

İnceleme Makalesi

**HİLELİ FİNANSAL TABLOLARIN TESPİTİNDE VERİ MADENCİLİĞİ UYGULAMALARI:
MEVCUT ARAŞTIRMA EĞİLİMLERİNİN İNCELENMESİ (2006-2024)**

(DATA MINING PRACTICES IN THE DETECTION OF FRAUDULENT FINANCIAL STATEMENTS:

A REVIEW OF THE CURRENT RESEARCH TRENDS (2006-2024))

Tuğrul KANDEMİR¹ Zafer KARDEŞ²

ÖZ

Finansal piyasaların giderek karmaşıklaşması ve mevcut verilerin katlanarak artması, hileli finansal tablo tespitini kritik ve zor bir konu haline getirmiştir. Büyük hacimli verilerden ve finansal verilerin karmaşıklığından ilgili bilgiyi çıkarıp hileli finansal tablo tespit etmeye olanak sağlayan çeşitli veri madenciliği yöntemleri geliştirilmiştir. Son yıllarda veri madenciliği yöntemleri, hileli finansal tablo tespiti için en güvenilir yöntemlerden biri haline gelmiştir. Bu araştırmanın amacı, hileli finansal tablo tespitinde veri madenciliğinin uygulanmasına ilişkin mevcut araştırmalar hakkında bir inceleme sağlamak ve bulguları karşılaştırmaktır. Bu araştırmada bilimsel dergilerde yayınlanan, hileli finansal tablo tespitinde veri madenciliği yöntemlerini uygulayan araştırmaların nicelik açısından değerlendirilmesiyle birlikte araştırmacılara ve uygulayıcılara bir veri tabanı oluşturulması amaçlanmıştır. Araştırma sonucunda, en yaygın kullanılan veri madenciliği yöntemlerinin sırasıyla karar ağacı, destek vektör makineleri, lojistik regresyon ve yapay sinir ağı olduğu görülmüştür. Araştırmada en iyi performans gösteren yöntemlerin ise lojistik regresyon, hibrit modeller, sinir ağı ve XGBoost algoritmasının olduğu ve ele alınan araştırmaların %56,10'unda örneklem büyüklüklerinin 1.000'den az olduğu tespit edilmiştir.

Anahtar Kelimeler: Hile Tespiti, Hileli Finansal Tablolar, Veri Madenciliği, Makine Öğrenimi, Literatür Taraması

JEL Kodları: M40, M42, C38

ABSTRACT

The increasing complexity of financial markets and the exponential increase in available data have made fraudulent financial statement detection a critical and difficult issue. Various data mining methods have been developed that make it possible to extract relevant information from large amounts of data and the complexity of financial data in order to recognize fraudulent financial statements. In recent years, data mining methods have become one of the most reliable methods for fraudulent financial statement detection. This study aims to provide a review of existing research on the application of data mining in fraudulent financial statement detection and compare the findings. This study aims to quantitatively evaluate research published in scientific journals that applies data mining methods to fraudulent financial statement detection and to create a database for researchers and practitioners. The study revealed that the most frequently used data mining methods were decision trees, support vector machines, logistic regression, and artificial neural networks. It was also detected that the best performing methods were logistic regression, hybrid models, neural network and XGBoost algorithm, and 56.10% of the studies had sample sizes less than 1,000.

Keywords: Fraud Detection, Fraudulent Financial Statements, Data Mining, Machine Learning, Literature Review

JEL Classification: M40, M42, C38

¹ Prof. Dr., Afyon Kocatepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, kandemir@aku.edu.tr, orcid.org/0000-0002-3544-7422

² Öğr. Gör. Dr., Afyon Kocatepe Üniversitesi, Emirdağ Meslek Yüksekokulu, zaferkardes@aku.edu.tr, orcid.org/0000-0002-5719-8551

1. GİRİŞ

Finansal tablolar, bir işletmenin finansal faaliyetlerini ve finansal durumunu ortaya koyan gelirlerini, giderlerini, kredilerini, alacaklarını, zararlarını, kârlarını vb. içeren kayıtlar olarak tanımlanmaktadır. Bu kayıtlar finansal tablo kullanıcılarına yönetimin işletme performansı ve olası sorunları için yaptığı bazı açıklamaları içermektedir. Temel finansal tablolar, işletme durumunun net bir resmini sunarak, işletmenin ne düzeyde başarılı olduğunu göstermekte, bunun bir sonucu olarak da, işletmenin güvenilir olup olmadığını belirlemeye yardımcı olmaktadır. Finansal tablo hilesi, işletmenin avantajlı görünmesini sağlamak için yanlış beyanlarda bulunarak finansal tablo kullanıcılarını yanıltmayı içermektedir. Bu tür hilenin arkasındaki itici gücün hisse senedi fiyatlarını artırmak, kişisel ya da kurumsal banka kredileri almak, vergi yükümlülüklerini azaltmak veya mümkün olduğu kadar çok yatırımcı çekmek olduğu görülmektedir (Al-Hashedi & Magalingam, 2021).

Association of Certified Fraud Examiners (ACFE) tarafından yayınlanan rapora göre muhasebe hilesi varlıkların suistimal edilmesi, yolsuzluk ve finansal tablo hilesi olmak üzere üç temel kategoriye ayrılmaktadır. ACFE tarafından yayınlanan raporda, tüm hilelerin yalnızca % 10'unu finansal tablo hileleri oluşturmalarına rağmen, finansal tablo hilelerinin neden olduğu finansal kaybın oldukça yüksek olduğu belirtilmiştir (ACFE, 2020). Finansal tablo hilesi genellikle kurumsal hile (Corporate fraud) olarak adlandırılmaktadır. Rezaee (2005) tarafından finansal tablo hilesi, işletmelerin önemli ölçüde yanlış beyanda bulunan finansal tabloları sunarak, finansal tablo kullanıcılarını özellikle yatırımcıları ve alacaklıları aldatmaya veya yanıltmaya yönelik kasıtlı girişimler olarak tanımlanmıştır. American Institute of Certified Public Accountant (AICPA) tarafından hileli finansal raporlama, finansal tabloların tüm önemli yönleriyle genel kabul görmüş muhasebe ilkelerine uygun olarak sunulmamasına neden olan, finansal tablo kullanıcılarını aldatmak için tasarlanmış, finansal tablolardaki tutar veya açıklamaların kasıtlı olarak yanlış beyan edilmesi veya ihmal edilmesi şeklinde ifade edilmiştir (AICPA, 2021).

SPK tarafından yayımlanan tebliğe göre hileli finansal raporlama; “finansal tablolarda yer alması gereken tutarları veya açıklamaları atlayarak veya yanlış sunarak finansal tablo kullanıcılarını kasten yanıltmak” olarak ifade edilmiştir. Hileli finansal raporlama aşağıdaki eylemleri de içermektedir: (a) Muhasebe kayıtları veya dokümanlar üzerinde sahtecilik (belge sahtekarlığı), tahrifat, manipülasyon ve/veya gerçeğe aykırı belge düzenleme, (b) Finansal tablolara yansıtılması gereken olayların, işlemlerin ve diğer önemli bilgilerin kasıtlı olarak ihmal edilmesi veya yanlış beyan edilmesi ve (c) Finansal tablolara yansıtılması gereken olayların, işlemlerin ve diğer önemli bilgilerin muhasebeleştirilmesi, ölçülmesi, sınıflandırılması, sunumu ve açıklanmasında muhasebe ilkelerinin kasıtlı olarak yanlış uygulanması (SPK, Seri X, No.22).

Veri madenciliği ve muhasebe hile tespiti arasındaki bağlantı, gelişmiş bir analitik araç olarak veri madenciliğinin denetçilere karar verme ve hileleri tespit etmede yardımcı olmasından kaynaklanmaktadır. Veri madenciliği yöntemleri, hile tespitinin etkisi ve verimliliği arasındaki çelişkiyi çözme potansiyeline sahiptir. Veri madenciliği, genellikle çok büyük veri kümelerindeki gizli kalıpları çıkarmak ve keşfetmek için kullanıldığından, hileli finansal tabloların tespitinde önemli bir rol oynamaktadır. Bir denetçinin hileli bir işlemin varlığından ve arkasındaki niyete ilişkin mutlak bir yargıya ulaşması, her zaman mümkün değildir. Denetçi açısından bakıldığında en uygun maliyetli seçenek, mevcut verilerden yeterli hile kanıtı elde etmek için özel matematiksel ve veri işleme algoritmalarını kullanmaktır (Sharma & Panigrahi, 2013). Veri madenciliği yöntemleri, son yıllarda muhasebe hilelerini tespit etmek için en güvenilir metodolojilerden biri olarak kabul edilmektedir (Ngai ve diğerleri, 2011; Zhou & Kapoor, 2011).

Bu araştırmanın amacı, hileli finansal tablo tespitinde veri madenciliğinin uygulanmasına ilişkin mevcut araştırmalar üzerine bir inceleme yaparak bulguları karşılaştırmaktır. Araştırma, özellikle aşağıdaki soruları ele almaktadır.

- Araştırmalarda yer alan veri madenciliği, makine öğrenimi, derin öğrenme ve yapay zekâ yöntemleri nelerdir?
- En iyi performans gösteren yöntemler nelerdir?
- Araştırmalarda kaç farklı yöntem kullanılmıştır?
- Araştırmaların veri seti örneklem büyüklükleri nedir?
- Araştırmaların yayın yılına göre dağılımı nasıldır?
- Mevcut araştırmaların ülkelere göre dağılımı nedir?
- Araştırmaların yayınlandığı dergilere göre dağılımı nasıldır?

Tuğrul KANDEMİR, Zafer KARDEŞ

Bu araştırmanın, mevcut olan literatür hakkında gelecekte yapılacak araştırmalar için bir gösterge sağlaması beklenmektedir. Ayrıca, araştırmacıların çakışan çabalardan kaçınmasına ve uygulamalarını yeni gelişmelerle uyumlu hale getirmesine yardımcı olacağı düşünülmektedir.

Araştırmanın geri kalan kısmı şu şekilde organize edilmiştir. İkinci bölümde, literatürde bulunan başlıca araştırmalara yer verilmiştir. Üçüncü bölümde, araştırma metodolojisi sunulmuştur. Dördüncü bölümde tartışmaya yer verilirken, son bölümde ise araştırmanın sonuçları ve önerileri açıklanmıştır.

2. LİTERATÜR TARAMASI

Literatürde farklı finansal hile türlerinin tespitinde veri madenciliği yöntemlerini kullanan çalışmalar bulunmaktadır. Literatürde bu kapsamda yer alan mevcut araştırma eğilimlerini inceleyen başlıca çalışmalar aşağıda Tablo 1’de sunulmuştur.

Tablo 1. İlgili Literatür İncelemelerinin Karşılaştırılması

Yazarlar	Yıl	Kapsadığı Dönem	İncelenen Çalışma Sayısı	Hile Kategorisi
Yue ve diğerleri	2007	1995-2007	18	Finansal hile
Ngai ve diğerleri	2011	1997-2008	49	Banka, sigorta, menkul kıymetler ve emtia, diğer ilgili finansal hileler
Sharma & Panigrahi	2013	1995-2011	45	Finansal hile
Albashrawi	2016	2004-2015	65	Finansal tablo, banka, sigorta, ilgili diğer finansal hileler
Barman ve diğerleri	2016	1997-2014	27	Banka, finansal tablo, sigorta hilesi
Esen	2016	1994-2015	79	Banka, sigorta, kıymetli evrak, diğer hileler
West & Bhattacharya	2016	2004-2014	29	Banka, finansal tablo, sigorta hilesi
Selimoğlu & Altunel	2019	2008-2018	65	Finansal tablo hilesi ve hile denetimi
Al-Hashedi & Magalingam	2021	2009-2019	75	Banka, sigorta, finansal tablo, kripto para hilesi
Ashtiani & Raahemi	2021	2020 Öncesi	47	Finansal tablo hilesi
Gupta & Mehta	2021	1995-2019	28	Banka, finansal tablo, sigorta hilesi
Shahana ve diğerleri	2023	1995-2022	84	Finansal tablo hilesi
Beemamol	2024	1989-2022	24	Finansal tablo hilesi
Ramzan & Lokanan	2024	2011-2022	88	Finansal tablo hilesi
Gupta & Mehta	2024	1995-2020	28	Finansal tablo hilesi
Rao & Mandhala	2024	2016-2018	72	Finansal tablo hilesi

Literatür incelendiğinde finansal hile alanının önceki araştırmalarda büyük ilgi gördüğü ve farklı finansal hile türlerinin ele alındığı görülmektedir. Finansal hile faaliyetleri sektörlere göre değişiklik göstermektedir. Finansal hilenin farklı araştırmalara göre (Albashrawi, 2016; Al-Hashedi & Magalingam, 2021; Ashtiani & Raahemi, 2021; Esen, 2016; Ngai ve diğerleri, 2011) finansal tablo hilesi, banka hilesi, sigorta hilesi, menkul kıymet hilesi, vergi hilesi, kripto para hilesi gibi farklı türlere ayrıldığı görülmektedir. Finansal hile türlerine göre literatürü inceleyen çeşitli çalışmalar (Albashrawi, 2016; Al-Hashedi & Magalingam, 2021; Barman ve diğerleri, 2016; Esen, 2016; Gaffaroglu & Alp, 2023; Gupta & Mehta 2021; Ngai ve diğerleri, 2011; West & Bhattacharya, 2016) mevcuttur. Albashrawi (2016) tarafından yapılan araştırma sonucunda, incelenen 65 makalenin 41’i (%63) hile tespitinde araştırılan en büyük iki finansal uygulama alanı olarak finansal tablo hilesi ve banka hilesi olduğu tespit edilmiştir.

Literatürde sadece hileli finansal tabloların tespitine yönelik yapılan araştırmaları inceleyen (Beemamol, 2024; Shahana ve diğerleri, 2023) çalışmalar bulunmaktadır. Bununla birlikte, hileli finansal tabloların tespitinde veri madenciliği, makine öğrenimi, yapay zekâ ve metin madenciliği gibi yöntemlerin sonuçlarını değerlendiren (Ashtiani & Raahemi, 2021; Gupta & Mehta, 2021; Ramzan & Lokanan, 2024; Rao & Mandhala 2024; Selimoğlu & Altunel, 2019; Sharma & Panigrahi, 2013; Uğurlu & Sevim 2015a; Yue ve diğerleri, 2007) araştırmalar da mevcuttur.

Mevcut literatür incelendiğinde en az 3 yıllık (2016-2018) dönemi (Rao & Mandhala, 2024) ve en fazla 34 yıllık (1989-2022) dönemi (Beemamol, 2024) kapsayan araştırmaların olduğu tespit edilmiştir. Söz konusu araştırmalarda incelenen

Tuğrul KANDEMİR, Zafer KARDEŞ

çalışma sayıları 18 (Yue ve diğerleri, 2007) ile 88 (Ramzan & Lokanan, 2024) arasında değişiklik göstermektedir. Veri madenciliğine dayalı finansal hile tespiti için genel bir çerçeve tasarımı önerilerine bazı çalışmalarda (Sharma & Panigrahi, 2013; Yue ve diğerleri, 2007) yer verilmiştir. İlgili literatürde, ilk inceleme araştırmasının 2007 yılında (Yue ve diğerleri, 2007) yapıldığı görülmektedir. Son yıllarda ise literatürde (Ramzan & Lokanan, 2024; Rao & Mandhala 2024; Beemamol, 2024) önemli bir artış olduğu görülmektedir. Mevcut literatürde konuya ilişkin artan ilgi ve önem açık bir şekilde gözlenmektedir.

Mevcut araştırmalardan bazıları (Ashtiani & Raahemi, 2021; Gupta & Mehta 2021; Ramzan & Lokanan, 2024; Shahana ve diğerleri, 2023) analize dâhil edilecek çalışmaları belirlerken bir araştırma protokolünü takip etmiştir. Konu kapsamında Türkiye örnekleminde yapılan literatür incelemelerinde, sistematik bir araştırma protokolü izleyen araştırmaya rastlanmamıştır. Bu araştırmada PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2020 araştırma protokolü izlenmiştir.

Mevcut araştırmalarda kullanılan veri tabanları incelendiğinde ACM DL, EBSCO, Emerald, IEEE Xplore, JSTOR, ProQuest, Sage Publications, ScienceDirect, Scopus, Springerlink, Taylor and Francis, Web of Science, Wiley, World Scientific ve Yök Tez Merkezi veri tabanlarında yapıldığı tespit edilmiştir. Bu araştırma hem İngilizce hem de Türkçe literatürü kapsamı nedeniyle ScienceDirect, Scopus, Web of Science, DergiPark ve TRDizin veri tabanları araştırma kapsamına alınmış ve ilgili veri tabanlarında taramalar gerçekleştirilmiştir.

Literatürde yer alan araştırmalar genel olarak değerlendirildiğinde, hile türlerine göre kategorilere ayrılarak veri madenciliği uygulamalarına yer verildiği gözlenmektedir. Albashrawi (2016), finansal hile türlerini finansal tablo hilesi, banka hilesi, sigorta hilesi ve diğer ilgili finansal hileler olmak üzere dört kategoride sınıflandırmıştır. Araştırmada hile türleri içerisinde en çok finansal tablo hilelerinin incelendiği tespit edilmiştir. Bu bulgu neticesinde bu araştırmada finansal hile türlerinden hileli finansal tablo türüne odaklanılmıştır. Bununla birlikte, elde edilen literatür taraması ışığında, Türkiye örnekleminde mevcut inceleme araştırmalarının spesifik olarak hileli finansal tablo ve veri madenciliği performanslarının karşılaştırmalı olarak sunulmadığı tespit edilmiş ve araştırma boşluğu belirlenmiştir. Dolayısıyla bu araştırmanın, hileli finansal tablo tespitinde veri madenciliği yöntemlerine yönelik araştırma eğilimlerini inceleyerek literatüre katkı sağlaması beklenmektedir.

3. METODOLOJİ

Bu bölümde, araştırmanın amacı, kapsamı ve bulgularına ilişkin bilgilere yer verilmiştir.

3.1. Araştırmanın Amacı ve Kapsamı

Araştırmanın amacı, hileli finansal tabloları tespit etmek için veri madenciliği yöntemlerini uygulayan araştırmalar üzerine inceleme yaparak ortaya konulan bulguları karşılaştırmaktır. Sistematik İncelemeler ve Meta Analizler için Tercih Edilen Raporlama Öğeleri (Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)) kılavuzu, sistematik incelemelerde incelemenin neden yapıldığını, yazarların ne yaptığını ve ne bulduklarını şeffaf bir şekilde raporlamalarına yardımcı olmak amacıyla hazırlanmıştır. Bu araştırmada sonuçları raporlamak ve tüm aşamalardaki bilgi akışını kanıta dayalı ayrıntılarla şeffaf bir yöntemle açıklamak için PRISMA 2020 akış şeması izlenmiştir. PRISMA 2020, araştırmaları belirleme, seçme, değerlendirme ve sentezleme yöntemlerindeki gelişmeleri yansıtan raporlama rehberliği içermektedir (Page ve diğerleri, 2021). Araştırmada, ilgili araştırmaları belirleme süreci Şekil 1’de gösterilen PRISMA 2020 akış şemasında sunulmuştur.

Literatür taramasına yönelik anahtar kelimeler ve arama dizeleri bulmak, sistematik bir arama sürecinin ilk adımıdır (Tranfield ve diğerleri, 2003). Bu nedenle, literatür taraması ScienceDirect, Scopus, Web of Science, DergiPark ve TRDizin olmak üzere toplam beş veri tabanı üzerinden 15 Şubat 2025 tarihinde (arama bitiş tarihi) gerçekleştirilmiştir. İlk arama sonuçlarına göre 600 araştırmaya ulaşılmıştır. İlgili araştırmaları belirlemek için çok sayıda anahtar kelime kullanarak arama kriterleri gerçekleştirilmiştir. Ancak ilgili sonuçlara ulaşan en kabul edilebilir arama kriterleri Tablo 2’de sunulmuştur.

Tablo 2. Veri Tabanlarında Kullanılan Arama Dizesi ve Arama Sonuçları

Veri Tabanı	Uygulanan Sınır	Arama Sorgusu	Ulaşılan Araştırma Sayısı
ScienceDirect	Title, abstract or author-specified keywords	("fraud") AND ("financial statement" OR "financial reporting") AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence")	36
Scopus	Article title, Abstract, Keywords	("fraud*") AND ("financial statement*" OR "financial reporting") AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence")	230
Web of Science	All Fields	("fraud*") AND ("financial statement*" OR "financial reporting") AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence")	176
DergiPark	DergiPark Arama Motoru	("financial") AND ("fraud") AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence")	102
		("finansal") AND ("hile") AND ("veri madenciliği" OR "makine" OR "derin öğrenme" OR "yapay zeka")	25
TRDizin	TRDizin Arama Motoru	(abstract : ("financial" AND "fraud" AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence")))	19
		(title : ("financial" AND "fraud" AND ("data mining" OR "machine learning" OR "deep learning" OR "artificial intelligence")))	6
		(abstract : ("finansal" AND "hile" AND ("veri madenciliği" OR "makine" OR "derin öğrenme" OR "yapay zeka")))	5
		(title : ("finansal" AND "hile" AND ("veri madenciliği" OR "makine" OR "derin öğrenme" OR "yapay zeka")))	1
Toplam Ulaşılan Araştırma Sayısı:			600

Araştırma kapsamı aşağıda Tablo 3'te sunulan dâhil etme ve hariç tutma kriterlerine göre belirlenmiştir.

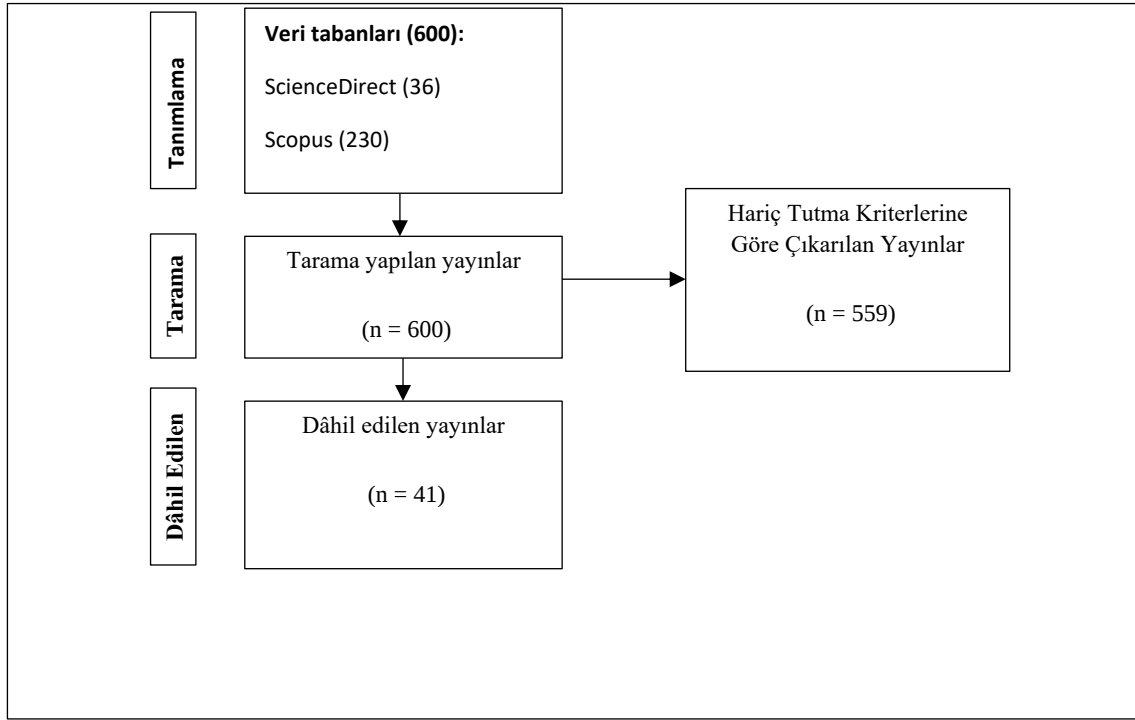
Tablo 3. Araştırma Dâhil Etme ve Hariç Tutma Kriterleri

Dâhil Etme Kriterleri	Hariç Tutma Kriterleri
Taranan veri tabanlarında yinelenen araştırmaların bir kez ele alınması	Taranan veri tabanlarında yinelenen araştırmalar
2006-2024 yılları arasında yayınlanan araştırmalar	2006-2024 yılları arasında yer almayan araştırmalar
Veri madenciliği, makine öğrenimi, derin öğrenme ve yapay zekâ yöntemlerini hileli finansal tablo tespitine uygulayan araştırmalar	Veri madenciliği, makine öğrenimi, derin öğrenme ve yapay zekâ yöntemlerini hileli finansal tablo tespitine uygulamayan araştırmalar
Doğruluk (accuracy) performans değerlendirme ölçütü bulunan araştırmalar	Doğruluk (accuracy) performans değerlendirme ölçütü bulunmayan araştırmalar
Veri tabanlarında yer alan araştırmalardan erişilebilen araştırmalar	Veri tabanlarında yer alan araştırmalardan erişilemeyen araştırmalar
Yalnızca hakemli dergilerde yer alan araştırma makaleleri	Hakemli dergilerde ve araştırma makalesi olarak sınıflandırılmayan araştırmalar
Yayın dili İngilizce ve Türkçe olan araştırmalar	Yayın dili İngilizce ve Türkçe olmayan araştırmalar

Tuğrul KANDEMİR, Zafer KARDEŞ

Araştırmada, dâhil etme kriterlerini karşılayan araştırma makalelerini belirlemek için titiz bir seçim süreci yürütülmüştür. Bu sürecin sonunda dâhil etme kriterlerini karşılayan araştırma makaleleri analize dâhil edilmiştir. Analize dâhil edilecek araştırma makalelerinin belirlenmesine yönelik sürecin ayrıntıları Şekil 1’de PRISMA 2020 akış şemasında sunulmuştur.

Şekil 1. PRISMA 2020 Akış Şeması



Araştırmada veri toplama süreci, makalenin incelenmesini ve ardından araştırma yılı, hileli / hilesiz veri seti dağılımı, yöntemlerin doğruluk oranları, kullanılan toplam yöntem sayısı ve en iyi performans gösteren yöntem gibi belirli bilgilerin çıkarılmasını içermektedir. Veriler, ilk tarama sonucunda elde edilen 600 araştırmanın başlıkları, veri tabanı, yayın türü, araştırma yılı ve yazar(lar) dikkatli bir şekilde taranması için Microsoft Excel dosyası olarak indirilmiştir. Tarama sonucunda belirlenen 41 araştırma makalesi, analize dâhil edilmiştir. Analize dâhil edilen araştırmalara ilişkin bilgiler (Ek 1: Hileli Finansal Tabloların Tespitinde Veri Madenciliği Yöntemlerini Kullanan Başlıca Araştırmaların Özeti (2006-2024))’de sunulmuştur.

3.2. Araştırmanın Bulguları

İncelenen araştırmaların veri madenciliği yöntemlerine göre dağılımı, en iyi performans gösteren yöntemlerin karşılaştırılması, kullanılan yöntem sayıları, örneklem büyüklükleri, yayınların veri tabanlarına göre dağılımı, yayın yılına göre dağılımı, ülkelere göre dağılımı ve yayınlanan dergilere göre dağılımı ile ilgili bulgular bu bölümde açıklanmıştır.

3.2.1. İncelenen Araştırmaların Veri Madenciliği Yöntemlerine Göre Dağılımı

Veri Madenciliği terimi istatistik, makine öğrenimi, veri tabanları ve görselleştirme alanlarından türetilen çok sayıda algoritma, model ve yöntemi ifade etmektedir. Bu yöntemlerin birçoğu hileli finansal tablo tespitinde uygulanmıştır. Tablo 4, en sık kullanılan 10 veri madenciliği yöntemini göstermektedir.

Tuğrul KANDEMİR, Zafer KARDEŞ

Tablo 4. En Sık Kullanılan 10 Veri Madenciliği Yöntemi

S.N.	Yöntem	Araştırmalarda Toplam Kullanılma Sayısı	Yüzde (%)
1	DT (Karar Ağacı)	31	14,76
2	SVM (Destek Vektör Makineleri)	27	12,86
3	LR (Lojistik Regresyon)	24	11,43
4	ANN (Yapay Sinir Ağı)	20	9,52
5	RF (Rastgele Orman)	19	9,05
6	K-NN (K-En Yakın Komşu)	11	5,24
7	XGBoost	9	4,29
8	NB (Naive Bayes)	7	3,33
9	AdaBoost	6	2,86
10	BBN (Bayes İnanç Ağları)	6	2,86
11	Diğer Yöntemler	50	23,81
Toplam:		211	100,00

Hileli finansal tablo tespiti için kullanılan veri madenciliği yöntemlerinin sistematik bir incelemesi Tablo 4'te sunulmuştur. İncelenen 41 adet araştırmada, toplam 46 farklı veri madenciliği yöntemi kullanılmıştır. Yöntemlerin araştırmalarda toplam kullanılma sayısının 211 olduğu tespit edilmiştir. En sık kullanılan “sınıflandırma” kategorisine giren veri madenciliği yöntemi sırasıyla karar ağacı, destek vektör makineleri, lojistik regresyon, yapay sinir ağları ve rastgele ormandır. İncelenen 41 adet araştırmada, %14,76 oranla “karar ağaçları” en çok kullanılan yöntemdir. Karar ağaçlarını, %12,86 ile destek vektör makineleri, %11,43 ile lojistik regresyon, %9,52 ile yapay sinir ağları ve %9.05 ile rastgele orman yöntemi izlemektedir.

Literatürde yer alan araştırmalar finansal hile tespitinde veri madenciliği yöntemlerinin kullanım sıklığına göre karşılaştırıldığında, yapay sinir ağlarının (Barman ve diğerleri, 2016; Ngai ve diğerleri, 2011; Sharma & Panigrahi, 2013; West & Bhattacharya, 2016) en fazla kullanılan yöntem olduğu tespit edilmiştir. Yapay sinir ağlarını lojistik regresyon (Albashrawi, 2016; Ngai ve diğerleri, 2011; Sharma & Panigrahi, 2013) ve destek vektör makineleri (Al-Hashedi & Magalingam, 2021; Barman ve diğerleri, 2016; West & Bhattacharya, 2016) yöntemleri izlemiştir. Bazı araştırmalarda (Ngai ve diğerleri, 2011; Sharma & Panigrahi, 2013) bayes ağları ve karar ağacı yöntemlerinin finansal tablo hileleri tespitinde yaygın olarak kullanıldığı tespit edilmiştir. Gupta & Mehta (2021) diğer araştırmalardan farklı olarak araştırma makalelerinin meta-analizine dayanarak, istatistiksel yaklaşımlara kıyasla makine öğrenimi yaklaşımlarının, özellikle örnek verilerin sayısı az olduğunda daha iyi sınıflandırma doğruluğu sağladığını ifade etmiştir. Geleneksel istatistiksel yöntemlere dayalı araştırmaların aksine, hileli finansal tabloları tespit etmek için literatürde veri madenciliği ve makine öğrenimi yöntemleri kullanılmaktadır. Veri madenciliği ve makine öğrenimi yöntemleri geleneksel istatistiksel yöntemlerden daha yüksek tahmin doğruluğu sunmaktadır. Ayrıca, veri madenciliği ve makine öğrenimi yöntemleri geleneksel istatistiklerin gerektirdiği varsayımlara ihtiyaç duymamakta ve doğrusal olmayan sorunları etkili bir şekilde ele alabilmektedir (Jan, 2021). Yapay zekâ çağında, hileli finansal tabloları tespit etmek için bazı çalışmalarda (Craja ve diğerleri, 2020; Jan, 2021; Li ve diğerleri, 2024; Tatar & Kıymık, 2021) derin öğrenme yöntemleri de uygulanmıştır. Derin öğrenme, sinir ağlarına dayalı olmakla birlikte, bunlardan daha fazla katmana sahiptir ve özellikleri etkili bir şekilde yakalayarak karmaşık sorunlarla başa çıkabilmektedir. Craja ve diğerleri (2020) tarafından derin öğrenme, makine öğrenimiyle karşılaştırıldığında, hileli finansal tabloları daha etkili bir şekilde tespit edebilme yeteneğine sahip olduğu ifade edilmiştir.

3.2.2. En İyi Performans Gösteren Yöntemlerin Karşılaştırılması

Hileli finansal tabloların gerçek bir endişe kaynağı olduğu iş dünyasında, performansın ölçülmesinde çeşitli faktörler ele alınmaktadır. Ancak performans ölçümünde genellikle doğruluk değerlendirme ölçütü dikkate alınmaktadır. Doğruluk, başarılı sınıflandırma sayısının toplam sınıflandırma sayısına oranı olarak hesaplanır. Bu performans ölçümlerine ek olarak, birçok araştırmacı tarafından başka bazı ölçümler de (kesinlik, hassasiyet, F-ölçüsü, tip I hata, tip-II hata, AUC, ROC, kappa gibi) kullanılmıştır. Tablo 3'te ifade edildiği gibi doğruluk ölçüsü belirtilmeyen araştırmalar, yapılacak karşılaştırmalı değerlendirme sonuçlarının daha sağlıklı olabilmesi açısından hariç tutulmuştur. Araştırmalar bazında, doğruluk ölçüsüne göre en iyi performans gösteren veri madenciliği yöntemleri Tablo 5'te sunulmuştur.

Tuğrul KANDEMİR, Zafer KARDEŞ

Tablo 5. Doğruluk Ölçüsüne Göre En İyi Performans Gösteren Yöntemler

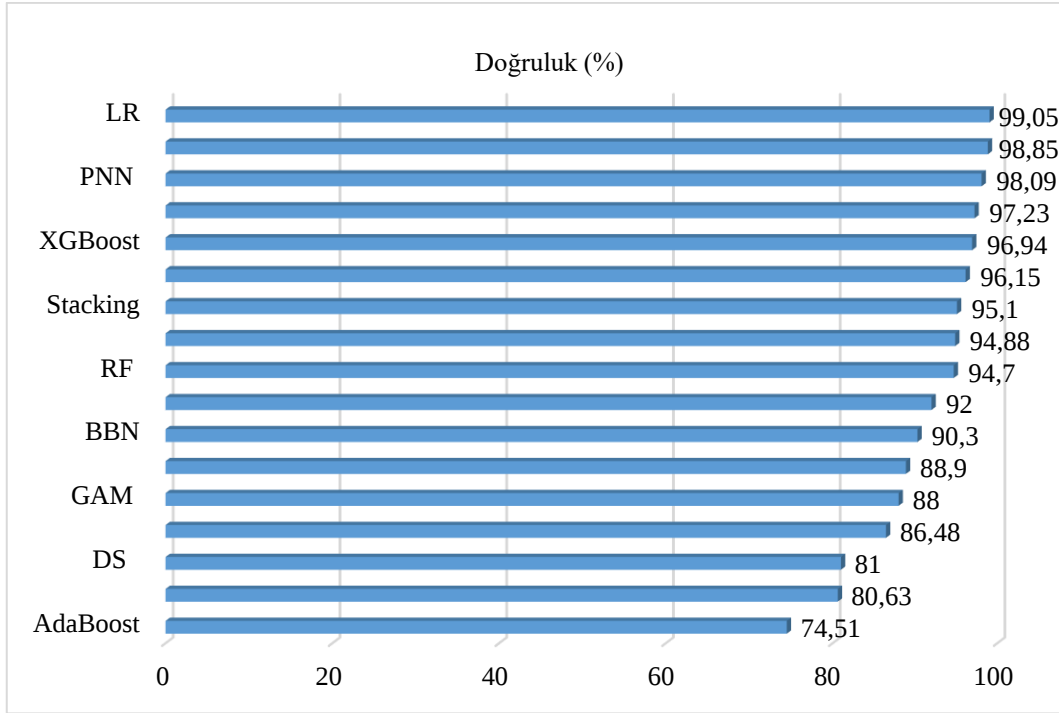
En İyi Performans Gösteren Yöntem	Miktar	Yazar(lar), Yıl (Doğruluk (%) Oranı)
RF (Rastgele Orman)	9	An & Suh, 2020 (79,90); Craja ve diğerleri, 2020 (86,53); Hamal & Senvar, 2021 (93,74); Kırdı & Özçelik, 2021 (89,83); Papík & Papíková, 2021 (83,75); Papík & Papíková, 2022 (85,50); Phong ve diğerleri, 2024 (72,3); Wyrobek, 2020 (94,70); Xiuguo & Shengyong, 2022 (87,74)
ANN (Yapay Sinir Ağı)	8	Ata & Seyrek, 2009 (77,36); Aksoy, 2021 (96,15), Kılıç & Önal, 2022 (88,89); Lin ve diğerleri, 2015 (92,8); Nhien ve diğerleri, 2024 (97,23); Terzi & Şen, 2012 (91,0); Uğurlu & Sevim, 2015a (90); Wang & Chen, 2024 (67,73)
DT (Karar Ağacı)	5	Aksoy, 2021 (96,15); Chen ve diğerleri, 2014 (85,71); Chen, 2016 (87,97); Gupta & Gill, 2012 (94,74); Tatar & Kıymık (84,25)
XGBoost	4	Ali ve diğerleri, 2023 (93,66); Li ve diğerleri, 2024 (96,94); Sun ve diğerleri, 2023 (68,01); Sun ve diğerleri, 2024 (86,3)
DL (Derin Öğrenme)	3	Li ve diğerleri, 2024 (90,85); Jan, 2021 (94,88); Tatar & Kıymık (84,25)
BBN (Bayes İnanç Ağları)	2	Hajek & Henriques, 2017 (90,09); Kirkos ve diğerleri, 2007 (90,3)
Hybrid	2	Javadian Kootanaee ve diğerleri, 2021 (80,0); Singh Yadav & Sora, 2022 (98,85)
AdaBoost	1	Zhu ve diğerleri, 2024 (74,51)
CatBoost	1	Lokanan & Sharma, 2024 (92)
DS (Karar Kütüğü)	1	Sorkun & Toraman, 2017 (81,0)
Ensemble	1	Song ve diğerleri, 2014 (88,9)
ERT (Aşırı Rastgele Ağaçlar)	1	Riskiyadi, 2024 (86,48)
GAM (Genelleştirilmiş Katkı Modeli)	1	Tragouda ve diğerleri, 2024 (88,0)
LR (Lojistik Regresyon)	1	Liou, 2008 (99,05)
PNN (Olasılık Sinir Ağı)	1	Ravisankar ve diğerleri, 2011 (98,09)
Stacking	1	Kotsiantis ve diğerleri, 2006 (95,1)
SVM (Destek Vektör Makineleri)	1	Yao ve diğerleri, 2019 (80,63)

Tablo 5 incelendiğinde, araştırmalarda en iyi performans gösteren yöntemin rastgele orman yöntemi olduğu görülmektedir. İncelenen 41 adet araştırmanın dokuzunda en iyi performansı rasgele orman yöntemi göstermiştir. Rastgele orman yöntemini sırasıyla, sekiz adetle yapay sinir ağları, beş adetle karar ağacı ve dört adetle XGBoost algoritması izlemiştir.

Aksoy (2021) araştırmasında hem yapay sinir ağları hem de karar ağacı yöntemleri aynı oranda (%96,15) en iyi performansı göstermiştir. Tatar ve Kıymık araştırmalarında hem karar ağacı hem de derin öğrenme yöntemleri aynı oranda (%84,25) en iyi performansı göstermiştir. Dolayısıyla iki araştırmada iki farklı yöntem aynı oranda en iyi performansı göstermiştir. Uygulanan yöntemler bazında en yüksek doğruluk oranlarının karşılaştırılması aşağıda Şekil 2’de sunulmuştur.

Tuğrul KANDEMİR, Zafer KARDEŞ

Şekil 2. En İyi Performans Gösteren Yöntemlerin Doğruluk Oranlarına Göre Karşılaştırılması



Şekil 2’de araştırmalar bazında yapılan değerlendirmede, en yüksek doğruluk oranına sahip 17 farklı veri madenciliği yöntemi olduğu görülmektedir. En yüksek performansı gösteren yöntem, %99,05 ile lojistik regresyon yöntemi olmuştur. Lojistik regresyonu sırasıyla %98,85 ile hibrit model, %98,09 ile olasılık sinir ağı ve %97,23 ile yapay sinir ağı yöntemi takip etmiştir. Doğruluk performans ölçütlerinin %74,51 ile %99,05 aralığında olduğu görülmektedir. Dolayısıyla ilgili araştırmalarda doğru sınıflandırma başarılarının yüksek olduğu tespit edilmiştir.

Hileli finansal tablo tespitinde öne çıkan farklı veri madenciliği yöntemleri, Tablo 6’da avantajları ve dezavantajları ile karşılaştırılmıştır.

Tablo 6. Hileli Finansal Tablo Tespiti İçin Farklı Veri Madenciliği Yöntemlerinin Karşılaştırılması

Veri Madenciliği Yöntemi	Avantajları	Dezavantajları
Rastgele Orman	Çalışma süreci hızlıdır ve dengesiz verilerle kullanılabilir.	Eğitim verisi aralığının dışında yeni örüntüleri tahmin etmek için verimsizdir.
Yapay Sinir Ağı	Algoritmik olmayan ve ikili sınıflandırma problemleri için uygundur.	Eğitim için yüksek hesaplama gücü gereklidir.
Karar Ağacı	Uygulanması basit ve kolaydır, eğitim için çok düşük hesaplama gücü gerektirir.	İlk kurulum yüksek hesaplama gücü gerektirir ve problem alanının yüksek temsiline sahip bir eğitim seti gerektirir.
Bayes İnanç Ağları	Algoritmik olmayan ve ikili sınıflandırma problemleri için uygundur.	Dolandırıcılık türüyle ilgili tipik ve anormal davranışlar hakkında eksiksiz bilgi gereklidir.
Lojistik Regresyon	Uygulaması kolaydır ve çoğunlukla kredi kartı ve finansal tablo hile tespiti için kullanılır.	Daha yüksek hesaplama karmaşıklığı ancak daha düşük performans göstermektedir.
Destek Vektör Makineleri	Doğrusal olmayan sınıflandırma problemini çözebilir ve eğitim için düşük hesaplama gücü gerektirir.	Dönüştürülmüş girdi seti nedeniyle sonuçları işlemek zordur.

Son zamanlarda bazı çalışmalar, daha iyi hile tespiti doğruluğu elde etmek ve örnekleri sınıflandırmak için birden fazla algoritmanın güçlü yönlerini birleştiren hibrit yöntemler kullanmıştır (Javadian Kootanaee ve diğerleri, 2021; Singh Yadav & Sora, 2022). Javadian Kootanaee ve diğerleri (2021), destek vektör makinesi ile geliştirilmiş bir ID3 karar ağacı hibrit model olarak sunulmuştur. Singh Yadav & Sora (2022), hibrit evrişimli sinir ağını kullanarak finansal hile beyanlarını tespit etmek için denetimsiz bir öğrenme modeli sunmuştur. Sunulan hibrit model, finansal tablolardan veri madenciliği tabanlı metin özelliği çıkarma ile büyük miktarda veriyi hızlı ve etkili bir şekilde işleyebilmektedir. Finansal tablo hile tespitinde karşılaşılan yaygın sorunlardan bazıları aşırı uyum, sınıflandırma doğruluğunun düşüklüğü, hesaplamaların karmaşık süreci, verinin kıt olması, değişen ve karmaşık kalıpların analiz edilmesinin zor olmasıdır. Ayrıca; sınıf dengesizliğinin yüksekliği en büyük zorluklardan birisi olup, veri analizlerinde sorunlar ortaya çıkarabilmektedir. Bu sorunların üstesinden gelebilmek için hibrit modeller geliştirilmektedir. Hibrit modeller, büyük hacimli verilerin içerisinde gizli olan ayrıntıları ortaya çıkarabilir ve finansal araştırmalardaki karmaşık sorunları çözebilir (Singh Yadav & Sora, 2022).

3.2.3. İncelenen Araştırmalarda Kullanılan Yöntem Sayısı

Bazı araştırmalarda sadece bir adet veri madenciliği yöntemi kullanırken bazı araştırmalarda birden fazla veri madenciliği yöntemi kullanılmıştır. Aşağıdaki Tablo 7’de incelenen araştırmalarda kullanılan yöntem sayıları sunulmuştur.

Tablo 7. İncelenen Araştırmalarda Yöntem Sayısı

Toplam Kullanılan Yöntem Sayısı	Araştırma Sayısı	Yüzde (%)
1	3	7,32
2	5	12,20
3	5	12,20
4	5	12,20
5	6	14,63
6	5	12,20
7	6	14,63
8	2	4,88
9	1	2,44
10	1	2,44
12	1	2,44
14	1	2,44
Toplam:	26	100,00

Tablo 7, incelenen araştırmalarda toplam kullanılan veri madenciliği yöntem sayılarını vermektedir. Araştırmalarda 1 ile 14 arasında farklı sayılarda yöntem kullanılmıştır. İncelenen araştırmalarda sadece bir yöntem kullanan üç araştırma, iki yöntem kullanan beş araştırma, üç yöntem kullanan beş araştırma, dört yöntem kullanan beş araştırma, beş yöntem kullanan altı araştırma, altı yöntem kullanan beş araştırma, yedi yöntem kullanan altı araştırma, sekiz yöntem kullanan iki araştırma vardır. 9, 10, 12 ve 14 farklı yöntem kullanan birer araştırma bulunmaktadır. Araştırmalarda en sık iki ile yedi arasında farklı yöntem kullanıldığı görülmektedir.

3.2.4. İncelenen Araştırmaların Örneklem Büyüklüğü

İncelenen araştırmalardaki veri setleri, hileli/hilesiz olmak üzere iki kategoride toplanmıştır. Araştırmaların örneklem büyüklükleri aşağıda Tablo 8’de sunulmuştur.

Tuğrul KANDEMİR, Zafer KARDEŞ
Tablo 8. Örneklem Büyüklükleri

Örneklem Büyüklüğü	Araştırma Sayısı	Yüzde (%)
>1.000	18	43,90
>500	5	12,20
>200	1	2,44
<=200	17	41,46
Toplam:	41	100,00

Tablo 8 incelendiğinde, ele alınan araştırmaların yarısından fazlasının (%56,10) 1.000’den az örnekleme sahip olduğu görülmektedir. Bunun en önemli sebebi, hileli finansal tablo verilerine ulaşmadaki kısıtlılıklardır. Örneklem büyüklüğü 200’den az 17 araştırma, 200-500 arası bir araştırma, 500-1000 arası beş araştırma ve 1.000 üzeri 18 araştırma olduğu görülmektedir. Örneklem büyüklüğü 1.000 üzeri olan araştırmaların özellikle son beş yılda yapıldığı tespit edilmiştir. Araştırma veri setlerine ilişkin hileli ve hilesiz dağılımı (Ek 1: Hileli Finansal Tabloların Tespitinde Veri Madenciliği Yöntemlerini Kullanan Başlıca Araştırmaların Özeti (2006-2024))’de sunulmuştur.

3.2.5. İncelenen Yayınların Veri Tabanlarına Göre Dağılımı

İnceleme kapsamına giren araştırmalar belirlenirken, beş farklı veri tabanı kullanılmıştır. Literatürde yer alan bazı araştırmalarda (Al-Hashedi & Magalingam, 2021; Ashtiani & Raahemi, 2021; Ngai ve diğerleri, 2011; Ramzan & Lokanan, 2024; Shahana ve diğerleri, 2023) kullanılan veri tabanları belirtilmiştir. Bu doğrultuda veri tabanı kapsamı da dikkate alınarak ScienceDirect, Scopus ve Web of Science veri tabanları belirlenmiştir. Bu araştırmada, mevcut literatürdeki araştırmalardan farklı olarak Türkçe araştırmaları da kapsamından dolayı DergiPark ve TRDizin veri tabanları da kullanılmıştır. Veri tabanlarında belirlenen arama dizelerine göre ilk taramada ulaşılan araştırma sayıları aşağıda Tablo 9’da sunulmuştur.

Tablo 9. Veri Tabanlarına Göre Yayın Sayısı

Veri Tabanı	Yayın Sayısı	Yüzde (%)
ScienceDirect	36	6,00
Scopus	230	38,33
Web of Science	176	29,33
DergiPark	127	21,17
TRDizin	31	5,17
Toplam:	600	100,00

Tablo 9 incelendiğinde, araştırmaların %38,33’üne Scopus, %29,33’üne Web of Science ve %21,17’sine Dergipark veri tabanı aracılığıyla ulaşıldığı görülmektedir. Erişim sağlanan araştırmaların %88,83’üne bu üç veri tabanı aracılığıyla ulaşılmıştır. Diğer veri tabanlarına bakıldığında, araştırmaların %6’sına ScienceDirect, %5,17’sine TRDizin veri tabanından ulaşıldığı görülmektedir.

3.2.6. İncelenen Araştırmaların Yayın Yılına Göre Dağılımı

Yayın yılı 2006’dan 2024’e kadar olan araştırmaların yayın yıllarına göre dağılımı aşağıda Tablo 10’da sunulmuştur.

Tablo 10. İncelenen Araştırmaların Yayın Yılına Göre Dağılımı

Yıl	Yayın Sayısı	Yüzde (%)
2006	1	2,44
2007	1	2,44
2008	1	2,44
2009	1	2,44
2010		
2011	1	2,44
2012	2	4,88
2013		
2014	2	4,88

Tuğrul KANDEMİR, Zafer KARDEŞ

2015	2	4,88
2016	1	2,44
2017	2	4,88
2018		
2019	1	2,44
2020	3	7,32
2021	4	9,76
2022	7	17,07
2023	2	4,88
2024	10	24,39
Toplam:	41	100,00

Tablo 10, incelenen araştırma makalelerinin yayın yılına göre dağılımını göstermektedir. 2006'dan 2024'e kadar incelenen 41 adet araştırmanın yıllara göre dağılımı verilmiştir. 2010, 2013 ve 2018 yıllarında herhangi bir araştırmaya rastlanılmamıştır. Araştırma yapılan yıllara bakıldığında yayın sayıları 1 ile 10 arasında dağılım göstermiştir. 10 adet araştırmayla en çok 2024 yılında araştırma yapılmıştır. Dolayısıyla literatürde hileli finansal tablo tespitinde veri madenciliği yöntemlerinin uygulandığı araştırmalarda son yıllarda artış olduğu tespit edilmiştir.

3.2.7. İncelenen Araştırmaların Ülkelere Göre Dağılımı

İncelenen hileli finansal tablo tespitinde veri madenciliği uygulamaları araştırmalarının ülkelere göre dağılımı Tablo 11'de sunulmuştur.

Tablo 11. İncelenen Araştırmaların Ülkelere Göre Dağılımı

Ülke	Yayın Sayısı	Yüzde (%)
ABD	10	24,39
Türkiye	9	21,95
Çin	8	19,51
Tayvan	5	12,20
Yunanistan	3	7,32
Vietnam	2	4,88
Afrika	1	2,44
Endonezya	1	2,44
Güney Kore	1	2,44
İran	1	2,44
Toplam:	41	100,00

Ele alınan 41 araştırmanın ülkelere göre dağılımı incelendiğinde en çok araştırmanın ABD'de yapıldığı görülmektedir. Türkiye dokuz adet araştırmayla ikinci sırada yer almaktadır. Tablo 3'te ifade edildiği gibi ele alınan araştırmaların yayın dili İngilizce ve Türkçedir. Türkiye örneğinde yapılan dokuz araştırmanın yayın diline bakıldığında dört araştırmanın İngilizce, beş araştırmanın Türkçe olarak yayınlandığı görülmektedir. Diğer ülkelerde yapılan araştırmaların yayın dili sadece İngilizce olmasından dolayı Türkiye örneğinde yapılan araştırma sayısı ikinci sırada yer almıştır. Türkiye'yi sekiz adet araştırmayla Çin izlemiştir. Araştırma kapsamında ele alınan araştırmaların %65,85'i ilk üç ülkede yapıldığı tespit edilmiştir. Ayrıca (Ek 1: Hileli Finansal Tabloların Tespitinde Veri Madenciliği Yöntemlerini Kullanan Başlıca Araştırmaların Özeti (2006-2024))'nde yer aldığı gibi Çin örneğinde yapılan araştırmaların yarısı 2024 yılında yapılmıştır. Dolayısıyla son yıllarda Çin örneğinde yapılan araştırmalardaki artış dikkat çekmektedir.

3.2.8. İncelenen Araştırmaların Dergilere Göre Dağılımı

İncelenen hileli finansal tablo tespitinde veri madenciliği uygulamaları araştırmalarının dergilere göre dağılımı Tablo 12'de sunulmuştur. Tablo 12 incelendiğinde, araştırmaların daha çok bilgisayar bilimleri alanında yer alan dergilerde yayınlandığı görülmektedir. Veri madenciliği, makine öğrenimi, derin öğrenme ve yapay zekâ kapsamında yer alan dergilerin daha çok tercih edildiği, muhasebe ve denetim kapsamında yer alan dergilerin daha az tercih edildiği ifade edilebilir.

Tuğrul KANDEMİR, Zafer KARDEŞ

Tablo 12. İncelenen Araştırmaların Dergilere Göre Dağılımı

Dergi Adı	Yayın Sayısı	Yüzde (%)
Decision Support Systems	2	4,88
IEEE Access	2	4,88
International Journal of Computational Intelligence	2	4,88
Knowledge-Based Systems	2	4,88
Muhasebe ve Vergi Uygulamaları Dergisi	2	4,88
Procedia Computer Science	2	4,88
Sustainability	2	4,88
Applied Sciences	1	2,44
Asian Review of Accounting	1	2,44
Calitatea	1	2,44
Concurrency and Computation: Practice and Experience	1	2,44
Data Technologies and Applications	1	2,44
Economics & Politics	1	2,44
Equilibrium, Quarterly Journal of Economics and Economic Policy	1	2,44
Expert Systems with Applications	1	2,44
Intelligent Systems and Applications in Engineering	1	2,44
Intelligent Systems in Accounting, Finance and Management	1	2,44
International Journal of Accounting Information Systems	1	2,44
International Journal of Advanced Computer Science and Applications	1	2,44
İşletme Araştırmaları Dergisi	1	2,44
Journal of Combinatorial Mathematics and Combinatorial Computing	1	2,44
Journal of Forecasting	1	2,44
Journal of International Commerce, Economics and Policy	1	2,44
Journal of Network Intelligence	1	2,44
Journal of Optimization in Industrial Engineering	1	2,44
Maliye ve Finans Yazıları	1	2,44
Managerial Auditing Journal	1	2,44
Muhasebe ve Denetime Bakış Dergisi	1	2,44
SpringerPlus	1	2,44
Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi	1	2,44
Systems	1	2,44
The British Accounting Review	1	2,44
The Scientific World Journal	1	2,44
Yaşar Üniversitesi E-Dergisi	1	2,44
Toplam:	41	100,00

Tablo 12, analizde yer alan 41 araştırmanın dergilere göre dağılımını listelemektedir. Liste, araştırmanın yayınlandığı dergi adını, yayın sayısını ve yüzdesini içermektedir. 41 adet araştırma, 34 farklı dergide yayınlanmıştır. Yayın sayılarına bakıldığında, yedi adet dergide ikişer adet, diğer 27 dergide ise birer adet yayınlandığı görülmektedir.

4. TARTIŞMA

Bu çalışmada, 2006'dan 2024'e kadar hileli finansal tablo tespitinde veri madenciliği uygulamalarını ele alan bilimsel dergilerde yayınlanan 41 adet araştırma makalesi incelenmiştir. Söz konusu makaleler kullanılan veri madenciliği yöntemleri, en iyi performans gösteren yöntemler, kullanılan yöntem sayısı, veri seti örneklem büyüklüğü, yayınların veri tabanlarına göre dağılımı, yayın yılı, yayınlandığı ülke ve yayınlandığı dergi kriterleri bakımından incelenmiştir. İncelenen 41 adet araştırma sonucunda tespit edilen durumlar aşağıdaki gibidir:

- İncelenen 41 adet çalışmada 46 farklı veri madenciliği yöntemi kullanılmıştır. Yöntemlerin toplam kullanım sayısı 211 adettir. En sık kullanılan 10 veri madenciliği yönteminin kullanım sıklıkları ve kullanım yüzdeleri

Tuğrul KANDEMİR, Zafer KARDEŞ

Tablo 4'te gösterilmiştir. Hileli finansal tablo tespiti için karar ağacı, destek vektör makineleri, lojistik regresyon, yapay sinir ağları ve rastgele orman yöntemlerinin esnek hesaplama avantajı ve yüksek doğruluk performanslı modeller sunması açısından daha çok tercih edildiği (>%50) tespit edilmiştir.

- Karar ağacı yönteminin %14,76 ile veri madenciliğinde en yaygın kullanılan hileli finansal tablo tespit yöntemi olduğu ve bunu %12,86 ile destek vektör makineleri, %11,43 ile lojistik regresyon, %9,52 ile yapay sinir ağları ve %9,05 ile rastgele orman yöntemi izlemiştir. Bu sonuç karar ağacı yönteminin, hileli finansal tablo tespitinde kullanılan önde gelen veri madenciliği yöntemi olduğunu göstermektedir. Literatürde (Ngai ve diğerleri, 2011; Sharma & Panigrahi, 2013) ilgili yöntemlerin finansal tablo hileleri tespitinde yaygın olarak kullanıldığı görülmekte olup, bu durum çalışma bulguları ile örtüşmektedir.
- Veri madenciliği uygulama sınıfı, kümeleme, sınıflandırma, aykırı değer tespiti, tahmin, regresyon, görselleştirme olmak üzere altı kategoriye ayrılmaktadır (Ngai ve diğerleri, 2011). Literatürde yer alan araştırmalar, finansal tablo hilelerinin tipik bir sınıflandırma sorunu olduğunu göstermektedir.
- İncelenen 41 araştırmanın dokuzunda, kullanılan veri madenciliği yöntemleri arasında en fazla performans gösteren rastgele orman yöntemi olmuştur. Bunu sırasıyla, sekiz çalışmada yapay sinir ağları, beş çalışmada karar ağacı yöntemi izlemiştir. Dolayısıyla hileli finansal tablo tespitinde rastgele orman yönteminin, kullanılan yöntemler arasında en fazla performans gösteren yöntem olduğu sonucuna ulaşılmıştır.
- İncelenen araştırmalar bazında değerlendirildiğinde, doğruluk ölçüsüne göre en yüksek performansı gösteren yöntem %99,05 ile lojistik regresyon olmuştur. Bunu sırasıyla %98,85 ile hibrit model, %98,09 ile olasılık sinir ağı, %97,23 ile yapay sinir ağı izlemiştir. Doğruluk performans ölçütlerinin %74,51 ile %99,05 aralığında olmasından dolayı, doğru sınıflandırma başarılarının yüksek olduğu tespit edilmiştir.
- Yapılan araştırmalar, veri madenciliği yöntemlerinin finansal raporlarda hileleri tespit etmede başarılı olduğunu göstermektedir. Terzi (2012), veri madenciliği yöntemlerinin kullanılmasının işletmelere hile ve hataların önlenmesinde büyük avantajlar sağlayacağını ifade etmiştir.
- Araştırmalarda kullanılan veri madenciliği yöntem sayısı 1 ile 14 arasında değişmektedir. Araştırmalarda en sık iki ile yedi arasında farklı yöntem kullanıldığı tespit edilmiştir. Aynı veri seti üzerinde birden fazla yöntemin kullanılması yöntem performanslarını karşılaştırma bakımından fayda sağlamaktadır. Bu açıdan birden fazla yöntemin kullanılmasıyla en yüksek performans gösteren yöntemin tercih edilmesi mümkün olmaktadır.
- Veri seti büyüklüklerine bakıldığında araştırmaların yarısından fazlası (%56,1) 1.000'den az örnekleme sahip olduğu görülmektedir. Buradaki temel sorunun, araştırma yapmak için temel oluşturacak kamuya açık verilerin eksikliğinden kaynaklandığı gözlenmiştir.
- Literatürdeki araştırma sonuçları çalışmalarda kullanılan yöntem performansının veri setine, veri madenciliği yöntemine ve seçilen değişkenlere büyük ölçüde bağlı olduğunu göstermektedir. İncelenen araştırmaların tamamında finansal değişkenler kullanılmıştır. Kullanılan finansal değişkenlerin sayısına bakıldığında üç (Phong ve diğerleri, 2024) ile 99 (Zhu ve diğerleri, 2024) arasında değiştiği görülmüştür. Finansal olmayan değişkenler ise sadece dokuz çalışmada kullanılmış olup diğer araştırmalarda kullanılmamıştır. İlgili çalışmalara ait detay bilgileri (Ek 1: Hileli Finansal Tabloların Tespitinde Veri Madenciliği Yöntemlerini Kullanan Başlıca Araştırmaların Özeti (2006-2024))'nde sunulmuştur. Ayrıca ele alınan literatür kapsamında Türkiye örnekleminde finansal olmayan değişken kullanan araştırmaya rastlanılmamıştır.
- Araştırma ScienceDirect, Scopus, Web of Science, DergiPark ve TRDizin olmak üzere beş veri tabanı üzerinde yapılmıştır. İlk tarama sonucunda 600 adet araştırmaya erişilmiş olup Tablo 3'te ifade edilen hariç tutma kriterlerinden sonra 41 adet araştırma belirlenmiştir. Araştırmaların %88,83'üne Web of Science, Scopus ve DergiPark veri tabanı aracılığıyla ulaşılmıştır.
- Tespit edilen araştırmaların yayın yılı 2006'dan 2024'e kadar olan zaman aralığını kapsamaktadır. 19 yıllık periyoda bakıldığında, 2006-2014 arasında dokuz araştırma, 2015-2019 arasında altı araştırma, 2020-2024 arasında 26 araştırma olduğu tespit edilmiştir. Son yıllarda, hileli finansal tablo tespitinde veri madenciliği yöntemlerinin giderek daha fazla uygulandığı gözlenmiştir.

Tuğrul KANDEMİR, Zafer KARDEŞ

- Tespit edilen araştırmaların ülkelere göre dağılımına bakıldığında, en çok ABD’de araştırma yapıldığı görülmektedir. ABD’yi yayın dili hem Türkçe hem de İngilizce olan araştırmaları kapsamasından dolayı dokuz adet araştırmayla Türkiye izlemiştir. Çin’de sekiz adet, Tayvan’da beş, Yunanistan’da üç, Vietnam’da iki araştırma yapılırken, Afrika, Endonezya, Güney Kore ve İran ülkelerinde birer araştırma yapıldığı gözlenmiştir.
- Analize dâhil edilen 41 araştırma, 34 farklı dergide yayınlanmıştır. Yayın sayılarına bakıldığında, yedi dergide ikişer araştırma ve 27 dergide birer araştırma yayınlandığı gözlenmiştir.

Yapılan araştırma sonucuna göre, hileli finansal raporlama tespiti için veri madenciliği uygulamalarının büyük bir potansiyele sahip olduğu gözlenmiştir.

5. SONUÇ VE ÖNERİLER

Hileli finansal tabloların tespit edilmesinde veri madenciliği yöntemlerinin uygulanması işletmenin bilgi kullanıcıları olarak ifade edilen yöneticiler, denetçiler, vergi otoriteleri, finansal analistler ve diğer ilgili taraflar için önemli bulgular sunmaktadır. Örneğin denetçiler bu bulguları kullanarak daha etkili kanıtlar toplayabilir ve denetim planını geliştirebilirler. Ek olarak, denetçiler bu sonuçları kullandıkları yazılıma dâhil ederek işletmelerin finansal tablolarını analiz edebilir ve “kırmızı bayrak” olarak bilinen risk faktörlerini tespit edebilirler (Terzi & Şen, 2012: 38). Veri madenciliği uygulamaları çeşitli iş, denetim ve muhasebe süreçlerinin ayrılmaz bir parçası haline gelmekte ve şirketler veri madenciliği uygulamalarının geliştirilmesine giderek daha fazla sermaye ayırmaktadır. Hileli finansal tablo tespitinde veri madenciliğinin kullanılmasıyla birlikte denetçilerin, kredi derecelendirme uzmanlarının ve yatırımcıların işleri kolaylaştırılabilir ve karar verme süreçlerinde zaman ve maliyet avantajı sunabilir.

İlerleyen yıllarda bu araştırma alanlarına ilişkin yeni sorunlar ve tehditlerin ortaya çıkması muhtemeldir, ancak aynı zamanda daha fazla yenilik, olası fırsatlar ve etkili çözümler getireceği de açıkça görülmektedir. Daha fazla araştırma çabalarıyla birlikte veri madenciliğinin muhasebe ve denetim alanında daha da değerli bir araç haline geleceği açıktır. Bu araştırma sonuçlarının muhasebeciler, denetçiler, finansal analistler, yöneticiler ve gelecekte yapılacak akademik araştırmalar için önemli bir referans kaynağı olacağı öngörülmektedir.

Gelecekteki araştırmalar için öneriler aşağıda sunulmaktadır:

- Hileli finansal tablo tespiti üzerine yapılan önceki araştırmalar, belirli veri madenciliği yöntemlerine odaklanmıştır. Ancak gelecekteki araştırmalar farklı yöntemleri birleştiren hibrit modeller, derin öğrenme ve metin madenciliği alanlarına yönelerek aralarında karşılaştırmalar gerçekleştirebilir.
- Uygulayıcılar, kullanım sıklığı ve performansı dikkate alarak en uygun veri madenciliği yöntemini seçebilirler. Böylece hileli finansal tablo tespitinde daha yüksek bir doğruluk düzeyi elde edilmesi sağlanabilir.
- Araştırmaların %78,05’inde iki ile yedi arasında farklı veri madenciliği yöntemi kullanılmıştır. Gelecekte araştırmalarda, en az yedi farklı veri madenciliği yöntemi kullanarak karşılaştırma yapılması önerilmektedir.
- Araştırmalarda kullanılan değişkenlerin standart olmadığı farklı değişkenler kullanıldığı görülmektedir. Uğurlu & Sevim (2015b) araştırmalarında hileli finansal tablo tespiti için genel kabul görmüş değişken setinin olmadığı sonucuna ulaşmış olup, bu araştırma sonuçları ile paralellik göstermektedir. Gelecekteki araştırmalar için özellikle denetçi, iç denetim, yönetim, metinsel ve dilsel gibi finansal olmayan değişkenlerin kullanılması önerilmektedir.
- Veri madenciliği uygulama sınıfı olarak araştırmalarda “sınıflandırma” analizi üzerinde yoğunlaşma olmuştur. “Sınıflandırma” analizleri dışında “kümeleme” yöntemleri de analizlerde kullanılabilir. İncelenen araştırmalar içerisinde sadece bir araştırmada (Singh Yadav & Sora, 2022) kümeleme yöntemi kullanıldığı, diğer araştırmalarda sınıflandırma kategorisi kullanıldığı tespit edilmiştir. Türkiye örneğinde yer alan araştırmalarda, kümeleme kategorisinde yer alan araştırmaya rastlanılmamıştır. Özellikle Türkiye örneğinde yapılacak araştırmalarda sınıflandırma dışındaki kategorilerin ele alınması önerilmektedir.
- Araştırmaların %56,10’u 1.000’den az örnekleme sahiptir. Araştırmalarda veri setinin düşük olmasının en önemli sebebi hileli kategorisine giren şirket sayısının hilesiz kategorisine giren şirket sayısından çok daha düşük olmasıdır. Dolayısıyla, veri setinin az olması, sonuçların yanlışlığı problemine neden olabilir. Bu nedenle örneklem sayılarının dengeli dağılımı gözetilerek arttırılmasına yönelik, Hamal & Senvar (2021)’in

Hileli Finansal Tabloların Tespitinde Veri Madenciliği Uygulamaları: Mevcut Araştırma Eğilimlerinin İncelenmesi (2006-2024)

Tuğrul KANDEMİR, Zafer KARDEŞ

araştırmalarında da ifade edildiği gibi, Sentetik Azınlık Örnekleme Arttırma Yöntemi (Synthetic Minority Over-sampling Technique (SMOTE)) tekniği önerilmektedir.

- Araştırmalarda en çok ABD örnekleme kullanılmaktadır. Gelecekteki araştırmalar diğer ülkelerden elde edilecek veri setlerini de kullanıp, aynı veri madenciliği yöntemlerini uygulayarak sonuçları karşılaştırabilir.

Bu araştırmanın iki önemli kısıtlılığı vardır. İlk olarak araştırmada, 2006 ile 2024 yılları arasında yayınlanan makaleler için birden çok anahtar kelime ile çevrimiçi veri tabanları taranmıştır. İleride yapılacak araştırmalarda, konferans bildirileri, ders kitapları, tezleri kapsayacak şekilde ve yıl aralığını artırarak araştırma genişletilebilir. İkinci olarak, yalnızca hileli finansal tabloların tespitinde veri madenciliği yöntemlerini uygulayan araştırma makaleleri dikkate alınmıştır. Gelecekteki araştırmalar, bankacılık, sigortacılık, kripto para vb. diğer alanlarda hileleri tespit etmek için ilgili veri madenciliği uygulamalarını içerecek şekilde genişletilebilir.

Bu araştırmada, hileli finansal tablo tespitinde veri madenciliği yöntemlerini uygulayan araştırmalar sistematik olarak incelenmiştir. Araştırma kapsamında akademik makalelerin bir incelemesi yapılmış ve veri madenciliği yöntemlerinin hileli finansal tablo tespitinde uygulanmasına yönelik araştırmaların bir bibliyografyası sunulmuştur. Bu araştırmayla hileli finansal tablo tespiti ile ilgili gelişen literatürün sistematik bir incelemesi ortaya konularak, hileli finansal tablo tespitinde veri madenciliği yöntemlerinin uygulanması hakkında hem akademisyenlerin hem de uygulayıcıların bilgilendirilmesi amaçlanmıştır. Bu nedenle; bu alandaki literatürü zenginleştirerek, gelecekte yapılacak araştırmalara bir temel oluşturacağı düşünülmektedir. Araştırmanın, hileli finansal tablo tespiti alanında çalışanlar için değerli içgörüler sağlayacağı ve alana daha fazla ilgi uyandırmaya yardımcı olacağı umulmaktadır.

Ekler

Ek 1: Hileli Finansal Tabloların Tespitinde Veri Madenciliği Yöntemlerini Kullanan Başlıca Araştırmaların Özeti (2006-2024)

S.N.	Yazar(lar)	Yıl	Veri Seti (Hileli/Hilesiz)	Ülke	Değişken Özellikleri*	Kullanılan Yöntemler (Doğruluk %)**	Toplam Yöntem Sayısı	En İyi Performans Gösteren Yöntem
1	Kotsiantis ve diğerleri	2006	41/123 (164)	Yunanistan	F (25)	DT (91,2), ANN (73,4), BBN (74,1), K-NN (79,7), Ripper (86,8), LR (75,3), SVM (78,6), Stacking (95,1), Voting (92,1), BestCV (91,2), Grading (93,3), Stacking (93,9)	12	Stacking (95,1)
2	Kirkos ve diğerleri	2007	38/38 (76)	Yunanistan	F (27)	DT (73,6), ANN (80,0), BBN (90,3)	3	BBN (90,3)
3	Liou	2008	1.515/1.515 (3.030)	Tayvan	F (52)	LR (99,05), DT (95,59), ANN (95,82)	3	LR (99,05)
4	Ata & Seyrek	2009	50/50 (100)	Türkiye	F (24)	DT (67,92), ANN (77,36)	2	ANN (77,36)
5	Ravisankar ve diğerleri	2011	101/101 (202)	Çin	F (35)	PNN (98,09), LR (66,86), SVM (70,41), GP (94,14), GMDH (93,00), MLFF (78,36)	6	PNN (98,09)
6	Gupta & Gill	2012	29/85 (114)	ABD	F (35)	DT (94,74), NB (87,72), GP (87,72)	3	DT (94,74)
7	Terzi & Şen	2012	25/25 (50)	Türkiye	F (15)	DT (80,0), ANN (91,0)	2	ANN (91,0)
8	Chen ve diğerleri	2014	66/66 (132)	Tayvan	F (24), FO (5)	DT (85,71), LR (80,95), SVM (72,02)	3	DT (85,71)
9	Song ve diğerleri	2014	110/440 (550)	Çin	F (23)	LR (77,9), DT (78,6), BPNN (85,1), SVM (85,5), Ensemble (88,9)	5	Ensemble (88,9)
10	Lin ve diğerleri	2015	129/447 (576)	Tayvan	F (27), FO (6)	LR (88,5), DT (90,3), ANN (92,8)	3	ANN (92,8)
11	Uğurlu & Sevim	2015	97/192 (289)	Türkiye	F (10)	ANN (90)	1	ANN (90)
12	Chen	2016	44/132 (176)	Tayvan	F (23), FO (7)	DT (CHAID) (75,28), DT (CART) (87,97), ANN (82,40), SVM (79,05), BBN (81,01)	5	DT (CART) (87,97)

Hileli Finansal Tabloların Tespitinde Veri Madenciliği Uygulamaları: Mevcut Araştırma Eğilimlerinin İncelenmesi
(2006-2024)

Tuğrul KANDEMİR, Zafer KARDEŞ

13	Hajek & Henriques	2017	311/311 (622)	ABD	F (32)	LR (71,98), NB (57,08), BBN (90,09), DTNB (89,70), SVM (77,36), JRIP (87,15), DT (C4.5) (85,83), DT (CART) (86,36), LMT (85,06), MLP (79,48), VP (51,07), Bagging (87,13), RF (88,07), AdaboostM1 (77,01)	14	BBN (90,09)
14	Sorkun & Toraman	2017	72/0 (72)	Türkiye	F (9)	RF (77,0), ANN (65,0), K-NN (37,0), SVM (60,0), DS (81,0), DT (M5P) (80,0), D. Table (80,0)	7	DS (81,0)
15	Yao ve diğerleri	2019	134/402 (536)	Çin	F (17), FO (7)	SVM (80,63), DT (CART) (74,38), BPNN (73,13), LR (80,00), BBN (73,13), K-NN (78,13)	6	SVM (80,63)
16	An & Suh	2020	1.591/31.628 (33.219)	Güney Kore	F (23)	DT CART (74,76), RF (78,06), Bagging CART (77,50), Boosting CART (77,97), LR (71,95), SVM (71,35), ANN (78,51), Modified RF (79,90)	8	Modified RF (79,90)
17	Craja ve diğerleri	2020	208/7.341 (7.549)	ABD	F (47)	LR (82,52), SVM (82,80), RF (86,53), XGBoost (84,81), ANN (67,90)	5	RF (86,53)
18	Wyrobek	2020	54/58 (112)	ABD	F (15)	Logit (84,70), XGBoost (93,50), RF (94,70), DNN (91,04), BBN (62,00), LDA (92,20), SVM (89,70)	7	RF (94,70)
19	Aksoy	2021	44/44 (88)	Türkiye	F (17)	ANN (96,15), DT (CART) (96,15), SVM (80,77), LR (80,77)	4	ANN (96,15), DT (CART) (96,15)
20	Hamal & Senvar	2021	321/1.384 (1.705)	Türkiye	F (32)	SVM (85,87), NB (58,81), ANN (88,34), K-NN (89,92), RF (93,74), LR (88,08), Bagging (91,27)	7	RF (93,74)
21	Jan	2021	51/102 (153)	Tayvan	F (14), FO (4)	DL-RNN (87,18), DL-LSTM (94,88)	2	DL-LSTM (94,88)
22	Javadian Kootanaee ve diğerleri	2021	45/135 (180)	İran	F (23)	Hybrid (80,0), SVM (73,9), DT (75,0), NB (78,9)	4	Hybrid (80,0)
23	Kırda & Özçelik	2021	19/20 (39)	Türkiye	F (11)	RF (89,83), K-NN (88,87), XGBoost (86,44), SVM (85,49), ANN (84,0), DT (CART), (80,33), NB (67,17)	7	RF (89,83)
24	Papík & Papíková	2021	40/40 (80)	ABD	F (9)	DT (80,05), RF (83,75)	2	RF (83,75)
25	Tatar & Kıymık	2021	35/92 (127)	Türkiye	F (12)	DL (84,25), ANN (81,89), DT (J48) (84,25), RF (80,31), SVM (82,68), K-NN (81,89), AdaBoost (81,10), NB (75,59), LR (82,68), DT (79,53)	10	DT (J48) (84,25), DL (84,25)
26	Kılıç & Önal	2022	48/96 (144)	Türkiye	F (37)	ANN (88,89)	1	ANN (88,89)
27	Papík & Papíková	2022	320/1.200 (1.520)	ABD	F (17)	DA (75,44), DT (79,94), JRip (78,75), K-NN (77,88), LR (72,69), NN (82,25), RF (85,50), SVM (81,19)	8	RF (85,50)
28	Singh Yadav & Sora	2022	739/4.994 (5.733)	ABD	F (6), FO (14)	Hybrid CNN-MRFO (98,85)	1	Hybrid CNN-MRFO (98,85)

Hileli Finansal Tabloların Tespitinde Veri Madenciliği Uygulamaları: Mevcut Araştırma Eğilimlerinin İncelenmesi
(2006-2024)

Tuğrul KANDEMİR, Zafer KARDEŞ

29	Xiuguo & Shengyong	2022	244/5.130 (5.374)	Çin	F (58)	LR (81,09), RF (87,74), SVM (83,92), XGBoost (84,18), ANN (74,56), CNN (84,19)	6	RF (87,74)
30	Ali ve diğerleri	2023	102/1798 (1.900)	Afrika	F (26)	LR (73,88), DT (82,22), SVM (88,88), RF (80,55), AdaBoost (83,33), XGBoost (93,66)	6	XGBoost (93,66)
31	Sun ve diğerleri	2023	76/8.923 (8.999)	ABD	F (28)	LR (64,68), SVM (67,52), RF (62,64), XGBoost (68,01)	4	XGBoost (68,01)
32	Li ve diğerleri	2024	27/2.815 (2.842)	Çin	F (8)	LR (80,87), SVM (83,68), DT (79,12), DL (90,85)	4	DL (90,85)
33	Li ve diğerleri	2024	98/18.536 (18.634)	ABD	F (15)	DT (77,67), LR (71,70), SVM (93,63), RF (78,10), XGBoost (96,94)	5	XGBoost (96,94)
34	Lokanan & Sharma	2024	1.816/2.138 (3.954)	ABD	F (26), FO (9)	LR (60), LDA (92), RF (91), Boosting (91), CatBoost (92), DT (90)	6	CatBoost (92)
35	Nhien ve diğerleri	2024	467/1.768 (2.235)	Vietnam	F (15), FO (10)	RF (88), ANN (97,23)	2	ANN (97,23)
36	Phong ve diğerleri	2024	1.630/1.507 (3.137)	Vietnam	F (3)	ANN (66,9), SVM (66,3), RF (72,3), LR (63,6), KNN (67,1), NB (62,8), DT (61,5)	7	RF (72,3)
37	Riskiyadi	2024	1.058/2.412 (3.470)	Endonezya	F (27)	SGD (61,07), SVM (57,97), K-NN (60,87), DT (77,99), RF (86,25), ERT (86,48), AdaBoost (85,51), Boosting (84,06), NN (81,99)	9	ERT (86,48)
38	Sun ve diğerleri	2024	366/30.636 (31.002)	Çin	F (99)	LR (63,3), XGBoost (86,3), AdaBoost (83,2), SVM (61,9)	4	XGBoost (86,3)
39	Tragouda ve diğerleri	2024	192/560 (752)	Yunanistan	F (8), FO (6)	NN (82,8), LR (81,8), GAM (88,0), RF (87,1), MLKNN (85,6)	5	GAM (88,0)
40	Wang & Chen	2024	120/9.208 (9.328)	Çin	F (60)	LR (59,72), DT (59,93), SVM (60,52), K-NN (60,61), ANN (67,73)	5	ANN (67,73)
41	Zhu ve diğerleri	2024	2.686/26.568 (29.254)	Çin	F (11)	SVM (68,87), RF (69,44), K-NN (74,17), AdaBoost (74,51), XGBoost (71,62), LightGBM (70,22), DT (70,66)	7	AdaBoost (74,51)

Not: Literatür taramasına göre yazarlar tarafından oluşturulmuştur.

Değişken Özellikleri*: Finansal, (F); Finansal Olmayan, (FO).

Kullanılan Yöntemler**: ANN, Artificial neural networks; BBN, Bayesian belief networks; BestCV, Best cross validation; BPNN, Back propagation neural network; CNN, Convolutional neural network; D. Table, Decision table; DA, Discriminant analysis; DL, Deep learning; DNN, Deep neural networks; DS, Decision stump; DT, Decision tree; DTNB, Decision table/naive bayes hybrid; ERT, Extremely randomized trees; GAM, Generalized additive model; GMDH, Group method of data handling; GP, Genetic programming; JRIP, Ripper; K-NN, K-nearest neighbor; LDA, Linear discriminant analysis; LightGBM, Light gradient boosting machine; LMT, Logistic model tree; LR, Logistic regression; MLFF, Multi lane free flow ; MLKNN, Multi label k nearest neighbor; MLP, Multi layer perceptron; NB, Naive bayes; NN, Neural network; PNN, Probabilistic neural network; RF, Random forest; SGD, Stochastic gradient descent; SVM, Support vector machine; VP , Voted perceptron.

KAYNAKÇA

- ACFE. (2020). Report to the nations: 2020 global study on occupational fraud and abuse. <https://legacy.acfe.com/report-to-the-nations/2020/> (Erişim Tarihi:09.04.2023).
- AICPA. (2021). Consideration of fraud in a financial statement audit. <https://us.aicpa.org/content/dam/aicpa/research/standards/auditattest/downloadabledocuments/au-00316.pdf> (Erişim Tarihi: 10.04.2023).
- Aksoy, B. (2021). Finansal tablo hileleri'nin makine öğrenmesi yöntemleri ve lojistik regresyon kullanılarak tahmin edilmesi: borsa İstanbul örneği. *Maliye ve Finans Yazıları*, (115), 27-58. <https://doi.org/10.33203/mfy.733855>
- Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, 14(3), 553-569. [https://doi.org/10.6339/JDS.201607_14\(3\).0010](https://doi.org/10.6339/JDS.201607_14(3).0010)
- Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, 40, 100402. <https://doi.org/10.1016/j.cosrev.2021.100402>
- Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A powerful predicting model for financial statement fraud based on optimized xgboost ensemble learning technique. *Applied Sciences*, 13(4), 2272. <https://doi.org/10.3390/app13042272>
- An, B., & Suh, Y. (2020). Identifying financial statement fraud with decision rules obtained from Modified Random Forest. *Data Technologies and Applications*, 54(2), 235-255. <https://doi.org/10.1108/DTA-11-2019-0208>
- Ashtiani, M. N., & Raahemi, B. (2022). Intelligent fraud detection in financial statements using machine learning and data mining: A systematic literature review. *IEEE Access*, 10, 72504-72525. <https://doi.org/10.1109/ACCESS.2021.3096799>
- Ata, H. A. & Seyrek, İ. H. (2009). Hileli finansal tabloların tespitinde veri madenciliği tekniklerinin kullanımı: imalat firmaları üzerine bir uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14 (2) , 157-170.
- Barman, S., Pal, U., Sarfaraj, M. A., Biswas, B., Mahata, A., & Mandal, P. (2016). A complete literature review on financial fraud detection applying data mining techniques. *International Journal of Trust Management in Computing and Communications*, 3(4), 336-359. <https://doi.org/10.1504/IJTMCC.2016.084561>
- Beemamol, M. (2024). Mapping the trends of financial statement fraud detection research from the historical roots and seminal work. *Journal of Economic Criminology*, 6, 100096. <https://doi.org/10.1016/j.jeconc.2024.100096>
- Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. *SpringerPlus*, 5(1), 89. <https://doi.org/10.1186/s40064-016-1707-6>
- Chen, S., Goo, Y. J. J., & Shen, Z. D. (2014). A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent financial statements. *The Scientific World Journal*, 1, 1-9. <https://doi.org/10.1155/2014/968712>
- Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, 139, 113421. <https://doi.org/10.1016/j.dss.2020.113421>
- Esen, M. F. (2016). Finansal suçların tespitinde veri madenciliği yaklaşımı ve literatüre bakış. *Eskişehir Osmangazi Üniversitesi İktisadi Ve İdari Bilimler Dergisi*, 11(2), 93-118.

- Gaffaroglu, S., & Alp, S. (2023). Detecting frauds in financial statements: A comprehensive literature review between 2019 and 2023 (June). *Pressacademia*, 18(1), 47-51. <https://doi.org/10.17261/Pressacademia.2023.1849>
- Gupta, R., & Gill, N. S. (2012). Prevention and detection of financial statement fraud – An implementation of data mining framework. *International Journal of Advanced Computer Science and Applications*, 3(8). <https://doi.org/10.14569/IJACSA.2012.030825>
- Gupta, S., & Mehta, S. K. (2021). Data mining-based financial statement fraud detection: Systematic literature review and meta-analysis to estimate data sample mapping of fraudulent companies against non-fraudulent companies. *Global Business Review*, 25(5), 1290-1313. <https://doi.org/10.1177/0972150920984857>
- 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. <https://doi.org/10.1016/j.knosys.2017.05.001>
- Hamal, S., & Senvar, O. (2021). Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for Turkish SMEs. *International Journal of Computational Intelligence Systems*, 14(1), 769-782. <https://doi.org/10.2991/ijcis.d.210203.007>
- Jan, C. L. (2021). Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry. *Sustainability*, 13(17), 9879. <https://doi.org/10.3390/su13179879>
- Javadian Kootanaee, A., Poor Aghajan, A. A., & Hosseini Shirvani, M. (2021). A hybrid model based on machine learning and genetic algorithm for detecting fraud in financial statements. *Journal of Optimization in Industrial Engineering*, 14(2), 169-186. <https://doi.org/10.22094/joie.2020.1877455.1685>
- Kılıç, İ., & Önal, S. (2022). Finansal hilelerin tespit edilmesinde kullanılan veri madenciliği yöntemleri ve borsa İstanbul'da bir uygulama. *Muhasebe ve Denetime Bakış*, 22(67), 181-208. <https://doi.org/10.55322/mdbakis.1068503>
- Kırda, K., & Özçelik, M. K. (2021). Finansal tablo hilesi riski taşıyan şirketlerin veri madenciliği ile belirlenmesi. *Muhasebe ve Vergi Uygulamaları Dergisi*, 14(2), 609-639. <https://doi.org/10.29067/muvu.802703>
- 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. <https://doi.org/10.1016/j.eswa.2006.02.016>
- 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.
- Li, B., Yen, J., & Wang, S. (2024). Uncovering Financial Statement Fraud: A Machine Learning Approach With Key Financial Indicators and Real-World Applications. *IEEE Access*, 12, 194859-194870. <https://doi.org/10.1109/ACCESS.2024.3520249>
- Li, W., Liu, X., & Zhou, S. (2024). Deep learning model based research on anomaly detection and financial fraud identification in corporate financial reporting statements. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 123(1), 343-355. <https://doi.org/10.61091/jcmcc123-24>
- Lin, C. C., Chiu, A. A., Huang, S. Y., & Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, 89, 459-470. <https://doi.org/10.1016/j.knosys.2015.08.011>
- Liou, F. M. (2008). Fraudulent financial reporting detection and business failure prediction models: a comparison. *Managerial Auditing Journal*, 23(7), 650-662. <https://doi.org/10.1108/02686900810890625>
- Lokanan, M., & Sharma, S. (2024). The use of machine learning algorithms to predict financial statement fraud. *The British Accounting Review*, 57(1), 101560. <https://doi.org/10.1016/j.bar.2025.101560>

- 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. <https://doi.org/10.1016/j.dss.2010.08.006>
- Nhien, C. T., Hung, D. N., & Binh, V. T. T. (2024). Using random forest and artificial neural network to detect fraudulent financial reporting: Data from listed companies in Vietnam. *Calitatea*, 25(202), 160-173. <https://doi.org/10.47750/qas/25.202.17>
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Med* 18(3): e1003583. <https://doi.org/10.1371/journal.pmed.1003583>
- Papík, M., & Papíková, L. (2021). Application of selected data mining techniques in unintentional accounting error detection. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 16(1), 185–201. doi: 10.24136/eq.2021.007
- 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. <https://doi.org/10.1016/j.accinf.2022.100559>
- Phong, N. A., Tam, P. H., & Tung, N. T. (2024). Identifying fraud financial reports based on signs of income management using machine learning technology: the case of listed companies in Vietnam. *Journal of International Commerce, Economics and Policy*, 15(02), 2450013. <https://doi.org/10.1142/S1793993324500133>
- Ramzan, S., & Lokanan, M. (2024). The application of machine learning to study fraud in the accounting literature. *Journal of Accounting Literature*. <https://doi.org/10.1108/JAL-11-2022-0112>
- Rao, R. K., & Mandhala, V. N. (2024). Unveiling financial fraud: A comprehensive review of machine learning and data mining techniques. *Ingenierie des Systemes d'Information*, 29(06), 2309-2334. <https://doi.org/10.18280/isi.290620>
- 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. <https://doi.org/10.1016/j.dss.2010.11.006>
- Rezaee, Z. (2005). Causes, consequences, and deterrence of financial statement fraud. *Critical perspectives on Accounting*, 16(3), 277-298. [https://doi.org/10.1016/S1045-2354\(03\)00072-8](https://doi.org/10.1016/S1045-2354(03)00072-8)
- Riskiyadi, M. (2024). Detecting future financial statement fraud using a machine learning model in Indonesia: A comparative study. *Asian Review of Accounting*, 32(3), 394-422. <https://doi.org/10.1108/ARA-02-2023-0062>
- Selimoğlu, S., & Altunel, M. (2019). Finansal tablo hileleri ve hile denetimi konusunda Türkiye’de hazırlanmış yüksek lisans/ doktora tezleri kapsamında akademik literatür taraması: 2008-2018. *Muhasebe ve Finansman Dergisi*, 127-140. <https://doi.org/10.25095/mufad.606008>
- Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. *Technological Forecasting and Social Change*, 192, 122527. <https://doi.org/10.1016/j.techfore.2023.122527>
- Sharma, A., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. *International Journal of Computer Applications*, 39(1), 37-47. <https://doi.org/10.5120/4787-7016>
- Singh Yadav, A. K., & Sora, M. (2022). Unsupervised learning for financial statement fraud detection using manta ray foraging based convolutional neural network. *Concurrency and Computation: Practice and Experience*, 34(27), e7340. <https://doi.org/10.1002/cpe.7340>

- Song, X., Hu, Z., Du, J., & Sheng, Z. (2014). Application of machine learning methods to risk assessment of financial statement fraud: Evidence from China. *Journal of Forecasting*, 33(8), 611-626. <https://doi.org/10.1002/for.2294>
- Sorkun, M. C., & Toraman, T. (2017). Fraud detection on financial statements using data mining techniques. *Intelligent Systems and Applications in Engineering*, 5(3), 132-134. <https://doi.org/10.18201/ijisae.2017531428>
- SPK. Sermaye piyasasında bağımsız denetim standartları hakkında tebliğ, Seri: X, No:22. [https://spk.gov.tr/data/635b6d211b41c62c78189ce9/Ba%C4%9F%C4%B1ms%C4%B1z%20Denetim%20Tebliğ%20\(2\).pdf](https://spk.gov.tr/data/635b6d211b41c62c78189ce9/Ba%C4%9F%C4%B1ms%C4%B1z%20Denetim%20Tebliğ%20(2).pdf), (Erişim Tarihi: 14.04.2023).
- Sun, H., Li, J., & Zhu, X. (2023). Financial fraud detection based on the part-of-speech features of textual risk disclosures in financial reports. *Procedia Computer Science*, 221, 57-64. <https://doi.org/10.1016/j.procs.2023.07.009>
- Sun, Y., Zeng, X., Xu, Y., Yue, H., & Yu, X. (2024). An intelligent detecting model for financial frauds in Chinese A-share market. *Economics & Politics*, 36(2), 1110-1136. <https://doi.org/10.1111/ecpo.12283>
- Tatar, B., & Kıymık, H. (2021). Finansal tablolarda hile riskinin tespit edilmesinde veri madenciliği yöntemlerinin kullanılmasına yönelik bir araştırma. *Journal of Yaşar University*, 16(64), 1700-1719. <https://doi.org/10.19168/jyasar.956623>
- Terzi, S. (2012). Hile ve usulsüzlüklerin tespitinde veri madenciliğinin kullanımı. *Muhasebe ve Finansman Dergisi*, (54), 51-64.
- Terzi, S., & Şen, İ. K. (2012). Finansal tablo hilelerinin veri madenciliği yardımıyla tespit edilmesi: üretim sektöründe bir araştırma. *Muhasebe ve Vergi Uygulamaları Dergisi*, 5(2), 25-40.
- Tragouda, M., Doumpos, M., & Zopounidis, C. (2024). Identification of fraudulent financial statements through a multi-label classification approach. *Intelligent Systems in Accounting, Finance and Management*, 31(2), e1564. <https://doi.org/10.1002/isaf.1564>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207-222. <https://doi.org/10.1111/1467-8551.00375>
- Uğurlu, M., & Sevim, Ş. (2015a). Artificial neural network methodology in fraud risk prediction on financial statements; an empirical study in banking sector. *İşletme Araştırmaları Dergisi*, 7(1), 60-89. <https://doi.org/10.20491/isader.2015115752>
- Uğurlu, M., & Sevim, Ş. (2015b). Finansal tablolardaki hile riskinin tahmin edilmesinde karma modellerin nispi başarısı üzerine karşılaştırmalı bir analiz. *Gaziantep University Journal of Social Sciences*, 14(1), 65-88. <https://doi.org/10.21547/jss.256778>
- Wang, D., & Chen, L. (2024). Financial intelligence forecasting model on regression analysis and support vector machine. *Journal of Network Intelligence (JNI)*, 9(3), 1388-1404.
- West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, 57, 47-66. <https://doi.org/10.1016/j.cose.2015.09.005>
- Wyrobek, J. (2020). Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture. *Procedia Computer Science*, 176, 3037-3046. <https://doi.org/10.1016/j.procs.2020.09.335>
- Xiuguo, W., & Shengyong, D. (2022). An analysis on financial statement fraud detection for Chinese listed companies using deep learning. *IEEE Access*, 10, 22516-22532. <https://doi.org/10.1109/access.2022.3153478>

Tuđrul KANDEMİR, Zafer KARDEŐ

Yao, J., Pan, Y., Yang, S., Chen, Y., & Li, Y. (2019). Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: a multi-analytic approach. *Sustainability*, 11(6), 1579. <https://doi.org/10.3390/su11061579>

Yue, D., Wu, X., Wang, Y., Li, Y., & Chu, C. H. (2007). A review of data mining-based financial fraud detection research. *International Conference on Wireless Communications, Networking and Mobile Computing*, 5519-5522. <https://doi.org/10.1109/WICOM.2007.1352>

Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision support Systems*, 50(3), 570-575. <https://doi.org/10.1016/j.dss.2010.08.007>

Zhu, S., Wu, H., Ngai, E. W. T., Ren, J., He, D., Ma, T., & Li, Y. (2024). A financial fraud prediction framework based on stacking ensemble learning. *Systems*, 12(12), 588. <https://doi.org/10.3390/systems12120588>