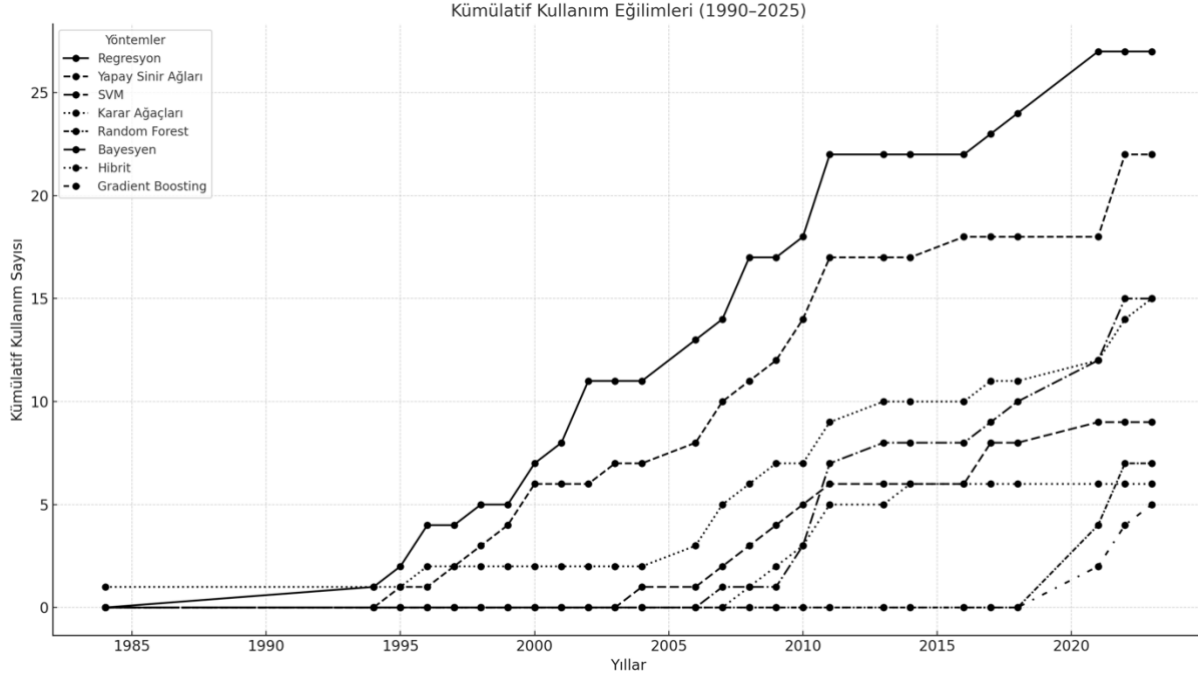
Tablo 5’te içerisinde incelenen dönemde aynı isimli çalışmalar harici tutulmuştur. Çalışma da tüm makaleler incelenmiş, her bir birincil makalede kullanılan yöntemler sıralanmış, kaynakları ve kullandıkları yöntemler göre tablolar oluşturulmuş ve sınıflandırılmıştır. Elde edilen bulgular, VM yöntemlerinden “Regresyon (Logit/Logistic/Probit)” toplam 27 makale ile en çok muhasebe dolandırıcılığının tespitinde kullanılan yöntem olmuştur. LR yönteminin tercih edilme nedeni olarak; erişilebilirlik, kullanım kolaylığı, dengesiz verilerde istikrarlı tutarlılığı ve ikili bağımlı değişken için yaygın kullanımından dolayı olduğu söylenebilir. İkinci olarak en çok kullanılan yöntem 23 çalışmada kullanılan Yapay Sinir Ağları (ANN / NN-Artificial Neural Network / Neural Network, MLP -Multilayer Perceptron, LSTM-Long Short-Term Memory, GNN -Graph Neural Network) olmuştur.

Tablo 5: Analize Sonucuna Göre Kullanılan VM Yöntemleri, Sayısı ve Kaynaklar

Veri Madenciliği Yöntemi	Kullanım Sayısı	Kaynaklar (Detaylı)
Lojistik Regresyon (Regresyon (Logit / Logistic / Probit vb.)	28	Persons (1995); Beasley (1996); Hansen vd. (1996); Summers ve Sweeney (1998); Abbott vd. (2000); Bell ve Carcello (2000); Spathis (2002); Spathis vd. (2002); Owusu-Ansah vd. (2002); Huang (2006); Ren (2006); Guoxin vd. (2007); Bai vd. (2008); Yuan vd. (2008); Liou (2008); Feroz vd. (2000); Lin vd. (2003); Ravisankar vd. (2011); Perols (2011); Zhou ve Kapoor (2011); Yue vd. (2010); Deng (2017); Bhattacharyya vd. (2011); Pinquet vd. (2007); Viaene vd. (2007); Bermúdez vd. (2008); Balagolla vd. (2021); Runnan Tan vd. (2021)
Yapay Sinir Ağları (ANN / NN / PNN / MLP / LSTM / GNN vb.)	23	Fanning vd. (1995); Green ve Choi (1997); Fanning ve Cogger (1998); Cerullo ve Cerullo (1999); Feroz vd. (2000); Koskivaara (2000); Lin vd. (2003); Kotsiantis vd. (2006); Kirkos vd. (2007); Bose ve Wang (2007); Liou (2008); Krambia-Kapardis vd. (2010); Ravisankar vd. (2011); China (2011); Perols (2011); Zhou ve Kapoor (2011); Cecchini vd. (2010); Whitrow vd. (2009); Liang ve Lv (2016); Yang Zheng vd. (2022); Bingxue Fu vd. (2022); Baran Kılıç vd. (2022); Martin vd. (2022)
Karar Ağaçları (DT / CART / RF / J48 vb.)	15	Breiman vd. (1984); Hansen vd. (1996); Kotsiantis vd. (2006); Kirkos vd. (2007); Bai vd. (2008); Perols (2011); Sahin vd. (2013); Bhattacharyya vd. (2011); Whitrow vd. (2009); Bose ve Wang (2007); Hajek ve Henriques (2017); Eduardo vd. (2021); Chaurasia vd. (2022); Baran Kılıç vd. (2022); Yong Ren vd. (2023)
Destek Vektör Makineleri (SVM)	14	Kirkos vd. (2007); Ravisankar vd. (2011); Perols (2011); Yue vd. (2011); Deng (2010, 2017); Bhattacharyya vd. (2011); Huang (2013); Cecchini vd. (2010); Bingxue Fu vd. (2022); Elmougy vd. (2021); Baran Kılıç vd. (2022); Bui Tong Nha vd. (2022); Balagolla vd. (2021)
Bayesyen Yöntemler (BBN / NB)	9	Viaene vd. (2004); Kirkos vd. (2007); Zhou ve Kapoor (2011); Holton (2009); Bermúdez vd. (2008); Deng (2010, 2017); Hajek ve Henriques (2017); Tan vd. (2021)
Kümeleme (Clustering – K-Means, Hierarchical, Fuzzy, Hibrit)	9	Virdhagrishwaran (2006); Deng (2009); Tangod ve Kulkarni (2015); Meenakshi ve Sivaranjani (2016); Murcia (2008); Glancy (2011); Li (2015); Li (2021); Deng (2009)
Bulanık Mantık (Fuzzy Logic)	9	Deshmukh vd. (1997); Deshmukh ve Talluru (1998); Ammar vd. (2000); Lenard ve Alam (2004); Pathak vd. (2005); Chai vd. (2006); Lenard vd. (2007); Liang ve Lv (2016 – Fuzzy GA-BPNN)
Random Forest (RF)	7	Eduardo vd. (2021); Balagolla vd. (2021, 2022); Elmougy vd. (2021); Baran Kılıç vd. (2022); Yang vd. (2022); Tan vd. (2021)
Genetik Algoritmalar / Programlama (GA/GP)	6	Welch vd. (1998); Kiehl vd. (2005); Hoogs vd. (2007); Ravisankar vd. (2011); Tsai ve Yen (2008); Liang ve Lv (2016)
Hibrit Yöntemler (karma)	6	Panigrahi vd. (2009); Duman ve Özelçi (2011); Bermúdez vd. (2008); Cecchini vd. (2010); Dong vd. (2014); Humpherys vd. (2011)
Gradient Boosting (XGBoost, LightGBM, GBDT)	5	Ren vd. (2023); Baran Kılıç vd. (2022); Tan vd. (2021); Bartoletti vd. (2021); Yang vd. (2022)
İstatistiksel Testler (Beneish, GAAP vb.)	3	Beneish (1997); Beneish (1999); Beasley vd. (2000)

Metin Madenciliği (Text Mining)	3	Glancy ve Yadav (2011); Cecchini vd. (2010); Zaki ve Theodoulidis (2013)
Self-Organising Map (SOM)	3	Quah ve Sriganesh (2008); Sánchez vd. (2009); Olszewski (2014)
Yapay Bağışıklık Sistemleri (AIS)	3	Wu ve Banzhaf (2008); Wong vd. (2012); Soltani Halvaice ve Akbari (2014)
K-En Yakın Komşu (KNN)	3	Kotsiantis vd. (2006); Kirkos vd. (2007); Yang vd. (2022)
Uzman Sistemler (Expert Systems)	2	Eining vd. (1997); Pacheco vd. (1996)
Stacking (ensemble)	2	Kotsiantis vd. (2006); Kirkos vd. (2007)
Process Mining	1	Jans vd. (2011)
Group Method of Data Handling (GMDH)	1	Ravisankar vd. (2011)
Response Surface Methodology	1	Zhang ve Zhou (2004)
RIPPER (rule-based)	1	Kirkos vd. (2007)
MCDA / UTADIS	1	Spathis vd. (2002)
Autoencoder (AE)	1	Bui Tong Nha vd. (2022)
REP (Reduced Error Pruning)	1	Chaurasia vd. (2022)
Stochastic Gradient Descent (SGD)	1	Martin vd. (2022)

Bu yöntemin tercih edilmesinin nedeni olarak; doğrusal olmayankarmaşık kalıplar, sıralı davranışın modellemesi ve ilişkisel yapıların yakalanması sayılabilir. Nadir olarak kullanılan ancak son yıllarda karşılaşılan modeller incelendiğinde; Otoenkoder (Autoencoder, AE), REP (Reduced Error Pruning) ve Stokastik Gradient Descent (SGD) karşımıza çıkmaktadır. Nha ve Thuan (2022) derin öğrenme tabanlı bir otoenkoder kullanarak finansal anomali tespit etmeyi araştırmıştır. Bu güncel çalışma, denetimsiz öğrenme ile özellik çıkarımı yaparak sahtekârlık tespiti yapmıştır. Chaurasia vd. (2022) karar ağaçlarında bir budama stratejisi olan Reduced Error Pruning (REP) yöntemini finansal dolandırıcılık verisine uygulamıştır. Martin vd. (2022) çalışmasında, makine öğrenimi modellerini eğitirken “Stokastik Gradyan İnişi” optimizasyon tekniğini kullanmıştır. Genel olarak tablo değerlendirildiğinde, derin ve denetimsiz yaklaşımların kullanımı artsa da, en sık kullanılan yaklaşım denetimli olanlardır (Aros vd., 2024: 2). Lojistik regresyon ve SVM gibi doğrusal/yarı-doğrusal sınıflandırıcılar, yorumlanabilirlik ve tablo/rasyo temelli çalışmalara uygunluk nedeniyle en üstte yer almaktadır. Yöntem seçimi veri tipi (metin, işlem akışı, tablolar), etiketlenme seviyesi ve yorumlanabilirlik gereğine göre değişmiştir.



Şekil 2: Yıllara göre VM yöntemlerinin Kullanım Eğilimleri

Şekil 2’de VM yöntemlerinin literatürde kullanım eğilimleri gösterilmektedir. 1985 yılından itibaren çalışmalarda yerini alan veri madenciliği yöntemleri içinde, en yaygın kullanılan ve her zaman popülaritesini koruyan regresyon ve yapay sinir ağları olmuştur. Regresyon yöntemi, 2000’lerin başında ikinci bir sıçrama yaşamış ve bu yöntem neredeyse her dönemde ilgi görmüştür. Yapay Sinir Ağları erken dönem popülarlığından sonra model karmaşıklığı ve veri gereksinimi gibi nedenlerden dolayı seyrinde bir azalma olduğu düşünülmektedir. Karar Ağaçları 2000’ li yılların başında hızlı bir yükseliş yakalamıştır. Ancak tekil ağaçların yerini topluluk yöntemleri aldığından dolayı yavaşlama meydana gelmiştir. Hibrit modeller genel olarak 2018 sonrasında görülür. Ana akım olmamıştır. Nedeni olarak, kurulumu ve bakımının karmaşık olması ve daha çok veri/uzmanlık ve hesaplama gerektirmesi sayılabilir. Gradient Boosting 2020 ve sonrasında önemli bir yükseliş yakalamıştır. Yapılandırılmış (tablolu) veri üzerinde son dönemin “go-to” ailesi (ağaç tabanlı boosting ailesi-XGBoost, LightGBM, CatBoost) içindedir ve doğrusal olmayan çok iyi yakaladığı belirtilir.

Çalışmanın ikinci araştırma sorusu olan “RQ2: Bu yöntemlerin zaman içinde kullanım eğilimleri nasıl değişmiştir?” cevabını bulmak amacıyla oluşturulan zaman grafinin sonuçları genel olarak değerlendirildiğinde, alanın temel istatistiksel modellerden (Regresyon) daha karmaşık makine öğrenmesi modellerine (SVM, Karar Ağaçları) ve oradan da topluluk yöntemlerine (Random Forest, Gradient Boosting) ve nihayetinde derin öğrenmeye doğru evrimi gözlemlenmiştir.

TARTIŞMA VE SONUÇ

Kuruluşlar rekabet avantajı, yönetsel kararlar, denetim ve değerlendirme için bilgiye ihtiyaç duyarlar. İhtiyaç duyulan bilgilerin verilerden çıkarılmasında gelişmiş analitik yöntemlerden biri olan VM yöntemleri, en etkili araçlardan biridir. Şirketin ihtiyaç duyduğu çeşitli bilgiler büyük miktardaki verilerden VM yöntemleri kullanılarak farklı bilgi türlerine göre çıkarılabilmektedir. Bu nedenle kullanıcıların veri madenciliğini daha verimli ve etkili bir şekilde kullanabilmeleri için kapsamlı eğitim ve kurumsal hedeflere ilişkin iyi bir algıya sahip olması gereklidir. Veri türlerine ve ihtiyaç olunan bilgi türlerine göre VM yöntemleri değişiklik göstermektedir. Bu yöntemlerinin seçimi zaman ve maliyet açısından da kritik öneme sahiptir. Özellikle denetim faaliyetlerinde hileli uygulamaların tespitinde kullanılacak VM yöntemlerinin belirlenmesi, şirketlerin uğrayacağı büyük zararların engellenmesini sağlayacaktır. Çalışmada, muhasebede veri madenciliği uygulamalarına ilişkin literatür incelemiş ve finansal muhasebe dolandırıcılığının tespiti için en yaygın olarak kullanılan veri madenciliği tekniklerine dair kavramsal bir çerçeve sunulmuştur. Yapılan analiz sonucunda, literatürde Lojistik Regresyon Yapay Sinir Ağları ve Karar Ağaçları' nin en çok kullanılan yöntemler olduğu tespit edilmiştir. Bu yöntemlerin 1980'li yıllardan beri kullanıldığı görülmüştür. Bu yöntemlerin Liou (2008) tarafından belirtildiği gibi yaygın kullanımının nedeni hem lojistik regresyon hem de sinir ağları yöntemlerinin dolandırıcılığa maruz kalan firmaları doğru bir şekilde sınıflandırabildiğinden dolayı olduğu düşünülmektedir. Ayrıca Liu (2021) tarafından belirtildiği şekilde karar ağacının ağaç benzeri makine öğrenmesi modeli; görüntü tanıma, kelime işleme vb. alanlarda yaygın olarak kullanılmasından dolayı muhasebe davranışlarını ve faaliyetlerini daha iyi denetleyebildiği algısından dolayı talep gören yöntemler arasında yer aldığı söylenebilir. Yine çalışmada yapılan diğer bir analiz sonucunda, kullanılan VM yöntemlerinin zaman içindeki eğiliminin günümüze doğru derin öğrenmeye doğru kaydı görülmüştür.

Kullanılan yöntemlerdeki değişimin sebebi olarak; verilere erişim, verilerdeki yüksek sınıf dengesizliği, dolandırıcılıkların değişen doğası ve çok sayıda yanlış alarm, çeşitli alanlardaki karmaşık problemler ve teknolojinin gelişimi sayılabilir. Çalışmanın alan yazınına katkısı, gelecekteki araştırmalar için potansiyel fırsatları ortaya koyması ve hem akademisyenler hem de uygulayıcılar için çeşitli veri madenciliği teknikleri hakkında yararlı bir bilgiler sağlaması beklenmektedir. Gelecekte yapılacak çalışmalarda; veri sınırlamaları, teknik uzmanlık eksikliği ve kısıtlı bakış açıları gibi zorlukların önüne geçilmesi, bu yöntemlerin pratik uygulamasını engelleyen unsurların ortadan kaldırılması, VM yöntemlerinin uygulamaları ile ilgili ayrıntılı açıklamalar ve yöntem seçim kriterleri gibi alan yazınındaki bazı boşluklarla ilgili çalışmalar yapılabilir.

KAYNAKÇA

- Abbott, L. J., Park, Y., & Parker, S. (2000). The effects of audit committee activity and independence on corporate fraud. *Managerial Finance*, 26(11), 55–67.
- Alagöz, A., Öge, S., & Ortakarpuz, M. (2014). Bir Kurumsal Zekâ Teknolojisi Olarak Veri Madenciliği ile Muhasebe Bilgi Sistemi İlişkisi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (31.1), 1-21.
- Alsharif, M. and Alvi, A. (2021). The impact of data mining on accounting profession: evidence from emerging economies. *International Journal of Emerging Markets*, 16(1), pp. 129-146.
- Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, 24, 32-58.
- Ammar, S., Dunlap, D., & Wright, R. (2000). A neural network approach to detecting financial statement fraud. *Proceedings of the International Conference on Artificial Intelligence*.
- Association of Certified Fraud Examiners. 2018. Global study on occupational fraud and abuse, available at: <https://s3-us-west-2.amazonaws.com/acfepublic/2018-report-to-the-nations.pdf> (accessed 16th February 2020).
- Aslantas, M. (2024). The Effect of Talent Management Strategies on Work Engagement in the Finance Sector: A Study on Bank Employees. *Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi*, 11(2), 290-317
- Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management science*, 49(3), 312-329.
- Bai, Y., Chen, J., & Wang, Y. (2008). A comparative study of data mining methods for financial statement fraud detection. *Journal of Accounting and Finance*, 12(3), 45-59.
- Balagolla, E. M. S. R., Fernando, P. G. T., & Rathnayake, H. M. S. P. (2021). A hybrid model for financial fraud detection using machine learning techniques. *Journal of Financial Crime*.
- Baran Kılıç, M., Akar, G. B., & Güzeliş, C. (2022). A comparative analysis of deep learning models for financial fraud detection. *Expert Systems with Applications*, 207, 117945.
- Bartoletti, M., Carta, S., & Onnio, M. (2021). A survey on fraud detection with machine learning. *ACM Computing Surveys*, 54(6), 1–38.
- Beasley, M. S. (1996). An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review*, 71(4), 443–465.
- Beasley, M. S., Carcello, J. V., & Hermanson, D. R. (2000). Fraudulent financial reporting: Consideration of industry traits and corporate governance mechanisms. *Accounting Horizons*, 14(4), 441–454.
- Bell, T. B., & Carcello, J. V. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, 19(1), 169–184.

- Beneish, M. D. (1997). Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy*, 16(3), 271–309.
- Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24–36.
- Bermúdez, L., Pérez, J. M., Ayuso, M., & Gómez, E. (2008). A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics*, 42(2), 779–786.
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613.
- Bose, I., & Wang, J. (2007). Data mining for detection of financial statement fraud in China. *Journal of Emerging Technologies in Accounting*, 4(1), 1–20.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.
- Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7), 1146–1160.
- Cerullo, M. J., & Cerullo, V. (1999). Using neural networks to predict financial reporting fraud: Part 1. *The CPA Journal*, 69(5), 42–47.
- Chahadah, A. A., Refae, G. A. E., & Qasim, A. (2018). The use of data mining techniques in accounting and finance as a corporate strategic tool: an empirical investigation on banks operating in emerging economies. *International Journal of Economics and Business Research*, 15(4), 442–452.
- Chai, S., Chen, Y., & Huang, B. (2006). A fuzzy logic approach for fraud detection in financial statements. *Journal of Intelligent & Fuzzy Systems*, 17(4), 345–356.
- Chaurasia, V., Pal, S., & Tiwari, B. B. (2022). A hybrid machine learning model for credit card fraud detection. *International Journal of Information Technology*, 14(2), 1103–1113.
- Chen, L. D., Sakaguchi, T., & Frolick, M. N. (2000). Data mining methods, applications, and tools. *Information systems management*, 17(1), 65–70.
- Chye Koh, H., & Kee Low, C. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal*, 19(3), 462–476.
- Coman, D. M., Mieiă, M., Voinea, C. M., Tănase, L. C., & Necula, A. I. (2025). Determinants of the Accounting Services' Outsourcing Amid the Business' Digitization in Romania: An Analytical Examination. *Journal of East European Management Studies*, 30(2), 39038.
- Debreceeny, R. S., & Gray, G. L. (2010). Data mining journal entries for fraud detection: An exploratory study. *International Journal of Accounting Information Systems*, 11(3), 157–181.
- Deng, X. (2009). *An integrated framework for financial fraud detection using data mining* [Doctoral dissertation, University of Technology, Sydney].

- Deng, X. (2017). A hybrid model for financial fraud detection: An empirical study in China. *Journal of Forensic & Investigative Accounting*, 9(1), 145–160.
- Deshmukh, A., & Talluru, L. (1998). A rule-based fuzzy logic system for fraud detection in financial statements. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (Vol. 4, pp. 3723-3728).
- Deshmukh, A., Romine, J., & Siegel, P.H. (1997). Measurement and combination of red flags to assess the risk of management fraud: a fuzzy set approach, *Managerial Finance* 23 (6) 35–48.
- Dong, W., Liao, S., & Zhang, Z. (2014). The impact of management integrity on audit efficiency: Evidence from China. *China Journal of Accounting Research*, 7(1), 59–78.
- Duman, E., & Özelçi, Ü. (2011). Detecting credit card fraud by genetic algorithm and scatter search. *Expert Systems with Applications*, 38(10), 13057–13063.
- Eduardo, S., Pérez, M., & López, V. (2021). A novel ensemble method for fraud detection using machine learning. *Applied Soft Computing*, 100, 106991.
- Eining, M. M., Jones, D. R., & Loebbecke, J. K. (1997). Reliance on decision aids: An examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice & Theory*, 16(2), 1–19.
- Elmougy, S., Tolba, A., & Hamdy, N. (2021). A hybrid feature selection model for financial fraud detection. *IEEE Access*, 9, 128310–128322.
- Fanning, K. M., & Cogger, K. O. (1998). Neural network detection of management fraud using published financial data. *Intelligent Systems in Accounting, Finance and Management*, 7(1), 21–41.
- Fanning, K. M., Cogger, K. O., & Srivastava, R. P. (1995). Detection of management fraud: A neural network approach. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 4(2), 113–126.
- Farooq Aziz. (2023). Data analytics impacts in the field of accounting. *World Journal of Advanced Research and Reviews*, 18(2), 946–951. doi:10.30574/wjarr.2023.18.2.0863.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- Feng, M., Zheng, J., Han, Y., Ren, J., & Liu, Q. (2018). Big data analytics and mining for crime data analysis, visualization and prediction. In *International conference on brain inspired cognitive systems* (pp. 605-614). Cham: Springer International Publishing.
- Feroz, E. H., Kwon, T. M., Pastena, V. S., & Park, K. (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 & Management*, 9(3), 145–157.
- Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, 50(3), 595–601.
- Gray, G. L., & Debreceeny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, 15(4), 357-380.

- Green, B. P., & Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing: A Journal of Practice & Theory*, 16(1), 14–28.
- Guoxin, L., Hua, Z., & Wei, L. (2007). The application of data mining in financial statement fraud detection. *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing*.
- Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud—A comparative study of machine learning methods. *Knowledge-Based Systems*, 128, 139–152.
- Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan kaufmann*, 340(1), 94104-103205.
- Hansen, J. V., McDonald, J. B., & Stice, J. D. (1996). Artificial intelligence and generalized qualitative-response models: An empirical test on two audit decision-making domains. *Decision Sciences*, 27(2), 229–255.
- Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: a review of anomaly detection techniques and recent advances. *Expert systems With applications*, 193, 116429.
- Hoogs, B., Kiehl, T., Lacombe, C., & Senturk, D. (2007). A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance and Management*, 15(1-2), 41–56.
- Huang, S. Y. (2006). An empirical study on the detection of financial statement fraud using data mining. *Journal of American Academy of Business*, 9(2), 89–95.
- Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), 585–594.
- Jain, N., & Srivastava, V. (2013). Data mining techniques: a survey paper. *IJRET: International Journal of Research in Engineering and Technology*, 2(11), 2319-1163.
- Jans, M., Lybaert, N., & Vanhoof, K. (2011). A framework for internal fraud risk reduction: IT support for forensic accountants. *International Journal of Digital Accounting Research*, 11, 1–29.
- Jun Lee, S., & Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Systems*, 101(1), 41-46.
- Kandemir, T., & Kardeş, Z. (2025). Hileli Finansal Tabloların Tespitinde Veri Madenciliği Uygulamaları: Mevcut Araştırma Eğilimlerinin İncelenmesi (2006-2024). *Denetışim* (33), 333-355. <https://doi.org/10.58348/denetisim.1674341>.
- Karami, M., Baber, W. W., & Ojala, A. (2022). The effectual process of business model innovation for seizing opportunities in frontier markets. *Technovation*, 117, 102595.
- Kiehl, T., Hoogs, B., & LaComb, C. (2005). A genetic algorithm for detecting temporal patterns indicative of financial statement fraud. *Proceedings of the Genetic and Evolutionary Computation Conference*.
- 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). Artificial neural networks for analytical review in auditing. *Turku Centre for Computer Science*.
- 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.
- Krambia-Kapardis, M., Christodoulou, C., & Agathocleous, M. (2010). Neural networks: The panacea in fraud detection? *Managerial Auditing Journal*, 25(7), 659–678.
- Kumar, M. K., & Kar, P. K. (2020). A Study on Privacy Preserving in Big Data Mining Using Fuzzy Logic Approach. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 11, 2108-2116.
- Laourou, A. B. F. (2025). The adoption and implementation of data mining in accounting information systems (AIS) within the public sector of Nigeria. *IIARD Journals*. <https://doi.org/10.56201/jafm>, 11.
- Lenard, M. J., & Alam, P. (2004). An evolutionary approach to data mining for fraud detection in financial statements. *Journal of Emerging Technologies in Accounting*, 1(1), 69–82.
- Li, J. (2015). *Financial fraud detection based on text mining and sentiment analysis* [Doctoral dissertation, University of Maryland].
- Li, J. (2021). Advanced machine learning techniques for financial fraud detection: A review. *Journal of Financial Analytics*, 4(1), 45-62.
- Liang, D., & Lv, J. (2016). A fuzzy GA-BPNN model for financial fraud detection. *Journal of Intelligent & Fuzzy Systems*, 30(4), 2309–2318.
- 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.
- Liou, F. M. (2008). Fraudulent financial reporting detection using business intelligence and data mining. *Journal of Modern Accounting and Auditing*, 4(11), 38–47.
- Liu, Q., & Vasarhelyi, M. A. (2014). Big questions in AIS research: Measurement, information processing, data analysis, and reporting. *Journal of information systems*, 28(1), 1-17.
- Liu, X. (2021). Empirical analysis of financial statement fraud of listed companies based on logistic regression and random forest algorithm. *Journal of Mathematics*, 2021(1), 9241338.
- Martin, K., Sanders, E., & Baird, J. (2022). Deep learning for anomaly detection in financial networks. *Journal of Financial Stability*, 60, 100992.
- Meenakshi, S., & Sivaranjani, S. (2016). A survey on credit card fraud detection using data mining techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(5), 268–272.
- Mercan, M., Kitesashvili, D., Aslantas, M. (2023). Effect of Financial Literacy on Financial Well-Being in Georgia. *Journal of Business*, Vol. 12 No. 2

- Mraović, B. (2008). Relevance of data mining for accounting: social implications. *Social Responsibility Journal*, 4(4), 439-455.
- Murcia, F. C. (2008). *Fraud detection in financial statements: An empirical study of Brazilian public companies* [Doctoral dissertation, University of São Paulo].
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3), 559-569.
- Nha, B. T., & Thuan, N. D. (2022, December). Methodology Interaction by Machine Learning Model to Detect Vulnerability in Smart Contract of Blockchain. In 2022 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 112-117). IEEE.
- Olszewski, D. (2014). Fraud detection using self-organizing maps. *Proceedings of the International Conference on Artificial Intelligence and Soft Computing*.
- Owusu-Ansah, S., Moyes, G. D., Oyelere, P. B., & Hay, D. (2002). An empirical analysis of the likelihood of detecting fraud in New Zealand. *Managerial Auditing Journal*, 17(4), 192–204.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2(3), 65-82.
- Pacheco, R., Salgado, L., & Eining, M. (1996). The impact of decision aids on auditor's fraud risk assessments. *Advances in Accounting Information Systems*, 4, 141–159.
- Page MJ, McKenzie JE, Bossuyt PM, ve ark. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372 (71):1-9. doi:10.1136/bmj.n71
- Panigrahi, S., Kundu, A., Sural, S., & Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Information Fusion*, 10(4), 354–363.
- Papík, M., & Papíková, L. (2022). Detecting accounting fraud in companies reporting under US GAAP through data mining. *International Journal of Accounting Information Systems*, 45, 100559.
- Pathak, J., Vidyarthi, N., & Summers, S. L. (2005). A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. *Managerial Auditing Journal*, 20(6), 632–644.
- Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, 30(2), 19–50.
- Pinquet, J., Ayuso, M., & Guillén, M. (2007). Selection bias and auditing policies for insurance claims. *The Journal of Risk and Insurance*, 74(2), 425–440.
- Popat, R. R., & Chaudhary, J. (2018, May). A survey on credit card fraud detection using machine learning. In 2018 2nd international conference on trends in electronics and informatics (ICOEI) (pp. 1120-1125). IEEE.
- Quah, J. T. S., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications*, 35(4), 1721–1732.

- Ramzan, M., & Ahmad, M. (2014, March). Evolution of data mining: An overview. In 2014 Conference on IT in Business, Industry and Government (CSIBIG) (pp. 1-4). IEEE.
- Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50(2), 491–500.
- Ren, Y. (2006). *An empirical study on financial fraud detection in Chinese listed companies* [Doctoral dissertation, Shanghai University of Finance and Economics].
- Ren, Y., Zhang, L., & Wang, P. (2023). A novel hybrid deep learning model for financial fraud detection in imbalanced datasets. *Information Sciences*, 621, 703-720.
- Rezaee, Z. (2005). Causes, consequences, and deterrence of financial statement fraud. *Critical perspectives on Accounting*, 16(3), 277-298.
- Ryan, C. L., Cant, R., McAllister, M. M., Vanderburg, R., & Batty, C. (2022). Transformative learning theory applications in health professional and nursing education: An umbrella review. *Nurse Education Today*, 119, 105604. <https://doi.org/10.1016/j.nedt.2022.105604>
- Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in society*, 24(4), 483-502.
- Sağlar, J., & Kefe, İ. (2021). A Review on Data Mining Methods used in Internal Audit and External Audit. *EKEV Akademi Dergisi*. (88), 259-274.
- Sahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, 40(15), 5916–5923.
- Sánchez, D., Vila, M. A., Cerda, L., & Serrano, J. M. (2009). Association rules applied to credit card fraud detection. *Expert Systems with Applications*, 36(2), 3630–3640.
- Seifert, J. W. (2004). Data mining: An overview. *National security issues*, 201-217.
- Seng, J. L., & Chen, T. C. (2010). An analytic approach to select data mining for business decision. *Expert Systems with Applications*, 37(12), 8042-8057.
- Sharma, A., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. *arXiv preprint arXiv:1309.3944*.
- Soltani Halvaiee, N., & Akbari, M. K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. *Applied Soft Computing*, 24, 40–49.
- Söylemez, Y. & Türkmen, S. Y. (2017). Yapay Sinir Ağları Modeli ile Finansal Başarısızlık Tahmini. *Finans Ekonomi ve Sosyal Araştırmalar Dergisi*, 2(4), 270-284.
- Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: A comparative study using multicriteria analysis and multivariate statistical techniques. *European Accounting Review*, 11(3), 509–535.
- Summers, S. L., & Sweeney, J. T. (1998). Fraudulently misstated financial statements and insider trading: An empirical analysis. *The Accounting Review*, 73(1), 131–146.
- Tan, R., Zhang, X., & Li, Y. (2021). A hybrid machine learning approach for financial fraud detection using multi-source data. *Electronic Commerce Research and Applications*, 48, 101070.

- Tangod, S., & Kulkarni, S. (2015). A survey on credit card fraud detection using data mining techniques. *International Journal of Computer Applications*, 118(15), 31–34.
- Theodorakopoulos, L., Thanasas, G., & Halkiopoulou, C. (2024). Implications of big data in accounting: Challenges and opportunities. *Emerging Science Journal*, 8(3), 1201-1214.
- Tsai, C. F., & Yen, D. C. (2008). A comparative study of data mining techniques for corporate bankruptcy prediction. *Expert Systems with Applications*, 35(3), 756–763.
- Tüzüntürk, S. (2010). Veri madenciliği ve istatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 29(1), 65-90.
- Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: An overview. *Accounting Horizons*, 29(2), 381-396.
- Viaene, S., Ayuso, M., Guillen, M., Gheel, D. V., & Dedene, G. (2004). Strategies for detecting fraudulent claims in the automobile insurance industry. *European Journal of Operational Research*, 176(1), 565–583.
- Virdhagriswaran, S. (2006). *Data mining for insurance fraud detection* [Doctoral dissertation, University of Texas at Arlington].
- Warren, J. D., Moffitt, K. C., & Byrnes, P. (2015). How big data will change accounting. *Accounting horizons*, 29(2), 397-407.
- Welch, J. R., Reeves, T. P., & Welch, S. A. (1998). Using genetic algorithms to create a network for detecting management fraud. *Intelligent Systems in Accounting, Finance and Management*, 7(3), 161–174.
- West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, 57, 47-66.
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1), 30–55.
- Wong, N., Ray, P., Stephens, G., & Lewis, L. (2012). Artificial immune systems for the detection of credit card fraud: An assessment. *International Journal of Computer Applications*, 42(17), 26–32.
- Wu, S. X., & Banzhaf, W. (2008). A hybrid method for credit card fraud detection based on evolutionary computation. *Proceedings of the 10th annual conference on Genetic and evolutionary computation*.
- Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1), 97-107.
- Xiao, M., Xiaoli, H., & Gaojin, L. (2010). Research on application of data mining technology in financial decision support system. In 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering (Vol. 4, pp. 381-384). IEEE.
- Yang, Z., Liu, Y., & Chen, X. (2022). Financial fraud detection using graph neural networks on transaction data. *Knowledge-Based Systems*, 240, 108079.

- Yong Ren, V., Zhang, L., & Wang, P. (2023). A novel hybrid deep learning model for financial fraud detection in imbalanced datasets. *Information Sciences*, 621, 703-720. (Not: Bu kaynak "Ren vd. (2023)" ile aynı olabilir, kontrol edilmelidir)
- Yuan, L., Chen, Y., & Zhang, S. (2008). The application of data mining in financial fraud detection: Evidence from China. *China Economic Review*, 19(4), 693–700.
- Yue, D., Wu, X., & Wang, Y. (2007). A review of data mining-based financial fraud detection research. *International Conference on Wireless Communications, Networking and Mobile Computing*.
- Zaki, M. J., & Theodoulidis, B. (2013). Fraud detection in financial statements using meta-learning. *Intelligent Data Analysis*, 17(4), 689–704.
- Zhang, J., Li, T., & Chen, H. (2014). Composite rough sets for dynamic data mining. *Information Sciences*, 257, 81-100.
- Zhang, Y., & Zhou, J. (2004). A novel model for credit card fraud detection using ensemble learning. *Proceedings of the International Conference on Machine Learning and Cybernetics*.
- Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems*, 50(3), 570–575.