

Sosyal Medya Alanında Yazılan Büyük Veri Konulu Tezlerin Metin Madenciliğiyle İncelenmesi

Sadullah ÇELİK*

Fatma ZEREN**

Öz

Metinler, birçok konuda yararlı bilgiler içerebilir. Metinleri analiz etmek, insanların daha iyi kararlar vermesine, daha etkili işler yapmasına ve daha fazla bilgiye ulaşmasına yardımcı olabilir. Sosyal medya gibi zengin kaynaklardan elde edilen veriler, bu metinlere ait büyük verilerdir. Bu verileri anlamak ve yorumlamak için çeşitli yöntemler kullanılmaktadır. Metin madenciliği ve veri analitiği gibi teknikler, en çok kullanılan yöntemlerdir. Ayrıca, bu metin verilerinin içerdiği bilgiyi ortaya çıkarabilmek için yapılandırılmış verilerden daha fazla veriye ihtiyaç duyulabilir. Bu çalışmada, Türkiye’de hazırlanmış sosyal medyadan sağlanan büyük veri yaklaşımı ile ilgili lisansüstü tezler ele alınmıştır. Bu tezler, çeşitli anabilim dalları tarafından hazırlanmıştır ve bu nedenle büyük metinler ile ilgili pek çok boyut incelenmiştir. Bu bağlamda bu tezlerden bazılarında kısa özetler sunulmuştur. En fazla bilgisayar mühendisliği alanında konu ile ilgili tez yazıldığı gözlenmiştir. Ancak bu tezlerin nitelikleri birbirinden ayrılmaktadır. Bir kısmında hedef yazılım boyutuna yönelik olurken, bir kısmında ise sosyal medya informasyonu hakkında analiz yapılmasına yöneliktir. Bilgisayar mühendisliğinden sonra en fazla iletişim alanındaki çeşitli anabilim dallarında çok fazla tez yazılmıştır. Yıllara göre büyük veri konusunda yazılan tez sayısının giderek attığı gözlemlenmiştir. Çalışmada metin madenciliği yöntemi kullanılarak 2008-2022 yılları arasında yazılmış tezlere kelime analizi yapılmıştır. Analiz sonucunda

* Arş. Gör. Dr., Aydın Adnan Menderes Üniversitesi, Nazilli İktisadi ve İdari Bilimler Fakültesi, Uluslararası Ticaret ve Finansman Bölümü, sadullah.celik@adu.edu.tr, Orcid: 0000-0001-5468-475X.

** Prof. Dr., İnönü Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, fatma.zeren@inonu.edu.tr, Orcid: 0000-0002-1661-3587.

tezlerin kelime dağılımlarının kuvvet yasası dağılımına uygun olduğu bulunmuştur. Elde edilen bulgulardan tezlerde aşırı odaklanma sorunu olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Büyük Veri, Sosyal Medya, Kuvvet Yasası Dağılımı, Kelime Bulutu, Aşırı Odaklanma.

The Analysis of Graduate Studies on Big Data on Social Media through Text Mining

Sadullah ÇELİK - Fatma ZEREN

Abstract

Texts may contain useful information on many topics. Analyzing texts can help people make better decisions, do more effective work, and access more information. Data obtained from rich sources such as social media constitute big data belonging to these texts. Various methods are employed to understand and interpret these data. Among them, text mining and data analytics are the most widely used techniques. In addition, there may be need for more data than available through structured data to excavate the information contained in a given text data. This article examines graduate theses prepared in Turkey employing big data approach obtained from social media. These studies have been prepared by various departments and hence big data has been examined from various aspects. In this regard, this paper provides brief summaries of some these theses. The findings reveal that the majority of related theses were written in the field of computer engineering. However, their characteristics differ from each other. While some target the software aspects, others analyze social media information. The next most popular field is the various departments in the field of communications. It has been observed that the number of theses written on big data has increased over the years. This study has applied word analysis on theses written between 2008 and 2022 through the text mining method. The results confirm the congruity of the word distribution in theses to the power law distribution. The overall findings point to the problem of excessive focus in theses.

Keywords: Big Data, Social Media, Power Law Distribution, Word Cloud, Excessive Focus.

Giriş

XXI. yüzyıl, verilerin her zamankinden daha önemli olduğu bir çağdır. Bunun nedeni, dünyanın artık her zamankinden daha fazla birbirine bağlı olması ve verilerin bu birbirine bağlılığı anlamının ve manipüle etmenin anahtarı olmasıdır. Bu veriler, devlet kurumları, özel şirketler ve bireyler dahil olmak üzere birçok kaynaktan gelmektedir. Veriler, eğilimleri ve kalıpları anlamamıza ve gelecek hakkında tahminler yapmamıza imkan sağlar. Aynı zamanda etkili pazarlama ve reklamcılığın anahtarıdır. Kısacası, modern dünyayı anlamak ve gezinmek için veriler gereklidir. Ayrıca veriler bireylerin, işletmelerin ve hükümetlerin daha iyi kararlar almasına yardımcı olmak için kullanılabilir. Örneğin, birçok kişi tarafından kullanılan bir sosyal ağın verileri incelenerek, insanların nerede ve ne zaman etkileşime girecekleri tahmin edilebilir. Bu tür çalışmalar, gelecekte çok daha fazla bilginin elde edilmesini ve bu bilginin doğru şekilde kullanılmasını sağlayacaktır.

Metin madenciliği, bir metnin anlamını belirlemek için bir metni inceleme sürecidir¹. Bu süreç, romanlar, makaleler, şiirler, konuşmalar veya denemeler gibi her tür metne uygulanabilmektedir. Bir metni analiz etmek için önce amacını, hedef kitesini ve türünü belirlemek gerekir. Bu unsurlar belirlendikten sonra, metin anlamı açısından analiz edilebilir. Metin madenciliği, bir metni anlamak için değerli bir araçtır ve kişinin okuduğunu anlamasını geliştirmek için kullanılabilir. Okuyucular bir metni analiz ederek içeriğini daha iyi anlayabilir ve ana fikirlerini belirleyebilir². Ayrıca metin analizi, okuyucuların yazarın metni yazma amacını daha iyi anlamalarına yardımcı olabilir. Yazarlar bir metnin yapısını ve anlamını anlayarak kendi düşünce ve fikirlerini daha iyi organize edebilirler. Ayrıca metin analizi, yazarların izleyicilerinin beklentilerini daha iyi anlamalarına yardımcı olabilir. Metin analizi hem okuma hem de yazma için önemli bir araçtır. Bir metnin anlamını anlayarak, okuyucular ve yazarlar içeriğini ve amacını daha iyi anlayabilirler.

Büyük veri çağında, metin madenciliği, büyük miktarda yapılandırılmamış veriden ilgili bilgileri çıkarmak için bir araç olarak giderek daha önemli hale geliyor³. Metin madenciliği, büyük miktarda yapılandırılmamış veriden değerli bilgileri çıkarma yeteneğinden dolayı önemlidir⁴. Büyük veri çağında, işletmeler

1 Sangno Lee, Jaeki Song ve Yongjin Kim, "An Empirical Comparison of Four Text Mining Methods", *Journal of Computer Information Systems*, c. 51, sy. 1, 2010, s. 1-10.

2 Gail Dines, "Gender, Race, and Class in Media: A Text-reader", *Sage*, 2003.

3 Amir Gandomi ve Murtaza Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics", *International Journal of Information Management*, c. 35, sy. 2, 2015, s. 137-144.

4 Andreas Hotho, Andreas Nürnberger ve Gerhard Paaß, "A Brief Survey of Text Mining", *Ldv Forum*, c. 20, s. 1, 2005.

sürekli büyüyen veri hacmini anlamlandırmaya çalıştıkça, bu yetenek giderek daha önemli hale geliyor. Metin madenciliği, analistlerin diğer yöntemlerle tespit edilmesi zor olan eğilimleri ve kalıpları ortaya çıkarmasına yardımcı olabilir ve ayrıca yeni hipotezler oluşturmak veya mevcut hipotezleri test etmek için kullanılabilir.

Bu çalışma da Türkiye’de sosyal medya alanında büyük veri konulu tezler istatistiksel olarak analiz edilerek bu alanda çalışacak kişilere yol gösterici bilgiler sunulması amaçlanmıştır. Çalışma da 2008-2022 yılları arasında Türkiye’de sosyal medya alanında büyük veri konulu tezler metin madenciliği yöntemi kullanılarak incelenmiştir. Metin verilerinin göselleştirmesinde R yazılımının “wordcloud” paketi kullanılmıştır. Çalışmanın giriş bölümünden sonra literatür kısmında sosyal medya alanında büyük veri konusunda yazılmış bazı tezler hakkında, ardından büyük veri ve sosyal medya konusunda bilgiler verilmiştir. Bunu müteakiben metin madenciliği yöntemi ve kuvvet yasası dağılımı hakkında bilgiler verilmiş ve son bölümde analiz sonuçlarından elde edilen bulgular derlenerek kısa bir değerlendirmeye çalışma sonlandırılmıştır.

Seçilmiş Literatür

Sosyal medyadan sağlanan büyük veriye yönelik pekçok konuyu ele alan tezler yazılmıştır. Çalışmanın bu bölümünde bu tezlerden bazıları ele alınarak makalenin amacı çerçevesinde değerlendirilecektir. Bu çalışma için incelenen tezlerin yıllara ve anabilim dallarında göre dağılımları ise EK bölümünde Tablo 1’de verilmiştir.

Muhammed Şen⁵ tarafından bilgisayar bilimleri anabilim dalında hazırlanan “Sosyal Medyadan Elde Edilen Büyük Veri Baz Alınarak Sosyal Olay Analizi Gerçekleme” isimli tezde microblog sitelerden-Twitter’dan elde edilen büyük veri ile etki analizi yapılmış, sosyal medyanın, Arap baharı gösterileri ve Gezi Parkı olaylarından hareketle, örgütlenme ve iletişim için önemli bir etmen olduğundan bahsedilmiştir. 10 Temmuz-11 Ağustos 2014 tarih aralığındaki Türkçe 355 milyon 360 bin 510 tweet veri olarak kullanılmıştır. Daha önce yapılmış çalışmalardan farklı olarak tweetlerin etkisinin aynı düşünceyi paylaşan gruplardan ziyade farklı düşüncelerdeki gruplara etkisinin daha gerçekçi olduğu vurgulanmıştır. Gruplama işlemi için ise gözlemlerin tümünden, hashtag kullanımına bağlı üç grup oluşturulmuştur. Bu bağlamda her bir tweetin etki değeri hesaplanmıştır. Farklı düşüncedeki gruplar tarafından retweet veya yanıtlanma durumuna öncelik tanınarak bu etki değeri ölçülmüştür. Ölçüm amacıyla tek boyutlu Analitik Hiyerarşi Süreci- AHP kullanılmıştır. Yapılan analiz sonrasında etki değeri en yüksek olan tweet ve bu tweeti atan kullanıcının grubu tespit edilmiştir. Örneklem döneminde sosyal medyadaki paylaşım sayısı en yüksek olan görseller (video ve resim) ile bu

5 Muhammed Şen, “Sosyal Medyadan Elde Edilen Büyük Veri Baz Alınarak Sosyal Olay Analizi Gerçekleme”, Yüksek Lisans tezi, Gazi Üniversitesi, Bilgisayar Bilimleri Anabilim Dalı, 2016.

tweetin eşleştiği tespit edilmiştir. Ayrıca her bir grup için etki değerleri ayrı ayrı hesaplanarak, etki değeri yüksek grup da tespit edilmiştir.

Gamze Yıldız Erduran⁶ tarafından işletme anabilim dalında hazırlanan “Online Müşteri Şikayetlerinin Veri Madenciliği ile İncelenmesi” isimli doktora tezinde bankaların ekonomik kalkınmada bir lokomotif olması nedeniyle, onların faaliyetlerinin etkin ve üretken olmasında müşteri memnuniyetinin en temel etmenlerden biri olduğu vurgulanmıştır. Bu amaç doğrultusunda müşteri şikayetleri bu tezin ana veri kaynağını oluşturmuştur. Müşteri şikayet kanallarındaki metinlerin metin ve veri madenciliği yaklaşımı analizinden sağladığı bulgular doğrultusunda bankalar için politika ve strateji önerilerinde bulunulmuştur. Tezde kullanılan veriler JAVA dilinde ECLIPSE kodları ile veriler elde edilmiş ve bu doğrultuda 100.000 müşteri şikayeti incelenmiştir. Şikayet konularının zamanla farklılaşabilme olasılığına dayalı olarak iki ayrı zaman sürecindeki veriler ele alınmıştır. Bu zaman aralıkları 14 Temmuz 2015-14 Aralık 2015 ve 14 Aralık 2015-6 Eylül 2016’dır. Metin madenciliği ile şikayetlerde kullanım sıklığı yüksek olan ortak kelime ve kelime öbekleri tespit edilmiştir. Tespiti yapılan sınıflandırmalara bağlı olarak şikayetler kategorize edilmiş ve en çok şikayete konu olan durumlar belirlenmiştir. Veri ön işleme sürecinden sonra veri madenciliği yaklaşımları kümeleme ve biriktelik analizleri uygulanmıştır. Şikayetlerin kümelenme analizinde şikayet kelimelerine bağlı olarak kümeler oluşturulmuştur. Böylece bankaların şikayetlerin her birini ayrı ayrı incelemeleri yerine kategorilere odaklanarak daha hızlı çözüm sunma olanağı sağlanmıştır. Yapılan analiz sonucunda, her iki zamana ait şikayet verilerinde adı en çok geçen bankaların azalan sırayla Garanti Bankası, Akbank, Yapı Kredi ve Ziraat Bankası olduğu tespit edilmiştir. Ancak bu bulgunun bu bankaların şikayetleri bertaraf etmedeki beceri durumlarının bir göstergesi olmadığı, bunun, söz konusu bankaların müşteri sayılarının çokluğu ve müşterilerine teknolojik hizmetleri daha başarılı bir şekilde sunmalarının bir sonucu olduğu belirtilmiştir. İkinci zaman dilimine ait analiz sonucunda ise birinci zaman diliminden farklı olarak ING bankası yalnızca 5 şikayet metninde bahsedilmiştir. Bunun nedeni olarak da bu bankanın çözüme ulaştırdığı şikayetleri kaldırmasının bir sonucu olduğu vurgulanmıştır. Twitter, Facebook vs gibi ağlardaki ve üçüncü partideki şikayetler için strateji geliştirmenin zor olduğu belirtilerek, kontrol edebilecekleri mekanizmaları artırmaları önerilmiştir. Bankacılık sektörü ürünlerinden “kredi kartı”nın en çok şikayet edilen ürün olduğu tespit edilmiş ve “tüketici hakem heyeti, kredi kartı aidatı, üyelik ücreti”nin ise kelime grubu olarak en fazla şikayet edilen gruplar olduğu ortaya konulmuştur. Şikayete konu olan bu ürün ve bu ürünlere ilişkin uygulamalar için sunulacak çözümler ile müşteri portföylerinin artırılabilceği vurgulanmıştır. Müşteriler şikayet konusu ile ilgili olarak kendi

6 Gamze Yıldız Erduran, “Online Müşteri Şikayetlerinin Veri Madenciliği ile İncelenmesi”, Doktora tezi, Trakya Üniversitesi, İşletme Anabilim Dalı, 2017.

durumlarını en çok “mağdur” ve “haksız” kelimeleri ile tanımlamışlardır. Tavsiye olarak işletmelerin kendi bünyelerinde veri analizi birimleri oluşturarak şikayetleri daha hızlı bir şekilde yönetebilme imkanına erişebilecekleri dile getirilmiştir. Müşteri sadakatinin giderek azalması ve online ortamlarda ve özellikle sosyal medya kanallarındaki işleyişin oldukça hızlı olması nedeniyle zaman periyodu kısa ya da anlık verilere bağlı bilgilerin analizlerine odaklanılmasının işletmelerin başarılarına oldukça katkı sağlayacağı belirtilmiştir.

Onurcan Güden⁷ tarafından iletişim anabilim dalında hazırlanan “Hakikat Sonrası Dönemde Siyasal İletişim Yöntemini Yeniden Düşünmek Üzere Niteliksel Bir Araştırma” isimli çalışmada “Post-truth”, yani hakikat sonrası olarak ifade edilen siyasal iletişimin XXI. yüzyıldaki hem teknik hem de söylemsel bağlamdaki biçimi, niteliksel veri analizi yaklaşımlarından derinlemesine görüşme tekniği ile incelenmiştir. Yapılan görüşmeler sonrasında sosyal ağlardaki yalan haber, filtre baloncğu ve yankı odası olarak tanımlanan etmenlerin hakikate erişimi engelleyici etkenler olduğu ortaya konulmuştur. Sosyal ağların manipülasyon ve dezenformasyon nedeniyle bireyi hakikatten uzak tutacak olumsuz niteliklerine vurgu yapılmıştır. Bu bağlamda dijitalleşmenin siyasal iletişim tekniklerini dönüştürdüğü ve büyük veri aracılığı ile hedefe erişebilecek mikro içerikler üretme olanağına sahip olduğu gözlenmiştir. Verilerden hareketle, siyasi parti ve adayların kampanyalarını kitlelere ihtiyaçları ve beklentileri doğrultusunda içerikler hazırlayarak yürüttükleri görülmüştür. Kişisel verilerden yararlanılarak yürütülen bu süreçte gözetim toplumu sorunun ortaya çıktığı ve suistimal ve manipülasyon yapabilme durumunun da sözkonusu olduğu belirtilmiştir. Doğru olmayan ya da doğruluğu kanıtlanmamış bilginin tespit edilmesi için doğrulama platformlarının bu sürecin olumsuz şartlarını nötrleştirilmesi üzerinde durulmuş ve gelişmişlik düzeyi en yüksek olan toplumlarda bile bu platformlardan haberdar olma düzeyinin yetersiz olduğu tespit edilmiştir. Ayrıca bilginin güvenilirliğini sorgulama ihtiyacına sahip birey sayısının azlığı sebebiyle siyasal iletişimdeki manipülatif etkilerin denetlenemeyeceği ortaya çıkmıştır.

Fikriye Ataman⁸ tarafından istatistik anabilim dalında hazırlanan “Büyük Veri Tabanlı Sistemlerde Duygu Analizine Yönelik Bir Uygulama: Avrupa’da Twitter Kullanıcılarının Mülteci Algısı” isimli doktora tezinde yeni bir duygu analizi modeli tasarlanmıştır. Hadoop ortamında geliştirilen bu yeni model, duygu analizinde sözcüğe dayalı yaklaşımlar ile makine öğrenmesine dayalı yaklaşımların birleşiminden oluşturulmuştur. Bu modelin büyük veri işleme sürecindeki etiketleme

7 Onurcan Güden, “Hakikat Sonrası Dönemde Siyasal İletişim Yönetimini Yeniden Düşünmek Üzere Niteliksel Bir Araştırma”, Yüksek Lisans tezi, İletişim Anabilim Dalı, Galatasaray Üniversitesi, 2020.

8 Fikriye Ataman, “Büyük veri Tabanlı Sistemlerde Duygu Analizine Yönelik Bir Uygulama: Avrupa’da Twitter Kullanıcılarının Mülteci Algısı”, Doktora tezi, Van Yüzüncü Yıl Üniversitesi, İstatistik Anabilim Dalı, 2020.

aşamasında önemli ölçüde bir performans artışı gösterdiği tespit edilmiştir. Ayrıca duygu sınıflaması bağlamında etkin olduğu ve modelin başarımlarının istatistiklerinin geçerliliği tespit edilmiştir.

Tamer Türegün⁹ tarafından iletişim sanatları ve tasarım sanatları anabilim dalında yüksek lisans tezi olarak hazırlanan “Büyük Verinin Yeni Medya Sanatında Kullanımı; Veri Resmi ve Veri Heykeli” isimli çalışmada büyük veri, yeni medya sanatı, büyük verinin yeni medya sanatındaki yerini açıklayan, büyük veriyi yeni medya sanatında kullanan Refik Anadol’un eserleri ele alınmıştır. Bilim ve teknolojideki hızlı ilerlemeler nedeniyle veri resmi ve veri heykeli gibi dijital eserlere ait çalışmalarda hem niteliğin hem de niceliğin hızla artacağı belirtilmiştir. Büyük verilerden yararlanılarak veriler resmedilip heykelleştirilebilmektedir. Verinin görsel sunumu için sanatçının estetik yaklaşımına paralel yapay zeka araçlarından yararlanıldığı ifade edilmiş, bu dijital eserlerin yansıtılması için yüzey seçiminin uygunluğunun önemli olduğu ve sergilenen mekânın ise eserin algı düzeyine etki edeceğinin altı çizilmiştir. Anadol’un çeşitli kurum ve kuruluşlardan edindiği verileri kendi estetik anlayışı ile yazılım veya yapay zeka araçları ile işlediği ve böylece bu büyük verilerden görsel, işitsel ve duysal eserler oluşturduğu vurgulanmıştır.

Seda Perihan Oflas’ın¹⁰, sosyoloji anabilim dalında hazırlanmış olduğu “Büyük Veri Teknolojisi Çağında Kullanıcı Verilerinin Gizlilik ve Mahremiyetine İlişkin Sosyolojik Analiz” isimli yüksek lisans tezinde veri teknolojilerinin tehlikelerine yönelik farkındalık düzeyini ölçmek, bireylerin internet ile etkileşim alışkanlıklarını ve kullanıcılara özgü verilere ilişkin bireylerin tedbir ve kaygı düzeylerini belirlemek amacıyla anket çalışması yapılmıştır. Büyük veri teknolojisine sahip birçok küresel ve yerel şirket, kullanıcılara ilişkin elde ettikleri bilgi düzeyini artırdıkça daha fazla reklam geliri elde etmektedirler. Çünkü bu şirketlerin hedefli reklamcılık için yarattıkları gözetim-endüstri teknolojilerinin, yeni tür üretici olan internet kullanıcılarının ücretsiz çevrim-içi emeklerini sömürdükleri öne sürülmüştür. Ayrıca teknolojik algoritmaların etik, ahlak gibi insani niteliklere sahip olmadığı belirtilmiştir. Bu durumun kullanıcılarda dijital ayak izi olarak tanımlanan kişisel ve davranışsal verilerin kötüye kullanabileceği endişesi yarattığı ve bu endişenin bireyden topluma doğru geçişinin güç eşitsizlikleri doğurduğu vurgulanmıştır. Tezde, aralarında teknolojik uzmanlık düzeyi bakımından farklar olan üç ayrı örneklem oluşturulmuştur. Sosyoloji, felsefe öğrencilerinden oluşan birinci örneklem teknolojik uzmanlık düzeyi düşük gruptur. Bilgisayar mühendisliği bölümü öğrencileri teknolojik uzmanlık düzeyi orta derece olan ikinci gruptur

9 Tamer Türegün, “Büyük Verinin Yeni Medya Sanatında Kullanımı; Veri Resmi ve Veri Heykeli”, Yüksek Lisans tezi, Beykent Üniversitesi, İletişim ve Tasarım Anasanat Dalı, 2020.

10 Seda Perihan Oflas, “Büyük Veri Teknolojisi Çağında Kullanıcı Verilerinin Gizlilik ve Mahremiyetine İlişkin Sosyolojik Analizi”, Yüksek Lisans tezi, Sosyoloji Anabilim Dalı, Hacettepe Üniversitesi, 2020.

ve bir teknoloji firmasının elemanları ise yüksek düzeyde teknolojik uzmanlığa sahip üçüncü gruptur. Elde edilen sonuçlar ise üç kategoriye ayrılmıştır: Veri gizliliğine verilen önem, gizlilik endişeleri ve veri paylaşımına ilişkin farkındalık. Veri gizliliğine verilen önem bağlamında kişisel veri gizliliğine verilen önem arttıkça (azaldıkça) davranışsal veri gizliliğine verilen önem azalmaktadır (artmaktadır). Kişisel veri gizliliğine verilen önem bağlamında birinci ve ikinci grup arasında istatistiksel olarak anlamlı fark olduğu tespit edilmiştir. Davranışsal veri gizliliği önemi bağlamında ise gruplar arasında istatistiksel açıdan anlamlı bir fark olduğu bulgusuna erişilmemiştir. Gizlilik endişesi kategorisinde elde edilen bulgular ise şöyledir: Günde 9 saatten fazla internet kullanımı, hassas ve davranışsal veriler bağlamında bireylerde kaygılanma olasılığını artırmaktadır. Hem hassas verilerine (tubbi bilgiler, etnik bilgi, siyasi eğilim, cinsel kimlik vb) ve hem de davranışsal verilerine yönelik alışkanlıklara etkisi bağlamında günlük internet kullanım sıklığı ve kişisel verilerine özgü alışkanlık kıyaslanmış ve günlük internet kullanım sıklığının daha etkili olduğu bulgusuna ulaşılmıştır. Davranışsal veri kategorisinde bireylerin tutumlarının onların teknolojik uzmanlık düzeyi ile ilişkili olduğu belirtilmiştir. Davranışsal verilere ilişkin kaygı düzeyinin teknolojik uzmanlık bilgisine göre farklılık gösterdiği bulunmuştur. Veri paylaşımında farkındalık kategorisinde ise güvenilir bilgiye erişim becerisi yükseldikçe kişisel verilerini muhafaza etmede tedbir bilincinin giderek arttığı ortaya çıkarılmıştır.

Erdem Başhan¹¹ tarafından radyo-tv-sinema anabilim dalında hazırlanan “Gelişen İletişim Teknolojileriyle Değişen Anlatım Yolları Üzerinden Yeni Medyanın İncelenmesi” isimli yüksek lisans tezinde teknolojik gelişmelerin insan iletişimi ve kültürel yapıya etkisi etik bağlamda ele alınmıştır. Anlatı insan duygu ve düşüncelerinin kalıcı olabilmesi amacıyla kullanılan bir araç olmasının yanı sıra toplumdaki yaşam biçimi, geçmişe ait taşıdıkları ile sosyal yapıya ait bilgileri de nakleden bir araçtır. Bu bağlamda da anlatı ile veri arasındaki olumsuz ilişki örnekler aracılığı ile ortaya koyulmuştur. Anlatının katmansal bir yapısı olduğu vurgulanmış ve gösterge ve sembollerin tarihsel olarak ait oldukları dönemin olgularını temsil ettiği ve bireyin o dönemde yaşadığı toplumsal ve kültürel özelliklerinin bir yansıması olduğu açıklanmış, teknolojik gelişim ile birlikte anlatının da dijitalleştiğine dikkat çekilmiştir. Büyük veri teknolojisine sahip otoriteler, verinin taşıdığı bilgileri analiz ederek tahmin yapma ve strateji geliştirdiği vurgulanan çalışmada yanlış enformasyon içeren bir veriden beslenerek oluşturulan algoritmalar ile hatalı tahminler ve öngürüler yapılabilineceği tehlikesi sözkonusu olduğu, yani veriye verilen öncüllüğün hukuki aksamalara neden olacağı belirtilmektedir. Aynı zamanda teknolojik sistem yaratıcılarının manipülatif davranışları nedeniyle eşitsizlikler yaratabilme tehlikeleri bünyelerinde barındırdıkları ve bunu önlemek amacıyla

11 Erdem Başhan, “Gelişen İletişim Teknolojileriyle Değişen Anlatım Yolları Üzerinden Yeni Medyanın İncelenmesi”, Yüksek Lisans tezi, İstanbul Üniversitesi, Radyo TV Sinema Anabilim Dalı, 2020.

yasal düzenlemelerle algoritmaların denetlenmesinin gerektiği tespit edilmiştir. Aynı zamanda sosyal bilimlerin çeşitli disiplinlerinin, büyük veri teknoloji üretici ve denetçileriyle birlikte çalışılması durumunda doğabilecek olumsuz sonuçların önüne geçme potansiyeline sahip olduğu ifade edilmiştir.

Emre Cihan Ateş¹² tarafından hazırlanan disiplinlerarası adli bilimlerde “Önleyici Kolluk Faaliyetleri Kapsamında Sosyal Medyada Veri Madenciliği: Suçları Önlemek İçin Twitter Analytic’i Kullanma” isimli doktora tezinde dijitalleşmenin artışı ile birlikte suç oranlarının da giderek arttığı vurgulanmıştır. Bu amaçla twitter paylaşımları veri madenciliği yaklaşımı ile analiz edilip suçu engellemeye yönelik politikalar geliştirilmesi hedeflenmiştir. Bu bağlamda Türkiye, İtalya, Fransa ve İspanya jandarma ve polis teşkilatlarının Twitter paylaşımları örneklem olarak ele alınmıştır. Ülkeler karşılaştırılarak sosyal medya kullanım farklılıkları ve bu teşkilatların vatandaşlarına ne düzeyde etki ettikleri belirlenmiştir. Analiz amacıyla tüm ülke grubuna ait tweetler makine çevirisi ile ingilizceye dönüştürülmüştür. Makine çevirisi sonrası makine öğrenmesi yaklaşımına dayalı duygu analizi gerçekleştirilmiştir. Duyguların doğruluk sınaması için kullanılan rassal orman algoritmasına göre % 79.565 F1 skoru elde edilmiştir. Toplum desteği için sosyal ağ paylaşımlarının hem ülke içinde hem de uluslararası düzlemde farklılaştığı hipotezi doğrulanmış, Türkiye jandarma teşkilatının diğer ülkelerden ayrıştığı tespit edilmiştir. Ayrıca bu çalışmada siber zorbalık tespitleri için 19 makine öğrenimi algoritması kullanılmıştır. Light Gradyan Artırma Makine algoritması ile en iyi sonuç elde edilmiştir. Türkçe makine öğrenme ile siber zorbalığın başarılı şekilde ortaya çıkarıldığı ifade edilmiştir. Veri madenciliği yaklaşımlarından keşfedici veri analizi ile sosyal ağlar üzerinde açık kaynak istihbaratının sağlanabileceği doğrulanmıştır. Herhangi bir konu ile ilgili mesajlardan duygu tespiti yapılarak bazen anomilerin de tespit edilebileceği belirtilmiş en fazla negatif mesajın 17:00-24:00 saatleri arasında gönderildiği belirlenmiştir. Tezde bunların yanısıra denetimsiz öğrenme ile büyük veriden anahtar kelime ile ilintili bilgilerin elde edilebileceği vurgulanmıştır.

Onur Erkan¹³ tarafından hazırlanan “Kentsel Değişimin Konum Tabanlı Sosyal Medya Ağları ile İncelenmesi: Taksim Meydanı Örneği” isimli yüksek lisans tezinde Taksim meydanı ile ilgili Foursquare uygulamasında elde edilen veriler ile kentsel değişim ve demografik değişimler analiz edilmiş, Taksimi yayalaştırma projesinin (TYP) etkileri belirlenmeye çalışılmıştır. Foursquare uygulaması yorumlarında meydanın fiziksel çevresinin yeni hali hakkındaki yorumların genellikle

12 Emre Cihan Ateş, “Önleyici Kolluk Faaliyetleri Kapsamında Sosyal Medyada Veri Madenciliği: Suçları Önlemek İçin Twitter Analytic’i Kullanma”, Doktora tezi, Ankara Üniversitesi, Disiplinlerarası Adli Bilimler, 2021.

13 Onur Erkan, “Kentsel Değişimin Konum Tabanlı Sosyal Medya Ağları ile İncelenmesi: Taksim Meydanı Örneği”, Yüksek Lisans tezi, İstanbul Teknik Üniversitesi, Kentsel Tasarım Anabilim Dalı, 2019.

olumsuz olduğu tespit edilmiştir. 2013-2017 yılları için turistler ile ilgili yorumların arttığı, özellikle ortadoğulu turistlere yönelik yaklaşımların olumsuz olduğu gözlenmiştir. TYP nin trafik sorununu çözeceği beklentisi de doğrulanmış, TYP sonrası taksim meydana ile ilgili algı, sembol ve yüklenen anlamlarda değişimler farkedildiği belirtilmiştir.

Sonuç olarak büyük veri, iş ve akademik dünyanın en önemli konularından biri haline geldi. Nasıl karar verdiğimizde, müşterilerle nasıl etkileşimde bulunduğumuzda ve işlerimizi nasıl yürüttüğümüzde devrim yaratma potansiyeline sahiptir. Potansiyeline rağmen, özellikle Türkiye bağlamında, büyük verileri çevreleyen birçok soru var. Bu makale, Türkiye'deki büyük verilerle ilgili tezleri, özellikle potansiyel uygulamalarını ve sonuçlarını incelemeyi amaçlamaktadır.

Büyük Veri

Bilgi işlem, herhangi bir veri, hareket veya işlemin bilgiye dönüştürülmesidir. Bir bilgisayarda, veriler ikili sayılar (1 veya 0) olarak dizilir. Veriler tekil veya çoğul nesnelere olarak kullanılabilir. Ham veriler, verilerin en temel dijital biçimini tanımlayan bir terimdir.¹⁴ Veriler, girdi olarak kullanılan kaynak verilerinden yeni veriler üretmek için CPU (Central Processing Unit) tarafından gerçekleştirilen mantıksal işlemler sonucu elde edilir.¹⁵ Bilgi işlem sistemleri, yalnızca verileri işlerler veya üretirler; aynı zamanda, sisteme giren verilerden yeni veriler üreten süreçler de vardır.

Bir veri kümesinin büyüklüğü, ölçeği, yapısı, saklanma şekli, analiz edilebilirliği ve kullanılabilirliğiyle ilgilidir. Büyük veri yüksek hızda toplanan, saklanan ve analiz edilen büyük veri kümeleridir. Bu büyük veriyi işlemek için geleneksel veri işleme teknikleri yeterli değildir. "Büyük veri" terimi, geleneksel veri işleme uygulamalarının bunlarla başa çıkmak için yetersiz kaldığı büyük veya karmaşık veri kümelerini ifade eder.¹⁶ Büyük veri, veri kümelerinin büyüklüğü, ölçeği, yapısı, saklanma şekli, analiz edilebilirliği ve kullanılabilirliğiyle ilgilidir. Büyük veri, değerli içgörüler ve bilgiler elde etmek için onu yönetmek, işlemek ve analiz etmek için yeni teknolojiler ve yaklaşımlar gerektirir.

Büyük veriler, daha fazla verinin saklanmasına, daha fazla bilginin toplanmasına ve daha fazla bilginin yapılandırılmasına imkan sağlayan verilerdir. Büyük veri veritabanları, web tarayıcıları ve mobil cihazlar gibi farklı kaynaklardan toplanabilir. Bu veri, kuruluşların karar verme süreçlerini iyileştirmelerine, operasyonlarını optimize etmelerine ve rekabet avantajı elde etmelerine yardımcı

14 Vaughan, "What Is Data? - Definition from Whats.Com", July 2019, <https://www.techtarget.com/searchdatamanagement/definition/data> [Erişim Tarihi: 16.04.2022].

15 Svetlin A. Manavski, "CUDA Compatible GPU As an Efficient Hardware Accelerator for AES Cryptography", *2007 IEEE International Conference on Signal Processing and Communications*, 2007, s. 65-68.

16 Mereena Thomas, "A Review Paper on BIG Data", *International Research Journal of Engineering and Technology (IRJET)*, c. 2, sy. 9, 2015.

olma yeteneğinde yatmaktadır. Büyük veri analitiği, kuruluşların yeni eğilimleri, kalıpları ve korelasyonları belirlemesine yardımcı olabilir. Ayrıca dolandırıcılık ve anormalliklerin tespit edilmesine de yardımcı olabilir. Bu veriler, müşteri hizmetlerini iyileştirmek, pazarlama çabalarını hedeflemek ve tüketici davranışını anlamak için kullanılabilir. Tedarik zincirlerinin ve üretim süreçlerinin verimliliğini artırmak için de kullanılabilir. Sağlık sektöründe, hasta bakımını iyileştirmek ve hastalıklar için yeni tedaviler belirlemek için büyük veriler kullanılıyor. Ayrıca büyük veri, işletmelere operasyonlarını iyileştirmeleri ve daha iyi kararlar almaları için yeni fırsatlar sunuyor.

Büyük veri, küresel olarak kullanılan bir veri kaynakları kümesidir ve birçok veri türünden oluşmaktadır. Bir kullanıcının internet tarayıcısında yaptığı aramaları, konumunu, tıklamalarını, harcadığı zamanı, açtığı web sayfalarını ve daha fazlasını içerebilir.¹⁷ Bu veri, yalnızca bir veri kaynağına sahip değildir; aynı zamanda, bu verilerin anlamlı bir şekilde yorumlanmasını sağlayan bir yapıya da sahiptir. Bu yapı, büyük veriyi yönetmek ve anlamak için gereken teknikleri, araçları ve yöntemleri içermektedir. Bu teknikler arasında en yaygın olanları; verileri toplama, saklama, analiz etme ve yorumlamadır.

Büyük veri, çok yeni bir kavram olmasa da yakın zamanlarda popüler hale gelmiştir. Veri toplama ve analizi, yıllar boyunca birçok farklı amaç için yapılmıştır.¹⁸ Örneğin, bir ülkenin ekonomik durumunu değerlendirmek için istatistikler toplanır ve bu verilerin yorumlanması yoluyla kararlar alınır. Aynı şekilde, bir hastanenin hastalarının sağlık durumunu değerlendirmek için veriler toplanır ve bu verilerin yorumlanmasıyla kararlar alınır. Ancak, büyük veri, bu verilerin çok daha büyük bir kütleye sahip olduğu anlamına gelir. Bu veri, yalnızca birkaç yüz kaynaktan toplanan verilerden oluşmaz; aynı zamanda, yıllarca toplanan milyonlarca kaynaktan oluşmaktadır. Bu nedenle, büyük veri setlerinin anlamlı bir şekilde yorumlanması oldukça zordur.

Araştırma ve uygulama düzeyinde yorumlamalarda yaşanan bu zorluğu rağmen büyük verinin popüler hale gelmesinin birçok nedeni vardır. Birincisi, bu verinin toplanması ve yönetilmesi için geliştirilen teknolojilerin yaygınlaşmasıyla birlikte, büyük veri toplama ve yönetme konusunda daha fazla bilgiye sahip olunmuştur¹⁹. İkincisi, bu verinin toplanması ve yönetilmesi için gereken maliyetler düşmüştür.²⁰

17 Che-Hung Liu, Jen Sheng Wang ve Ching-Wei Lin, "The Concepts of Big Data Applied in Personal Knowledge Management", *Journal of Knowledge Management*, 2017.

18 Michael Gallagher, "Data Collection and Analysis", *Researching with Children and Young People: Research Design, Methods and Analysis*, 2009, s. 65-127.

19 Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma ve Sandeep Kaushik "Big Data in Healthcare: Management, Analysis and Future Prospects" *Journal of Big Data* c. 6, sy. 1, 2019, s. 1-25.

20 Knud Erik Skouby ve Anders Henten "Interview with Jørgen Abild Andersen, Chairman of Committee on Digital Economy Policy (CDEP) of OECD, the Organisation for Economic Co-operation and Development", *Journal of NBICT*, c. 1, 2014, s. 139-146.

Son olarak, büyük verinin toplanması ve yönetilmesi için gereken birtakım yeni teknikler ve yöntemler geliştirilmiştir.²¹ Büyük veri, genellikle birkaç terabayt bulan, yüksek hızda toplanan, yapılandırılmamış veya yapılandırılmış verilerden oluşur. Yalnızca büyük şirketlerin değil, aynı zamanda küçük ve orta ölçekli şirketlerin de kullanabileceği bir veri kaynağıdır. Bu şirketler, bu veri setlerini kullanarak, müşterilerinin ihtiyaçlarını daha iyi anlayabilir, davranışlarını daha iyi tahmin edebilir ve onlara daha iyi hizmet sunabilirler. Büyük verinin kullanımı, birçok farklı alanda yaygınlaşmıştır²². Örneğin, sağlık alanında hastalıkların tedavisinde; eğitim alanının da eğitim süreçlerinin iyileştirilmesinde kullanılmaktadır.

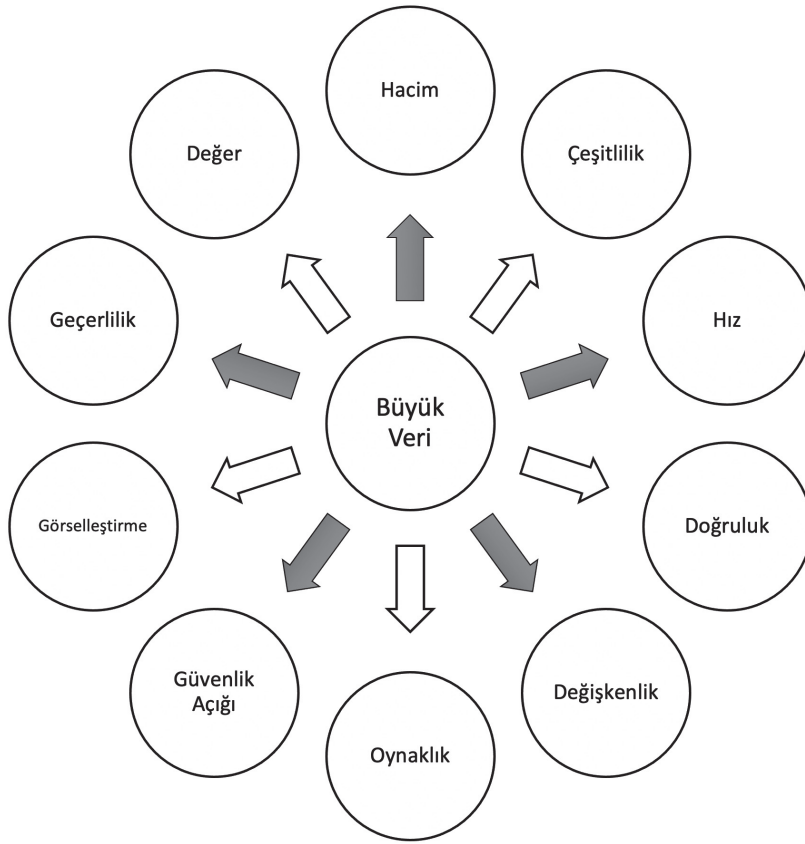
Şekil 1 'de gösterildiği gibi Büyük Veri'nin hacim (volume); çeşitlilik (variety), hız (velocity), doğruluk (veracity), değişkenlik (variability), oynaklık (volatility), güvenlik açığı (vulnerability), görselleştirme (visualization), geçerlilik (validity) ve değer (value) ve olmak üzere on karakteristik özelliği vardır.²³

Hacim: Günlük olarak üretilen büyük miktardaki veriyi ifade eder. Bu veriler, sosyal medya, sensörler ve işlem verileri dahil olmak üzere çeşitli kaynaklardan gelebilmektedir. Veri hacmi o kadar büyük olabilir ki geleneksel yöntemlerle depolanması ve yönetilmesi zor olabilir.

Çeşitlilik: Analiz için mevcut olan farklı veri türlerini ifade eder.²⁴ Büyük Veri, sosyal medya verilerinden işlem verilerine ve sensör verilerine kadar her şeyi içerebilir.²⁵ Veri kaynaklarının ve türlerinin çeşitliliği, bu veri analizini çok karmaşık ama aynı zamanda çok güçlü hale getirebilmektedir.

Hız: Büyük verilerin hız özelliği, verilerin üretilme, toplanma ve işleme hızını ifade eder. Hız, bu verilerin önemli bir özelliğidir çünkü kuruluşların gerçek zamanlı olarak daha iyi kararlar almasına yardımcı olabilmektedir. Kuruluşlar veriyi anlık analiz ederek hızlı kararlar alabilmekte ve bu sayede büyük kazançlar elde edebilmektedir.

-
- 21 Yuji Roh, Geon Heo, and Steven Euijong Whang, "A Survey on Data Collection for Machine Learning: A Big Data-ai Integration Perspective", *IEEE Transactions on Knowledge and Data Engineering*, c. 33, sy. 4, 2019, s. 1328-1347.
- 22 Laika Satish ve Norazah Yusof, "A Review: Big Data Analytics for Enhanced Customer Experiences with Crowd Sourcing", *Procedia Computer Science*, c. 116, 2017, s. 274-283.
- 23 Gunasekaran Manogaran, Chandu Thota ve Daphne Lopez, "Human-Computer Interaction With Big Data Analytics", *Research Anthology on Big Data Analytics, Architectures, and Applications* içinde, Hershey, United States: IGI Global, 2022, s. 1578-96; Nagham Saeed ve Laden Husamaldin, "Big Data Characteristics (V's) in Industry", *Iraqi Journal of Industrial Research*, c. 8, sy. 1, 2021, s. 1-9; Jayanthi Ranjan, "The 10 Vs of Big Data Framework in the Context of 5 Industry Verticals", *PRODUCTIVITY*, c. 59, sy. 4, 2019, s. 324-42.
- 24 Suhail Sami Owais ve Nada Sael Hussein, "Extract Five Categories CPIVW from the 9V's Characteristics of the Big Data", *International Journal of Advanced Computer Science and Applications*, c. 7, sy. 3, 2016.
- 25 Necmi Gürsakal ve Sadullah Çelik, "Büyük Veri ve Pazarlama", 1. baskı, Bursa: Dora Yayınevi, 2021, s. 21.



Şekil 1. Büyük Verinin Karakteristik Özellikleri

Doğruluk: Büyük verilerin doğruluk özelliği, verilerin doğruluğunu ve güvenilirliğini ifade eder.²⁶ Bu veriler, sensörler, sosyal medya ve işlem verileri dahil olmak üzere çeşitli kaynaklardan gelebilir. Verilerin karar vermede kullanılabilmesi için doğru ve güvenilir bir kaynaktan elde edilmiş olması oldukça önemlidir.

Değişkenlik: Büyük verinin yapı ve içerik açısından ne derece değişebileceğini ifade eder. Bu verinin değişkenliği, büyüklüğüne, karmaşıklığına ve üretildiği kaynakların çeşitliliğine bağlı olabilmektedir.²⁷

Oynaklık: Büyük verinin oynaklık özelliği, sürekli değişmesi ve büyümesi anlamına gelmektedir. Oynaklığın yüksek olması verilerin zaman içerisinde çok

26 George Firican, "The 10 Vs of Big Data. Transforming Data with Intelligence", February 18, 2017, <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx> [Erişim Tarihi: 10.04.2022].

27 Raghav Toshniwal, Ghosh Kanishka Dastidar ve Asoke Nath, "Big Data Security Issues and Challenges" *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, c. 2, sy. 2, 2015, s. 2349-2163.

fazla değişmesine sebep olur. Oynaklık, bu verilerin yönetilmesini ve depolanmasını zorlaştırmaktadır.

Güvenlik Açığı: Büyük verinin güvenlik açığı özelliği, büyük miktarda veriyi korumanın ve güvence altına almanın zor olabileceği gerçeğini ifade eder. Bunun nedeni, birden fazla sistem ve konuma yayıldığına verileri izlemenin ve yönetmenin zor olabileceğidir. Ayrıca neyin hassas bilgi olup neyin olmadığını belirlemek daha zor olabileceğinden, büyük verilerin korunması geleneksel veri kümelerinden daha zor olabilir.

Görselleştirme: Büyük verilerin görselleştirmesi, verilerde gizli olan kalıpların görülmesini sağlar. Görselleştirme, bu veri kümelerinin analizine izin veren araçlar kullanılarak yapılır²⁸. Bu araçlar, kalıpları ve eğilimleri görme, aykırı değerleri belirleme, farklı veri kümelerini karşılaştırma ve uzaklaştırma, veri kümelerini belirli kriterlere göre filtreleme ve veri kümeleriyle gerçek zamanlı olarak etkileşim kurma yeteneğini içerebilmektedir.

Geçerlilik: Büyük verilerin geçerlilik özelliği, doğru ve gerçek dünyayı temsil etmesidir. Geçerlilik, gelecek hakkında güvenilir tahminler yapmak için bu verilerin kullanılabilirliği anlamına gelir. Bununla birlikte, genel olarak, büyük verinin, daha büyük boyutu ve daha çeşitli bileşimi nedeniyle geleneksel veri kümelerinden daha geçerli olduğu söylenir. Bunun nedeni, bu veri kümelerinin önyargılı olma veya sonuçları çarpıtabilecek hatalar içerme olasılığının daha düşük olmasıdır.²⁹

Değer: Büyük verinin değer özelliği, daha önce mevcut olmayan içgörüler sağlama yeteneğidir. Bu veriler, geleneksel veri kümelerinde gizlenecek olan eğilimleri, korelasyonları ve kalıpları belirlemek için kullanılabilir. Kuruluşlar, büyük verileri analiz ederek daha iyi kararlar alabilir, operasyonel verimliliği artırabilir ve rekabet avantajı elde edebilirler.

Sonuç olarak verilerin hacmi ve çeşitliliği artmaya devam ettikçe, veriler akademik araştırmalarda giderek daha önemli hale geliyor. Büyük veri, daha önce analiz edilmesi çok zor veya zaman alıcı olan büyük ve karmaşık veri kümelerinin analizine izin verdiği için, araştırmacının yürütülme biçiminde devrim yaratma potansiyeline sahiptir. Akademik araştırmalarda büyük veri kullanımı hala erken aşamalarında, ancak araştırmacının yürütülme şeklini değiştirmeye başladı bile. Gelecekte, veri hacmi ve çeşitliliği artmaya devam ettikçe büyük verinin daha da önemli hale gelmesi muhtemeldir.

28 Sadullah Çelik, "Büyük Veri", *Gece Kitaplığı*, Ankara, 2018, s. 127.

29 Judith S. Hurwitz, Alan Nugent, Fern Halper ve Marcia Kaufman, "How to Ensure the Validity, Veracity, and Volatility of Big Data", March 26, 2016, <https://www.dummies.com/article/technology/information-technology/data-science/bigdata/how-to-ensure-the-validity-veracity-and-volatility-of-big-data-167357/> [Erişim Tarihi: 10.04.2022].

Sosyal Medya

Sosyal medyanın gücü, dünyanın her yerinden insanları birbirine bağlama ve bilgi ve deneyimlerini paylaşmalarına izin verme yeteneğinde yatmaktadır. Büyük veri, sosyal medya için önemlidir çünkü kullanıcı davranışlarını ve tercihlerini anlamak için kullanılabilir.³⁰ Bu anlayış daha sonra kullanıcı deneyimini geliştirmek için kullanılabilir. Örneğin Facebook, insanların siteyi nasıl kullandığını anlamak ve kullanıcı deneyimini iyileştiren değişiklikler yapmak için bu veriyi kullanıyor.

Sosyal medyalarda bulunan ağlar, kişilerin birbirleriyle etkileşim kurabilecekleri, bilgi alışverişi yapabilecekleri ve paylaşımlarda bulunabilecekleri çevrimiçi ortamlardır. Sosyal ağlar, kullanıcılarının diğer kullanıcılarla bağlantı kurmasına, mesajlaşmasına ve paylaşımlarda bulunmasına olanak tanır.³¹ Bazı sosyal ağlar, kullanıcılarına profiller oluşturma, arkadaş listeleri oluşturma ve blog gibi özellikler sunar. Sosyal ağlar ayrıca, kullanıcıların diğer kullanıcıların profillerini görmesine ve bu profillerdeki bilgileri kullanmasına, kişilerin fotoğraflarını, videolarını, blog yazılarını inceleyebilmesine ve başka kişilerle iletişim kurmalarına ve paylaşımlarda bulunmalarına imkan sağlar.

Büyük veri, sosyal ağların önemli bir parçasıdır. Çünkü sosyal ağlar, insanların daha fazla bilgiye ulaşmasını sağlar. Bu veri, sosyal ağlardan gelen bilgileri işlemeye ve anlamaya yardımcı olur.³² Sosyal ağların daha etkili bir şekilde kullanılmasını sağlar. Sosyal ağların önemini artırarak, insanların daha fazla bilgiye ulaşmalarını sağlar. Büyük verinin sosyal ağlar için önemi, reklamların ve içeriğin daha etkili hedeflenmesine olanak sağlamasıdır. Kullanıcılar hakkında toplanan çok miktarda veri ile sosyal ağlar, her bireye uyarlanmış daha özelleştirilmiş deneyimler sunabilmektedir. Bu, kullanıcıların kendileriyle alakalı reklamları ve içeriği görme olasılıklarının daha yüksek olmasına neden olur ve bu da daha fazla etkileşim ve dönüşüm sağlar.

Büyük veri kümelerini sosyal ağlar olarak görselleştirmenin birkaç temel nedeni vardır.³³ Birincisi, sosyal ağ görselleştirmeleri, verilerdeki gizli kalıpları ve eğilimleri ortaya çıkarabilir. Örneğin, büyük bir veri kümesinin bir sosyal ağ görselleştirmesi, birbirine yakından bağlı varlık kümelerini ortaya çıkarabilir. Bu tür bilgileri, verilerin geleneksel bir tablo veya grafik temsilinden ayırt etmek zor

30 Vasile-Daniel Păvăloaia, Elena-Mădălina Teodor, Doina Fotache ve Magdalena Danileț, "Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences", *Sustainability*, c. 11, sy. 16, 2019, s. 4459.

31 Chei Sian Lee ve Long Ma, "News Sharing in Social Media: The Effect of Gratifications and Prior Experience", *Computers in Human Behavior*, c. 28, sy. 2, 2012, s. 331-39.

32 Doyd Danah ve Kate Crawford, "Critical Questions for Big Data", *Information, Communication & Society*, c. 15, sy. 5, 2012, s. 662-79.

33 Jeffrey Heer ve Danah Boyd "Vizster: Visualizing Online Social Networks", *IEEE Symposium on Information Visualization* içinde, Minneapolis, MN, USA: IEEE, 2005, s. 32-39.

veya rızasına gerek olmaksızın üçüncü taraf uygulamalar tarafından verilerine erişilebileceğinden endişe duymaktadır³⁴. Bu endişeler, sosyal medya şirketlerine olan güvenin azalmasına neden olmaktadır. Diğer taraftan bazı şirketlerin kullanıcı verilerini toplayıp sattığı doğru olsa da, veriler genellikle anonimleştirilir ve reklamları hedeflemek veya ürün satmak için kullanılmaz. Ayrıca, çoğu kullanıcı verilerinin nasıl toplandığının veya kullanıldığının farkında değildir ve bu nedenle bu konuda endişe duymazlar.³⁵

Sonuç olarak büyük veri ve metin madenciliği için sosyal medyanın önemi iki nedenden dolayı önemlidir. Birincisi, sosyal medya, büyük veri ve metin madenciliği uygulamaları için kullanılabilir zengin bir veri sağlar. Bu veriler, sosyal eğilimleri anlamak, tüketici tercihlerini belirlemek ve gelecekteki olayları tahmin etmek için kullanılabilir. İkincisi, sosyal medya, işletmelerin müşterileriyle etkileşime girmesi ve geri bildirim toplaması için bir platform sağlar. Bu geri bildirim, ürün ve hizmetleri geliştirmek ve müşteri ihtiyaç ve isteklerini daha iyi anlamak için kullanılabilir.

Metodoloji ve Veri Seti

Veri Seti

Bu çalışmada kullanılan veriler YÖK Ulusal Tez Merkezi'nden çekilmiştir. Çalışmada 2008 yılından 2022 yılına kadar sosyal medya alanında büyük veri konulu yazılmış 72 tez incelenmiştir. Bu tezlerdeki kelimeler metin madenciliği yöntemi kullanılarak literatür incelemesi yapılmıştır.

Metin Madenciliği

Metin madenciliği, bir veri kümesindeki anlamlı ve önemli bilgileri keşfetmeye yönelik bir yöntemdir. Bu yöntem, genellikle bir veri kümesinin yapısını ve içeriğini anlamaya yöneliktir. Veri kümelerinin büyüklüğü ve karmaşıklığı, metin madenciliğinin ne kadar zor olduğunu göstermektedir.³⁶ Veri kümelerindeki bilgilerin anlamlı şekilde keşfedilmesi için, çeşitli yöntemlerin geliştirilmesi gerekmektedir. Metin madenciliği, metin verilerinden bilgi çıkarma işlemidir. Bu süreç, web sitelerinden, kitaplardan ve diğer kaynaklardan gelen metinler gibi yapılandırılmamış verileri çıkarmak ve bunları çeşitli amaçlarla analiz edilebilecek ve kullanılabilir yapılandırılmış bir biçime dönüştürmek için kullanılabilir.³⁷

34 Luci Pangrazio ve Neil Selwyn, "It's not Like It's Life or Death or Whatever": Young People's Understandings of Social Media Data", *Social Media+ Society*, c. 4 sy. 3, 2018.

35 Philip M. Napoli, "Special Issue Introduction: Big Data and Media Management", *International Journal on Media Management*, c. 18, sy. 1, 2016, s. 1-7.

36 IBM, "What Is Text Mining?", November 16, 2020, <https://www.ibm.com/cloud/learn/text-mining> [Erişim Tarihi: 15.04.2022].

37 Han Jiawei ve K.C.-C. Chang, "Data Mining for Web Intelligence", *Computer*, c. 35, sy. 11, 2002, s. 64-70.

Metin madenciliği, metin verilerinden değerli içgörüler elde etmek için kullanılabilir bir araçtır. Duygu analizi, konu modelleme ve metin sınıflandırma gibi çeşitli amaçlar için kullanılabilir.³⁸ Duygu analizi, metin verilerinden duyguları veya görüşleri çıkarma işlemidir.³⁹ Bu, örneğin müşteri duyarlılığını analiz etmek için kullanılabilir. Konu modelleme, metin verilerinden konu çıkarma işlemidir. Bu, bir belgenin veya belge grubunun ana konularını belirlemek için kullanılabilir. Metin sınıflandırma, metin verilerine bir sınıf etiketi atama işlemidir.⁴⁰ Bu, belgeleri otomatik olarak farklı kategorilere ayırmak için kullanılabilir.

Metin madenciliği nispeten yeni bir alandır ve yeni yöntemler geliştirmek ve mevcut olanları iyileştirmek için hala çok sayıda araştırma yapılıyor. Bununla birlikte, metin madenciliğinin potansiyel uygulamaları çok geniş ve halihazırda pazarlama, sosyal medya analizi ve müşteri hizmetleri gibi çeşitli alanlarda kullanılmaktadır.⁴¹

Kuvvet Yasası Dağılımı

Kuvvet yasası, bir sözcüğün bir dil içindeki ortalama sıklığı ile o sözcüğün görülme sıklığı arasındaki ilişkiyi açıklar. Sözcüklerin görülme sıklıkları, sözcüklerin dil içindeki yaygınlığını gösterir. Bu yasa, bir dilin kurallarının, sözcüklerin sıklığına göre değişebileceğini gösterir.⁴² Kuvvet yasası dağılımının kelime frekansları, tahmin edilebilir bir model izler. En yaygın kelimeler metinde en sık geçen kelimelerdir, daha az kullanılan kelimeler ise daha az sıklıkla geçen kelimelerdir. Bu model Zipf yasası olarak bilinir ve bir metindeki kelimelerin sıklığını tahmin etmek için kullanılabilir. Zipf yasası, adını ilk kez 1930'larda öneren dilbilimci George Zipf'ten almıştır. Zipf yasası, bir kelimenin sıklığının metindeki sıralamasıyla ters orantılı olduğunu belirtir.⁴³ Başka bir deyişle, en yaygın kelime, ikinci en yaygın kelimedenden iki kat ve en yaygın üçüncü kelimedenden üç kat daha sık geçecektir. Bu model, kelime frekanslarının kuvvet yasası dağılımında görülebilir. Metindeki en

38 Dandan Tao, Pengkun Yang ve Hao Feng, "Utilization of Text Mining as a Big Data Analysis Tool for Food Science and Nutrition", *Comprehensive Reviews in Food Science and Food Safety*, c. 19, sy. 2, 2020, s. 875-94.

39 Shahnawaz ve Parmanand Astya, "Sentiment Analysis: Approaches and Open Issues", *2017 International Conference on Computing, Communication and Automation (ICCCA)* içinde, Greater Noida, India: IEEE, 2017, s. 154-58.

40 Vikas K Vijayan, K. R. Bindu ve Latha Parameswaran, "A Comprehensive Study of Text Classification Algorithms", *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* içinde, Udupi, India: IEEE, 2017, s. 1109-13.

41 Sadullah Çelik, "Metin Madenciliği ile Shakespeare Külliyyatının İncelenmesi", *MANAS Sosyal Araştırmalar Dergisi*, c. 9, sy. 3, 2020, s. 1344.

42 Mark EJ Newman, "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary physics*, c. 46, sy.5, 2005, s.323-351.

43 Steven T. Piantadosi, "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions", *Psychonomic Bulletin & Review*, c. 21, sy. 5, 2014, s. 1112-30.

yaygın kelimeler en sık geçen kelimelerdir, daha az kullanılan kelimeler ise daha az sıklıkla geçen kelimelerdir. Bu model, kuvvet yasası dağılımı olarak bilinir. Denklem (1) de kesikli kuvvet yasası formülü verilmiştir.⁴⁴

$$p(x) = Cx^{-\alpha} \quad (1)$$

Denklem (1)'deki α ve C parametreleri maksimum olabilirlik yöntemi ile tahmin edilir. Kolmogorov Smirnov istatistiği minimize edilerek x_{min} (kesmeli değer) tahmin edilir.

Kelime frekanslarının kuvvet yasası dağılımının dil öğrenimi için önemli etkileri vardır. Bir dildeki en önemli kelimelerin en sık meydana gelen kelimeler olduğunu öne sürer.⁴⁵ Bu kelimeler yüksek frekanslı kelimeler olarak bilinir ve öğrencilerin yeni bir dil öğrenirken odaklanması gereken kelimelerdir. Kelime frekanslarının kuvvet yasası dağılımının da bilgi alımı için önemli etkileri vardır. Belirli bir sorgu için en alakalı belgelerin, sorgu terimlerini en sık içerenler olduğunu önerir. Bu ilke, frekans-ters döküman frekansı (Term Frequency-Inverse Document Frequency-TF-IDF) algoritması terimi olarak bilinir ve arama motorları tarafından belgeleri sıralamak için kullanılır.⁴⁶

Kelime frekanslarının kuvvet yasası dağılımı, dilbilimciler, dil öğrenenler ve bilgi alma uzmanları için değerli bir araçtır. Dilin yapısını ve yüksek frekanslı kelimelerin önemini daha iyi anlamamıza yardımcı olabilir. Bu amaçla çalışmada metin madenciliği yöntemi ile edile kelimelerin frekanslarının kuvvet yasasına uygunluğu Kolmogorov-Smirnov testi ile araştırılmıştır.

Bulgular

Bu çalışmada analizler R yazılımının “wordcloud” paketi kullanılarak yapılmıştır. “wordcloud” paketi, belirli bir metindeki en yaygın kelimelerin grafiksel temsilleri olan kelime bulutlarını oluşturmak için kullanılır. Şekil 3'te 2008-2022 dönemleri arasında büyük veri konulu 72 teze ait anahtar kelimelerin kelime bulutu verilmiştir.

44 Aaron Clauset, Cosma Rohilla Shalizi ve M. E. J. Newman, “Power-Law Distributions in Empirical Data”, *SIAM Review*, c. 51, sy. 4, 2009, s. 661-703.

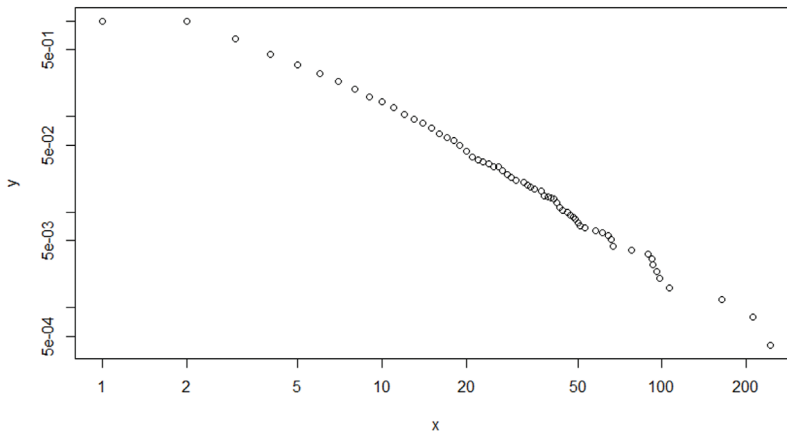
45 Álvaro Corral, Gemma Boleda ve Ramon Ferrer-i-Cancho, “Zipf's Law for Word Frequencies: Word Forms Versus Lemmas in Long Texts”, *PLOS ONE*, c. 10, sy. 7, 2015.

46 Yanchang Zaho ve Yonghua Cen, *Data Mining Applications with R*, Elsevier, 2014.



Şekil 4. Tezlerin Kelime Bulutu

Şekil 3 dikkate alındığında tezlerde en çok “veri”, “büyük”, “Sosyal”, “medya”, “analiz”, “tahmin” ve “iletişim” kelimeleri en çok kullanılmıştır. Şekil 4’te geçen kelimelerin frekans dağılımları Şekil 5’te verilmiştir.



Şekil 5. Kelimelerin Frekans Dağılımı

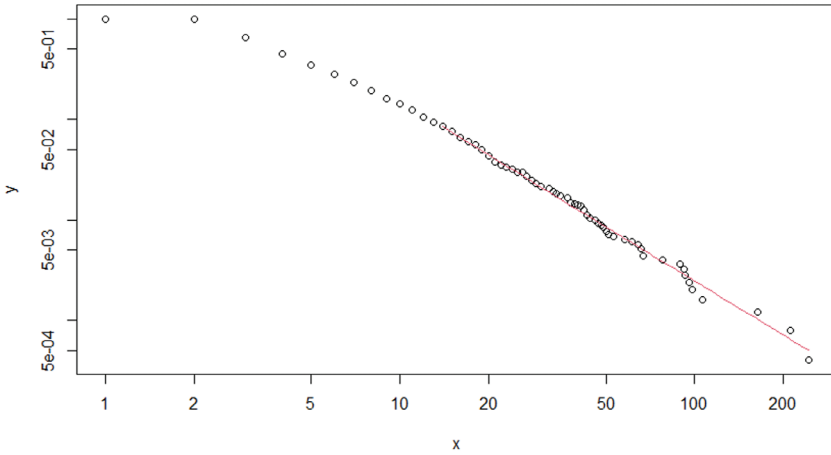
Şekil 5 incelendiğinde grafiğin sol yukarıdan sağ aşağıya doğru azalan bir eğri şeklinde olduğu görülmektedir. Yani makalelerde az sayıda kelime yüksek frekanslı, çok sayıda kelime ise düşük frekanslıdır. Bu görsel biçim kelime dağılımlarının kuvvet yasasına uygun olabileceğine işaret etmektedir. Tabi ki bunun bir test ile sınanması gerekmektedir

Çalışmada verilerin kuvvet yasasına uygunluğunu test etmek için R yazılımının “powerLaw” paketi kullanılmıştır. “powerLaw” bir kuvvet yasası dağılımının üssünü tahmin etme, bir veri kümesinin bir kuvvet yasası dağılımını takip edip etmediğini test etme ve sentetik kuvvet kanunu verileri üretme fonksiyonlarını içerir. Verilerin kuvvet yasasına uygunluğunu test etmek için Kolmogorov-Smirnov Uyum İyiliği Testi kullanılmıştır.

H_0 : Verilerin dağılımı kuvvet yasasına uygundur.

H_1 : Verilerin dağılımı kuvvet yasasına uygun değildir.

Şekil 6’da verilerin Kümülatif Yoğunluk Fonksiyonu grafiği verilmiştir.



Şekil 6. Kümülatif Yoğunluk Fonksiyonu

Şekil 6 incelendiğinde grafiğin sol yukarıdan sağ aşağıya doğru bir eğri şeklinde tanımlandığı görülmektedir. Bu grafik yüksek olasılıkla kelimelerin frekans dağılımlarının kuvvet yasasına uyabileceğini göstermektedir. Fakat yine de bu yeterli bir kanıt değildir. Bunun bir teste tabi tutulması gerekir. Tablo 1 de Kolmogorov-Smirnov test sonuçları verilmiştir.

Tahminci	C	x_{min}	α	n_{tail}	D(KS)	bs_p
Tahmin Sonuçları	1.254	14	2.772	211	0.032	0.4

Tablo 1. Kolmogorov-Smirnov Test Sonuçları

Tablo 1’de Kolmogorov-Smirnov testi sonucunda elde edilen istatistikler verilmiştir. Test sonuçlarına göre $p = 0,4$ bulunmuştur. Anlamlılık seviyesi 0,05 seçilirse $p = 0.4 > 0,05$ olduğundan H_0 temel hipotezi red edilemez. Dolayısıyla verilerin dağılımının kuvvet yasasına uygun olduğu sonucuna varılır.

Analiz sonucunda, yapılan tezlerde genellikle belli kelimelerin etrafında çok fazla tekrarin olduğu sonucuna varılmıştır. Bu nedenle tezlerde aşırı odaklanma sorunu olduğu söylenebilir. Aşırı odaklanma, bir bilimsel çalışmada bir konuyu çok derinlemesine ele alma durumudur. Bu, çalışmanın bir kısmının diğer kısımlarının önemini gölgede bırakabilecek kadar çok detaya girmesi demektir. Bu, makalenin okunabilirliğini ve anlaşılabilirliğini zorlaştırabilir.

Sonuç ve Değerlendirme

“Büyük veri” olarak adlandırılan büyük kütleleri içeren veriler, birçok farklı kaynaktan gelen ve hızla büyüyen verilerdir. Bu veriler, yapılacak işleri ve ürünleri daha iyi anlamak için kullanılabilir. “Büyük veri” çalışmaları, Türkiye deki akademisyenlerin verileri toplama, analiz etme ve yorumlama yeteneklerini geliştirmelerine yardımcı olabilmektedir. Ayrıca, ülkedeki bilimsel araştırmaların “büyük veri” kullanılarak yapılması, Türkiye’nin daha fazla bilgiye sahip olmasını ve daha etkili çözümler üretmesini sağlayabilir. Ayrıca, “büyük veri” sayesinde, ülkeler arasında bilgi alışverişinin de kolaylaşacağı düşünülmektedir. Büyük veri çalışmaları, yalnızca bilimsel araştırmaları değil, aynı zamanda ekonomik ve sosyal alanlardaki çalışmaları da etkileyebilir. Büyük veri çalışmaları, Türkiye’nin daha etkin bir şekilde çalışmalar yapmasına ve daha verimli çözümler üretmesine yardımcı olabilir.

Büyük veri, geleneksel yöntemlerle işlenmesi zor olan büyük miktarda veriyi tanımlamak için kullanılan bir terimdir.⁴⁷ Büyük veri genellikle yapılandırılmamıştır ve geleneksel yöntemlerle madencilik yapmak zor olabilir. Ancak büyük veri, metin madenciliği için benzersiz bir fırsat sunuyor. Metin madenciliği, faydalı bilgileri çıkarmak için büyük verileri işlemek ve analiz etmek için kullanılabilir. Metin madenciliği, bulunması zor olacak verilerdeki kalıpları ve eğilimleri belirlemek için kullanılabilir. Büyük veriler genellikle yapılandırılmamış olduğu için

47 SAS, “Big Data: What It Is and Why It Matters | SAS”, April 30, 2022, https://www.sas.com/en_us/insights/big-data/what-is-big-data.html [Erişim Tarihi: 13.04.2022].

geleneksel yöntemlerle işlenmesi zordur. Bununla birlikte, metin madenciliği, büyük verilerden faydalı bilgiler çıkarmak için kullanılabilir. Metin madenciliği yaklaşımına göre tezlerde en sık geçen kelimeler şöyledir: “veri”, “büyük”, “sosyal”, “medya”, “analiz”, “tahmin” ve “iletişim”dir. Metin madenciliği analizi sonucunda kelime frekansları dikkate alındığında tezlerde en çok “Veri Madenciliği”, “Metin Madenciliği”, “Duygu Analizi”, “Makine Öğrenmesi”, “Derin Öğrenme”, “Sosyal Ağ Analizi”, “Veri Analitiği” ve “Doğal Dil İşleme” tekniklerinin kullanıldığı tespit edilmiştir.

Kolmogrov-Smirnov tekniğine göre ise verilerin kuvvet yasasına göre dağılmış olduğu belirlenmiştir. Tezlerin yıllara göre dağılımı incelendiğinde 2019 yılları civarında daha fazla tez yazıldığı tespit edilmiştir. Sosyal medyadan elde edilen büyük veriler ile ilgili en çok tez bilgisayar mühendisliği anabilimdalında yazılmıştır. Fakat bu tezlerin bir kısmı yazılıma yönelik teknik boyuttadır. Beklenildiği üzere, iletişim alanındaki çeşitli disiplinler bir arada ele alındığında bilgisayar mühendisliğinden sonra en çok bu alanda tez yazılmıştır. Diğer anabilim dallarında yapılan çalışmalar sosyal medyanın çeşitli boyutlarından elde edilen büyük veri analizine yöneliktir. Suçları önleme ve tespit etme, sosyal medya aracılığı ile anlatının boyutunun değişkenliğine vurgu, sosyal medyaki paylaşımlar bağlamında gizlilik ve mahremiyet boyutu, büyük verinin ruhunu görsele aktaran veri resmi, sosyal medyanın siyasi amaçlı kullanımının çeşitli yönleri, sosyal medya paylaşımlarında etkili öğelerin tespiti gibi pekçok konu incelenerek farkındalık ve bilinç oluşturma eğilimine katkıda bulunulmuştur.

Çalışmanın önemli sonuçlarından birisi de yazılan tezlerde aşırı odaklanma sorununun saptanmış olmasıdır. Aşırı odaklanma sorunu, bir kişinin tek bir konu üzerinde çalışırken aşırı derecede yoğunlaşması ve diğer konuları gözden kaçırmasıdır. Bu sorun, bir kişinin akademik başarısını etkileyebilir ve onların çalışmalarını tamamlamalarını zorlaştırabilir. Odaklanma sorunuyla başa çıkmak için, kişiler tek bir konuda aşırı odaklanmamaya çalışmalı ve diğer konularla da ilgilenmelidir. Kişiler ayrıca, konuları daha küçük parçalara bölerek daha kolay odaklanabilirler. Bir kişinin aşırı odaklanma sorununu düzeltmesi için, öncelikle konuyu düşündüğü veya çalıştığı konuyla ilgili biraz ara vermesi gerekir. Ardından, kişinin konuyu daha iyi anlaması için başka kaynakları kullanması veya konuyla ilgili daha fazla bilgi edinmesi gerekir. Son olarak, kişinin konuyla ilgili daha iyi bir karar vermesi için problemi daha objektif bir şekilde değerlendirmesi gerekir.

Sonuç olarak Türkiye’de akademik alanda büyük veri kullanımı henüz emekleme döneminde. Erken benimseyen birkaç kişi büyük veri analitiğini denemeye başlamış olsa da, akademisyenlerin çoğu hala potansiyelinin farkında değil. Bu öncelikle büyük verinin ne olduğu ve araştırmayı geliştirmek için nasıl kullanılabileceğinin anlaşılmasından kaynaklanmaktadır. Ayrıca, büyük veriyi etkin bir şekilde kullanmak için gereken altyapı ve kalifiye personel eksikliği vardır.

Potansiyeline rağmen, büyük verinin akademide tam olarak kullanılabilmesi için ele alınması gereken birkaç zorluk var. Birincisi, büyük verinin ne olduğu ve nasıl kullanılabileceği konusunda bir anlayış eksikliği var. İkincisi, büyük veriyi etkin bir şekilde kullanmak için gereken altyapı ve kalifiye personel eksikliği. Son olarak, ele alınması gereken etik kaygılar vardır. Bu zorluklara rağmen, büyük veri akademi için büyük umut vaat ediyor. Araştırma yöntemlerini geliştirme ve uzun süredir cevaplanamaz olduğu düşünülen soruları yanıtlamaya yardımcı olma potansiyeline sahiptir.

Kaynakça

- Aaron Clauset, Cosma Rohilla Shalizi ve M. E. J. Newman, "Power-Law Distributions in Empirical Data", *SIAM Review*, 2009, c. 51, sy. 4, s. 661-703.
- Álvaro Corral, Gemma Boleda ve Ramon Ferrer-i-Cancho, "Zipf's Law for Word Frequencies: Word Forms versus Lemmas in Long Texts", *PLOS ONE*, 2015, c. 10, sy. 7.
- Andreas Hotho, Andreas Nürnberger ve Gerhard Paaß, "A Brief Survey of Text Mining", *Ldv Forum*, c. 20, s. 1, 2005.
- Anıl Durmuş Ahmet, "Ekonomi Politik Yaklaşım Çerçevesinde Yeni Medyanın Büyük Veri Üzerinden İncelenmesi", Yüksek Lisans tezi, İstanbul Üniversitesi, Halkla İlişkiler ve Tanıtım Anabilim Dalı, 2019.
- Amir Gandomi ve Murtaza Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics", *International Journal of Information Management*, c. 35, sy. 2, 2015, s. 137-144.
- Ayhan Ünlü, "Habermasçı Kamusal Alanın İmkânı: Sosyal Medya Üzerine Bir İnceleme", Yüksek Lisans tezi, Ankara Üniversitesi, Siyaset Bilimi Ve Kamu Yönetimi Anabilim Dalı, 2020.
- Ayten Aysenur Şentürk, "Coğrafi Konumu Etiketli Sosyal Medya Görselleri ile Kent- sel Kimlik Analizi: Kadıköy Moda Örneği", Yüksek Lisans tezi, İstanbul Teknik Üniversitesi, Kentsel Tasarım Anabilim Dalı, 2019.
- Burcu Karaöz, "Büyük Veri ve İşletme Analitiği: Sosyal Medya ve Duygu Analizi ile Bir Öngörü Modeli", Doktora tezi, İstanbul Üniversitesi, İşletme Anabilim Dalı- Sayısal Yöntemler Bilim Dalı, 2018.
- Che-Hung Liu, Jen Sheng Wang ve Ching-Wei Lin, "The Concepts of Big Data Applied in Personal Knowledge Management", *Journal of Knowledge Management*, 2017.
- Chei Sian Lee ve Long Ma, "News Sharing in Social Media: The Effect of Gratifications and Prior Experience", *Computers in Human Behavior*, 2012, c. 28, sy. 2, s. 331-39.
- Dandan Tao, Pengkun Yang ve Hao Feng, "Utilization of Text Mining as a Big Data Analysis Tool for Food Science and Nutrition", *Comprehensive Reviews in Food Science and Food Safety*, 2020, c. 19, sy. 2, s. 875-94.
- Dilek Şahinci, "Yapay Zekâ ve Reklamcılığın Geleceği", Doktora tezi, İstanbul Üniversitesi, Radyo-Televizyon- Sinema Anabilim Dalı, 2021.

- Doyd Danah ve Kate Crawford, "Critical Questions for Big Data", *Information, Communication & Society*, 2012, c. 15, sy. 5, s. 662-79.
- Duygu Furuncu, "Haber Üretim Pratikleri Bağlamında Veri Gazeteciliği: Türkiye'deki Gazetecilerin Veri Gazeteciliğine Yaklaşımı", Doktora tezi, Maltepe Üniversitesi, İletişim Bilimleri Anabilim Dalı, 2019.
- Elif Tiryaki, "Açık Kaynak İstihbaratında Sosyal Medya Kullanımının Analizi", Yüksek Lisans tezi, Milli Savunma Üniversitesi, İstihbarat Çalışmaları Anabilim Dalı, 2020.
- Emre Cihan Ateş, "Önleyici Kolluk Faaliyetleri Kapsamında Sosyal Medyada Veri Madenciliği: Suçları Önlemek İçin Twitter Analytic'i Kullanma", Doktora tezi, Ankara Üniversitesi, Disiplinlerarası Adli Bilimler, 2021.
- Erdem Başkan, "Gelişen İletişim Teknolojileriyle Değişen Anlatım Yolları Üzerinden Yeni Medyanın İncelenmesi", Yüksek Lisans tezi, İstanbul Üniversitesi, Radyo TV Sinema Anabilim Dalı, 2020.
- Esra Demiroğlu, "Sosyal Medyada Online Tüketici Yorumlarının Performansını Öngörme", Yüksek Lisans tezi, Gebze Teknik Üniversitesi, Strateji Bilimi Anabilim Dalı, 2021.
- Fatih Aykul "Veri Madenciliği Teknikleri ile Hava Yolu Firmalarının Tweetleri Üzerinden Duygu Analizi", Yüksek Lisans tezi, Burdur Mehmet Akif Ersoy Üniversitesi, Yönetim Bilişim Sistemleri Anabilim Dalı, 2019.
- Fikriye Ataman, "Büyük Veri Tabanlı Sistemlerde Duygu Analizine Yönelik Bir Uygulama: Avrupa'daki Twitter Kullanıcılarının Mülteci Algısı", Doktora tezi, Van Yüzüncü Yıl Üniversitesi, İstatistik Anabilim Dalı, 2020.
- Gail Dines, "Gender, Race, and Class in Media: A Text-reader", *Sage*, 2003.
- Gamze Yıldız Erduran, "Online Müşteri Şikayetlerinin Veri Madenciliği ile İncelenmesi", Doktora tezi, Trakya Üniversitesi, İşletme Anabilim Dalı, 2017.
- George Firican, "The 10 Vs of Big Data. Transforming Data with Intelligence", February 18, 2017, <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx> [Erişim Tarihi: 10.04.2022].
- Gökay Yılmaz, "E-Ticaret Sistemlerinde Yapılan Ürün Yorumlarının Metin Madenciliği Uygulaması ile İncelenmesi", Yüksek Lisans tezi, İstanbul Aydın Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, 2021.
- Gözde Bozkurt, "Sosyal Medyada Rasyonel Olmayan Yatırımcı Eğilimlerinin Nitel Tercih Modelleriyle Belirlenmesi: Twitter Örneği", Yüksek Lisans tezi, Marmara Üniversitesi, Ekonometri Anabilim Dalı, 2018.
- Gunasekaran Manogaran, Chandu Thota ve Daphne Lopez, "Human-Computer Interaction With Big Data Analytics", *Research Anthology on Big Data Analytics, Architectures, and Applications* içinde, 2022, s. 1578-96.
- Güneş Çalık, "İnternet Kullanıcılarının Ürün Satın Alma Eğiliminin Veri Madenciliği Algoritmaları ile Tahmin Edilmesi", Yüksek Lisans tezi, Sakarya Üniversitesi, Endüstri Mühendisliği, 2019.
- Han Jiawei ve K.C.-C. Chang, "Data Mining for Web Intelligence", *Computer*, 2002, c. 35, sy. 11, s. 64-70.

- IBM, "What Is Text Mining?", November 16, 2020, <https://www.ibm.com/cloud/learn/text-mining> [Erişim Tarihi: 15.04.2022].
- Jayanthi Ranjan, "The 10 Vs of Big Data Framework in the Context of 5 Industry Verticals", *PRODUCTIVITY*, 2019, c. 59, sy. 4, s. 324-42.
- Jeffrey Heer ve Danah Boyd, "Vizster: Visualizing Online Social Networks", *IEEE Symposium on Information Visualization* içinde, 2005, s. 32-39.
- Judith S. Hurwitz, "Alan Nugent, Fern Halper ve Marcia Kaufman, "How to Ensure the Validity, Veracity, and Volatility of Big Data", March 26, 2016, <https://www.dummies.com/article/technology/information-technology/data-science/big-data/how-to-ensure-the-validity-veracity-and-volatility-of-big-data-167357/> [Erişim Tarihi: 10.04.2022].
- Kaan Toprak, "Metin Madenciliği Yöntemleri Kullanarak İllere Göre Haber Analizi", Yüksek Lisans tezi, Karadeniz Teknik Üniversitesi, İstatistik ve Bilgisayar Bilimleri Anabilim Dalı, 2018.
- Khondoker Zahidul Hossain Choyan, "Güney Çin Denizi Politikaları Üzerine Twitter Verileriyle Duygu Analizi", Yüksek Lisans tezi, Konya Teknik Üniversitesi, Bilgisayar Mühendisliği, 2021.
- Knud Erik Skouby ve Anders Henten "Interview with Jørgen Abild Andersen, Chairman of Committee on Digital Economy Policy (CDEP) of OECD, the Organisation for Economic Co-operation and Development", *Journal of NBICT*, c. 1, 2014, s. 139-146.
- Laika Satish ve Norazah Yusof, "A review: Big Data Analytics for Enhanced Customer Experiences with Crowd Sourcing", *Procedia Computer Science*, c. 116, 2017, s. 274-283.
- Luci Pangrazio ve Neil Selwyn, "'It's not Like It's Life or Death or Whatever': Young People's Understandings of Social Media Data", *Social Media+ Society*, c. 4 sy. 3, 2018.
- Mark EJ Newman, "Power Laws, Pareto Distributions and Zipf's Law", *Contemporary Physics*, c. 46, sy.5, 2005, s.323-351.
- Mereena Thomas, "A Review Paper on BIG Data", *International Research Journal of Engineering and Technology (IRJET)*, 2015, c. 2, sy. 9.
- Michael Gallagher, "Data Collection and Analysis", *Researching with Children and Young People: Research Design, Methods and Analysis*, 2009, s. 65-127.
- Mısra Gül, "Dijital Bir Terim Olarak Sosyal Çöplüğün Tanımlanması ve Tüketici Davranışlarına Etkisi", Yüksek Lisans tezi, Işık Üniversitesi, İşletme Anabilim Dalı, 2018.
- Muhammed Salih Karakaşlı, "Sosyal Medyada Büyük Veri Analizi", Yüksek Lisans tezi, İstanbul Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, 2020.
- Muhammed Şen, "Sosyal Medyadan Elde Edilen Büyük Veri Baz Alınarak Sosyal Olay Analizi Gerçekleme", Yüksek Lisans tezi, Gazi Üniversitesi, Bilgisayar Bilimleri Anabilim Dalı, 2016.

- Murat Taşyürek, “Büyük Verilerin Analizi İçin Coğrafik Ağırlıklı Regresyon Yaklaşımlarının Geliştirilmesi”, Doktora tezi, Erciyes Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, 2020.
- Mustafa Ahmed Mahmood ve Sefer Kurnaz, “Sentiment Analysis in Data of Twitter Using Machine Learning Algorithms”, *International Journal of Computer Science and Mobile Computing*, c. 8, sy. 3, 2019, s. 31-35.
- Naghm Saeed ve Laden Husamaldin, “Big Data Characteristics (V’s) in Industry”, *Iraqi Journal of Industrial Research*, 2021, c. 8, sy. 1, s. 1-9.
- Necmi Gürsakar, ve Sadullah Çelik, *Büyük Veri ve Pazarlama*, 1. baskı, Bursa: Dora Yayınevi, 2021.
- Neşe Gülmez, “Üniversite Kariyer Merkezi Hizmetlerinde Dijital Platform ve Sosyal Medya Kullanımının Etkisi”, Doktora tezi, Marmara Üniversitesi, Yönetim ve Organizasyon Bilim Dalı, 2021.
- Nubar İbrahimzade, “Türkiye’de Halkla İlişkiler Alanında Büyük Verinin Kullanımı”, Yüksek Lisans tezi, İstanbul Ticaret Üniversitesi, Halkla İlişkiler Anabilim Dalı, 2017.
- Nurdan Kayım, “Yapay Zekanın Halkla İlişkiler Uygulamalarındaki Yeri”, Doktora tezi, İstanbul Üniversitesi, Halkla İlişkiler Anabilim Dalı, 2021.
- Oğuz Kuş, “Suriyeli Mültecilerin Sosyal Entegrasyonunda Halkla İlişkilerin Rolü”, Doktora tezi, İstanbul Üniversitesi, Halkla İlişkiler ve Tanıtım Anabilim Dalı, 2019.
- Onur Can Sert, “Dinamik Sosyal Ağlarda Akan ve Çok Boyutlu Veri Üzerinden Analiz ve Tahmin Yapılması”, Doktora tezi, Tobb Ekonomi ve Teknoloji Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, 2020.
- Onur Erkan, “Kentsel Değişimin Konum Tabanlı Sosyal Medya Ağları ile İncelenmesi: Taksim Meydanı Örneği”, Yüksek Lisans tezi, İstanbul Teknik Üniversitesi, Kentsel Tasarım Anabilim Dalı, 2019.
- Onur Ertürk, “Dizi Distopyasından Açık Hava Hapishanesine: Sosyal Skor Uygulaması”, Yüksek Lisans tezi, Akdeniz Üniversitesi, Radyo-Televizyon ve Sinema Ana Bilim Dalı, 2019.
- Onur Seveli, “Yapısal Olmayan Verilerin Büyük Veri Analizi ile İşlenmesi ve Yapısal Olan Verilerle İlişkilendirilmesi Yönelik Bir Platform: Sosyal Medya Temelli Tavsiye Motoru Geliştirme”, Doktora tezi, Süleyman Demirel Üniversitesi, Bilgisayar Mühendisliği, 2016.
- Onurcan Güden, “Hakikat Sonrası Dönemde Siyasal İletişim Yönetimini Yeniden Düşünmek Üzere Niteliksel Bir Araştırma”, Yüksek Lisans tezi, Galatasaray Üniversitesi, İletişim Anabilim Dalı, 2020.
- Orkun Güneş, “Türkiye’de Siyasal Partilerin Seçim Kampanyalarında Büyük Veri Kullanımı Üzerine Niteliksel Bir Araştırma”, Galatasaray Üniversitesi, İletişim Anabilim Dalı-Stratejik İletişim Yönetimi, 2021.
- Philip M. Napoli, “Special Issue Introduction: Big Data and Media Management”, *International Journal on Media Management*, 2016, c. 18, sy. 1, s. 1-7.

- Raghav Toshniwal, Ghosh Kanishka Dastidar ve Asoke Nath, "Big Data Security Issues and Challenges", *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, 2015, c. 2, sy. 2, s. 2349-2163.
- Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma ve Sandeep Kaushik "Big data in Healthcare: Management, Analysis and Future Prospects" *Journal of Big Data* c. 6, sy. 1, 2019, s. 1-25.
- Sadık Ahmet Çavuşoğlu, "Sosyal Medya Platformlarında Veri Gizliliği Politikaları", Yüksek Lisans tezi, Üsküdar Üniversitesi, Yeni Medya ve Gazetecilik Anabilim Dalı, 2020.
- Sadullah Çelik, "Büyük Veri", *Gece Kitaplığı*, Ankara, 2018.
- Sadullah Çelik, "Metin Madenciliği ile Shakespeare Külliyatının İncelenmesi", *MANAS Sosyal Araştırmalar Dergisi*, c. 9, sy. 3, 2020, s. 1343-57.
- Sangno Lee, Jaeki Song ve Yongjin Kim, "An Empirical Comparison of Four Text Mining Methods", *Journal of Computer Information Systems*, c. 51, sy. 1, 2010, s. 1-10.
- SAS, "Big Data: What It Is and Why It Matters | SAS", April 30, 2022, https://www.sas.com/en_us/insights/big-data/what-is-big-data.html [Erişim Tarihi: 13.04.2022].
- Seda Perihan Oflas, "Büyük Veri Teknolojisi Çağında Kullanıcı Verilerinin Gizlilik ve Mahremiyetine İlişkin Sosyolojik Analizi", Yüksek Lisans tezi, Hacettepe Üniversitesi, Sosyoloji Anabilim Dalı, 2020.
- Sedat Öztürk, "Sosyal Medya ve Arama Motorları Trend Karşılaştırması", Yüksek Lisans tezi, Bahçeşehir Üniversitesi, Bilgisayar Mühendisliği, 2019.
- Serbest Ziyanak, "Web Sunucu Loglarının Web Madenciliği Yöntemleri ile Analizi", Yüksek Lisans tezi, Yüzüncü Yıl Üniversitesi, İstatistik Anabilim Dalı, 2017.
- Shahnawaz ve Parmanand Astya, "Sentiment Analysis: Approaches and Open Issues", *2017 International Conference on Computing, Communication and Automation (ICCCA)* içinde, 2017, s. 154-58.
- Sinan Kul, "Türkçe Tweetlerin Analiz Edilebilmesi İçin Hadoop/Hive Kullanan Melez Bir Sistemin Geliştirilmesi", Doktora tezi, Atatürk Üniversitesi, Yönetim Bilişim Anabilim Dalı, 2018.
- Steven T. Piantadosi, "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions", *Psychonomic Bulletin & Review*, 2014, c. 21, sy. 5, s. 1112-30.
- Suhail Sami Owais ve Nada Sael Hussein, "Extract Five Categories CPIVW from the 9V's Characteristics of the Big Data", *International Journal of Advanced Computer Science and Applications*, 2016, c. 7, sy. 3.
- Svetlin A. Manavski, "CUDA Compatible GPU as an Efficient Hardware Accelerator for AES Cryptography", *2007 IEEE International Conference on Signal Processing and Communications*, 2007, s. 65-68.
- Tamer Türegün, "Büyük Verinin Yeni Medya Sanatında Kullanımı; Veri Resmi ve Veri Heykeli", Yüksek Lisans tezi, Beykent Üniversitesi, İletişim ve Tasarım Anasanat Dalı, 2020.

- Tohid Javadi, "Sosyal Ağlarda Yapay Zekâ ile Veri Analizi", Yüksek Lisans tezi, Gazi Üniversitesi, Bilgisayar Mühendisliği Ana Bilim Dalı, 2019.
- Uğur Mertoğlu, "Türkçe İçin Sahte Haber Tespit Modelinin Oluşturulması", Doktora tezi, Hacettepe Üniversitesi, 2020.
- Vasfi Tataroğlu, "Derin Öğrenmeye Dayalı Sosyal Medya Profillemesi", Yüksek Lisans tezi, Pamukkale Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, 2019.
- Vasile-Daniel Păvăloaia, Elena-Mădălina Teodor, Doina Fotache ve Magdalena Danileț, "Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences", *Sustainability*, 2019, c. 11, sy. 16, s. 4459.
- Vaughan, "What Is Data? - Definition from WhatIs.Com", July 2019, <https://www.techtarget.com/searchdatamanagement/definition/data> [Erişim Tarihi: 16.04.2022].
- Vikas K Vijayan, K. R. Bindu ve Latha Parameswaran, "A Comprehensive Study of Text Classification Algorithms", *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* içinde, 2017, s. 1109-13.
- Yanchang Zaho ve Yonghua Cen, *Data Mining Applications with R*, Elsevier, 2014.
- Yuji Roh, Geon Heo, and Steven Euijong Whang, "A Survey on Data Collection for Machine Learning: B Big Bata-ai Integration Perspective", *IEEE Transactions on Knowledge and Data Engineering*, 2019, c. 33, sy. 4, s. 1328-1347.
- Zühal Mollaoğulları, "Uluslararası Çatışmaların Sosyal Ağ Analizi ile İncelenmesi", Yüksek Lisans tezi, Aydın Adnan Menderes Üniversitesi, Uygulamalı Ekonomi Programı, 2020.

EK

Tablo 1: Tezlerin Yıllara ve Anabilim Dallarına Göre Dağılımı

Yıllar	Hazırladığı Bilim Dalı	Lisansüstü Düzey
2008	Bilgisayar Mühendisliği	Yüksek lisans
2013	Bilgi Teknolojileri	Yüksek lisans
2014	Bilgisayar Mühendisliği	Doktora
2014	Bilgisayar Mühendisliği	Yüksek lisans
2014	Bilgisayar Bilimi ve Mühendisliği	Yüksek lisans
2015	Bilgisayar Mühendisliği	Yüksek Lisans
2015	Gazetecilik/Bilişim	Yüksek Lisans
2015	Bilgisayar Mühendisliği	Yüksek Lisans
2016	İletişim Sanat	Yüksek Lisans
2016	Bilgisayar bilimleri	Yüksek Lisans
2016	Doğal ve Uygulamalı Bilim	Yüksek Lisans
2016	Siber Güvenlik Programı	Yüksek Lisans
2016	Bilgisayar Mühendisliği	Doktora
2017	İstatistik	Yüksek Lisans
2017	Halkla İlişkiler	Yüksek Lisans
2017	İşletme	Doktora
2017	Bilgisayar Grafıği	Yüksek Lisans
2017	İşletme	Doktora
2017	İşletme	Yüksek Lisans
2017	Bilgisayar Mühendisliği	Yüksek Lisans
2018	Finansal Ekonometri	Yüksek Lisans
2018	Ekonometri	Yüksek Lisans
2018	Coğrafi Bilgi Teknolojileri	Yüksek Lisans
2018	İktisat	Yüksek Lisans
2018	Bilgisayar Mühendisliği	Yüksek Lisans
2018	Bilgisayar Mühendisliği	Yüksek Lisans
2018	İşletme	Doktora
2018	Yönetim Bilgi Sistemleri	Doktora
2018	İşletme	Yüksek Lisans
2018	Bilgisayar Mühendisliği	Yüksek Lisans
2018	İstatistik ve Bilgisayar Bilimleri	Yüksek Lisans
2018	Yazılım Mühendisliği	Yüksek Lisans
2018	Yeni Medya Disiplin	Yüksek Lisans
2018	Bilgisayar Mühendisliği	Yüksek Lisans

Yıllar	Hazırladığı Bilim Dalı	Lisansüstü Düzey
2018	Yönetim Bilişim	Doktora
2018	İktisat	Yüksek Lisans
2019	İletişim Bilimleri	Doktora
2019	Bilgisayar Mühendisliği	Yüksek Lisans
2019	Bilgi Teknolojileri	Yüksek Lisans
2019	Yönetim Bilişim	Yüksek Lisans
2019	Kentsel Tasarım	Yüksek Lisans
2019	Bilgi Teknolojileri	Yüksek Lisans
2019	Bilgisayar Mühendisliği	Yüksek Lisans
2019	Halkla İlişkiler ve Tanıtım	Doktora
2019	İktisat	Doktora
2019	Yönetim ve Bilgi Sistemleri	Yüksek Lisans
2019	Bilgisayar Mühendisliği	Yüksek Lisans
2019	Bilgisayar Mühendisliği	Yüksek Lisans
2019	Bilgisayar Mühendisliği	Yüksek Lisans
2019	Halkla İlişkiler ve Tanıtım	Yüksek Lisans
2019	Kentsel Tasarım	Yüksek Lisans
2019	Endüstri Mühendisliği	Yüksek Lisans
2019	Bilgisayar Mühendisliği	Yüksek Lisans
2019	Radyo-TV-Sinema	Yüksek Lisans
2019	Radyo-TV-Sinema	Yüksek Lisans
2020	Bilgisayar Mühendisliği	Yüksek Lisans
2020	İstatistik	Doktora
2020	Radyo-TV-Sinema	Yüksek Lisans
2020	İş Analitiği	Yüksek Lisans
2020	İstihbarat Çalışmaları	Yüksek Lisans
2020	Bilgisayar Mühendisliği	Doktora
2020	İletişim Sanatları ve Tasarım Sanat	Yüksek Lisans
2020	Sosyoloji	Yüksek Lisans
2020	Yeni Medya ve Gazetecilik	Yüksek Lisans
2020	Elektrik ve Bilgisayar Mühendisliği	Yüksek Lisans
2020	Bilgisayar Mühendisliği	Doktora
2020	İletişim	Yüksek Lisans
2020	Uygulamalı Ekonometri	Yüksek Lisans
2020	Geomatik Mühendisliği	Doktora
2021	Strateji Bilimi	Yüksek Lisans

Yıllar	Hazırladığı Bilim Dalı	Lisansüstü Düzey
2021	Yönetim ve Organizasyon	Doktora
2021	Elektronik ve Bilgisayar Mühendisliği	Yüksek Lisans
2021	Adli Bilimler	Doktora
2021	İşletme ve Yönetim	Yüksek Lisans
2021	Radyo-TV	Yüksek Lisans
2021	Halkla İlişkiler	Doktora
2021	Bilgisayar Mühendisliği	Yüksek Lisans
2021	Radyo-TV-Sinema	Doktora
2021	Stratejik İletişim	Yüksek Lisans
2021	Bilgisayar Mühendisliği	Yüksek Lisans
2021	Bilgisayar Mühendisliği	Yüksek Lisans
2022	Uluslararası İlişkiler	Yüksek Lisans
2022	İstatistik	Doktora

