

MUHASEBE ALANINA GÜNCEL YAKLAŞIMLAR: METİN MADENCİLİĞİ*

Dr. Öğr. Üyesi Hüseyin ÖZYİĞİT^a

Teorik Araştırma
(Theoretical Research)

*Muhasebe ve Vergi
Uygulamaları Dergisi*
Kasım 2022; 15 (3): 637-663

ÖZ

Metin madenciliği; bilgi bilimleri, dilbilim, bilgisayar bilimleri ve veri bilimleri gibi farklı alanlardan kavram ve teknikleri içeren çok disiplinli bir bilgi dalıdır. Kuruluşların kâğıt verilerden elektronik belgelere ve dijital kayıtlara geçmesiyle birlikte, iş süreçlerinin hızla dijitalleşmesi metin madenciliğine olan ilgiyi artırmıştır. Muhasebe alanındaki verilerin giderek büyümesinden dolayı metin madenciliği teknolojisi bu alan için önemli bir araştırma konusu olmuştur. Bu çalışmanın amacı; muhasebe alanında, metin madenciliğinin kullanımına yönelik bilgiler verilerek gelecekte bu teknolojinin kuruluşlara ve kişilere etkisini özlü bir şekilde ortaya koymaktır. Sonuç olarak metin madenciliği teknolojisinin muhasebe alanında kullanımı; muhasebe otomasyonu, denetim otomasyonu, vergi otomasyonu ve iş danışmanlığı otomasyonu şeklinde ele alınmış ve açıklamalar yapılmıştır. Ayrıca yapay zekâ ve makine öğrenmesi yaklaşımlarıyla birleştirilen metin madenciliğinin, işlemleri çok daha fazla otomatikleştirmesinden dolayı kuruluşlara ve muhasebe meslek mensuplarına önemli fırsatlar sunacağı ön görülmektedir.

Anahtar Sözcükler: Metin Madenciliği, Muhasebe, Teknoloji.

JEL Kodları: M40, M42.

APA Stili Kaynak Gösterimi:

Özyiğit, H. (2022). Muhasebe Alanına Güncel Yaklaşımlar: Metin Madenciliği. *Muhasebe ve Vergi Uygulamaları Dergisi*. 15 (3), 637-663.

* Makalenin gönderim tarihi: 16.04.2022; Kabul tarihi: 21.08.2022, iThenticate benzerlik oranı %2
^a Erzincan Binali Yıldırım Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, huseyinozyigit@erzincan.edu.tr ORCID: [0000-0002-0632-7931](https://orcid.org/0000-0002-0632-7931).

CURRENT APPROACHES TO ACCOUNTING: TEXT MINING

ABSTRACT

Text mining; It is a multidisciplinary branch of knowledge that includes concepts and techniques from different fields such as information sciences, linguistics, computer science and data science. With the transition of organizations from paper data to electronic documents and digital records, the rapid digitization of business processes has increased the interest in text mining. Due to the growing data in the field of accounting, text mining technology has become an important research topic for this field. The aim of this study; In the field of accounting, by giving information on the use of text mining, it is to reveal the effect of this technology on organizations and individuals in the future in a concise way. As a result, the use of text mining technology in the field of accounting; accounting automation, audit automation, tax automation and business consultancy automation. In addition, it is predicted that text mining combined with artificial intelligence and machine learning approaches will offer significant opportunities to organizations and accounting professionals, as it automates processes much more.

Keywords: Text Mining, Accounting, Technology.

JEL Codes: M40, M42.

EXTENDED SUMMARY

Introduction

Text mining; It is the process of extracting information from various text sources (Word documents, PDF files, social media posts, emails, websites, articles, XML files, etc.) to discover patterns, trends and themes. The texts found in these sources are not generally structured, meaning they are not in a format that can be predefined and analyzed by software such as IDEA or ACL. Text mining takes place in two stages. These stages are; using data mining techniques to structure text data sources and extract relevant information. Text mining is used not only in accounting, but also in areas containing large amounts of textual data. For example, auditors review contracts, invoices, legal letters, conference calls, and news articles. Professor Aldhizer of Wake Forest University suggests that forensic and auditing applications are considering using text analytics for high-risk tasks. Aldhizer states that text analytics can be used for concept extraction to identify offensive words from social media posts and emails. It is also used to generate appropriate reporting for text mining, rental and revenue transactions. Recently, text mining has been gaining significant momentum due to the exponential growth of content on the Internet with web-enabled applications and social networks with unstructured texts. As organizations move from paper to electronic documents and other digital records, the rapid digitization of business processes is increasing interest in text mining. Finally, current technological developments in cloud computing, artificial intelligence, internet of things and big data analytics have a very significant impact on the amount of data available for text mining. As a process of

discovering information from unstructured texts, text mining requires natural language interpretation. Therefore, natural language processing, which is the process of interpreting a computer program or natural language such as text, is one of the most important methodologies that enables text mining to extract actionable information from texts effectively and efficiently. This study, which aims to provide information on the use of text mining technology in the field of accounting, primarily focuses on the definition of text mining, the methods used in text mining, and the advantages and disadvantages of text mining. Then, explanations about the use of text mining in the field of accounting were made in the form of accounting automation, tax automation, audit automation and business consultancy automation. Finally, in parallel with the data obtained from the study, the effect of text mining on the field of accounting was evaluated in the conclusion part.

Literature on Research

In this study, it is aimed to give information about the use of text mining technology in the field of accounting. In the literature review conducted for this purpose, it has been determined that text mining has just begun to be used in the field of accounting, mostly from foreign sources. In this context, it is thought that this study will make a significant contribution to the literature in terms of text mining being a current technology area and associating it with accounting. Text mining in early research; It is expressed as the process of discovering knowledge from comprehensive collections of unstructured textual data (Feldman & Dagan, 1995). Professor Marti Hearst from the University of California, in his article on text mining (Hearst, 2003), defines text mining as the computer discovery of new previously unknown information by automatically extracting information from different written sources. Professor Catherine Blake of the University of Illinois at Urbana-Champaign describes text mining as the process of identifying new, interesting and understandable patterns from a collection of texts (Blake, 2011). text mining for Professor Stephane Tuffery of the University of Rennes in France; It is a set of techniques and methods used to automatically process large amounts of text data in the form of computer files, extract and structure content and themes, rapid analysis, discover hidden data, and make automated decisions (Tuffery, 2011). A 2019 study (Lewis & Young, 2019) reviewed articles using text analysis methods published in major accounting and finance journals during the 2010-2018 period. In the research; - the accounting and finance profession is slow to adopt natural language processing, - accounting and finance professionals prefer simple approaches such as dictionaries that measure certain attributes such as keyword searches, word counts and readability scores, - at the document level, machine learning classifiers and tools for determining subject structure Determinations were made that there was a small study

group applying certain natural language processing techniques, including statistical methods. Text mining is also called text analytics. These terms are often used interchangeably in the literature and in practice. However, there is a slight difference between the two words. The difference is that text analytics focuses more on the visualization method (graphs, reports) using the results from the analysis performed by text mining.

Method of The Research

In the study, the data were obtained from the examination of text mining applications in the field of accounting, the research of periodicals and resources on the internet, and the evaluation of domestic and foreign literature. Since text mining is a very new and up-to-date technology in the field of accounting, this study has been discussed from a theoretical perspective.

Findings of The Research

In the study; In the field of accounting, it is aimed to reveal the effect of this technology on organizations and individuals in the future by giving information about the use of text mining. Although text mining has made its way into the field of accounting, the use of text mining among professional accountants is a very new process. Although “Big Four (Deloitte, PwC, Ernst&Young, KPMG)” companies are notable early adopters of text mining technologies, we can say that other accounting institutions have started to adopt or try these technologies as well. Use of text mining technology in accounting; accounting automation, audit automation, tax automation and business consultancy automation. As a result of the investigations, the following determinations were made regarding these automations:

- **Audit Automation:** Some accounting and auditing firms state that they use text mining software to automate contract reviews for audit or compliance assessment. Automated contract reviews; It is carried out as part of field-specific audits such as fraud detection, internal audits, information technology audits, which are within the scope of annual financial audits of organizations.
- **Accounting Automation:** accounting applications; information extraction is integrated into machine learning software combined with optical character recognition technology and robotic process automation solutions. These applications automatically record and process transactions (such as accounts payable invoices) or make recommendations for journal entries based on extracted data, a knowledge base, and patterns learned by the system from previous transactions. Text mining applications are also used in the supply chain to manage vendor contract compliance.

- **Tax Automation:** Text mining powered by machine learning and artificial intelligence has a significant impact on tax automation. The use of these technologies; It will continue to evolve and evolve to further automate tax preparation and tax planning tasks with the advent of question-answering systems and potentially virtual assistants that empower people to handle customer-specific tax questions. Text mining; It is also used for tax audits, taxation (income tax, sales tax, property tax, international taxation and others) and analysis of tax cases.
- **Business Consulting Automation:** Professional accountants; Apart from financial accounting, auditing and tax, he also serves in various business advisory roles. In general, many accountants; He works in functions such as management accounting, financial planning and analysis, internal auditing, compliance, forensic accounting or information systems auditing and makes recommendations. These functions have been significantly influenced by artificial intelligence and machine learning technologies, which include text mining of both textual and non-textual data.

The use of text mining in the field of accounting is very current. In the adoption of text mining by organizations and related professionals;

- Improvements in data access,
- Increased collaboration between researchers from various disciplines and
- It is thought that factors such as big data analytics, machine learning and increased availability of artificial intelligence resources will contribute.

Conclusion

Combined with artificial intelligence and machine learning approaches, text mining offers significant opportunities to automate the work of accountants and auditors much more than they do today. While the use of these technologies is only developing in the accounting profession, they are expected to significantly change the way accountants and auditors work. These changes; more robotization of accounting processes (which means more time spent analyzing data instead of collecting and recording data while providing significant efficiency gains), analyzing more information (meaning more accuracy, better insights and better decision making), internal more proactive and continuous monitoring of controls and audits (meaning early detection of anomalies or risks and improved business performance). The study helps answer questions about what text mining is and how to use it in accounting. In addition, in order to provide detailed information to organizations and individuals, explanations were made on

accounting automation, audit automation, tax automation and business consultancy automation, and text mining application examples in the field of accounting were given. Considering the role that text mining technology will play in the field of accounting in the coming years, it is thought that this study will provide significant benefits to organizations and individuals. Finally, it is expected that the use of natural language processing technology for text mining in financial reporting processes will be on the agenda of researchers in the future. The most important factor that will ensure the development and progress of text mining in the field of accounting is that inter-sectoral and interdisciplinary cooperation should be established on correct and applicable foundations. It is foreseen that this study can be used in different fields by setting an example for similar or more comprehensive studies.

1. GİRİŞ

Kuruluşların kâğıttan elektronik belgelere ve diğer dijital kayıtlara geçmesiyle birlikte, iş süreçlerinin hızla dijitalleşmesi metin madenciliğine olan ilgiyi artırmaktadır. Son olarak yapay zekâ, bulut bilişim, nesnelerin interneti ve büyük veri analitiğindeki güncel teknolojik gelişmeler, metin madenciliği için mevcut veri miktarı üzerinde çok önemli bir etkiye sahiptir. Metin madenciliği sadece muhasebede alanında değil aynı zamanda büyük miktarda metinsel veri içeren alanlar içinde kullanılmaktadır. Denetçilerin sözleşmeleri, faturaları, yasal mektupları, konferans görüşmelerini ve haber makalelerini incelemesi bu duruma örnek olarak verilebilir. Metin madenciliği; bilgi bilimleri, dilbilim, bilgisayar bilimleri ve veri bilimleri (örneğin istatistik ve veri madenciliği) gibi farklı alanlardan kavram ve teknikleri içeren çok disiplinli bir bilgi dalıdır. Metin madenciliği teknikleri; tıbbi araştırma, bibliyometrik çalışmalar, pazarlama, hükümet yönetimi, politik araştırma ve teknoloji gibi birçok alanda uygulanmakta ve kullanılmaktadır. Bu alanların çoğunda metin madenciliğine yönelik önemli araştırmalar yayınlanmaktadır. Nagarkar ve Kumbhar (2015), Web of Science veri tabanında “Bilgi Bilimi” konu kategorisi altında yayınlanan metin madenciliği literatürüne ilişkin çalışmalarında; biyoloji, teknoloji, kimya, fizik, tıp bilimleri ve sosyal bilimler alanlarına yönelik 36.000'den fazla araştırma makalesi olduğunu tespit etmiştir. Wake Forest Üniversitesi'nden Profesör Aldhizer (2017), adli tıp ve denetim uygulamalarının yüksek riskli görevler için metin analitiği kullanmayı düşündüğünü öne sürmektedir. Aldhizer, metin analitiğinin sosyal medya gönderilerinden ve e-postalardan suçlayıcı kelimeleri belirlemek için kavram çıkarımı amacıyla kullanılabileceğini belirtmektedir. Ayrıca metin madenciliği, kiralama ve gelir işlemlerine yönelik uygun raporlamayı oluşturmak için de kullanılmaktadır (Iowa State University, 2018, p.6). 2019 yılındaki bir araştırma (Lewis & Young, 2019, p. 597), 2010-2018 döneminde önemli muhasebe ve finans dergilerinde yayınlanan metin analizi yöntemlerini kullanan makaleleri gözden geçirmiştir. Araştırmada;

- Muhasebe ve finans mesleğinin doğal dil işleme yöntemini benimsemeye yavaş olduğu,
- Muhasebe ve finans uzmanlarının anahtar kelime aramaları, kelime sayıları ve okunabilirlik puanları gibi belirli nitelikleri ölçen sözlükler gibi basit yaklaşımları tercih ettiği,
- Belge düzeyinde, makine öğrenmesi sınıflandırıcıları ve konu yapısını belirlemeye yönelik istatistiksel yöntemler de dahil olmak üzere, belirli doğal dil işleme tekniklerini uygulayan küçük bir çalışma grubunun olduğu gibi tespitler yapılmıştır.

Genel olarak literatür taraması değerlendirildiğinde, çoğunluğu yabancı kaynak olmak üzere gerçekleştirilen çalışmalardan metin madenciliğinin

muhasabe alanında yeni kullanılmaya başlandığı tespit edilmiştir. Bu bağlamda metin madenciliğinin güncel bir teknoloji alanı olması ve muhasabe ile bağdaştırılması bakımından bu çalışmanın literatüre önemli katkı sağlayacağı düşünülmektedir. Metin madenciliği teknolojisinin muhasabe alanında kullanımına yönelik bilgiler vermeyi amaçlayan bu çalışmada; öncelikle metin madenciliğinin tanımı, metin madenciliğinde kullanılan yöntemler ve teknolojiler, metin madenciliğinin avantajları ve dezavantajları konusu üzerinde durulmuştur. Daha sonra metin madenciliğinin muhasabe alanında kullanımına ilişkin muhasabe otomasyonuna, vergi otomasyonuna, denetim otomasyonuna ve iş danışmanlığı otomasyonuna yönelik açıklamalar yapılmıştır. Son olarak çalışmadan elde edilen veriler paralelinde metin madenciliğinin muhasabe alanına etkisi sonuç kısmında değerlendirilmiştir.

2. METİN MADENCİLİĞİ VE METİN MADENCİLİĞİNDE KULLANILAN YÖNTEM VE TEKNOLOJİLER

Metin madenciliği; kalıpları, eğilimleri ve temaları keşfetmek için çeşitli metin kaynaklarından (Word belgeleri, PDF dosyaları, sosyal medya gönderileri, e-postalar, web siteleri, makaleler, XML dosyaları vb.) bilgi çıkarma işlemidir. Bu kaynaklarda bulunan metinler genel olarak yapılandırılmamıştır, yani IDEA veya ACL gibi yazılımlar tarafından önceden tanımlanarak analiz edilebilecek bir formatta değildirler. Literatürde metin madenciliğinin birçok tanımı bulunmaktadır. Bu tanımlarda dikkat edilmesi gereken unsur artık metin madenciliğinin bilgi erişiminin ötesine geçtiğidir. Bilgi erişiminin amacı, bilgiyi elde etmeyi sağlamak veya bilgiyi daha hızlı ve daha doğru hale getirmek için ilgili ve çeşitli metinler arasından ayırım yapmaktır. Metin madenciliğinin amacı ise öncelikli olarak yeni faaliyetlerde kullanılabilir bilgiler elde etmek veya karar vermeyi geliştirmek için metinleri analiz etmek ve onlardan yeni öngörüler ortaya çıkarmaktır. Başka bir deyişle, metin madenciliği, veri tabanları gibi yapılandırılmış kaynakların aksine, yapılandırılmamış metin kaynaklarından gelen kalıpları analiz eden bir veri madenciliği biçimidir. Metin madenciliği, iki aşamada gerçekleşmektedir. Bu aşamalar sırasıyla (Sharda vd., 2014, p. 22);

- Metin veri kaynaklarını yapılandırmak ve
- İlgili bilgileri çıkarmak için veri madenciliği tekniklerini kullanmak şeklinde sıralanmaktadır.

Metin madenciliği, yapılandırılmamış metinler içeren web özellikli uygulamalar ve sosyal ağlar ile internetteki içeriğin katlanarak büyümesi nedeniyle önemli bir ivme kazanmaktadır. Yapılandırılmamış metinlerden bilgi keşfetme süreci olarak metin madenciliği, doğal dil yorumunu

gerektirmektedir. Bu nedenle, bilgisayar programını veya metin gibi doğal dili yorumlama süreci olan doğal dil işleme, metin madenciliğinin metinlerden eyleme dönüştürülebilir bilgileri etkili ve verimli bir şekilde çıkarmasını sağlayan en önemli metodolojilerden biridir. Doğal dil işleme, bilgisayarların ve insanların dillerini kullanan insan etkileşimine odaklanan bir yapay zekâ alt alanıdır. Bu alt alan doğal dil anlayışı ve doğal dil üretimi olmak üzere ikiye ayrılmaktadır.

Metin madenciliği sistemleri, genellikle programlama dillerinin kitaplıklarında (örneğin Python) veya yazılım uygulamalarında (örneğin SAS Text Miner, IBM Watson NLU) bulunan çok çeşitli teknikler ve algoritmalarından (bilgisayar programları) oluşmaktadır. Gartner'ın 2018 (Davis vd., 2018, p. 4) yılında yaptığı bir araştırmada, metin madenciliği hizmeti sunan tedarikçileri; Expert System (Cogito), SAP (SAP HANA Text Mining XS classic), Microsoft Azure, Google (Cloud Natural Language API), IBM, Verint Systems, SAS (Text Miner), Lexalytics, Amazon Comprehend ve OpenText şeklinde sıralamaktadır. İsviçre Federal Teknoloji Enstitüsü Yapay Zekâ Laboratuvarı araştırmacıları Martin Rajman ve Martin Vesely tarafından metin madenciliği teknikleri ve algoritmaları ön belge işleme, madencilik (metinsel veriler için) ve görselleştirme olmak üzere üç kategoride sınıflandırılmaktadır (Rajman & Vesely, 2004, p. 9).

2.1. Ön Belge İşleme

2.1.1. Veri Seçimi ve Filtreleme

Metinleri analiz etmek veya en alakalı öğelerine indirgemek için veri seçimi ve filtreleme teknikleri ve algoritmaları kullanılmaktadır. Daha spesifik olarak veri seçimi; anahtar veya tanımlayıcı kelimeler gibi ilişkili oldukları açık meta verilere dayalı olarak ilgili belgelerin tanımlanması ve alınması konusunda kullanıcılara yardımcı olmaktadır. Veri filtreleme daha sonra uygunluk ölçütlerini kullanarak belgelerin alaka düzeyini gerçek içeriklerine göre değerlendirmektedir.

2.1.2. Veri Temizleme

Veri temizleme araçları; metinsel verilerden yazım hatalarını, tutarsızlıkları ve gereksiz öğeleri kaldırmak ve üst dilsel bilgileri tanımlamak için kullanılmaktadır. Bu araçlar, kullanıcılara aşağıdaki konularda yardımcı olmaktadır (Rajman & Vesely, 2004, p. 12):

- Hataları düzeltmek
- Metni normalleştirmek (örneğin büyük/küçük harf normalleştirme, kısaltma normalleştirme)
- İşlenen dile ait olmayan kısımları çıkarmak,

- Gereksiz sözcükleri, noktalama işaretlerini ve özel karakterleri kaldırmak
- Kelimelere üst dil etiketleri atamak. Üst dil etiketleri aşağıdaki gibidir:
 - Adlandırılmış Varlık Tanıma: Kişilerin, yerlerin, kuruluşların ve tarihlerin adlarını tanımlayan bir algoritma türüdür.
 - Sözcük Türü Etiketleme: Her kelimeye isim, fiil, sıfat ve herhangi bir tanımlayıcı atayan bir algoritma türüdür.

2.1.3. Belge Temsili

Belge temsil algoritmaları, yapılandırılmamış metni, sistemin analiz veya görselleştirme için yorumlayabileceği bir temsile dönüştürmektedir. Basit olarak temsil yaklaşımı; bir belge veya cümle gibi bir metnin, metindeki ve dilbilgisindeki sıralarına bakılmaksızın bir dizi kelime olarak temsil edildiği kelime torbası (BoW) modeli yaklaşımıdır. Metin belgeleri için sıklıkla kullanılan popüler bir BoW temsili, vektör uzay modelidir (VSM). Bir vektör uzay modelinde, her metin belgesi bir vektör olarak temsil edilir ve vektör uzayı, metin belgesindeki sözcüklerle, önemliliği ile temsil edilir ve oluşma sıklığı (yani, sözcüğün metin belgesinde görünme sayısı) bakımından ölçülür (Sharda vd., 2014, p. 41). Ayrıca daha karmaşık temsil modelleri, daha yapılandırılmış anlamsal modelleri içermektedir.

Keika vd., (2008), dört farklı belge temsil türünü ifade etmektedir. Bunlar; N-gramlar, tek terimler, cümleler ve zengin belge yaklaşımı olarak adlandırılmaktadır:

- N-gram yaklaşımı; dilsel işleme içermeyen dize tabanlı bir gösterimdir. Belgelerin n kelimelik diziler olarak temsil edildiği en basit gösterimdir.
- Tek terimli yaklaşım; minimum dilsel işlemeye sahip kelimelere dayanmaktadır. Bu yaklaşımda belgeler, daha önce açıklandığı gibi farklı sözcüklerinin vektörleri ve önemleri olarak temsil edilmektedir (vektör uzay modeli yaklaşımı). Çoğu zaman, belge eşleştirme sonuçlarını artırmak için kelimelerin kendileri yerine kelimelerin kökü kullanılmaktadır.
- Tümler yaklaşımı; dilsel olarak oluşturulmuş tümcelere ve tek sözcüklere dayanmaktadır. İstatistiksel veya dilsel ifadelerin çıkarılmasını ve belgelerin kök kelime ve kelime öbekleriyle temsil edilmesini içeren daha karmaşık bir yaklaşımdır.
- Zengin belge yaklaşımı; bir belgenin daha anlamsal bir temsilini sağlamaktadır. Dilsel işlemeye (konuşma parçası ve eşleştirme kuralları tanımlama gibi) dayanmaktadır ve belgeleri, metindeki

ilişkileri tanımlayan bir dizi mantıksal terim ve ifade olarak temsil etmektedir. Örneğin, “.....kişisel bilgisayarlar için işletim sistemleri.....” cümlesindeki “için” sözcüğü, “işletim sistemleri” ile “kişisel bilgisayarlar” arasında bir ilişki olduğunu düşündürmektedir.

Ayrıca sadece sözcük ve kelimelere değil aynı zamanda kavramlara dayalı belge temsil modelleri de vardır. Kavram temelli bir temsil yaklaşımında bir belge, bir kavram vektörü olarak temsil edilmektedir. Kavramların önemi, ortaya çıkma sıklığı açısından ölçülmektedir. Hiyerarşik veya örgü yapısı, kavramın belgede hiyerarşik bir "is-a (katılım)" ilişkisi olarak görünme sayısını temsil etmektedir. Örneğin, bir belge "satış", "alış" ve "muhasabe" olmak üzere üç ilgili kavramı içeriyorsa, muhasabe hem satışın hem de alışın bir üst sınıfı olacaktır (Da Costa Pereira & Tettamanzi, 2006, p. 217). Yazılımlar bu karmaşık yaklaşımların, metinlerin anlamını daha iyi pekiştirmesini, sistemlerden bilgi almasını, analizler gerçekleştirmesini ve görselleştirme açısından daha doğru ve kullanışlı kararlar vermesini sağlamaktadır.

2.1.4. Morfolojik Normalleştirme ve Ayrıştırma

Morfolojik normalleştirme; kök ayırma, lemmatizasyon (sözcüklerin köklerine indirgenmesi) ve konuşmanın bir bölümünü etiketleme gibi doğal dil görevlerini gerçekleştirmektedir. Ayrıştırma; metin bölümlenme, cümle öbekleri gibi daha küçük sözdizimsel birimlere ayırma ve belirlenen birimler arasındaki sözdizimsel ilişkilerin tanımlanması dahil olmak üzere normalleştirilmiş metne sözdizimsel yapı atama sürecini kapsamaktadır (Rajman & Vesely, 2004, p. 14).

2.1.5. Semantik Analiz

Semantik analiz araçları, anlamsal belirsizlikleri çözmektedir. Bu araçlar, kelime anlamı belirsizleştirme (WSD), anafora çözünürlüğü ve ortak referans çözünürlüğü gibi teknikleri kullanmaktadır. Semantik analiz aynı zamanda belgeler arasındaki ve içindeki güncel yakınlığı da değerlendirmektedir. Sistematik analize yönelik araçlar sistem tarafından metinlerin anlamlarını yorumlama yeteneğini daha da artırmaktadır. Bu teknikler ve algoritmalar aşağıda yer almaktadır (Da Costa Pereira & Tettamanzi, 2006, p. 223):

- Kelime anlamı netleştirme; sistemlerin birden çok anlamı olan kelimeleri yorumlamasını sağlamaktadır. Kelime anlamı netleştirme araçları, metindeki o kelimenin etrafındaki diğer kelimeler tarafından belirlenen bir kelimeye uygun anlamı atamak için kullanılmaktadır.
- Öncül çözümleme araçları; bir metinde zamirlerin veya isim tamlamalarının nelere atıfta bulunduğunu çözmeye yardımcı olmaktadır (isim öbeği, ilgili ismi değiştirmektedir, örneğin;

“muhasabe ilkeleri”, ilke isimdir ve onun deęiřtiricisi muhasabe kavramıdır).

- Ortak referans araçları; aynı varlığa atıfta bulunan ifadelerin bulunmasına yardımcı olmaktadır (Random House Kernerman Webster, 2020).
- Gizli semantik analiz; eş anlamlıları veya kelime grupları gibi terim bağımlılıklarını ortaya çıkarmaya yardımcı olan teknikleri ve algoritmaları ifade etmektedir.
- Başka bir anlamsal model ve algoritma türü, bir belge koleksiyonunun sözcüklerinden konuları tanımlayan, bir belgeyi konuların bir karışımı olarak temsil eden tekniklere ve sistemlere atıfta bulunan gizli Dirichlet tahsisidir (LDA). Gizli Dirichlet tahsisi (LDA), doğal dil işlemede önemli bir konu modelleme tekniğidir.
- Güncel yakınlığı değerlendirmek için kullanılan ölçüler; genelleştirilmiş Öklid uzaklığı, kosinüs benzerliği ve Ki-kare mesafesi gibi vektör uzayı benzerlik ölçütlerini ve ifade oluşumuna dayalı ölçütler, değerlendirilen belgenin uzunluğuna dayalı ölçütler gibi diğer ölçütleri içermektedir (Rajman ve Vesely, 2004, p. 18).

Semantik analiz, doğal dil işleme sürecinde zor bir alan olmasına rağmen hızlı bir şekilde gelişmektedir. Bu durum, yazılımın doğal dili yorumlama becerisini daha da geliştirirken güncel metodolojilere makine öğrenmesi uygulamalarını kullanma imkânı sağlamaktadır.

2.2. Madencilik

2.2.1. Kümeleme

Belge kümeleme, belgeleri benzerliklerine göre kümeler halinde gruplayan bir makine öğrenmesi tekniğidir. Metin madenciliğinde belge seçimi, organizasyon, özetleme ve görselleştirme gibi çeşitli işlevler için kümeleme kullanılmaktadır. Kümeleme sürecinde birden fazla yaklaşım ve çok çeşitli algoritmalar yer almaktadır. En iyi bilinen kümeleme algoritması “k-ortalamar” algoritmasıdır. Bu algoritmada, her küme en yakın veri noktalarının ortalaması ile temsil edilmektedir (yani her küme, benzerlik ölçülerine veya uzaklık ölçülerine göre algoritma tarafından gruplandırılmış bir belge kümesini temsil eder). Benzer kümeleme teknikleri, medyan veya mod gibi diğer merkezi eğilim ölçülerini kullanmaktadır. Farklı kümeleme metodolojileri, yoğunluğa dayalı kümelemeyi ve hiyerarşik kümelemeyi içermektedir (Tang vd., 2008, p. 14).

2.2.2. Sınıflandırma

Belge sınıflandırma, belgelere önceden tanımlanmış sınıflar atayan bir makine öğrenmesi tekniğidir. Kümeleme algoritmalarının aksine,

sınıflandırma algoritmaları denetlenmektedir. Araştırmacılar, algoritmaya doğru sınıfı ve belgeyi temsil etmek için kullanılan özellikleri (örneğin vektör tabanlı temsil) içeren eğitim örnekleri sağlamaktadır. Sınıflandırma algoritması daha sonra verilen özellikleri her sınıfa en iyi şekilde eşleyen bir model oluşturmaktadır. Eğitim verileri yalnızca iki sınıfa sahip olduğunda, ikili sınıflandırıcı oluşturulmaktadır. İki'den fazla sınıfın olduğu durumlarda ise çok sınıflı bir sınıflandırıcı gerekmektedir. Sınıflandırma algoritmalarına (Blake, 2011, p. 132);

- K-en yakın komşuluk,
- Naive Bayes (Bayes'in koşullu olasılık teoremine dayalı olasılık modeli),
- Destek vektör makineleri (eğitim verilerinin uzayda kategorilere ayrılmış noktalar olarak temsil edildiği modeller),
- Karar ağaçları (eğitim verilerinden öğrenilen “if-then” kuralları ile yukarıdan aşağıya bir ağaç yapısına dayalı) ve diğer karar kuralı tabanlı modeller ve
- Sinir ağları (insan beyninin bilgiyi işleme şeklini taklit eden modeller) gibi örnekler verilebilir.

İstenmeyen posta filtreleme, e-posta yönlendirme, içerik etiketleme, müşteri görüşü ve duygu analizi gibi belge sınıflandırmanın çeşitli uygulamaları da vardır.

2.2.3. Varlık ve İlişki Çıkarımı

Varlık çıkarma algoritmaları; belgelerden kişi adları, kuruluş adları, konumlar, tarihler, telefon numaraları, referans numaraları, fiyatlar, tutarlar ve diğer öğeler gibi varlıkları çıkarmak için kullanılmaktadır. İlişki çıkarma algoritmaları; kişi-kuruluş, kişi-konum veya kuruluş-mekân gibi varlıklar arasındaki ilişkileri tanımlamak ve karakterize etmek için kullanılmaktadır. Bazı algoritmalar, bir olayla ilgili varlıkları tanımlamayı amaçlayan çıkarımlara odaklanmaktadır. Bilgi çıkarma (IE) algoritmaları; kural öğrenme tabanlı yöntemler, sınıflandırma tabanlı yöntemler ve sıralı etiketleme tabanlı yöntemler dahil olmak üzere çeşitli makine öğrenmesi yaklaşımlarını kullanmaktadır (Tang vd., 2008, p. 19).

- Kural öğrenmeye dayalı sistemler, metinden istenen bilgiyi (kelimeler veya metin parçaları) çıkarmak için önceden tanımlanmış talimatları kullanmaktadır.
 - Sözlük tabanlı sistemler: Bu sistemler önce bir şablon sözlüğü oluşturur ve ardından metinden gerekli bilgileri çıkarmak için bu sözlüğü kullanmaktadır.

- Kural tabanlı sistemler: Bu sistemler, metinden bilgi çıkarmak için sözlük yerine genel kuralları kullanmaktadır.
- Sarma sistemleri: Web sayfaları gibi yapılandırılmış ve yarı yapılandırılmış belgelerden bilgi çıkarmak için bir dizi bilgi çıkarım kuralı ve program kodlarından oluşan prosedürdür.
- Sınıflandırmaya dayalı sistemler, bilgi çıkarımını bir sınıflandırma görevi olarak kullanır (yani, çıkarma kuralları bir sınıflandırma modeline dayalı olarak oluşturulur).
- Sıralı etiketlemeye dayalı çıkarım sistemleri; sıralı etiketlemenin bir görevi olarak bilgi çıkarımını gerçekleştirmektedir. Sıralı etiketlemede bir belge, bir dizi kelime ve konuşma bölümü dizisi olarak görülmektedir. Bu sistemler, hedef bilgiler arasındaki bağımlılıkların tanımlanmasını sağlamaktadır. Bu bağımlılıklar, çıkarımların doğruluğunu artırmak için kullanılmaktadır.

Günümüzün dijital iş ortamında belgelerden bilgi çıkarmanın çok sayıda uygulaması yer almaktadır. Bunlara (Torres-Moreno, 2014, p. 48);

- Dijital kütüphaneler için otomatik meta veri üretimi,
- Veri girişi otomasyonu için günlük iş uygulamalarında otomatik bilgi çıkarma (örneğin, özgeçmişlerden, makbuzlardan, faturalardan, yasal belgelerden ve diğerlerinden bilgi çıkarma),
- Bilgi teknolojisi güvenliği, uyumluluk, izleme veya pazarlama araştırması gibi amaçlar için e-postalardan, sosyal medyadan veya diğer metin kaynaklarından otomatik bilgi çıkarma,
- Uyum, dolandırıcılık riski, kredi riski tespiti, denetim, finansal araştırma, bilimsel araştırma ve patent analizi gibi çeşitli amaçlar için otomatik belge incelemesi ve analizi,
- Satın alma siparişlerinden veya sosyal medya içeriklerinden çıkarılan kalıplara dayalı ürün veya film tavsiye sistemleri
- Borsa analizi, firma finansal performansı, hisse senedi fiyat hareketi tahminleri ve dolandırıcılık tespiti için SEC dosyalarından ve diğer yatırımcı iletişim materyallerinden (örneğin basın bültenleri) otomatik bilgi çıkarma gibi örnekler verilebilir.

2.3. Görselleştirme

2.3.1. Çok Boyutlu Veriler İçin Görselleştirme Teknikleri

Görselleştirme teknikleri, metinlerden çıkarılan verilerin grafiksel temsillerine dayanmaktadır. Örnekler arasında kelime bulutları, etiket bulutları, histogramlar, dağılım grafikleri, çizgi grafikleri, kutu grafikleri, ağ grafikleri ve diğer grafikler bulunmaktadır. Görselleştirme; kümeleme,

güncel yakınlık analizi, duygu analizi ve diğer analizlerden elde edilen sonuçları değerlendirmek için yaygın olarak kullanılmaktadır. Görselleştirme, metin madenciliği sürecinin ve genel olarak veri madenciliğinin önemli bir parçasıdır. Araştırmacı Daniel Keim'in (2002) belirttiği gibi; görsel veri keşfi, algısal yetenekleri günümüz bilgisayar sistemlerinde bulunan büyük veri setlerine uygulayarak insanı veri keşif sürecine entegre etmeyi amaçlamaktadır. Görsel veri keşfinin temel işlevi; veriyi görsel bir biçimde sunarak insanın veriyi kavramasını, sonuçlar çıkarmasını ve verilerle doğrudan etkileşime girmesini sağlamaktır. Ayrıca, görsel veri keşfinin istatistik, makine öğrenmesi ve otomatik veri madenciliği tekniklerine göre başlıca avantajları şunlardır:

- Görsel veri keşfi, homojen olmayan ve detay işlem gerektiren verilerle kolayca başa çıkabilir.
- Görsel veri keşfi sezgiseldir ve karmaşık matematiksel veya istatistiksel algoritmaların veya parametrelerin anlaşılmasını gerektirmez.

Genellikle görselleştirme daha hızlı veri keşfine olanak sağlamak ve özellikle otomatik algoritmaların başarısız olduğu durumlarda daha iyi sonuçlar vermektedir. Daniel Keim (2002), daha basitten daha karmaşığa doğru görselleştirme tekniklerinin kapsamlı bir sınıflandırmasını yapmaktadır. 2D/3D ekranlar gibi daha basit metodolojiler, küçük ve düşük boyutlu veri kümeleri için uygundur. Geometrik olarak dönüştürülmüş görüntüler, simge tabanlı görüntüler, yoğun piksel görüntüler ve diğer gelişmiş teknikler çok boyutlu, metin, hiper metin ve diğer veri kümeleri için kullanılmaktadır. Dinamik projeksiyonlar, yakınlaştırma, filtreleme, etkileşim ve bozulma teknikleri bu görselleştirme yöntemleriyle ilişkilidir.

2.3.2. Metin Özetleme

Metin özetleme; kaynakta neyin önemli olduğunu, seçme, genelleştirme ve içerik indirgeme yoluyla kaynak metnin özet metne dönüştürülmesinden oluşmaktadır (Sparck-Jones, 1999). Otomatik metin özetleme alanı, son yıllarda yoğun araştırmaların konusu olmaktadır. Metin özetleme (anlamsal analiz gibi) bilgisayar sistemleri için en zorlu görevlerden biridir. İnternetteki yapılandırılmamış metinlerin katlanarak büyümesi, doğal dil işleme ve makine öğrenmesindeki son gelişmeler metin özetleme teknolojilerine olan ilgi ve ihtiyacı artırmaktadır. Otomatik metin özetleme sistemleri birçok fayda sağlamaktadır. Bunlar;

- Belgeleri okumak veya taramak için manuel emeği azaltmak,
- Özetleme sürecinde verimliliği ve etkinliği artırmak,
- Belge seçimini ve indekslemeyi kolaylaştırmak,
- Soru cevaplama sistemlerinde cevapları zenginleştirmek ve

- Metin özetlemesindeki olası insan hatalarını veya eksiklikleri ortadan kaldırmak şeklinde sıralanmaktadır.

Otomatik metin özetleme; özetleme sürecinde insan önyargılarını ortadan kaldırma potansiyeline sahiptir. Kumar vd., (2016), iki tür otomatik özetleme uygulamasından bahsetmektedir. Bunlar:

- Çıkarımsal özetler; doğrudan belgeden seçilen anlamlı cümleleri tanımlayarak üretilmektedir. Önemli cümleler, kelime sıklığı, cümle konumu, başlık ve diğer kriterler kullanılarak bir öğrenme modeli (makine öğrenmesi yaklaşımı) otomatik olarak gerçekleştirilmektedir.
- Soyutlayıcı özetler; seçilen belge cümlelerinin tutarlı bir şekilde birleştirilmesi ve cümlelerin önemsiz kısımlarının hariç tutulması şeklinde üretilmektedir. Soyutlayıcı özetler oluşturmak, oldukça karmaşık bir dil modellemesi gerektirir bu nedenle bilgisayar sistemleri için soyutlayıcı özet görevi çıkarımsal özet görevinden daha karmaşıktır.

Torres-Moreno çalışmasında (2014), otomatik metin özetlemenin birçok çeşidi olduğunu belirtmiştir. Bunlardan biri olan varyantlar; genel veya kişiselleştirilmiş, tek belgeli, çok belgeli ve çok dilli özetlemeyi içermektedir. Torres-Moreno tarafından tanımlanan otomatik metin özetleme algoritmalarının ek varyantları; alana özgü özetlemeyi (örneğin kimya, biyomedikal, muhasebe), güncelleme özetlemeyi (yalnızca yeni gerçekler), cümle sıkıştırmayı, çok cümleyi birleştirmeyi, anlamsal özetlemeyi ve ultra özetlemeyi (kısa metinlerin özetlenmesi) kapsamaktadır. Metin özetlemede derin öğrenmenin kullanımı son zamanlarda büyük ilgi görmekte ve yeni araştırmalar için önemli bir konu olmaktadır. Derin öğrenme; metin madenciliği bağlamında metindeki verilerden kalıpları keşfetmek için yapay sinir ağlarını kullanan bir makine öğrenmesi alt kümesidir. Araştırmacı Sukriti Verma & Vagisha Nidhi (2019), derin öğrenme modeli kullanarak olgusal raporlar oluşturmak amacıyla çıkarımsal bir metin özetleme yaklaşımını önermişlerdir. Bu model, özet için seçilen cümle setini ve özetin doğruluğunu geliştirmek için çeşitli özellikleri (en sık geçen kelimeler, cümle yapısı, cümle uzunluğu ve diğerleri) araştırmaktadır.

Zhang vd., (2016), evrişimsel sinir ağı modeline dayalı olarak bir belge özetleme çerçevesi geliştirmiştir. Bu çerçeve, cümle özelliklerini öğrenir ve önceden programlanan kelime vektörleri için bir regresyon süreci kullanarak cümle sıralamasını oluşturmaktadır. Evrişimsel sinir ağı, başlangıçta bilgisayar görüntüsü görevleri ve daha yakın zamanda doğal dil işleme görevleri için kullanılan bir tür derin sinir ağıdır. Sonuç olarak, otomatik özetleme algoritmaları, canlı ve dinamik bir araştırma alanını temsil etmektedir. Bu algoritmalar, performanslarını iyileştirmek ve yeni zorlukların üstesinden gelmek için sürekli olarak gelişmektedir. Zorluklar;

sistemlerin küresel, çok dilli ortamlarda giderek daha büyük miktarlarda yapılandırılmamış metinleri (bloglar, e-postalar, sosyal medya gönderileri ve diğerleri) ve diğer multimedya içeriklerini (görüntüler, ses ve video) işleme becerisini içermektedir.

3. METİN MADENCİLİĞİNİN AVANTAJLARI VE DEZAVANTAJLARI

Metin madenciliği çeşitli endüstrilerde yaygın olarak kullanılmaktadır. Bunun sebebi, kullanıcıların büyük miktarda yapılandırılmamış metinlerden veri madenciliği yapmalarını sağlamasıdır. Veri madenciliği yapılandırılmamış metin; tüm endüstrilerde e-ticaret, sosyal medya ve dijital iş modellerinin genişlemesiyle birlikte ve internetteki yapılandırılmamış metin içeriklerinin son zamanlarda katlanarak büyümesi göz önüne alındığında önemli bir noktaya gelmiştir. Birçok kuruluş, kâğıt tabanlı işlemlerden dijital işlemlere, fiziksel satış ve tedarik zinciri kanallarından dijital platformlara ve şirket içi uygulamalardan bulut tabanlı uygulamalara gelişim göstermiştir. Bu bağlamda internette her geçen gün daha fazla elektronik metinsel veri üretilmektedir. Metin madenciliği, kullanıcıların pazarlama, araştırma ve geliştirme, müşteri hizmetleri, operasyonlar, tedarik zinciri yönetimi, finans, muhasebe ve idari hizmetler gibi tüm iş fonksiyonlarında ön görüler oluşturmak için bu geniş veri havuzunu analiz etmelerini sağlamaktadır. Metin madenciliği, özellikle büyük veri çağında, veri madenciliğinin önemli bir parçasıdır. Yapılandırılmamış metinsel veriler (belgeler, web siteleri, e-postalar, mesajlar, sosyal medya gönderileri, ses ve video dosyalarının açıklamaları, transkriptler, konuşmayı metne dönüştürme dosyaları ve diğerleri) günümüzün dijital dünyasında veri madenciliği için mevcut içeriğin çoğunluğunu temsil etmektedir. Metin madenciliği aşağıdakiler gibi farklı avantajlarda sunmaktadır (Zhang vd., 2016, p. 918):

- Diğer alternatiflere (manuel veya yarı manuel işlemler gibi) kıyasla içeriğin seçiminde, gözden geçirilmesinde ve özetlenmesinde verimliliği artırmak
- Kullanıcıların kalıpları ve öngörülerini zamanında ortaya çıkarmasını sağlamak (işleme ve analiz hızı)
- Geleneksel araştırma yöntemlerinin örneklere dayandığı durumlarda, tüm içeriğin işlenmesini gerçekleştirmek
- Görevlerini tutarlı bir şekilde yürüterek insan hatalarını ve insan önyargılarını ortadan kaldırmak
- Metin madenciliği sonuçlarını soru cevaplama sistemlerine entegre etmek (Optik karakter tanıma teknolojisi ile birlikte bilgi çıkarma

algoritmalarının; bir denetim veya uygunluk sistemine otomatik sözleşme girişlerini, bir iş başvuru sisteminde otomatik özgeçmiş girişlerini ve bir borç hesapları sisteminde otomatik fatura girişlerini kurumsal sistemlere entegre etmesi)

Metin madenciliği aynı zamanda önemli dezavantajları veya sınırlamaları bünyesinde bulundurmaktadır. Metin madenciliğinin temel dezavantajları; yüksek karmaşıklığı, semantik analiz ve soyutlayıcı özetlemedeki sınırlamalarıdır. Teknolojinin, yapay zekânın ve makine öğrenmesinin hızla gelişmesi birlikte metin madenciliğinde önemli ilerlemeler gerçekleşmiştir. Metin madenciliği teknolojileri olan semantik analiz ve soyutlayıcı özetlemeyi içeren görevlerde insanlara ihtiyaç duyulmaktadır. Ayrıca, algoritmalar ve belgelerden bilgi çıkarma görevleri esnasında sonuçların gözden geçirilmesi ve düzeltilmesi için insan çabası ön plana çıkmaktadır. Otomasyondan elde edilen bilgiler genellikle doğrulama ve çalışan personelin ek veri girmesini gerektirdiği için bu da süreçlerde makine ve insan etkileşiminin bir kombinasyonunu zorunlu kılmaktadır.

Metin madenciliğinin bir başka potansiyel dezavantajı, bazı durumlarda şeffaflık veya yorumlanabilirlik eksikliğidir. Algoritmalar, kendi iç mantığının veya işleyişinin gizlendiği kara kutu yaklaşımlarını kullandığında şeffaflık eksikliği oluşur ve sistemin belirli bir sonuca veya tahmine nasıl ulaştığını açıklamasını zorlaştırır. Ayrıca, metin madenciliğinin kullanımı, mahremiyetin ve düzenlemelere uyumun dikkate alınmasını gerektirmektedir. Örneğin, bir bireyin e-postalarında veya sosyal medya içeriklerinde metin madenciliği kullanmak, tüketici mahremiyet yasalarını ihlal edebilir veya başka bir şekilde etik kaygılara yol açabilir. Metin madenciliğinde analiz edilen bilgiler, etik olmayan bir şekilde kötüye kullanılabilir böylece kuruluşu güvenlik, telif hakkı ihlali veya ayrımcılık riskleri gibi risklere maruz bırakabilir (Nagarkar & Kumbhar, 2015, p. 257). Son olarak, metin madenciliği algoritmalarını ve makine öğrenmesi eğitim setlerini sürdürmek zorunlu bir görev olmaktadır. Teknoloji, organizasyon ve diğer faktörlerin hızlı değişimleri çevre ve endüstrileri önemli derecede etkilemektedir. Örneğin geçerli metin madenciliği setlerini düzgün bir şekilde sürdürmek; özellikle standartların, kuralların ve düzenlemelerin sürekli değiştiği muhasebe dünyasında çok önemlidir.

4. MUHASEBEDE METİN MADENCİLİĞİNİN MEVCUT VE POTANSİYEL UYGULAMALARI

Metin madenciliği muhasebe alanına doğru yol almasına rağmen muhasebe meslek mensupları arasında metin madenciliğinin kullanımı çok yeni bir süreçtir. “Big Four (Deloitte, PwC, Ernst&Young, KPMG)” şirketleri metin madenciliği teknolojilerinin kayda değer ilk uygulayıcıları olsa da diğer muhasebe kuruluşlarının da bu teknolojileri benimsemeye veya denemeye

başladığını görmekteyiz. Metin madenciliği teknolojilerinin muhasebedeki mevcut ve potansiyel uygulamaları dörde ayrılmaktadır (Torres-Moreno, 2014, p. 71). Bu uygulamalar; denetim otomasyonu, muhasebe otomasyonu, vergi otomasyonu ve iş danışmanlığı otomasyonu şeklinde sınıflandırılmaktadır.

4.1. Denetim Otomasyonu

Bazı muhasebe ve denetim firmaları, denetim veya uygunluk değerlendirmesi için sözleşme incelemelerini otomatikleştirmek amacıyla metin madenciliği yazılımı kullandığını belirtmektedir. Bunlara örnek olarak Deloitte, Ernst&Young ve EisnerAmper firmaları verilebilir (Alarcon vd., 2019, p. 4). Otomatik sözleşme incelemeleri; kuruluşların yıllık mali denetim kapsamına giren dolandırıcılık tespiti, iç denetimler, bilgi teknolojisi denetimleri gibi alana özgü denetimlerin bir parçası olarak gerçekleştirilmektedir. Örneğin, kira işlemlerinin muhasebe standartlarına uygunluğun değerlendirilmesine yönelik sözleşme incelemeleri, kira (Deloitte, 2020) sözleşme inceleme yazılımları kullanılarak otomatikleştirilmektedir.

Sözleşme incelemelerine ek olarak, metin madenciliğinin denetimdeki mevcut veya potansiyel uygulamaları; SEC dosyalarının ve diğer kurumsal iletişim malzemelerinin, iş sözleşmelerinin, finansman sözleşmelerinin, müşteri sözleşmelerinin, satıcı sözleşmelerinin ve işlemlerinin, ilgili taraf açıklamalarının ve mali tabloları destekleyen diğer belgelerin incelenmesini içermektedir. 2018 yılında, Rutgers Üniversitesi'nden Profesör Ting Sun ve Miklos Vasarhelyi, denetimde metinsel verilerin önemini vurgulayarak denetçiler tarafından derin öğrenme tekniklerinin kullanılması gerektiğini savunmuşlardır. Sun ve Vasarhelyi ayrıca çeşitli metinsel veriler tarafından sağlanan bilgilerin denetimdeki yararlılığını da analiz etmişlerdir. Makaleleri, derin öğrenmenin planlama, iç kontrol değerlendirmesi, maddilik testi ve denetimin tamamlanma sürecinde dahil olmak üzere tüm denetim aşamalarında karar verme sürecini destekleyebileceğini öne sürerek denetçilere derin öğrenmeyi uygulamak için bir kılavuz oluşturmuşlardır.

4.2. Muhasebe Otomasyonu

Muhasebe otomasyonu, akıllı bilgi çıkarımını (makine öğrenmesi yaklaşımlarını kullanarak bilgi çıkarımı) çeşitli alanlarda birleştirmeye başlamıştır. Bu alanlardan biri de veri girişi otomasyonudur. Nitekim muhasebe uygulamaları; bilgi çıkarma, optik karakter tanıma teknolojisi ve robotik süreç otomasyonu çözümleriyle birleştirilmiş makine öğrenmesi yazılımlarına entegre olmaktadır. (Pejic-Bach vd., 2019, p. 4) Bu uygulamalar, işlemleri (borç hesapları faturaları gibi) otomatik olarak kaydeder ve işler ya da çıkarılan verilere, bir bilgi tabanına ve sistem tarafından önceki işlemlerden öğrenilen kalıplara dayalı olarak yevmiye kayıtlarına yönelik önerirler de bulunmaktadır.

Benzer şekilde, bu teknolojilerin uygulamaları, sistemlerin veri girişi ve süreç raporlarını otomatikleştirmek için optik karakter tanıma teknolojisinden ve akıllı bilgi çıkarımından (makbuzlardan) yararlanabildiği süreç raporlamasında artık mevcuttur. Metin madenciliği uygulamaları, satıcı sözleşmesi uyumluluğunu yönetmek için tedarikte zincirinde de kullanılmaktadır. Bu durumlarda metin madenciliği, politikalara uyumu sağlamak veya tedarik modellerinin verimliliğini artırmak için satıcı sözleşmelerini analiz edebilir (Sun & Vasarhelyi, 2018, p. 53). Metin madenciliği teknolojilerinin potansiyel kullanımları aşağıdaki gibi sıralanmaktadır:

- Sözleşmeleri muhasebe politikalarıyla uzlaştırmak
- Gelirleri yönetmek
- Fiyatlandırmayı veya diğer sözleşme şartlarını ve yükümlülüklerini kontrol etmek
- Sözleşme yenilemelerini veya uzatmalarını kontrol etmek
- Müşteri sözleşmelerinin otomatik analizini yapmak
- Borç finansman anlaşmalarının ve borç sözleşmesi yönetimi için değişiklikleri otomatik olarak gözden geçirmek
- Muhasebe standartlarına uygunluğun veya muhasebe durum tespitinin sürdürülmesi için kira sözleşmelerinin veya diğer belge türlerinin (iş sözleşmeleri, çalışan teklif mektupları, taşeron sözleşmeleri ve diğerleri) otomatik analizini yapmak

4.3. Vergi Otomasyonu

H&R Block ve KPMG firmaları, vergi işlemleri için IBM Watson'ı kullanan vergi hizmeti sağlayıcılarının örneklerindedir. H&R Block'ta vergi uzmanları, müşterilerin her bir kesinti ve vergi kredisine yardımcı olmak amacıyla IBM Watson destekli bir uygulama ile çalışmaktadır (H&R Block, 2022). KPMG, müşterilerin Ar-Ge kredilerini güvence altına almasına yardımcı olmak için IBM Watson ile bir uygulama geliştirmiştir. Uygulama ile kullanıcılar, doğal dil işleme teknolojisini kullanarak Ar-Ge kredisi almaya uygun projeleri belirlemeye yardımcı olmak için binlerce belgeyi inceleyebilir ve yapılandırılmış veya yapılandırılmamış verileri çok hızlı bir şekilde analiz edebilir (Brown & Rainey, 2018).

Vergi süreci için yapılandırılmış ve yapılandırılmamış verileri kullanıldığı makine öğrenmesi destekli uygulamaların benimsenmesine bir başka örnek Intuit Inc'dir. Intuit Inc, TurboTax kullanıcılarının vergi işlemlerini kolaylaştırmasına yardımcı olan Vergi Bilgi Motoru (TKE) adlı bir uygulama sağlamaktadır. Sistem, bir bireyin mali durumuna dayalı olarak 80.000 sayfadan fazla vergi gerekliliklerini ve talimatlarını ilişkilendirerek

ve birbiriyle bağlantı kurarak her bir vergi mükellefine göre uyarlanmış cevaplar sunmaktadır. Ayrıca, kullanıcı verilerine dayanarak hangi soruların sorulacağını önerir ve hesaplamalara yardımcı olmaktadır. Son olarak, herhangi bir vergi kavramı için vergi bilgi motorunun istenilen zamanda hesaplamaları açıklayabilmesi için bilgilendirme özelliğine sahiptir (Wang, 2019).

Makine öğrenmesi ve yapay zekâ tarafından desteklenen metin madenciliği vergi otomasyonunu önemli derecede etkilemektedir. Bu teknolojilerin kullanımı; soru yanıtlama sistemlerinin ve potansiyel olarak insanları müşteriye özel vergi sorularını ele almada güçlendiren sanal asistanların ortaya çıkmasıyla birlikte vergi hazırlama ve vergi planlama görevlerini daha da otomatikleştirmek için ilerlemeye ve gelişmeye devam edecektir. Metin madenciliği; vergi denetimleri, vergilendirme (gelir vergisi, satış vergisi, emlak vergisi, uluslararası vergilendirme ve diğerleri) ve vergi davalarının analizi için de kullanılmaktadır. Örneğin metin madenciliği; işlemlerin veya varlıkların uygun vergi sınıflandırmasını ve vergilendirme için faturaları, sözleşmeleri veya diğer belgeleri analiz etmektedir.

Vergi kapsamlı yetkili otoritelerinde vergi incelemeleri, vergi kaçakçılığı tespiti, vergi kaçırma, hükümetler veya düzenleyiciler açısından vergi politikaları için bu ileri teknolojilerin önemli kullanıcıları olmaları beklenmektedir. Internal Revenue Service'in (2022) bu duruma ilişkin beyanı aşağıdaki gibidir:

- Gelir idaresi karar verme sürecini iyileştirmek için teknolojiden tam olarak yararlanmalıdır. Modern teknolojiler, özel ve kamu sektörlerindeki kuruluşların misyonlarını, ürünlerini ve hizmetlerini sunma biçimlerini değiştirmeye devam etmektedir. Devlet yöneticileri; dijital teknolojilerin gelir tahsilatı, denetimler, nakit yönetimi ve hasar yönetimi gibi finansal hizmetlerin iyileştirilmesi gerektiğini düşünmektedir. Gelir idaresi, diğer değişikliklere (süreç robotiği, blok zinciri ve yapay zekâ) yanıt vermeli ve daha verimli görev süreci sağlayan teknolojileri bünyesine entegre etmelidir. Örneğin gelir idaresi; kimlik hırsızlığı algılama modellerini, geri ödeme sahtekarlığını ve uyumsuzluğu tespit etmek için tasarlanmış iş kuralı setlerini analiz edebilir ve faaliyetleri sürekli izleyerek kimlik hırsızlığını tespit etme ve önleme konusunda bir iyileştirme modeli sağlayabilir.

4.4. İş Danışmanlığı Otomasyonu

Muhasebe meslek mensupları; finansal muhasebe, denetim ve vergi dışında çeşitli ticari danışmanlık rollerinde de hizmet vermektedir. Genelde birçok muhasebeci; yönetim muhasebesi, finansal planlama ve analiz, iç denetim, uygunluk, adli muhasebe veya bilgi sistemleri denetimi gibi işlevlerde çalışmakta ve önerilerde bulunmaktadır. Bu işlevler hem metinsel hem de

metinsel olmayan verilerin veri madenciliğini içeren yapay zekâ ve makine öğrenmesi teknolojilerinden önemli ölçüde etkilenmiştir. Örneğin, sosyal medya ve diğer metinsel veri kaynaklarıyla birleştirilmiş kurumsal belgeler; bir kuruluşun gizlilik ve güvenlik, iş kanunlarına uygunluk veya sektöre özgü düzenlemelere uygunluk gibi düzenleyici gereksinimlere bağlılığını değerlendirmek için zengin bir bilgi kaynağı olabilir.

İç denetim, bilgi sistemleri denetimi veya kurumsal risk yönetimi gibi işlevler, iç denetçilerin çeşitli kaynaklardan gelen bol miktarda yazılı belgeyi incelemesini gerektirmektedir. Bu durumu Ernst & Young'dan Daniel Torpey ve Vincent Walden (2009) çalışmalarında;

- Kurumsal riski çevreleyen tüm veri kaynakları yelpazesini daha verimli bir şekilde ele almak için iç denetçiler artık yapılandırılmamış veri veya metin analizi araçlarını iş planlarına dahil edebilir ve
- Metin analizi araçları, kontrollerin veya iş uygulamalarının adli incelemesinin bir parçası olarak veya fiili bir soruşturma sırasında riske dayalı bir iç denetim kapsamında kullanılabilir şekilde ifade etmişlerdir.

Daniel Torpey ve Vincent Walden (2009), çoğu ticari işlemin veya olayın e-posta işlemlerine dayalı olduğunu söylemektedir. E-postalar zengin meta veriler (kaynak, sürüm ve erişim tarihi gibi veriler hakkında depolanan bilgiler) içermektedir. Bu meta veriler, iç denetçiler için önemli bir bilgi kaynağıdır. Ayrıca bu araştırmacılar, metin analitiği teknolojilerinin kullanımına ilişkin örnekler vererek ve metin madenciliği araçlarının bir kuruluş içinde riskleri anlamak ve anormallikleri belirlemek için proaktif olarak da kullanılabilceğini belirtmektedir.

Küresel bir teknoloji firmasında iç denetim direktörü örneğinden bahseden Daniel Torpey ve Vincent Walden (2009), ABD Yurtdışı Yolsuzluk Uygulamaları Yasası ile bağlantılı uygulama faaliyetlerinde son zamanlarda yaşanan artış bağlamında, uyumluluk riskini değerlendirmek ve satışlar için ihlallerin önlenmesine yardımcı olmak amacıyla metin analizi araçlarını kullanmayı önermişlerdir. Ayrıca iç denetim metodolojilerini metin analitiği çevresinde birleştirerek, denetçilerin proaktif risk çabalarını geliştirebileceğini ve hizmet verdikleri müşteriler için iş performansını potansiyel olarak iyileştirebileceğini ifade etmişlerdir.

5. SONUÇ

Yapay zekâ ve makine öğrenmesi yaklaşımlarıyla birleştirilen metin madenciliği, muhasebecilerin ve denetçilerin çalışmalarını bugün olduğundan çok daha fazla otomatikleştirmek için önemli fırsatlar

sunmaktadır. Bu teknolojilerin kullanımı yalnızca muhasebe mesleğinde gelişiyor olsa da bunların muhasebecilerin ve denetçilerin çalışma şeklini de önemli ölçüde değiştirmesi beklenmektedir. Bu değişikliklerden bazıları aşağıdaki gibidir:

- Muhasebe İşlemlerinin Daha Fazla Robotlaştırılması: Önemli verimlilik kazanımları sağlarken, verilerin toplanması ve kaydedilmesi yerine verilerin analizine daha fazla zaman harcanması anlamına gelmektedir.
- Daha Fazla Bilginin Analiz Edilmesi: Daha fazla doğruluk, daha iyi ön görümler ve daha iyi karar verme anlamına gelmektedir.
- İç Kontrollerin ve Denetimlerin Daha Proaktif ve Sürekli İzlenmesi: Anormalliklerin veya risklerin erken bir şekilde tespit edilmesi ve iyileştirilmiş iş performansı anlamına gelmektedir.

Ayrıca metin madenciliğinin muhasebe alanında kullanımı çok günceldir. Metin madenciliğinin kuruluşlar ve ilgili meslek mensupları tarafından benimsenmesinde;

- Veri erişiminde gelişmeler,
- Çeşitli disiplinlerden araştırmacılar arasında artan iş birliği ve
- Yapay zekâ, makine öğrenmesi ve büyük veri analitiği kaynaklarının artan kullanılabilirliği gibi etmenlerinde katkısı olacağı düşünülmektedir.

Çalışma, metin madenciliğinin ne olduğuna ve muhasebe alanında nasıl kullanılacağına ilişkin soruları yanıtlamaya yardımcı olmaktadır. Ayrıca kuruluşlara ve kişilere detaylı bilgiler vermek amacıyla, muhasebe otomasyonu, denetim otomasyonu, vergi otomasyonu ve iş danışmanlığı otomasyonuna yönelik açıklamalar yapılmış ve muhasebe alanında metin madenciliği uygulama örnekleri verilmiştir. Muhasebe alanında metin madenciliği teknolojisinin, önümüzdeki yıllarda oynayacağı rol dikkate alındığında, bu çalışmanın kuruluşlara ve kişilere önemli faydalar sağlayacağı düşünülmektedir. Son olarak gelecek dönemlerde finansal raporlama işlemlerinde metin madenciliğine yönelik doğal dil işleme teknolojisi kullanımının, araştırmacıların gündeminde olması beklenmektedir. Metin madenciliğinin muhasebe alanında gelişimini ve ilerlemesini sağlayacak en önemli unsurun, sektörler arası ve disiplinler arası iş birliğinin doğru ve uygulanabilir temeller üzerine kurulması gerektiğidir. Bu çalışmanın benzer ya da daha kapsamlı çalışmalara örnek teşkil ederek farklı alanlarda da kullanılabileceği öngörülmektedir.

KAYNAKÇA

Alarcon, J., Fine T. & Ng, C. (2019). Accounting AI and machine learning: Applications and challenges. *Accounting and Technology: PICPA's Guide to an Evolving Profession*, 3-7. Çevrimiçi <http://onlinedigeditions.com/publication/?m=14667&i=583202&p=0>.

Aldhizer, G. R. (2017). Visual and text analytics. *The CPA Journal*, 87(6), 30-33. Çevrimiçi <https://www.cpajournal.com/2017/06/20/visual-textanalytics/>.

Blake, C. (2011). Text mining. *Annual Review of Information Science and Technology*, 45(1), 121-155. <http://doi.org/10.1002/aris.2011.1440450110>.

Brown, B. & Rainey, S. (2018). Driving faster, more accurate and more beneficial tax decisions. *IBM*. Çevrimiçi <https://www.ibm.com/blogs/watson/2018/04/driving-faster-more-accurate-and-more-beneficial-tax-decisions/>.

Chopra, S., Auli, M. & Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California, United States. Çevrimiçi <https://www.aclweb.org/anthology/N16-1012/>.

Da Costa Pereira, C. & Tettamanzi, A. G. B. (2006). *An ontology-based method for user model acquisition*. Studies in fuzziness and soft computing: Soft computing in ontologies and semantic web, Springer.

Davis, M., Vashisth, S., Emmott, S. & Brethenoux, E. (2018). *Market guide for text analytics (ID: G00361404)*. Retrieved From Gartner Database.

Deloitte Harnesses The Power of Kira for Lease Accounting Contract Review. (2020). *Kira Systems*. Retrieved. Çevrimiçi <https://kirasystems.com/resources/case-studies/deloitte/>.

Feldman, R. & Dagan, I. (1995). Knowledge discovery in textual databases. *The First International Conference on Knowledge Discovery and Data Mining (KDD-95)*. Montreal, Quebec, Canada.

Internal Revenue Service (2022). Advance data and analytics. Çevrimiçi <https://www.irs.gov/about-irs/strategic-goals/advance-data-analytics>.

H&R Block (2017). H&R Block with IBM Watson reinventing tax preparation. Çevrimiçi <https://www.hrblock.com/tax-center/newsroom/around-block/partnership-with-ibm-watson-reinventing-tax-prep/>.

Hearst, M. A. (2003). What is text mining? [Unpublished Essay]. Çevrimiçi <http://people.ischool.berkeley.edu/~hearst/text-mining.html>.

Keikha, M., Razavian, N. S., Oroumchian, F. & Razi, H. S. (2008). *Document representation and quality of text: An analysis*. In Berry M. W. &

Castellanos M. (Eds.), Survey of Text Mining II: Clustering, Classification, and Retrieval. Springer.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8.

KPMG and IBM (2020). KPMG. Retrieved. Çevrimiçi <https://home.kpmg/xx/en/home/about/alliances/ibm.html>.

Kumar, B. S. & Ravi, V. (2016). A survey of the applications of text mining in the financial domain. *Knowledge-Based Systems*, 114, 128-147. <https://doi.org/10.1016/j.knosys.2016.10.003>.

Kumar, Y. J., Goh, O. S., Basiron, H., Choon, N. H. & Suppiah, P. C. (2016). A review on automatic text summarization approaches. *Journal of Computer Science*, 12(4), 178-190. <https://doi.org/10.3844/jcssp.2016.178.190>.

Lewis, C. & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, 49(5), 587-615. <https://doi.org/10.1080/00014788.2019.1611730>.

Lowa State University. (2018). Textual analytics for accounting and auditing. Çevrimiçi <https://www.ivybusiness.iastate.edu/files/2018/12/Janvrin-Textual-Analysis-Presentation-Dec-14-2018.pdf>.

Nagarkar, S. & Kumbhar, R. (2015). Text mining: An analysis of research published under the subject category 'information science library science' in web of science database during 1999-2013. *Library Review*, 64(3), 248-262. <https://doi-org.libproxy.temple.edu/10.1108/LR-08-2014-0091>.

Nallapati, R., Zhou, B., Dos Santos, C. N., Gulcehre, C. & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. *The SIGNLL Conference on Computational Natural Language Learning (CoNLL)*. Çevrimiçi <https://arxiv.org/abs/1602.06023>.

Pejic-Bach, M., Krstic Z., Seljan, S. & Turulja, L. (2019). Text mining for big data analysis in financial sector: A literature review. *Sustainability*, 11(5), 1277. <http://doi.org/10.3390/su11051277>.

Rajman, M. & Vesely, M. (2004). From text to knowledge: Document processing and visualization: A Text Mining Approach. In S. Sirmakessis (Ed.), *Text Mining and its Applications – Results of the NEMIS Launch Conference*. Springer.

Random House Kernerman Webster. (2020). Coreference. In *Random House Kernerman Webster's College Dictionary*. Çevrimiçi <https://www.thefreedictionary.com/coreference>.

- Rush, A. M. Chopra, S. & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *Cornell University Library*. Çevrimiçi <https://arxiv.org/abs/1509.00685>.
- Sharda, R., Delen, D. & Turban, E. (2014). *Business intelligence: A managerial perspective on analytics (3rd ed.)*. Pearson Prentice Hall.
- Song, S., Huang, H. & Ruan, T. (2019). Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications*, 78, 857-875. <https://doi.org/10.1007/s11042-018-5749-3>.
- Sparck-Jones, K. (1999). Automatic summarizing: Factors and directions. In I. Mani & M. T. Maybury (Eds.), *Advances in automated text summarization*, 1-12. MIT Press. Çevrimiçi <https://www.cl.cam.ac.uk/archive/ksj21/ksjdigipapers/sumbook99.pdf>.
- Sun, T. & Vasarhelyi, M. A. (2018). Embracing textual data analytics in auditing with deep learning. *International Journal of Digital Accounting Research*, 18, 49-67. https://doi.org/10.4192/1577-8517-v18_3.
- Tang, J., Hong, M., Zhang, D. L. & Li, J. (2008). Information extraction: Methodologies and applications. In H. do Prado & E. Farneda (Eds.), *Emerging technologies of text mining: Techniques and applications*, 1-33. IGI Global.
- Torpey, D. & Walden, V. (2009). Accounting for words; Text analytics technology may help internal auditors uncover hidden risks and gain greater insight on business performance. *Internal Auditor*, 66(4), 40-44.
- Torres-Moreno, J. M. (2014). *Automatic Text Summarization*. John Wiley & Sons, Inc.
- Tuffery, S. (2011). *Text mining*. In Wiley Series in Computational Statistics, Data Mining and Statistics for Decision Making, 627-636. John Wiley & Sons, Ltd.
- Verma, S. & Nidhi, V. (2019). Extractive summarization using deep learning. *Cornell University Library*. Çevrimiçi <http://libproxy.temple.edu/login?url=https://searchproquest.com.libproxy.temple.edu/docview/2075709212?accountid=14270>.
- Wang, G. (2019). Tech talk: Intuit's AI-powered tax knowledge engine boosts filers' confidence. *Intuit Blog*. Çevrimiçi <https://www.intuit.com/blog/social-responsibility/tech-talk-intuits-ai-powered-tax-knowledge-engine-boosts-filers-confidence/>.
- Zhang, Y., Er, M. J. & Pratama, M. (2016). Extractive document summarization based on Convolutional neural networks. *IECON 2016-42nd*

Annual Conference of the IEEE Industrial Electronics Society. Florence, 918-922.