

MUHASEBEDE ANALİZ YÖNTEMİ OLARAK METİN MADENCİLİĞİ*

Prof.Dr. Birol YILDIZ**

Arş.Gör.Dr. Şafak AĞDENİZ***

Muhasebe Bilim Dünyası Dergisi Haziran 2018; 20(2); 286-315

ÖZ

İşletme ile ilgili alınacak kararlarda menfaat sahiplerinin ilk başvurduğu kaynak genel amaçlı finansal tablolarıdır. Genel amaçlı finansal tablolardan işletme ile ilgili tüm bilgilerin elde edilmesi mümkün değildir. Bu nedenle menfaat sahipleri başka kaynaklara yönelmektedir. Faaliyet raporları, sürdürülebilirlik raporları, entegre raporlar bu kaynaklara örnek olarak verilebilir. Ancak burada bu raporlarda yer alan verilerin analizi menfaat sahipleri için bir sorun olmaktadır. Çünkü büyük oranda yapısal olmayan veri içeren bu raporların analizinde mevcut istatistiksel yöntemler yetersiz kalmaktadır. Metin madenciliği bu soruna çözüm getiren ve muhasebe alanında da son yıllarda sıklıkla kullanılan bir büyük veri analiz yöntemidir. Bu çalışmada muhasebe alanında metin madenciliği çalışmaları incelenerek, metin madenciliğinin muhasebe alanında uygulama alanları hakkında araştırmacılara yol gösterilmesi amaçlanmaktadır.

Anahtar Kelimeler: Metin Madenciliği, Büyük Veri, Yapısal Olmayan Veri, Finansal Rapor, Faaliyet Raporu, Finansal Olmayan Veri

JEL Sınıflandırması: M41, C49

TEXT MINING AS AN ANALYZING METHOD IN ACCOUNTING**ABSTRACT**

General purpose financial statements are the first reference guide of the stakeholders. All of the information about the company can not be attained from general purpose financial statements. So, stakeholders turn to other sources. Annual reports, sustainability reports, integrated reporting can be given as an example for these sources. But, analyzing the data that included in these reports has been a problem

* Bu çalışma "Finansal Raporların Analizinde Metin Madenciliğinin Kullanımı: Borsa İstanbul Şirketlerinin Kurumsal Yönetim Niteliklerinin Tahmini" adlı doktora tezinden üretilmiştir. Makale gönderim tarihi: 07.11.2017; kabul tarihi: 24.04.2018.

**Eskişehir Osmangazi Üniversitesi, İİBF, İşletme Bölümü, Muhasebe ve Finansman ABD, orcid.org/0000-0002-9599-8904, birol.yildiz@gmail.com.

***Eskişehir Osmangazi Üniversitesi, İİBF, İşletme Bölümü, Muhasebe ve Finansman ABD, orcid.org/0000-0003-0373-4694, agdenizsafak@gmail.com.

for the stakeholders. Because, available statistical techniques are insufficient in analyzing these reports that contain too much unstructured data. Text mining is a big data analyzing method that used to solve this problem. Also, it has been using in accounting frequently in recent years. The aim of this study is to guide researchers about text mining application fields in accounting by searching the studies.

Keywords: Text Mining, Big Data, Unstructured Data, Financial Report, Annual Report, Non-Financial Information

JEL Classification: M41, C49

1. GİRİŞ

Verinin, özellikle boyutunda meydana gelen üstel artış günümüzde büyük veri kavramı ile ifade edilmektedir. Büyük verinin yaklaşık %80-90'ı metin, ses, görüntü gibi yapısal olmayan veriden oluşmaktadır. Yapısal verinin analizinde birçok analiz yöntemi başarılı bir şekilde kullanılmakta iken söz konusu bu yöntemler yapısal olmayan verinin analizinde yetersiz kalmaktadır. Çünkü yapısal olmayan veri bu yöntemler dışında ek bir takım analiz yöntemleri gerektirmektedir. Metin madenciliği, literatürde sıklıkla veri madenciliğinin bir alt dalı olarak anılsa da veri madenciliğinden farklı bir analiz yöntemi olup, yarı yapısal ve yapısal olmayan veri analizinde bu soruna çözüm getiren yöntemlerden biridir.

Muhasebenin temel amacı, işletme ile ilgili gerekli faydalı bilgiyi menfaat sahiplerine sağlamaktır. Genel amaçlı finansal tablolar, muhasebenin bu işlevinin yerine getirilmesinde en sık kullanılan araçlardır. Bu noktada zaman zaman birbiri yerine kullanılan finansal tablo ile finansal rapor kavramları arasındaki farkın ortaya konulması önemli olacaktır. TMS (Türkiye Muhasebe Standartları) 1 Finansal Tabloların Sunuluşu Standardında genel amaçlı finansal tablolar, *“bir işletmenin, belirli bilgi ihtiyaçları doğrultusunda uyarlanmış finansal tablo hazırlamasına gerek duymayan kullanıcıların ihtiyaçlarını karşılamak üzere hazırlanan mali tablolar”* olarak tanımlanmıştır ve beş adet genel amaçlı finansal tablo belirlenmiştir. Bu tablolar;

- Dönem sonu finansal durum tablosu,
- Döneme ait kâr veya zarar ve diğer kapsamlı gelir tablosu,
- Döneme ait özkaynak değişim tablosu,
- Döneme ait nakit akış tablosu,
- Önemli muhasebe politikalarını ve diğer açıklayıcı bilgileri içeren dipnotlardır.

Bu finansal tablolar işletme ile ilgili temel bilgileri bir takım oranlar, rakamlar aracılığıyla sunmaktadır. Sunulan bu bilgiler işletmenin varlıkları, kaynakları, gelirleri, giderleri ve nakit akışlarına ilişkin temel finansal bilgileri içermektedir. Muhasebenin temel kavramlarından tam açıklama kavramı, 1 Seri No’lu Muhasebe Sistemi Uygulama Genel Tebliği’nde şu şekilde tanımlanmıştır:

“Mali tabloların bu tablolardan yararlanacak kişi ve kuruluşların doğru karar vermelerine yardımcı olacak ölçüde yeterli, açık ve anlaşılır olmasını ifade eder. Mali tablolarda finansal bilgilerin tam olarak açıklanması yanında, mali tablo kalemleri kapsamında yer almayan ancak alınacak kararları etkileyebilecek, gerçekleşmesi muhtemel olaylara da yer verilmesi bu kavramın gereğidir.”

Genel amaçlı finansal tabloların tam açıklama kavramı gereğince menfaat sahiplerine tüm bilgileri yeterince sunamadığı gerek kavramsal çerçevede gerekse TMS 1’de belirtilmiştir. Söz konusu bu maddeler aşağıda verilmiştir:

• Kavramsal Çerçevenin A6. Maddesinde finansal raporlar ile ilgili aşağıdaki husus belirtilmiştir:

“Finansal raporlar mevcut ve potansiyel yatırımcıların, borç verenlerin ve kredi veren diğer tarafların ihtiyaç duyduğu bütün bilgileri sağlamazlar. Ayrıca, genel amaçlı finansal raporların ihtiyaç duyulan tüm bilgileri sağlaması da beklenemez. Söz konusu kullanıcıların diğer kaynaklardan elde edilecek bilgileri de göz önünde bulundurması gerekir (örneğin genel ekonomik duruma, ekonomik beklentilere, siyasi olaylara, siyasi duruma, işletmenin bulunduğu endüstriye ve işletmeye ilişkin bilgiler gibi).”

• TMS 1 Madde 13’de, işletmelerin finansal tabloların dışında, yönetimin, işletmenin finansal performansı ve finansal durumu ve karşılaştığı temel belirsizliklerin ana esaslarını açıklayan, bir finansal değerlendirme sunduğu belirtilmiştir. Böyle bir rapor aşağıdaki değerlendirmeleri içermektedir:

○ *İşletmenin faaliyette bulunduğu çevredeki değişiklikler de dâhil olmak üzere, finansal performansı belirleyen temel unsurlar ve etkiler, işletmenin bu değişikliklere tepkisi ve bunların etkisi ve temettü politikası da dâhil, finansal performansın sürdürülmesi ve iyileştirilmesi için işletmenin yatırım politikası,*

○ *İşletmenin fon kaynakları ve hedeflenen borç özkaynak oranı,*

○ *İşletmenin TFRS (Türkiye Finansal Raporlama Standartları)'lere göre finansal durum tablosunda (bilançosunda) muhasebeleştirilmeyen kaynakları.*

• TMS 1 Madde 14'de ise “*özellikle çevresel faktörlerin önemli olduğu sanayi kollarında çalışan ve personelin önemli bir kullanıcı grubu olarak görüldüğü sektörlerdeki birçok işletmenin, finansal tabloların dışında, çevresel raporlar ve katma değer raporları gibi raporlar da sunduğu, finansal tablolardan ayrı olarak sunulan bu raporların ve tabloların TFRS'lerin kapsamı dışında olduğu*” belirtilmiştir.

Belirtilen bu nedenlerden dolayı, bu tablolarda yer almayan bilgiler; faaliyet raporları, sürdürülebilirlik raporları, analist raporları, ilk halka arz izahnamaları gibi raporlar ile sunulmaktadır ve genel amaçlı finansal tablolar ile birlikte finansal raporlama kavramı adı altında anılmaktadır. Dolayısıyla finansal raporlama daha geniş bir kavramdır ve genel amaçlı finansal tablolarda sunulamayan verileri de içermektedir.

Genel amaçlı finansal tabloların sınırlılığı ile ilgili çalışmalar 1990'lı yıllarda başlamıştır. Rimerman (1990, 79) ile Amir ve Lev (1996) tarafından yapılan çalışmalarda finansal raporlar tarafından sağlanan bilgiler ile kullanıcıların ihtiyaçları arasındaki farklılığın artmasıyla birlikte finansal tablo kullanıcılarının bilgi ihtiyaçlarını karşılamak için başka bilgi kaynaklarına yöneldikleri belirtilmiştir. Genel amaçlı finansal tabloların maddi olmayan duran varlıkları tam olarak sunamaması (Lev 2003; Alwert ve diğerleri 2009,309; Lev ve Gu, 2016), bu tabloların tarihi maliyet bilgilerini kullanması (Eccless ve diğerleri, 2010, 79; Lev,1989; Lew ve Zarowin, 1999; Francis ve Schipper, 1999) ve genişleyen işletme çevresi genel amaçlı finansal tablolara getirilen eleştirilerin ana başlıklarını oluşturmaktadır. Söz konusu durum işletme değerinin tam olarak belirlenememesine ve menfaat sahiplerinin yanlış kararlar almalarına neden olmaktadır. Bu nedenle, işletme ile ilgili finansal tablolarda sunulamayan bilgilerin¹ sunulmasına yönelik çeşitli raporlama çerçeveleri üzerinde çalışılmaktadır. Bu raporlarda işletmelerin finansal sonuçları ile birlikte faaliyetlerine ve bu faaliyetlerinin etkilerine yönelik çok çeşitli bilgiler sunulmaktadır. Bu verilere işletmenin kurumsal yönetim uygulamaları, insan kaynaklarına ilişkin politikaları, faaliyetlerinin çevreye olan etkileri örnek olarak verilebilmektedir.

¹ Literatürde finansal tablo dışında sunulan her türlü veri için finansal olmayan (nonfinancial) veri kavramının kullanıldığı görülmektedir. Ancak finansal karar süreçlerinde kullanılacak her türlü verinin finansal olduğu kanaatinde olduğumuz için çalışmada finansal tablolarda sunulamayan veri kavramı kullanılmıştır.

Genişleyen finansal raporlama çerçevesi ile birlikte veri analizinde yaşanan sorun muhasebe alanında da kendini göstermektedir. Genel amaçlı finansal tablolar oran analizi, trend analizi, karşılaştırmalı tablolar analizi gibi finansal tablolar analizi ile analiz edilmektedir. Ancak, finansal raporlarda, metin, resim gibi birçok nitel ve yapısal olmayan veri sunulmaktadır. Dolayısıyla finansal raporlarda yer alan finansal tablo dışı verilerin analizi menfaat sahipleri için bir sorun olmaktadır. Metin madenciliği analiz yöntemi ile finansal tablolarda sunulamayan yapısal olmayan veriler analiz edilebilmektedir. Böylece menfaat sahiplerine daha faydalı bilgi sunulabilmektedir. Yapılan metin madenciliği çalışmalarının sonuçları, finansal tablolarda sunulamayan verilerin analizinin menfaat sahiplerine finansal tablo analizinin yanında faydalı bilgi sağladığını göstermektedir.

Verilen bu bilgiler kapsamında, çalışma ile muhasebe alanında yapısal olmayan veri analizinde metin madenciliği kullanımı konusunda bir farkındalık oluşturulması amaçlanmıştır. Bu amaç doğrultusunda izleyen bölümde metin madenciliği konusu ele alınmış, üçüncü bölümde ise literatür incelemesi yapılarak muhasebe alanında metin madenciliği kullanım alanları hakkında bilgi verilmiştir.

2. METİN MADENCİLİĞİ

Bilgisayar ve internet teknolojisi geliştikçe üretilen, saklanan ve işlenen veri miktarında meydana gelen üstel artış neticesinde “büyük veri” olgusu ve yeni bir alan olan “veri bilimi” gelişmeye başlamıştır (Gürsakar 2014, 1).

Günümüzde bilgi değişiminde en çok kullanılan araçlardan birisi metinlerdir (Jusoh ve Alfawareh 2012, 431). Var olan verinin %80-90'ı web ortamı, e-posta, sosyal medya gibi ortamlardan elde edilmektedir. Elde edilen bu verilerin de metin, resim, ses vb biçimde olan yapısal olmayan veri olduğu görülmektedir. İstatistiksel yöntemler ile yapısal veri analizi başarılı bir şekilde gerçekleştirilmektedir. Ancak yapısal olmayan ve yarı yapısal verilerin analizi ayrı bir yöntem gerektirmektedir. Çünkü istatistiksel yöntemler temel olarak yapısal olmayan büyük veri analizinde yetersiz kalmaktadır.

Metin madenciliği yapısal olmayan veya yarı yapısal verilerin analizinde kullanılan büyük veri analiz yöntemlerinden birisidir. Bugün Dünya’da büyük veri analizi sağlık, ekonomi, güvenlik, işletmecilik gibi birçok alanda sıklıkla kullanılmaktadır.

2.1. Metin Madenciliğinin Tanımı

Metin madenciliği, metin verilerini kullanarak bunlardan anlamlı bilgiler üreten bir büyük veri analiz yöntemidir. Analize konu olan veri sayısı fazla olduğunda, analiz çoğunlukla bilgisayar destekli yapılmaktadır. Bu nedenle makine öğrenmesinin sıklıkla kullanıldığı bir analiz metodudur.

Metin madenciliği ile ilgili literatür incelendiğinde, metin madenciliği kavramının yerine metin veri madenciliği (text data mining), veri tabanlarından bilgi keşfi (knowledge discovery in databases), metin analizi (text analytics) gibi kavramların da kullanıldığı görülmektedir.

Metin madenciliği, özellikle 2000’li yıllarda daha fazla ilgi gören, serbest formatta bulunan metinler içerisinde yer alan fakat daha önceden bilmediğimiz bilgileri ortaya çıkarmamızı sağlayan işlemler bütünüdür (Oğuzlar 2011, 4). Metin madenciliği çalışmaları, metni veri kaynağı olarak kabul eden veri madenciliği (data mining) çalışmasıdır. Diğer bir tanımla metin üzerinden yapılandırılmış (structured) veri elde etmeyi amaçlamaktadır (Şeker 2015, 30).

Oğuzlar (2011, 7) metin madenciliğini, belirgin bir formatı olmayan, yazı içerisindeki veriler içerisinden gizli nitelikli bilgilerin çıkarılması ve düzensiz haldeki verinin formatlanması süreci olarak tanımlamıştır. Metin madenciliği yarı yapısal ve yapısal olmayan verilerin analiz edilerek bu verilerden bilgi elde etmek olarak tanımlanabilir. Miner ve diğerleri (2012, 30) metin madenciliğini “metinleri sayılara dönüştürme” teknolojileri olarak tanımlamıştır. Başka bir tanıma göre ise metin madenciliği metin verilerinden yüksek kalitede bilginin elde edilmesi amacıyla yapılan anlam analizidir (Gaikwad ve diğerleri 2014, 42). Metin madenciliğinin amacı, yapılandırılmamış bilgiyi işleyerek metinden anlamlı sayısal indeksler çıkararak, çeşitli veri madenciliği (istatistiksel ve makine öğrenmesi) algoritmaları tarafından metnin içerdiği enformasyonu erişilebilir kılmaktır (Gürsakar 2014, 46).

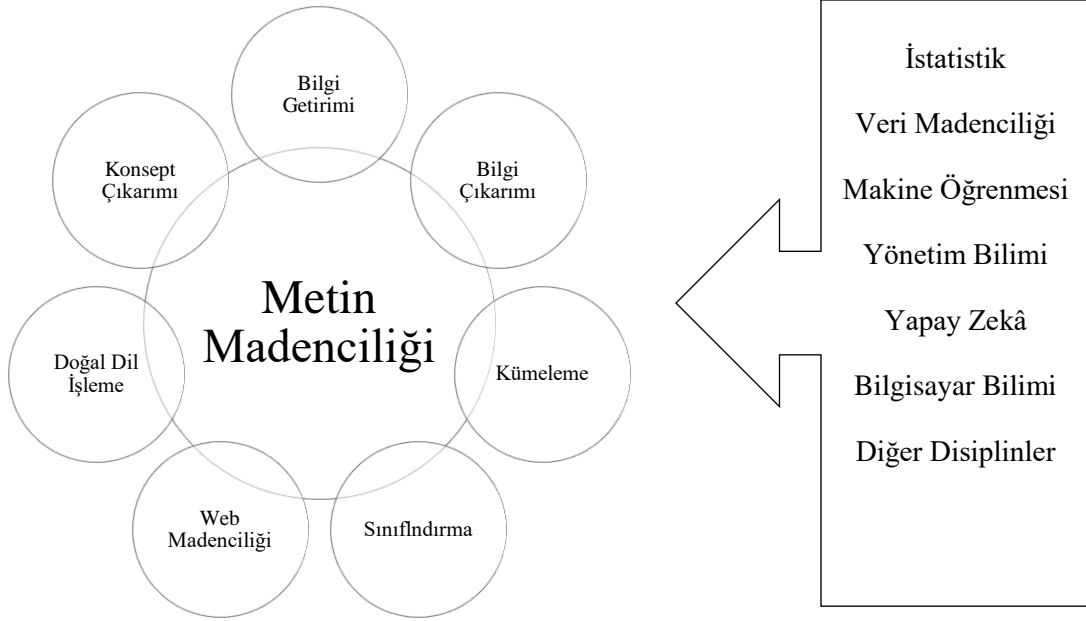
İnsanların ne düşündüğünü, neler hissettiklerini anlamak için geleneksel yöntemler olan anketler, odak görüşmeler ve diğer araştırma yöntemlerinin ötesinde, yayınlanmış yazılı metinleri ele almak ve bu metinleri de tıpkı bir büyük bir veri tabanı gibi analiz etmek ve sayısal değerlere dönüştürmek metin madenciliğinin temel fonksiyonudur (Atan 2016,128).

Metin madenciliği veri madenciliğinin bir çeşidi olan, yeni bir disiplindir (Gao ve diğerleri 2007, 2654). Bazı çalışmalarda metin madenciliği veri madenciliğinin alt dalı gibi görülmektedir. Aslında metin madenciliği, içine veri madenciliğini de alan daha geniş bir kavramdır. Metin madenciliği yapısal olmayan verileri analiz etmede kullanılan bir yöntem, veri

madenciliği ise yapısal verileri analiz etmede kullanılan bir yöntemdir. Metin madenciliği ile veriler yapısal hale getirilmekte, bu aşamadan sonra yapısal hale gelen verilere veri madenciliği algoritmaları uygulanmaktadır.

2.2. Metin Madenciliği Uygulamaları

Metin madenciliği uygulamaları yedi başlık altında incelenebilmektedir. Bu uygulama alanları birbirinden farklı olmakla birlikte aynı zamanda birbiriyle ilişkili alanlardır.



Şekil 1. Metin Madenciliği Uygulamaları ve Metin Madenciliğinin İlişkili Olduğu Alanlar

Kaynak: Miner ve diğerleri 2012, 40.

Şekil 1’de metin madenciliği uygulamaları ve bu uygulamalarda yararlanılan çeşitli disiplinler yer almaktadır. Bu uygulamalar hakkında bilgilere devam eden bölümlerde yer verilmektedir.

2.2.1. Bilgi Getirimi/Erişimi

Bir bilgi erişim sisteminin temel işlevi, kullanıcıların bilgi ihtiyaçlarını karşılaması mümkün olan belge koleksiyonundaki ilgili belgelerin tümüne erişmek, ilgili olmayanları ise ayıklamaktır (Oğuzlar 2011, 15). Bilgi erişiminde, anahtar sözcükler kullanılarak büyük metin verilerinden istenilen özellikte olan metinlerin bulunması amaçlanmaktadır. Günümüzde bilgi erişimine

verilebilecek en güzel örnek birçok insan tarafından sıklıkla kullanılan Google gibi arama motorlarıdır.

Bilgi erişim sistemlerinin performansının değerlendirilmesinde duyarlılık veya hassasiyet olarak adlandırılan (precision), anma veya çağırma olarak da isimlendirilen (recall) ve bu iki ölçütün kombinasyonundan oluşturulan F-skoru ölçütleri kullanılmaktadır (Oğuzlar 2011, 16).

2.2.2. Belge Kümeleme

Belge kümeleme, benzer belgelerin gruplandırılmasıdır. Bu uygulamada daha çok veri madenciliği algoritmalarından yararlanılmaktadır.

Kümeleme işleminde sınıflandırma yapılacak kategoriler önceden belli değildir. Bu nedenle, kümeleme analizi gözetimsiz (denetimsiz) öğrenme (unsupervised learning) yöntemidir (Tunalı 2011, 11). Gözetimsiz öğrenme modelinde kullanılan tüm değişkenler bağımsız değişkenlerdir. Bir çıktı veya aynı anlama gelmek üzere bir bağlı değişken bulunmadığından, bu bağlı değişken ile kıyaslanacak bir tahmin değeri de bulunmamaktadır. Denetimsiz öğrenmede ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır (Oğuzlar 2011, 69).

2.2.3. Belge Sınıflandırma

Belge sınıflandırma işlemi nesnelerin daha önceden bilinen sınıflara ya da kategorilere dâhil edilmesidir (Tunalı 2011, 9). Sınıflandırma işlemi gözetimli (denetimli) bir öğrenme (supervised learning) yöntemidir.

Örnekten öğrenme olarak da isimlendirilen gözetimli öğrenimde bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilmektedir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin örneklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Öğrenme süreci tamamlandığında tanımlanan kural cümleleri verilen yeni örneklerle uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir (Oğuzlar 2011, 62).

Belge sınıflandırma teknikleri büyük verinin saklanması, paylaşılması, transfer edilmesi, analizi ve görselleştirilmesi gibi zorlukları ortadan kaldıran etkili yöntemler sağlamaktadır (Jeong ve diğerleri 2016, 222).

Çoğu sınıflandırma algoritması pozitif ve negatif örnekleri içeren iki eğitim seti gerektirmektedir. Birinci kategoride istediğimiz veriler, ikincisinde ise istemediğimiz veriler yer

almaktadır (Bogaerd ve Aerts 2011, 13416). İstenmeyen maillerin spam kutusunda gösterilmesi sınıflandırma işleminin en çok bilinen uygulamasıdır.

2.2.4. Web Madenciliği

Web madenciliği, web üzerinde yer alan büyük hacimli verilerin analiz edilmesi olarak tanımlanmaktadır. Web madenciliği temel olarak üç grupta incelenmektedir (Miner ve diğerleri 2012, 995):

- *Web kullanım madenciliği:* Kullanıcıların web sayfaları ile olan ilişkileri analiz edilmektedir.
- *Web içerik madenciliği:* Bazı kaynaklarda web metin madenciliği olarak da adlandırılmakta olan web içerik madenciliği ile internet kullanıcılarının ne ile ilgilendikleri ve neyi araştırdıkları analiz edilmektedir.
- *Web yapı madenciliği:* Web yapı madenciliği çalışmaları ise web sitelerinin içeriğine ve birbirleriyle olan ilişkilerine yoğunlaşmaktadır.

Facebook, Twitter gibi sosyal medya kanallarının kullanımının giderek artması web madenciliğinin öneminin artmasına neden olmaktadır (Miner ve diğerleri 2012, 37). Web madenciliğinin günümüzde birçok alanda kullanılmasının en önemli nedeni, kişilerin web sayfalarında göstermiş oldukları davranışların, hareketlerin ve yapmış oldukları işlem bilgilerinin var olan iş süreçlerine entegrasyonunu sağlayarak müşterinin en iyi şekilde anlaşılmasını sağlayan müşteri odaklı bir sistem oluşturmalarıdır (Dolgun ve diğerleri 2009, 52).

2.2.5. Bilgi Çıkarımı

Bilgi çıkarımı, metin madenciliği uygulamalarında günümüzde kullanılan belki de en önemli tekniktir. Bu sistemler yapısal olmayan metinleri veri tabanı tablosuna aktarabilecek bir formata dönüştürmektedir (Oğuzlar 2011, 17).

Bilgi çıkarımı sistemleri kullanıcılara, dokümanların “makine tarafından okunabilir” görünümünden, bunların “makine tarafından anlaşılabilir” biçimine geçiş yapma imkânı vermektedir (Oğuzlar 2011, 18).

2.2.6. Doğal Dil İşleme

Doğal dil işleme, ana işlevi bir doğal dili çözümleme, anlama, yorumlama ve üretme olan bilgisayar sistemlerinin tasarımını ve gerçekleştirilmesini konu alan mühendislik alanıdır (Oğuzlar 2011, 11).

Doğal dil işleme, doğal dildeki metinlerin ve/veya seslerin bilgisayar ortamında işlenmesi üzerine çalışmalar yürütülen, bilgisayar bilimi ve dil bilimciliğinin bir alt bilim dalıdır (www.bb.itu.edu.tr).

Doğal dil işleme, doğal dillerin kurallı yapısının çözümlenerek anlaşılması veya yeniden üretilmesi amacını taşır. Bu çözümlenmenin insana getireceği kolaylıklar, yazılı dokümanların otomatik çevrilmesi, soru-cevap makineleri, otomatik konuşma ve komut anlama, konuşma sentezi, konuşma üretme, otomatik metin özetleme, bilgi sağlama gibi birçok başlıkla özetlenebilmektedir. Bilgisayar teknolojisinin yaygın kullanımı, bu başlıklardan üretilen uzman yazılımların gündelik hayatımızın her alanına girmesini sağlamıştır. Örneğin, tüm kelime işlem yazılımları birer imla düzeltme aracı taşımaktadır. Bu araçlar aslında yazılan metni çözümlenerek dil kurallarını denetleyen doğal dil işleme yazılımlarıdır (www.wikipedia.org).

Bu alandaki temel araştırmalar aşağıda verilmiştir (Oğuzlar 2011, 12):

- Doğal dillerin işlev ve yapısının daha iyi anlaşılması,
- Bilgisayarlar ile insanlar arasında arabirim olarak doğal dil kullanmak ve bu şekilde bilgisayarlar ile insanlar arasındaki iletişimi kolaylaştırmak,
- Bilgisayar ile dil çevirisi yapmak.

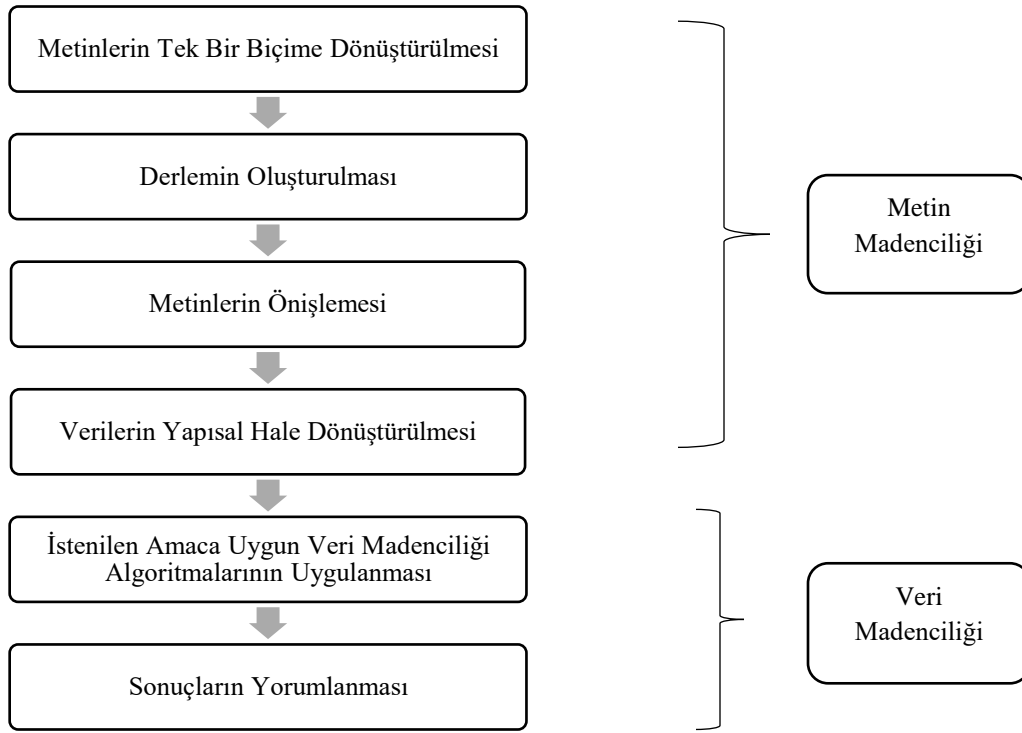
Doğal dil işlemenin temelleri bilgisayar ve bilgi bilimleri, matematik, elektrik elektronik mühendisliği, yapay zekâ ve psikoloji gibi birçok disipline dayanmaktadır (Vjayanani ve diğerleri 2016, 8).

2.2.7. Kavram Çıkarımı

Kavram çıkarımı, kelimelerin ve deyimlerin anlamsal olarak benzer gruplara ayrılarak metin verilerin ana temasının belirlenmesidir. Dolayısıyla, metin madenciliğinin en zor uygulama alanlarından birisidir. Makinelerin insan gibi anlamasına yönelik yapılan çalışmalar ile bu alanda ilerleme kaydedilmektedir (Miner ve diğerleri 2012, 38).

2.3. Metin Madenciliği Aşamaları

Metin madenciliğinde analize konu veriler yapısal olmayan veya yarı yapısal verilerdir. Metin madenciliği ile bu veriler yapısal hale dönüştürülerek analize hazır hale getirilmektedir. Genel olarak bir metin madenciliği uygulaması Şekil 2’de verilen adımları içermektedir:



Şekil 2. Metin Madenciliği Aşamaları

Şekil 2’de verilen bu aşamalarda genellikle makine öğrenmesinden yararlanılmaktadır. Çünkü veri sayısı az olduğunda insan gücü yeterli iken, veri sayısının artması olası hataları arttırmakta ve çok zaman almaktadır.

3. MUHASEBE ALANINDA METİN MADENCİLİĞİ KULLANIMI

Metin madenciliği arařtırmalarının ilk yıllarında bu alan daha çok veri uzmanlarının ilgi alanı iken günümüzde birçok alandan arařtırmacının ilgisini çeken bir alan haline gelmiştir (Jusoh ve Alfawareh 2012, 431). Muhasebe ve finansal analizde metin verilerin artması ve finansal tabloların işletme ile ilgili verilecek kararlarda yetersiz kalması ile birlikte metin madenciliği ile ilgili çalışmaların bu alanda da başladığı görülmektedir.

Li (2010, 143-144) işletmeler tarafından açıklanan raporlarda yer alan metin verilerin anlaşılmasının muhasebe arařtırması için önemini iki başlıkta belirtmektedir:

- Metinlerde yer alan veriler finansal verilerin içeriğine ilişkin bilgi verebilir. Örneğin hasılat iki farklı şekilde rapor edilebilir. Hasılatın farklı raporlanmasının nedeni benimsenen hasılat tanımı olabileceği gibi yönetimin farklı yaklaşımından da kaynaklanabilir.

• Yöneticilerin iletişim şekilleri, belirli bir takım yönetsel özellikleri dolayısıyla kurumsal kararların anlaşılmasını sağlar. Yönetimin bir iletişim şekli olarak metin veriler, araştırmacılara yöneticilerin davranışsal yaklaşımlarını ve işletme davranışını anlamayı sağlar.

Muhasebe alanında metin madenciliği kullanan çalışmalar tarihsel boyutta incelendiğinde, analiz açısından sürecin okunabilirlik çalışmaları ile başladığını söylemek mümkündür. Söz konusu gelişim metin madenciliğinin gelişim süreci ile de paralellik göstermektedir. Metin madenciliğinin muhasebe alanındaki gelişim süreci Şekil 3’deki gibi özetlenebilmektedir:



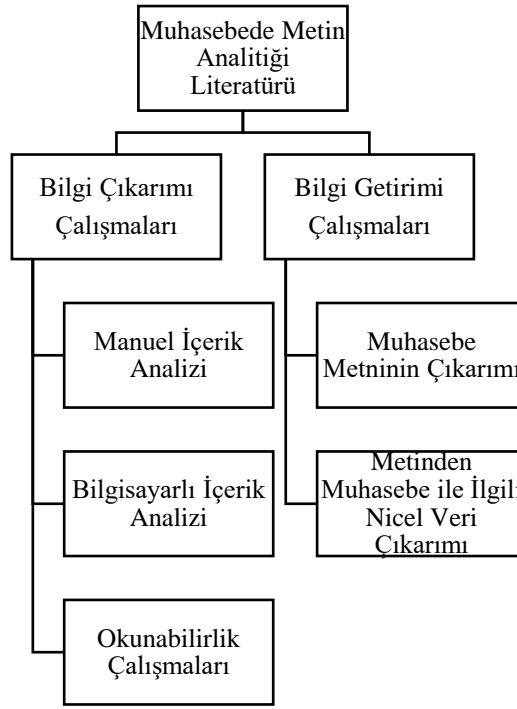
Şekil 3. Muhasebede Metin Madenciliği Çalışmaları Süreci

Muhasebe alanında yapılan çalışmalar incelendiğinde metin madenciliği analizinde kullanılan belgeler aşağıdaki gibi sıralanabilmektedir:

- Faaliyet ve çeyrek raporları,
- Finansal tablo dipnotları,
- Sürdürülebilirlik raporları,
- Yönetici mektupları,
- SEC belgeleri,
- 10-K belgelerinin yönetici tartışmaları ve analizleri bölümü,
- İlk halka arz izahnameleri,
- Analist raporları,
- Gazete ve sosyal medya haberleri.

Görüldüğü üzere karar alıcılara faydalı bilgi sağlayacak finansal tablolar dışında farklı ve karar alıcılar için değerli bilgiler içeren birçok veri kaynağı mevcuttur.

Fisher ve diğerleri (2010, 3) muhasebe alanında metin analitiği konusunda yapılan çalışmaları iki gruba ayırmışlardır:



Şekil 4. Muhasebe Alanında Metin Analitiği ve Bilgi Çıkarımı Araştırma Alanları

Kaynak: Fisher ve diğerleri 2010, 3.

Başka bir çalışmada ise Fisher ve diğerleri (2016, 159) muhasebe, denetim ve finans alanında metin madenciliği kullanan çalışmalar ile ilgili yaptıkları literatür taramasında aşağıdaki sonuçlara ulaşılmıştır:

Tablo 1. Yöntemine Göre Sınıflandırılmış Metin Madenciliği Çalışmaları

Kullanılan Yöntem	Muhasebe	Denetim	Finans	Toplam
Manuel Metin Madenciliği	48	10	28	86
Temel Metin Madenciliği	19	7	55	81
Literatür Taraması	2	-	2	4
NLP+Makine Öğrenmesi/Yapay Zekâ	20	15	60	95
Toplam	89	32	145	266

Muhasebe literatüründe metin madenciliği çalışmalarını çeşitli şekillerde sınıflandırmak mümkündür. Kullanılan metin madenciliği tekniği, kullanılan veri kaynağı, çalışmanın amacı, kullanılan algoritma sınıflandırma şekillerine örnek olarak verilebilir.

3.1. Çalışmanın Yöntemi

Bu çalışmada muhasebe alanında yapılmış metin madenciliği çalışmaları hakkında bilgi verilmesi amaçlanmıştır. Öncelikle bu alanda literatür taraması yapan Fisher ve diğerleri (2010) ve Fisher ve diğerleri (2016)² tarafından yapılan çalışmalar incelenmiştir. Ayrıca Science Direct, EbscoHost, Google Scholar veri tabanları “text mining/analytic and accounting”, “text mining/analytic and audit” anahtar kelimeleri ile taranmış ve içerik analizi yapılmıştır. Çalışmada, 1960 yılı ile 2017 yılı arasında yapılmış çalışmalar ele alınmıştır.

3.2. Çalışmanın Bulguları

Muhasebe alanında Türkiye’de metin madenciliği kullanılarak yapılan bir çalışma bulunmamaktadır. Finans alanında ise Atan (2016) işletme ile ilgili internet haberlerini kullanarak bu haberlerin kişilerin yatırım tercihleri üzerine etkilerini duygu (sentiment) analizi ile analiz etmiştir. Yabancı literatürde ise metin madenciliği ile ilgili olarak özellikle hile, iflas, işletme performansı gibi konularda yapılmış çalışmalar söz konusudur.

Bu çalışmada literatür taraması aşağıda verilen beş başlık altında özetlenmiştir:

- Hileli işlemlerin tespiti,
- İşletme performansının tahmini,
- İflas tahmini,
- Sürdürülebilirlik ve kurumsal yönetim,
- Diğer çalışmalar.

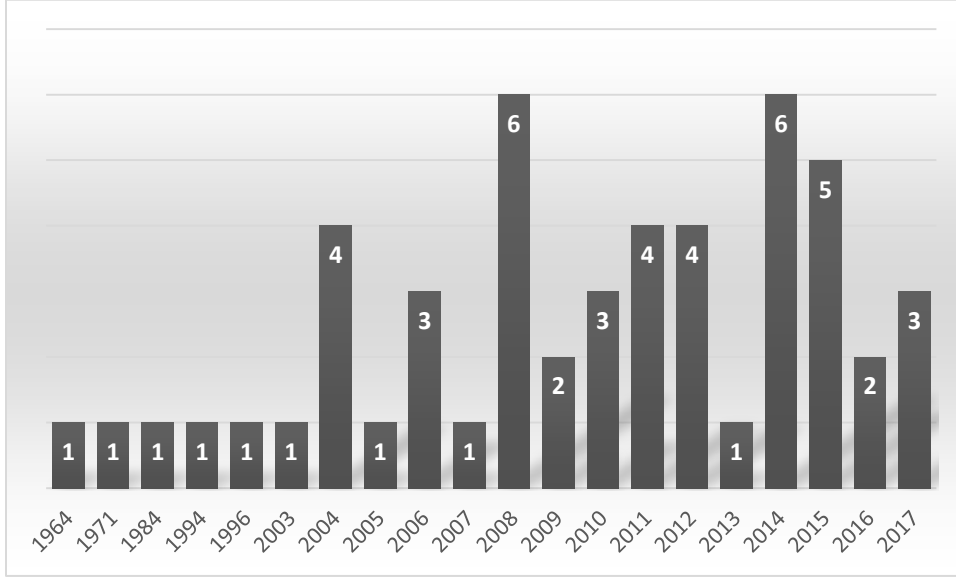
İncelenen çalışmaların dağılımına ilişkin bilgiler Tablo 2’de verilmektedir.

Tablo 2. Amacına Göre Sınıflandırılmış Metin Madenciliği Çalışmaları

	Okunabilirlik	İçerik Analizi	Metin Madenciliği	TOPLAM
Hileli İşlem Tespiti	-	1	8	9
İşletme Performansının Tahmini	1	11	5	17
İflas Tahmini	-	-	2	2
Sürdürülebilirlik ve Kurumsal Yönetim	-	-	7	7
Diğer Çalışmalar	2	3	11	16
TOPLAM	3	15	33	51

² Bu literatür araştırmasında yazarlar toplamda 266 adet çalışmayı bibliyometrik olarak analiz etmişlerdir. Bu makalede ise çalışmaların içeriklerini incelemek amaçlandığından tam metnine ulaşılabilen ve en çok atıf alan çalışmalar kullanılmıştır.

Tablo 2’de görüleceği üzere metin madenciliği en çok işletme performansının tahmini alanında kullanılmaktadır. Söz konusu bu çalışmalar yayınlandığı yıllar itibariyle Grafik 1’de gösterilmektedir.



Grafik 1. Yıllar İtibariyle Yapılan Çalışmalar

Grafik 1’de görüleceği üzere finansal tabloların sınırlılıklarının sorgulanmaya başladığı 1990’lı yıllardan sonra metin madenciliği çalışmaları artmıştır. Okunabilirlik çalışmaları metin madenciliği gelişim sürecinin ilk basamağıdır. Dolayısıyla muhasebe alanında yapılan ilk çalışmalar da okunabilirlik çalışmalarıdır. Yapılan çalışmalar yayın yerlerine göre Tablo 3’de gösterilmektedir.

Tablo 3. Çalışmaların Yayınlandıkları Dergiler

Çalışma Kâğıdı (Working Paper)	2
Bildiri	4
Doktora Tezi	1
Decision Support Systems	3
Journal of Emerging Technologies in Accounting	6
Intelligent Systems in Accounting Finance and Management	2
Knowledge Based Systems	1
Contemporary Accounting Research	2
Academy of Management Journal	1
Journal of Business Finance and Accounting	1

Journal of Accounting Research	3
The Accounting Review	3
Accounting and Business Research	1
Accounting Auditing and Accountability Journal	1
Journal of Accounting Literature	1
The Journal of Finance	3
European Journal of Operational Research	2
Review Accounting Studies	1
Journal of Applied Corporate Finance	1
The International Journal of Digital Accounting Research	3
International Journal of Accounting Information Systems	1
Diğer Dergiler*	8
*International Journal of Advanced Computer Science and Applications, Tourism and Management Studies, The Journal of Corporate Citizenship, Computers in Industry, Environmental Modelling and Software, International Journal of Computational Intelligence and Applications, Intelligent Data Analysis, Corporate Ownership and Control Journal	

51 adet çalışmanın 37'sinin muhasebe alanında yayın yapan dergilerde yayımlandığı görülmektedir. Bu dergiler içinde ise “Journal of Emerging Technologies in Accounting” adlı dergi metin madenciliği çalışmalarının en fazla yayımlandığı dergi olmuştur. Muhasebe genişleyen finansal raporlama çerçevesi birçok alandan araştırmacı için değerli bilgiler içerdiğinden özellikle bilgisayar ve veri bilimi alanında yapılan çalışmalarda kullanılmaktadır.

Bu bölümde ele alınan çalışmalar hakkında bibliyografik bilgilere yer verilmiştir. Devam eden bölümde ise muhasebe alanında yapılan bu çalışmalar detaylı bir şekilde incelenmektedir.

3.2.1. Hileli İşlemlere İlişkin Yapılmış Metin Madenciliği Çalışmaları

Denetim alanında yapılan bu çalışmalarda finansal tablo hilelerine ilişkin çalışmaların çokluğu göze çarpmaktadır.

Cecchini (2005) iflas ve hileli işlemlerin metin veriler kullanılarak daha doğru tahmin edilebileceğini gösterdiği doktora tezinde metin veriler için 10-K raporlarını kullanmıştır. Kernel yöntemlerinin ve bilgi getiriminin kullanıldığı çalışmada iflas tahmini ve hileli işlemlerin tespiti yüksek doğrulukla gerçekleştirilmiştir. Cecchini ve diğerleri ise (2010) 10-K raporlarının yönetici tartışmaları ve analizleri kısmı için metin analizi yaparak hileli finansal raporlama yapan ve yapmayan işletmeler ile iflas riski taşıyan işletmeleri sınıflandırmışlardır. Destekçi Vektör Makineleri (Support Vector Machine-SVM) algoritması ile sınıflandırma

yapılmış ve hileli finansal raporlamada %75, iflas riski taşıyan işletmeleri sınıflandırmada ise %80 sınıflandırma başarısı elde etmişlerdir.

Goel ve diğerleri (2010) faaliyet raporlarını doğal dil işleme araçları kullanarak analiz ederek hileli finansal raporlama yapan işletmeleri tespit etmişlerdir. Ayrıca bag of words yaklaşımı kullanılarak destekçi vektör makineleri ile hileli finansal raporlama yapan ve yapmayan işletmeleri %89,51 doğrulukla sınıflandırmışlardır. Çalışma sonucunda hileli finansal raporlama yapan işletmelerin faaliyet raporlarının dil bilimi açısından da daha farklı olduğu tespit edilmiştir.

Glancy ve Yadav (2011) işletmelerin faaliyet raporlarının yönetim tartışmaları ve analizleri bölümünü kullanarak hileli finansal raporlama yapan işletmeler ile yapmayan işletmeleri sınıflandırmışlardır. Finansal raporlama hileleri tespiti için bilgisayar destekli model önerisi sunan araştırmacılar bu modelin hileli finansal raporlama yapan işletmeleri tespit için kullanılabileceğini göstermişlerdir.

Humpherys ve diğerleri (2011) faaliyet raporlarının yönetici mektupları ve yönetim tartışmaları bölümünü metin madenciliği ile analiz ederek hileli finansal raporları tespit etmeye çalışmışlardır. Naive Bayes ve C.4.5 algoritmaları %67,3'lük sınıflandırma başarısı ile en iyi sınıflandırma performansını gösteren algoritmalar olmuştur.

Goel ve Gangolly (2012) işletmelerin faaliyet raporlarını metin madenciliği kullanarak bu raporların hileli işlemlere ilişkin ipuçları içerip içermediğini araştırmışlardır. Çalışmada hileli finansal raporlama yapan işletmelerin faaliyet raporlarında hileli finansal raporlama yapmayan işletmelerin raporlarından dil bilimi açısından farklılıklar tespit edilmiştir. Hileli işlem yapan işletmelerin faaliyet raporlarının daha fazla edilgen cümle ve belirsizlik ifadeleri içerdiği ve bu raporların daha zor okunabilirliğinin daha zor olduğu çalışmanın bulguları arasındadır. Tespit edilen bu dilbilimsel farklılıklar kelime torbası (bag of words) kullanılarak %89,51 doğrulukla hileli işlemlerin tespiti yapılmıştır.

Gupta ve Gill (2012) finansal tablolarda yer alan metin verileri kullanarak finansal tablo hilesi bulmaya yönelik bir öneri sundukları çalışmada kelime torbası ve destekçi vektör makineleri kullanmışlardır. Zaki ve Theodoulidis (2013) ise SEC dava dosyalarını metin madenciliği ile analiz ettikleri çalışmalarında finansal pazarlarda hileli işlemlerin tespitine yönelik bir model önerisi sunmuşlardır.

Purda ve Skillicorn (2014), Destekçi Vektör Makineleri sınıflandırma algoritmasını kullandıkları çalışmada işletmelerin yıllık ve çeyrek faaliyet raporlarının yönetici tartışmaları ve analizleri metin madenciliği ile analiz etmişlerdir. %82 sınıflandırma başarısı ile hileli ve hileli olmayan raporlar sınıflandırılmıştır.

Hajek ve Henriques (2017) finansal tablo hilelerinin tahmininde kullanılabilir finansal veriler ile faaliyet raporlarının birlikte kullanılabilirliği bir model geliştirmişlerdir. Faaliyet raporlarında geçen negatif kelimelerin sıklığı ile finansal tablo hilesi yapan işletmeler arasında doğrusal bir ilişki tespit edilmiştir. Çalışmada ayrıca birçok makine öğrenmesi algoritması kullanılarak performansları karşılaştırılmıştır. Bayesian Belief Network (BBN) ve Karar Ağacı/Naive Bayes (DTNB) birleşimi hibrit algoritma en yüksek performansı göstermiştir.

3.2.2. İşletme Performansının Tahminine Yönelik Yapılmış Metin Madenciliği Çalışmaları

İşletme performansının tahmininde yapılan çalışmalarda içerik analizi ve okunabilirlik çalışmalarının daha yoğun olduğu görülmektedir.

Abrahamson ve Park (1994), yıllık faaliyet raporlarının yönetici mektupları kısmını bilgisayar destekli içerik analizi ile analiz etmiş ve yöneticilerin işletmenin kötü performansını gizlediği ve bu gizleme işleminin de istenerek yapıldığı sonucuna varmışlardır. Abrahamson ve Amir (1996) ise çalışmada yönetici mektuplarını manuel ve bilgisayar destekli içerik analizi kullanılarak analiz etmişlerdir. Yönetici mektuplarının geçmişe değil, geleceğe yönelik faydalı bilgiler içerdiği tespit edilmiştir.

Baginski ve diğerleri (2004) yönetici tahminlerini veri olarak kullandıkları çalışmada söz konusu bu açıklamaların hisse senedi fiyatı ile ilişkilerini içerik analizi kullanarak incelemişlerdir. Büyük ve yasal düzenlemelerin daha az olduğu endüstrilerde yer alan işletmelerin geleceğe yönelik tahmin açıklamalarını daha fazla yaptıkları ve hisse senedi fiyatlarının bu açıklamalardan etkilendiği tespit edilmiştir.

Kothari ve diğerleri (2009), yaklaşık 100,000 adet yöneticiler, analistler ve gazeteciler tarafından açıklanan raporu içerik analizi ile inceledikleri çalışmada içerik analizinin olumlu olması durumunda sermaye maliyeti, hisse senedi getirisindeki değişiklikler gibi işletme riskini gösteren değişkenlerde azalma tespit edilmiştir.

Clatworthy ve Jones (2003a, 2006b) İngiltere’de işletmelerin faaliyet raporlarında yer alan yönetici mektuplarını inceledikleri çalışmalarda finansal performansı iyi olmayan işletmelerin

daha az sayısal/nicel sonuç kullandıkları, daha çok edilgen cümle kullandıkları ve cümlelerin gelecek odaklı olmadıklarını tespit etmişlerdir. Ayrıca yöneticilerin işletme ile ilgili olumlu durumları kendilerine mâl ettikleri, olumsuz durumlarda ise dış kaynakları suçladıkları tespit edilmiştir.

Schumaker ve Chen (2006) finansal haber makalelerini kullandıkları çalışmalarında, bu belgelerin hisse senedi getirisi tahmininde kullanılabilirliğini incelemişlerdir. Kullanılan metin madenciliği algoritmalarının haber yayınlandıktan 20 dakika sonra hisse senedi getirisi performansının ölçüldüğü çalışmada, hisse senedi getirilerinin doğrusal regresyon modeline kıyasla destekçi vektör makineleri ile daha doğru tahmin edildiği görülmüştür.

Li (2006), faaliyet raporlarının risk duyarlılığı ile işletmenin gelecek dönem hasılatı ve hisse senedi getirileri arasında ters yönlü bir ilişki tespit etmiştir. Faaliyet raporlarında risk ve belirsizlik ifadeleri fazla olan işletmelerin gelecek dönem kazançlarında azalma tespit edilmiştir. Li (2008a) faaliyet raporlarının okunabilirliği ile işletme performansı arasındaki ilişkiyi incelediği çalışmada finansal performansı düşük işletmelerin faaliyet raporlarının daha zor okunduğunu tespit etmiştir. Li tarafından yapılan başka bir çalışma ise (2008b), 10-K belgelerinin geleceğe yönelik ifadeler içerip içermediği ile ilgilidir. Naive Bayes makine öğrenmesi algoritmasının kullanıldığı çalışmada geleceğe yönelik ifadelerin işletmenin gelecek performansı ile pozitif bir ilişkisi olduğu tespit edilmiştir. Ayrıca kurumsal belgeleri analizde makine öğrenmesi algoritmalarının sözlük-bazlı metin madenciliği çalışmalarından daha başarılı sonuçlar elde ettiği görülmüştür.

Tettlock (2007), Tettlock ve diğerleri (2008) tarafından içerik analizinin kullanıldığı bu çalışmalarda “finansal haber içeriklerinden işletme performansı tahmini yapılabilir mi?” sorusuna cevap aranmaktadır. Çalışmalarda, haberlerde geçen olumsuz kelimelerin düşük işletme performansını tahmin edebildiği tespit edilmiştir.

Balakrishnan ve diğerleri (2010) 10-K raporlarının işletme performansı tahmininde kullanılabilirliğini metin sınıflandırma teknikleri kullanarak analiz etmişlerdir. Çalışmada bu raporların işletme performansı tahmininde yatırımcıların kararını etkileyen değerli bilgiler sağladığı tespit edilmiştir.

Davis ve diğerleri (2012) yöneticiler tarafından yapılan gelir açıklama basın bültenlerini metin madenciliği yazılım programlarından DICTION ile analiz ettikleri çalışmalarında 1998-2003 yılları arası 23.000 adet basılı bülteni kullanmışlardır. Çalışmada yöneticilerin basın

bültenlerinde kullandıkları dilin işletmenin gelecek performansı ile ilgili olup olmadığı araştırılmıştır. Çalışma sonucunda yöneticilerin basın bültenlerini işletmenin gelecek performansı hakkında ilgilileri bilgilendirmek amacıyla iletişim aracı olarak kullandıklarının ve ayrıca işletme ilgililerinin de yönetici söylemlerini dikkate aldıkları bulunmuştur.

Campbell ve diğerleri (2014) 10-K belgelerini duygu analizi ile incelemiş ve bu raporların karar alıcılar üzerindeki etkilerini analiz etmişlerdir. Çalışmada rapor içeriği ile verilecek karar arasında pozitif bir ilişki tespit edilmiştir.

Gemar ve Jimenez-Quintero (2015) sosyal medya haberleri ile işletme performansı arasındaki ilişkiyi metin madenciliği ile analiz ettikleri çalışmada aktif kârlılığını kullanmışlardır. Metin madenciliği rekabetçi zekâ (competitive intelligence) geliştirmede bir araç olarak kullanılmıştır. Rekabetçi zekâ kullanarak elde edilen veriler ile otellerin işletme performanslarını geliştirebildikleri tespit edilmiştir.

Wuthrich ve diğerleri (2008) internet haberlerini kullanarak hisse senedi fiyatı tahmini için bir model geliştirmişlerdir. Knn makine öğrenmesi algoritmasının kullanıldığı çalışmada %43,6 doğrulukla hisse senedi fiyatı doğru tahmin edilmiştir.

3.2.3. İflas Tahminine Yönelik Yapılmış Metin Madenciliği Çalışmaları

Shirata ve Sakagami (2008), Japon işletmelerinin faaliyet raporlarını işletmelerin sürekliliği ile ilgili tahminde bulunmak amacıyla metin madenciliğini kullanmışlardır. Çalışmada 20 işletme ele alınmıştır. Çalışma sonucunda yazarlar finansal veriler olmaksızın işletmenin sürekliliğinin tespitini işletmelerin faaliyet raporları üzerinden gerçekleştirmişlerdir.

Shirata ve diğerleri (2011), 1999-2005 yılları arası Tokyo Borsasında faaliyet gösteren 90 adet iflas eden ve 90 adet iflas etmeyen işletmenin faaliyet raporlarının metin madenciliği ile analiz etmişlerdir. Çalışmada “kâr payı” ve “dağıtılmayan kazançlar” ifadeleri ile birlikte kullanılan bazı ifadelerin, iflas eden ve etmeyen işletmeler için anlamlı farklılık gösterdikleri tespit edilmiştir.

3.2.4. Sürdürülebilirlik ve Kurumsal Yönetim Alanında Yapılmış Metin Madenciliği Çalışmaları

Sürdürülebilirlik ve kurumsal yönetim alanında yapılmış metin madenciliği çalışmaları aşağıdaki gibi özetlenebilmektedir.

Barkemeyer ve diğerleri (2009) sürdürülebilirlikte anahtar trendleri (eğilimleri) bulmaya ve değişik ülke ve bölgelerde medya içeriklerinin ve konularının zamanla nasıl değiştiğini

göstermeye yönelik olarak 39 ülkenin 115 ulusal gazetesini metin madenciliği ile analiz etmişlerdir. Çalışma sonucunda özellikle 1990 yılından sonra sürdürülebilirlikle ilgili haber sayısının arttığı bulunmuştur.

Liew ve diğerleri (2014) yaptıkları çalışmada işletmelerin sürdürülebilirlik konusundaki eğilimlerini incelemişlerdir. Çalışmada değerlendirmeye alınan işletmelerin sürdürülebilirlik raporları metin madenciliği kullanılarak analiz edilmiştir. Sürdürülebilirlik ile ilgili terimler işletmelerin sürdürülebilirlik raporlarında geçip geçmemesine göre ağırlıklandırılmıştır. Böylelikle sektörel olarak sürdürülebilirlik trendi oluşturulmaya çalışılmıştır.

Rivera ve diğerleri (2014) tarafından yapılan çalışmada sürdürülebilirlik göstergeleri metin madenciliği kullanılarak bulunmaya çalışılmıştır. Çalışma pilot bir çalışma olup Kaliforniya eyaletinin San Mateo County bölgesinde gerçekleştirilmiştir. Analize o bölgedeki gazete haberleri dâhil edilmiş olup gazete haberleri metin madenciliği kullanılarak sınıflandırılmış ve bölgesel olarak sürdürülebilirlik göstergeleri tespit edilmiştir.

Bala ve diğerleri (2015) Sürdürülebilirlik Trend Analizi olarak adlandırdıkları, doğal dil işleme ve makine öğrenmesinin birleşimini kullanarak sürdürülebilirlik konusundaki gelişmeleri anında izlemeye yönelik bir model geliştirmişlerdir. Analiz sonucunda bulunan veriler dinamik sürdürülebilirlik göstergeleri olarak adlandırılmış ve bu veriler ile işletmenin gelecek kurumsal finansal performansını ilişkilendirmişlerdir. Çalışan katılımını güçlendirme yönünde girişimleri bulunan 12 işletmenin 6 aylık dönemi analize dâhil edilmiştir. Çalışma sonucunda dinamik sürdürülebilirlik göstergeleri ile işletmenin finansal performansı arasında güçlü bir ilişki tespit edilememiştir.

Shahi ve diğerleri (2014), sürdürülebilirlik raporlarını makine öğrenmesi algoritması ile sınıflandırdıkları çalışmada Naive Bayes ve Karar Ağacı algoritmalarının en başarılı sonuçları verdiğini tespit etmişlerdir. Ayrıca çalışmada G3 raporlama çerçevesi kullanılarak 100 adet rapor derecelendirilmiştir.

Zheng ve Zhou (2012) Amerikan Sermaye Piyasası (SEC) belgelerini metin madenciliği ile analiz ettikleri çalışmalarında, kurumsal yönetim ilkelerine uyan firmaları tespit eden bir model sunmuşlardır. ICGA (Intelligent Corporate Analysis and Rating System) adını verdikleri model ile işletmelerin kurumsal yönetim nitelikleri otomatik olarak belirlenmektedir.

Pencle ve Mălăescu (2016) işletmelerin ilk halka arz izahnamelerinden kurumsal sosyal sorumluluklarına ilişkin tahminde bulunmak üzere bilgisayar destekli metin madenciliği

(Computer-Aided Text Analysis-CATA) kullanarak içerik analizi yapmışlardır. Araştırmacıların geliştirdiği sözlük birçok farklı belgeden işletmelerin kurumsal sosyal sorumluluğuna ilişkin yapılacak tahminlerde kullanılabilirlerdir.

3.2.5. Muhasebe ve Finansal Analizde Metin Madenciliği Yapılmış Diğer Çalışmalar

Bu bölümde muhasebe ve finansal analizde yapılmış farklı amaçlara yönelik çalışmalara yer verilmiştir.

Soper ve Dolphin (1964), Smith ve Smith (1971) faaliyet raporlarının okunabilirliğini incelemişlerdir. Soper ve Dolphin 1948 ve 1961 yıllarını karşılaştırdıkları çalışmalarında okunabilirlikte bir ilerleme olmadığını tespit etmişlerdir. Smith ve Smith ise Flesch ve Dale-Chall okunabilirlik modellerini karşılaştırdıkları çalışmalarında Flesch modelinin daha başarılı olduğu sonucuna ulaşmışlardır.

Frazier ve diğerleri (1984) 74 adet faaliyet raporunu içerik analiz ile analiz etmişlerdir. WORDS içerik analizi ile finansal tablolar dışında açıklanan muhasebe verilerinden faydalı bilgi çıkarılabildiği gösterilmiştir.

Antweiler ve Frank (2004), internet üzerinden hisse senedi mesaj panosunda yer alan bilgilerin piyasayı etkileyip etkilemediğini bilgisayar destekli dilbilimi ile analiz ettikleri çalışmalarında 45 işletme hakkında çıkan 1,5 milyondan fazla mesajı incelemişlerdir. Çalışmada bu panolarda yer alan haberler ile hisse senedi getirileri arasında istatistiksel olarak anlamlı bir ilişki tespit edilmiştir. Ancak ekonomik olarak çok önemli bir fark tespit edilememiştir.

Chakraborty ve diğerleri (2014) muhasebe ve muhasebe bilgi sistemlerini otomatik sınıflandırabilmek amacıyla metin madenciliğini kullanmışlardır. Çeşitli makine öğrenmesi algoritmalarının kullanıldığı çalışmada Karar Ağacı algoritması ile % 82,27'lik sınıflandırma başarısı sağlanmıştır.

Aureli (2017) yaptığı çalışmada 4 uluslararası kuruluşun sürdürülebilirlik raporlarını metin madenciliği ve içerik analizi ile incelemiştir. Çalışmanın amacı endüstride yaşanan kriz sonrası işletmelerin yaptığı açıklamalarda meydana gelen değişimlerin incelenmesidir. Çalışmada ayrıca her iki yöntem de karşılaştırmalı olarak verilmiştir. Çalışma sonucunda her iki yöntemin birbirinden çok farklı olmadığı tespit edilmiştir.

Matthies ve Coners (2015) 114 işletmenin faaliyet raporlarını bilgisayar destekli metin madenciliği ile analiz ederek finansal ekonomik/krizlerin işletme raporlaması üzerindeki etkisini

incelemişlerdir. Çalışmanın asıl amacı ampirik bir uygulama olmayıp bilgisayar destekli metin madenciliği uygulamasını tanıtmaktır.

Kloptchenko ve diğerleri (2004a) 2000-2001 yılları için Motorola, Ericsson ve Nokia'nın çeyrek raporlarını analiz ettikleri çalışmada bu raporlarda en sık geçen önemli kelimeleri ortaya çıkarmak için bilgisayar destekli metin madenciliği tekniklerini kullanmışlardır. Çalışmada önemli bilgiler içeren faaliyet raporlarının tamamının okunmadan metin madenciliği ile analizi ile faydalı bilgiler çıkarılabileceği tespit edilmiştir. Kloptchenko ve diğerleri (2004b) tarafından yapılan bir diğer çalışmada ise faaliyet raporları ile finansal tablolar birlikte analiz edilmiştir. Nicel veriler için Self Organizing Maps tekniği, metin veriler için ise metin kümeleme analizi uygulanmıştır. İşletme performansında kayda değer bir değişiklik olmadan önce finansal raporların yazım sitilinde değişiklikler olduğu tespit edilmiştir. Eğer işlemenin durumu gelecek dönemde zayıflayacaksa rapor içeriklerinin de kötümser olduğu bulunmuştur.

Heidari ve Felden (2015), çalışmalarında finansal tablo dipnotlarını metin madenciliği ile analiz etmişlerdir. Çalışma gelir vergisi dipnotu üzerinde yapılmış olup amacı metin verilerinden faydalı bilgi çıkarımında metin madenciliği uygulamasının gösterilmesidir. Sınıflandırma algoritmalarının uygulandığı çalışmada en başarılı algoritma Naive Bayes olmuştur.

Kamaruddin ve diğerleri (2015) banka finansal raporlarını metin madenciliği kullanarak analiz etmiş ve cümle bazında sapmaları tespit etmeye çalışmışlardır. Söz konusu çalışma finansal raporlar üzerinde yapıldığı için geliştirilen model hem sayısal hem de metin veriler için uygulanabilir bir model olmuştur.

Rich ve diğerleri (2016), faaliyet raporlarının yönetim tartışmaları ve analizleri bölümünü finansal raporlama kalitesinin belirlenmesi amacıyla metin madenciliği ile analiz etmişlerdir. Çalışmada yönetim tartışmaları ve analizlerinin pozitif cümleler içermesinin finansal raporlama kalitesinin bir göstergesi olduğu tespit edilmiştir.

Liu ve Moffitt metin madenciliğini SEC tarafından gönderilen mektupların yoğunluğunu ölçmede kullanmışlardır. Yoğunluk ölçütü olarak güçlü/zayıf gereklilik fiillerinin kullanılmıştır. Yönetici mektuplarında geçen güçlü kelimelerin oranının fazlalığı ile bu ifadelerin 10-K belgelerinde geçmesi arasında pozitif bir ilişki tespit edilmiştir.

Demers ve Vega (2008) işletmelerin kazanç basın açıklamalarında yer alan metin verileri DICTION metin madenciliği ile analiz ettikleri çalışmalarında yöneticilerin kullandıkları net

iyimser ifadelerin kısa dönemli hisse senedi fiyatını etkilediğini bulmuşlardır. Yazarlar söz konusu metin verileri “yumuşak veri” olarak ifade etmişlerdir. Çalışmalarında yumuşak veri kullanan diğer yazarlar ise Tsai ve Wang (2017), metin verileri risk analizi açısından incelemişlerdir. Kelime torbası (bag of words) ve kelimelerin duygu analizinin yapıldığı çalışmada işletmenin gelecek risk analizlerinde metin verilerin kullanımının anlamlı sonuçlar verdiği tespit edilmiştir.

Loughran ve McDonald (2011) 1994-2008 yılları arası yayınlanmış 10-K belgelerini analiz ederek muhasebe ve finans alanında metin madenciliği ile yapılacak çalışmalarda referans olabilecek bir sözlük oluşturmuşlardır. Ayrıca oluşturulan diğer sözlüklerin içerdiği kelimelerin çoğunluğunun muhasebe ve finans alanında farklı anlam içerdiğini, dolayısıyla kullanılmaması gerektiğini belirtmişlerdir.

4. SONUÇ

Muhasebe, işletme içi ile işletme dışı ilgililer arasındaki iletişimi sağlayan önemli bir işleve sahiptir. Muhasebe bu işlevini temel olarak genel amaçlı finansal tablolar aracılığıyla yerine getirmektedir. Ancak değişen çevre şartları, teknoloji gibi etmenler, muhasebenin varlık tanımındaki değişme, menfaat sahiplerinin ihtiyaçları gibi üzerine oturduğu kavramsal temel olguları derinden etkilemiştir. Teknoloji ile birlikte işletmelerin varlık yapısında değişimler meydana gelmiş, maddi varlıklardan ziyade maddi olmayan varlıklar ağırlık kazanmıştır. Toplumun çevresel ve sosyal olayları yakından takip etmesi ve bu konulara gelecek nesiller için daha fazla önem verilmesi sonucu, menfaat sahiplerini toplumun genelini de içine alarak genişletmiştir. Temelde bu iki neden geleneksel finansal raporlama çerçevesini son 20 yılın tartışılan konularından biri haline getirmiştir. Bu alanda gerek uluslararası, gerek ulusal düzenlemelerle, işletmelerin daha fazla veri sunmasını amaçlayan finansal raporlama çerçeveleri ortaya çıkmaya başlamıştır. İşletmeler genel amaçlı finansal tabloları aracılığıyla sunamadıkları ancak işletmeyi ilgilendiren birçok bilgiyi faaliyet raporları, sürdürülebilirlik raporları, internet siteleri gibi çeşitli araçlar ile menfaat sahipleriyle paylaşmaya başlamışlardır. Finansal raporlamada gelinen son nokta ise üzerinde halen çalışmaların devam ettiği entegre raporlamadır.

Muhasebe, genişleyen raporlama çerçeveleri ile işletme ve menfaat sahipleri arasındaki bilgi asimetrisini azaltmaya çalışmaktadır. Ancak sunulan bu kadar fazla veri menfaat sahipleri

tarafından faydalı bilgiye dönüştürülememektedir. Çünkü söz konusu rapor içerikleri incelendiğinde sayısal verilerin dışında bu verilerin metin ve resim gibi verileri de içerdiği görülmektedir. Bu verilerin miktar olarak çokluğunun yanı sıra, bu veriler yapısal olmayan verilerdir. Söz konusu bu verilerin analizi yeni yöntemler gerektirmektedir. Bu çalışmada bu yöntemlerden metin madenciliği ele alınmıştır.

Metin madenciliği yarı yapısal ve yapısal olmayan verilerin analizinde kullanılan bir büyük veri analiz yöntemidir. Metin madenciliği, genişleyen finansal raporlama çerçevesi ile birlikte özellikle 1990'lı yıllardan itibaren muhasebe alanında da kullanılmaktadır. Çalışmada, yapılan literatür taraması sonucunda metin madenciliğinin hileli işlemlerin tespiti, işletme performansının tahmini gibi alanlarda sıklıkla kullanıldığı görülmektedir. Yapılan çalışmalar, muhasebe ve finansal analizde finansal tablolarla sunulamayan verilerin de analiz edilerek finansal tablo analizi ile birlikte menfaat sahiplerine daha faydalı bilgi sağlandığını göstermektedir.

KAYNAKÇA

- Abrahamson, E. ve C. Park. 1994. "Concealment of Negative Organizational Outcomes: An Agency Theory Perspective", *Academy of Management Journal*, 37(5).
- Abrahamson, E. ve E. Amir. 1996. "The Information Content of the President's Letter to Shareholders", *Journal of Business Finance and Accounting*, 23(8).
- Alwert, K., M. Bornemann ve M. Will. 2009. "Does Intellectual Capital Reporting Matter to Financial Analysts?", *Journal of Intellectual Capital*, 10(3).
- Amir, E., ve B. Lev. 1996. "Value-Relevance of Nonfinancial Information: The Wireless Communications Industry", *Journal of Accounting and Economics*, 22.
- Antweiler, W. ve M.Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards", *The Journal of Finance*, LIX(3).
- Atan, S. 1996. "Metin Madenciliği ile Sentiment Analizi ve Borsa İstanbul Uygulaması", Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Basılmamış Doktora Tezi, Ankara.
- Atan, S. 2016. "Veri, Büyük Veri ve İşletmecilik", Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 19(35).
- Aureli, S. 2017. "A Comparison of Content Analysis Usage and Text Mining in CSR Corporate Disclosure", *The International Journal of Digital Accounting Research*, 17.
- Bala, G., B., H., J.P. Hawley ve Y.J. Lee. 2015. "Tracking "Real Time Corporate Sustainability Signals Using Cognitive Computing", *Journal of Applied Corporate Finance*, 27(2).
- Balakrishnan, R., X.Y. Qiu ve P. Srinivasan. 2010. "On The Predictive Ability of Narrative Disclosures in Annual Reports", *European Journal of Operational Research*, 202(3).

- Baginski, S.P., J.M. Hassell ve M.D. Kibrough. 2004. "Why Do Managemrs Explain Their Earnings Forecasts?", *Journal of Accounting Research*, 42(1).
- Barkemeyer, R., F. Figge, D. Holt ve T. Hahn. 2009. "What the Papers Say: Trends in Sustainability", *The Journal of Corporate Citizenship*, 33.
- Bogaerd, M.V.D. ve W. Aerts. 2011. "Applying Machine Learning in Accounting Research", *Expert Systemes with Applications*, 38.
- Cecchini, M. 2005. "Quantifying the Risk of Financial Events Using Kernel Methods and Information Retrieval", University of Florida, Doctorate Dissertation.
- Cecchini, M., H. Aytuğ, G. Koehler ve P. Pathak. 2010. "Making Words Work: Using Financial Text as a Predictor of Financial Events", *Decision Support Systems*, 50(2).
- Chakraborty, V., V. Chiu ve M. Vasarhelyi. 2014. "Automatic Classification of Accounting Literatüre", *International Journal of Accounting Information Systems*, 15(2).
- Clatworthy, M.A. ve M.J. Jones. 2003a. "Financial Reporting of Good News and Bad News: Evidence from Accounting Narratives", *Accounting, and Business Research*, 33(3).
- Clatworthy, M.A. ve M.J. Jones. 2006b. "Differential Patterns of Textual Characteristics and Company Performance in the Chairman's Statement", *Accounting, Auditing and Accountability Journal*, 19(4).
- Davis, A.K., J.M. Piger ve L.M. Sedor. 2012. "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language", *Contemporary Accounting Resarch*, 29(3).
- Demers, E. ve C. Vega. 2008. "Soft Information in Earnings Announcements:News or Noise?", Board of Governors of the Federal Reserve System International Finance Discussion Papers, No:951.
- Dolgun, M.Ö., T.G. Özdemir ve D. Oğuz. 2009. Veri Madenciliğinde Yapısal Olmayan Verinin Analizi:Metin ve Web Madenciliği", *İstatistikçiler Dergisi*, 2.
- Eccles, R.G., M. Krzus ve D. Tapscott. 2010. One Report, John Wiley&Sons Incorporated.
- Fisher, I.E., R.G. Margaret, G. Sunita ve T. Kinsun. 2010. "The Role of Text Analytics and Information Retrieval in the Accounting Domain", *Journal of Emerging Technologies in Accounting*, 7.
- Fisher, I.E., M.R. Garnsey ve M.E. Hughes. 2016. "Natural Language Processing in Accounting, Auditing and Finance:A Synthesis of the Literature with a Roadmap for Future Research", *Intelligent Systems in Accounting, Finance and Management*, 23.
- Fraizer, K.B, R.W. Ingram ve B.M. Tennyson. 1984. "A Methodology for the Analysis of Narrative Accounting Disclosures", *Journal of Accounting Research*, 22(1).
- Francis, J. ve K. Schipper. 1999. "Have Financial Statements Lost Their Relevance?", *Journal of Accounting Research*, 37(2).
- Gaikwad, S.V, A. Chaugule ve P. Pramod. 2014. "Text Mining Methods and Techniques", *International Journal of Computer Applications*, 85(17).

- Gao, L., E. Chang ve S. Han. 2007. "Powerful Tool to Expand Business Intelligence: Text Mining", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 1(8).
- Gemar, G. ve J.A. Jimenez-Quintero. 2015. "Text Mining Social Media for Competitive Analysis", *Tourism ve Management Studies*, 11(1).
- Glancy, F.H. ve S.B. Yadav. 2011. "A Computational Model for Financial Reporting Fraud Detection", *Decision Support Systems*, 50(3).
- Goel, S., J. Gangolly, S.R. Faerman ve O. Uzuner. 2010. "Can Linguistic Predictors Detect Fraudulent Financial Filings?", *Journal of Emerging Technologies in Accounting*, 7.
- Goel, S. ve J. Gangolly. 2012. "Beyond the Numbers: Mining the Annual Reports for Hidden Cues Indicative of Financial Statement Fraud", *Intelligent Systems in Accounting, Finance and Management*, 19.
- Gupta, R. ve N.S. Gill. 2012. "Financial Statement Fraud Detection Using Text Mining", *International Journal of Advanced Computer Science and Applications*, 3(12).
- Gürsakar, N. 2014. *Büyük Veri, Dora Yayınları, Bursa*.
- Hajek, P. ve R. Henriques. 2017. "Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud – A Comparative Study of Machine Learning Methods", *Knowledge- Based Systems*, 128.
- Heidari, M. ve C. Felden. 2015. "Financial Footnote Analysis: Developing a Text Mining Approach", *Proceedings of the International Conference on Data Mining (DMIN)*; Athens: 10-16. Athens: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing.
- Holton, C. 2009. "Identifying Disgruntled Employee Systems Fraud Risk Through Text Mining: A Simple Solution for a Multi-Billion Dollar Problem", *Decision Support Systems*, 46(4).
- Humpherys, S.L., K.C. Moffitt, M.B. Burns, J.K. Burgoon ve W.F. Felix. 2011. "Identification of Fraudulent Financial Statements Using Linguistic Credibility Analysis", *Decision Support Systems*, 50(3).
- https://tr.wikipedia.org/wiki/Do%C4%9Fal_dil_i%C5%9Fleme, (Erişim Tarihi: 15.11.2016).
- <http://www.bb.itu.edu.tr> (Erişim Tarihi: 11.11.2016).
- Jeong, H., Y. Ko ve J. Seo. 2016. "How to Improve Text Summarization and Classification by Mutual Cooperation on an Integrated Framework", *Expert Systems With Applications*, 60.
- Jusoh, S. ve H.M. Alfawareh. 2012. "Techniques, Applications and Challenging Issue in Text Mining", *International Journal of Computer Sciences Issues*, 9(6).
- Kamaruddin, S.S., A.A. Bakar, A.R. Hamdan, F.M. Nor, M.Z.A. Nazri, Z.A. Othman ve Hussein, G.S. 2015. "A Text Mining System for Deviation Detection in Financial Documents", *Intelligent Data Analysis*, 19.
- Kloptchenko, A., C. Magnusson, B. Back, A. Visa ve H. Vanharanta. 2004a. "Mining Textual Contents of Financial Reports", *The International Journal of Digital Accounting Research*, 4(7).

- Kloptchenko, A., T. Eklund, J. Karlsson, B. Back, H. Vanharanta ve A. Visa. 2004b. "Combining Data and Text Mining Techniques for Analysing Financial Reports", *Intelligent Systems in Accounting, Finance and Management*, 12.
- Kothari, S.P., X. Li ve J.E. Short. 2009. "The Effect of Disclosures by Management, Analysts, and Business Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Case Study Using Content Analysis", *The Accounting Review*, 84(5).
- Kumar, B.S. ve V. Ravi. 2016. "A Survey of the Applications of Text Mining in Financial Domain", *Knowledge Based Systems*, 114.
- Lev, B. 1989. "On the Usefulness of Earnings and Earnings Research: Lessons and Directions from Two Decades of Empirical Research", *Journal of Accounting Research*, 27 (Supplement).
- Lev, B. 2003. "Remarks on the Measurement, Valuation, and Reporting of Intangible Assets", *Economy policy Review*, 9(3).
- Lev, B. ve P. Zarowin. 1999. "The Boundaries of Financial Reporting and How to Extend Them", *Journal of Accounting Research*, 37(2).
- Lev, B. ve F. Gu. 2016. *The End of Accounting and the Path Forward for Investors and Managers*, John Wiley and Sons Inc.
- Li, F. 2006. "Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?", Unpublished Working Paper, University of Michigan, Available at SSRN: <https://ssrn.com/abstract=898181> or <http://dx.doi.org/10.2139/ssrn.898181>
- Li, F. 2008a. "Annual Report Readability, Current Earnings and Earnings Persistence", *Journal of Accounting and Economics*, 45(2-3).
- Li, F. 2008b. "The Determinants and Information Content of the Forward-Looking Statements in Corporate Filings- A Naive Bayesian Machine Learning Approach", *Journal of Accounting Research*, 48(5).
- Li, F. 2010. "Textual analysis of Corporate Disclosures: A Survey of the Literature", *Journal of Accounting Literature*, 29.
- Liew, W.T., A. Adhitya ve R. Srinivasan. 2014. "Sustainability Trends in the Process Industries: A Text Mining-Based Analysis", *Computers in Industry*, 65.
- Liu, Y. ve K.C. Moffitt. 2016. "Text Mining to Uncover the Intensity of SEC Comment Letters and Its Association with the Probability of 10-K Restatement", *Journal of Emerging Technologies in Accounting*, 13(1).
- Loughran, T. ve B. Mc Donald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks", *The Journal of Finance*, LXVI(1).
- Matthies, B. ve A. Coners. 2015. "Computer-Aided Text Analysis of Corporate Disclosures- Demonstration and Evaluation of Two Approaches", *The International Journal of Digital Accounting Research*, 15.

- Menon, R., L.H. Tong, S. Sathiyakeerthi, A. Brombacher ve C. Leong, C. 2004. “The Needs and Benefits of Applying Textual Data Mining Within the Product Development Process”, *Quality and Reliability Engineering International*, 20.
- Miner, G., D. Delen, A. Fast, T. Hill, J. Elder ve B. Nisbet. 2012. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, Elsevier, USA
- Oğuzlar, A. 2011. *Temel Metin Madenciliği*, Dora Yayınları, Bursa
- Pencle, N. ve I. Mălăescu. 2016. “What’s in the Words? Development and Validation of a Multidimensional Dictionary for CSR and Application Using Prospectuses”, *Journal Emerging Technologies in Accounting*, 13(2).
- Purda, L. ve D. Skillicorn. 2014. “Accounting Variables, Deception, and A Bag of Words:Assesing the Tools of Fraud Detection”, *Contemporary Accounting Research*, 32(3).
- Rich, K.T., B.L. Roberts, ve J.X. Zhang. 2016. “Linguistic Tone of Municipal Management Discussion and Analysis Disclosures and Future Financial Reporting Delays”, *Journal of Emerging Technologies in Accounting*, 13(2).
- Rimerman, T.W. 1990. “The Changing Significance of Financial Statements”, *Journal of Accountancy*, 169(4).
- Rivera, S.J., B.S. Minsker, D.B. Work ve D. Roth. 2014. “A Text Mining Framework for Advancing Sustainability Indicators”, *Environmental Modelling&Software*, 62.
- Schumaker, R.P. ve H. Chen. 2006. “Textual Analysis of Stock Market Prediction Using Financial News Articles”, *Proceedings of the Twelfth Americas Conference on Information Systems*, Acapulco, Mexico.
- Shahi, A.M., B. Issac ve J.R. Modapothala. 2014. “Automatic Analysis of Corporate Sustainability Reports and Intelligent Scoring”, *International Journal of Computational Intelligence and Applications*, 13(1).
- Shirata, C.Y. ve M. Sakagami. 2008. “An Analysis of the “Going Concern Assumption”: Text Mining from Japanese Financial Reports”, *Journal of Emerging Technologies in Accounting*, 5.
- Shirata, C.Y., H. Takeuchi, S. Ogino ve H. Watanebe. 2011. “Extracting Key Phrases as Predictors of Corporate Bankruptcy:Empirical Analysis of Annual Reports by Text Mining”, *Journal of Emerging Technologies in Accounting*, 8.
- Smith, J.E. ve N.P. Smith. 1971. “Readability:A Measure of the Performance of the Communication Function of Financial Reporting”, *The Accounting Review*, 46(3).
- Soper, F.J. ve R. Dolphin. 1964. “Readability and Corporate Annual Reports”, *The Accounting Review*, 39(2).
- Şeker, Ş.E. 2015. “Metin Madenciliği (Text Mining)”, *YBS Ansiklopedi*, 2(3).
- Tettlock, P.C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”, *The Journal of Finance*, LXII(3).
- Tettlock, P.C., M. Saar-Tsechansky ve S. Macskassy. 2008. “More Than Words:Quantifying Language to Measure Firms’ Fundamentals”, *The Journal of Finance*, LXIII(3).

- Tsai, M.F., ve C.J. Wang. 2017. "On the Risk Prediction and Analysis of Soft Information in Finance Reports", *European Journal of Operational Research*, 217.
- Tunalı, V. 2011. "Metin Madenciliği İçin İyileştirilmiş Bir Kümeleme Yapısının Tasarımı ve Uygulaması", Marmara Üniversitesi, Fen bilimleri Enstitüsü, Doktora Tezi.
- Wuthrich B., V Cho., S. Leung, D. Permunetilleke, K. Sankaran, J. Zhang ve W. Lam. 2008. "Daily Stock Market Forecast from Textual Web Data", *Systems, Man, and Cybernetics*, 1998 IEEE International Conference.
- Zaki, M. ve B. Theodoulidis. 2013. "Analyzing Financial Fraud Cases Using a Linguistics-Based Text Mining Approach", Available at SSRN: <https://ssrn.com/abstract=2353834> or <http://dx.doi.org/10.2139/ssrn.2353834>
- Zheng, Y. ve H. Zhou. 2012. "An Intelligent Text Mining System Applied to SEC Documents", *IEEE/ACIS 11th International Conference on Computer and Information Science*.
- TMS 1 Finansal Tabloların Sunuluşu Standardı Türkiye Muhasebe ve Finansal Raporlama Standartları Kavramsal Çerçeve 1 Seri Nolu Muhasebe Sistemi Uygulama Genel Tebliği