



**FİNANSAL PERFORMANS TAHMİNİNDE  
METİN MADENCİLİĞİNİN KULLANIMI: BİST  
İMALAT SANAYİ İŞLETMELERİNDE  
BİR ARAŞTIRMA**

**Doktora Tezi**

**Sait PEKİN**

**Eskişehir 2020**

**FİNANSAL PERFORMANS TAHMİNİNDE METİN MADENCİLİĞİNİN  
KULLANIMI: BİST İMALAT SANAYİ İŞLETMELERİNDE BİR ARAŞTIRMA**

**Sait PEKİN**

**DOKTORA TEZİ**

**İşletme Anabilim Dalı**

**Danışman: Prof. Dr. Saime ÖNCE**

**Eskişehir**

**Anadolu Üniversitesi**

**Sosyal Bilimler Enstitüsü**

**Şubat 2020**

## JÜRİ VE ENSTİTÜ ONAYI

Sait PEKİN'in "Finansal Performansın Tahmininde Metin Madenciliğinin Kullanımı: BİST İmalat Sanayi İşletmelerinde Bir Araştırma" başlıklı tezi 06 Şubat 2020 tarihinde, aşağıdaki jüri tarafından Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca İşletme Anabilim Dalı Muhasebe Bilim Dalında, Doktora tezi olarak değerlendirilerek kabul edilmiştir.

Üye (Tez Danışmanı) : Prof. Dr. Saime ÖNCE

Üye :Prof. Dr. Tunç KÖSE

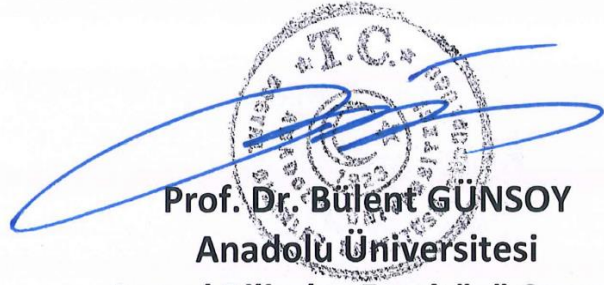
Üye :Prof. Dr. Vedat EKERGİL

Üye :Prof. Dr. Ayşe Banu BAŞAR

Üye :Dr.Öğr.Üy. Ahmet YÜCEL

İmza





Prof. Dr. Bülent GÜNŞOY  
Anadolu Üniversitesi  
Sosyal Bilimler Enstitüsü Sosyal  
Bilimler Enstitüsü Müdürü

## ÖZET

### FINANSAL PERFORMANS TAHMİNİNDE METİN MADENCİLİĞİNİN KULLANIMI: BİST İMALAT SANAYİ İŞLETMELERİNDE BİR ARAŞTIRMA

Sait PEKİN

İşletme Anabilim Dalı

Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, Şubat, 2020

Danışman: Prof. Dr. Saime ÖNCE

Ekonomik ve teknolojik gelişmenin etkisiyle değişen finansal raporlama yaklaşımı ile birlikte sayısal veriler yanında metin verilerde finansal raporlarda ve diğer işletme açıklamalarında yer almaya başlamıştır. Bu durum geleneksel yöntemlerle analiz edilmesi mümkün olmayan metin veri içeren veri yığınları sorununu beraberinde getirmiştir. Bu noktadan hareketle bu çalışmada finansal raporlarda yer alan metin verilerin analizi üzerine odaklanılmış ve işletmelerin finansal raporlama stratejilerinin finansal performansa göre şekillenip şekillenmediği araştırılmıştır. Çalışmanın temel amacı işletmelerin finansal performansına ilişkin analizlerde sayısal veriler yanında metin verilerin de kullanılabilmesini sağlayacak pratik analiz teknikleri geliştirmektir. Bu amaca yönelik olarak finansal performans ile faaliyet raporlarındaki metin veriler arasında bir ilişki bulunup bulunmadığı sorusuna cevap aranmıştır. Ayrıca çalışmada; işletmenin gelecekteki finansal performansı hakkında faaliyet raporlarındaki açıklamalarda bilgi bulunup bulunmadığı araştırılmıştır. Araştırma yöntemi metin madenciliğidir. Araştırma modeli olarak ise veri madenciliği sınıflandırma algoritmalarından destek vektör makineleri algoritması kullanılmıştır. Araştırmanın veri seti BİST imalat sanayi sektöründe faaliyet gösteren işletmelerin 2010-2017 yılları arasındaki 8 yıllık dönemde yayınladıkları faaliyet raporları ve finansal performans göstergeleridir. Araştırma sonuçları finansal performans ile faaliyet raporlarındaki metin veriler arasında bir ilişki bulunduğunu göstermektedir. Bununla birlikte araştırma sonuçlarına göre, faaliyet raporlarındaki metin veri içeren açıklamalarda işletmenin gelecekteki finansal performansı hakkında bilgi bulunduğu söylenebilir. Ayrıca araştırma sonuçları finansal raporlarda sunulan metin verilerin analizinde metin madenciliğinin kullanışlı ve pratik bir yöntem olduğunu göstermektedir.

**Anahtar Kelimeler:** Metin madenciliği, finansal performans, veri madenciliği

## **ABSTRACT**

### **USE OF TEXT MINING IN PREDICTION OF FINANCIAL PERFORMANCE: A RESEARCH IN BIST MANUFACTURING INDUSTRY**

Sait PEKİN

Department of Business Administration

Anadolu University, Graduate School of Social Sciences, February, 2020

Adviser: Prof. Dr. Saime ÖNCE

Along with the financial reporting approach that has changed due to economic and technological development, textual data as well as numerical data have been included in financial reports and other disclosures of companies. This has led to the problem of data stacks containing text data that cannot be analyzed by conventional methods. Starting from this point, this study focuses on the analysis of text data in financial reports and examines whether financial reporting strategies of companies are shaped according to financial performance. The main objective of the study is to develop practical analysis techniques that will enable the use of text data as well as numerical data in the analysis of financial performance of companies. For this purpose, it is sought to answer the question whether there is a relationship between financial performance and the text data in the annual reports. In addition, the study investigated whether there is information about the future financial performance of the companies in the explanations in the annual reports. The research method is text mining. As the research model, support vector machines algorithm, which is one of the data mining classification algorithms, was used. The data set of the research is the annual reports and financial performance indicators published by the companies in the BIST manufacturing industry sector in the 8 yearly period between 2010-2017 years. The results of the research show that there is a relationship between financial performance and text data in annual reports. On the other hand, according to the results of the research, it can be said that the explanations containing the text data in the annual reports contain information about the future financial performance of the companies. In addition, the results of the research show that text mining is a useful and practical method in the analysis of text data presented in financial reports.

**Keywords:** Text mining, financial performance, data mining

## ÖNSÖZ

Bu çalışmanın gerçekleştirilmesinde bilgi ve tecrübesiyle değerli katkılar sunan, doktora sürecimin her aşamasında beni cesaretlendiren, anlayışını ve desteğini hiçbir zaman esirgemeyen tez danışmanım, değerli hocam Prof. Dr. Saime ÖNCE'ye, fikirleri ve önerileriyle çalışmanın bu halini almasında önemli katkıları olan Prof. Dr. Ayşe Banu BAŞAR'a ve Prof. Dr. Tunç KÖSE'ye, çalışmanın uygulama bölümünde desteğini esirgemeyen arkadaşım Dr. Öğr. Üyesi Ahmet YÜCEL'e teşekkürlerimi sunarım.

Doktora eğitimim ve tez çalışmam esnasındaki sabrı, desteği ve anlayışı için sevgili eşim Ümran PEKİN'e, varlıkları ile hayatımı güzelleştiren çocuklarım Muhammed ve Ömer'e, bugünlere gelmemde emeklerinin karşılığını ödeyemeyeceğim anneme, babama ve kardeşlerime şükranlarımı sunarım.

06.02.2020

## ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmamın Anadolu Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

  
Sait PEKİN

## İÇİNDEKİLER

|                                                               |     |
|---------------------------------------------------------------|-----|
| BAŞLIK SAYFASI.....                                           | i   |
| JÜRİ VE ENSTİTÜ ONAY SAYFASI .....                            | ii  |
| ÖZET .....                                                    | iii |
| ABSTRACT.....                                                 | iv  |
| ÖNSÖZ.....                                                    | v   |
| ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....              | vi  |
| İÇİNDEKİLER .....                                             | vii |
| TABLolar DİZİNİ .....                                         | x   |
| ŞEKİLLER DİZİNİ.....                                          | xi  |
| KISALTMALAR DİZİNİ .....                                      | xii |
| 1. GİRİŞ.....                                                 | 1   |
| 1.1. Sorun.....                                               | 1   |
| 1.2. Amaç.....                                                | 2   |
| 1.3. Araştırma Soruları .....                                 | 2   |
| 1.4. Önem.....                                                | 3   |
| 1.5. Sınırlılıklar .....                                      | 3   |
| 2. ALAN YAZIN .....                                           | 4   |
| 2.1. Finansal Raporlama .....                                 | 4   |
| 2.1.1. Finansal tablolar .....                                | 6   |
| 2.1.2. Faaliyet raporları .....                               | 6   |
| 2.1.3. Sürdürülebilirlik raporları .....                      | 10  |
| 2.1.4. Entegre raporlar.....                                  | 11  |
| 2.2. Finansal Performans .....                                | 13  |
| 2.2.1. Muhasebe bazlı finansal performans göstergeleri.....   | 13  |
| 2.2.1.1. Aktif kârlılık oranı (ROA-return on assets) .....    | 14  |
| 2.2.1.2. Öz kaynak kârlılık oranı (ROE–return on equity)..... | 14  |

|          |                                                              |    |
|----------|--------------------------------------------------------------|----|
| 2.2.1.3. | Net kâr marjı (ROS-return on sales).....                     | 14 |
| 2.2.1.4. | Hisse senedi başına kâr (EPS-earnings per share).....        | 14 |
| 2.2.2.   | Değer bazlı finansal performans göstergeleri .....           | 14 |
| 2.2.3.   | Piyasa bazlı finansal performans göstergeleri.....           | 15 |
| 2.3.     | Büyük Veri ve Veri Madenciliği .....                         | 15 |
| 2.3.1.   | Veri Madenciliği (Data Mining) .....                         | 17 |
| 2.3.1.1. | Veri madenciliği modelleri .....                             | 19 |
| 2.4.     | Metin Madenciliği (Text Mining) .....                        | 22 |
| 2.4.1.   | Metin Madenciliği Çalışma Alanları .....                     | 23 |
| 2.4.1.1. | Bilgiye erişim .....                                         | 23 |
| 2.4.1.2. | Bilgi çıkarımı.....                                          | 23 |
| 2.4.1.3. | Doğal dil işleme.....                                        | 24 |
| 2.4.1.4. | Web madenciliği.....                                         | 24 |
| 2.4.1.5. | Duygu analizi .....                                          | 25 |
| 2.4.2.   | Metin Madenciliğinin Adımları .....                          | 25 |
| 2.4.2.1. | Derlem (Corpus) oluşturulması .....                          | 26 |
| 2.4.2.2. | Metin ön işleme .....                                        | 26 |
| 2.4.2.3. | Metin dönüşümü .....                                         | 28 |
| 2.5.     | Muhasebe Alanında Yapılan Metin Madenciliği Çalışmaları..... | 28 |
| 3.       | YÖNTEM.....                                                  | 38 |
| 3.1.     | Araştırma Modeli.....                                        | 38 |
| 3.2.     | Araştırmanın Evren ve Örneklemi.....                         | 40 |
| 3.3.     | Veri Toplama Tekniği ve Araçları .....                       | 40 |
| 3.4.     | Veri Analizi .....                                           | 40 |
| 3.4.1.   | Tablo Arama Yöntemi .....                                    | 40 |
| 3.4.2.   | Vektör Uzay Modeli ve Binary Algoritması.....                | 41 |
| 3.4.3.   | Tekil Değer Ayrışımı (Singular Value Decomposition) .....    | 42 |

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| 3.4.4. Destek Vektör Makineleri (Support Vector Machines).....                                  | 42 |
| 3.4.5. Değişken Seçimi.....                                                                     | 45 |
| 3.4.5.1. Bağımlı değişken.....                                                                  | 45 |
| 3.4.5.2. Bağımsız değişkenler.....                                                              | 52 |
| 3.4.6. Verilerin Birleştirilmesi ve Eğitim ve Test Verisinin Oluşturulması                      | 53 |
| 3.4.7. Model Performansının ve Sonuçların Değerlendirilmesi .....                               | 54 |
| 4. BULGULAR VE YORUM.....                                                                       | 56 |
| 4.1. Giriş .....                                                                                | 56 |
| 4.2. Öz kaynak Kârlılığının Tahminine İlişkin Destek Vektör Makineleri<br>Analiz Sonuçları..... | 57 |
| 4.3. Net Kâr Marjının Tahminine İlişkin Destek Vektör Makineleri Analiz<br>Sonuçları.....       | 58 |
| 4.4. Hisse Başına Kârın Tahminine ilişkin Destek Vektör Makineleri Analiz<br>Sonuçları.....     | 59 |
| 5. SONUÇ .....                                                                                  | 61 |
| KAYNAKÇA.....                                                                                   | 65 |
| ÖZGEÇMİŞ                                                                                        |    |

## TABLÖLAR DİZİNİ

|                                                                                     |    |
|-------------------------------------------------------------------------------------|----|
| <b>Tablo 2.1:</b> Faaliyet raporunun içeriği .....                                  | 9  |
| <b>Tablo 2.2:</b> CRISP-DM veri madenciliği aşamaları ve uygulanacak işlemler ..... | 19 |
| <b>Tablo 3.1:</b> Arama tablosu örneği.....                                         | 41 |
| <b>Tablo 3.2:</b> Belge terim matrisi örneği.....                                   | 42 |
| <b>Tablo 3.3:</b> Temel Kernel fonksiyonları ve parametreleri .....                 | 44 |
| <b>Tablo 3.4:</b> Karışıklık matrisi .....                                          | 54 |
| <b>Tablo 4.1:</b> ÖKK karışıklık matrisi.....                                       | 57 |
| <b>Tablo 4.2:</b> ÖKK duyarlılık oranları.....                                      | 57 |
| <b>Tablo 4.3:</b> NKM karışıklık matrisi .....                                      | 58 |
| <b>Tablo 4.4:</b> NKM duyarlılık oranları .....                                     | 59 |
| <b>Tablo 4.5:</b> HBK karışıklık matrisi.....                                       | 59 |
| <b>Tablo 4.6:</b> HBK duyarlılık oranları .....                                     | 60 |

## ŞEKİLLER DİZİNİ

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| Şekil 2.1: Finansal raporlamanın gelişimi.....                             | 5  |
| Şekil 2.2: CRISP-DM veri madenciliği aşamaları.....                        | 18 |
| Şekil 2.3: Veri madenciliği modelleri.....                                 | 20 |
| Şekil 3.1: Araştırma süreci.....                                           | 39 |
| Şekil 3.2: İki sınıflı bir problem için hiper-düzlemler.....               | 43 |
| Şekil 3.3: Optimum hiper-düzlem ve destek vektörleri.....                  | 43 |
| Şekil 3.4: Doğrusal olarak ayrılabilen veri setleri için hiper-düzlem..... | 44 |
| Şekil 3.5: ÖKK'nın grafiksel özeti.....                                    | 48 |
| Şekil 3.6: NKM'nın grafiksel özeti.....                                    | 49 |
| Şekil 3.7: HBK'nın grafiksel özeti.....                                    | 51 |
| Şekil 3.8 : Analize hazır hale getirilmiş verilerin görünümü.....          | 54 |

## KISALTMALAR DİZİNİ

|                |                                                              |
|----------------|--------------------------------------------------------------|
| <b>BİST</b>    | : Borsa İstanbul                                             |
| <b>CDP</b>     | : Carbon Disclosure Project                                  |
| <b>CRISPDM</b> | : Cross-Industry Standard Process for Data Mining            |
| <b>DVM</b>     | : Destek Vektör Makineleri                                   |
| <b>EPS</b>     | : Earnings Per Share                                         |
| <b>EVA</b>     | : Economic Value Added                                       |
| <b>GRI</b>     | : Global Reporting Initiative                                |
| <b>HBK</b>     | : Hisse Başına Kâr                                           |
| <b>IASB</b>    | : Uluslararası Muhasebe Standartları Kurulu                  |
| <b>IASC</b>    | : Uluslararası Muhasebe Standartları Komitesi                |
| <b>IIRC</b>    | : International Integrated Reporting Council                 |
| <b>İTÜ NLP</b> | : İstanbul Teknik Üniversitesi Doğal Dil İşleme              |
| <b>KAP</b>     | : Kamuyu Aydınlatma Platformu                                |
| <b>ML</b>      | : Machine Learning                                           |
| <b>MVA</b>     | : Market Value Added                                         |
| <b>NKM</b>     | : Net Kâr Marjı                                              |
| <b>OECD</b>    | : The Organisation for Economic Co-operation and Development |
| <b>ÖKK</b>     | : Öz kaynak Kârlılığı                                        |
| <b>ROA</b>     | : Return on Assets                                           |
| <b>ROE</b>     | : Return on Equity                                           |
| <b>ROS</b>     | : Return on Sales                                            |
| <b>SEC</b>     | : ABD Menkul Kıymetler ve Borsa Komisyonu                    |
| <b>SEMMA</b>   | : Sample, Explore, Modify, Model, and Assess                 |
| <b>SOM</b>     | : Self Organizing Map                                        |
| <b>SVM</b>     | : Support Vektör Machines                                    |
| <b>TDA</b>     | : Tekil Değer Ayırışımı                                      |
| <b>TMS</b>     | : Türkiye Muhasebe Standartları                              |
| <b>UNGC</b>    | : United Nations Global Compact                              |
| <b>US GAAP</b> | : ABD, Genel Kabul Görmüş Muhasebe Standartları              |

## 1. GİRİŞ

Muhasebe bilgi kullanıcıları genellikle işletmelerin finansal tablolarda yayınladıkları sayısal verilere odaklanmaktadır. Ancak son dönemlerde finansal raporlamada meydana gelen önemli değişmelerle birlikte sayısal veriler yanında finansal raporlardaki metin veriler de değerli bilgi kaynağı haline gelmiştir. İşletmeler günümüzde finansal tablolar yanında işletme ile ilgili önemli nitel bilgileri içeren faaliyet raporları, kurumsal sürdürülebilirlik raporları, entegre raporlar vb. raporlar yayınlamaya başlamışlardır. Bu raporların bir kısmı yasal zorunluluktan, bir kısmı ise işletmenin kendi isteğine bağlı olarak yayınlanmaktadır. Ayrıca gelişen teknoloji ile birlikte işletmelerin internet siteleri, haber siteleri, sosyal medya ve benzeri ortamlar işletmeler ile ilgili önemli metin veriler barındırmaktadır. Finansal veri kaynaklarındaki bu gelişmelere rağmen muhasebe araştırmalarının birçoğunda hala finansal raporlardaki sayısal veriler önemli yer kaplarken metin veriler ile ilgili yeterince çalışma yapılmamaktadır.

Bu çalışma Türkçe yazılmış finansal raporlarda yer alan metin verilerin muhasebe analizlerinde kullanılabilmesinde veri ve metin madenciliği araçlarının potansiyelini değerlendirmeyi amaçlamaktadır. Özellikle işletmelerin finansal performansına yönelik analizlerde metin verilerin kullanımı üzerine odaklanılmıştır. Bu kapsamda çalışmada sayısal veriler ile metin verilerin birlikte değerlendirebileceği bir yaklaşım geliştirilmeye çalışılmıştır. Çalışma üç bölüm olarak şekillendirilmiştir.

Birinci bölümde finansal raporlama, finansal performans, veri madenciliği ve metin madenciliği konuları ele alınmıştır. Ayrıca muhasebe alanında metin madenciliği ile ilgili daha önce yapılmış çalışmalar anlatılmıştır.

İkinci bölümde araştırmanın yöntemi ve veri analizinde kullanılacak algoritmalar anlatılmıştır.

Üçüncü bölümde ise, analiz sonuçlarına ilişkin bulgular sunulmuş ve analiz sonuçları yorumlanmıştır.

### 1.1. Sorun

Günümüzde teknolojik gelişme ile birlikte veri oluşumu baş döndürücü bir hızda meydana gelmektedir. Verilerin büyük kısmı video, ses, resim ve metin gibi yapılandırılmamış veri olarak saklanmaktadır. Özellikle metin şeklindeki veri işletme paydaşları için önemli bilgiler içermektedir. Ancak metin şeklindeki verilerin geleneksel yöntemlerle analiz edilmesi ve faydalı bilgi elde edilmesi neredeyse imkânsızdır. Zira

veri boyutu çok hızlı bir şekilde büyümektedir. Büyük boyuttaki veriyi ise geleneksel yöntemlerle analiz etmek ve ondan faydalı bilgi elde etmek mümkün değildir. Bu bağlamda günümüzde metin verilerin analizi için metin madenciliği önemli bir araç haline gelmiştir. Ülkemizde Türkçe yazılmış metinler için metin madenciliği çalışmaları çeşitli alanlarda yapılmakla birlikte çalışmaların henüz istenen seviyede olduğu söylenemez. Diğer alanlarda olduğu gibi muhasebe alanında da metin veriler önemli bir yer tutmaktadır. Ancak bu metinler sayısal veriler kadar hızlı ve kolay analiz edilememektedir. Yapılan çalışmalarda geleneksel yöntemler kullanılmakta ve kısıtlı bilgi elde edilmektedir. Metin verilerin analizi, çok fazla emeği ve zamanı gerektirmekte ve çok büyük bir maliyeti beraberinde getirmektedir. Bu açıdan muhasebe açıklamalarındaki metin verilerin analizi önemli bir sorundur. Bu sorunun çözümü için bir araç olabilecek büyük veri teknolojisi, veri madenciliği ve metin madenciliği ile ilgili, İngilizce ve diğer dillerde oluşturulmuş metin veriler için çalışmalar bulunmakla birlikte Türkçe metin veriler için muhasebe alanında bir çalışmaya rastlanılmamıştır.

## **1.2. Amaç**

Çalışmanın amacı işletmelerin finansal performansına ilişkin analizlerde sayısal veriler yanında metin verilerin de kullanılabilmesini sağlayacak analiz teknikleri geliştirmektir. Bu motivasyonu sağlayan ise işletmelerin yayınladıkları metin açıklamaların finansal performansla ilgili önemli bilgiler içerdiği düşüncesidir. Buradan hareketle finansal performans ile işletmelerin metin açıklamaları arasında bir ilişki bulunup bulunmadığı ve sayısal veriler ile metin verilerin birlikte analiz edilip edilemeyeceği araştırılmıştır.

## **1.3. Araştırma Soruları**

“İşletmelerin finansal performansı ile faaliyet raporlarındaki metin veriler arasında bir ilişki var mıdır?” sorusu çalışmanın temel araştırma sorusunu oluşturmaktadır. Diğer araştırma soruları ise aşağıdaki şekildedir:

1. Faaliyet raporundaki açıklamalar işletmenin gelecekteki finansal performansı hakkında bilgi içermekte midir?
2. Faaliyet raporlarının analizinde metin madenciliği kullanılabilir mi?
3. İşletmelerin finansal raporlarda yayınladıkları sayısal veri ile metin veri birlikte değerlendirilebilir mi?

#### **1.4. Önem**

Çalışma, muhasebe alanında Türkçe metinler üzerine yapılmış ilk metin madenciliği çalışması olması, ayrıca sayısal veriler ile metin verilerin birlikte değerlendirildiği ender çalışmalardan biri olması nedeniyle önem arz etmektedir. Çalışma, veri ve metin madenciliği alanındaki teknolojik gelişmelerin muhasebe çalışmalarında uygulanabildiğinin bir göstergesidir. Ayrıca muhasebe ve finans alanındaki akademik çalışmalarda dikkati metin veriler üzerine çekme potansiyeli taşımaktadır. Bu nedenle çalışmanın muhasebe literatürüne önemli katkıları olacağı düşünülmektedir.

İşletme performansının tahmininde, sayısal göstergelerle birlikte metin açıklamaların kullanılmasında metin madenciliğinin önemli bir araç olduğu düşüncesi bağlamında metin madenciliği kullanılarak atıl olarak duran büyük miktardaki metin verinin kullanılmasını sağlayacak bir yöntem geliştirmek suretiyle, finansal performansa yönelik analizlerde farklı bir bakış açısı sunulması da diğer önemli bir husustur. Dolayısıyla faaliyet raporlarındaki metinlerden işletmelerin gelecekteki finansal performansının tahmin edilmesini sağlayacak bilgi elde edilmesi mümkün olacaktır.

#### **1.5. Sınırlılıklar**

Türkçe için özel olarak geliştirilmiş metin madenciliği programlarının bulunmaması ve Türkçe gövdeleme algoritması geliştirme çalışmalarının yeterli düzeyde olmaması nedeniyle otomatik gövdeleme algoritmasının kullanılmaması çalışmanın bir sınırlılığıdır.

Analiz dönemi 2010 ve 2017 yılları arasını kapsamakla birlikte, her işletmenin bu tarihler arasında düzenli olarak faaliyet raporu yayınlamaması çalışmanın diğer sınırlılığıdır.

## 2. ALAN YAZIN

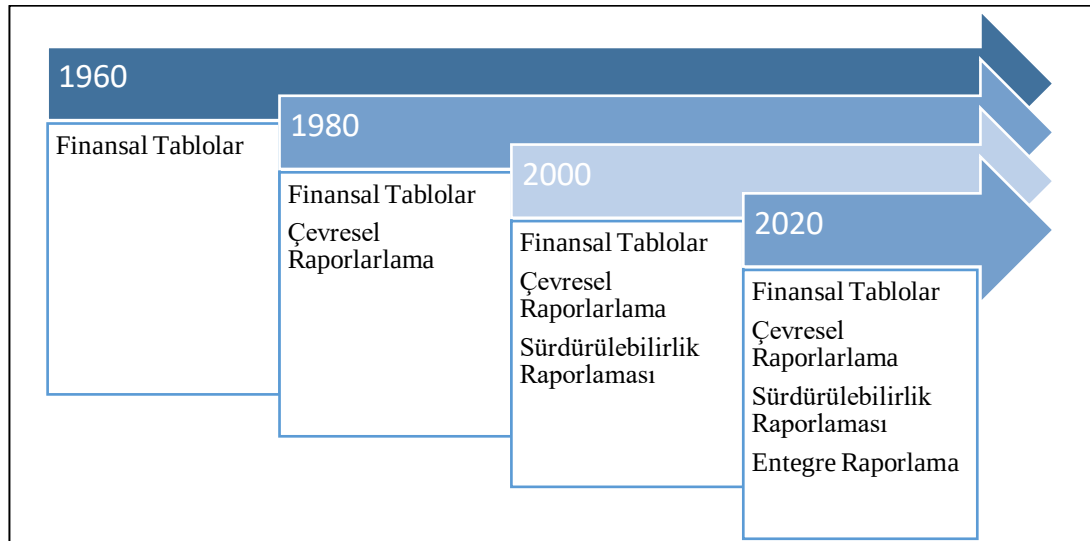
### 2.1. Finansal Raporlama

Geleneksel olarak muhasebe sürecinde finansal raporlar deyince akla ilk olarak finansal tablolar gelmektedir. Finansal tablolar işletme dışına işletme ile ilgili finansal bilgi sunmanın temel araçlarıdır ve işletmenin finansal durumunu ve faaliyet sonuçlarını para ile ifade edilmiş şekilde sunmaktadırlar. Ancak işletme ile ilgili bilgilerin tümünün bu finansal tablolarda sunulması mümkün değildir. Muhasebe sürecinde üretilen finansal bilgilere dayanılarak elde edilen temel finansal tablolara ek olarak finansal muhasebe süreciyle doğrudan ya da dolaylı ilişkili olan bilgilerin de raporlanması gerekir. Bu da finansal tabloların yanı sıra muhasebe süreciyle doğrudan ya da dolaylı ilişkili olan bilgilerin iletilmesini sağlayan araçları da kapsayan finansal raporlamayı ifade eder. Bu açıdan finansal raporlama daha geniş kapsamlıdır ve finansal tablolarla birlikte finansal tablolarda sunulmayan finansal olmayan verileri de kapsar (Cemalcılar ve Önce, 1999, s. 13-15).

Finansal raporlamanın temelleri ilk tarım faaliyetlerinin gerçekleştirildiği yıllarda tutulan muhasebe kayıtlarına kadar uzanmaktadır. Bununla birlikte, bu kayıtların yayımlanmasının yasal zorunluluk haline gelmesi 1929 Dünya Ekonomik Bunalımı sonrasında gerçekleşmiştir. Yaşanan bu büyük kriz yatırımcı güvenini önemli ölçüde zedelemiş ve yatırımcı güveninin yeniden tesis edilmesi için başta ABD olmak üzere dünyanın birçok ülkesinde yasal düzenlemelere gidilmiştir. ABD, Genel Kabul Görmüş Muhasebe Standartları'nı (US GAAP) oluşturmuş ve ABD Menkul Kıymetler ve Borsa Komisyonu'nu (SEC) kurarak payları borsada işlem gören işletmelerin finansal bilgilerini yayımlamasını zorunlu hale getirmiştir. İşletmelerin finansal bilgilerine ilişkin şeffaflığın yatırımcı güvenini artırdığını gören diğer ülkelerin de benzer düzenlemelere gitmesi sonucunda finansal raporlama kısa sürede dünya genelinde yaygınlaşmıştır. Ancak, farklı ülkelerden işletmelerin farklı prensiplere göre raporlama yapması yatırımcılar açısından karşılaştırma yapmayı zorlaştırmıştır. Yatırımcıların ve yasal düzenleyicilerin daha sağlıklı karşılaştırma yapabilmesi ve uluslararası yatırımın ve ticaretin artırılabilmesi amacıyla 1973 yılında Uluslararası Muhasebe Standartları Komitesi (IASC) tarafından Uluslararası Muhasebe Standartları geliştirilmiştir (Aras ve Sarıoğlu, 2015, s. 21). 1990'lı yılların sonlarına doğru IASC'nin yapısal bir değişikliğe ihtiyacı olduğuna karar verilmiş ve komitenin tüm sorumlulukları 2001 yılında, Londra'da kurulan Uluslararası Muhasebe

Standartları Kurulu (IASB)'na devredilmiştir. Kurul, Uluslararası Muhasebe Standartları'nı güncelleyerek bugün dünyanın birçok ülkesinde ortak raporlama dili olarak kullanılan Uluslararası Finansal Raporlama Standartlarına dönüştürmüştür.

Finansal raporlamanın ilk uygulaması 1930'lu yıllarda işletmenin geçmiş finansal bilgilerine odaklanan finansal tabloların sunulmasıyla başlamıştır. 1980'li yıllardan itibaren ise, işletme faaliyetlerinin sosyal ve çevresel etkileri yoğun olarak tartışılmaya başlanmış ve finansal tabloların sosyal ve çevresel etkileri içermemesi nedeniyle işletmelerin gerçek durumunu yansıtmada yetersiz kaldığı görülmüştür. Bu nedenle işletmeler finansal tabloların yanında yönetimin açıklamalarına, çevresel ve kurumsal yönetim ve ücretlendirmeye ilişkin bilgileri de içeren raporlar yayınlanmaya başlamıştır (Aydın, 2015, s. 21). Bu raporlama yöntemi ilk olarak çevresel raporlama ile başlamış, zaman içerisinde gelişerek 2000'li yıllardan itibaren işletmelerin sosyal, ekonomik ve çevresel faaliyetlerinin sonuçlarının bir bütün olarak yer aldığı sürdürülebilirlik raporlamasına dönüşmüştür. Finansal raporlamanın anlatılan süreç içerisinde gelişimi görüldüğü üzere işletmelerin birbirinden bağımsız birçok rapor yayınlamasına ve finansal raporlamanın dağınık bir görüntü vermesine neden olmuştur. Son dönemlerde ise finansal raporlamada ortaya çıkan bu dağınık yapının ortadan kaldırılması ve işletmelerin finansal bilgileri ile finansal olmayan bilgileri arasındaki bağlantının daha doğru bir şekilde bilgi kullanıcılarına yansıtılabilmesi için işletmenin finansal bilgileri ile finansal olmayan bilgilerinin birlikte yer aldığı entegre raporlamaya doğru bir yöneliş söz konusudur. Bu bağlamda finansal raporlama, tarihsel süreç içinde finansal tablolar ile başlayan ve günümüzde entegre raporlamaya doğru evrilen bir gelişim içindedir.



Şekil 2.1: Finansal raporlamanın gelişimi (Aras ve Sarıoğlu, 2015, s. 38)

### 2.1.1. Finansal tablolar

Finansal tablolar muhasebe sürecinde toplanan finansal bilgilerin özetlenerek gösterildiği tablolardır. 1 Seri No'lu Muhasebe Sistemi Uygulama Genel Tebliği, 7 adet finansal tablo saymış ve bu finansal tabloları temel ve ek finansal tablolar şeklinde iki sınıfa ayırmıştır. Tebliğde sayılan finansal tablolar; Bilanço, Gelir Tablosu, Satışların Maliyeti Tablosu, Fon Akım Tabloları, Nakit Akım Tablosu, Kâr Dağıtım Tablosu ve Öz Kaynaklar Değişim Tablosudur. Bu tablolardan bilanço ve gelir tablosu dipnotları ve ekleri ile birlikte temel finansal tabloları, diğerleri ise ek finansal tabloları oluşturur. Türkiye Muhasebe Standartları'nda ise finansal tablolar, genel amaçlı finansal tablolar ve özel amaçlı finansal tablolar şeklinde sınıflandırılmıştır.

Muhasebe standartlarında genel amaçlı finansal tablolar, kendi özel bilgi ihtiyaçlarını karşılamaya yönelik raporlar hazırlanmasını talep edebilme imkânı bulunmayan finansal tablo kullanıcılarının ihtiyaçlarını karşılamaya yönelik finansal tablolar olarak ifade edilmektedir. TMS1 Finansal Tabloların Sunuluşu Standardına göre tam bir finansal tablolar seti aşağıdakilerden oluşmaktadır:

- (a) Dönem sonuna ait finansal durum tablosu,
- (b) Döneme ait kâr veya zarar ve diğer kapsamlı gelir tablosu,
- (c) Döneme ait öz kaynak değişim tablosu,
- (d) Döneme ait nakit akış tablosu,
- (e) Önemli muhasebe politikalarını ve diğer açıklayıcı bilgileri içeren dipnotlar

Özel amaçlı finansal tablolar ise, özellikle bilgi sunmak için talep edilmesi halinde hazırlanan tablolardır. Özel amaçlı tablolara, kâr dağıtım tablosu ve satışların maliyeti tablosu örnek olarak verilebilir.

### 2.1.2. Faaliyet raporları

Günümüzde teknolojik gelişmenin de etkisiyle ticari hayat baş döndürücü bir hızla ilerlemekte ve bilgi üretmektedir. Ancak üretilen bilginin fazlalığı ve üretilme hızı bilgilerin karmaşıklaşmasını ve manipülatif bilginin hızlı bir şekilde yayılmasını beraberinde getirmiştir. Bu durum bilgi kullanıcılarının güvenilir bilgiye zamanında erişim sağlayabileceği kaynaklara olan ihtiyacını da artırmaktadır. Bu bağlamda faaliyet raporları işletmelerin kendisi ile ilgili güvenilir finansal bilgiler yanında finansal olmayan bilgileri de sunabileceği önemli bir kaynak olarak karşımıza çıkmaktadır.

Faaliyet raporu, şirketin mevcut ve potansiyel iş ortaklarını ve menfaat sahiplerini bilgilendirmek ve etkilemek için yıllık olarak hazırlanan ve kapsadığı hesap dönemine ilişkin bilgileri, şirketin önümüzdeki dönemdeki hedeflerini ve bu hedefleri nasıl yerine getireceğini anlatan bir dokümandır (Deloitte, 2009, s. 2).

Ülkemizde faaliyet raporları ile ilgili çeşitli yasal düzenlemeler bulunmaktadır. Bu düzenlemelerden en kapsamlı olanı 28.08.2012 tarihinde resmi gazetede yayınlanan Şirketlerin Yıllık Faaliyet Raporunun Asgari İçeriğinin Belirlenmesi Hakkında Yönetmelik'tir. Bunun yanında Türk Ticaret Kanunu'nda ve Kurumsal Yönetim Tebliği'nde de faaliyet raporları ile ilgili düzenlemeler bulunmaktadır.

Şirketlerin Yıllık Faaliyet Raporunun Asgari İçeriğinin Belirlenmesi Hakkında Yönetmelik'te tüm sermaye şirketlerinin yıllık faaliyet raporu düzenlenmek zorunda olduğu ifade edilmektedir. Bu açıdan, anonim şirketlerin, limited şirketlerin ve sermayesi paylara bölünmüş komandit şirketlerin her yıl faaliyet raporu düzenlemesi gerekir.

Yönetmelikte faaliyet raporlarının hazırlanmasında dikkat edilmesi gereken hususlar konusunda da detaylı açıklamalar bulunmaktadır. Buna göre; yıllık faaliyet raporu, işletmenin ilgili hesap dönemine ait iş ve işlemlerinin akışını, her yönüyle finansal durumunu, şirketin hak ve yararını da gözetecek şekilde, gerçeğe uygun ve dürüst bir şekilde ifade etmelidir. İşletmenin karşı karşıya bulunduğu temel riskler yıllık faaliyet raporunda belirtilmeli ve ayrıca finansal olmayan risklere de faaliyet raporunda yer verilmelidir. İşletmenin faaliyet ve hizmetlerinin etkin, güvenilir ve kesintisiz bir şekilde yürütülmesini, muhasebe ve mali raporlama sisteminden sağlanan bilgilerin bütünlüğünü, tutarlılığını, güvenilirliğini, zamanında elde edilebilirliğini ve güvenliğini sağlamak amacıyla yapılan iç kontrollerin etkinliği, yeterliliği ve uyumluluğu konularında açıklamalara yer verilmelidir. Yıllık faaliyet raporunda, geleceğe yönelik bilgi verildiği veya tahminlerde bulunulduğu durumlarda, bunların dayandığı gerekçelere ve istatistiki bilgilere de yer verilmesi zorunludur. Ayrıca geleceğe yönelik bilgi ve tahminler şirketin finansal durumu ve faaliyet sonuçları ile uyumlu olarak sunulmalıdır.

Faaliyet raporlarının denetlenebilir olması ve vermiş olduğu bilgilerin yasal bir çerçeveye bağlanmış olması faaliyet raporlarındaki bilgilere olan güveni artırmakta ve menfaat sahipleri tarafından kullanılan önemli bir araç olmasını sağlamaktadır.

Faaliyet raporları, finansal analistler, yatırımcılar ve düzenleyiciler açısından en önemli ve değerli bilgileri içeren kaynaklardan biridir. Faaliyet raporlarındaki zorunlu açıklamalar, fiyat ve satış değişikliklerinin nedenleri, satış geliri ve maliyet

değişikliklerinin nedenleri, planlanan harcamalar, gelecekteki likidite durumu ve geçmiş yıl performans ve gelecek beklentilerinin görünümü gibi performansla ilgili bilgileri içerir. Faaliyet raporlarını okuyanlar sezgisel olarak işletmenin şu andaki ve gelecekteki performansı, stratejileri ve kârlılığı ile ilgili açıklamalardan görüşler çıkarmayı beklemektedir (Qiu, 2007, s. 1).

Faaliyet raporunun sağladığı temel yararları şu şekilde özetleyebiliriz (Deloitte, 2009, s. 3):

- a. Düzenleyici kuruluşların standartlarına uyumu sağlar ve bu kuruluşlarla olan iletişime katkıda bulunur
- b. Hisse senedi fiyatının, işletmenin gerçek değerini yansıtmasını sağlar
- c. İşletme menfaat sahipleri (yatırımcılar, tedarikçiler vb.) ile işletme arasındaki ilişkileri geliştirir
- d. İşletmenin bilinirliğini artırır
- e. İşletme ilgili fırsatların ve işletmenin büyüme potansiyelinin anlaşılmasını sağlar
- f. İşletmenin sermaye piyasalarından yeni fon temin etme yeteneğini artırır.

Şirketlerin Yıllık Faaliyet Raporunun Asgari İçeriğinin Belirlenmesi Hakkında Yönetmelik faaliyet raporunda yer alması gereken bölümleri ve içeriği hakkında bilgi vermektedir. Buna göre faaliyet raporu aşağıda gösterilen yedi bölümden oluşmaktadır.

- a) Genel bilgiler
- b) Yönetim organı üyeleri ile üst düzey yöneticilere sağlanan mali haklar
- c) Şirketin araştırma ve geliştirme çalışmaları
- d) Şirket faaliyetleri ve faaliyetlere ilişkin önemli gelişmeler
- e) Finansal durum
- f) Riskler ve yönetim organının değerlendirmesi
- g) Diğer hususlar

Yıllık faaliyet raporunun yukarıda sayılan altı bölümünde yer alması gereken hususlar Tablo 2.1'de belirtilmiştir. Raporun diğer hususlar bölümünde ise, faaliyet yılının sona ermesinden sonra şirkette meydana gelen ve ortakların, alacaklıların ve diğer ilgili kişi ve kuruluşların haklarını etkileyebilecek nitelikteki özel önem taşıyan olaylara ilişkin açıklamalara yer verilmesi beklenir. Bu bölümde ayrıca, söz konusu yönetmelik hükümlerine aykırı olmamak kaydıyla yönetim organının uygun gördüğü ilave bilgilere de yer verilebilir.

**Tablo 2.1: Faaliyet raporunun içeriği**

| Genel bilgiler                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Yönetim organı üyeleri ile üst düzey yöneticilere sağlanan mali haklar                                                                                                                                                                                                           | Şirketin araştırma ve geliştirme çalışmaları:                                                    | Şirket faaliyetleri ve faaliyetlere ilişkin önemli gelişmeler:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Finansal Durum                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Riskler ve yönetim organının değerlendirmesi:                                                                                                                                                                                                                                                                                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>a) Raporun ilgili olduğu hesap dönemi, şirketin ticaret unvanı, ticaret sicili numarası, merkez ve varsa şubelerine ilişkin iletişim bilgileri ile varsa internet sitesinin adresi,</p> <p>b) Şirketin organizasyon, sermaye ve ortaklık yapıları, nihai gerçek kişi ortaklar (Ortaklık yapısının şeffaf olması açısından önemli) ile bunlara ilişkin hesap dönemi içerisindeki değişiklikler,</p> <p>c) Varsa imtiyazlı paylara ve payların oy haklarına ilişkin açıklamalar,</p> <p>d) Yönetim organı, üst düzey yöneticileri ve personel sayısı ile ilgili bilgiler,</p> <p>e) Varsa; şirket genel kurulunca verilen izin çerçevesinde yönetim organı üyelerinin şirketle kendisi veya başkası adına yaptığı işlemler ile rekabet yasağı kapsamındaki faaliyetleri hakkında bilgiler.</p> | <p>a) Sağlanan huzur hakkı, ücret, prim, ikramiye, kâr payı gibi mali menfaatlerin toplam tutarları,</p> <p>b) Verilen ödenekler, yolculuk, konaklama ve temsil giderleri ile ayni ve nakdi imkânlar, sigortalar ve benzeri teminatların toplam tutarlarına ilişkin bilgiler</p> | <p>a) Şirketin araştırma ve geliştirme çalışmaları ile bunların sonuçlarına ilişkin bilgiler</p> | <p>a) Şirketin ilgili hesap döneminde yapmış olduğu yatırımlara ilişkin bilgiler,</p> <p>b) Şirketin iç kontrol sistemi ve iç denetim faaliyetleri hakkında bilgiler ile yönetim organının bu konudaki görüşü,</p> <p>c) Şirketin doğrudan veya dolaylı iştirakleri ve pay oranlarına ilişkin bilgiler,</p> <p>d) Dolaylı ve doğrudan iştiraklerle dönem içinde gerçekleştirilen ticari ve mali ilişkilerin tutarlarını gösteren tablolar</p> <p>e) Şirketin iktisap ettiği kendi paylarına ilişkin bilgiler,</p> <p>f) Hesap dönemi içerisinde yapılan özel denetime ve kamu denetimine ilişkin açıklamalar,</p> <p>g) Şirket aleyhine açılan ve şirketin mali durumunu ve faaliyetlerini etkileyebilecek nitelikteki davalar ve olası sonuçları hakkında bilgiler,</p> <p>h) Mevzuat hükümlerine aykırı uygulamalar nedeniyle şirket ve yönetim organı üyeleri hakkında uygulanan idari veya adli yaptırımlara ilişkin açıklamalar,</p> <p>i) Geçmiş dönemlerde belirlenen hedeflere ulaşıp ulaşılmadığı, genel kurul kararlarının yerine getirilip getirilmediği, hedeflere ulaşamamışsa veya kararlar yerine getirilmemişse gerekçelerine ilişkin bilgiler ve değerlendirmeler,</p> <p>j) Yıl içerisinde olağanüstü genel kurul toplantısı yapılmışsa, toplantının tarihi, toplantıda alınan kararlar ve buna ilişkin yapılan işlemler de dâhil olmak üzere olağanüstü genel kurula ilişkin bilgiler,</p> <p>k) Şirketin yıl içinde yapmış olduğu bağış ve yardımlar ile sosyal sorumluluk projeleri çerçevesinde yapılan harcamalara ilişkin bilgiler,</p> <p>l) Şirketler topluluğuna bağlı bir şirkete; hâkim şirketle, hâkim şirkete bağlı bir şirketle, hâkim şirketin yönlendirmesiyle onun ya da ona bağlı bir şirketin yararına yaptığı tüm işlemler</p> | <p>a) Finansal duruma ve faaliyet sonuçlarına ilişkin yönetim organının analizi ve değerlendirmesi, planlanan faaliyetlerin gerçekleşme derecesi, belirlenen stratejik hedefler karşısında şirketin durumu,</p> <p>b) Geçmiş yıllarla karşılaştırmalı olarak şirketin yıl içindeki satışları, verimliliği, gelir oluşturma kapasitesi, kârlılığı ve borç/öz kaynak oranı ile şirket faaliyetlerinin sonuçları hakkında fikir verecek diğer hususlara ilişkin bilgiler ve ileriye dönük beklentiler,</p> <p>c) Şirketin sermayesinin karşılıksız kalıp kalmadığına veya borca batık olup olmadığına ilişkin tespit ve yönetim organı değerlendirmeleri,</p> <p>d) Varsa şirketin finansal yapısını iyileştirmek için alınması düşünülen önlemler,</p> <p>e) Kâr payı dağıtım politikasına ilişkin bilgiler ve kâr dağıtımı yapılmayacaksa gerekçesi ile dağıtılmayan kârın nasıl kullanılacağına ilişkin öneri.</p> | <p>a) Varsa şirketin öngörülen risklere karşı uygulayacağı risk yönetimi politikasına ilişkin bilgiler,</p> <p>b) Oluşturulmuş ise riskin erken teşhis ve yönetimi komitesinin çalışmalarına ve raporlarına ilişkin bilgiler,</p> <p>c) Satışlar, verimlilik, gelir yaratma kapasitesi, kârlılık, borç/öz kaynak oranı ve benzeri konularda ileriye dönük riskler.</p> |

### 2.1.3. Sürdürülebilirlik raporları

Sürdürülebilirlik raporları, temel olarak işletme dışındaki paydaşları işletmenin ekonomik, çevresel ve sosyal faaliyetlerinin sonuçları hakkında bilgilendirmek için hazırlanmaktadır. Ancak sürdürülebilirlik raporları temel etkisini, işletmenin sürdürülebilirlik amaçlarının belirlenmesinde, verilerin toplanmasında ve yöneticilerin ve çalışanların sürece dahil edilmesinde gösterir (Özsözügün Çalışkan, 2012, s. 59). Bu açıdan sürdürülebilirlik raporları, işletmelerin yürüttükleri faaliyetler sonucunda yarattıkları ekonomik, çevresel ve sosyal etkilerin olumlu yönlerinin geliştirilmesi ile olumsuz yönlerinin ortadan kaldırılmasına ilişkin hedefler, faaliyetler ve politikaların, işletmenin faaliyetlerinden etkilenen ve faaliyetlerini etkileyen tüm kesimlerle belirli standartlar çerçevesinde paylaşılması amacıyla hazırlanan raporlardır (Önce, Onay, Yeşilçelebi ve Özkes, 2015, s. 64).

Kurumsal sürdürülebilirlik gelişmelerini yakından takip ederek amaç ve sorumluluklarını yeniden belirleyen bazı işletmeler, 1990'lı yılların başından itibaren sürdürülebilirlik performanslarını raporlamaya başlamışlardır. Öncelikle enerji ve su kullanımı, karbon emisyonu vb. çevresel verilerin raporlaması ile başlayan bu süreç, 2000'li yıllardan itibaren çalışan hakları, kurumsal yönetim, sosyal sorumluluk gibi konuları da kapsayan sürdürülebilirlik raporlarına dönüşmüştür. Sürdürülebilirlik raporları işletmelerin faaliyetleri sonucunda ortaya çıkan değer veya zararların daha iyi ölçülmesini sağlar, uzun vadeli stratejilerini belirlemelerine yardımcı olur (Aras ve Sarioğlu, 2015, s. 27).

Sürdürülebilirlik raporlamasının işletmelere sağladığı faydalar şu şekilde sıralanabilir (WBCSD, 2002, s. 15):

**Paydaşlara karşı şeffaflık:** Sürdürülebilirlik raporlaması hissedarlar, yerel sivil toplum kuruluşları ve devlet kurumları gibi paydaşlara uygun bilgiler sunmayı hedefleyen bir iletişim kanalıdır. Bu sayede, kurumsal görünürlük artırılır ve şeffaf olunduğunun kanıtlanması sağlanır.

**Finansal değer yaratılması:** Sürdürülebilirlik raporlaması, dolaylı olarak, işletmenin maddi olmayan varlıklarının hisse değerini uzun dönemde artırma kabiliyetini ve istekliliğini yansıtmaktadır.

**Uzun vadeli sermaye çekilmesi ve elverişli finansal koşulların sağlanması:** Sürdürülebilirlik raporlaması uzun dönemli anlayışa sahip sabırlı yatırımcıların çekilmesi

ile finansör ve sigortacıların işletme için düşük risk primi yargısına sahip olmasını destekler.

**Çalışanların motivasyonunun ve farkındalığının artırılması ile yetenekli işgücünün işletmeye çekilmesi:** Sürdürülebilirlik raporlaması hem işletme içindeki çalışanlara hem de işgücü piyasasına işletmenin çevresel ve sosyal konulara ilişkin işletme değerleri ve prensiplerinin nasıl uygulayacağını göstermesini sağlayacaktır.

**Yönetim sistemlerinin geliştirilmesi:** Sürdürülebilirlik raporlaması ekonomik, çevresel ve sosyal etkilerin yönetilmesi için daha sert ve güçlü yönetim sistemleri uygulamalarının kolaylaştırılması ve desteklenmesini sağlayabilir. Kısaca, doğru verilerin daha iyi toplanmasını sağlayabilir.

**Risk farkındalığı:** İşletmenin riskleri nasıl yönettiğini yansıtır.

**İnovasyonun Desteklenmesi:** Sürdürülebilirlik raporlaması ileri görüşlü düşünme ve performansı destekleyerek işletmenin rekabet avantajı kazanmasına yardım eder.

**Sürekli Gelişme:** Sürdürülebilirlik raporlaması sürekli gelişme ve öğrenmeyi destekler. Raporlama gelecek yıl raporlanacak daha fazla ilerleme için üst düzey yöneticileri harekete geçirir.

**İtibarın artırılması:** Sürdürülebilirlik raporlaması ürün değeri, pazar payı, müşteri sadakatinin artırılmasına katkı sunan, uzun vadeli itibarın sağlanmasına yardım eder.

Dünyada sürdürülebilirlik raporlaması konusunda yol gösterici birçok kurum ve girişim bulunmaktadır. Bunların en önemlilerini şu şekilde sıralayabiliriz:

- 1) GRI (Global Reporting Initiative – Küresel Raporlama Girişimi)
- 2) IIRC (International Integrated Reporting Council - Uluslararası Entegre Raporlama Konseyi)
- 3) UNGC (UN Global Compact - Birleşmiş Milletler Küresel İlkeler Sözleşmesi)
- 4) OECD Çok Uluslu Şirketler Rehberi
- 5) CDP (Carbon Disclosure Project – Karbon Saydamlık Projesi)

#### **2.1.4. Entegre raporlar**

Entegre raporlama; bir işletmenin stratejisinin, kurumsal yönetiminin, performansının ve beklentilerinin işletmenin dış çevresi bağlamında kısa, orta ve uzun vadede değer yaratmayı nasıl sağlayacağını kısa ve öz bir şekilde bildirilmesidir. Bir entegre raporun temel amacı finansal sermaye sağlayan taraflara işletmenin zaman içinde

nasıl değer yaratacağını açıklamaktır. Dolayısıyla entegre rapor hem finansal hem de finansal olmayan konularla ilgili bilgiler içerir (IIRC, 2013, s. 7). Entegre rapor tüm paydaşlara finansal ve finansal olmayan bilgiler arasında bağlantı kurma, işletmenin süreçlerini ve stratejilerini, kısa, orta ve uzun vadede ölçülebilir risk ve fırsatlarını görme imkânı vermektedir. Entegre raporda işletmenin geçmiş performansını ortaya koyulmakla birlikte, şirketin hedeflerine nasıl ulaşacağı, nasıl hayatta kalacağı ve nasıl değer yaratacağı da açıklanmaktadır (Öztürk, 2019, s. 3).

Nelson Mandela'nın 1994 yılında, Güney Afrika'daki kurumlara duyulan güvensizliği azaltabilmek amacıyla şeffaflık ve bilgi paylaşımını teşvik etmek üzere Mervyn King'i King Komitesi'ni kurması için görevlendirmesi entegre raporlamanın başlangıcı olarak kabul edilmektedir. Bu süreçte 1994'de yayımlanan King I, 2002'de yayımlanan King II ve 2009 yılında yayımlanan King III raporları ile işletmelerin sosyal, çevresel ve ekonomik faaliyetlerinin sonuçlarının rapor edilmesi, finansal bilgiler ile birlikte finansal olmayan bilgilerin de raporlanmasını sağlamıştır. Entegre raporlamaya ilişkin küresel boyutta ciddi çalışmalar ise, 2010 yılında Uluslararası Entegre Raporlama Konseyi'nin (International Integrated Reporting Council - IIRC) kurulması ile başlamıştır. IIRC uzun vadeli vizyonunu, "kamu sektöründe ve özel sektörde kurumsal raporlama normu olarak entegre raporlamanın kullanıldığı, yaygın kurumsal uygulamalarda entegre düşünce yapısının uygulandığı bir dünya" şeklinde ifade etmektedir. IIRC, şirketlere entegre rapor hazırlamalarında yön gösterecek, uluslararası kabul görmüş bir entegre raporlama çerçevesinin hazırlanmasını misyonu olarak belirlemiş ve Türkiye'nin de yer aldığı çok sayıda ülkeden özel sektör, kamu, sivil toplum temsilcileri ve akademisyenlerden gelen görüşler doğrultusunda hazırlanan Entegre Raporlama Uluslararası Çerçevesini (International IR Framework) 2013 yılında yayımlamıştır. Entegre raporlamaya başlayan kurumlar raporlamalarında bu çerçeveye göre hareket etmektedir (Darman, Taştan, Seçkin ve Kır, 2018, s. 63).

2013 yılında yayımlanan Uluslararası Entegre Raporlama Çerçevesinde bir entegre raporun içermesi gereken öğeler 8 başlık altında incelemiştir. Bunlar;

- A.** Kurumsal genel görünüm ve dış çevre
- B.** Kurumsal Yönetim
- C.** İş modeli
- D.** Riskler ve fırsatlar
- E.** Strateji ve kaynak aktarımı

- F. Performans
- G. Genel görünüş
- H. Hazırlık ve sunum temeli ve
- İ. Genel raporlama ilkeleridir.

## 2.2. Finansal Performans

Finansal performans, belli bir dönemde bir işletmenin varlıklarının kullanım etkinliğinin ve varlıklarından gelir elde etme düzeyinin nesnel bir ölçüsüdür. Finansal performans aynı zamanda belirli bir zaman zarfında bir işletmenin finansal durumunun genel bir ölçüsü olarak kullanılır ve işletmeyi aynı sektördeki benzer işletmelerle veya diğer sektörlerdeki işletmeler ile karşılaştırma olanağı sağlar (Kenton, 2018).

Bir işletmenin temel finansal görevi, kaynaklarını değer yaratmak için kullanmaktır. Finansal performans ise bu değer yaratma görevinin uygulama sonucudur. Karar vericiler, değer yaratmada işletmenin kaynaklarının etkin olarak kullanılıp kullanılmadığını görebilmek için bilgiye ihtiyaç duyarlar. İhtiyaç duyulan bu bilgi ise finansal performans göstergeleri aracılığıyla sunulmaktadır (Holbrook, 2010, s. 5).

Finansal performans göstergeleri; geçmişin ve bugünün ekonomik sonuçlarını belirlemede kullanılan, işletme faaliyetlerinin etkinlik derecesini ve hedeflerle olan uyum durumunu tespit eden ve aynı zamanda ölçen süreçlerdir (Düzer, 2018, s. 85). Muhasebe ve finans literatüründe finansal performans göstergelerinin sınıflandırılması ile ilgili farklı yaklaşımlar bulunmakla birlikte, genel itibarıyla finansal performans göstergelerinin; **geleneksel muhasebe bazlı göstergeler** ve **değer odaklı modern performans göstergeleri** şeklinde kategorize edildiği görülmektedir. Ancak bunun yanında **piyasa bazlı göstergeler** de literatürde önemli bir yer tutmaktadır. Muhasebe bazlı göstergeler daha çok maliyet ve verimlilik üzerine odaklanırken, modern performans göstergeleri işletmenin değer yaratan faaliyetleri üzerine odaklanmaktadır (Gökbulut, 2009, s. 43). Piyasa bazlı performans göstergeleri ise işletmelerin sermaye piyasalarındaki hareketleri üzerine odaklanmaktadır.

### 2.2.1. Muhasebe bazlı finansal performans göstergeleri

Muhasebe bazlı finansal performans göstergeleri için kârlılık temel kaynaktır. Bu açıdan en sık kullanılan muhasebe bazlı finansal performans göstergeleri Aktif Kârlılık Oranı, Öz kaynak Kârlılık Oranı, Net Kâr Marjı ve Hisse Başına Kârdır.

#### 2.2.1.1. *Aktif kârlılık oranı (ROA-return on assets)*

Bu oran, aktiflerin işletmede ne ölçüde kârlı kullanıldığını belirlemek amacıyla kullanılır. Oran hem vergiden sonraki kâr hem de vergiden önceki kâr esas almak suretiyle hesaplanabilir. Ancak vergi tutarının işletme yönetiminin kontrolünde olmaması ve vergi uygulamalarından işletmelerin farklı etkilenmesi nedeniyle kâr olarak vergiden önceki kârın kullanılması daha uygun olacaktır.

$$\text{Aktif Kârlılık Oranı} = \frac{\text{Dönem Kârı}}{\text{Aktif Toplamı}} \quad (2.1)$$

#### 2.2.1.2. *Öz kaynak kârlılık oranı (ROE-return on equity)*

Öz kaynak kârlılık oranı, işletmeye ortaklarca tahsil edilmiş bulunan fonların ne ölçüde etkin ve verimli kullanıldığını belirlemek için kullanılan orandır. Bu orana bakarak ortaklar işletmeye koydukları paranın alternatif maliyetini hesaplar.

$$\text{Öz kaynak Kârlılık Oranı} = \frac{\text{Dönem Kârı}}{\text{Öz kaynaklar}} \quad (2.2)$$

#### 2.2.1.3. *Net kâr marjı (ROS-return on sales)*

Net kâr marjı oranı faaliyet kârı elde etmiş bir işletmede net satışların yüzde kaçlık bir net kârı kapsadığını göstermektedir. Bu oran sayesinde faiz ve vergi de dahil olmak üzere bütün giderler karşılandıktan sonraki net satışların kârlılık yüzdesini görmek mümkün olmaktadır. Oranın yüksek çıkması olumlu olarak değerlendirilmektedir.

$$\text{Net kâr marjı} = \frac{\text{Net Kâr}}{\text{Net Satışlar}} \quad (2.3)$$

#### 2.2.1.4. *Hisse senedi başına kâr (EPS-earnings per share)*

Bu oran işletmenin her bir hisse senedi başına ne kadar kâr elde ettiğini tespit etmek için kullanılır.

$$\text{Hisse senedi başına kâr} = \frac{\text{Dönem Net Kârı}}{\text{Hisse Senedi Sayısı}} \quad (2.4)$$

### 2.2.2. *Değer bazlı finansal performans göstergeleri*

Değere dayalı yönetim anlayışının gelişmesi ile birlikte işletme değerinin ve dolayısı ile hissedar refahının maksimize edilmesi hedefinin önem kazanmasını takiben işletme değerlemesinde ve finansal performans ölçümünde değer bazlı finansal performans göstergeleri kullanılmaya başlanmıştır. Değere dayalı finansal performans

ölçüm kriterlerinin en önemli özelliği hissedar odaklı olmalarıdır. Geleneksel performans göstergeleri, işletmenin kârlılığını ön plana alırken değer bazlı performans göstergeleri hissedar değerine odaklanmakta ve işletme başarısını kârlılığın artmasının yanında hissedarların refahına yapılan katkı ile ölçmektedir (Koyuncu, 2010, s. 44).

Değere bazlı finansal performans göstergeleri geleneksel performans göstergelerini reddetmezler. Aksine onları tamamlayıcı ve anlamlandırıcı nitelikte göstergelerdir. Kâr veya kârlılık gibi finansal performansın tek yönüne odaklanmak yerine, işletmenin finansal performansını bir bütün halinde ortaya koymaya çalışırlar (Kırlı, Kuğu ve Kara, 2013, s. 102). Bu açıdan geleneksel muhasebe performans göstergelerinin eksik kaldıkları noktalarda, yatırım çevrelerine ve işletmelere yol göstermede değer bazlı performans göstergeleri popülerite kazanmıştır. Literatürde ön plana çıkan değere dayalı performans göstergeleri Ekonomik Katma Değer (EVA-Economic Value Added) ve Pazar Katma Değeri (MVA-Market Value Added)'dir (Şamiloğlu, 2005, s. 79).

Ekonomik Katma Değer, net işletme kârının, o kârı yaratmak için kullanılan sermayenin maliyetinden arındırılması sonucu ulaşılan değerdir. Ekonomik Katma Değer gerçek ekonomik kârın veya gelirlerin, yatırımcıların karşılaştırılabilir risk düzeyindeki başka yatırımlardan elde etmek isteyecekleri minimum getiri oranından ne kadar az veya çok olduğunu göstermektedir. Pazar Katma Değeri ise, şirketin piyasa değerinden defter değerinin çıkarılmasıyla bulunan değerdir ve yatırılan sermayenin etkin kullanılıp kullanılmadığını gösterir (Topal, 2008, s. 251-254).

### **2.2.3. Piyasa bazlı finansal performans göstergeleri**

Piyasa bazlı yaklaşım, herhangi bir varlığın veya işletmenin değerinin karşılaştırılabilir bir benzerinin fiyatına göre belirlenebileceği düşüncesini temel almaktadır (Düzer, 2008, s. 37). Piyasa bazlı finansal performans ölçütü olarak genellikle Tobin's Q, Piyasa Değeri/Defter Değeri ve Fiyat/Kazanç oranları kullanılmaktadır.

### **2.3. Büyük Veri ve Veri Madenciliği**

Genel manada veri kavramı bir sorun hakkında toplanan ve henüz çözümlenmemiş bilgi topluluğunu ifade etmektedir. Bu bilgi yığını metinsel, sayısal, görsel vb. olabilmektedir (Doğanay, 2012, s. 8).

Geçmişte veri kaynakları sınırlı olduğundan, elde edilen verilerin boyutu çok büyük değildi ve nispeten işlenmesi ve veriden faydalı bilginin elde edilmesi günümüze göre çok daha kolaydı. Ancak internet ve bilgisayar teknolojisinin gelişmesiyle birlikte bilgi

üretimi hızlanmış ve devasa boyutlara ulaşmıştır. Zira internet ve GPS bağlantılı akıllı telefonlar, bilgisayarlar ve sensörler gibi elektronik nesneler birçok farklı alanda kullanılan milyonlarca veriyi baş döndürücü bir şekilde üretmekte ve depolamaktadır. Her gün milyonlarca tweet atılmakta, video paylaşım ortamlarında milyonlarca video paylaşılmakta, sosyal ağlarda milyonlarca görsel ve yazılı bilgi paylaşılmaktadır. Ayrıca iklim ölçüm cihazları, trafik kameraları ve ölçüm sensörleri gibi cihazlarda her gün ciddi manada veri üretmektedir. Verinin bu kadar hızlı ve yoğun üretildiği bir ortamda geleneksel metotlarla verileri analiz etmek ve bu veri yığınlarından faydalı bilgi çıkarmak mümkün değildir. Bu aşamada 2000’li yıllardan itibaren bu veri yığınlarını tanımlamak için büyük veri kavramı kullanılmaya başlanmış ve yeni veri işleme teknolojileri ortaya çıkmaya başlamıştır.

Büyük veri geleneksel olarak kullanılan veri işleme uygulamalarının saklama, yönetme ve işleme kapasitesini aşan genişlikte ve karmaşıklıkta veri kümelerini anlatmak için kullanılan bir terimdir. Büyük verinin devasa boyutları ile bundan fayda sağlamak için gereken analizlerin karmaşıklığının birleşmesi, yeni sınıf teknolojilerin ve bunları yönetecek araçların gelişmesini sağlamıştır. Bu açıdan büyük veri terimi, hem yönetilen verinin türünü, hem de onu depolamak ve işlemek için kullanılan teknolojiyi kapsayan bir ifadedir. Bu teknolojiler çoğunlukla, Google, Amazon, Facebook ve LinkedIn vb. şirketlerin büyük miktardaki sosyal medya verisi ile uğraşırken, kendileri için geliştirdikleri teknolojiden doğmuştur (Cackett, 2013, s. 14). Büyük veri kavramı içerisinde değerlendirilen veriler çok büyüktür, çok hızlı hareket etmekte ve geleneksel veri tabanı sistemlerinin kısıtlarına uymamaktadırlar (Syed, Gillela and Venugopal, 2013, s. 2446). Bu bağlamda büyük verinin beş bileşene sahip olduğu düşünülmektedir. Bunlar; büyüklük (volume), hız (velocity), çeşitlilik (variety), doğruluk (veracity) ve değer (value) dir.

**Büyüklük:** Veri miktarını belirtmektedir. Büyüklükten kasıt, verinin geleneksel veri işleme sistemlerinin kapasitesini aşan miktarda olmasıdır. Bu açıdan büyük verinin boyutu Terabyte, Petabyte, Exabyte ve Zetabyte ile ifade edilmektedir.

**Hız:** Verinin üretilme ve ulaşılma hızını ifade etmektedir. Büyük veri akışının önemli bir bölümü anlık ve gerçek zamanlı olarak gerçekleşmektedir. Büyük veri teknolojisi ile verilen çok hızlı bir şekilde üretilmekte, yayılmakta, işlenmekte ve ulaşılabilir.

**Çeşitlilik:** Verinin türünü belirtmektedir. Fotoğraflar, sesler, videolar, hasta kayıtları, istatistiki sonuçlar, sosyal medya paylaşımları, uydu görüntüleri, trafik akış bilgileri gibi birçok çeşitte ve sayıdaki yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerin tamamı büyük veri kavramı içinde değerlendirilmektedir.

**Doğruluk:** Verinin kalite, güvenilirlik ve belirsizlik durumunu ifade etmektedir (Song and Zhu, 2016, s. 365).

**Değer:** Değer veriden anlamlı bilginin çıkarılmasını ifade etmektedir. Bu açıdan değer, büyük veri alanındaki en önemli bileşen olarak kabul edilmektedir. Büyük veriden değer çıkartılamıyorsa büyük verinin bir anlamı yoktur. Çeşitlilik ve doğruluk büyük verinin zorlayıcı boyutları olarak görülmekle birlikte en zorlayıcı boyut değerdir. Eğer büyük verinin zorlukları ile başa çıkabilir ve büyük veriden değer oluşturabilirsek yenilikçi çözümler üretebilir ve teknoloji, toplum ve iş yaşamı üzerinde önemli etkiler oluşturma şansı yakalayabiliriz (Song and Zhu, 2016, s. 365).

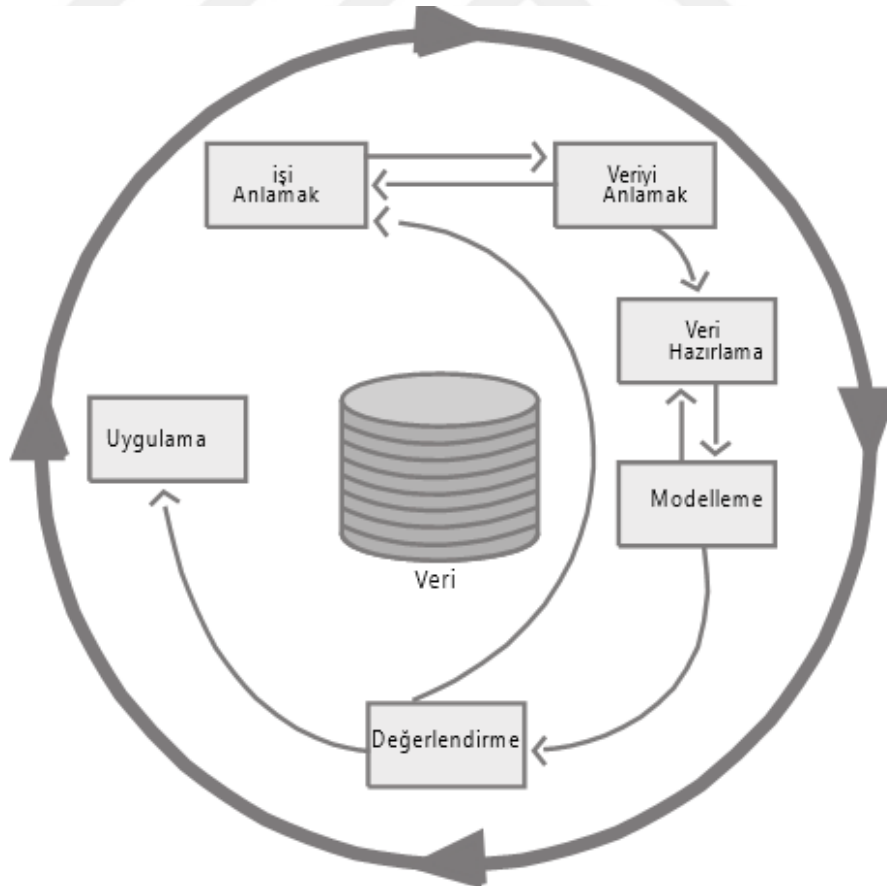
### 2.3.1. Veri Madenciliği (Data Mining)

Büyük veri teknolojisinin hayatımıza girmesiyle birlikte veriler kolaylıkla toplanmaya ve geniş veri havuzları oluşturmaya başlamıştır. Ancak bu verilerin önemli bir kısmı işlenmeden ham veri olarak saklanmaktadır. Veriler işlenmediği sürece anlamsız bir veri yığını olarak kalmaya devam edecektir. Dolayısıyla veri ambarlarında biriken büyük verilerin işlenmesi ve faydalı bilgi haline getirilmesi günümüzün önemli konularından biridir. Büyük verilerin geleneksel veri işleme yöntemleri ile işlenmesinin mümkün olmaması nedeniyle birçok yeni veri işleme yöntemi ve uygulaması hayatımıza girmiştir. Yeni gelişen ve gelişmekte olan bu veri işleme yöntemleri ve uygulamaları bir bütün olarak veri madenciliği başlığı altında incelenmektedir.

Veri Madenciliği, istatistik ve matematiğe dayalı teknikleri ve teknolojileri kullanarak veri yığınları arasında gizlenmiş bilgileri, fark edilmesi güç ilişkileri ortaya çıkaran ve ileriye yönelik tahminler yapılmasına imkan veren uygulamalardır (Özkul ve Pektekin, 2009, s. 71). Veri madenciliği veri tabanı teknolojisi, istatistik, yapay zeka, makine öğrenimi ve veri görselleştirmesi gibi pek çok teknik alan arasında köprü görevi gören çok disiplinli bir alandır (Özekes, 2003, s. 67). Bu açıdan günümüzde karar verme sürecine ihtiyaç duyulan pazarlama, biyoloji, bankacılık, borsa, telekomünikasyon, genetik, sağlık, mühendislik, kriminoloji ve istihbarat gibi birçok alanda veri madenciliği uygulamalarının yaygın olarak kullanıldığı görülmektedir. Son 20 yıldır Amerika Birleşik

Devletleri'nde çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılıklarının ortaya çıkartılmasına kadar geniş bir alanda kullanıldığı bilinmektedir (Savaş, Topaloğlu ve Yılmaz, 2012, s. 5).

Veri madenciliği çeşitli aşamalardan oluşan bir süreçtir. Veri madenciliği sürecinin standartlaştırılmasına yönelik dünyada farklı kurum ve kişiler tarafından çeşitli çalışmalar yapıldığı görülmektedir. Bu çalışmalar arasında literatürde ön plana çıkanlar SEMMA and CRISP-DM metodolojileridir. Bu metodolojiler arasında yaklaşım farklılığı olmasına karşın her ikisi de birbirine benzer adımlardan oluşmaktadır (Azevedo and Santos, 2008, s. 182). Bununla birlikte veri madenciliği sürecini en iyi ifade edebilen metodoloji CRISP-DM'dir. CRISP-DM metodolojisi NCR, SPSS ve Daimler-Benz şirketlerinin konsorsiyumu tarafından oluşturulmuştur. Bu metodoloji, önemli aşamaları, genel görevleri, özel görevleri ve süreç örneklerini içeren bir hiyerarşiyi tanımlamaktadır (Nisbet, Elder and Miner, 2009, s. 35). CRISP-DM'ye göre veri madenciliği Şekil 2.2'de gösterildiği üzere 6 aşamadan oluşmaktadır. Bunlar; işi anlamak, veriyi anlamak, veri hazırlama, modelleme, değerlendirme ve uygulama aşamalarıdır.



Şekil 2.2: CRISP-DM veri madenciliği aşamaları (Chapman ve diğerleri, 2000, s. 10)

Veri madenciliğinde işin anlaşılması aşaması, problemin tanımlanmasıdır. Veri madenciliğine konu olan ve çözüm bulunması istenen problem, doğru ve eksiksiz bir şekilde tanımlanmalıdır. Problemin tanımlanmasından sonra veri setinin anlaşılması aşamasına geçilir. Veri setinin problemin çözümüne ilişkin değişkenleri içermesi beklenmektedir. Bir sonraki aşamada verilerin analize uygun şekilde hazırlanması gerekmektedir. Analize hazır hale getirilen veriler modellenmekte, modeller ve sonuçları değerlendirilmekte, en iyi model seçilerek veri setine uygulanmaktadır (Eskici ve Koçak, 2018, s. 12). CRISP-DM metodolojisinde belirtilen veri madenciliği sürecinin aşamaları ve bu aşamalarda uygulanacak işlemler Tablo 2.2’de gösterilmiştir.

**Tablo 2.2:** CRISP-DM veri madenciliği aşamaları ve uygulanacak işlemler (Chapman ve diğerleri, 2000, s. 12)

| İş Anlamak                                                                                                                                                                                                | Veriyi Anlamak                                                                                                                                                                   | Veri Hazırlığı                                                                                                                                                                 | Modelleme                                                                                                                                                                                    | Değerlendirme                                                                                                                                               | Uygulama                                                                                                                                                                                                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• İş hedeflerini belirleme</li> <li>• Mevcut durumu değerlendirme</li> <li>• Veri madenciliğinin amacının belirlenmesi</li> <li>• Proje planı oluşturma</li> </ul> | <ul style="list-style-type: none"> <li>• Verinin toplanması</li> <li>• Verinin açıklanması</li> <li>• Verinin keşfedilmesi</li> <li>• Veri kalitesinin teyit edilmesi</li> </ul> | <ul style="list-style-type: none"> <li>• Veri seçimi</li> <li>• Veri temizliği</li> <li>• Veri oluşturma</li> <li>• Veri entegrasyonu</li> <li>• Veri biçimlendirme</li> </ul> | <ul style="list-style-type: none"> <li>• Modelleme tekniğinin Seçilmesi</li> <li>• Test tasarımının oluşturulması</li> <li>• Modelin inşa edilmesi</li> <li>• Model değerlendirme</li> </ul> | <ul style="list-style-type: none"> <li>• Sonuçları değerlendirme</li> <li>• Sürecin gözden geçirilmesi</li> <li>• Sonraki adımların belirlenmesi</li> </ul> | <ul style="list-style-type: none"> <li>• Planın uygulanması</li> <li>• Planın gözetlenmesi, bakım ve sürdürülmesi</li> <li>• Sonuç raporunun oluşturulması</li> <li>• Projenin gözden geçirilmesi</li> </ul> |

### 2.3.1.1. Veri madenciliği modelleri

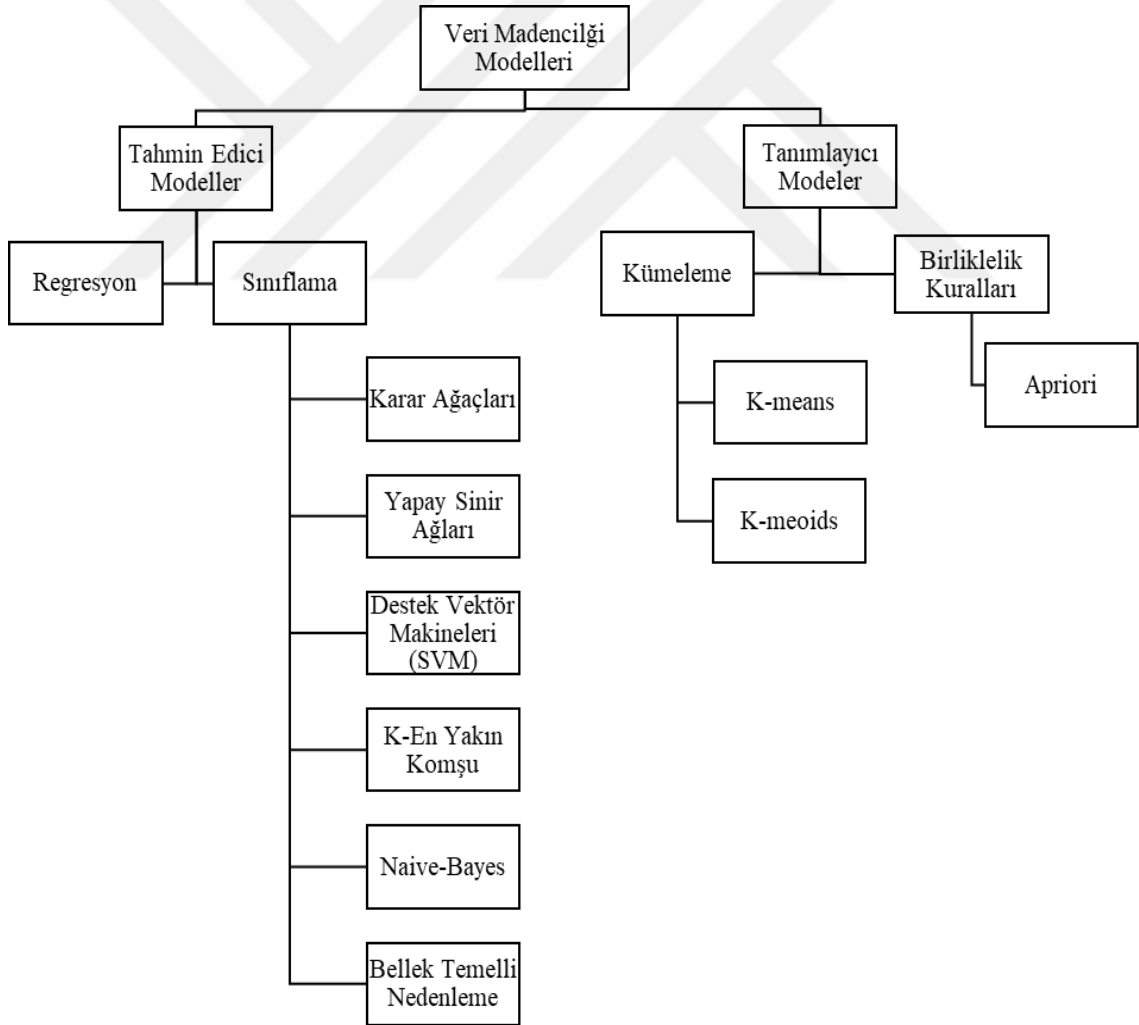
Veri madenciliği ile ilgili geçmişten günümüze kullanılan ve hemen her geçen gün yeni eklenen birçok yöntem ve algoritma bulunmaktadır. Bunlardan bir kısmı yıllardır kullanılan klasik teknikler diyebileceğimiz, ağırlıklı olarak istatistiksel yöntemlerdir. Diğer bir kısmı ise genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zekâ destekli yeni nesil yöntemlerdir (Savaş, Topaloğlu ve Yılmaz, 2012, s. 9). Bu açıdan veri madenciliğinde kullanılan modellerin tamamından bahsetmek mümkün değildir. Ancak literatürde genel kabul gören ve veri madenciliği uygulamalarında sıklıkla kullanılan modellerden kısaca bahsetmekte yarar vardır.

Veri madenciliğinde kullanılan modeller, **tahmin edici (Predictive)** ve **tanımlayıcı (Descriptive)** olmak üzere iki ana başlık altında toplanmaktadır. Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmekte ve

kurulan bu modelden hareketle sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntüler ortaya çıkarılmaktadır. Veri madenciliği modelleri gördükleri işlevlere göre ise üç başlık altında incelenmektedir. Bunlar (Akpınar, 2000, s. 5);

- Sınıflama ve Regresyon,
- Kümeleme,
- Birliklilik Kuralları ve Ardışık Zamanlı Örüntülerdir.

Şekil 2.3’de görüldüğü üzere sınıflama ve regresyon modelleri tahmin edici modeller, kümeleme ve birliklilik kuralları ve ardışık zamanlı örüntü modelleri ise tanımlayıcı modellerdir (Akpınar, 2000, s. 5).



Şekil 2.3: Veri madenciliği modelleri

Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri içeren iki veri analiz yöntemidir. Sınıflama kategorik değerlerin tahmin edilmesinde kullanılırken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır. Yani, regresyon kategorik değerler yerine eksik veya mevcut olmayan sayısal veri değerlerini tahmin etmek için kullanılır (Han, Kamber and Pei, 2001, s. 19). Örneğin, bir sınıflama modeli bir işletmenin müşterilerini yüksek gelirli, orta gelirli ve düşük gelirli gibi üç gruba kategorize etmek amacıyla kurulurken, regresyon modeli, gelir, meslek, cinsiyet gibi özellikleri verilen potansiyel müşterilerin herhangi bir ürünü alırken yapacakları harcamaları tahmin etmek için kurulabilir. Sınıflama için kullanılan temel teknikler; Karar Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri (DVM), K-En Yakın Komşu, Naive-Bayes ve Bellek Temelli Nedenleme'dir. Bu teknikler dışında sınıflama için kullanılan farklı teknikler de bulunmaktadır.

Kümeleme, nesnelerin küme adı verilen gruplara ayrıldığı denetimsiz bir süreçtir. Bu yöntemde nesnelere ilişkin etiketler yalnızca verilerden elde edilir. Sınıflandırma problemlerinde, modele önceden sınıflandırılmış eğitim örnekleri koleksiyonu sağlanmıştır ve sistemin görevi, yeni bir etiketlenmemiş nesneyi sınıflandırabilmek için sınıfların açıklamalarını öğrenmektir. Kümelemede ise, verilen etiketlenmemiş nesneler herhangi bir ön bilgi olmadan anlamlı kümeler halinde gruplandırılmaktadır. Kümeleme, veri madenciliği, belge alımı, görüntü bölümlendirme ve örüntü sınıflandırması dahil çok çeşitli veri analizi alanlarında uygulanmaktadır. Bu gibi birçok problemde, veriler hakkında önceden çok az bilgi vardır ve karar vericinin veriler hakkında mümkün olduğunca az varsayım yapması gerekir (Feldman and Sanger, 2007, s. 82). Kümeleme için yaygın olarak kullanılan teknikler; K-Means ve K-Medoids kümeleme algoritmalarıdır. Bunun dışında SOM (Self Organizing Map) gibi çeşitli çalışmalarda kullanılan kümeleme algoritmaları da bulunmaktadır.

Birliktelik kuralları, eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır. Pazar sepeti analizi, birliktelik kuralı yöntemin en iyi örneklerinden biridir. Bu analizde amaç, alışveriş esnasında müşterilerin satın aldıkları ürünler arasındaki birliktelik ilişkisini bulmak ve bu ilişki verisi doğrultusunda müşterilerin satın alma alışkanlıklarını tespit etmektir. Örneğin, düşük yağlı peynir ve yağsız yoğurt alan müşterilerin, %85 ihtimalle diyet süt de satın alması durumu bir birliktelik ilişkisidir. Ardışık zamanlı örüntüler ise birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde

gerçekleşen ilişkilerin tanımlanmasında kullanılır. Çekiş satın alan bir müşterinin ilk üç ay içerisinde %15, bu dönemi izleyen üç ay içerisinde %10 ihtimalle çivi satın alması ardışık zamanlı bir örüntüdür (Akpınar, 2000, s. 7). Birliktelik kuralları analizi için kullanılan en popüler algoritma Apriori Algoritmasıdır. Bunun yanında Eclat, AIS, SETM, Partition gibi algoritmalarda birliktelik kuralları analizine ilişkin olarak geliştirilmişlerdir.

#### **2.4. Metin Madenciliği (Text Mining)**

Metin madenciliği, veri madenciliğinin içerisinde yer almakta ve veri madenciliğinin bir parçası olarak görülmekle birlikte alışlagelmiş veri madenciliğinden farklı bir yaklaşıma sahiptir. Yaklaşımdaki temel farklılık, metin madenciliğinin yapılandırılmamış veri kaynağı olan metinleri incelemesidir (Ergün, 2012, s. 22). Bu bakımdan metin madenciliği veri madenciliği uygulamasının gelişmiş halidir. Zira metin madenciliğinde yapılandırılmamış veri kaynağı olan metinler, çeşitli işlemlerden geçirilerek yapılandırılmış veri haline getirilmekte, daha sonra yapılandırılmış bu verilere veri madenciliği modelleri uygulanmak suretiyle bilgiye ulaşılmaktadır.

Metin madenciliği işletme raporları, elektronik postalar, gazete haberleri, sosyal medya yorumları, müşteri yorumları, web sayfaları, sınav kâğıtları vb. yazılı belgelerden kolaylıkla çıkarılamayan içerikler, ilişkiler ve örüntüler gibi değerli bilgiler elde etmek için kullanılan bir makine öğrenme yaklaşımıdır (Yucel, 2016, s. ii). Metin madenciliği ile insanlar açısından çok zaman alacak işlemler kısa bir sürede tamamlanmakta ve fark edilmesi belli bir uzmanlık ve tecrübe gerektiren bilgiler hızlı ve basit bir şekilde elde edilmektedir. Ayrıca çok büyük boyuttaki metin verileri de hızlı bir şekilde analiz edilebilmektedir. Örneğin metin madenciliği uygulamaları sayesinde dil çevirileri, müşteri yorumlarının değerlendirilmesi, yazılı belgelerin sınıflandırılması gibi birçok işlem hızlı bir şekilde yapılabilmekte ve bunun yapılması için bir uzmana ihtiyaç duyulmamaktadır.

Metin madenciliği süreci, veri madenciliği sürecine benzer şekilde işlemektedir. Ancak metin madenciliği sürecindeki, özellikle veri hazırlığı aşamasındaki işlemler daha karmaşık ve zahmetlidir. Metin madenciliği süreci analiz edilecek belgelerin toplanması ile başlar. Toplanan bu belgelerde yer alan yapılandırılmamış veri kaynağı metinlerden öncelikle noktalama işaretleri, sayılar, alfabe dışı karakterler çıkartılır. Daha sonra metinde yer alan etkisiz (gereksiz) kelimeler çıkartılır ve tüm harfler küçük harfe

dönüştürülür. En son işlem olarak kelimeler köklerine (gövdelere) ayrılır. Tüm bu işlemler sonucunda her bir belgede yer alan metinler yapılandırılmış veri haline dönüşmektedir. Bu aşamadan sonra yapılandırılmış verilere veri madenciliği modelleri uygulanarak ulaşılmak istenen bilgi elde edilmektedir.

#### **2.4.1. Metin Madenciliği Çalışma Alanları**

##### **2.4.1.1. Bilgiye erişim**

Bilgiye erişim, kullanıcının sorgusuyla eşleşen bir koleksiyondaki kullanıcının ilgilendiği belgeleri tanımlar. En bilinen bilgiye erişim sistemleri World Wide Web’te yer alan ve belirli bir kelime gurubuyla ilgili belgelerin bulunmasını sağlayan Google gibi arama motorlarıdır. Bilgiye erişim sistemleri belirli bir sorun ile ilgili doküman kümesinin daraltılmasına olanak tanımaktadır (Singh and Ghosh, 2013, s. 24). Bilgiye erişim, belirli bir sorguyla eşleşen belgeleri getirir ancak yine de kullanıcının ilgili bilgileri bulmak için bu belgeleri okumasını gerektirir (Feldman and Sanger, 2007, s. 62).

##### **2.4.1.2. Bilgi çıkarımı**

Bilgi çıkarımının görevi, daha önceden açıkça tanımlanmış bir bilgi gereksinimi ile ilgili olarak bir metin belgesinden, alakasız bilgileri göz ardı ederek, belirli bir alanda ilgili bilgi kümesine ulaşmaktır. Başka bir deyişle, bilgi çıkarımı, yapılandırılmamış metinden yapılandırılmış gerçek bilgiyi türetmekle ilgilidir (Piskorski and Yangarber, 2013, s. 24).

Bilgiye erişim ile bilgi çıkarımı çoğu zaman birbiriyle karıştırılmaktadır. Ancak bilgiye erişim ile bilgi çıkarımı benzer kavramlar olarak gözükmekle birlikte birbirinden farklı işleve sahiptirler. Bilgiye erişim sorguya göre ilgili belgelere erişim olanağı tanımakta iken bilgi çıkarımı dokümanlardaki kişi isimleri, tarihler gibi spesifik bilgilere ulaşılmasını sağlamaktadır (Eskici ve Koçak, 2018, s. 35). Bilgiye erişim süreci genellikle sorguya cevap olarak sistemin belgeye tayin ettiği alaka düzeyi puanına karşılık gelen sıralı bir belge listesini döndürür. Ancak, sıralanan belge listesi, bu belgelerin içeriği hakkında ayrıntılı bilgi sağlamaz. Bilgi çıkarımının amacı, belgeleri sıralamak veya seçmek değil, onların anlamsal içeriği için daha anlamlı ve zengin temsiller oluşturmak için belgelerden önceden belirlenmiş olay türleri, varlıklar veya ilişkiler hakkındaki belirgin gerçekleri ortaya çıkarmaktır (Piskorski and Yangarber, 2013, s. 25).

#### **2.4.1.3. Doğal dil işleme**

Doğal dil işleme, ana işlevi doğal bir dili çözümleme, anlama, yorumlama ve üretme olan bilgisayar sistemlerinin tasarımını ve gerçekleştirilmesini konu alan bir bilim alanıdır. Doğal dil işlemede, yapay zeka, biçimsel diller kuramı, kuramsal dilbilim ve bilgisayar destekli dilbilim gibi çok değişik alanlarda geliştirilmiş kuram, yöntem ve teknolojiler bir arada kullanılmaktadır (Delibaş, 2008, s. x).

Doğal dil işlemenin temel amacı bir dili insan gibi işlemeyi başarmaktır. Tüm alan doğal dil işleme olarak adlandırılırken aslında iki farklı odaklanma alanı vardır. Bunlar dil işleme ve dil üretmedir. Dil işleme, anlamlı bir temsil üretme amacıyla dilin analizini, dil üretme ise bir temsilden dilin üretimini ifade etmektedir. Doğal dil işlemenin görevi, okuyucunun veya dinleyicinin rolüne eşittir, Doğal dil üretiminin görevi ise yazar veya konuşmacının görevidir (Liddy, 2001). Doğal dil işleme, karmaşık bir yapıya sahip olan insan dilini anlama çalışması olmasından dolayı yapay zekâ alanında karşılaşılan en güç sorunlardan biridir (Hamde, 2018, s. 39-40).

Google ve Yahoo gibi bazı arama motorları tarafından kullanılan makine çevirisi, otomatik özetleme, anahtar kelime çıkarma, anlamsal arama, soru cevaplama gibi görevler, doğal dil işleme algoritmalarının kullanıldığı uygulamalara örnek olarak gösterilebilir (Güvenç, 2016, s. 1).

#### **2.4.1.4. Web madenciliği**

Web madenciliği, World Wide Web belgeleri ve servislerinden bilgileri otomatik olarak bulmak ve çıkarmak için veri madenciliği tekniklerinin kullanılmasıdır (Etzioni, 1996, s. 65). Web madenciliği temel olarak yapı, içerik ve kullanım olmak üzere üç grupta incelenmektedir. Web yapı madenciliği ile internetin temel yapısını oluşturan web siteleri, web sayfaları arası ya da web sayfasındaki bağlantılar arasındaki ilişkiler incelenir. Web içerik madenciliği ile web sayfalarının içeriğinde yer alan metin, resim ve video gibi veriler analiz edilir. Web kullanım madenciliği ile ise web sunucularında tutulan kullanıcı erişim kayıtları incelenerek anlamlı ve faydalı kalıplar bulunabilir (Dolgun, Özdemir ve Oğuz, 2009, s. 51).

#### **2.4.1.5. Duygu analizi**

Duygu analizi, metin madenciliğinin önemli bir alanı olup, doğal dil işleme ve metin analizi kullanılarak bir metnin olumlu, olumsuz veya tarafsız bir içeriğe sahip olup olmadığını incelenmesi işlemini kapsamaktadır (Can ve Alataş, 2017, s. 75).

Literatürde, metin gruplarını kategorilere ayırmak için duygu analizini kullanan çok sayıda araştırma vardır. Bu tür araştırmalarda yaygın olarak kullanılan yaklaşımlar iki grupta değerlendirilebilir. “Anlamsal Yönelim” olarak adlandırılan ilk yaklaşım çoğunlukla alana özgü bir sözlük kaynağı kullanarak, kelimeleri “iyi” veya “kötü” gibi ikili sınıflara ayırmayı amaçlamaktadır. Bu nedenle, eğer belirli bir metin belgesi daha önceden belirlenmiş “iyi” terimleri, “kötü” terimlerden daha fazla içeriyorsa, o metin parçası “iyi” olarak kabul edilir ve bunun tersi de geçerlidir. Yapay zekâya dayalı olan “Machine Learning” (ML) adlı ikinci yaklaşım, yalnızca metin analitiği için değil, finanstan sağlık bilişimine kadar çeşitli alanlarda da kullanılmıştır. Makine öğrenme modelleri, etiketli tarihsel verileri öğrenerek metni ya ikili ya da çoklu kategorilerde sınıflandırır. Bu iki ana akım arasındaki kavramsal fark, makine öğrenme modellerinin, duyguları daha ince bir şekilde ifade eden cümleleri analiz etmede, anlamsal yönetime göre daha etkili olmasıdır (Yucel, 2016, s. 6).

#### **2.4.2. Metin Madenciliğinin Adımları**

Metin madenciliği gelişmiş bir veri madenciliği uygulaması olması nedeniyle, metin madenciliği süreci veri madenciliği ile benzer bir şekilde işlemektedir. Bu açıdan metin madenciliğinde de klasik veri madenciliği sürecinde takip edilen adımlar takip edilmektedir. Ancak metin madenciliği veri hazırlama aşamasında klasik veri madenciliği sürecinden farklılaşmaktadır. Zira yapılandırılmamış veri olan metinler üzerinde çalışıldığından, metinlerin yapılandırılmış veri haline getirilmesi bu aşamada yapılmakta ve önemli bir emek harcanmaktadır. Metin verilerinin, üzerinde veri madenciliği işlemi yapılabilecek hale getirilmesini sağlayacak metin işleme süreci şu üç temel adımda özetlenebilir.

1. Derlem (Corpus) Oluşturulması
2. Metin Ön işleme
3. Metin Dönüşümü

#### 2.4.2.1. Derlem (Corpus) oluşturulması

Derlem (Corpus) twitter, facebook, kurumsal internet sitesi, veri tabanı gibi alanlardan alınan ve üzerinde herhangi bir işlem yapılmamış veri kütlesini ifade etmektedir. Metin madenciliğinde ilk işlem bu derlemin oluşturulması ile başlar.

#### 2.4.2.2. Metin ön işleme

Metin madenciliğinde derlemin oluşturulmasından sonra elde edilen metnin çeşitli ön işleme adımlarından geçirilmesi gerekmektedir. Metin ön işleme adımları yapılacak çalışmanın durumuna göre farklılaşabilmekle birlikte genel itibarıyla standartlaştırma, temizleme, filtreleme, dizgeciklere ayırma ve gövdeleme işlemlerinden oluşmaktadır.

**Standartlaştırma:** Standartlaştırma, toplanan html, pdf, doc ve xml gibi farklı uzantılardaki dokümanların analizde kullanılacak formata dönüştürülmesi işlemidir. Bu işlem ön işleminin ilk aşamasını oluşturmaktadır. Bu aşamada ayrıca metnin tamamının küçük harfe dönüşümü sağlanır.

**Temizleme (Cleanning):** Temizleme işleminde standartlaştırılan dokümanlardaki metinlerde yer alan ve çalışma için herhangi bir anlam ifade etmeyen noktalama işaretleri, sayılar ve diğer karakterler silinir. Ayrıca metin içerisinde çalışmayı ilgilendirmeyen bölümler bulunuyor ise bunlarda silinir.

**Filtreleme:** Filtreleme, metin hakkında pek bir bilgi içermeyen ve literatürde gereksiz kelimeler (Stop Word) olarak ifade edilen kelimelerin sözlükten çıkarma işlemidir. Örneğin bağlaçlar ve edatlar gibi metin içeriğine etkisi olmayan kelimelerin analizde yer almasına gerek yoktur; bu yüzden filtrelenir. Ayrıca çok sık tekrar eden veya çok az tekrar eden kelimeler de uçbirim (outlier) olarak görülüp filtreleme işlemine tabi tutulabilir (Hotho, Nürnberger and Paass, 2005, s. 25)

**Dizgeciklere Ayırma (Tokenization):** Dizgeciklere ayırma işleminde, metin kelimeler, deyimler veya cümleler gibi anlamlı bileşenlere bölünür (Yucel, 2016, s. 14).

**Gövdeleme (Stemming ve Lemmatization):** Metin işlemede bir kelimenin ekli halinden daha çok kök hali önem arz etmektedir. Zira metin işleme teknolojisinin kullanıldığı durumlarda daha anlamlı ve doğru sonuçlara hızlı bir şekilde ulaşabilmek için aynı köke sahip kelimelerin birlikte değerlendirilmesi gerekmektedir. Çünkü gövdeleme yapılmaksızın yapılan bir işlemde, bir kelimenin birbirinden farklı biçimleri bir araya getirilemeyecek ve aynı köke sahip kelimelerin birbiriyle eşleşememesi problemi ortaya çıkacaktır. Örnek vermek gerekirse “banka” ile ilgili belgelerin arandığı bir sorguda,

kullanıcı “banka” kelimesi ile arama yaptığında “bankaya” kelimesini içeren belge ile bir eşleşme olmayacaktır ve bu belgeye erişilemeyecektir. Fakat kelime kökü dikkate alınarak “bankaya” kelimesi “banka” olarak işlem görür ise belgeye rahatlıkla erişilebilecektir.

Kök bulma (stemming) ve anlamsal köküne inme (lemmatization) işlemlerinin amacı; bir kelimenin çekim hallerini ve türetilmiş biçimlerini azaltarak, temel ortak bir biçime sokmaktır. Ancak bu iki işlemin amacı aynı olsa da, kelimeleri kök veya gövdelerine indirme mantıkları birbirlerinden farklıdır. Kök bulma, genel olarak kelimelerin son eklerini keserek amacına ulaşmaya çalışan yüzeysel bir işlemdir. Anlamsal köke inme ise bir sözlük kullanarak ve kelimelerin anlamsal analizlerini yaparak en temel hallerine ulaşmayı amaçlayan detaylı bir işlemdir (Manning, Raghavan and Schütze, 2009, s. 32).

Gövdeleme yöntemleri iki grupta incelenmektedir. Birinci grup, gövdelemenin elle yapılmasıdır. İkinci grup ise çeşitli gövdeleme algoritmaların kullanıldığı otomatik yöntemlerdir. Otomatik yöntemler 4 başlık altında incelenmektedir. Bunlar Tablo Arama Yöntemi, Halef Çeşitlilik Yöntemi, N-Grams Yöntemi ve Ek Atma Yöntemleridir (Frakes, 1992, s. 132).

Dünya genelinde başta İngilizce olmak üzere çeşitli dillerde kullanılmak üzere oluşturulmuş gövdeleme algoritmaları bulunmaktadır. En popüler algoritmalarından birisi Porter algoritmasıdır. Algoritma karmaşık son ekleri basit son eklerden oluşan bileşikler olarak ele alır ve basit son eklerin birbirini izleyen birkaç adımda silinmesi şeklinde çalışır (Porter, 1980, s. 130). Bunun dışında Lovins (1968), Dawson (1974), Krovetz (1993) gibi bilim insanları tarafından geliştirilen gövdeleme algoritmaları da literatürde önemli yer tutmaktadır ve metin madenciliği çalışmalarında kullanılmaktadır.

Türkçe sondan eklemeli bir dil olduğundan gövdeleme için kullanılabilecek çok fazla algoritma bulunmamaktadır. Var olan algoritmaların da henüz istenen verimlikte sonuç verdiği söylenemez. Ancak son dönemlerde Türkçe gövdeleme algoritmalarının geliştirilmesine yönelik önemli çalışmalar da göz ardı edilemez. Bunlardan en fazla bilineni açık kaynak kodlu Türkçe doğal dil işleme kütüphanesi olan zemberektir. Bunun yanında İstanbul Teknik Üniversitesi Doğal Dil İşleme grubu (İTÜ NLP) Türkçe için gövdeleme algoritması oluşturmaya yönelik önemli çalışmalar yapmaktadır.

#### **2.4.2.3. Metin dönüşümü**

Metin dönüşümü ön işleme süreci sonucunda elde edilen metin verilerinin yapısal veri temsiline oluşturulması sürecidir. Bu süreçte metinler sayısal forma dönüştürülmektedir. Belge gösterimi olarak da ifade edilen metin dönüşümü sürecinde vektör uzay modeli, olasılıksal konu modeli ve istatistiksel dil modeli gibi değişik belge gösterim yöntemleri bulunmakla birlikte en sık kullanılan yöntem vektör uzay modelidir. Vektör uzay modelinde belgeler n-boyutlu bir alanda vektörler olarak temsil edilir; burada n, kelime haznesindeki benzersiz terimlerin sayısıdır (Singh, Devi and Mahanta, 2017, s. 1780).

Vektör uzay modelinde her bir belge, kelime haznesini oluşturan her bir kelimenin değerleri bütününden oluşan bir vektör ile temsil edilir. Her bir belgeyi temsil eden vektörler de birleşerek bütün belgeleri içeren bir matris oluştururlar. Bu matris, belgelerden ve belgelerdeki terimlerden oluştuğu için belge terim matrisi olarak ifade edilmektedir. Belge terim matrisinin her bir satırı bir belge vektörünü gösterirken, sütunları terim vektörlerini göstermektedir. Bu vektörlerdeki kelime değerlerinin belirlenmesinde ise genellikle üç yöntem kullanılmaktadır. Bunlar; ikili gösterim (binary), terim sıklığı (term frequency) ve ters doküman sıklığıdır (inverse document frequency).

#### **2.5. Muhasebe Alanında Yapılan Metin Madenciliği Çalışmaları**

Muhasebe Literatüründe, faaliyet raporları ve sürdürülebilirlik raporları gibi büyük miktarda metin veri içeren raporların analizi ile ilgili birçok çalışma bulunmaktadır. Bu çalışmaların büyük bölümü içerik analizi gibi geleneksel yöntemleri kullanmaktadır. Bunun yanında son yıllarda teknolojik gelişmenin de etkisiyle bu raporlarının analizinde metin madenciliği kullanımının yaygınlaştığı görülmektedir.

Muhasebe literatüründe metin madenciliği alanında Frazier, Ingram ve Tennyson (1984) tarafından yapılan çalışmayı, metin madenciliğini bugünkü manada değerlendiren ilk çalışma olarak değerlendirebiliriz. Frazier, Ingram ve Tennyson (1984) yaptıkları çalışmada metin verilerin istatistiksel olarak değerlendirilmesi amacıyla geliştirilen WORDS adlı bir içerik analiz sisteminden oluşan bir yöntemi faaliyet raporlarındaki metin verilerine uygulamışlardır. Çalışmada WORDS'ün, tipik olarak incelenmesi zor olan verilerin objektif, sistematik sınıflandırması ve sayısallaştırılması için potansiyel olarak faydalı olabileceği ifade edilmektedir.

Kohut ve Segars (1992) yaptıkları çalışmada kurumsal iletişim stratejisindeki kalıpları incelemek üzere işletme yöneticilerinin hissedarlara sunduğu mektupları incelemişlerdir. Sonuçlar yüksek ve düşük performans gösteren firmaların, kendi yöneticilerinin yazılarında vurgulanan temalara dayanarak doğru bir şekilde sınıflandırılabilceğini göstermektedir.

Subramanian, Insley ve Blackwell (1993) performans ile yıllık raporların okunabilirliği arasındaki ilişkiyi test etmişlerdir. Bilgisayara dayalı analiz yöntemi kullanılarak yapılan 60 adet faaliyet raporunun stil analizi sonuçları, iyi performans gösteren işletmelerin faaliyet raporlarının, düşük performans gösteren işletmelerin faaliyet raporlarından daha kolay okunduğunu göstermiştir.

Back ve diğerleri (2001) dünya çapındaki orman şirketlerinin performansını karşılaştırmak için faaliyet raporlarından elde edilen sayısal verilerle metin verilere ilişkin sonuçları akıllı kodlama, SOM ve belge histogramları kullanarak karşılaştırmışlardır. Araştırma 1985-1989 dönemi için uluslararası kağıt hamuru ve kağıt endüstrisindeki 76 şirkete uygulanmıştır. Uygulamada öncelikle, şirketler bir taraftan sayısal veriye ve diğer taraftan metin verisine göre, kendi kendini düzenleyen haritalar (SOM) kullanarak kümelennmiştir. Daha sonra, kümeleme yöntemleri ile üretilen sonuçlar karşılaştırılmıştır. Karşılaştırma sonuçları, sonuçlar arasında bir fark olduğunu göstermektedir. Araştırmada metin ve sayısal veri kümeleme sonuçlarındaki farklılıkların, metinde performansı abartma yönünde hafif bir eğilime bağlı olduğu sonucuna ulaşılmıştır.

Clatworthy ve Jones (2003) İngiltere’de vergi öncesi kârlarındaki yüzde değişime göre listelenen performansı iyileşen 50 işletme ile performansı kötüleşen 50 işletmenin faaliyet raporlarında yer alan yönetici mektuplarını incelemişlerdir. Araştırma sonuçlarına göre, performansını iyileştiren işletmeler kötü haberlerden ziyade iyi haberlere odaklanır. Ancak, performansı düşüş gösteren işletmeler, performanslarının düşüklüğünün niteliğini ve sebeplerini tartışmamakta ve analiz etmemektedir. En iyi ihtimalle, hem iyi hem de kötü haberi eşit olarak tartışır; en kötüsü iyi haberlere odaklanırlar. Bu sonuçlar göstermektedir ki yönetim, yönetici mektuplarını finansal performansa olumlu bir vurgu yapmak için kullanır ve dikkati düşük finansal sonuçların sorumluluğundan uzaklaştırmaya çalışır.

Kloptchenko ve diğerleri (2004) yaptıkları çalışmada faaliyet raporunun metin bölümünün gelecekteki finansal performans hakkında bazı göstergeler içerip içermediğini görmek için finansal raporlardaki sayısal ve metin verilerin analizi için veri

ve metin madenciliği yöntemlerini bir arada kullanmışlardır. Sayısal verilerin analizi, kendi kendini düzenleyen haritalar kullanılarak yapılmıştır. Metin verilerin analizi ise prototip eşleştirmeye dayalı metin kümeleme kullanılarak yapılmıştır. Analiz, telekomünikasyon sektöründeki üç önde gelen şirketin üç aylık raporlarına uygulanmıştır. Üç aylık raporlardan elde edilen sayısal ve metin bilgileri analiz edildikten sonra elde edilen sonuçlar, raporlardaki metinleri analiz ederek finansal performansta gelecekte meydana gelecek bazı değişikliklerin tahmin edilebileceğini kanıtlamıştır. Çalışma sonuçlarına göre, bir işletmenin finansal performansında çarpıcı bir değişiklik meydana gelmeden önce, finansal raporun yazılı tarzında bir değişiklik meydana gelmektedir. Raporun tonu, şirketin gelecekteki performansına daha yakın olma eğilimindedir. İşletmenin pozisyonu önümüzdeki çeyrek boyunca sayısal olarak daha zayıf olacaksa, cari çeyrek raporu, gerçek finansal performans aynı kalsa da, daha karamsar olmaya meyillidir.

Cecchini (2005) yaptığı çalışmada hile ve iflasın tespitinde metin verilerinin kullanımının başarı sağladığı sonucuna ulaşmıştır. Cecchini sağlıklı şirketler ile iflas eden şirketlerin 10-K raporlarındaki yönetici tartışmaları ve analizleri kısmındaki metin verileri arasında ciddi farklar bulunduğunu ifade etmektedir.

Li (2006) faaliyet raporlarındaki risk duyarlılığının işletmenin gelecek dönem kazancı ve hisse senedi getirileri üzerindeki etkilerini incelemiştir. Faaliyet raporlarındaki risk duyarlılığı, 10-K raporlarında risk veya belirsizlikle ilgili kelimelerin sıklığını sayarak ölçülmüştür. Li risk duyarlılığında artış olan işletmelerde gelecek yıl daha fazla olumsuz kazanç değişikliği olduğu sonucuna ulaşmıştır. Ayrıca araştırma sonuçları, borsa faaliyetlerinin yıllık raporların metin bölümündeki halka açık bilgilere tepki verdiğini göstermektedir.

Nelson ve Pritchard (2007) firmaların 1995 tarihli Özel Menkul Kıymetler Dava Reformu Yasası'nın güvenli liman hükmü bağlamında işletmelerin açıklamalarındaki ihtiyatlı dil kullanımını araştırmışlardır. Çalışma sonucunda daha yüksek dava riskine maruz kalan işletmelerin daha ihtiyatlı bir dil kullandıkları, açıklamayı yıldan yıla daha fazla güncelledikleri ve daha okunaklı bir dil kullandıkları tespit edilmiştir. Sonuçlar, işletmelerin beklenen dava maliyetlerini azaltmak için açıklama politikaları benimsemiş olduklarını göstermektedir.

Qiu (2007) faaliyet raporlarından şirket finansal performansını tahmin etmek için bir sınıflandırma modeli oluşturup oluşturmayacağını araştırmıştır. Araştırma sonuçları,

faaliyet raporlarının metin içeriği kullanılarak finansal performans tahmin modellerinin başarıyla oluşturulabileceğini doğrulamaktadır.

Shirata ve Sakagami (2008) yaptıkları çalışmada, anahtar kelimeler kullanılarak Japonya'daki finansal raporlarda (Japonya 10-K) açıklanan metin verileri analiz ederek, süreklilik konusunda sorunlu işletmeler ile sorunsuz işletmeler arasındaki farkı tespit etmeye çalışmışlardır. Araştırma sonuçları, finansal raporlardaki bazı finansal olmayan anahtar kelimelerin kurumsal finansal pozisyonu değerlendirmek için kullanılabileceğini göstermektedir.

Li (2008) Naive Bayesian makine öğrenme algoritması kullanarak işletmelerin 10-K ve 10-Q dosyalarında ileriye dönük ifadelerin tonunu ve içeriğini incelemiştir. Araştırma sonucunda, ileriye dönük ifadelerin tonunun gelecekteki finansal performansla pozitif bir ilişkisi olduğu tespit edilmiştir.

Barkemeyer ve diğerleri (2009) sürdürülebilirlikteki ana eğilimleri tespit etmek ve farklı ülke ve bölgelerde zaman içinde kavramların ve konuların medyadaki sunumlarının nasıl geliştiğini analiz etmek amacıyla bir çalışma yapmışlardır. Çalışma dünya genelinde önde gelen 115 ulusal gazetede, Ocak 1990 ile Temmuz 2008 arasındaki döneme ait yaklaşık 340.000 gazete sayısında yayınlanan 20.500.000 makaleyi kapsamaktadır. Yapılan çalışma sonucunda, sürdürülebilirlikle ilgili medya haberlerinde 1990'dan itibaren bir artış olduğu görülmüştür. Ayrıca, sürdürülebilirlikle ilgili kavramları kapsayan haberlerde belirgin bölgesel ve ulusal farklılıklar bulunduğu ifade edilmektedir.

Balakrishnan ve diğerleri (2010) 10-K ve 10K-405 raporlarındaki metin verilerinin piyasa performansını tahmin etmeyi sağlayacak bilgiler içerip içermediğini araştırmışlardır. Beş yıl boyunca 1236 işletme tarafından yayınlanan 4280 rapordaki metin verilerine makine sınıflandırma tekniklerini uygulanmıştır. Çalışma sonuçları metin verilerinin gelecekteki muhasebe ve piyasa performansının tahmininde kullanılabileceği göstermiştir.

Goel ve diğerleri (2010) hile tespitinde yıllık raporlardaki metin verilerinin kullanılabilirliğini araştırmışlardır. Bu çalışmada, yıllık raporların metin verileri doğal dil işleme araçları kullanılarak incelenmiştir. Çalışma sonuçları, dilsel özelliklerin kullanılmasının hileleri tespit etmede etkili bir araç olduğunu göstermektedir.

Cecchini ve diğerleri (2010) iflas, hile gibi kötü finansal olayları tahmin etmek için finansal metni otomatik olarak analiz eden bir metodoloji geliştirmeyi amaçlamışlardır.

Çalışmada 10-K raporlarındaki yönetim tartışma ve analiz bölümlerinden oluşturulan sözlükler kullanılmıştır. Çalışma sonuçları oluşturulan sözlüklerin hileyi %75 oranında, iflası ise %80 oranında doğru tahmin etmeyi sağladığını göstermektedir.

Shirata ve diğerleri (2011) Japonya'daki işletmelerin faaliyet raporlarındaki metin verilerinden iflasın tahmin edilmesini sağlayacak temel ifadelerin belirlenmesine yönelik bir araştırma yapmışlardır. Araştırma sonuçları, faaliyet raporunun aynı bölümünde bazı ifadelerin “temettü” veya “birikmiş kazanç” kelimeleri ile birlikte kullanılması durumunda, bu kelimelerin iflas tahmininde etkili olabileceğini ortaya koymuştur.

Glancy ve Yadav (2011) finansal raporlama hilelerinin tespiti için faaliyet raporlarındaki metin verilerinin kullanımı için bir hile tespit modeli önermişlerdir. Çalışma 10-K raporlarındaki yönetim tartışma ve analiz bölümü ile sınırlandırılmıştır. Araştırma sonucunda oluşturulan model finansal raporlama hilelerini, faaliyet raporlarının metninden tespit etmenin mümkün olduğunu göstermektedir.

Humpherys ve diğerleri (2011) Dilbilimsel Güvenilirlik Analizi kullanarak hileli finansal tabloların belirlenmesine yönelik bir çalışma yapmışlardır. Çalışmada en iyi performansı, Naif Bayes ve C4.5 karar ağacı sınıflandırıcıları 10 değişkenli modeli kullanarak elde edilmiştir. Her ikisi de %67,3 doğruluk elde etmiştir. Sınıflandırma sonuçlarındaki başarı, dilbilimsel analiz modellerinin finansal tablolarda hileyi tespit etmede kullanılabilir olduğunu göstermektedir. Ayrıca sonuçlar, denetçilerin şüpheli finansal açıklamaları işaretlemek ve hile riskini değerlendirmek için dilbilimsel analizlerin potansiyel kullanımını desteklemektedir.

Davis, Piger ve Sedor, (2012) yaptıkları çalışmada 1998-2003 yılları arasında PR Newswire'da yayınlanan yaklaşık 23.000 adet çeyreklik gelir açıklama basın bülteninin bir örneğini analiz etmişlerdir. Çalışmada iyimser ve karamsar olarak nitelendirilen kelimeleri sayan, yerleşik bir metin analiz yazılımı olan DICTION kullanılmıştır. Araştırma sonuçları basın bültenlerinde kullanılan net iyimser dil seviyelerinin, gelecek çeyreklerdeki işletme performansının tahmini için kullanılabileceğini göstermiştir.

Goel ve Gangolly (2012) metin verilerin hile tespitinde kullanılabilirliği test etmek ve hile yapan işletmeler ile yapmayan işletmeler arasında yazım ve sunum stilleri açısından görülebilir farklılıklar olup olmadığını görmek için yıllık raporlardaki metin verileri incelemişlerdir. Araştırmada, hile tespitinde yıllık raporların metin içeriğinin kullanılabilir olduğu sonucuna ulaşılmıştır.

Gupta ve Gill (2012) yaptıkları çalışmada, finansal tablolarda yer alan metinlerdeki gizli ipuçlarını analiz ederek finansal tablo hilelerini tespit etmek için bir metin madenciliği yaklaşımı önermişlerdir. Çalışmada destek vektör makineleri ve kelime haznesi (Bag of Words) yaklaşımı kullanılmıştır.

Zaki ve Theodoulidis (2013) ABD Menkul Kıymetler ve Borsa Komisyonu (SEC) dava bültenlerinden finansal kavramların çıkarılması sürecini otomatikleştirmek için dil temelli bir metin madenciliği yaklaşımı önermektedir. Bu çalışma ile, farklı piyasa manipülasyon türlerini anlamak ve analistlere araştırma sürecinde yardımcı olabilecek SEC dava bültenlerinden gelen bilgileri yakalamak için dil tabanlı bir metin madenciliği yaklaşımı geliştirilmiştir.

Chakraborty, Chiu ve Vasarhelyi (2014) muhasebe araştırmalarını otomatik olarak sınıflandırmak için anlamsal ayrıştırma, bilgi alma ve veri madenciliği tekniklerini kullanma olasılığını araştırmışlardır. Çalışmanın örnek makaleleri on muhasebe dergisinde 1984 ve 2008 arasında yayınlanmış makalelerden oluşmaktadır. Araştırmada sınıflandırma sürecini incelemek ve hassaslaştırmak için üç çalışma yapılmıştır. Araştırma makalelerinin anahtar sözcükleri ve özetleri ilk iki çalışmaya uygulanmış, son çalışmada ise yalnızca özetler kullanılmıştır. Bulgular anlamsal ayrıştırma ve veri madenciliği araçlarının muhasebe literatürünün otomatik sınıflandırmasını sağladığını göstermektedir.

Liew, Adhitya ve Srinivasan (2014) proses endüstrilerinde sürdürülebilirlik eğilimlerini ve uygulamalarını tanımlamak için metin madenciliğini kullanmışlardır. Çalışmada petrol/petrokimyasallar, dökme/özel kimyasallar, ilaç ve tüketici ürünleri olmak üzere endüstrinin dört ana sektörü incelenmiştir. Çalışma bulguları, dört sektörün en üst sürdürülebilirlik odağının çok benzer olduğunu ortaya koymaktadır. Bunlar; sağlık ve güvenlik, insan hakları, sera gazı azaltma, enerji/enerji verimliliğini koruma ve topluluk yatırımdır. Ayrıca çalışmada sektöre özgü sürdürülebilirlik sorunları da belirlenmiştir.

Rivera ve diğerleri (2014) yapılandırılmamış dijital haber makalelerini metin madenciliği yöntemleriyle analiz ederek sürdürülebilirlik göstergelerini tanımlamanın, izlemenin ve raporlamanın uygulanabilirliğini araştırmışlardır. Çalışma sonuçları metin madenciliğinin sürdürülebilirlik göstergelerinin belirlenmesi, izlenmesi ve raporlanması için yararlı bilgiler sağlayabileceğini göstermiştir.

Shahi, Issac ve Modapothala (2014) kurumsal sürdürülebilirlik raporlarının değerlendirmesini otomatikleştirmeye yönelik bir çalışma yapmışlardır. Çalışma kurumsal sürdürülebilirlik raporlarının denetimli öğrenmeye dayalı bir metin madenciliği yazılımı kullanılarak analiz edilebileceğini savunmaktadır. Çalışma sonuçları yazılımın kurumsal sürdürülebilirlik raporlarının değerlendirilmesinde ciddi başarı gösterdiğini ortaya koymuştur.

Purda ve Skillicorn (2015) faaliyet raporlarının yönetim tartışmaları ve analizler bölümünde kullanılan dile dayanarak hileli ve hileli olmayan raporlar arasında ayrım yapmak için bir yöntem geliştirmeye çalışmışlardır. Çalışmada destek vektör makineleri (DVM) sınıflandırma algoritması kullanılmıştır. Çalışma sonuçları geliştirilen yöntemin hileli raporlar ile hileli olmayan raporları başarılı bir şekilde sınıflandırdığını kanıtlamıştır.

Kamaruddin ve diğerleri (2015) bir finansal belgeler koleksiyonundan cümle sapmalarını saptayabilen bir metin madenciliği sistemi önermektedirler. Sistem, bir bankanın finansal tablolarını uygulanmıştır. Bulgular, önerilen sistemin belgelerde meydana gelen sapma cümleleri tanımlayabildiğine dair geçerli kanıtlar sunmaktadır. Yapılan çalışmada, önerilen sistem ile tespit edilen sapmaların, iş kararlarını iyileştirmek için yetkililere fayda sağlayabileceği ifade edilmektedir.

Bala ve diğerleri (2015) Sürdürülebilirlik Trend Analizi olarak adlandırdıkları bilgisayar tabanlı bir uygulama için model önermişlerdir. Bu çalışmanın amacı gerçek zamanlı, büyük veri tabanlı dinamik sürdürülebilirlik sinyallerini elde etmektir. Çalışmada, dinamik sürdürülebilirlik sinyalleri, kamu şirketleri ve diğer büyük kuruluşların sürdürülebilirlik girişimlerinin ve yatırımlarının miktar ve kalitesinin ileriye dönük göstergeleri olarak ifade edilmektedir.

Heidari ve Felden (2015) metin madenciliği yaklaşımının, finansal tablo dipnotlarında yer alan metin verilerin analizine yardımcı olmak için uygun bir çözüm olup olmadığını araştırmışlardır. Çalışmada finansal tablo dipnotlarının otomatik analizi için sınıflandırma algoritmaları kullanılmıştır. Uygulama sonuçları, metin madenciliğinin finansal raporların yapılandırılmamış kısımlarının analizini kolaylaştırmak için uygun bir çözüm olabileceğini göstermektedir.

Gémar ve Jiménez-Quintero (2015) yaptıkları çalışmada, Twitter, Facebook, LinkedIn, YouTube gibi sosyal medya sitelerinde işletmeler ile ilgili verileri analiz etmek için bir metin madenciliği aracı kullanmışlardır. Araştırma bir otel örneğine üzerine

odaklanmıştır. Araştırma sonucunda, metin madenciliği ve finansal performans yoluyla elde edilen çeşitli değişkenler arasında bir bağımlılık bulunmuştur. Sonuçlar, metin madenciliği teknikleri kullanılarak yapılan sosyal medya analizinin finansal performansı artırmanın bir yöntemi olabileceğini göstermektedir.

Liu ve Moffitt (2016) SEC yorum mektuplarının yoğunluğunu incelemek için metin incelemesi kullanmışlardır. Yorum mektuplarının tarzına dayalı bir yoğunluk ölçüsü geliştirilmiştir. 10-K raporları ile ilgili ilk yorum mektubu örneği üzerinde deneysel analiz yapılmıştır. Sonuçlar, yoğunluğun, incelenen 10-K raporlarının yeniden ifade edilme olasılığı ile pozitif ilişkili olduğunu göstermektedir.

Rich, Roberts ve Zhang (2016) belediye yönetim tartışma ve analiz açıklamalarının bilgi içeriğini incelemişlerdir. Çalışmada, 364 belediyenin 2011 yılı yönetim tartışma ve analiz bildirim örneği kullanılmıştır. Çalışma sonuçlarına göre, belediye yönetim tartışma ve analiz açıklamalarındaki olumlu dilin güçlü bir finansal raporlamanın işareti olduğu değerlendirilmektedir. Buna göre finansal açıklamaların metinsel analizi, gelecekteki finansal raporlama kalitesini tahmin etmede faydalı bilgi sağladığı ifade edilmektedir.

Hajek ve Henriques (2017) kurumsal faaliyet raporlarında yer alan finansal bilgiler ve yönetsel yorumlardan elde edilen spesifik özellikleri birleştirerek geliştirilmiş bir finansal hile tespit sisteminin geliştirilip geliştirilemeyeceğini araştırmışlardır. Çalışmada, yıllık raporlardaki düşük nispi negatif kelime sıklığının hile yapmayan işletmeleri gösterebileceği sonucuna ulaşılmıştır.

Tsai ve Wang (2017) finansal tablolardaki metin veriler ile finansal risk arasındaki ilişkileri incelemişlerdir. Çalışmada halka açık şirketler tarafından yıllık olarak yayınlanan geniş bir finansal rapor koleksiyonu kullanılmıştır. Çalışma bulguları, finansal tablolardaki metin verilerin risk tahmini açısından önemini ortaya koymaktadır.

Ağdeniz (2017) metin madenciliğini kullanarak işletmelerin kurumsal yönetim niteliklerinin tahminine yönelik bir araştırma yapmıştır. Çalışmada veri seti olarak BIST 100 ve XKURY'de yer alan 83 işletmenin faaliyet raporları kullanılmıştır. Makine öğrenmesi algoritmalarından KNN, Naive Bayes ve Karar Ağacı algoritmalarının kullanıldığı çalışmada, işletmelerin kurumsal yönetim niteliklerinin belirlenmesinde en başarılı performansı KNN algoritması göstermiştir. Uygulama sonuçları, faaliyet raporlarının içerdiği yapısal olmayan verilerin analizinin menfaat sahiplerine faydalı bilgi sağladığını göstermektedir.

Wahyuni, Febrianto ve Rahman (2018) işletme performansı ile faaliyet raporlarının okunabilirliği arasındaki ilişkiyi incelemişlerdir. Çalışmada işletme performansı göstergesi olarak kârlılık oranları kullanılmıştır. Araştırma 2013-2017 yılları arasında Endonezya Menkul Kıymetler Borsası'na kote olmuş işletmelere uygulanmıştır. Araştırma modeli çoklu doğrusal regresyon analizidir. Araştırma sonuçları işletmelerin finansal performansı ile faaliyet raporlarının okunabilirliği arasında negatif bir ilişki olduğunu göstermektedir. Buna göre firmanın performansı ne kadar yüksek olursa, faaliyet raporunun okunmasının o kadar kolay olduğu sonucuna ulaşılmıştır.

Duan, He ve Zhong (2018) yaptıkları çalışmada, metin madenciliği araçlarına dayanarak kurumsal sosyal sorumluluk bilgilerinin açıklanmasının öznelliğini ve nesnelliğini değerlendirmek için yeni bir yöntem önermeyi amaçlamışlardır. Sonuçlar, Çinli borsa şirketlerinin sosyal sorumluluk raporlarının öznel puanlarının genellikle normal dağılımda olduğunu göstermektedir.

Krishnamoorthy (2018) performans göstergelerini kullanarak finansal haber makalelerinin duygu analizini yapmıştır. Çalışmada, duygu tahmini için performans göstergeleri kullanılarak hiyerarşik bir duygu sınıflandırıcısı önerilmiştir. Önerilen yöntem, finansal metni olumlu, tarafsız veya olumsuz olarak sınıflandırmayı amaçlamaktadır. Önerilen yaklaşım ile çalışılan tüm veri setlerinde %68'in üzerinde f-ölçümü ve %80'in üzerinde bir doğruluk ile oldukça başarılı bir sonuç elde edilmiştir.

Lin ve Hsu (2018) Tayvan'da işletmelerin kurumsal sosyal sorumluluk haber raporlarının kurumsal işletme performansı tahminleri üzerindeki etkisini incelemişlerdir. Sonuçlar önerilen modelin kurumsal performans tahmini için ümit verici bir alternatif olduğunu ortaya koymaktadır.

Hájek (2018) anormal hisse senedi getirilerini tahmin etmek için yıllık raporlarda yer alan metin verileri ve finansal tablo bilgilerini birlikte kullanmışlardır. Çalışmada, tahmin doğruluğunu artırmak için finansal göstergeler, okunabilirlik, duygu kategorileri ve kelime haznesi (sözlük) yaklaşımlarının bir birleşimi kullanılmıştır. Çalışma sonuçları, uygulanan yöntemin tahmin kalitesini önemli ölçüde artırdığını göstermektedir.

Boskou, Kirkos ve Spathis (2019) yaptıkları çalışmada otomatik metin analizini kullanarak iç denetim kalitesini değerlendirmeyi amaçlamaktadır. Çalışmanın veri kaynağı faaliyet raporlarındaki iç denetim mekanizması açıklamalarıdır. Sonuçlar, metin analizi kullanılarak geliştirilen sınıflandırma modellerinin iç denetim kalitesini değerlendirmesinde ümit verici bir alternatif olabileceğini göstermektedir.

Rahrovi Dastjerdi, Foroghi, ve Kiani (2019) dışbükey optimizasyon (CVX) yöntemi ve en az mutlak büzülme ve seçim operatörü (LASSO) regresyon yöntemi olmak üzere iki yöntem kullanarak yönetim kurulu raporlarının metinlerini analiz etmişlerdir. Öncelikle, yöneticinin yüksek hile risk endeksini açıklamada en büyük güce sahip olan kelimeler belirlenmiştir. Bu kelimeler kullanılarak, işletme yöneticilerinin yüksek hile riski endeksini tespit edebilecek bir model sunulmuştur. Sonuçlar, her iki yöntemin de yöneticinin yüksek hile risk endeksini yüzde 82,55 ile 91,25 arasında kesin bir şekilde tespit edebileceğini göstermiştir.

Wei ve diğerleri (2019) yaptıkları çalışmada banka risk faktörlerini finansal tablolarda bildirilen metinsel risk açıklamalarından kapsamlı bir şekilde çıkarmayı amaçlamaktadır. Metinsel risk açıklamalarını analiz etmek için yeni bir yarı denetimli metin madenciliği algoritması kullanılmıştır. Veri kaynağı olarak 2010'dan 2016'ya kadar olan 2.189 ABD ticari bankasının 10-K raporlarında yer alan risk açıklamaları kullanılmıştır. Çalışma sonucunda toplamda 21 adet banka risk faktörü belirlenmiştir.

Fenyves ve diğerleri (2019) Macar bilgi-teknoloji hizmetleri sektöründeki işletmelerin muhasebe kanunu tarafından öngörülen mali tablo dipnotlarında bilgi sağlama yükümlülüğünü yerine getirme derecelerini incelemişlerdir. Analiz metin madenciliği yöntemi kullanılarak yapılmıştır. Çalışmanın bulguları, yayınlanan verilerin miktarının yasal olarak öngörülen asgari yükümlülükleri bile tamamıyla karşılamadığını ortaya koymaktadır.

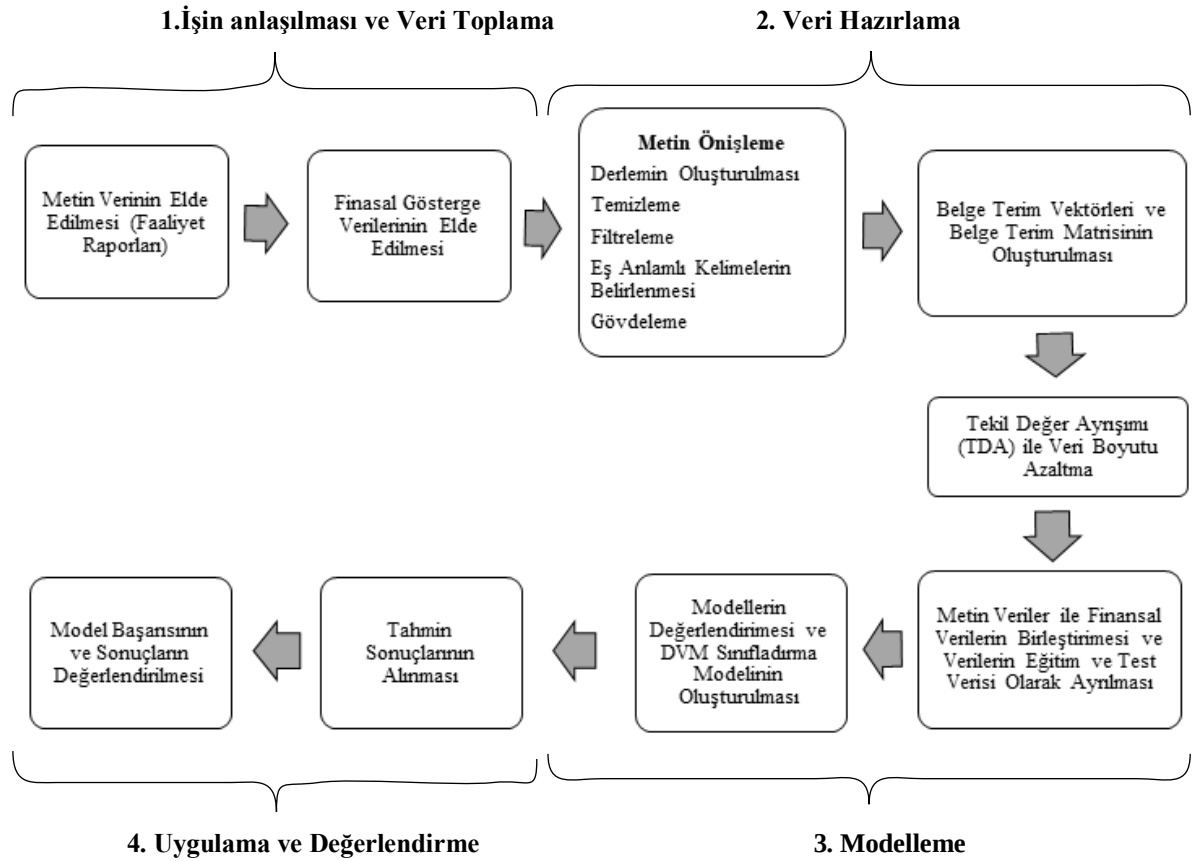
### 3. YÖNTEM

#### 3.1. Araştırma Modeli

Araştırma yöntemi metin madenciliğidir. Araştırmada model olarak ise, kategorik değişkenlerin tahmin edilmesinde kullanılan bir sınıflama modeli olan destek vektör makineleri (DVM) algoritması kullanılmıştır. Araştırma süreci klasik bir metin madenciliği aşamaları takip edilecek şekilde tasarlanmıştır. Buna göre araştırma süreci Şekil 3.1’de gösterildiği üzere 4 aşamadan oluşmaktadır. İlk aşama işin anlaşılması ve verilerin toplanmasıdır. Öncelikle işin değerlendirilmesi ve örneklemin belirlenmesi ile başlanmıştır. Daha sonra faaliyet raporları Kamuyu Aydınlatma Platformu (KAP)’dan ve işletmelerin internet sitelerinden indirilmiştir, finansal performans oranları ise finnet veri tabanından alınmıştır. Bu aşamadan sonra veri hazırlama aşamasına geçilmiştir. Veri hazırlama aşamasında ilk olarak “pdf” şeklinde toplanan faaliyet raporları “txt” uzantılı metin dosyalarına dönüştürülerek derlem (corpus) oluşturulmuştur. Daha sonra metin ön işleme ve metin dönüşümü işlemleri yapılarak metin verileri yapısal veri haline getirilmiştir. Metin madenciliği aracı olarak Statistica programı metin madenciliği aracı (text mining) kullanılmıştır.

Metin ön işlemede temizleme, filtreleme, eş anlamlı kelime belirleme, gövdeleme, dizgeciklere ayırma işlemleri sırasıyla yapılmıştır. Temizleme işleminde çalışma için herhangi bir anlam ifade etmeyen noktalama işaretleri, sayılar ve diğer karakterler silinmiştir. Ayrıca faaliyet raporları içinde yer alan ancak finansal veri niteliği taşıdığı için çalışmada yer almaması gereken finansal tablolar, finansal tablo dipnotları, finansal performansla ilgili açıklamalar (finansal oranlar, temel finansal göstergeler v.b.) ve bağımsız denetim raporları silinmiştir. Filtrelemede metin hakkında bilgi içermeyen ve literatürde gereksiz kelimeler olarak ifade edilen “ve”, “veya”, “bu” gibi kelimeler belirlenerek analiz dışı bırakılmıştır. Ayrıca çok sık tekrar eden veya çok az tekrar eden kelimeler de uçbirim (outlier) olarak görülüp filtreleme işlemine tabi tutulmuştur. Gövdeleme işlemi için ise otomatik bir gövdeleme yöntemi olan tablo arama yöntemi kullanılmıştır. Metin dönüşümü için binary (ikili gösterim) algoritması kullanılarak belge terim vektörleri ve belge terim matrisi oluşturulmuştur. Daha sonra oluşturulan matristeki kelime değerleri tekil değer ayrışımı (TDA) algoritması kullanılarak boyut azaltma işlemine tabi tutulmuştur. Bu işlemin sonucunda veri hazırlama aşaması tamamlanmıştır. Bu aşamadan sonra modelleme aşamasına geçilmiştir. Modelleme aşamasında ilk olarak

yapısal metin veriler ile finansal veriler birleştirilerek eğitim ve test verisi olarak ayrılmıştır. Daha sonra çeşitli seçenekler gözden geçirilerek kategorik değişkenlerin tahmin edilmesinde kullanılan bir sınıflandırma modeli olan destek vektör makineleri (DVM) algoritmasının kullanılmasına karar verilmiştir. Son aşamada ise, Statistica programı kullanılarak veriler analiz edilmiş ve sonuçlar alınarak değerlendirilmiştir.



Şekil 3.1: Araştırma süreci

### **3.2. Araştırmanın Evren ve Örneklemi**

Araştırmanın evreni yıllık faaliyet raporu düzenlemek zorunda olan işletmelerdir. Örneklem olarak verilere ulaşım imkânının kolay olması, çalışmayı ciddi oranda etkileyebilecek bir sektörel farklılığın bulunmaması ve tüm sektörlerde uygulanabilecek ortak bir kelime haznesine (Bag of Words) ulaşılma imkânının daha yüksek olması gibi nedenlerle BİST imalat sanayi işletmeleri seçilmiştir. Ayrıca BİST imalat sanayi sektörünün veri boyutunun büyüklüğünün veri madenciliği doğasına uygun olması da bu sektörün seçilmesinde önemli bir nedendir. Çalışma dönemi olarak ise 2010 ve 2017 yılları seçilmiştir. 2010 ve 2017 yılları arasının seçilmesinin nedeni, bu dönemde işletmelerin finansal göstergelerinin değerlendirilmesini önemli derecede etkileyebilecek ciddi bir ekonomik kriz, doğal afet gibi bir durumun gerçekleşmemesidir. Bu bağlamda BİST imalat sanayi sektöründe işlem gören 179 işletmenin 2010-2017 yılları arasındaki 8 yıllık döneme ait 1249 adet dokümanına ulaşılabilmiş ve çalışmaya dahil edilmiştir.

### **3.3. Veri Toplama Tekniği ve Araçları**

Finansal performans göstergeleri olan oranlar; finnet internet sitesi veri tabanından elde edilmiştir (Finnet 2000 Plus, 2018).

Faaliyet raporları ise, kamuyu aydınlatma platformu ve işletmelerin internet sitelerinden elde edilmiştir.

### **3.4. Veri Analizi**

Veri analizinde Statistica programı kullanılmıştır. Veri analiz sürecinde gövdeleme işlemi için tablo arama yöntemi, belge terim vektörünün oluşturulması aşamasında binary algoritması ve veri boyutunun azaltılması için ise TDA algoritması kullanılmıştır. Veri analizinin son aşamasında ise sınıflandırma modeli olarak DVM algoritması kullanılmıştır.

#### **3.4.1. Tablo Arama Yöntemi**

Tablo arama yönteminde kelime kökü, gerek yapım eki, gerek çekim eki almış kelimeler, gerekse ek almamış kelimeler için kendisine karşılık gelen kök kelimeyi içeren (Tablo 3.1’de gösterilen örnekteki gibi) bir tabloya göre belirlenir (Frakes, 1992). Bu yöntemde kelimenin kökünü bulmak için tablo, program tarafından kontrol edilir ve tabloda kelimenin eşleştiği sorgu kelime kökü olarak kabul edilir. Bu yöntem aynı

zamanda sözlük tabanlı algoritmalar olarak da adlandırılır. Bunlar istisnai durumlara dikkat edebilecek basit ve kullanımı kolay tekniklerdir. Ancak yöntem başka bazı kök bulma yöntemleri gerektirebilir ve yöntemde sözlüğün (tablonun) dışındaki kelimelerin kullanılması mümkün değildir (Singh and Gupta, 2017, s. 161).

**Tablo 3.1:** Arama tablosu örneği

| Kelime     | Kök     |
|------------|---------|
| denetim    | denetim |
| denetimden | denetim |
| denetimler | denetim |
| madde      | madde   |
| maddesi    | madde   |
| maddesinin | madde   |

Tablo arama yöntemi yukarıda ifade edildiği üzere olumsuz yönleri bulunmakla birlikte Türkçe için kullanışlı bir yöntemdir. Zira Türkçe için, zemberek gibi gövdeleme algoritmaları geliştirilmiş olsa da, gövdeleme çalışmalarının henüz istenen düzeye ulaştığı söylenemez (Sönmez, 2017, s.76; Yücebaş ve Tintin, 2017). Ayrıca bu algoritmaların metin madenciliği programları ile uyumu ayrı bir sorun olarak karşımıza çıkmaktadır. Bu nedenle faaliyet raporlarındaki muhasebe ve finans alanlarına özgü ifadelerde göz önünde bulundurulduğunda, gövdeleme işlemi için tablo arama yönteminin kullanılması uygun olacaktır.

### 3.4.2. Vektör Uzak Modeli ve Binary Algoritması

Vektör uzak modeli (vector space model), bir indeks teriminin vektörü olarak metin belgelerini sunan cebirsel bir modeldir. Vektörler metinlerin temsilinde kullanılmaktadır. Metin içindeki her bir kelime, vektörlerin elemanıdır. Vektörün boyutu belgeden çıkacak olan kelime sayısı ile orantılıdır. Vektör içinde kelimelere ait sayısal bir değer olur. Bu sayısal değerlerin tespiti için farklı algoritmalar kullanılabilir. Bunlardan birisi de binary algoritmasıdır.

Binary algoritması, belgede ilgili kelimenin olması veya olmaması durumuna göre (1,0) olacak şekilde belge terim vektörü oluşturulması esasına dayanır. Burada her bir terim bir kelimeye karşılık gelmektedir.

Vektör uzak modelinde Tablo 3.2’de görüldüğü üzere her bir belgedeki metin, dizgeciklere ayrılarak ve sayısallaştırılarak Belge Terim Vektörü oluşturulur. Oluşturulan bu Belge Terim Vektörleri’nin birleşiminden ise Belge Terim Matrisi elde edilir.

**Tablo 3.2:** Belge terim matrisi örneği

|        | Terim1 | Terim2 | ...  | TerimN |
|--------|--------|--------|------|--------|
| Belge1 | 1      | 0      | .... | 1      |
| Belge2 | 0      | 1      | .... | 1      |
| .....  | 1      | 1      | .... | 0      |
| BelgeN | 1      | 0      | .... | 0      |

### 3.4.3. Tekil Değer Ayırışımı (Singular Value Decomposition)

Tekil değer ayrışımı (TDA) boyut küçültme biçimidir, ancak bilgi kaynaklarını ortadan kaldırmak yerine değişkenlerde yerleşik olan toplam bilgiye eklenir. TDA bunu mevcut değişkenlerin doğrusal kombinasyonlarını hesaplayarak yapar. Değişkenlerden bilginin çoğunu alan çok daha küçük ve daha yönetilebilir bir dizi doğrusal kombinasyonla çok sayıda değişkenin azaltılmasını veya özetlenmesini sağlar. Veri matrisinin boyutunu azaltmak için veri madenciliği ve tahmine dayalı modellemede TDA yaygın olarak kullanılan bir tekniktir (Miner ve diğerleri, 2012, s. 935-936).

TDA en basit anlatımla bir matrisi 3 parçaya ayırarak tutar ve bu üç parçayı kullanarak aynı matrisi daha küçük boyutta yeniden oluşturur.

$$M = U\Sigma V \quad (3.1)$$

Burada;

U: Üniter matris (unitary matrix) olmaktadır

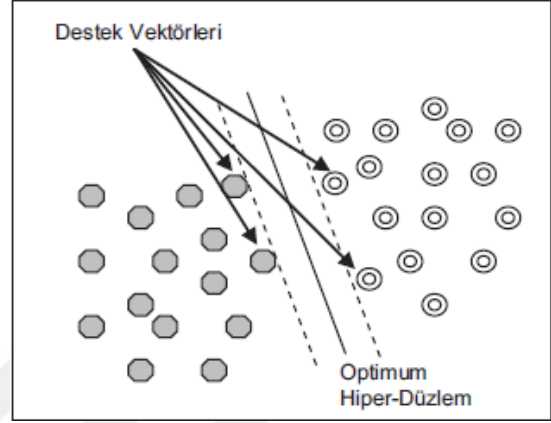
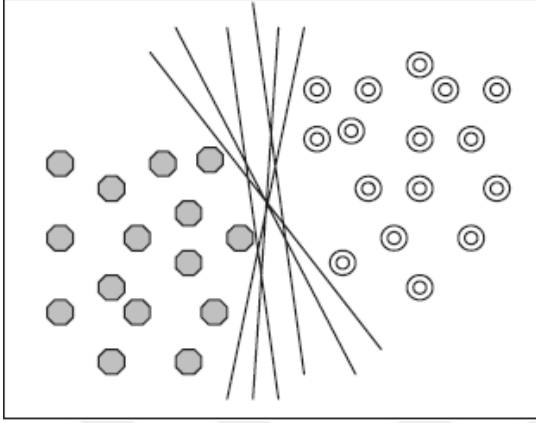
V: M matrisinin birimdik (orthonormal) özelliklerini tutan matristir

$\Sigma$ : köşegen matris (diagonal matrix) olup tekil değerleri tutmaktadır

### 3.4.4. Destek Vektör Makineleri (Support Vector Machines)

Destek vektör makineleri (DVM), 1995 yılında Vladimir Vapnik, Bernhard Boser ve Isabelle Guyon tarafından geliştirilmiştir (Cortes and Vapnik, 1995, s. 276). DVM istatistiksel öğrenme teorisine dayalı kontrollü bir sınıflandırma algoritmasıdır. DVM'nin sahip olduğu matematiksel algoritmalar başlangıçta iki sınıflı doğrusal verilerin sınıflandırılması için tasarlanmış, daha sonra çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması için genelleştirilmiştir. DVM ile sınıflandırmada genellikle  $\{-1, +1\}$  şeklinde sınıf etiketleri ile gösterilen iki sınıfa ait örneklerin, eğitim verisi ile elde edilen bir karar fonksiyonu yardımıyla birbirinden ayrılması hedeflenmektedir. Söz konusu karar fonksiyonu ile eğitim verisini en uygun şekilde ayırabilecek hiper-düzlem bulunur. Şekil 3.2'de gösterildiği üzere iki sınıflı verileri birbirinden ayırabilen birçok hiper-düzlem çizilebilir. Ancak DVM'nin amacı kendisine en yakın noktalar arasındaki

uzaklığı maksimuma çıkaran hiper-düzlemi bulabilmektir. Şekil 3.3’de görüldüğü üzere sınırı maksimuma çıkararak en uygun ayrımı yapan hiper-düzleme optimum hiper-düzlem denir ve sınır genişliğini sınırlandıran noktalar ise destek vektörleri olarak isimlendirilir (Kavzoğlu ve Çölkesen, 2010).



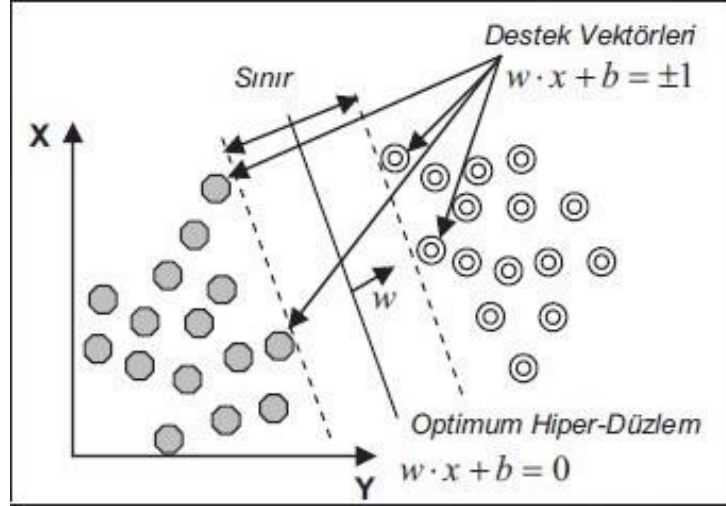
**Şekil 3.2:** İki sınıflı bir problem için hiper-düzlemler **Şekil 3.3:** Optimum hiper-düzlem ve destek vektörleri

Doğrusal olarak ayrılabilen iki sınıflı bir sınıflandırma probleminde DVM’nin eğitimi için  $k$  sayıda örnekten oluşan eğitim verisinin  $\{x_i, y_i\}$ ,  $i=1, \dots, k$  olduğu kabul edilirse, optimum hiper-düzleme ait eşitsizlikler aşağıdaki şekilde olur:

$$w \cdot x_i + b \geq +1 \text{ her } y = +1 \text{ için} \quad (3.2)$$

$$w \cdot x_i + b \leq -1 \text{ her } y = -1 \text{ için} \quad (3.3)$$

Burada  $x \in R^N$  olup  $N$ -boyutlu bir uzayı,  $y \in \{-1, +1\}$  ise sınıf etiketlerini,  $w$  ağırlık vektörünü (hiper-düzlemin normali) ve  $b$  eğilim değerini göstermektedir. Optimum hiper-düzlemin belirlenebilmesi için bu düzleme paralel ve sınırlarını oluşturacak iki hiper-düzlemin belirlenmesi gerekir. Şekil 3.4’de gösterildiği üzere Bu hiper-düzlemleri oluşturan noktalar destek vektörleri olarak adlandırılır ve bu düzlemler  $w \cdot x_i + b = \pm 1$  şeklinde ifade edilirler (Osuna, Freund and Girosi, 1997).



**Şekil 3.4:** Doğrusal olarak ayrılabilen veri setleri için hiper-düzlem

Özetle, doğrusal olarak ayrılabilen iki sınıflı bir problem için karar fonksiyonu aşağıdaki şekilde yazılabilir (Osuna, Freund and Girosi, 1997, s. 9):

$$f(x) = \text{sign} \left( \sum_{i=1}^k \lambda_i y_i (x * x_i) + b \right) \quad (3.4)$$

Doğrusal ayrılmayan veriler için destek vektör makineleri matematiksel olarak  $K(x_i, x_j) = \varphi(x) * \varphi(x_j)$  şeklinde ifade edilen bir kernel fonksiyonu yardımıyla doğrusal olmayan dönüşümler yapılabilmekte ve bu şekilde verilerin yüksek boyutta doğrusal olarak ayırılmasına imkân sağlamaktadır. Literatürde kernel fonksiyonu olarak en sık kullanılan polinom, radyal tabanlı fonksiyon, Pearson VII (PUK) fonksiyonu ve normalleştirilmiş polinom kernelleri Tablo 3.3’de parametreleri ve formülleriyle birlikte sunulmaktadır (Kavzoğlu ve Çölkesen, 2010, s. 78).

**Tablo 3.3:** Temel Kernel fonksiyonları ve parametreleri

| Kernel Fonksiyonu                 | Matematiksel İfadesi                                                                                                       | Parametre                                            |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|
| Polinom Kerneli                   | $K(x, y) = ((x \cdot y) + 1)^d$                                                                                            | Polinom derecesi (d)                                 |
| Normalleştirilmiş Polinom Kerneli | $K(x, y) = \frac{((x \cdot y) + 1)^d}{\sqrt{((x \cdot x) + 1)^d ((y \cdot y) + 1)^d}}$                                     | Polinom derecesi (d)                                 |
| Radyal Tabanlı Fonksiyon Kerneli  | $K(x, y) = e^{-\gamma \ x - x_i\ ^2}$                                                                                      | Kernel boyutu ( $\gamma$ )                           |
| Pearson VII (PUK) Kerneli         | $\frac{1}{\left[ 1 + \left( \frac{2 \cdot \sqrt{\ x - y\ ^2} \sqrt{2^{(1/\omega)} - 1}}{\sigma} \right)^2 \right]^\omega}$ | Pearson genişliği parametreleri ( $\sigma, \omega$ ) |

### 3.4.5. Değişken Seçimi

#### 3.4.5.1. Bağımlı değişken

Araştırmamızda nihai hedef, finansal performans tahmini olduğundan bağımlı değişken olarak finansal performans göstergeleri kullanılacaktır. Ancak muhasebe ve finans alanında bir işletmenin finansal performansını ölçmek için aktif kârlılığı, hisse başı kazanç, ekonomik katma değer (EVA) gibi birçok farklı gösterge bulunmaktadır. Burada cevaplandırılması gereken önemli bir soru; bu göstergelerden hangilerinin kullanılacağıdır. Cevaplanması gereken diğer bir soru ise bu göstergelerin gerçek değerlerinin kullanılmasının mı daha faydalı olacağı yoksa göstergelerin kategorize edilerek belli sınıflara ayrılmış değerlerinin kullanılmasının mı daha faydalı olabileceğidir. Diğer taraftan sınıflandırma çerçevesinde, işletme performansı için sınıfların nasıl tanımlanacağı sorusuyla da karşı karşıyayız. Bu soruların kesin bir cevabı bulunmamaktadır. Ancak bu noktada en fazla faydayı sağlayacak cevabı bulmak için çeşitli yöntemler geliştirilebilir. Bu çalışmada daha önce yapılan çalışmalar da göz önünde bulundurularak ve çeşitli istatistiki değerlendirmeler yapılarak sonuca ulaşılmaya çalışılmıştır.

Literatürde daha önce yapılan çalışmalar bu sorulara farklı açılardan cevaplar sunmaktadır. Kohut ve Segars (1992) yaptıkları çalışmada kurumsal iletişim stratejisindeki kalıpları incelemek üzere işletme yöneticilerinin hissedarlara sunduğu mektupları incelemişlerdir. Bu çalışmada işletmeler **öz kaynak kârlılığına** dayalı olarak **başarılı işletmeler** ve **daha az başarılı olan işletmeler** olarak iki sınıfa ayrılmıştır. Elde edilen örnek, Fortune dergisinin yıllık en büyük 500 Amerikan şirketi listesinden seçilen **ilk 25 ve en alttaki 25 işletmesinin** mektuplarından oluşuyordu.

Li (2006) risk duyarlılığındaki değişiklikler ile gelecekteki kazançlar arasındaki ilişkiyi incelediği çalışmasında **kârdaki değişimin gerçek değerlerini** dikkate almıştır.

Frazier ve diğerleri (1984) yaptıkları çalışmada **pozitif finansal performans** ve **negatif finansal performans** şeklinde iki sınıf belirlemişler ve performans ölçütü olarak da **kârdaki büyümenin yüzdesel değişim oranını** dikkate almışlardır.

Clatworthy ve Jones (2003) yaptıkları çalışmada finansal performans göstergesi olarak **vergilendirmeden önceki kârdaki yüzde değişim oranını** kullanmışlardır. Çalışma örneği İngiltere’de kayıtlı 200.000’den fazla işletmeyi kapsayan veri tabanı içinden performansı güçlü şekilde iyileşen ilk 50 işletme ile performansı güçlü bir şekilde

kötüleşen en alt 50 işletme olarak seçilmiştir ve bu ayırım göz önünde bulundurularak **“performansı iyileşen”** ve **“performansı kötüleşen”** şeklinde iki sınıf belirlenmiştir.

Wahyuni, Febrianto ve Rahman (2018) yaptıkları çalışmada işletme performansı göstergesi olarak **kârlılık oranlarını (aktif kârlılık oranı, öz kaynak kârlılık oranı ve net kâr marjı)** kullanmışlardır. Araştırmada çoklu doğrusal regresyon analizi yapıldığından **kârlılık oranlarının gerçek değerleri** kullanılmıştır.

Subramanian, Insley ve Blackwell (1993) performans ile yıllık raporların okunabilirliği arasındaki ilişkiyi test etmek için yaptıkları çalışmada finansal performans göstergesi olarak **net kârı** kullanmışlardır. Çalışmada net kâr açıklayan ve kârı bir önceki yıla göre artan işletmeleri **iyi performans** gösteren işletme, net zarar açıklayan ve net zararı bir önceki yıldan daha büyük olan işletmeleri ise **kötü performans** gösteren işletme şeklinde iki sınıfa ayırmışlardır.

Qiu (2007), finansal performans göstergesi olarak **büyükölçe göre düzeltilmiş kümülatif getiri, öz kaynak kârlılığı ve hisse başına kârdaki değişim oranlarını** kullanmıştır. Çalışmada işletmeler, değişim oranlarının dağılımlarına göre yüksek performanslı, ortalama performanslı ve düşük performanslı şeklinde üç sınıfa ayrılmıştır. En yüksek %25'lik işletmeler yüksek performans sınıfı, ortadaki %50 ortalama performans sınıfı ve en düşük %25 ise düşük performans sınıfı olarak tanımlanmıştır.

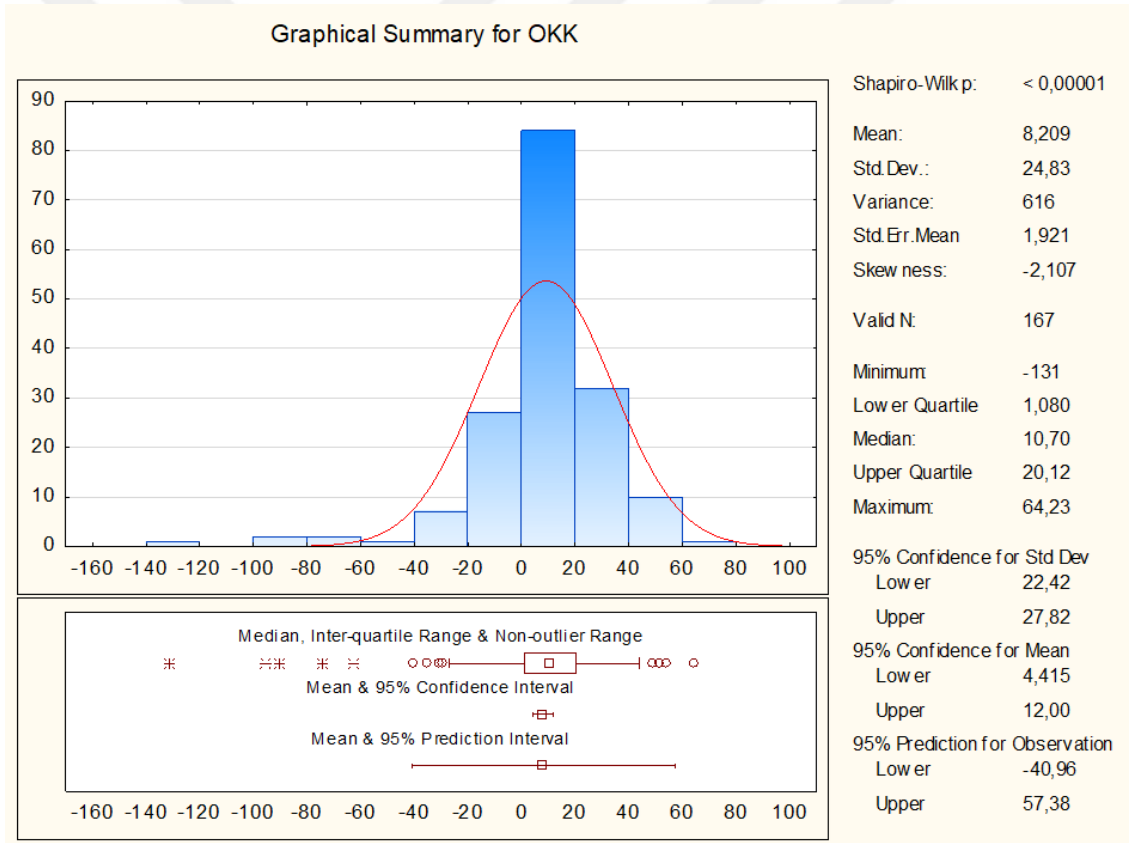
Görüldüğü üzere daha önce yapılan çalışmalarda finansal performans göstergelerinin seçilmesinde ve kullanılacak değerlerin belirlenmesinde belli bir kriter bulunmamaktadır. Ancak göstergelerin genel itibarıyla kâr ve kârlılık etrafında yoğunlaştığı ve kategorik ölçüt değerlerin kullanıldığı görülmektedir.

Kâr, bir işletmenin belirli bir dönemde elde ettiği hasılatın, bu hasılatı elde etmek için katlandığı giderlerin çıkarılmasından sonra kalan olumlu farktır. Kârlılık ise bir işletmenin kazanç yaratma yeteneği olarak ifade edilir. Kârlılık bir orandır ve kâr ile kâra etki eden faktörler arasındaki rakamsal ifadedir. Kârlılık analizinde işletmenin yeterli kâr elde edip edemediği anlaşılmaya çalışılır (Sevim, 2013, s. 134-135). Kâr farklı faktörlerden etkilendiğinden dolayı bir işletmenin kârlılığı birkaç farklı fakat birbiriyle ilişkili şekilde ölçülebilir. Birincisi, işletmenin kârı satışlarıyla ilgilidir. Yani bir liralık satış karşılığında kalan getirinin ne olduğu ölçülür. İkincisi, kâr elde etmek için gerekli yatırımlarla ilgili kârlarla ilgili olarak öz kaynakların getirisi veya toplam varlıkların getirisinin ne olduğu ölçülür. Ayrıca işletme ortaklarına yeterli bir gelir sağlanıp sağlanmadığının saptanmasında işletmenin hisse senetlerinin mali değerinin analizi ve

pay başına düşen kâr paylarının belirlenmesi de önemlidir (Önce , 2011, s. 242). Bu bakımdan bu çalışmada bağımlı değişken olarak, öz kaynak getirisini değerlendirmek için **öz kaynak kârlılığı (ÖKK)**, satışların getirisini değerlendirmek için **net kâr marjı (NKM)** ve işletmenin ortaklarına sağladığı geliri değerlendirmek için **hisse başına kâr (HBK)** değişkenleri finansal performans göstergesi olarak kullanılmıştır. Ayrıca üç farklı oranın da faaliyet raporlarındaki açıklamalar ile ilişkisinin ayrı ayrı değerlendirilmesi, çıkan sonuçları karşılaştırma ve yorumlama bakımından önem arz etmektedir.

Çalışmada, işletmelerin finansal performans göstergelerine ait gerçek değerlerin kullanılması yerine, işletme performanslarının sınıflama sonucunda oluşmuş kategorik değerlerinin kullanılmasına karar verilmiştir. Sınıflandırmada işletmelerin finansal performansı için sınıfların nasıl tanımlanacağı ve işletmelerin ayrılacağı sınıf sayısının kaç olması gerektiği üzerinde durulması gereken önemli bir konudur. Ayrıca işletmelerin finansal performanslarına göre farklı sınıflara ayırmak için uygun kriterlerin neler olduğunun belirlenmesi gerekir. Bu soruların kesin bir cevabı bulunmamakla birlikte muhasebe literatüründe kârlılık oranlarının değerlendirilmesinde oranların geçmiş yıllardaki durumu ve işletmenin bulunduğu sektör ortalaması önemli kriterlerdir. Özellikle de işletmenin bulunduğu sektördeki rakiplerine göre finansal performansı işletmenin başarısını değerlendirmekte önemli bir kriterdir. Bu noktada temel ölçüt, analiz kapsamındaki işletmelerin kârlılık oranlarının ortalamasıdır. Bu bakımdan sınıf sayısının ve sınırlarını belirlemek için tahmin edilecek 2017 yılı baz alınarak kârlılık oranlarının dağılım grafiği ve tanımlayıcı istatistik değerleri incelenmiştir. Şekil 3.5 ÖKK'nın grafiksel özeti, Şekil 3.6 NKM'nın grafiksel özeti ve Şekil 3.7 HBK'nın grafiksel özeti incelendiğinde bütün kârlılık oranlarının genel itibariyle sivri bir dağılım grafiğine sahip olduğu gözlemlenmiştir. Özellikle işletmelerin kârlılık oranlarının sıfırın sağına yakın noktada yoğunlaştığı görülmektedir. Her üç kârlılık oranının ortalama değerinin işletmelerin yoğunlaştığı bu noktada olduğu görülmektedir. Diğer taraftan pozitif finansal performans gösteren, yani sıfırın üzerinde kârlılık oranına sahip işletme sayısının her üç kârlılık oranında da sıfır ve sıfırın altında kârlılık oranına sahip işletme sayısından çok daha fazla olduğu görülmektedir. Dağılım grafikleri incelendiğinde yaklaşık olarak firmaların %50'sinin orta noktada olduğu diğerlerinin ise yaklaşık olarak %25'lik dilimlerde yer aldığı görülmektedir. Bu açıdan işletmelerin %25, %50, %25 olarak üç sınıfa ayrılması sağlıklı karar verebilme açısından daha faydalı olabilecektir. Ayrıca daha önce yapılan bazı çalışmalarda kullanılan **başarılı işletme-başarısız işletme**

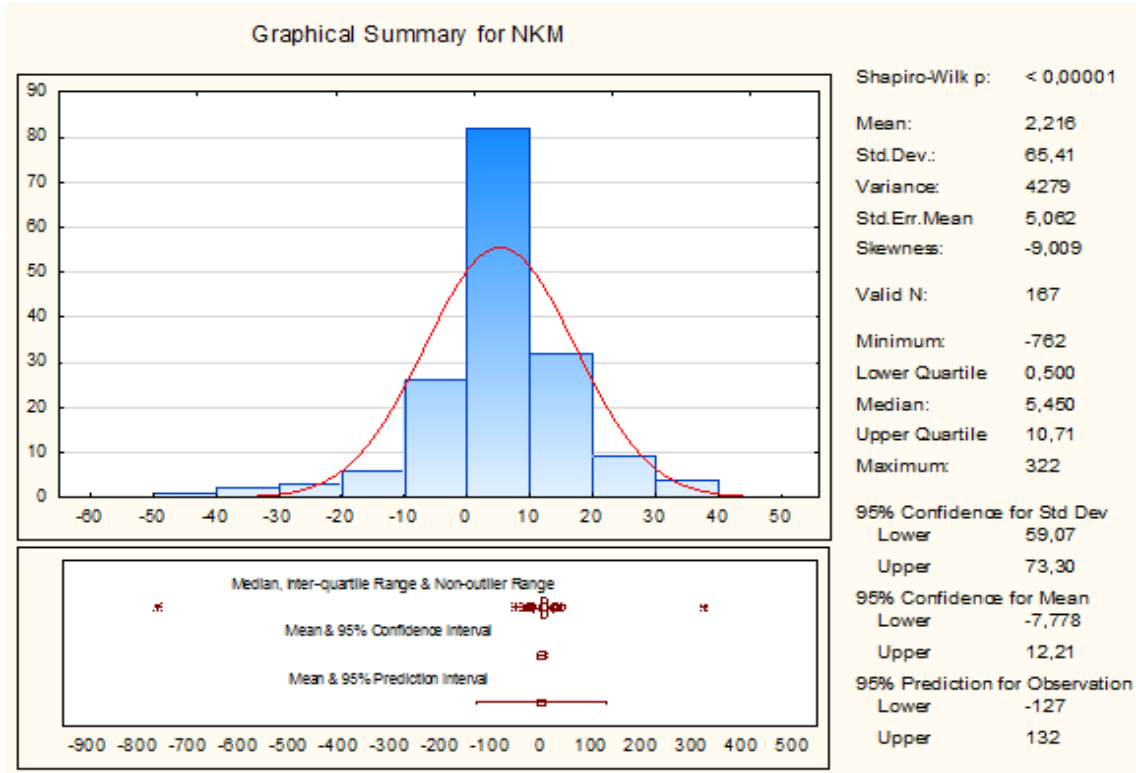
veya **iyi performans-kötü performans** şeklindeki ikili sınıflandırmanın ortalamasının üzerinde başarılı işletmeleri ayırmaya imkân vermediği değerlendirildiğinde üçlü sınıflandırma ikili sınıflandırmaya nazaran çok daha faydalı olabilecektir. Bu açıdan işletmeler finansal performanslarına göre **düşük performans, orta performans** ve **yüksek performans** şeklinde üç sınıfa ayrılmışlardır. Sınıf sınırları ise sektör ortalaması göz önünde bulundurularak belirlenmiştir. Buna göre sektör ortalamasının etrafında yer alan %50’lik kısım orta performanslı olarak belirlenmiştir. Alt çeyrek değerinin altında yer alan %25’lik kısım düşük performanslı ve üst çeyrek değerinin üzerinde yer alan %25’lik kısım ise yüksek performanslı olarak belirlenmiştir. Düşük performans sınıfına dahil işletmeler “0” değeri ile, orta performans sınıfına dahil işletmeler “1” değeri ile ve yüksek performans sınıfına dahil işletmeler “2” değeri ile ifade edilmiştir.



**Şekil 3.5: ÖKK'nın grafiksel özeti**

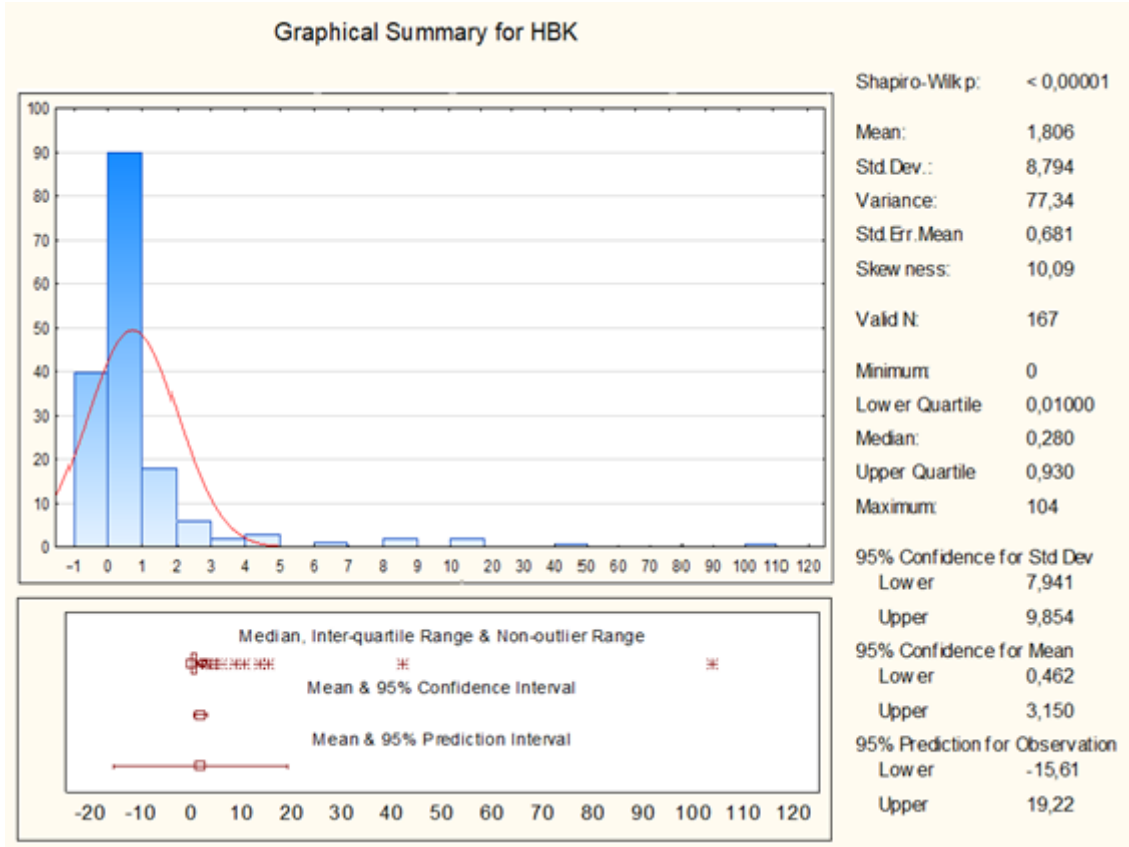
Çalışmada kullanılan ilk bağımlı değişken olan öz kaynak kârlılığı için sınıf sayısı ve sınıf sınırlarını tespit etmek amacıyla baz yıl olarak kabuk edilen 2017 yılı öz kaynak kârlılık oranı tanımlayıcı istatistik değerleri ve dağılım grafiği Şekil 3.5’de gösterildiği şekilde özet olarak oluşturulmuştur. Burada gösterilen değerler öz kaynak kârlılık oranının yüzde değerleridir. Şekil 3.5’deki ÖKK’nın grafiksel özeti incelendiğinde

ÖKK'nın ortalamasının (mean) %8,209 olduğu medyanın ise %10,70 olduğu görülmektedir. İşletmelerin öz kaynak kârlılık oranı dağılım grafiğinde de görüldüğü üzere işletmelerin yaklaşık yüzde ellisi %0 ile %20 öz kaynak kârlılık oranı arasında yer almaktadır. Kalan işletmelerin ise grafiğin sağına ve soluna yaklaşık olarak eşit oranda dağıldığı görülmektedir. Bu durum üç sınıflı önermemizi destekleyen bir dağılım olduğunu göstermektedir. Buna göre işletmeler öz kaynak kârlılığına göre düşük performans, orta performans ve yüksek performans olarak üç sınıfa ayrılmıştır. Sınıf sınırları ise öz kaynak kârlılık oranı üst çeyrek ve alt çeyrek değerlerine göre belirlenmiştir. Yüzde yirmibeşlik alt dilimde yer alan işletmeler için alt çeyrek değeri olan %1,08 oranı sınır kabul edilmiş ve bu oranın altında olan işletmeler düşük performanslı işletme sınıfına dâhil edilmiştir. Düşük performans “0” değeri ile ifade edilmiştir. Yüzde ellilik orta dilimde yer alan işletmeler için alt çeyrek değeri olan %1,08 oranı alt sınır, üst çeyrek değeri olan %20,12 oranı ise üst sınır kabul edilmiş ve bu sınırlar arasında kalan işletmeler orta performanslı işletme sınıfına dâhil edilmiştir. Orta performans “1” değeri ile ifade edilmiştir. Yüzde yirmibeşlik üst dilimde yer alan işletmeler için üst çeyrek değeri olan %20,12 oranı sınır kabul edilmiş ve bu oranın üzerinde olan işletmeler yüksek performanslı işletme sınıfına dâhil edilmiştir. Yüksek performans “2” değeri ile ifade edilmiştir.



**Şekil 3.6:** NKM'nın grafiksel özeti

Diğer bir bağımlı değişken olan net kâr marjı için baz yıl olarak kabul edilen 2017 yılı tanımlayıcı istatistik değerleri ve dağılım grafiği Şekil 3.6'da gösterildiği şekilde özet olarak oluşturulmuştur. Burada gösterilen değerler net kâr marjının yüzde değerleridir. Şekil 3.6'daki grafiksel özet incelendiğinde net kar marjının ortalamasının %2,216 olduğu medyanın ise %5,45 olduğu görülmektedir. İşletmelerin net kâr marjının dağılım grafiğinde de görüldüğü üzere işletmelerin yaklaşık yüzde ellisi %0 ile %10 net kâr marjı arasında yer almaktadır. Kalan işletmelerin ise grafiğin sağına ve soluna yaklaşık olarak eşit oranda dağıldığı görülmektedir. Buna göre işletmeler net kâr marjına göre düşük performans, orta performans ve yüksek performans olarak üç sınıfa ayrılmıştır. Sınıf sınırları ise net kâr marjı üst çeyrek ve alt çeyrek değerlerine göre belirlenmiştir. Yüzde yirmibeşlik alt dilimde yer alan işletmeler için alt çeyrek değeri olan %0,50 oranı sınır kabul edilmiş ve bu oranın altında olan işletmeler düşük performanslı işletme sınıfına dâhil edilmiştir. Düşük performans "0" değeri ile ifade edilmiştir. Yüzde ellilik orta dilimde yer alan işletmeler için alt çeyrek değeri olan %0,50 oranı alt sınır, üst çeyrek değeri olan %10,71 oranı ise üst sınır kabul edilmiş ve bu sınırlar arasında kalan işletmeler orta performanslı işletme sınıfına dâhil edilmiştir. Orta performans "1" değeri ile ifade edilmiştir. Yüzde yirmibeşlik üst dilimde yer alan işletmeler için üst çeyrek değeri olan %10,71 oranı sınır kabul edilmiş ve bu oranın üzerinde olan işletmeler yüksek performanslı işletme sınıfına dâhil edilmiştir. Yüksek performans "2" değeri ile ifade edilmiştir.



**Şekil 3.7: HBK'nın grafiksel özeti**

Hisse başına kazanç miktarı, net dönem karının şirket hissesinin dönem içindeki ağırlıklı ortalama pay adedine bölünmesiyle hesaplanır. Bu açıdan Şekil 3.7 HBK'nın grafiksel özetindeki değerler bir hisse başına düşen net karın TL tutarındır. Şekil 3.7'daki HBK'nın grafiksel özeti incelendiğinde hisse başına karın ortalamasının 1,80 TL olduğu medyanın ise 0,28 TL olduğu görülmektedir. Diğer taraftan hisse başına kazanç değerlerinin diğer oranlara göre biraz farklı dağıldığı görülmektedir. Hisse başına kar için en düşük değer sıfırdır yani hisse başına kazancın diğer oranlar gibi negatif değeri bulunmamaktadır. İşletmelerin ortalama değer etrafında değil daha alta ve sıfıra yakın noktada toplandığı görülmektedir. Ortalama değer üst noktada oluşmasının nedeni hisse başına kazancın negatif değer almaması ve uç değerlerin ortalamayı yükseltmesi olabilir. Ancak yine de hisse başına kazancında üçlü sınıflandırmaya uygun olduğu söylenebilir. Zira dağılım grafiği incelendiğinde işletmelerin yaklaşık %25'inin sıfır değerini aldığı görülmektedir. Yaklaşık %50'sinin ise orta noktada kümelendiği görülmektedir. Geri kalanlar ise yaklaşık %25'lik üst bölme dağılmış olduğu görülmektedir. Bu bakımdan işletmeler hisse başına kara göre düşük performans, orta performans ve yüksek performans olarak üç sınıfa ayrılmıştır. Sınıf sınırları ise diğer oranlarda olduğu gibi üst

çeyrek ve alt çeyrek değerlerine göre belirlenmiştir. Yüzde yirmibeşlik alt dilimde yer alan işletmeler için alt çeyrek değeri olan 0,01 sınır kabul edilmiş ve bu değerin altında olan işletmeler düşük performanslı işletme sınıfına dâhil edilmiştir. Düşük performans “0” değeri ile ifade edilmiştir. Yüzde ellilik orta dilimde yer alan işletmeler için alt çeyrek değeri olan 0,01 alt sınır, üst çeyrek değeri olan 0,93 ise üst sınır kabul edilmiş ve bu sınırlar arasında kalan işletmeler orta performanslı işletme sınıfına dâhil edilmiştir. Orta performans “1” değeri ile ifade edilmiştir. Yüzde yirmibeşlik üst dilimde yer alan işletmeler için üst çeyrek değeri olan 0,93 oranı sınır kabul edilmiş ve bu oranın üzerinde olan işletmeler yüksek performanslı işletme sınıfına dâhil edilmiştir. Yüksek performans “2” değeri ile ifade edilmiştir.

#### **3.4.5.2. Bağımsız değişkenler**

Bağımsız değişkenler, metin madenciliği süreci sonucunda elde edilen kelime haznesine (Bag of Words) göre oluşturulan ve bir veya birden fazla kelimedenden oluşan belge terim vektörlerinin, boyut küçültme işlemine tabi tutulması sonucunda elde edilen 22 adet TDA’ya dayalı belge skorları olarak belirlenmiştir.

Faaliyet raporlarından elde edilen veri Türkçe metinlerden oluşmaktadır. Metin işlemede Statistica metin madenciliği aracı kullanılmıştır. Statistica İngilizce ve diğer bazı diller için, kelimeleri köklerine indirgeyen (stemming ve lemmatization) ve cümleleri dizgeciklere ayıran (tokenization) algoritmaları da sağlamaktadır. Ancak Statistica Türkçe için kelimeleri köklerine indirgeyen hazır bir gövdeleme algoritması sunmamaktadır. Bunun yerine dil seçeneğini devre dışı bırakarak çalışma imkânı sunmaktadır. Çalışmada Statistica programının seçilmesinin nedeni, hazır gövdeleme algoritması olmaksızın çalışma imkânının bulunması nedeniyle Türkçe metinler için ideal bir araç olmasıdır. Zira Türkçe için oluşturulmuş gövdeleme algoritmaları bulunmakla birlikte bunların henüz istenen seviyede doğru gövdeleme işlemi yaptığı söylenemez (Sönmez, 2017, s. 76; Yücebaş ve Tintin, 2017). Ayrıca faaliyet raporlarında alana özgü ifadelerin varlığı da gövdeleme algoritması kullanmayı zorlaştıran başka bir nedendir. Diğer taraftan bu algoritmaların metin madenciliği araçları ile entegrasyonu da diğer bir sorun olarak karşımıza çıkmaktadır. Tüm bu nedenlerden dolayı daha öncede ifade edildiği üzere çalışmada gövdeleme yöntemi olarak tablo arama yöntemi kullanılmıştır. Tablo arama yönteminde ilk adım, çalışılacak metin göz önünde bulundurularak bir sözlük oluşturulmasıdır. Sözlük oluşturulduktan sonra eş anlamlı kelimeler tespit edilir.

Ayrıca kelimelerin kökleri de tespit edilerek kelimeleri ve kelime köklerini içeren bir arama tablosu oluşturulur. Çalışmada oluşturulacak sözlüğe dayanak olması için ilk olarak Statistica programında dil seçeneği devre dışı bırakıldıktan sonra faaliyet raporlarındaki metinler dizgeciklere ayrılarak ham bir kelime haznesi elde edilmiştir. Daha sonra Türkçe dil bilgisi ve muhasebe ve finans terimleri göz önünde bulundurmak suretiyle bu kelime haznesi incelenerek, gereksiz kelimeler, eş anlamlı kelimeler ve kelime kökleri manuel olarak belirlenmiştir. Son aşamada ise belirlen bu kelimelerden bir sözlük oluşturularak arama tablosu haline getirilmiştir. Oluşturulan bu arama tablosu dil analizinden önce programa yüklenerek çalışmada kullanılacak kelime haznesi elde edilmiştir. Kelime haznesinden ise ikili (binary) gösterim seçeneği kullanılmak suretiyle 2890 adet kelime ve kelime öbeğinden oluşan ve çalışmada ulaşılan 1249 adet faaliyet raporunu temsil eden belge terim vektörleri oluşturulmuştur. Belge terim vektörlerinin birleşiminden ise 2890 sütun ve 1249 satırdan oluşan belge terim matrisi elde edilmiştir. Daha sonra bu matris TDA'ya dayalı olarak yeniden düzenlemiş ve 22 adet değişken için TDA'ya dayalı belge skorlarına ulaşılmıştır. Bu şekilde belge terim matrisinin boyutu küçültülerek 22 sütun ile 1249 satırdan oluşan ve belge terim matrisinin temsili olan yeni bir matris elde edilmiştir.

#### **3.4.6. Verilerin Birleştirilmesi ve Eğitim ve Test Verisinin Oluşturulması**

Çalışma örneklemini oluşturan 1249 adet faaliyet raporuna yukarıda ifade edilen metin madenciliği adımları uygulanmak suretiyle metin veriler yapısal hale dönüştürülmüş ve TDA'ya dayalı belge skorları elde edilmiştir. TDA'ya dayalı belge skorları ile finansal performans göstergelerine ilişkin kârlılık oranlarından elde edilen kategorik finansal performans gösterge değerleri Şekil 3.8'de gösterildiği üzere bir excel dosyasında birleştirilmiş ve veriler eğitim ve test verisi olarak ayrılarak Statistica programına yüklenmiştir. Çalışmada işletmelerin 2017 yılına ait finansal performans değerleri tahmin edileceğinden işletmelerin 2017 yılı verileri test verisi olarak belirlenmiştir. İşletmelerin 2010 ile 2016 yılları arasındaki döneme ait verileri ise eğitim verisi olarak belirlenmiştir. Buna göre 2017 yılına ait 167 gözleme ilişkin veriler test verisi, 2010 ile 2016 yılları arasındaki döneme ait 1082 gözleme ilişkin veriler ise eğitim verisi olarak seçilmiştir.

| D_NO  | RAP_YIL | İSL_KOD | TESTING | NKM | NKM_O | HBK | HBK_O | OKK | OKK_O | Concept1 | Concept2 | Concept3 | Concept4 | Concept5 | Concept6 | Concept7 |
|-------|---------|---------|---------|-----|-------|-----|-------|-----|-------|----------|----------|----------|----------|----------|----------|----------|
| D0001 | 2013    | ACSEL   | TRAIN   | 1   | 6,06  | 2   | 0,95  | 1   | 15,57 | 0,02392  | -0,01015 | -0,01141 | -0,01054 | -0,00517 | 0,00863  | 0,02230  |
| D0002 | 2014    | ACSEL   | TRAIN   | 0   | 0,40  | 1   | 0,08  | 0   | 0,93  | 0,02325  | -0,01379 | -0,01467 | -0,00916 | 0,00093  | 0,00893  | 0,01857  |
| D0003 | 2015    | ACSEL   | TRAIN   | 1   | 2,84  | 1   | 0,03  | 1   | 3,48  | 0,02526  | -0,01283 | -0,01595 | -0,00728 | -0,00054 | 0,00895  | 0,02356  |
| D0004 | 2016    | ACSEL   | TRAIN   | 1   | 4,76  | 1   | 0,06  | 1   | 3,72  | 0,02495  | -0,01511 | -0,02186 | -0,01170 | 0,00345  | 0,00468  | 0,01794  |
| D0005 | 2017    | ACSEL   | TEST    | 1   | 8,69  | 1   | 0,17  | 1   | 8,14  | 0,02560  | -0,01328 | -0,01693 | -0,01084 | 0,00034  | 0,00432  | 0,02441  |
| D0006 | 2010    | ADANA   | TRAIN   | 2   | 33,34 | 1   | 0,63  | 2   | 34,26 | 0,02698  | -0,03579 | -0,00618 | 0,00930  | 0,09284  | -0,00964 | -0,07232 |
| D0007 | 2011    | ADANA   | TRAIN   | 2   | 24,46 | 1   | 0,47  | 2   | 25,34 | 0,02898  | -0,03830 | -0,01678 | 0,00981  | 0,09247  | -0,00809 | -0,07846 |
| D0008 | 2012    | ADANA   | TRAIN   | 2   | 30,52 | 1   | 0,54  | 2   | 29,38 | 0,03302  | -0,01956 | -0,04998 | 0,01371  | 0,09931  | -0,00331 | -0,05345 |
| D0009 | 2013    | ADANA   | TRAIN   | 2   | 17,09 | 1   | 0,41  | 2   | 21,88 | 0,03788  | -0,00618 | -0,08863 | 0,02035  | 0,12630  | -0,00341 | -0,08038 |
| D0010 | 2014    | ADANA   | TRAIN   | 2   | 32,51 | 1   | 0,88  | 2   | 41,56 | 0,03818  | -0,00035 | -0,09195 | 0,02160  | 0,13205  | -0,00472 | -0,08344 |
| D0011 | 2015    | ADANA   | TRAIN   | 2   | 33,65 | 1   | 0,88  | 2   | 38,15 | 0,03618  | -0,00034 | -0,08896 | 0,02123  | 0,11848  | -0,00365 | -0,10129 |
| D0012 | 2016    | ADANA   | TRAIN   | 2   | 34,10 | 1   | 0,80  | 2   | 34,27 | 0,03693  | -0,00289 | -0,08737 | 0,01943  | 0,11778  | -0,00298 | -0,10225 |
| D0013 | 2017    | ADANA   | TEST    | 2   | 26,74 | 1   | 0,77  | 2   | 33,25 | 0,03703  | 0,01023  | -0,09061 | 0,02702  | 0,12093  | -0,00666 | -0,08734 |
| D0014 | 2011    | ADEL    | TRAIN   | 2   | 19,07 | 2   | 3,30  | 2   | 31,13 | 0,00979  | -0,00704 | 0,01011  | -0,00292 | 0,00024  | -0,00340 | -0,00365 |
| D0015 | 2012    | ADEL    | TRAIN   | 2   | 15,30 | 2   | 3,10  | 2   | 24,91 | 0,02664  | -0,00744 | 0,00227  | 0,01223  | -0,00568 | 0,01401  | 0,02557  |
| D0016 | 2013    | ADEL    | TRAIN   | 2   | 26,69 | 2   | 6,31  | 2   | 40,16 | 0,02894  | -0,00171 | 0,00920  | 0,00856  | 0,00404  | 0,01922  | 0,04167  |
| D0017 | 2014    | ADEL    | TRAIN   | 2   | 12,69 | 2   | 3,57  | 1   | 19,30 | 0,03003  | 0,00531  | -0,01530 | 0,00934  | 0,00219  | 0,00950  | 0,04698  |
| D0018 | 2015    | ADEL    | TRAIN   | 2   | 28,42 | 2   | 9,08  | 2   | 40,39 | 0,02992  | 0,00045  | -0,01720 | 0,00586  | 0,00520  | 0,01024  | 0,04936  |
| D0019 | 2016    | ADEL    | TRAIN   | 1   | 6,93  | 1   | 0,80  | 1   | 9,40  | 0,02990  | 0,00285  | -0,01723 | 0,00643  | 0,00315  | 0,01315  | 0,04389  |
| D0020 | 2017    | ADEL    | TEST    | 1   | 8,33  | 2   | 1,14  | 1   | 13,41 | 0,02904  | 0,00803  | -0,01735 | 0,00865  | 0,00230  | 0,01514  | 0,04437  |
| D0021 | 2010    | AEFES   | TRAIN   | 2   | 12,08 | 2   | 1,12  | 1   | 19,39 | 0,03347  | 0,00616  | 0,06085  | 0,02421  | 0,00606  | -0,01984 | -0,02994 |
| D0022 | 2011    | AEFES   | TRAIN   | 1   | 7,17  | 1   | 0,76  | 1   | 11,54 | 0,03349  | 0,00669  | 0,05745  | 0,03528  | -0,00512 | -0,01885 | -0,02877 |
| D0023 | 2012    | AEFES   | TRAIN   | 2   | 14,12 | 2   | 1,03  | 1   | 12,39 | 0,04102  | 0,02619  | 0,02839  | 0,03290  | -0,00650 | 0,00056  | -0,01093 |
| D0024 | 2013    | AEFES   | TRAIN   | 2   | 28,37 | 2   | 4,41  | 2   | 32,06 | 0,04502  | 0,03825  | 0,02263  | 0,04037  | -0,00992 | 0,00998  | -0,00172 |
| D0025 | 2014    | AEFES   | TRAIN   | 0   | -5,11 | 0   | 0,00  | 0   | -5,96 | 0,04171  | 0,04766  | 0,01618  | 0,03880  | 0,00181  | -0,01257 | -0,00444 |
| D0026 | 2015    | AEFES   | TRAIN   | 0   | -1,94 | 0   | 0,00  | 0   | -2,58 | 0,04096  | 0,04837  | 0,01437  | 0,04084  | -0,00115 | -0,01312 | -0,00473 |
| D0027 | 2016    | AEFES   | TRAIN   | 0   | -0,68 | 0   | 0,00  | 0   | -0,83 | 0,04087  | 0,04752  | 0,01997  | 0,04349  | 0,00282  | -0,01536 | 0,00634  |
| D0028 | 2017    | AEFES   | TEST    | 1   | 1,15  | 1   | 0,25  | 1   | 1,55  | 0,04148  | 0,04833  | 0,01858  | 0,04244  | 0,00056  | -0,01623 | 0,00453  |

Şekil 3.8 : Analize hazır hale getirilmiş verilerin görünümü

### 3.4.7. Model Performansının ve Sonuçların Değerlendirilmesi

Veri madenciliğinde kullanılan sınıflandırma modellilerinin başarımının değerlendirilmesinde kullanılan temel yöntem, modele ait doğruluk oranıdır. Bunun yanında kesinlik, duyarlılık, f-ölçütü ve hata oranı da sonuçların değerlendirmesi için kullanılmaktadır. Ayrıca bu ölçütlere dayanak olan karışıklık matrisi (confusion matrix) önemli bir değerlendirme aracıdır.

Hata matrisi olarak da bilinen karışıklık matrisi, makine öğrenimi alanındaki sınıflandırma problemlerinde, algoritmanın performansının görselleştirilmesini sağlayan bir tablo düzenidir. Karışıklık matrisi Tablo 3.4’de gösterildiği şekilde düzenlenir.

Tablo 3.4: Karışıklık matrisi

|        |         | Tahmin              |                     |
|--------|---------|---------------------|---------------------|
|        |         | Pozitif             | Negatif             |
| Gerçek | Pozitif | Doğru Pozitif(DP)   | Yanlış Negatif (YN) |
|        | Negatif | Yanlış Pozitif (YP) | Doğru Negatif (DN)  |

Burada;

- DP(Doğru Pozitif); gerçekte pozitif iken, model tarafından pozitif tahmin edilen örnek sayısıdır
- YP(Yanlış Pozitif); gerçekte pozitif iken, model tarafından negatif tahmin edilen örnek sayısıdır

- DN(Doğru Negatif); gerçekte negatif iken, model tarafından negatif tahmin edilen örnek sayısıdır
- YN(Yanlış Negatif); gerçekte negatif iken, model tarafından pozitif tahmin edilen örnek sayısıdır

Karışıklık matrisinde yer alan değerler kullanılarak doğruluk (accuracy) – hata oranı, duyarlılık, kesinlik, ve f-ölçütü hesaplanır.

Doğruluk oranı modeldeki doğru sınıflandırılmış örnek sayısının, toplam örnek sayısına bölünmesi ile bulunur.

$$\text{Doğruluk Oranı} = \frac{DP+DN}{DP+DN+YP+YN} \quad (3.5)$$

$$\text{Hata Oranı} = 1 - \text{Doğruluk Oranı} \quad (3.6)$$

Duyarlılık (sensivity), modelin girdilerden pozitif sınıfa ait örnekleri tahmin etmekteki başarısını gösterir. Doğru sınıflandırılan pozitif örneklerin toplam pozitif örnek sayısına bölünmesi ile bulunur.

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3.7)$$

Kesinlik (precision), doğru sınıflandırılan pozitif örneklerin toplam pozitif tahmin edilen örneklerle bölünmesi ile bulunur.

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (3.8)$$

F-Ölçütü, kesinlik ve duyarlık ölçütlerinin harmonik ortalamasıdır. Bu ölçüt her iki ölçütün birlikte değerlendirmesini sağlar.

$$F - \text{Ölçütü} = \frac{2 \times \text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.9)$$

## 4. BULGULAR VE YORUM

### 4.1. Giriş

Çalışmanın bu bölümünde gerçekleştirilen analizlerin sonuçları ve sonuçların değerlendirilmesi üç başlık halinde verilmiştir. Bu başlıkların birincisi ÖKK, ikincisi NKM, üçüncüsü ise HBK'nın tahmin edilmesine ilişkin analiz sonuçları ve bunların değerlendirmesini kapsamaktadır.

Çalışmada tahmin modellerinin başarısı metin sınıflandırma araştırmalarındaki standart değerlendirme ölçütleri kullanılarak değerlendirilmiştir. Daha önce ifade edildiği üzere literatürde, karışıklık matrisi, doğruluk oranı, hata oranı, f-ölçütü ve kesinlik gibi değerlendirme ölçütleri standart ölçütler olarak kullanılmaktadır. Bu çalışmada da daha önceki çalışmalarda en sık kullanılan karışıklık matrisi, duyarlılık ve doğruluk oranı ölçütleri kullanılmıştır.

Tahmin modelinin başarı durumu sınıf kategorilerinde birim sayısının eşit olması durumunda “**1/Toplam kategori sayısı**” şeklinde hesaplanan temel doğruluk oranına göre değerlendirilir. Buna göre çalışmada 3 sınıf olduğundan temel doğruluk oranı 1/3 (yaklaşık %33) olarak hesaplanabilir. Ancak bu çalışmada kategorilerdeki birim sayısı eşit olmadığından temel doğruluk oranı için birim sayısı en çok olan kategori baz alınmıştır. Buna göre en çok birim sayısına sahip olan orta performans sınıfına ait %50 oranı temel doğruluk oranı olarak kabul edilmiştir. Qiu (2007)'nin yaptığı benzer çalışmada da çoğunluğa sahip olan kategoriye ait %50 oranı dayanak noktası olarak kabul edilmiştir (Qiu, 2007, s.14-15). Bu nedenle, tahmin modelinin başarısı duyarlılık ve doğruluğun %50'den yüksek olup olmamasına göre değerlendirilmiştir.

#### 4.2. Öz kaynak Kârlılığının Tahminine İlişkin Destek Vektör Makineleri Analiz Sonuçları

Öz kaynak kârlılığını tahmin etmek için yapılan DVM analizinde Tablo 4.1’de gösterilen sonuçlar elde edilmiştir.

**Tablo 4.1: ÖKK karışıklık matrisi**

| ÖKK          |                   | Tahmin           |                 |                   | Satır Toplam |
|--------------|-------------------|------------------|-----------------|-------------------|--------------|
|              |                   | Düşük Performans | Orta Performans | Yüksek Performans |              |
| Gerçek       | Düşük Performans  | 16               | 24              | 1                 | 41           |
|              | Orta Performans   | 18               | 57              | 9                 | 84           |
|              | Yüksek Performans | 5                | 24              | 13                | 42           |
| Sütun Toplam |                   | 39               | 105             | 23                | 167          |

Öz kaynak kârlılığının tahmin edilmesi için yapılan DVM analizi sonucunda oluşturulan model 2017 yılı için 167 işletmeyi finansal performanslarına göre %51,50 doğrulukla sınıflandırabilmiştir. Buna göre model 167 işletmeden 86 tanesini doğru tahmin etmiştir. Hedeflenen oran %100 olduğu göz önünde bulundurulduğunda %51,50 oranı başarı olarak istenen seviyede değildir. Ancak model başarısının değerlendirilmesinde taban değer olarak ifade edilen %50’den daha yüksek bir oranın elde edilmesi modelin finansal performanslarına göre işletme sınıflarını tahmin etmek için kullanılabileceğini göstermektedir. Diğer taraftan Tablo 4.1 ÖKK karışıklık matrisi ve Tablo 4.2 ÖKK duyarlılık oranları incelendiğinde %39,02 ve %30,95 duyarlılık oranları ile düşük performanslı işletme ve yüksek performanslı işletme sınıfında yer alan işletmeleri tahmin etmede modelin başarılı sonuçlar vermediği görülmektedir. Ancak %67,86 duyarlılık oranı ile orta performanslı işletmelerin tahmin edilmesinde başarı sağlandığı görülmektedir.

**Tablo 4.2: ÖKK duyarlılık oranları**

| ÖKK Sınıfı        | Duyarlılık |
|-------------------|------------|
| Düşük Performans  | 39,02      |
| Orta Performans   | 67,86      |
| Yüksek Performans | 30,95      |

#### 4.3. Net Kâr Marjının Tahminine İlişkin Destek Vektör Makineleri Analiz Sonuçları

Net kâr marjını tahmin etmek için yapılan DVM analizinde Tablo 4.3’de gösterilen sonuçlar elde edilmiştir.

**Tablo 4.3:** NKM karışıklık matrisi

| NKM          |                   | Tahmin           |                 |                   | Satır Toplam |
|--------------|-------------------|------------------|-----------------|-------------------|--------------|
|              |                   | Düşük Performans | Orta Performans | Yüksek Performans |              |
| Gerçek       | Düşük Performans  | 20               | 16              | 5                 | 41           |
|              | Orta Performans   | 24               | 48              | 12                | 84           |
|              | Yüksek Performans | 2                | 15              | 25                | 42           |
| Sütun Toplam |                   | 46               | 79              | 42                | 167          |

Net kâr marjının tahmin edilmesi için yapılan DVM analizi sonucunda oluşturulan model 2017 yılı için 167 işletmeyi finansal performanslarına göre %55,69 doğrulukla sınıflandırabilmiştir. Buna göre model 167 işletmeden 93 tanesini doğru tahmin etmiştir. Oran yüksek olmamakla birlikte öz kaynak kârlılığının tahmini için oluşturulmuş modele göre daha başarılıdır. Diğer yandan %50’lik taban değerinden yüksek olması modelin finansal performansına göre işletmelerin dahil oldukları sınıfları tahmin etmede kullanılabileceğini göstermektedir. Tablo 4.3 NKM karışıklık matrisi ve Tablo 4.4 NKM duyarlılık oranları incelendiğinde, yüksek performanslı işletme ve orta performanslı işletme sınıflarının duyarlılık oranlarının %59,52 ve %57,14 olarak gerçekleştiği ve dikkate değer şekilde %50 değerinin üzerine çıktığı görülmektedir. Bu durum faaliyet raporlarındaki metin veriler ile finansal performans arasında bir ilişki bulunduğuna yönelik ciddi bir kanıt sunmaktadır. Her ne kadar düşük performanslı işletmelerin tahmin edilmesinde başarı oranı %48,78 ile %50’nin altında olsa da bu oranında eşik değere yakın çıkması finansal performansın tahmin edilmesinde metin verilerin kullanılabilirliği konusunda umut vermektedir. Ayrıca karışıklık matrisi detaylı şekilde incelendiğinde hem yüksek performanslı işletmelerin düşük performanslı, hem de düşük performanslı işletmelerinde yüksek performanslı işletme olarak hatalı tahmin edilme oranının çok düşük olması, düşük performanslı işletmeler ile yüksek performanslı işletmelerin faaliyet raporlarındaki açıklamaların birbirinden farklılaştığının bir göstergesi olabilir. Zira

yüksek performanslı 42 işletmeden sadece 2 tanesi düşük performanslı işletme olarak, düşük performanslı 41 işletmeden ise sadece 5 tanesi yüksek performanslı işletme olarak hatalı tahmin edilmiştir.

**Tablo 4.4:** NKM duyarlılık oranları

| NKM Sınıfı        | Duyarlılık |
|-------------------|------------|
| Düşük Performans  | 48,78      |
| Orta Performans   | 57,14      |
| Yüksek Performans | 59,52      |

#### 4.4. Hisse Başına Kârın Tahminine ilişkin Destek Vektör Makineleri Analiz

##### Sonuçları

Hisse başına kâr tahmin etmek için yapılan DVM analizinde Tablo 4.5’de gösterilen sonuçlar elde edilmiştir.

**Tablo 4.5:** HBK karışıklık matrisi

| HBK          |                   | Tahmin           |                 |                   | Satır Toplam |
|--------------|-------------------|------------------|-----------------|-------------------|--------------|
|              |                   | Düşük Performans | Orta Performans | Yüksek Performans |              |
| Gerçek       | Düşük Performans  | 21               | 18              | 1                 | 40           |
|              | Orta Performans   | 20               | 59              | 6                 | 85           |
|              | Yüksek Performans | 6                | 18              | 18                | 42           |
| Sütun Toplam |                   | 47               | 95              | 25                | 167          |

Hisse başına kâr tahmin edilmesi için yapılan DVM analizi sonucunda oluşturulan model 2017 yılı için 167 işletmeyi finansal performanslarına göre %58,68 doğrulukla sınıflandırabilmiştir. Buna göre model 167 işletmeden 98 tanesini doğru tahmin etmiştir. HBK için kurulan modelin doğruluk oranı hem net kâr marjının hem de öz kaynak kârlılığının tahmini için oluşturulmuş modele göre daha başarılıdır. Ancak sınıflandırma başarısının yeterince yüksek olduğu söylenemez. Diğer yandan doğruluk oranının %50’lik taban değerinden yüksek olması bu modelinde finansal performansına göre işletmelerin dâhil oldukları sınıfları tahmin etmede kullanılabileceğini göstermektedir. Tablo 4.5 HBK karışıklık matrisi ve Tablo 4.6 HBK duyarlılık oranları incelendiğinde, düşük performanslı işletme ve orta performansı işletme sınıflarının duyarlılık oranlarının %52,50 ve %69,41 olarak gerçekleştiği görülmektedir. Burada

özellikle orta sınıftaki işletmelerin tahmin edilmesinde daha başarılı olduğu görülmektedir. Hisse başına kârın tahmin edilmesi için yapılan DVM analizi de diğer iki analizde olduğu gibi faaliyet raporlarındaki metin veriler ile finansal performans arasında bir ilişki bulunduğuna yönelik kanıt sunmaktadır.

HBK karışıklık matrisi detaylı şekilde incelendiğinde NKM'daki duruma benzer olarak, hem yüksek performanslı işletmelerin düşük performanslı, hem de düşük performanslı işletmelerinde yüksek performanslı işletme olarak hatalı tahmin edilme oranının çok düşük olduğu görülmektedir. Bu durum düşük performanslı işletmeler ile yüksek performanslı işletmelerin faaliyet raporlarındaki açıklamalarının birbirinden farklılaştığına yönelik kanıyı güçlendirmektedir. Zira yüksek performanslı 42 işletmeden sadece 6 tanesi düşük performanslı işletme olarak, düşük performanslı 40 işletmeden ise sadece 1 tanesi yüksek performanslı olarak hatalı tahmin edilmiştir.

**Tablo 4.6:** HBK duyarlılık oranları

| HBK Sınıfı        | Duyarlılık |
|-------------------|------------|
| Düşük Performans  | 52,50      |
| Orta Performans   | 69,41      |
| Yüksek Performans | 42,86      |

## 5. SONUÇ

Muhasebenin temel amacı, işletme hakkında karar verecek olan taraflara işletme hakkında faydalı finansal bilgi sağlamaktır. Muhasebenin kayıt tutma işlevi sonucunda üretilen finansal bilgilerin bilgi kullanıcılarına, özellikle işletme dışındaki taraflara iletilmesi ise finansal raporlar aracılığıyla gerçekleşmektedir. Geçmişte finansal raporlama sayısal veriler üzerine odaklanmaktaydı. Ancak zaman içinde gerek teknolojik gelişme, gerekse işletme paydaşlıklarının bilgi ihtiyaçlarında meydana gelen farklılaşma ile birlikte sayısal veriler yanında resim, metin vb. verileri de içeren finansal raporların yayınlanması zorunlu hale gelmiştir. Bu bakımdan finansal raporlama finansal tablolar ile başlayan ve entegre raporlamaya doğru evrilen bir gelişim süreci içerisinde. Bugün finansal tablolar yanında finansal tablo dipnotları, sürdürülebilirlik raporları, faaliyet raporları ve entegre raporlar işletme ile ilgili önemli bilgiler içeren raporlardır. Bu raporlarda sayısal veriler kadar metin veriler de önemli yer tutmaktadır. Ancak tüm bu gelişmelere paralel olarak sayısal veri analizinde önemli gelişmeler yaşanırken, metin verilerin analizinde teknolojik gelişmeden yeterince yararlanılmadığı görülmektedir. Metin verilerin analizi için halen büyük oranda geleneksel yöntemler kullanılmakta ve büyük miktarda metin veri atıl vaziyette durmaktadır. Bu bakımdan finansal raporlardaki metin verilerin analizi önemli bir sorun olarak karşımızda durmaktadır. Başta İngilizce olmak üzere farklı dillerde metin şeklindeki muhasebe verilerinin analizi için gelişen teknoloji olanaklarının kullanıldığı çeşitli çalışmalar yapılmaktadır. Bu çalışmalarda özellikle veri ve metin madenciliği önemli bir yer tutmaktadır. Ancak Türkçe olarak sunulan muhasebe raporlarının analizi için bu araçların kullanıldığı çalışmalara rastlanılmamıştır. Bu noktadan hareketle Türkçe yazılmış finansal raporlardaki metin verilerin analiz sorununun çözümünde veri ve metin madenciliği araçlarının kullanılabilirliğini araştırmak amacıyla bu çalışma yapılmıştır.

Çalışmada finansal performansa ilişkin analizlerde kullanılan sayısal verilerin, faaliyet raporlarında yer alan metin verilerden tahmin edilip edilemeyeceği üzerine odaklanılmıştır. Bu şekilde hem faaliyet raporlarında işletmenin gelecekteki finansal performansı hakkında veri bulunup bulunmadığı tespit edilmiş, hem de sayısal veriler ile metin verilerin birlikte nasıl değerlendirilebileceğine yönelik bir yöntem tanıtılmıştır. Çalışmada faaliyet raporunun seçilmesin nedeni faaliyet raporunun, gerek kapsadığı hesap dönemine ilişkin bilgileri, gerekse işletmenin önümüzdeki dönemdeki hedeflerini ve bu hedefleri nasıl yerine getireceğini anlatan bir belge olması açısından önem arz

etmesidir. Ayrıca daha önce yapılan çalışmalar, faaliyet raporlarındaki metin verilerin işletme performansını değerlendirmek için önemli bir kaynak olduğunu göstermiştir.

Çalışmada veri seti olarak, BİST imalat sanayi sektöründe yer alan 179 işletmenin 2010-2017 yılları arasındaki 8 yıllık dönemde yayınlan faaliyet raporları ve finansal tabloları kullanılmıştır. Tahmin edilecek yıl olarak ise 2017 yılı seçilmiştir. 2017 yılında 179 işletmeden 167 işletmenin faaliyet raporuna ulaşılmış ve analize dâhil edilmiştir.

Çalışmada öncelikle faaliyet raporlarında yer alan metin veriler yapısal hale dönüştürülmüştür. Yapısal hale dönüştürme işlemi metin madenciliği adımları takip edilerek yapılmıştır. Daha sonra yapısal hale dönüştürülen metin veriler ile finansal performans göstergelerine ilişkin sayısal veriler birleştirilerek tek bir veri dosyası elde edilmiştir. Bu aşamadan sonra veri madenciliği sınıflandırma algoritmalarından destek vektör makineleri (DVM) algoritması kullanılarak veriler analiz edilmiştir. Çalışmada işletmeler finansal performanslarına göre kategorize edilerek belli sınıflara ayrılmış ve işletmenin dâhil olduğu sınıf tahmin edilmeye çalışılmıştır. Daha önce yapılan çalışmalarda da işletmelerin sınıflandırılması yoğunlukla kullanılan bir yaklaşım olarak karşımıza çıkmaktadır. Sınıflandırma sınırları ve sınıf sayısı kârlılık oranlarının dağılımları ve ortalaması dikkate alınarak yapılmıştır. Zira kârlılık oranlarının değerlendirilmesinde sektör ortalaması önemli bir kriterdir. Bu açıdan işletmeler ortalamaya yakın olanlar ve ortalamadan önemli derecede farklılaşanlar göz önünde bulundurularak üç sınıfa ayrılmıştır. Sınıf sınırları ise alt çeyrek ve üst çeyrek değerlerine göre belirlenmiştir. Buna göre birinci sınıfta yer alanlar düşük performanslı işletme, ikinci sınıfta yer alanlar orta performanslı işletme ve üçüncü sınıfta yer alanlar yüksek performanslı işletme olarak ifade edilmiştir.

Bilinen bu üç işletme sınıfı göz önüne alındığında, bu araştırmadaki hedef, işletmelerin 2010-2016 yılları arasındaki geçmiş 7 yıllık döneme ait faaliyet raporlarındaki metin veriler ile oluşturulan modelleri kullanarak 2017 yılı finansal performanslarına göre işletmelerin ait oldukları sınıfları tahmin etmektir. Yapılan DVM analizleri sonucunda oluşturulan modeller işletmelerin ait oldukları sınıfları öz kaynak kârlılığına göre %51,50 doğrulukla, net kâr marjına göre %55,69 doğrulukla ve hisse başına kâra göre %58,68 doğrulukla tahmin etmiştir. Buna göre oluşturulan modeller 2017 yılı için 167 işletmeden öz kaynak kârlılığı açısından 86 tanesinin, net kâr marjı açısından 93 tanesinin, hisse başına kâr açısından ise 98 tanesinin sınıfını doğru tahmin etmiştir. Bu sonuçlara göre DVM algoritması düşük, orta ve yüksek finansal performansa

sahip işletmelerin sınıflarını tahmin etmede başarılı olmuştur, ancak başarı seviyesi çok yüksek değildir. Fakat çalışmanın Qiu (2007) tarafından yapılan benzer çalışmaya yakın sonuçlar vermesi başarı sağlandığı kanısını desteklemektedir. Zira Qiu (2007) yaptığı çalışmada hisse başına kâr için kurduğu DVM modelinin ortalama doğruluk oranını %53,89, öz kaynak kârlılık oranı için kurduğu DVM modelinin ortalama doğruluk oranını ise %51,93 olarak ifade etmiştir (Qiu, 2007, s. 47).

DVM makine öğrenme algoritmasının imalat sanayi işletmelerinin 2010 ile 2016 yılları arasındaki döneme ilişkin faaliyet raporlarındaki metin veriler ile işletmelerin 2017 yılındaki finansal performansını tahmin etmede başarılı olması, işletmelerin faaliyet raporlarındaki metin veriler ile finansal performans arasında bir ilişki bulunduğunu kanıtlamaktadır. Ayrıca 2010-2016 yılları arasındaki döneme ilişkin metin veriler ile oluşturulan modelin işletmelerin 2017 yılı finansal performansını tahmin etmede sağladığı başarı, faaliyet raporlarının işletmelerin gelecekteki finansal performansları hakkında bilgi içerdiğini göstermektedir.

Bu çalışmada finansal performans ile metin veriler arasında ilişki bulunup bulunmadığına odaklanılmış olduğundan bütünsel bir yaklaşım sergilenmiştir. Bu nedenle kelime bazlı değerlendirmeye gidilmemiştir. Zira kelime bazlı değerlendirme bir sonraki adımdır ve daha kapsamlı analizler gerektirmektedir. Gelecek çalışmalarda kelime bazlı derin analizler yapılabilir. Bu kapsamda finansal performanstaki farklılaşmanın işletmelerin dilinde ne gibi değişiklikler meydana getirdiği araştırılabilir. Daha karamsar bir dil veya iyimser bir dil özelliğinin meydana gelip gelmediği, işletme performansın da önemli değişiklikler olmadan önce faaliyet raporlarındaki metin verilerin farklılaşıp farklılaşmadığı, finansal performansı düşük ve finansal performansı yüksek işletmelerde hangi kelimelerin ağırlık kazandığının belirlenmesi gibi konularda analizler yapılabilir. Ayrıca metin madenciliği ile faaliyet raporlarında yer alan ve analistler tarafından tespit edilmesi çok zor olan gizli ilişkilerin, anlamlı sonuçların elde edilmesi mümkün olabilecektir. Çalışma sonuçları metin verileri içeren finansal raporların veri ve metin madenciliği araçları ile çok daha hızlı ve az emek harcayarak analiz edilebileceğini kanıtlamaktadır.

Türkçe metin madenciliği için henüz diğer diller kadar iyi sonuçlar elde edilebilecek pratik metin madenciliği araçları bulunmamasına rağmen kullanılan yöntemle elde edilen sonuçların başarısı umut verici bir noktadır. Türkçe için metin işleme araçlarının gelişmesi ile birlikte çok daha iyi sonuçlar elde edilecektir. Bu

bakımdan gerek muhasebe gerekse diğer alanlarda Türkçe için metin madenciliği çalışmalarının desteklenmesi gerektiği düşünülmektedir.



## KAYNAKÇA

- Ağdeniz, Ş. (2017). *Finansal Raporların Analizinde Metin Madenciliğinin Kullanımı: Borsa İstanbul Şirketlerinin Kurumsal Yönetim Niteliklerinin Tahmini*. Yayınlanmamış Doktora Tezi. Eskişehir: Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Enstitüsü .
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İ. Ü. İşletme Fakültesi Dergisi*, 29(1), 1-22.
- Aras, G. ve Sarioğlu, G. U. (2015). *Kurumsal Raporlamada Yeni Dönem: Entegre Raporlama*. İstanbul: TÜSİAD.
- Aydın, S. (2015). *Entegre Raporlama*. İstanbul: Türkmen Kitabevi.
- Azevedo, A. and Santos, M. F. (2008). KDD, semma and CRISP-DM: A parallel overview. *European Conference on Data Mining* (s. 182-185). Amsterdam: IADIS.
- Back, B., Toivonen, J., Vanharanta, H. and Visa, A. (2001). Comparing numerical Data and Text Information From Annual Reports Using Self-Organizing Maps. *International Journal of Accounting Information Systems*, 2(4), 249-270.
- Bala, G., Bartel, H., Labs, T., Hawley, J. P. and Lee, Y.-J. (2015). Tracking “Real Time Corporate Sustainability Signals Using Cognitive Computing. *Journal of Applied Corporate Finance*, 27(2), 95-102.
- Balakrishnan, R., Qiu, X. Y. and Srinivasan, P. (2010). On the Predictive Ability of Narrative Disclosures in Annual Reports. *European Journal of Operational Research*, 3(202), 789-801.
- Barkemeyer, R., Figge, F., Holt, D. and Hahn, T. (2009). What the Papers Say: Trends in Sustainability: A Comparative Analysis of 115 Leading National Newspapers Worldwide. *The Journal of Corporate Citizenship*(33), 69-86.
- Boskou, G., Kirkos, E. and Spathis, C. (2019). Classifying Internal Audit Quality Using Textual Analysis: The Case of Auditor Selection. *Managerial Auditing Journal*, 34(8), 924-950.
- Cackett, D. (2013). *Information Management and Big Data A Reference Architecture*. Redwood Shores: Oracle Corporation.
- Can, Ü. ve Alataş, B. (2017). Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi. *International Journal of Pure and Applied Sciences*, 3(1), 75-111.

- Cecchini, M. (2005). *Quantifying the Risk of Financial Events Using Kernel Methods and Information Retrieval*. Doctor of Philosophy Thesis: University of Florida .
- Cecchini, M., Aytug, H., Koehler, G. J. and Pathak, P. (2010). Making Words Work: Using Financial Text as a Predictor of Financial Events. *Decision Support Systems*, 50, 164-175.
- Cemalcılar, Ö. ve Önce, S. (1999). *Muhasebenin Kuramsal Yapısı*. Eskişehir: T.C. Anadolu Üniversitesi Yayınları.
- Chakraborty, V., Chiu, V. and Vasarhelyi, M. (2014). Automatic Classification of Accounting Literature. *International Journal of Accounting Information Systems*, 15(2), 122-148.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). *CRISP-DM 1.0 Step-by-Step Data Mining Guide*. SPSS.
- Clatworthy, M. and Jones, M. J. (2003). Financial Reporting of Good News and Bad News: Evidence from Accounting Narratives. *Accounting and Business Research*, 33(3), 171-185.
- Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.
- Darman, G. M., Taştan, B., Seçkin, S. S. ve Kır, C. (2018). *Kurumsal Yönetim*. İstanbul: Sermaye Piyasası Lisanslama Sicil ve Eğitim Kuruluşu.
- Davis, A. K., Piger, J. M. and Sedor, L. M. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research*, 29(3), 845-868.
- Dawson, J. L. (1974). Suffix Removal and Word Conflation. *ALLC Bulletin, Michaelmas*, 33-46.
- Delibaş, A. (2008). *Doğal Dil İşleme ile Türkçe Yazım Hatalarının Denetlenmesi*. Yüksek Lisans Tezi. İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Deloitte. (2009, Ocak). *Yıllık faaliyet raporu hazırlama rehberi*.
- Doğanay, A. (2012). Bilimsel Yönteme Giriş. A. Şimşek içinde, *Sosyal Bilimlerde Araştırma Yöntemleri* (s. 2-29). Eskişehir: T.C. Anadolu Üniversitesi Yayını.
- Dolgun, M. Ö., Özdemir , T. G. ve Oğuz, D. (2009). Veri Madenciliği'nde Yapısal Olmayan Verinin Analizi: Metin ve Web Madenciliği. *İstatistikçiler Dergisi*, 2(2), 48-58.

- Duan, Z., He, Y. and Zhong, Y. (2018). Corporate Social Responsibility Information Disclosure Objective or Not: An Empirical Research of Chinese Listed Companies Based on Text Mining. *Nankai Business Review International*, 9(4), 519-539.
- Düzer, M. (2008). Finansal Analizde Kullanılan Oranlar ve Firma Değeri İlişkisi: İMKB’de Bir Uygulama. Yayımlanmamış Yüksek Lisans Tezi. Sakarya: Sakarya Üniversitesi, Sosyal Bilimler Enstitüsü.
- Düzer, M. (2018). *Sürdürülebilirlik Performans Göstergelerine İlişkin Açıklamaların Finansal Performans Üzerine Etkisi*. Yayımlanmamış Doktora Tezi. Eskişehir: Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü.
- Ergün, K. (2012). *Metin Madenciliği Yöntemleri İle Ürün Yorumlarının Otomatik Değerlendirilmesi*. Yayımlanmamış Doktora Tezi. Sakarya: Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.
- Eskici, H. B. ve Koçak, N. A. (2018). *SAS Enterprise Miner ile Metin Madenciliği*. İstanbul: Pusula Yayıncılık.
- Etzioni, O. (1996). The World Wide Web: Quagmire or Gold Mine? *Comm. of ACM*, 39(11), 65-68.
- Feldman, R. and Sanger, J. (2007). *The Text Mining Handbook, Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Fenyves, V., Böcskei, E., Bács, Z., Zéman, Z. and Tarnóczy, T. (2019). Analysis of the Notes to the Financial Statement Related to Balance Sheet in Case of Hungarian Information-Technology Service Companies. *Scientific Annals of Economics and Business*, 66(1), 27-39.
- Finnet 2000 Plus*. (2018, Eylül 1). finnet:  
<https://www.finnet2000.com/f2000plus/FinansalAnaliz/SektorOranAnalizleri>  
adresinden alındı
- Frakes, W. B. (1992). Stemming Algorithms. W. B. Frakes and R. Baeza-Yates içinde, *Information Retrieval: Data Structures and Algorithms* (s. 131-160). Upper Saddle River, New Jersey: Prentice Hall.
- Frazier, K. B., Ingram, R. W. and Tennyson B., M. (1984). A Methodology for the Analysis of Narrative Accounting Disclosures. *Journal of Accounting Research*, 22(1), 318-331.

- Gémar, G. and Jiménez-Quintero, J. A. (2015). Text Mining Social Media for Competitive Analysis. *Tourism & Management Studies*, 11(1), 84-90.
- Glancy, F. H. and Yadav, S. B. (2011). A Computational Model for Financial Reporting Fraud Detection. *Decision Support Systems*, 50(3), 595-601.
- Goel, S. and Gangolly, J. (2012). Beyond the Numbers: Mining the Annual Reports for Hidden Cues Indicative of Financial Statement Fraud. *Intelligent Systems in Accounting, Finance and Management*, 19(2), 75-89.
- Goel, S., Gangolly, J., R. Faerman, S. and Ozlem, U. (2010). Can Linguistic Predictors Detect Fraudulent Financial Filings? *Journal of Emerging Technologies in Accounting*, 7(1), 25-46.
- Gökbulut, R. İ. (2009). *Hissedar Değeri İle Finansal Performans Ölçütleri Arasındaki İlişki ve İMKB Üzerine Bir Araştırma*. Yayınlanmamış Doktora Tezi. İstanbul: İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü.
- Gupta, R. and Gill, N. S. (2012). Financial Statement Fraud Detection using Text Mining. *International Journal of Advanced Computer Science and Applications*, 3(12), 189-191.
- Güvenç, B. (2016). *Machine Learning Methods in Natural Language Processing*. Yüksek lisans Tezi. İstanbul: Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü.
- Hájek, P. (2018). Combining Bag-of-Words and Sentiment Features of Annual Reports to Predict Abnormal Stock Returns. *Neural Computing and Applications*, 29(7), 343-35.
- Hajek, P. and Henriques, R. (2017). Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud - A Comparative Study of Machine Learning Methods. *Knowledge-Based Systems*, 128, 139-152.
- Hamde, M. A. (2018). *Kurumsal Belgelere (Metin Verilerine) Metin Madenciliği Tekniği ile Erişimin Değerlendirilmesi: Türk Özel Sektörüne Yönelik Bir İnceleme*. Doktora Tezi. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.
- Han, J., Kamber, M. and Pei, J. (2001). *Data Mining Concepts And Techniques* (Third Edition b.). USA: Elsevier.
- Heidari, M. and Felden, C. (2015). Financial Footnote Analysis: Developing a Text Mining Approach. *In Proceedings of International Conference on Data Mining*, 10-16.

- Holbrook, M. E. (2010). *Corporate Social Responsibility and Financial Performance: An Examination of Economic Benefits and Costs as Manifested in Accounting Earnings*. PhD (Doctor of Philosophy) Thesis. Kentucky: Kentucky University.
- Hotho, A., Nürnbergger, A. and Paass, G. (2005). A Brief Survey of Text Mining. *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*(20), 19-62.
- Humpherys, S. L., Moffitt, K. C., Burns, Mary, B., Burgoon, J. K. and Felix, W. F. (2011). Identification of Fraudulent Financial Statements Using Linguistic Credibility Analysis. *Decision Support Systems*, 50(3), 585-594.
- IIRC. (2013). *Uluslararası <ER> Çerçevesi*. IIRC. <http://integratedreporting.org/wp-content/uploads/2015/03/13-12-08-THE-INTERNATIONAL-IR-FRAMEWORK-Turkish.pdf> adresinden alındı
- Kamaruddin, S. S., Abu, B. A., Hamdan, A. R., Nor, F. M., Mohd, Z. A., Othman, Z. A. and Hussein, G. S. (2015). A Text Mining System for Deviation Detection in Financial Documents. *Intelligent Data Analysis*, 19, 19-44.
- Kavzoğlu, T. ve Çölkesen, İ. (2010). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. *Harita Dergisi*(144), 73-82.
- Kenton, W. (2018, Temmuz 5). investopedia: <https://www.investopedia.com/terms/f/financialperformance.asp> adresinden alındı
- Kırlı, M., Kuğu, T. ve Kara, İ. (2013). Finansal Performansın Ölçümünde Nakit Katma Değer Yöntemi (CVA) ve Borsa İstanbul'a Kote Şirketlerde Bir Uygulama. *SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi* , 13(26), 98-123.
- Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanharanta, H. and Visa, A. (2004). Combining Data and Text Mining Techniques for Analysing Financial Reports. *Intelligent Systems in Accounting, Finance and Management*, 12(1), 29–41.
- Kohut, G. F. and Segars, A. H. (1992). The President's Letter to Stockholders: An Examination of Corporate Communication Strategy. *The Journal of Business Communication*, 29(1), 7-21.
- Koyuncu, B. (2010). *Değere Dayalı ve Geleneksel Finansal Performans Ölçütlerinin Karşılaştırmalı Analizi: İMKB'de Bir Uygulama*. Doktora Tezi. Ankara: Ankara Üniversitesi, Sosyal Bilimler Enstitüsü.

- Krishnamoorthy, S. (2018). Sentiment Analysis of Financial News Articles Using Performance Indicators. *Knowledge and Information Systems*, 56(2), 373–394.
- Krovetz, R. (1993). Viewing Morphology as an Inference Process. *SIGIR '93 Proceedings of the 16th annual international ACM SIGIR Conference on Research and Development in Information Retrieval* (s. 191-202). Pittsburgh, Pennsylvania, USA: ACM.
- Li, F. (2006). *Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?* Available at SSRN: <https://ssrn.com/abstract=898181> or <http://dx.doi.org/10.2139/ssrn.898181>.
- Li, F. (2008). *The Determinants and Information Content of the Forward-looking Statements in Corporate Filings - A Naive Bayesian Machine Learning Approach*. AAA 2009 Financial Accounting and Reporting Section (FARS) Paper: Available at SSRN: <https://ssrn.com/abstract=1267235> or <http://dx.doi.org/10.2139/ssrn.1267235>.
- Liddy, E. D. (2001). *Natural Language Processing*. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.
- Liew, W. T., Adhityab, A. and Srinivasan, R. (2014). Sustainability Trends in the Process Industries: A Text Mining-Based Analysis. *Computers in Industry*, 65(3), 393-400.
- Lin, S.-J. and Hsu, M.-F. (2018). Decision Making by Extracting Soft Information from CSR News Report. *Technological and Economic Development of Economy*, 24(4), 1344-1361.
- Liu, Y. and Moffitt, K. C. (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), 85-94.
- Lovins, J. B. (1968). Development of a Stemming Algorithm. *Mechanical Translation and Computational Linguistics*, 11(1 and 2), 22-31.
- Manning, C. D., Raghavan, P. and Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Miner, G., Elder, J., Hill, T., Delen, D., Fast, A. and Nisbet, R. A. (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Elsevier Science.

- Nelson, K. K. and Pritchard, A. C. (2007). *Litigation Risk and Voluntary Disclosure: The Use of Meaningful Cautionary Language*. 2nd Annual Conference on Empirical Legal Studies Paper: Available at SSRN: <https://ssrn.com/abstract=998590> or <http://dx.doi.org/10.2139/ssrn.998590>.
- Nisbet, R., Elder, J. and Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier.
- Osuna, E. E., Freund, R. and Girosi, F. (1997). *Support Vector Machines: Training and Applications*.
- Önce , S. (2011). Oran Analizi. F. Çömlekçi içinde, *Muhasebe Denetimi ve Mali Analiz* (s. 221-250). Eskişehir: Anadolu Üniversitesi Yayınları.
- Önce, S., Onay, A., Yeşilçelebi, G. ve Özkes, S. (2015). Güçlü Şirket Olmada İpucu; Kurumsal Sürdürülebilirlik Raporlaması ve GRI. *Eskişehir Sanayi Odası Dergisi*(11), s. 62-66.
- Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*, 2(3), 65-82.
- Özkuş, F. U. ve Pektekin, P. (2009). Muhasebe Yolsuzluklarının Tespitinde Adli Muhasebecinin Rolü ve Veri Madenciliği Tekniklerinin Kullanılması. *Muhasebe Bilim Dünyası Dergisi*, 11(4), 57-88.
- Özsözgün Çalışkan, A. (2012). Sürdürülebilirlik Raporlaması. *Muhasebe ve Vergi Uygulamaları Dergisi*, 41-48.
- Öztürk, S. (2019). Geleceğin Kurumsal Raporlama Yaklaşımı Olarak Entegre Raporlama: Garanti Bankası Örneği. *Muhasebe ve Finansman Dergisi*(81), 1-20.
- Piskorski, J. and Yangarber, R. (2013). Information Extraction: Past, Present and Future. T. Poibeau, H. Saggion, J. Piskorski and R. Yangarber içinde, *Multi-source, Multilingual Information Extraction and Summarization* (s. 23-49). Berlin, Heidelberg: Springer.
- Porter, M. F. (1980). An Algorithm for Suffix Stripping. *Program*, 14(3), 130-137.
- Purda, L. and Skillicorn, D. (2015). Accounting Variables, Deception, and A Bag of Words: Assessing the Tools of Fraud Detection. *Contemporary Accounting Research*, 32(3), 1193-1223.
- Qiu, X. Y. (2007). *On building Predictive Models with Company Annual Reports*. PhD (Doctor of Philosophy) Thesis: University of Iowa.

- Rahrovi Dastjerdi, A., Foroghi, D. and Kiani, G. H. (2019). Detecting Manager's Fraud Risk Using Text Analysis: Evidence from Iran. *Journal of Applied Accounting Research*, 20(2), 154-171.
- Rich, K. T., Roberts, B. L. and Zhang, J. X. (2016). Linguistic Tone of Municipal Management Discussion and Analysis Disclosures and Future Financial Reporting Delays. *Journal of Emerging Technologies in Accounting*, 13(2), 93-107.
- Rivera, S. J., Minsker, B. S., Work, D. B. and Roth, D. (2014). A Text Mining Framework for Advancing Sustainability Indicators. *Environmental Modelling & Software*, 62, 128-138.
- Savaş, S., Topaloğlu, N. ve Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*(21), 1-23.
- Sevim, Ş. (2013). Kârlılık Analizi. S. Önce içinde, *Finansal Tablolar Analizi* (s. 132-157). Eskişehir: Anadolu Üniversitesi Yayını.
- Shahi, A. M., Issac, B. and Modapothala, J. R. (2014). Automatic Analysis of Corporate Sustainability Reports and Intelligent Scoring. *International Journal of Computational Intelligence and Applications*, 13(1).
- Shirata, C. Y. and Sakagami, M. (2008). An Analysis of the “Going Concern Assumption”: Text Mining from Japanese Financial Reports. *Journal of Emerging Technologies in Accounting*, 5(1), 1-16.
- Shirata, C. Y., Takeuchi, H., Ogino, S. and Watanabe, H. (2011). Extracting Key Phrases as Predictors of Corporate Bankruptcy: Empirical Analysis of Annual Reports by Text Mining. *Journal of Emerging Technologies in Accounting*, 8(1), 31-44.
- Singh, A. and Ghosh, U. (2013). Text Mining: A Burgeoning Technology for Knowledge Extraction. *International Journal of Scientific Research Engineering & Technology*, 1(12), 22-26.
- Singh, J. and Gupta, V. (2017). A Systematic Review of Text Stemming Techniques. *Artificial Intelligence Review*, 48(2), 157–217.
- Singh, K. N., Devi, H. M. and Mahanta, A. K. (2017). Document representation techniques and their effect on the document Clustering and Classification: A Review. *International Journal of Advanced Research in Computer Science*, 8(5), 1780-1784.

- Song, I.-Y. and Zhu, Y. (2016). Big Data and Data Science: What Should We Teach? *Expert Systems*, 33(4), 364-373.
- Sönmez, N. (2017). *Çevrimiçi Yorumların Metin Madenciliği ile Analizi: İstanbul'daki Alışveriş Merkezleri Üzerine Bir Çalışma*. Yüksek lisans Tezi. İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Subramanian, R., Insley, R. G. and Blackwell, R. D. (1993). Performance and Readability: A Comparison of Annual Reports of Profitable and Unprofitable Corporations. *The Journal of Business Communication*, 30(1), 49-61.
- Syed, A. R., Gillela, K. and Venugopal, C. (2013). The Future Revolution on Big Data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(6), 2446–2451.
- Şamiloğlu, F. (2005). Piyasa Katma Değeri ve Hisse Getirileri: İMKB'deki İmalat Sanayi Şirketlerinde Ampirik Bir Uygulama. *Muhasebe ve Finansman Dergisi*(25), 79-88.
- Topal, Y. (2008). Ekonomik Katma Değer (EVA), Pazar Katma Değer (MVA) Arasındaki İlişki İMKB İmalat İşletmelerinden Örnek. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 13(2), 249-261.
- Tsai, M.-F. and Wang, C.-J. (2017). On the Risk Prediction and Analysis of Soft Information in Finance Reports. *European Journal of Operational Research*, 257(1), 243-250.
- Wahyuni, R., Febrianto, R. and Rahman, A. (2018). Firm Performance and Readability of the Annual Report. *International Journals of Sciences and High Technologies*, 12(1), 25-36.
- WBCSD. (2002). *Sustainable Development Reporting: Striking the Balance*. Switzerland: WBCSD Reports.
- Wei, L., Li, G., Zhu, X. and Li, J. (2019). Discovering Bank Risk Factors from Financial Statements Based on A New Semi-Supervised Text Mining Algorithm. *Accounting & Finance*, 59(3), 1525-1558.
- Yucel, A. (2016). *Predictive Text Analytics and Text Classification Algorithms*. PhD (Doctor of Philosophy) thesis. Alabama: Auburn University.
- Yücebaş, S. C. ve Tintin, R. (2017). Gövde-Türk: Bir Türkçe Gövdeleme Yöntemi. *The 2nd International Conference on Computer Science and Engineering*, (s. 343-347). Antalya.

Zaki, M. and Theodoulidis, B. (2013). Analyzing Financial Fraud Cases Using a Linguistics-Based Text Mining Approach. *SSRN Electronic Journal*.



## ÖZGEÇMİŞ

Adı Soyadı : Sait PEKİN  
Yabancı Dil : İngilizce  
Doğum Yeri ve Yılı : Şereflikoçhisar /1981  
E-Posta : saitpekin@aksaray.edu.tr

### Eğitim Geçmişi:

- 2005, Niğde Üniversitesi, Aksaray İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü
- 2011, Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü, Muhasebe ve Finansal Yönetim Anabilim Dalı

### Mesleki Geçmişi:

- 2007-2012, Bilgisayar İşletmeni, Mehmet Akif Ersoy Üniversitesi, Bucak Emin Gülmez Teknik Bilimler Meslek Yüksekokulu
- 2012, Şef, Mehmet Akif Ersoy Üniversitesi, Bucak Emin Gülmez Teknik Bilimler Meslek Yüksekokulu
- 2012-2018, Öğretim Görevlisi, Aksaray Üniversitesi, Şereflikoçhisar Berat Cömertoğlu Meslek Yüksekokulu
- 2018 - devam ediyor, Öğretim Görevlisi, Aksaray Üniversitesi, Ortaköy Meslek Yüksekokulu

### Yayınları

- Dalğar, H. ve Pekin, S. (2011), “Kurumsal Yönetim İle Finansal Tablo Manipülasyonu Arasındaki İlişki: İMKB Kurumsal Yönetim Endeksinde Yer Alan Şirketlerde Bir Araştırma”, Mali Çözüm Dergisi, Sayı:107, 19–44.