

Yayın Geliş Tarihi (Submitted): 17/02/2022

Yayın Kabul Tarihi (Accepted): 23/04/2022

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Koşuk Ünlü, H. (2022), Başlığında “data science” ifadesi geçen uluslararası kongrelerde sunulan bildiri özetlerinin metin madenciliği yöntemleri ile incelenmesi, *Nicel Bilimler Dergisi*, 4(1), 1-21. doi:10.51541/nicel.1075225

BAŞLIĞINDA “DATA SCIENCE” İFADESİ GEÇEN ULUSLARARASI KONGRELERDE SUNULAN BİLDİRİ ÖZETLERİNİN METİN MADENCİLİĞİ YÖNTEMLERİ İLE İNCELENMESİ

Hande Koşuk Ünlü¹

ÖZ

Veri Bilimi kavramının ortaya çıkması ile birlikte, veriye dayalı karar verme süreçleri ve bu süreçlerde kullanılan yöntemler, klasik istatistiksel yöntemlerin yanı sıra bilişim teknolojilerinin sağladığı olanakların kullanıldığı makine öğrenme, derin öğrenme vb. algoritmaların gelişmesini sağlamıştır. Günümüzde veriyi elde etmeden daha çok, hızla artan bir veri yapısı için veriyi anlama, örüntülerini ortaya koyma, işlevsel biçimde modelleme ve sonuçlarının üretilmesi önem kazanmıştır. Bu nedenle “Veri Bilimi” kavramı istatistik kongrelerinin adı ve içeriğinde de değişime neden olmuştur. Bu çalışma, metin madenciliği kullanılarak istatistik kongrelerinde geçen “Veri Bilimi” kavramının incelemesine odaklanmıştır. Bu kongrelerde sunulan ve bildiri kitaplarında basılan bildirilerin özet içeriklerinin incelenmesi ve eğilimlerinin belirlenmesi, veri biliminde önemli bir yere sahip olan metin madenciliği yöntemleri bakımından incelenmiştir. Bu çalışma, veri bilimi kavramı ile birlikte kongrelerde yer alan bildiri özetlerinin içeriklerindeki eğilimleri ortaya koymaktadır.

Anahtar Kelimeler: Veri bilimi, Metin Madenciliği, Kelime Bulutu, Web Kazıma

¹Sorumlu yazar, Dr, Hacettepe Üniversitesi, Halk Sağlığı Enstitüsü, Sağlık Araştırmaları ABD, Ankara, Türkiye. ORCID ID: <https://orcid.org/0000-0003-3572-0254>

AN EVALUATION OF ABSTRACTS PRESENTED IN THE INTERNATIONAL CONGRESSES WHICH INCLUDES “DATA SCIENCE” EXPRESSION IN TITLE USING TEXT MINING METHOD

ABSTRACT

By the emergence of the concept of Data Science, the data-based decision-making processes and the methods used in these processes has enabled the development of algorithms by using the opportunities provided by information technologies as well as classical statistical methods, like machine learning, deep learning, etc. At the present time, understanding the data, revealing the patterns, modeling the data in a functional way and producing the results have gained importance for a rapidly increasing data structure rather than obtaining the data. For this reason, the concept of “Data Science” caused a change in the name and content of statistical congresses. The purpose of this research is to utilize text mining to examine the concept of "Data Science" in statistics conferences. Examining the summary contents of the presentations in these congresses and published in the proceedings and determining their tendencies were examined in terms of text mining methods, which have an important place in data science. This study reveals the trends in the contents of the abstracts in congresses with the concept of data science.

Keywords: Data Science, Text Mining, Word Cloud, Web Scrapping

1. GİRİŞ

“Veri Bilimi” kavramına gerek üniversiteler gerekse akademisyenler çok büyük önem vermektedir. “Veri Bilimi” uygulamalarıyla, özellikle makine öğrenmesi, veri madenciliği teknikleri, yapay zekâ, büyük veri, derin öğrenme konuları daha da ön plana çıkmıştır. Bu konuların iyi anlaşılması ve doğru kullanılması da ayrıca önem taşımaktadır. Veri bilimi kapsamında ele alınan konuların en güncel izlendiği yerler kongrelerdir. Kongrelerde veri biliminin ele alınış biçimi kavramsal olarak bu alanda çalışanların vizyonunu yönlendirebilir. Bu doğrultuda metin madenciliği gibi yeni yaklaşımlarla konuların nasıl ele alındığının ortaya konulması, tanımlanması, durum saptama ve planlamada yardımcı olabilir.

“Veri Bilimi” terimi ilk kez 1996 yılında, bir istatistiksel konferans (Uluslararası Sınıflandırma Dernekleri Federasyonu (IFCS) “Veri Bilimi, sınıflandırma ve ilgili yöntemler”) başlığına dahil edilmiştir (Press, 2022). Terim istatistikçiler tarafından önerilmiş olsa da “Veri

Bilimi” 'nin kamusal imajında, bilgisayar bilimi ve iş uygulamalarının önemi, özellikle de büyük veri döneminde, çok daha fazla vurgulanmaktadır (Weihs ve Ickstadt, 2018). 1970'lerde John Tukey'in fikirleri; verilerden hipotezler türetme şeklindeki istatistik bakış açısını, hipotez test edilmeden önce verileri anlamaya çalışma olarak değiştirmiştir (Cao, 2007; Weihs ve Ickstadt, 2018).

“Veri bilimi”nin Hayashi (1998) ve Dhar (2013) tarafından yapılan tanımı, “yapılandırılmış ve yapılandırılmamış verilerden bilgi ve öngörü elde etmek için bilimsel yöntemleri, süreçleri, algoritmaları ve sistemleri kullanan çok disiplinli bir alan” olarak geçmektedir. Ayrıca “Veri Bilimi” nin veri madenciliği ve büyük verilerle ilişkili olduğu ve gerçek olayları verilerle anlamak ve analiz etmek için istatistik, makine öğrenimini ve ilgili yöntemlerini birleştiren bir kavram” olduğu ifade edilmiştir (Dhar, 2013; Hayashi, 1998). Kauermann ve Seidl, “Veri Bilimi” için yapılan bu tanımı vurgulayarak istatistik ve bilgisayar bilimlerini, “Veri Bilimi” nin temel disiplinleri olduğunu belirtmişlerdir (Kauermann ve Seidl, 2018). Son zamanlarda “Veri Bilimi” için yapılan en kapsamlı tanımlarından biri Cao tarafından aşağıdaki şekilde formüle edilmiştir.

$$\text{Veri Bilimi} = (\text{istatistik} + \text{bilişim} + \text{bilgi işlem} + \text{iletişim} + \text{sosyoloji} + \text{yönetim}) \mid (\text{veri} + \text{çevre} + \text{düşünme})$$

Yukarıda verilen ifadede yer alan sosyoloji terimi, sosyal yönleri ifade etmektedir. (veri + çevre + düşünme) terimlerinden önce gelen “[” işareti ise bahsedilen tüm bilimlerin veriye, çevreye ve veriden bilgiye, bilgiden bilgeliğe düşüncesine dayandığını göstermektedir (Cao, 2007).

Veri analizi ve bilimi, yeni ekonominin ve internet, sosyal & mobil ağlar ve bulut bilişimi gibi etki alanlarının yürütülmesi; telekomünikasyon, biyoloji, sağlık ve sosyal bilimler gibi klasik disiplinlerin yeniden biçimlendirilmesi ve aynı zamanda temel iş ve ekonomik aktivitelerin iyileştirilmesi için önemli bir paradigma olarak ortaya çıkmıştır. Veriye dayalı bilimsel keşif, yenilik ve uygulamalar; büyük veri kümelerinin anlaşılabilirliğinin sağlanması ve verilerden elde edilen sonuçların eyleme dönüştürülebilmesi için gerekli hale gelmiştir (Cao, 2007).

Literatürde “Veri Bilimi” kavramıyla ilgili yapılmış birçok yayın mevcuttur. Yıldız, 2015 yılında Bilim ve Teknik’te çıkan yazısında büyük veri ve veri bilimi kavramlarından bahsetmiş; veri bilimcinin kimlere deneceğini açıklamıştır (Yıldız, 2015). Gökalp ve arkadaşlarının 2018 yılında yaptıkları çalışmalarında, veri bilimi ve büyük verinin hangi yönleriyle birbirine benzediği, hangi yönleriyle birbirinden ayrıldığı ve veri odaklı bir

organizasyon kurmak için veri bilimi alanından kaynaklanan sorunların neler olabileceğini araştırmışlardır (Gökalp vd, 2018). Kılıç ve arkadaşları, 2019 yılında yaptıkları çalışmalarında veri analizinde yaygın olarak kullanılan veri madenciliği ve istatistiksel yöntemler arasındaki ilişkiyi incelemişlerdir (Kılıç vd., 2019). Temmuz Gürcanok 2020 yılında yaptığı tez çalışmasında e-ticaret sayfasında tüketicilerin satın alma tercihlerini incelemiş; en çok satın alınan, birlikte alınan ve birlikte alınabilecek olan ürünlerin analizini ve tahminini veri madenciliği yöntemleri kullanarak yapmaya çalışmıştır (Temmuz Gürcanok, 2020). Albayrak, 2020 yılında yaptığı çalışmasında “*Data Science and Advanced Analytics*” adındaki konferansın 2015-2018 yıllarına ait bildirimlerinin başlık ve anahtar kelimelerini doğal dil işleme tekniklerini kullanarak incelemiş ve analiz sonuçlarından yararlanarak “Veri Bilimi ve Uygulamaları” adında bir lisansüstü ders içeriği oluşturmuştur (Albayrak, 2020). Bunlarla birlikte Türkiye’de 2018 yılında kurulmuş “Veri Bilimi” isimli bir dergi bulunmaktadır (Gök, 2022). Bu doğrultuda ülkemizde “Veri Bilimi” kavramına olan ilgi artışı ve oldukça popüler bir çalışma alanı haline gelmesi dikkat çekicidir. Ülkemizde bu konudaki eğitimler genel olarak “Veri Analitiği” adı altında verilmektedir ve birçok üniversite bünyesinde yüksek lisans programları açılmıştır.

“Veri Bilimi” odaklı incelemelerin nereye yöneldiğinin saptanması, ders içeriklerinin revize edilmesi ve bu alanda çalışan akademisyenlere ışık tutması açısından önemlidir. Bu çalışmanın amacı, Türkiye’de ve yurtdışında düzenlenen, başlığında “Veri Bilimi (Data Science)” ifadesi bulunan ve kongre dili İngilizce olan kongrelerde sunulan ve bildiri kitaplarında basılan bildirimlerin özet içeriklerinin metin madenciliği yöntemleri kullanılarak incelenmesi, tercih edilen veri analiz yöntemlerin neler olduğunun belirlenmesi, yıllara ve kongrenin yapıldığı yere göre değişip değişmediğinin araştırılmasıdır. Metin madenciliği, yapılandırılmamış formattaki verileri düzenli ve yapılandırılmış bir veri formatına dönüştürerek, elde edilen bu veriden anlamlı örüntüler çıkarmayı amaçlayan bir yöntem olduğu için bu çalışmada tercih edilmiştir. Bu bağlamda, araştırma kapsamında incelenecek bildiri özetleri, yıl, kullanılan yöntem ve kongrenin yapıldığı yer yönünden değerlendirilmiştir.

Çalışmanın 2. bölümünde, kullanılan veri seti ile metin madenciliği yöntemi ve uygulama adımlarından bahsedilmiştir. 3. bölümde analizde kullanılacak veri setinin toplanması, ön işlenmesi ve dönüştürülmesi ile özellik seçimi aşamaları detaylı olarak anlatılmış ve analiz sonuçları sunulmuştur. Son bölümde ise tartışma ve sonuç bölümü yer almaktadır.

2. MATERYAL VE METOT

2.1. Veri Seti

Çalışmada kullanılan veriler, 2012 yılında düzenlenen “1st International Conference on Data Science, Technology and Applications (DATA)” konferansından başlayarak 27 Ocak 2021 tarihine kadar Türkiye’de ve yurtdışında düzenlenen, başlığında “Veri Bilimi (Data Science)” ifadesi bulunan, kongre dili İngilizce olan ve IEEEExplore, ACM Proceedings ve Google veri tabanlarında taranan kongrelerden toplanmıştır. Bu amaçla, ilgili tüm kongrelerin özet kitabına elektronik olarak açık kaynak yoluyla erişilmeye çalışılmıştır. Çalışmaya dahil edilen kongrelerin listesi Tablo 1’de verilmiştir.

Tablo 1. Çalışmaya dahil edilen kongreler ve bildiri sayıları

Kongrenin Adı	Kongrenin Yılı	Kongrenin Düzenlendiği Yer	Bildiri Sayısı
1st International Conference on Data Science, Technology and Applications (DATA)	2012	Yurtdışı	32
2nd International Conference on Data Science, Technology and Applications (DATA)	2013	Yurtdışı	34
BigDataScience '14: International Conference on Big Data Science and Computing	2014	Yurtdışı	32
CoDS '14: 1st IKDD Conference on Data Sciences	2014	Yurtdışı	7
IEEE DSAA 2014: 1st IEEE International Conference on Data Science and Advanced Analytics	2014	Yurtdışı	92
3rd International Conference on Data Science, Technology and Applications (DATA)	2014	Yurtdışı	43
CoDS '15: Second ACM IKDD Conference on Data Sciences	2015	Yurtdışı	30
CoDS-IKDD'15: 2nd IKDD Conference on Data Sciences	2015	Yurtdışı	10
IEEE DSAA 2015: 2nd IEEE International Conference on Data Science and Advanced Analytics	2015	Yurtdışı	125
4th International Conference on Data Science, Technology and Applications (DATA)	2015	Yurtdışı	36
CoDS-IKDD'16: 3rd IKDD Conference on Data Sciences	2016	Yurtdışı	27
IEEE DSAA 2016: 3rd IEEE International Conference on Data Science and Advanced Analytics	2016	Yurtdışı	87
5th International Conference on Data Science, Technology and Applications (DATA)	2016	Yurtdışı	23
iDSC2017: 1st International Data Science Conference	2017	Yurtdışı	17
CoDS '17: Fourth ACM IKDD Conference on Data Sciences	2017	Yurtdışı	16
IEEE DSAA 2017: 4th IEEE International Conference on Data Science and Advanced Analytics	2017	Yurtdışı	83
6th International Conference on Data Science, Technology and Applications (DATA)	2017	Yurtdışı	32
CoDS-COMAD '18: ACM India Joint International Conference on Data Science and Management of Data	2018	Yurtdışı	49
DATA '18: International Conference on Data Science, E-learning and Information Systems 2018	2018	Yurtdışı	30
DSIT '18: 2018 International Conference on Data Science and Information Technology	2018	Yurtdışı	30

Tablo 1 (Devam). Çalışmaya dahil edilen kongreler ve bildiri sayıları

IEEE DSAA 2018: 5th IEEE International Conference on Data Science and Advanced Analytics	2018	Yurtdışı	79
1st International Conference on Data Science and Applications (ICONDATA'18)	2018	Türkiye	90
7th International Conference on Data Science, Technology and Applications (DATA)	2018	Yurtdışı	39
CoDS-COMAD '19: ACM India Joint International Conference on Data Science and Management of Data	2019	Yurtdışı	62
DATA '19: International Conference on Data Science, E-learning and Information Systems 2019	2019	Yurtdışı	30
DSIT '19: 2019 2nd International Conference on Data Science and Information Technology	2019	Yurtdışı	30
IEEE DSAA 2019: 6th IEEE International Conference on Data Science and Advanced Analytics	2019	Yurtdışı	65
2nd International Conference on Data Science and Applications (ICONDATA'19)	2019	Türkiye	60
DMS-2019, International Conference on Data Science, Machine Learning and Statistics 2019	2019	Türkiye	88
8th International Conference on Data Science, Technology and Applications (DATA)	2019	Yurtdışı	51
IDSES'19 - 1st International Data Science & Engineering	2019	Türkiye	63
CoDS COMAD 2020: 7th ACM IKDD CoDS and 25th COMAD	2020	Yurtdışı	67
DSIT '20: 2020 3rd International Conference on Data Science and Information Technology	2020	Yurtdışı	30
IEEE DSAA 2020: 7th IEEE International Conference on Data Science and Advanced Analytics	2020	Yurtdışı	108
3rd International Conference on Data Science and Applications (ICONDATA'20)	2020	Türkiye	44
9th International Conference on Data Science, Technology and Applications (DATA)	2020	Yurtdışı	34
CODS COMAD 2021: 8th ACM IKDD CODS and 26th COMAD	2021	Yurtdışı	88

2.2. Metin Madenciliği

Metin madenciliği, veri madenciliğinin bir alt türüdür. Metin madenciliği, pdf, Word formatındaki metin dosyaları, bloglar, sosyal medya paylaşımları gibi yapısal olmayan metinlerden oluşan veri tabanlarından anlamlı bilgi elde edilebilmesi amacıyla kullanılmaktadır. Bu amaçla metin içinde önemli, anlamlı, anlaşılabilir ve kullanılabilir örüntüler belirlenir ve analiz edilir. Metin madenciliği, veri madenciliğine benzemekle birlikte veri madenciliği yapısal veriyi analiz edebilecek şekilde geliştirilmiştir, fakat metin madenciliği yapısal olmayan ya da yarı-yapısal (e-postalar, HTML dosyaları vb.) veri setleri üzerinde de analiz yapılabilir (Gupta ve Lehal, 2009).

Metin madenciliği; metinlerin toplanmasıyla başlayan, toplanan metinlerin bir takım veri ön işleme ve dönüştürme süreçlerinden geçirilmesini içeren, istatistiksel ve veri

madenciliği yöntemleri aracılığıyla anlamlı bilgiye ulaşmayı hedefleyen, akış şeması Şekil 1’de sunulan beş adımlık bir süreçtir (Dang ve Ahmad, 2014).



Şekil 1. Metin madenciliği uygulama adımları

Adım 1: Metinlerin toplanması

Analiz edilecek veri, araştırma problemine göre uygun platformlardan (bloglar, sosyal medya, çeşitli internet sayfaları, vb.) toplanır. Veriler, internet kaynaklı veri tabanlarından elde edileceği gibi kişilerin sahip oldukları metin dosyalarından da elde edilebilir.

Adım 2: Metinlerin ön işlenmesi ve dönüştürülmesi

Metinlerin toplanmasıyla elde edilen veri, analiz edilebilir hale getirilebilmesi için bir takım ön işleme aşamalarından geçirilir. Toplanan metin içinde veri temizliği için; internet bağlantılarının ve “hashtag” lerin kaldırılması, noktalama işaretlerinin, simgelerin ve rakamların silinmesi, küçük harfe dönüştürme işlemleri yapılır (Gümüşsu ve Murat, 2019). Veri temizliği aşaması bittikten sonra metin dizgilere ayrılır, durak ve gereksiz kelimeler çıkartılır ve kelime kökleri ayrıştırılır.

Adım 3: Özellik seçimi

İstatistik ve veri madenciliği yöntemlerinin anlamlı sonuç verebilmesi için metin üzerinde araştırma konusunu en iyi açıklayan kelime ya da kelime öbeklerinin tespit edilmesi gerekir. Bu bölüm, çalışma konusuna ve amacına göre değişebilmektedir. Bir önceki adımda köklerine ayrıştırılan kelimelere ağırlıklandırma işlemi uygulanır. Ağırlıklandırma sonucu belirli bir eşik değerinin altında ağırlığa sahip olan kelimeler metinden çıkarılır. Bu işlemle boyut indirgeme yapılarak veri madenciliği yöntemlerinde analiz süresi kısaltılmaktadır. Çeşitli ağırlıklandırma yöntemleri bulunmaktadır (Başkaya ve Aydın, 2017).

- Binary ağırlıklandırma (Binary weighting):

Bu ağırlıklandırma yönteminde, kelime ya da kelime öbekleri metin içinde bulunuyorsa “1”, bulunmuyorsa “0” değerini alacak şekilde ağırlıklandırılır.

- Terim frekansına göre ağırlıklandırma (Term frequency weighting-TF):

Bu ağırlıklandırma yönteminde, kelime ya da kelime öbekleri metin içinde bulunma sıklıklarına göre ağırlıklandırılır.

- Ters belge frekansına göre ağırlıklandırma (Inverse document frequency-IDF):

Bu ağırlıklandırma yönteminde, daha çok tekrarlanan kelime ya da kelime öbeklerinin ağırlıkları azaltılırken, daha az kullanılan kelime ya da kelime öbeklerinin ağırlıkları arttırılır. Ters belge sıklığı, Eşitlik 1’deki gibi hesaplanır.

$$IDF = \log \left(\frac{\text{incelenen toplam metin sayısı}}{\text{kelime/kelime öbeklerinin geçtiği metin sayısı}} \right) \quad (1)$$

- Terim frekans-Ters belge frekans ağırlıklandırması (Term frequency-Inverse document frequency):

Bu ağırlıklandırma yönteminde, terim frekansı ve ters belge frekansı bir arada kullanılır. Bu ağırlıklandırma yöntemine göre, ayırt edici olan kelime ya da kelime öbekleri elde edilmeye çalışılır. Terim frekans – ters belge frekansı ağırlığı, Eşitlik 2'deki gibi hesaplanır.

$$TF - IDF = TF \times IDF \quad (2)$$

Adım 4: Veri madenciliği yöntemlerinin uygulanması

Özellik seçimi aşamasında ağırlıklandırılarak sayısal hale getirilen kelime ya da kelime öbeklerinden, anlamlı bilgiye ulaşılabilmesi için veri madenciliğinde kullanılan sınıflandırma ve kümeleme algoritmalarına başvurulur.

Adım 5: Bilgiye erişim

Veri madenciliği yöntemlerinin uygulanmasıyla elde edilen sonuçlar, araştırma sorusuna cevap olabilecek nitelikte görselleştirilir ve yorumlanır. Görselleştirme kapsamında grafikler, tablolar, haritalandırma gibi çeşitli görseller kullanılabilir.

3. BULGULAR

Bu çalışmada, Türkiye'de ve yurtdışında düzenlenen, başlığında “Veri Bilimi (Data Science)” ifadesi bulunan ve kongre dili İngilizce olan kongrelerde sunulan ve bildiri kitaplarında basılan bildiri özetleri incelenmiştir.

3.1. Metinlerin Toplanması

Araştırma kapsamında ilgili kriteri sağlayan kongrelere ilişkin erişim adresleri bulunarak listelenmiştir. Bu kongrelerde sunulan ve indirilebilen içeriğe sahip bildirimler R programlama dili versiyon 4.0.0 (R Core Team, 2022)'a aktarılmış ve “tm” (Feinerer, Hornik ve Meyer, 2008) ve “stringr” (Wickham, 2022) kütüphaneleri kullanılarak düzenlenip analize hazır hale getirilmiştir. Web ortamında bulunan ve indirilemeyen bildiri özetleri ise web

kazıma yöntemleri ile excel formatında indirilmiştir. İlgili web sitelerinden web kazıma yoluyla bildiri özetlerini otomatik olarak indirebilmek için “Chrome web driver” kurulmuş, Python versiyon 3.8.3 (Python Software Foundation, 2022) programlama dilindeki “BeautifulSoup” (Richardson, 2022) ve “Selenium” (Muthukadan, 2022) kütüphaneleri kullanılmıştır. Bu yöntemler ile ulaşılan bildiri özetleri, bildiri başlığı, kongrenin adı ve yılı ile düzenlendiği yer bilgileri toplanmıştır.

Bildiri özetlerinin Web kazıma yoluyla indirilebilmesi için araştırmacı tarafından bir kod parçacığı yazılmıştır. Örnek bir kongre için yazılan kod parçacığı genel hatlarıyla aşağıda verilmiştir.

```
browser=webdriver.Chrome("C:/.../chromedriver")
browser.get("https://dl.acm.org/doi/proceedings/10.1145/2640087")
...
soup = BeautifulSoup(browser.page_source, 'html.parser')
...
Big_Data14=pd.DataFrame(columns=("type", "year", "conference name", "abstract
title", "abstract"))

s=0
section_titles=soup.find_all(class_="section__title accordion-tabbed__control left-
bordered-title")

for sections in range(0,len(section_titles)):
    url=section_titles[sections]['href']

    browser=webdriver.Chrome("C:/.../chromedriver")
    browser.get(url)
    ...

    soup = BeautifulSoup(browser.page_source, 'html.parser')
    articles = soup.find_all("h5", class_="issue-item__title")

    for article in articles:
        ...
        Big_Data14.loc[s, "type"] = "International"
        Big_Data14.loc[s, "year"] = 2014
        Big_Data14.loc[s, "conference name"] = "BigDataScience '14: International
Conference on Big Data Science and
Computing"
        Big_Data14.loc[s, "abstract title"] = article.text
        Big_Data14.loc[s, "abstract"] = soup.find("div", class_="abstractSection").text
        s=s+1
        print(s)
        print(article)
```

Yukarıdaki kod parçacığı çalıştırıldığı elde edilen ekran görüntüleri Şekil 2’de verilmiştir.

```
1 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644151">Unlocking Value from Open Data</a></h5>
2
3 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644153">Computing Issues for Big Data: Theory, Systems and Applications</a></h5>
4
5 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644155">Scaling Distributed Machine Learning with the Parameter Server</a></h5>
6
7 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644151">Unlocking Value from Open Data</a></h5>
8
9 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644153">Computing Issues for Big Data: Theory, Systems and Applications</a></h5>
10
11 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644155">Scaling Distributed Machine Learning with the Parameter Server</a></h5>
12
13 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644156">Trending Words Based Event Detection in Sina Weibo</a></h5>
14
15 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644157">A feature set for streams and an application to high-frequency financial tick data</a></h5>
16
17 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644162">The Value of Using Big Data Technologies in Computational Social Science</a></h5>
18
19 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644163">Using Hadoop To Implement a Semantic Method Of Assessing The Quality Of Research Medical Dataset</a></h5>
20
21 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644151">Unlocking Value from Open Data</a></h5>
22
23 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644153">Computing Issues for Big Data: Theory, Systems and Applications</a></h5>
24
25 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644155">Scaling Distributed Machine Learning with the Parameter Server</a></h5>
26
27 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644150">Community Detection for Clustered Attributed Graphs via a Variational EM Algorithm</a></h5>
28
29 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644151">Unlocking Value from Open Data</a></h5>
30
31 <h5 class="issue-item__title"><a href="/doi/10.1145/2640087.2644153">Computing Issues for Big Data: Theory, Systems and Applications</a></h5>
32
```

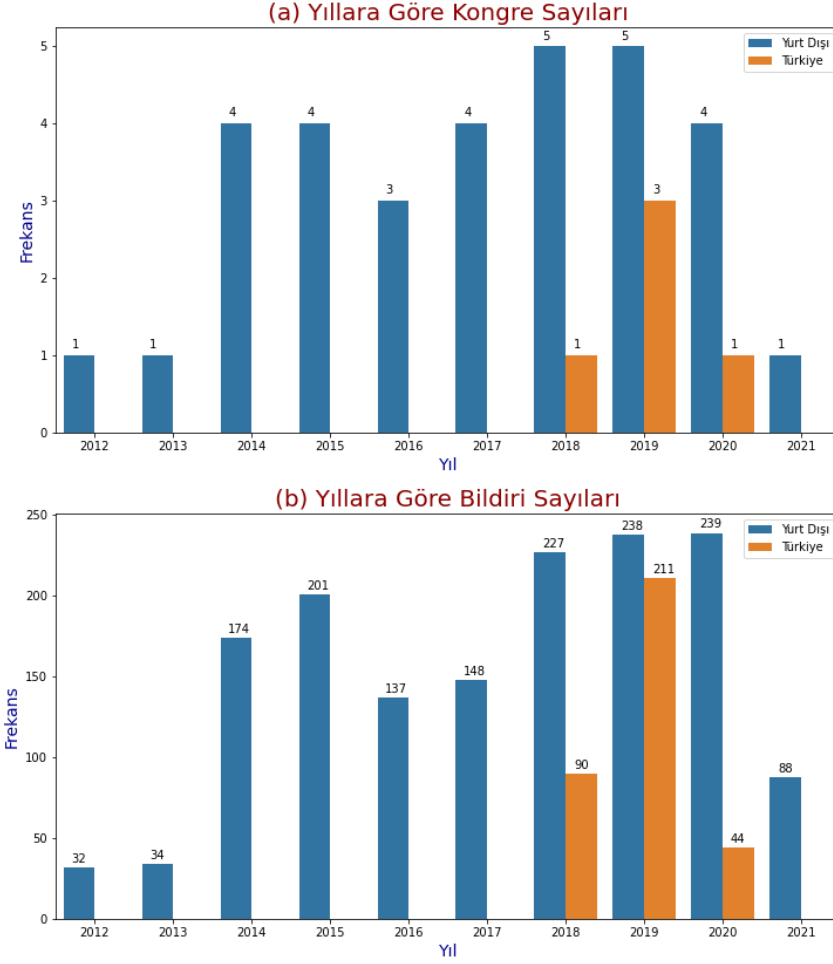
(a) Web kazıma yoluyla çekilen bildiri özeti listesi

	type	year	conference name	abstract title	abstract
0	International	2014	BigDataScience '14: International Conference o...	Unlocking Value from Open Data	Open data has become an important trend in ini...
1	International	2014	BigDataScience '14: International Conference o...	Computing Issues for Big Data: Theory, Systems...	Big data may contain big values, but also brin...
2	International	2014	BigDataScience '14: International Conference o...	Scaling Distributed Machine Learning with the ...	Big data may contain big values, but also brin...
6	International	2014	BigDataScience '14: International Conference o...	Trending Words Based Event Detection in Sina W...	Online social networks provide us an unprecede...
7	International	2014	BigDataScience '14: International Conference o...	A feature set for streams and an application t...	We propose a set of features to study the effe...
8	International	2014	BigDataScience '14: International Conference o...	The Value of Using Big Data Technologies in Co...	The discovery of phenomena in social networks ...
9	International	2014	BigDataScience '14: International Conference o...	Using Hadoop To Implement a Semantic Method Of...	In this paper a system for storing and queryin...
13	International	2014	BigDataScience '14: International Conference o...	Community Detection for Clustered Attributed G...	Community detection for attributed graphs, als...
17	International	2014	BigDataScience '14: International Conference o...	Detecting weak public transport connections fr...	Many modern and growing cities are facing decl...
18	International	2014	BigDataScience '14: International Conference o...	Galactica: A GPU Parallelized Database Acceler...	The amount of business data generated and coll...
19	International	2014	BigDataScience '14: International Conference o...	Scalable Hierarchical Recommendations Using Sp...	Collaborative Filtering (CF) is one of the mos...
20	International	2014	BigDataScience '14: International Conference o...	Recommending Friends in Local Social Networks...	As the proliferating of online social networks...
24	International	2014	BigDataScience '14: International Conference o...	Creating a Chemistry of Sciences with Big Data...	The Data Science Institute at Imperial College...
25	International	2014	BigDataScience '14: International Conference o...	Hidden Aspect Rating Discovery from Text Revie...	We investigate hidden aspect mining problem th...
26	International	2014	BigDataScience '14: International Conference o...	Data Readiness Level For Unstructured Data Wit...	As we entered into the so called big data era,...
27	International	2014	BigDataScience '14: International Conference o...	Clustering Experiments on Big Transaction Data...	This paper addresses the Volume dimension of B...
31	International	2014	BigDataScience '14: International Conference o...	SQLPhi: A SQL-Based Database Engine for Intel ...	Relational Database Management Systems (RDBMS...

(b) Listelenen bildirilerin analize hazır formata çevrilmesi

Şekil 2. Web kazıma sonucunda elde edilen ekran görüntüleri

Bu çalışmada toplam 1879 bildiriye ulaşılmıştır. Fakat incelemek bildiri özetlerindeki bilgi eksiklikleri nedeniyle analize 1863 bildiri özeti dahil edilmiştir. Yıllara ve kongrenin yapıldığı yere göre kongre ve bildiri sayılarını gösteren grafik Şekil 3’te verilmiştir.



Şekil 3. Kongrenin yapıldığı yıl ve yere göre (a) kongre ve (b) bildiri sayıları

Şekil 3 incelendiğinde “Data Science” kavramına olan ilginin son yıllarda daha çok arttığı görülmektedir. Başlığında “Data Science” ifadesinin geçtiği kongreler ilk olarak 2012 yılında yurt dışında düzenlenmiştir. Yıllara göre yurt dışında düzenlenen kongrelerin sayısında genel olarak bir artış gözlenmiştir. Benzer biçimde kongre sayılarındaki artış ilgili kongrelerde sunulan bildiri sayılarına da yansımıştır. Türkiye’de ise başlığında “Data Science” ifadesinin geçtiği kongreler ilk defa 2018 yılında gerçekleşmiştir. Türkiye’de yapılan kongrelerdeki bildiri sayılarında 2019 yılında ciddi bir artış gözlenmiş; fakat 2020 yılında tüm dünyayı etkisi altına alan COVID-19 pandemisi nedeniyle düzenlenen kongre ve kongrelerde sunulan bildiri sayılarında bir düşüş gözlenmiştir. Şekil 3’te 2021 yılında düzenlenen kongre sayısının yalnızca bir tane olduğu görülmektedir. Bu durumun nedeni bu çalışmanın 2021 yılının ilk ayında sona ermesi olup bu yıla ait bildiri sayıları henüz düşük sayıdadır (Şekil 3).

3.2. Metinlerin Ön İşlenmesi ve Dönüştürülmesi

Ulaşılan bildiri özetlerine ait yapılan tüm analizler için R programlama dili (R Core Team, 2022) kullanılmıştır. Elde edilen bildiri özetleri veri temizliği sürecine tabi tutulmuştur. Veri temizleme sürecinde “tm” (Feinerer, Hornik ve Meyer, 2008) ve “stringr” (Wickham, 2022) kütüphaneleri kullanılmıştır. Öncelikle metindeki tüm kelimeler küçük harfe dönüştürülmüş, noktalama işaretleri kaldırılmış, rakamlar silinmiş ve metinde yer alan gereksiz boşluklar kaldırılmıştır. Daha sonra tek harften oluşan kelimeler ile yazım hatası veya abartı anlamı içeren 3 ya da daha fazla tekrar eden harflerden oluşan kelimeler (Örneğin, deeeep, recurrent, goooood) de metinden çıkarılmıştır.

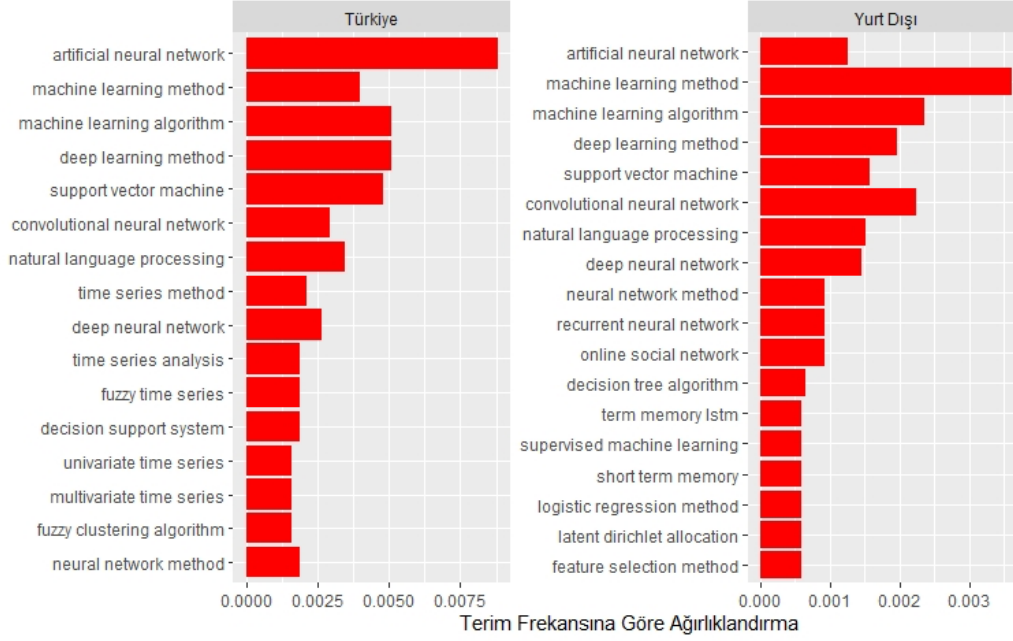
Veri temizliği aşaması bittikten sonra metin, dizgilere ayrılmıştır. Metni dizgilere ayırma işleminin yapılabilmesi için “tidytext” (Silge ve Robinson, 2016) kütüphanesi kullanılmıştır. Kullanılan kütüphanede yer alan duraklama kelimelerinden oluşan liste yardımıyla incelenen metin durak ve gereksiz kelimelerden arındırılmıştır. Ayrıca incelenen metne uygun olarak araştırmacı tarafından hazırlanan duraklama kelimeleri de metinden çıkarılmıştır.

3.3. Özellik Seçimi

Özellik seçiminin yapılabilmesi için “tidytext” (Silge ve Robinson, 2016) kütüphanesinde yer alan N-gram algoritması kullanılmıştır. N-gram algoritması, kelime ya da kelime öbeklerinin metin içinde ne sıklıkla geçtiğini elde etmek için kullanılmaktadır (Takcı ve Baktır, 2018). N-gram algoritmasında yer alan N değeri kelime öbeklerindeki kelime sayılarını göstermektedir. Örneğin, N değeri 1 olduğu (uni-gram) zaman tek kelimedenden oluşan ifadeler elde edilir.

Bu çalışmanın yapısı gereği daha anlamlı sonuçlara ulaşabilmek için N değeri 3 (tri-gram) olarak alınmıştır. N-gram yöntemine göre elde edilen kelime öbekleri terim frekansına (tf) göre ağırlıklandırılmıştır. Yapılan ağırlıklandırma yöntemlerine göre en sık geçen kelime öbekleri tespit edilmiş ve elde edilen sonuçlar “ggplot2” (Wickham vd., 2022) kütüphanesi ile görselleştirilmiştir.

N-gram algoritmasındaki N değeri 3 olarak alındığında ve terim frekans ağırlıklandırması yapıldığında elde edilen kelime öbekleri kongrenin yapıldığı yere göre Şekil 4’te verilmiştir.

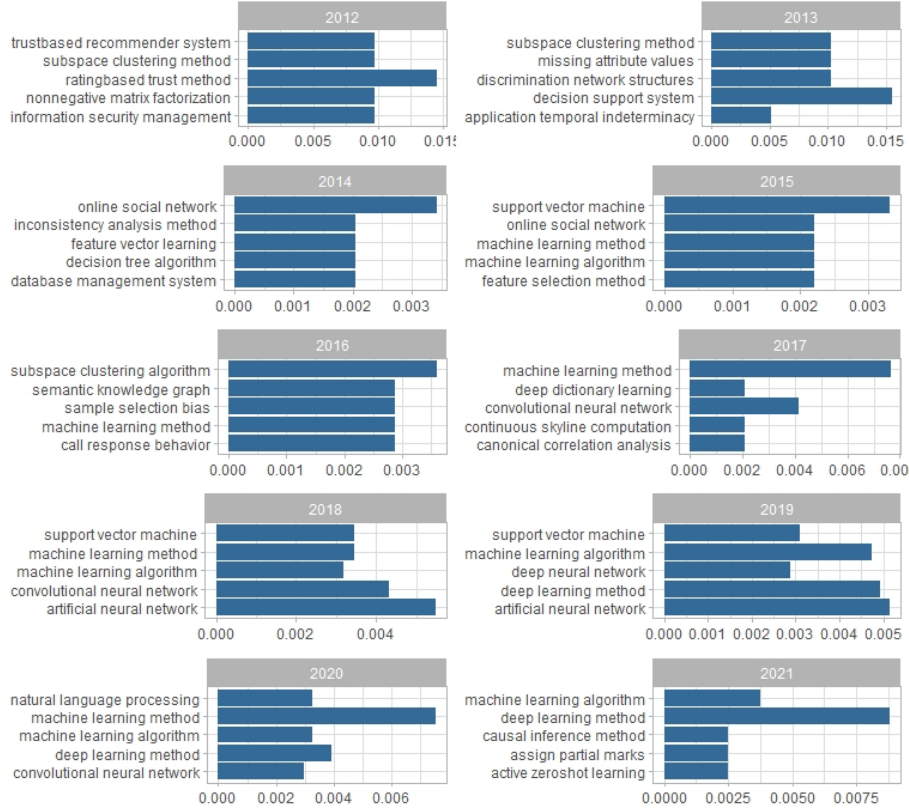


Şekil 4. Kongrenin yapıldığı yere göre terim frekans ağırlıklandırma yöntemiyle elde edilen kelime öbekleri

Şekil 4'te verilen grafik incelendiğinde Türkiye'de ve yurt dışında düzenlenen kongrelerin bildiri özetlerinde ilk 7 sırada geçen kelime öbekleri aynı olarak bulunmuştur. Türkiye'de yapılan kongre bildirilerinde yurt dışında yapılan kongrelerin bildirilerinden farklı olarak "time series analysis", "univariate time series", "multivariate time series", "fuzzy time series" kelime öbekleri öne çıkmaktadır. Yurt dışında yapılan kongre bildirilerinde ise "deep neural network", "recurrent neural network", "short term memory" gibi deep learning yöntemlerine ait kelime öbekleri öne çıkmaktadır (Şekil 4).

Kongrenin yapıldığı yıla göre terim frekansı ağırlıklandırma yöntemiyle elde edilen kelime öbekleri Şekil 5'de verilmiştir. Şekil 5'de verilen grafik incelendiğinde 2012 yılında düzenlenen kongre bildirilerinde "trustbased recommender system", "subspace clustering method" ve "nonnegative matrix factorization" gibi kelime öbekleri öne çıkmaktadır. 2013 yılında düzenlenen kongrelerde "decision support system" ve "missing attribute values" gibi kelime öbekleri öne çıkmaktadır. 2014-2016 yılları arasında düzenlenen kongrelerde "machine learning method", "support vector machine", "online social network", "decision tree algorithm" ve "subspace clustering method" gibi kelime öbekleri öne çıkmaktadır. 2017 yılından günümüze kadar geçen zaman içinde düzenlenen kongrelerde ise "deep learning

method”, “deep neural network”, “artificial neural network”, convolutional neural network” ve “natural language processing” gibi kelime öbekleri öne çıkmaktadır (Şekil 5).



Şekil 5. Kongrenin yapıldığı yıla göre terim frekans ağırlıklandırma yöntemiyle elde edilen kelime öbekleri

Kelime öbeklerinden elde edilen kelime bulutu Şekil 6’de verilmiştir. Tri-gram yöntemine göre elde edilen kelime bulutu incelendiğinde kongrelerin bildiri özetlerinde en sıklıkla karşılaşılan kelime öbeğinin “artificial neural network” olduğu görülmektedir. “deep neural network”, “support vector machine”, “natural language processing” kelime öbekleri de çok kullanılan ifadeler arasında yer almaktadır. Ayrıca düşük ağırlığa sahip olsa da “logistic regression method”, “latent dirichlet allocation”, “principal component analysis”, “hidden markov model”, “automatic text summarization”, “time series analysis”, “particle swarm optimization”, “matrix factorization method”, “canonical correlation analysis” ve “social network analysis” gibi ifadelerin bildiri özetlerinde yer aldığı görülmektedir (Şekil 6).

“machine learning algorithm”, “deep learning method”, support vector machine”, “convolutional neural network” ve “natural language processing” dir. İlk 6 sıradaki kelime öbeklerinin aynı olması, kongrelerde çalışılan konuların benzer olduğunu göstermektedir. Türkiye’de yapılan kongre bildirilerinde yurt dışında yapılan kongrelerin bildirilerinden farklı olarak zaman serisi yöntemleri öne çıkarken yurt dışında yapılan kongre bildirilerinde ise “deep learning” yöntemlerine ait kelime öbekleri öne çıkmaktadır. Tercih edilen yöntemlerin yurtdışında ve Türkiye’de farklı olmasının sebebi verinin kullanım amaçlarına bağlı olarak değişiklik göstermesi olabilir. Türkiye’de veriler daha çok ileriye dönük öngörülerde bulunabilmek amacıyla toplanmakta ve zaman serileri yöntemlerine başvurularak analiz edilmektedir. Ayrıca, her ne kadar ilk 6 sıradaki kelime öbekleri arasında “deep learning” konuları yer alsa da ülkemizde bu konuların, yurtdışındaki kadar yaygın çalışılmadığı; bu konuların gelecekte çok daha popüler bir çalışma konusu olacağı ve ülkemizde de daha yaygın olarak çalışılacağı düşünülmektedir.

Literatürde, 2020 yılında yapılan bir çalışmada, 2015-2018 yılları arasında düzenlenen “*Data Science and Advanced Analytics*” adlı konferans serisindeki bildiri başlıkları ve anahtar kelimeleri, doğal dil işleme teknikleriyle incelenmiş ve en sık geçen 20 kelime kullanılarak lisansüstü bir ders içeriği oluşturulmuştur (Albayrak, 2020). Bu çalışmada ise “Data Science” ifadesinin geçtiği kongrelerdeki bildiri özetleri incelenerek “Veri Bilimi” kavramının zaman içerisindeki değişimi ortaya konmuştur. Bu yaklaşım, literatürde yapılacak benzer çalışmalara örnek teşkil etmektedir.

Bu çalışma, “Veri Bilimi” alanında eğitim almak ve bu alanda uzmanlaşmak isteyen araştırmacıların gelecek vizyonlarına yardımcı olabilir. “Veri Bilimci”, 21. yüzyılın en havalı meslekleri arasında yer almaktadır (Davenport ve Patil, 2012). Literatürde 2018 yılında yapılan bir çalışmada iş başvurularında “Veri Bilimci” için istenilen gereksinim, beceri ve niteliklerin neler olduğu metin madenciliği yöntemi kullanılarak araştırılmıştır. Araştırma sonuçları, en çok istenilen özellikler arasında “Machine Learning”, “Statistics”, “Big Data”, “Data Mining”, “Natural Language”, “Python” ve “R” gibi terimlerinin yer aldığını göstermektedir (Radovitsky vd., 2018). Bu sonuçlar, “Veri Bilimi” alanında inter ve trans disiplinler çalışmanın önemine vurgu yapmaktadır. Bu bağlamda, bu çalışma ile “Veri Bilimi” konularının zaman içindeki seyrinin incelenmesi, gerek bu alanda uzmanlaşmak isteyen “Veri Bilimci” lere yön göstermekte gerekse değişen gereksinimlere göre bu konuda verilen eğitim içeriklerinin güncellenmesine önemli katkı sağlamaktadır. Bu alanda eğitim içeriklerinin oluşturulması ve

güncellenmesinin daha işlevsel hale getirilmesi için inter disiplinler çalışmalarının zenginleştirilmesi gerekmektedir.

Kongrelerde dijital ortamdaki bildirilerin minimum paylaşım standartlarının belirlenmesi bu tip çalışmalara kaynak sağlayarak, elde edilecek veri niteliğini arttıracak ve bilgi üretimini kolaylaştıracaktır. Bu çalışmadaki veri setinde kullanılan bildiri özetleri bazı kongreler için yapılandırılmış veri formatında sunulmuş olmasına rağmen bazı kongrelere ait bildiri özetleri yapılandırılmamış bir veri formatında sunulmuştur. Web kazıma yöntemleri kullanılarak veriler otomatik bir şekilde çekilmeye çalışılsa da hem bildiri özetlerinin elde edilişi hem de yapılandırılmış ve yapılandırılmamış olarak sunulan bildiri özetlerinin birleştirilerek standart bir veri formatı haline getirilmesi oldukça zaman alıcı bir süreç olmuştur. Bu süreçte, kongre internet sayfalarının kapanmış olması, bildiri özet kitapçıklarının kongre internet sitesinde yayımlanmamış olması, bildiri özet kitapçıklarının bağlantı adresinin çalışmaması ve internet ortamında yaşanabilecek sorunlardan kaynaklanan internet sayfalarına erişilememesi durumu araştırmanın kısıtlılıklarıdır. Bilimsel çalışmaların açık erişim ortamlarında paylaşılmasında ikincil veri analizine olanak sağlayacak şekilde standart bir formatta sunulması önerilir.

Bu çalışmanın sonunda elde edilen bilgiler araştırmacılar için, “Veri Bilimi” nin zaman içinde kat ettiği yolun ve son zamanlarda daha çok tercih edilen analiz yöntemlerinin belirlenmesi, kullanılan istatistiksel yöntemler ile bilgisayar algoritmalarının ortaya konulması açısından önem taşımaktadır. Veri biliminde kullanılan analiz yöntemlerinde özellikle büyük verinin ortaya çıkmasıyla bir paradigma değişikliği olmuştur. Bu değişiklik, yeni gereksinimler doğurması ve farklı yaklaşımlara temel teşkil edebilecek olması nedeniyle önemlidir. Bu doğrultuda veri analizinde kullanılan klasik yaklaşımlara (Örneğin, linear regression, logistic regression vb.) ek olarak “artificial neural network”, “convolutional neural network”, “deep learning”, “natural language processing” ve “recommender system algorithms” gibi veri analiz yöntemlerine başvurulması gerekmektedir.

Metin madenciliği, yapılandırılmamış metinler inceleyerek yeni bilgiler çıkarmayı, bağlamı keşfetmeyi veya metni yapılandırılmış bir veri formatına dönüştürmeyi amaçlayan bir yöntemdir. Bu çalışma, incelenen metinlerden önemli bilgilerin çıkarılmasında metin madenciliğinin önemini göstermektedir. Metin madenciliği yöntemleri, araştırmacıların ilgilendiklerin konularda metinleri inceleyerek metinlerdeki gizli bilgileri keşfetmelerine ve ilgilendiklerini alanı daha iyi anlamalarına yardımcı olmaktadır.

ETİK BEYAN

“Başlığında “Data Science” İfadesi Geçen Uluslararası Kongrelerde Sunulan Bildiri Özetlerinin Metin Madenciliği Yöntemleri İle İncelenmesi” başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

KAYNAKÇA

- Albayrak, A. (2020), Doğal dil işleme teknikleri kullanılarak disiplinler arası lisansüstü ders içeriği hazırlanması, *Bilişim Teknolojileri Dergisi*, 13(4), 373-383.
- Başkaya, F. ve Aydın, İ. (2017), Haber metinlerinin farklı metin madenciliği yöntemleriyle sınıflandırılması, *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Türkiye.
- Cao, L. (2017), Data science: A comprehensive overview, *ACM Computing Surveys*, 50(3), 1-42.
- Dang, S. and Ahmad, P. H. (2014), Text Mining: Techniques and its application, *IJETI International Journal of Engineering & Technology Innovations*, 1(4), 22-25.
- Davenport, T. H. and Patil, D. J. (2012). Data Scientist: The sexiest job of the 21st Century, *Harvard Business Review*, 90(10), 70–76.
- Dhar, V. (2013), Data science and prediction, *Communications of the ACM*, 56(12), 64-73.
- Feinerer, I., Hornik K. and Meyer, D. (2008), Text Mining Infrastructure in R, *Journal of Statistical Software*, 25(5), 1–54.
- Gök, M. (2022), *Veri Bilimi*, <https://dergipark.org.tr/en/pub/veri>. Erişim Tarihi: 14. 02.2022.
- Gökalp, M. O., Kayabay, K., Çoban, S., Yandık, Y. B. ve Eren, P. E. (2018), Büyük veri çağında işletmelerde veri bilimi, *5. International Management Information Systems Conference (IMISC)*, 24-26 Kasım 2018, Ankara, Türkiye.
- Gupta V. and Lehal, G. S. (2009), A survey of text mining techniques and applications, *Journal of Emerging Technologies in Web Intelligence*, 1(1), 60-76.

- Gümüüşsu, E. ve Murat, N. (2019), ‘Tamam’ ve ‘Devam’ etiketleriyle atılan tweetler ile cumhur ve millet ittifakı grupları arasındaki ilişkinin incelenmesi, *Bilişim Teknolojileri Dergisi*, 12(4), 287-298.
- Hayashi, C. (1998), What is data science? Fundamental concepts and a heuristic example, Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., Baba Y. (eds), *Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization*, 40-51, Springer, Tokyo.
- Kauermann, G. and Seidl, T. (2018), Data science: A proposal for a curriculum, *International Journal of Data Science and Analytics*, 6(3), 195–199.
- Kılıç, G. G., Turanlı, M. and Özden, Ü. H. (2019), Data mining and statistics in data science, *International Social Sciences Studies Journal*, 5(30), 960–968.
- Muthukadan, B. (2022), *Selenium Python documentation*. <https://selenium-python.readthedocs.io/installation.html>_Erişim Tarihi: 14 Şubat 2022.
- Press, G. (2022), *A Very Short History of Data Science*, <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=56de23555cfc>. Erişim Tarihi: 14.02.2022.
- Python Software Foundation (2022), *Python Language Reference*, version 3.8.3, <http://www.python.org>_Erişim Tarihi: 14.02.2022.
- R Core Team (2022), *R: A language and environment for statistical computing*, <https://www.R-project.org/>_Erişim Tarihi: 14.02.2022.
- Radovilsky, Z., Hegde, V., Acharya, A. and Uma, U. (2018), Skills requirements of business data analytics and data science jobs: A comparative analysis, *Journal of Supply Chain and Operations Management*, 16(1), 82-101.
- Richardson, L. (2022), *Beautiful soup documentation*, Python library version 4.10.0, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>_Erişim Tarihi: 14.02.2022.
- Silge, J. and Robinson, D. (2016), tidytext: Text mining and analysis using tidy data principles in R, *JOSS*, 1(3), 1-3.
- Sütçü, C. ve Aytekin, Ç. (2018), *Veri Bilimi*, 1. baskı, (137-140), Paloma Yayınevi, İstanbul, Türkiye.

- Takcı H. ve Baktır, N. (2018), Büyük veri yaklaşımıyla birden çok bilgi erişim merkezinin kolektif kullanımı, *Bilişim Teknolojileri Dergisi*, 11(2), 123-129.
- Temmuz Gürcanok, O. (2020), E-Ticaret satış verileri üzerinde bir veri bilimi vaka çalışması, Yüksek Lisans Tezi, Maltepe Üniversitesi, İstanbul, Türkiye.
- Weihs, C. and Ickstadt, K. (2018), Data science: The impact of statistics, *International Journal of Data Science and Analytics*, 6(3), 189–194.
- Wickham, H (2022), *stringr: Simple, Consistent Wrappers for Common String Operations Package*. R package version 1.4.0, <https://cran.r-project.org/web/packages/stringr/index.html>. Erişim Tarihi: 14.02.2022.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D. and RStudio (2022). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.3.5, <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>. Erişim Tarihi: 14.02.2022.
- Yıldız, U. A. (2015), Büyük veri kahramanı veri bilimci, *Bilim ve Teknik*, 569, 76-79.