

Citation: Aytaç, Z., & Bilge, H.Ş., Türkiye’de Kamu Ve Özel Kurumların Büyük Veri Alanında Mevcut Durum Ve Beklentilerinin Değerlendirilmesi Üzerine Nitel Bir Araştırma, BMIJ, (2020), 8(5): 4646-4679 doi: <http://dx.doi.org/10.15295/bmij.v8i5.1635>

TÜRKİYE’DE KAMU VE ÖZEL KURUMLARIN BÜYÜK VERİ ALANINDA MEVCUT DURUM VE BEKLENTİLERİNİN DEĞERLENDİRİLMESİ ÜZERİNE NİTEL BİR ARAŞTIRMA

Zeynep AYTAÇ ¹

Received Date (Başvuru Tarihi): 25/08/2020

Hasan Şakir BİLGE ²

Accepted Date (Kabul Tarihi): 12/10/2020

Published Date (Yayın Tarihi): 25/12/2020

Çalışmada ilk yazar Sorumlu Yazar (Corresponding Author) rolündedir.

ÖZ

Anahtar Kelimeler:

Büyük Veri,
Büyük Veri ve Eğitim

JEL Kodları:

M150

Bu çalışmanın amacı, Türkiye’de kamu ve özel kurumlarda büyük veri analitiğinin kullanımını inceleyerek, mevcut durumun ve beklentilerinin ne yönde olduğunun değerlendirilmesidir. Bu doğrultuda Türkiye’de en çok tercih edilen büyük veri araçlarını ve bilgi teknolojileri sektörünün öğrencilerden ve üniversitelerden beklentilerini belirlemektir. Araştırma Ankara ilinde büyük veri analitiği çözümleri sunan 10 kamu ve özel kurumun büyük veri ekip liderleriyle derinlemesine mülakat yapılarak gerçekleştirilmiştir. Verilerin analizi için nitel araştırma yöntemlerinden gömülü teori yaklaşımı ve tematik kodlama kullanılmıştır. Elde edilen bulgular, kurumların büyük veri analitiği mevcut durumu, üniversitelerin ilgili bölümlerinden ve öğrencilerden beklentileri ve öğrencilere ve üniversitelere önerileri olmak üzere üç ana başlıkta belirtilmektedir. Kurumların büyük veri analitiği mevcut durumu kategorisinde, büyük veri ekibinde hangi pozisyonların tercih edildiği, bu ekiplerde çalışan sayısı, bu sayının yeterliliği ve kullandıkları teknolojilere yer verilmektedir. Beklentiler kategorisinde, büyük veri alanında hangi bölüm mezunlarının tercih edildiği, mezunlardan yetenek ve teknik beceri anlamında beklentilerin ne yönde olduğu yer almaktadır. Öneriler kısmında ise kurumların, bu alanda çalışmayı düşünen öğrencilere ve bu alanla alakalı olarak üniversitelere olan önerilerine yer verilmektedir.

Keywords: Big Data Big Data and Education

JEL Codes: M150

¹ Öğr. Gör., Aksaray Üniversitesi, zduman@aksaray.edu.tr

<https://orcid.org/0000-0001-8051-3460>

² Prof. Dr., Gazi Üniversitesi, bilge@gazi.edu.tr

<https://orcid.org/0000-0002-4945-0884>

EXTENDED ABSTRACT

A QUALITATIVE RESEARCH ON EVALUATING PUBLIC AND PRIVATE INSTITUTIONS' EXISTING STATUS AND EXPECTATIONS ON BIG DATA AREA IN TURKEY

1. LITERATURE

Big data is defined as "Datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze (Manyika et al., 2011). Big data's 5V characteristics are volume, velocity, variety, veracity, and value (Gahi, Guennoun & Mouftah, 2016).

1.1. RESEARCH SUBJECT

In 2015, there were more than 2 million 350 thousand job postings for data science and analytics jobs in the USA. It is estimated that there will be close to 2 million 720 thousand job postings by 2020 (Miller & Hughes, 2017). Nearly 60 universities have data science programs at the undergraduate level in the USA (DataScience, 2020). In Turkey's Strategy and Budget Presidency's Strategic Plan (2019-2023), "Increasing the capacity of expert/ data scientists experienced in big data analysis and strengthening the infrastructure" is defined as a requirement (Strateji Ve Bütçe Başkanlığı, 2018). Accordingly, in this study, the expectations of public and private institutions from universities and students regarding the big data field were evaluated.

1.2. RESEARCH PURPOSE AND IMPORTANCE

This study aims to examine the use of big data analytics in Turkey's public and private institutions and determine the institutions' expectations. Within the scope, the most preferred big data tools and the expectations the field of big data have towards students and universities in Turkey were determined. This study is essential in terms of being a resource for determining the expectations of institutions from universities and students in big data.

1.3. CONTRIBUTION of the ARTICLE to the LITERATURE

The evaluation of public and private institutions' expectations from universities and students in Turkey is expected to contribute to immense data analytics literature.

2. DESIGN AND METHOD

2.1. RESEARCH TYPE

This study is an applied research and was carried out through exploratory research.

2.2. RESEARCH PROBLEMS

In this study, the following research questions are defined.

What are the big data tools and technologies preferred by public and private institutions that offer big data solutions?

What are the expectations of public and private institutions that provide big data solutions from students studying in computer science and related departments and universities in the field of big data?

2.3. DATA COLLECTION METHOD

The Snowball sampling method was chosen in the interviews. The criterion in sample selection is that the institution offers big data solutions. Meetings were held with the general managers or prominent data project managers of ten institutions. Considering the adoption of big data by sectors, institutions in different sectors have preferred the research. In this study, interviews were conducted with the prominent data department managers of two public and eight private institutions that offer big data analytics solutions in Ankara.

2.4. QUANTITATIVE/QUALITATIVE ANALYSIS

Qualitative research methods, namely the grounded theory approach, was used to analyze the data. The transcription of the voice recordings was collected to analyze the interview data. After the interview transcripts were examined, they were subjected to the coding process using NVivo 12 software.

2.5. RESEARCH MODEL

A qualitative research method was used in this study. The data obtained from the interviews were generalized inductively step by step. In this study, data collection and analysis were carried out together. After the data was collected, the concepts and processes that emerged with the analysis were included in the next data collection stage.

3. FINDINGS AND DISCUSSION

3.1. FINDINGS as a RESULT of ANALYSIS

The findings obtained with the data collected from the institutions were stated under three main parts: the current situation of the big data analytics of the institutions, their expectations from the universities and students, and their suggestions. The current status of the big data analytics category of the organizations covered the positions preferred by the big data team, the number of employees in these teams, the adequacy of this number, and the technologies they use. The category of expectation included the undergraduate programs preferred in big data and the abilities and skills expected from the graduates. The suggestion section included suggestions for students considering working in this field and for universities.

It was defined that a wide variety of position names were used, and it was revealed that the most preferred position name was a software engineer. It has been observed that most of the institutions use Hadoop Ecosystem. Spark, Cassandra, Hive, and Kafka are among the most preferred tools. The most preferred departments by institutions for recruitment are Statistics, Mathematics, and Computer Science. Among the skills expected from graduates, awareness and analytical thinking were the most expressed as the interviewed institutions' technical skills expectations are the database skills (SQL) and necessary programming skills. Among the suggestions of the interviewees to the students are to have awareness, to make an extra effort other than what is taught at school, to work part-time or internship, to have continuity of learning and to be eager to learning, to participate in online courses, to participate in seminars, courses, and events. The interviewees' recommendations for universities include computer science, statistics, mathematics, and management information systems to renew the curriculum, long-term internships, and open more graduate programs.

3.2. DISCUSSING the FINDINGS with the LITERATURE

The U.S. Bureau of Labor Statistics states that positions covering data science and analytics do not have a clear definition. It is stated that positions such as data scientists and data engineering in data science and analytics can not be followed (Miller & Hughes, 2017). It was revealed that the institutions interviewed can not make a clear distinction between positions such as data scientists, data engineers, and software engineers. It was concluded that the tools used by the institutions and the tools used

around the world (HDFS, Apache Hadoop, HBase, Cassandra, Hive, Oracle Big Data ve R) (Bhadani ve Jothimani, 2016) are similar. The institutions' recommendations are to be aware of the development of big data and analytics and follow innovations. The recommendations of the institutions for universities are to renew the curriculum and to open more graduate programs. Lawler and Molluzzo (2015) stated that with a good curriculum design, students studying in computer science and information system departments could contribute to big data analytics.

4. CONCLUSION, RECOMMENDATION, AND LIMITATIONS

4.1. RESULTS of the ARTICLE

It has been concluded that data scientists, data engineers, and data analysts could not be made clearly by institutions, yet when determining employees' positions in big data projects. It has been concluded that they generally prefer the position as a software engineer. Apache Hadoop, Spark, Hive, Kafka, and Cassandra were the most used tools by the interviewed institutions. It has been revealed that the institutions prefer the personnel who work in this field mostly from computer science, then mathematics, statistics, and management information systems. It is concluded that the institutions have suggested having awareness about the big data field for students and curriculum renewal for universities.

4.2. SUGGESTIONS BASED on RESULTS

It is recommended to open more graduate programs and undergraduate programs in interdisciplinary fields such as data science and redesign the relevant departments' curriculum to include big data analytics. The renewal of the curriculum will help institutions make more precise decisions in determining positions and recruitment. Another suggestion is to have more course alternatives in the field of data science in related departments. It is also recommended for universities to add long-term internships to relevant departments.

4.3. LIMITATIONS of the ARTICLE

In this research, interviews were done with the prominent data project team leaders of public and private institutions that offer big data solutions in Ankara. One of the limitations of this study is the limited number of institutions that produce solutions in big data. Another limitation is to find participants in the criteria determined due to the workload of the institutions.

1. GİRİŞ

Büyük veri, geleneksel veri tabanı teknolojileriyle saklanması, işlenmesi ve analiz edilmesi zor olan, veri hacmindeki artışı ifade etmek için kullanılan bir terimdir (Hashem vd, 2014). Uluslararası Veri Kurumu (IDC), dünyadaki verilerin her iki yılda bir, ikiye katlandığını iddia etmektedir. IDC istatistiklerine göre, 2018'den 2025'ye kadar ulaşılacak veri miktarının 33 zetabayt³'dan 175 zetabayt'a kadar artacağı öngörülmektedir (Rydning, 2018). Verinin eksponansiyel olarak büyümesi iş dünyasında dolayısıyla da veri tabanı sistemlerinde, veriye erişme, saklama, işleme konusunda çok büyük değişiklikleri beraberinde getirdiği ve getirmeye devam edeceği açıktır. Bu nedenle iş dünyasında bilgiye erişmek için artık çok farklı araçlar ve yazılımlar kullanılır hale gelmiştir. Geleneksel veritabanı sistemlerinin farklı kaynaklardan gelen, çok hızlı büyüyen ve farklı formatlarda olan bu veriler üzerinde işlem yapma kapasitesi yeterli olamamaktadır. IBM'de yapılan bir çalışmaya göre günümüz iş liderlerinin yarısından fazlası yapması gereken işleri kavrama ve öngörü erişimlerinin olmadığı kanaatinde olduklarını belirtmişlerdir (Zikopoulos ve Eaton, 2011). Bu eksiklikler veriye alternatif yöntemlerle yaklaşmayı gerektirdiğinden büyük veri kavramı ve sistemleri ortaya çıkmıştır. ProQuest araştırma kütüphanesinde büyük veri terimini içeren belgelerin sıklık oranlarının yıllara oranla değişimi gösterilirken, büyük veri teriminin 2011 yılında yaygınlaştığı ifade edilmektedir (Gandomi ve Haider, 2015).

Uzun yıllar bilişim dünyasında adından sıklıkla bahsedilen ve her geçen gün yeni geliştirilen araçlarla kullanımı yaygınlaşan büyük veri uygulamaları, Google, Facebook, LinkedIn, Yahoo ve daha birçok büyük teknoloji şirketleri tarafından yaygın olarak kullanılmaktadır. Kamu ve özel kurumların veri analizinde kullandıkları teknolojilerin büyük veriye adapte edilmesi konusunda kendilerini baskı altında hissettikleri belirtilmektedir (Wormer, 2014). McKinsey Global Enstitüsü⁴ raporuna göre, büyük verinin şirketlere çok büyük bir gelişme ve finansal fırsatlar sunduğu belirtilmiştir (Manyika vd., 2011). Obama Yönetimi, büyük veri teknolojilerini geliştirmek, büyük veri uygulamalarını göstermek ve yeni nesil veri bilimcilerini

³ 1 Zetabayt = 1024⁷ bayt

⁴ 1990 yılında kurulan McKinsey şirketinin işletme ve araştırma enstitüsüdür.

eğitmek için 2012 yılında Büyük Veri Araştırma ve Geliştirme Girişimi'ni başlatmış, aynı zamanda Beyaz Saray'a ilk kez *Veri Bilimi Yöneticisi* işe almıştır (Marzullo, 2016). ABD'de yayınlanan raporlar ve hükümetin bu alana verdiği önem sonucunda, eğitim alanında büyük veri ile ilgili gelişmeler başlamıştır. ABD'deki kolejler ve üniversiteler tarafından yürütülen federal destekli araştırmaların bir kısmını finanse eden ABD Ulusal Bilim Vakfı, 2012 yılında kendi başına bir disiplin olarak kurulan veri bilimine, büyük veri uygulamaları programını dâhil ederek bu alana katkıda bulunmuştur (National Science Foundation [NSF], 2019).

Büyük veri kamu ve özel kurumlarda yaygın bir şekilde kullanılmakta ve büyük veri analizinden anlayan nitelikli personel ihtiyacı da artmaktadır. 2015 yılında ABD'deki Temel Veri Bilimi ve Analitiği işleri için 2 milyon 350 binden fazla iş ilanı vardı ve 2020 yılına kadar 2 milyon 720 bine yakın iş ilanı olacağı tahmin edilmektedir (Miller ve Hughes, 2017). Eğitimciler ve eğitim sağlayıcıları, öğrencileri bugünün ve yarının analitikle ilgili pozisyonlarına hazırlayan programlarla analitik becerilere yönelik artan talebe proaktif bir şekilde yanıt vermek zorunda kalmışlardır (Miller ve Hughes, 2017). Büyük veri ve veri bilimi alanına 2010 yılından beri başta ABD ve Avrupa ülkelerinin farkındalıkları ve buna bağlı olarak sağlık, telekomünikasyon, finans, bankacılık, kentleşme, teknoloji, eğitim gibi alanlarda yatırımları günden günde artmıştır. Ancak bu alanda yetişmiş personel ihtiyacını karşılayacak olan üniversitelerde veri bilimi branşının ABD dışında çok az ülkede yaygın olduğu görülmektedir. ABD'de 60'a yakın üniversitede veri bilimi lisans programının olduğu belirtilmektedir. ABD dışında veri bilimi lisans programı olan üniversiteler İngiltere, Yeni Zelanda, Güney Kore, İrlanda, Hong-Kong, Fransa, İspanya ve Almanya'da yer almaktadır (DataScience, 2020).

Bütün dünyada olduğu gibi Türkiye'de de büyük veri alanında nitelikli personel ihtiyacı gün geçtikçe artmaktadır. Türkiye'de analitik yetenek ve yetkinlikler üzerine yapılan bir araştırmada, Kariyer.net'in veri tabanında yer alan 70 bine yakın firma içerisinde 2008 yılından itibaren 621 firma 2657 ilanında veri analitiği ve iş zekâsı gibi alanlara yer verdiğini açıklamaktadır (Analytics Center ve Kariyer.net, 2016). Türkiye'de gerçekleştirilen Bulut Bilişim ve Büyük Veri Çalıştayı anket sonuçlarına göre, ankete katılan kurumlarda veri işleme ve analizinde aşılması gereken sorun

olarak ilk sırada nitelikli personel eksikliği olarak belirlenmiştir (TÜBİTAK, 2016). T.C. Cumhurbaşkanlığı, Strateji Ve Bütçe Başkanlığı Stratejik Planında (2019-2023), “Hedef 4.2: Büyük veri temelli politika ve etki analizi çalışmaları yapılacaktır.” ifadesi ile büyük veriye verilen önem görülmektedir. Bu hedef kapsamında, “Büyük veri analizi konusunda tecrübeli uzman/veri bilimci kapasitesinin artırılması ve altyapının güçlendirilmesi” ihtiyaç olarak belirtilmiştir (Strateji Ve Bütçe Başkanlığı, 2018). Bu bağlamda, nitelikli personel ihtiyacını karşılayacak olan yükseköğretim kurumlarında Veri Bilimi ve Büyük Veri alanlarında yüksek lisans programları açıldığı görülmektedir. Ancak Türkiye’de kamu ve özel kurumların büyük veri analitiği konusunda üniversitelerden ve öğrencilerden beklentilerinin değerlendirilmesi konusunda bir çalışma bulunmamaktadır.

Bu doğrultuda Türkiye’de bilgi teknolojileri sektörünün büyük veri analitiği alanında mevcut durumunu ve kurumların beklentilerini değerlendirmeye yönelik bir çalışmanın yapılması gereksinimi duyulmuştur. Bu çalışma, kamu ve özel kurumların, büyük veri alanında ihtiyaç duydukları personel niteliğini tanımlamak, üniversite ve öğrencilerden beklentilerini belirlemek noktasında kaynak olması bakımından önem arz etmektedir. Bu çalışmada aşağıdaki araştırma sorularına cevap aranmaktadır.

- Büyük veri çözümleri sunan kamu ve özel kurumların tercih ettikleri büyük veri araçları ve teknolojileri nelerdir?
- Büyük veri çözümleri sunan kamu ve özel kurumların bilgisayar bilimleri ve ilgili bölümlerde okuyan öğrencilerden ve üniversitelerden büyük veri alanıyla ilgili beklentileri nelerdir?

2. BÜYÜK VERİ

2.1. Büyük Veri Tanımı ve Karakteristikleri

Büyük veri kelime olarak üretilmiş çok miktarda veri anlamındadır. Büyük veri, “Boyutu, geleneksel veri tabanı yazılım araçlarının kaydettiği, yönettiği ve analiz ettiği yeteneklerin ötesinde bir boyuta sahip olan veri setleri” olarak da tanımlanmıştır (Manyika vd., 2011). Gartner⁵ analistlerinden Laney (2001), büyük veriyi üç boyutuyla (3V⁶:

⁵ 1979 yılında kurulan araştırma ve danışmanlık şirketi

⁶ Volume, Velocity, Variety

Hacim, Hız, Çeşitlilik) ele alan ilk isim olmuştur. 2005 yılında O'Reilly Media'dan⁷ Roger Mougulas, Web 2.0 terimini oluşturduktan bir yıl sonra büyük veriyi, geleneksel iş zekâsı araçlarını kullanarak yönetimi ve işlenmesi neredeyse imkânsız olan geniş bir veri grubu olarak tanımlamıştır (Dontha, 2017). Davis ve Patterson (2012) ise "SQL gibi geleneksel veri tabanı protokolleri tarafından analiz edilemeyecek kadar büyük boyutlarda olan veri" olarak tanımlamışlardır. Büyük veri, bazı kaynaklarda, 3V'ye ek olarak değer veya gerçeklik özellikleri de eklenerek oluşturulan 4V modeli ile tanımlanmıştır (Singh ve Singh, 2017; Trifu ve Ivan, 2014). Ancak en çok kullanılan, hacim, hız, çeşitlilik, gerçeklik ve değer karakteristiklerinden oluşan 5V modelidir (Gahi, Guennoun ve Mouftah, 2016). Şekil 1'de büyük verinin 5V karakteristik özellikleri gösterilmektedir.

Volume (Veri Hacmi)	Velocity (Veri Hızı)	Variety (Veri Çeşitliliği)	Veracity (Veri Gerçekliği)	Value (Veri Değeri)
<ul style="list-style-type: none">•Terabaytlar•Kayıtlar•İşlemler•Tablolar•Dosyalar	<ul style="list-style-type: none">•Batch•Gerçek Zamanlı•Yakın Zamanlı•Süreçler•Akışlar	<ul style="list-style-type: none">•Yapısal•Yarı Yapısal•Yapısal Olmayan•Çok Faktörlü•Olasılıksal	<ul style="list-style-type: none">•Güvenilirlik•Geçerlilik•Kaynak, Saygınlık•Ulaşılabilirlik•Hesaplanabilirlik	<ul style="list-style-type: none">•İstatistiksel•Olgular•Korelasyonlar•Varsayımlar

Şekil 1. Büyük verinin 5V karakteristik özellikleri (Demchenko vd., 2013)

Hacim (Volume): Büyük veri hacmi, dosyalarda veya veri tabanlarında büyük miktarda kaydedilen ve çok çeşitli işlemlerle toplanan verinin miktar, boyut, ölçek ve hacmini ifade eden özelliklerini içermektedir. İşletmeler her geçen gün artan, terabaytlar hatta petabaytlar seviyelerinde bilgi toplamaktadırlar. Üretilen verinin çok hızlı artmasıyla, dünya çapındaki veri hacmi çok devasa boyutlara ulaşmaktadır. Mobil cihazlar ve kablosuz veri sensörleri her bir dakikada çok fazla veri üretmekte ve toplu veri alışverişleri milyarlarca internet servisleri arasında her saniye oluşmaya devam etmektedir. Bilimsel uygulamalar, videolar, tıbbi kayıtlar, operasyonel ticari veriler ve e-ticaret verileri büyük veri hacminin ciddi boyutlara ulaşmasına neden olan kaynaklardır (Demchenko vd., 2013).

Hız (Velocity): Büyük verinin işlenebildiği süreyi ifade etmektedir. Sensör verileri ve çoklu işlemlerden gelen veriler de dâhil olmak üzere büyük veriler, gerçek

⁷ Tim O'Reilly tarafından kurulmuş olan bu şirket, kitap yayınlamakta, teknik konferanslar düzenlemekte ve çevrimiçi öğrenme platformu sağlamaktadır.

zamanlı, gerçek zamanlıya yakın, toplu olarak veya akışlar halinde, yüksek hızda işlenmektedir (Rydning, 2018). Bazı faaliyetler çok önemlidir ve hızlı yanıt verilmesine ihtiyaç duyulmaktadır, bu nedenle veriyi hızlı işlemenin, verimliliği en üst düzeye çıkardığı görülmektedir. Örneğin Walmart⁸ saatte bir milyondan fazla işlem gerçekleştirmektedir (Cukier, 2010).

Çeşitlilik (Variety): Büyük verinin içerebileceği veri türünü ifade etmektedir. Bu veriler, yapısal, yarı yapısal veya yapısal olmayan veriler olabilmektedir. Mevcut verilerin sadece %5'ini oluşturan yapısal veriler (Cukier, 2010), elektronik tablolarda veya ilişkisel veri tabanlarında bulunan tablo verilerinden oluşmaktadır. Metin, resim, ses, video, sensör verileri, tıklama kayıtları, log dosyaları yapılandırılmamış verilere örnek olarak verilmektedir (Gahi vd., 2016).

Gerçeklik (Veracity): IBM, bu özelliği veri kaynaklarında bulunan güvenilirliği temsil eden dördüncü V olarak kullanmıştır. Örneğin, sosyal medyadaki müşteri uygulamaları belirsiz olabilmektedir, insani yargılamayı gerektirir yine de değerli bilgiler içermektedir. Gerçeklik, kesin olmayan ve belirsiz verilerin yönetimi ve madenciliği için geliştirilen araçlar ve analizler kullanılarak ele alınan büyük verinin bir özelliğidir (Gandomi ve Haider, 2015).

Değer (Value): Oracle, değeri büyük verilerin tanımlayıcı bir özelliği olarak tanıtmıştır. Oracle'in tanımına göre, büyük veriler genellikle "*düşük değer yoğunluğu*" ile karakterize edilmektedir. Yani orijinal formda alınan veriler genellikle hacmine göre düşük bir değere sahiptir ama büyük hacimlerde bu tür veriler analiz edilerek yüksek bir değer elde edilebilmektedir (Gandomi ve Haider, 2015). Değer, toplanan verilerin amaçlanan sürece, aktiviteye veya analize getirebileceği katma değer ile tanımlanan önemli bir özelliktir. Veri değeri, stokastik, olasılıklı, düzenli veya rastgele olarak temsil edilen işlemlere bağlı olmaktadır. Bu bağlamda veri değeri, veri hacmi ve çeşitliliği ile yakından ilgilidir (Gahi vd., 2016).

⁸ Dünya çapında şubeleri olan bir Amerikan perakende satış yapan mağazalar zinciridir.

2.2. Büyük Veri Altyapısı ve Hadoop

Hadoop, Java programlama dili ile yazılmış üst düzey bir bilgi işlem ortamı sunan açık kaynak kodlu bir yazılım ekosistemidir. Hadoop, çok büyük ölçekli veri uygulamaları için özel olarak geliştirilmiş ve tasarlanmış, dağıtık kümelenmiş bir dosya sisteminin üzerine inşa edilmiş bir bilgi işlem ortamıdır. Hadoop geliştirilirken, Google'ın dağıtık dosya sistemi (GFS-Google File System) ve MapReduce programlama paradigmasından ilham alınmıştır. Hadoop, büyük veri kümelerini analiz etmek ve yüksek oranda ölçeklenebilir, dağıtık toplu işlem sisteminden sonuçlar üretmek için tasarlanmıştır. Hadoop'un, Hadoop Dağıtık Dosya Sistemi (HDFS) ve Hadoop MapReduce olmak üzere iki temel bileşeni vardır (Jyothirmayee, Reddy ve Akbar, 2014). Hadoop, güvenilir ve paylaşılabilen depolama ve analiz sistemi sağlamaktadır. MapReduce temel olarak, anahtarlar ve değerlerle hesaplama yapan bir programlama modeli sağlayarak disk okuma ve yazma sorununu azaltmaktadır. Depolama işlemi Hadoop Dağıtık Dosya Sistemi (HDFS) tarafından, analiz işlemi ise MapReduce tarafından gerçekleştirilmektedir (Katal vd., 2013).

Büyük veri altyapısının önemli rollerinden birini de NoSQL veri tabanları oluşturmaktadır (Trifu ve Ivan, 2014). 'Not Only Sql' yani 'Sadece Sql Değil' anlamına gelen NoSQL, geleneksel veri tabanları yapısıyla kurulamayan geniş sütun kayıtları, doküman temelli, anahtar-değer yapısı veya birçok türden yapının yönetilmesine ve özelleştirilmesine olanak sağlayan farklı bir veri tabanı yaklaşımıdır (Warden, 2011). NoSQL veri tabanları, geleneksel olarak bilinen ve kullanılan ilişkisel veri tabanı yönetim sistemlerinden (RDBMS)⁹ farklı bir yapıdadır. NoSQL'in tasarımcısı Carlo Strozzi, ilişkisel veri tabanı modelinden farklı yapıda olduğundan dolayı, ilişkisel olmayan anlamında 'NoREL'¹⁰ veya benzer bir ad vermek gerektiğini ileri sürmektedir (Gökşen ve Aşan, 2015).

⁹ RDBMS: Relational Database Management Systems

¹⁰ NoREL: No RELation

3. YÖNTEM

Çalışmada toplanan veriler 1 Ocak 2020 tarihinden önce toplandığı için etik kurul izin belgesi bu çalışmaya eklenmemiştir.

3.1. Araştırma Deseni

Bu çalışmada nitel araştırma yöntemi kullanılmıştır ve gerçekleştirilen mülakatlardan elde edilen veriler tümevarımsal olarak adım adım geliştirilmiştir. Nitel araştırmalar çoğu zaman değişime açık ve esneklik ile yeniden düzenlemeye açık olma anlayışına sahiptir (Yıldırım ve Şimşek, 2013:53). Bu çalışmada, veri toplama ve analiz birlikte yürütülmüş, veriler toplandıktan sonra analizle ortaya çıkan kavramlar ve süreçler sonraki veri toplama aşamalarına dahil edilmiştir.

3.2. Örneklem

Görüşmelerde, kartopu örnekleme yöntemi seçilmiştir. Örneklem seçimindeki ölçüt, kurumun büyük veri çözümleri sunuyor olmasıdır. Kartopu örnekleme yöntemi sonucunda on kurumun genel müdürü ya da büyük veri proje müdürleri ile görüşmeler gerçekleştirilmiştir. Sektörlere göre büyük veri benimsenmesi dikkate alınarak farklı sektörlerde hizmet veren kurumlar tercih edilmeye çalışılmıştır.

Bu aşamada araştırmalar, Ankara ili genelinde büyük veri çözümleri sunan 2 kamu kurumu ve 8 özel kurumla görüşmeler yapılarak gerçekleştirilmiştir. Nitel araştırmada örneklem sayısı, toplanan verilere, sorulan sorulara, analizlere ve problemin durumuna göre değişmektedir. Temel ölçüt verinin doygunluğa ulaşmasıdır (Merriam, 2013:79). Bu çalışmada, görüşülecek kişi sayısı nitel araştırma metodolojisine uygun olacak şekilde önceden belirlenmemiş, verinin doyuma ulaştığı aşamada görüşmeler yeterli bulunup durdurulmuştur. Tablo 1’de görüşmecilerin demografik özellikleri; cinsiyeti, görevi, tecrübesi ve kurumların özellikleri; adı, kurum türü ve kuruluş yılı gösterilmiştir.

Tablo 1. Görüşülen Kurum ve Görüşmecilerin Profili

Şirket	Kurum	Kuruluş Yılı	Görüşmeci	Cinsiyet	Görev	Tecrübe (Yıl)
Kurum-1	Özel	2007	Kişi-1	Erkek	Genel Müdür	12
Kurum-2	Özel	2008	Kişi-2	Erkek	Proje Yöneticisi	Belirtilmemiş
Kurum-3	Özel	1998	Kişi-3	Erkek	Ürünler Direktörü	13
Kurum-4	Özel	1999	Kişi-4	Erkek	Çözüm Yöneticisi	10
Kurum-5	Özel	2004	Kişi-5	Erkek	Veri Analitiği Departmanı Yöneticisi	10
Kurum-6	Özel	2014	Kişi-6 Kişi-7	Erkek/ Kadın	Genel Müdür/ Yönetici Ortağı	6
Kurum-7	Özel	2010	Kişi-8	Erkek	Genel Müdür	9
Kurum-8	Özel	2018	Kişi-9	Erkek	Büyük Veri Yöneticisi	1
Kurum-9	Kamu	2017	Kişi-10	Erkek	Karar Destek Sistemleri Analisti	2
Kurum-10	Kamu	2013	Kişi-11	Erkek	B3Lab Yöneticisi	5

3.3. Veri Toplama

Bu araştırmada veri toplama, yarı yapılandırılmış görüşmeler üzerinden gerçekleştirilmiştir. Gerçekleştirilen yarı yapılandırılmış görüşmelerde, katılımcıya belirli bir esneklik tanınmıştır ancak katılımcının tamamen esnek olduğu yapılandırılmamış görüşme türünden de kaçınılmıştır. Görüşme süresince önceden belirlenen sorular üzerinde durularak, yukarıda değinilen araştırma sorularına yanıt alınabilecek noktalara odaklanılmış ve katılımcıların bu çerçeveye içerisinde kalmalarına dikkat edilmiştir. Yüz yüze yapılan görüşmeler katılımcıların kendi tercihleriyle kendi ofislerinde gerçekleştirilmiştir. Katılımcıların izni dâhilinde ses kaydı alınmıştır.

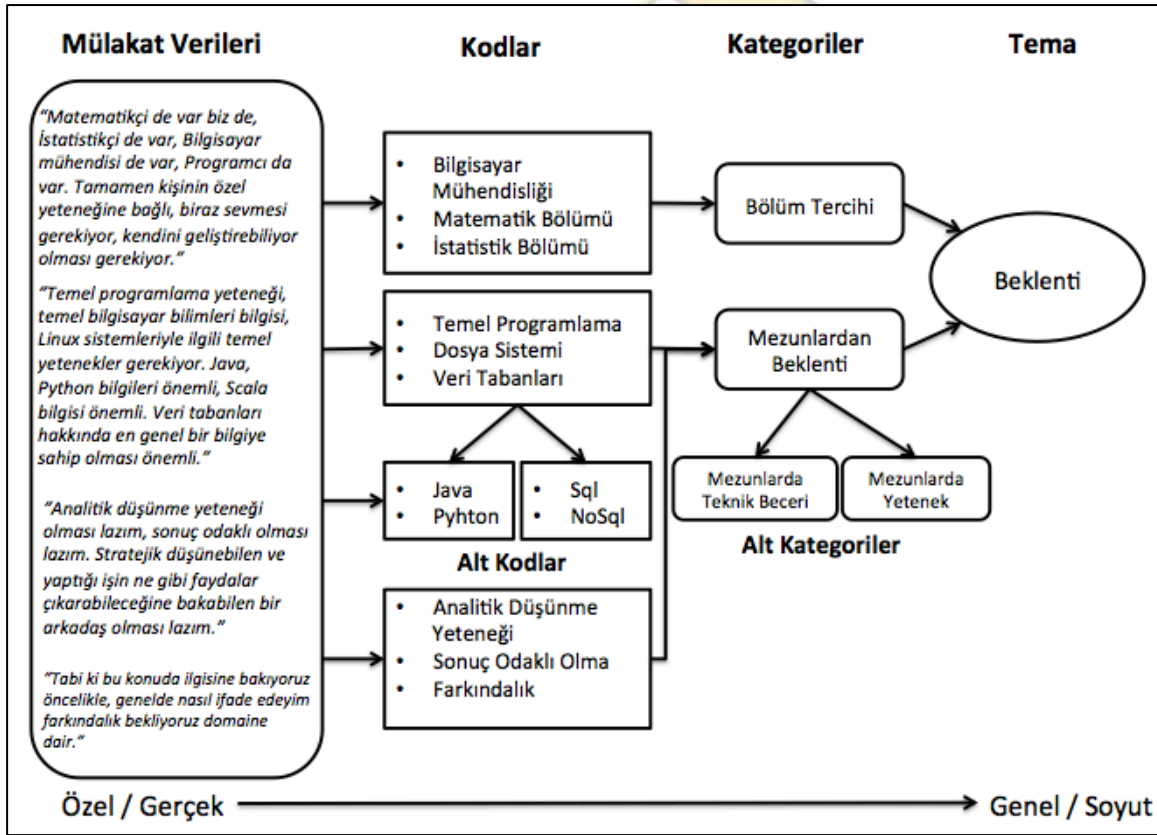
Araştırmanın etik açıdan değerlendirilmesi Gazi Üniversitesi Ölçme Değerlendirme Etik Alt Çalışma Grubu tarafından gerçekleştirilmiş olup 28.11.2019 tarihli E.150460 sayılı yazı ile onaylandığı görüşmecilere iletilmiştir.

3.4. Verilerin Analizi

Araştırmada, görüşmecilerden toplanan verilerin analizi için en uygun teknik, Glaser ve Strauss (1967), tarafından ortaya atılan gömülü teori yöntemi olarak belirlenmiştir. Gömülü teoride, ilk veriler toplandıktan sonra, araştırmacı verileri analiz eder ve analizden elde edilen kavramlar sonraki veri toplama işleminin temelini

oluşturur. Veri toplama ve analizi, araştırma süresi boyunca bir döngü içinde devam eder (Corbin ve Strauss, 2015).

Araştırmada, toplanan mülakat verilerinin analizi için öncelikle ses kayıtlarının transkripsiyon işlemleri yapılmıştır. Görüşme dökümleri incelendikten sonra NVivo 12 programı kullanılarak kodlama işlemine tabi tutulmuştur. Kodlama benzer olarak kodlanmış verilerin kategoriler şeklinde gruplanmasını sağlamaktadır (Saldana, 2013:39).



Şekil 2. Örnek Alıntılar Üzerinden Nitel Araştırmalarda Koddan Kategoriye Akış Modeli

Şekil 2'de, katılımcıların görüşme verilerinden, kod, kategori ve tema oluşturma sürecini gösteren örnek bir kodlama gösterilmektedir. En solda yer alan görüşmelerin transkripsiyonlarından (verilerden) kodlar, ilgili kodlar bir araya getirilerek kategoriler, kategorilerden tema oluşturulmuştur. Bu gösterimde sadece bir temanın belirli bir kısmı akış modeli üzerinde örnek olarak gösterilmiştir. Kodlar ve kategoriler oluşturulurken alt kodlar ve alt kategoriler belirlenmiştir.

3.5. Güvenilirlik ve Geçerlilik

Nitel araştırmanın geçerliliği ve güvenilirliği konusunda altı doğrulama ölçütü belirtilmektedir; yol ve tekniklerin belgelenmesi, yorumlama biçiminin argümantasyonu, kurallı olması, olguya yakınlık, iletişimsel geçerleme ve sacayağı (Mayring, 2011:145-148). Bu çalışmada, Mayring'in doğrulama ölçütleri temel alınarak araştırmanın bütün süreçleri açık bir şekilde ortaya konulmuştur.

4. BULGULAR

Bu çalışmada, bulgular ifade edilirken basit yüzde hesapları, sözcük sıklık hesapları kullanılmış ve yapılan açıklamaları destekleyen katılımcılardan birebir alıntılar yapılmıştır. Tablo 2'de görüşülen kurumlarla ilgili detaylı bilgiler yer almaktadır. Veri miktarı, görüşmecilerin ifadesiyle genel bir şekilde değerlendirilip, petabayt veya terabayt seviyesinde olduğu ifade edilirken, projeden projeye değişiklik gösterdiği de belirtilmiştir.

Tablo 2. Görüşülen Kurumların Detaylı Bilgileri

Şirket	Hizmet Ettiği Sektör	Veri Miktarı	Veri Türü	Kullanılan Sunucu	Kullanılan Yöntemler	Toplam/Büyük Veride Çalışan Sayısı (%Oran)
Kurum-1	Savunma, Kamu	Terabyte	Yapısal-Yapısal Olmayan	Sunucu ve Bulut Sunucu	Veri Madenciliği, Metin Madenciliği	67/17 (%25)
Kurum-2	Kamu, Sağlık	Petabyte-Terabyte	Yapısal-Yapısal Olmayan	Sunucu	Veri Analiz Yöntemleri	40/10 (%25)
Kurum-3	Askeri Sistemler	Petabyte-Terabyte	Yapısal Olmayan	Sunucu ve Bulut Sunucu	Veri Madenciliği, Görüntü İşleme	217/17 (%8)
Kurum-4	Telekomünikasyon, Sağlık, Finans	Terabyte	Yapısal-Yapısal Olmayan	Sunucu	-	1400/15 (%1)
Kurum-5	Telekomünikasyon, Bankacılık, Finans	Terabyte	Yapısal-Yapısal Olmayan	-	Veri Madenciliği, Veri Ambarı, Makine Öğrenmesi, İş Zekâsı	750/45 (%6)
Kurum-6	İletişim, Reklamcılık	Terabyte	Yapısal-Yapısal Olmayan	Sunucu	Veri Analizi Yöntemleri	7/2 (%29)
Kurum-7	Kamu, Özel, Perakende, Sağlık, Dağıtım	Petabyte-Terabyte	Yapısal	Sunucu ve Bulut Sunucu	Veri Ambarı, İş Zekâsı	40/16 (%40)
Kurum-8	Telekomünikasyon, Bankacılık, Finans, Sigorta, Perakende	Petabyte-Terabyte	Yapısal-Yapısal Olmayan	Sunucu	Veri Madenciliği, Veri Bilimi, Makine Öğrenmesi, ETL	6/2 (%33)
Kurum-9	Sağlık	Petabyte-Terabyte	Yapısal	Sunucu	Veri Madenciliği, İş Zekâsı, ETL	20/8 (%40)
Kurum-10	Kamu	Terabyte	Yapısal-Yapısal Olmayan	Sunucu ve Bulut Sunucu	Bulut Bilişim, Büyük Veri, Makine Öğrenmesi, Yapay Zekâ	33/20 (%61)

Görüşülen kurumlar, kamu, sağlık, telekomünikasyon, bankacılık ve finans, perakende, dağıtım, savunma, askeri sistemler, iletişim ve reklamcılık gibi sektörlerle hizmet vermektedir. Kurumların hepsinin büyük veri projelerinde kullandıkları veri miktarı terabayt seviyesindedir, yarısının ise çalıştıkları bazı projelerde kullandıkları veri miktarı petabayt seviyesine ulaşmaktadır. Görüşülen kurumların çoğunun kullandıkları veri türü hem yapısal hem de yapısal olmayan veri türüdür. Kurumların birkaçının büyük veri ekipleri, normal sunucuların yanında bulut sunucular ile de çalışmaktadır. Kurumlar, veri madenciliği, veri ambarı, iş zekâsı, ETL,¹¹ makine öğrenmesi, bulut bilişim, metin madenciliği, görüntü işleme ve büyük veri teknikleri gibi veri analizi yöntemlerinden birkaçını bir arada kullanmaktadır.

İlk olarak, görüşme yapılan kurumlardan toplanan verilere göre elde edilen bulgular, kurumların büyük veri analitiği mevcut durumu, üniversitelerin ilgili bölümlerinden ve öğrencilerden beklentileri, öğrencilere ve üniversitelere önerileri olmak üzere üç ana başlıkta belirtilmektedir. Kurumların büyük veri analitiği mevcut durumu kategorisinde, büyük veri ekibinde hangi pozisyonların tercih edildiği, bu ekiplerin çalışan sayısı ve yeterliliği ve kullandıkları teknolojilere yer verilmektedir. Beklentiler kategorisinde, büyük veri alanında hangi bölüm mezunlarının tercih edildiği, mezunlardan teknik ve beceri anlamında beklentilerin ne yönde olduğu yer almaktadır. Öneriler kısmında ise kurumların, bu alanda çalışmayı düşünen öğrencilere ve bu alanla alakalı olarak üniversitelere olan tavsiyelerine yer verilmektedir (Tablo 3).

Tablo 3. Temel Kategoriler

Büyük Veri Mevcut Durum	Çalışanların Pozisyonları
	Çalışan Sayısı ve Yeterliliği
	Kullanılan Teknolojiler
Beklentiler	Bölüm Tercihi
	Mezunlardan Beklentiler
Öneriler	Öğrencilere Öneriler
	Üniversitelere Öneriler

¹¹ ETL (Extract, Transform, Load): Hesaplama, çıkarma, dönüştürme, yükleme işlemlerinde, bir veya daha fazla kaynaktan verileri, kaynaktan farklı bir şekilde veya kaynaktan farklı bir şekilde temsil eden bir hedef sisteme kopyalamanın genel prosedürüdür.

4.1. Büyük Veri Mevcut Durum

Bu kategoride, görüşülen kurumların büyük veri projelerinde çalışanların hangi pozisyonda çalıştıkları, büyük veri projelerinde çalışan sayıları, bu sayıların yeterliliği ve büyük veri projelerinde kullanılan araçlar ve teknolojiler belirtilmektedir.

4.1.1. Büyük Veri Projelerinde Çalışanların Pozisyonları

Görüşülen kurumların verdikleri cevaplar doğrultusunda, çok çeşitli pozisyon isimleri kullanıldığı görülmüş ve en çok tercih edilen pozisyon isminin yazılım mühendisi olduğu ortaya çıkmıştır. Ardından, en çok tercih edilen pozisyon isimleri, veri bilimci, veri mühendisi, veri analisti, büyük veri yöneticisidir. Bunların haricinde az sayıda da olsa ifade edilen diğer pozisyonlar, büyük veri geliştirici, büyük veri mimarı, ETL geliştirici, iş zekâsı geliştirici olarak belirtilmiştir. Burada pozisyonu, genel ifadeyle yazılım mühendisi olarak belirten kurumlardan biri şu ifadelerle yer vermektedir;

“Farklı farklı pozisyonlar var, ama bir veri analisti veya işte bir veri tabanı uzmanı gibi uzmanlık alanlarına göre bizde bir sınıflandırma yok ne yazık ki. Ama onda uzmanlaşan onda çalışan insanlar var, hepsinin genel adı bizde yazılım mühendisi olarak geçiyor.”
(Kişi-3)

Bu üç temel rol ayrımına bakıldığında görüşülen kurumlar arasında B3Lab, büyük veri alanında yaptığı çalışmalarda veri mühendisi ve veri bilimci pozisyonlarını net bir şekilde ortaya koyan tek kurum olarak göze çarpmaktadır. B3Lab, bulut bilişim, büyük veri, makine öğrenmesi ve bunların kullanımıyla ilgili, araştırma departmanı gibi çalışan bir laboratuvar olması dolayısıyla veri mühendisi ve veri bilimci pozisyonlarında ayrım olması önem arz etmektedir. Kurumların büyük veri projelerinde çalışanların pozisyonlarını öğrenmenin yanı sıra, bu alanda çalışacaklara iş ilanı açmak istediklerinde pozisyonu aynı mı yoksa farklı mı açacakları sorulduğunda, en çok tercih edilen pozisyon yazılım mühendisi veya yazılım geliştirici olarak belirtilmiştir. Ancak ilana ek olarak hangi büyük veri araçlarını kullanıyorlarsa o araçlarda tecrübeli olma, daha önce bu araçlarla çalışma gibi ölçütler de belirtilmiştir. Kurumlardan çoğu ilanını, yazılım mühendisi veya yazılım geliştirici olarak belirtirken, sadece TÜBİTAK B3Lab, veri mühendisi ve veri bilimci olarak iki

ayrı kategoride ifade etmektedir. Veri mühendisini, veri uygulamalarını tasarlayan ve veri akışıyla ilgili konfigürasyonları sağlayan kişi, veri bilimciyi ise makine öğrenmesi ve derin öğrenme metotlarını kullanan, Python, R gibi dillerle destekleyerek veri analizi yapan kişi olarak belirtmektedir.

ABD İşgücü İstatistikleri Bürosu¹², veri bilimi ve analitiği alanını kapsayan pozisyonların net bir tanıma sahip olmadığını, veri bilimcisi ve veri mühendisliği gibi birçok yeni gelişmekte olan veri bilimi ve analitiği alanındaki pozisyonların takip edilemediğini belirtmektedir (Miller ve Hughes, 2017). Bir şirkette 'veri bilimcisi' olarak adlandırılan bir çalışan ile başka bir şirkette 'veri bilimcisi' olarak adlandırılan bir çalışanın belirgin bir şekilde farklı teknik becerilere sahip olduğundan dolayı iş unvanlarının veri bilimi ve analitiği pozisyonlarının çoğunda tutarlı olmadığını belirtmektedir (Miller ve Hughes, 2017). Bu çalışmada görüşülen kurumlarda iş unvanlarının, büyük veri analitiği alanında çok da tutarlı olmadığı, Miller ve Hughes (2017)'in görüşünü desteklemektedir. Görüşülen kurumlarda büyük veri alanında çalışanların farklı teknik beceri ve yeteneklere sahip olmalarından pozisyon ismi olarak genelde yazılım mühendisi ve benzer isimler tercih edilmiştir.

4.1.2. Büyük Veri Projelerinde Çalışan Sayısı ve Yeterliliği

Görüşülen kurumlarda, toplam çalışan sayısının, büyük veri projelerinde çalışan sayısına oranı Tablo 3'de gösterilmektedir. Bu oranlar, kurumun hizmet ettiği sektöre, toplam çalışan sayısına, gerçekleştirilen ürüne ve projelere bağlı olarak değişmektedir. Görüşülen kurumların yarısı, büyük veri projelerinde çalışan sayısını yeterli görmektedir. Bu sayının yeterli görülmesi, kurumlardan bazılarının ifadesiyle, proje tabanlı çalıştıklarından dolayı bu projelere göre yeterli oldukları şeklindedir. Kimola şirketinin kurucu ortağı ile yapılan görüşmede, ürünlerin yıllar içerisinde geliştiğini ve şu an ürünlerin tamamlandığı için sayının yeterli görüldüğü belirtilmiştir. Büyük veri projelerinde çalışan sayısının yeterli görülmediği kurumlar içerisinde dikkat çeken B3Lab, %61'lik oranla fazla sayıda büyük veri projelerinde çalışan olmasına rağmen sayıyı yeterli görmemektedir. B3Lab'ın bu oranı yeterli bulmaması, özünde bulut bilişim ve büyük veri teknolojileri geliştirmek üzere

¹² ABD'de işgücü piyasasının temel metriklerini izlemekle görevli federal kurum

kurulmuş bir laboratuvar olmasından kaynaklanmaktadır. Büyük veri projelerinde çalışan sayısının yeterli görülmediği, bir görüşmeci tarafından şu şekilde ifade edilmiştir;

“Başlangıçta ihtiyaç analizi ve kurulum için yeterli oluyor ancak idame ve bu büyük veri ortamı üzerinde proje geliştirmek gerektirdiği zaman yeterli olmuyor.” (Kişi-9)

4.1.3. Büyük Veri Projelerinde Kullanılan Teknolojiler

Bu kategoride, projelerde hangi araçların kullanıldığına ve bu araçlar kullanılırken karşılaşılan avantajlara ve sorunlara yer verilmiştir. Kurumların çoğunun Hadoop ve Hadoop’un ekosistemini kullandığı görülmüştür. Kullanılan araçlar arasında Spark, Cassandra, Hive ve Kafka en çok tercih edilenler arasındadır. Bunlardan sonra, Elastic Search, HBase, HDFS, Impala, Kudu ve MongoDB gelmektedir. Cloudera, Hortonworks, CouchBase, Sqoop, Drill, Flume, Nifi, R, Solr, Tez, Yarn ve ZooKeeper da görüşülen kurumlar tarafından en az sayıda kullanılan diğer araçlar ve platformlar olarak kaydedilmiştir.

Kurumların çoğu, büyük veri araçlarını kullanırken en önemli avantajın, performans ve daha hızlı analiz olduğunu belirtmekte ve görüşmecilerden biri bu avantajı şu şekilde ifade etmektedir;

“Hadoop, normal veri tabanlarına kıyasla sütun bazlı veri işlenebildiği için daha hızlı çalışabiliyor.” (Kişi-10)

Maliyet ve açık kaynaklı olması da, performanstan sonra gelen avantajlar olarak belirtilmiştir. Kullanılan araçların açık kaynaklı ve lisanssız olması maliyet avantajı sağlamaktadır. Aynı zamanda araçların açık kaynaklı olması ile ürünün altyapısına hâkim olmaları ve herhangi bir sorunla karşılaştığı zaman, koda müdahale ederek, kendi çözümlerini üretebilmeleri de avantaj olarak görülmektedir. Bu konuda görüşmecilerden biri şu ifadelerle yer vermiştir;

“Karar destek sistemlerinde hızlı karar verme, detaylı analizler gibi birçok alanda avantaj sağlıyor. Bir de mümkün olduğu kadar kullandığımız ürünün altyapısına da hâkim olmamız gerekiyor. Çünkü bir sorunla karşılaştığımız zaman, en kötü ihtimalle kendi içimizde kendi arkadaşlarımızın çözebileceği, müdahale edebileceği bir şey olması gerekiyor. Yoksa yarı yolda kalırız. Açık kaynak olması bu yüzden önemli.” (Kişi-3)

Büyük veri projelerinde kullanılan araçların alternatiflerinin çok olması ve verinin dağıtık saklanması da bu kategoride belirtilmiştir. Alternatiflerin çok olması, kurumlara kendi çözümlerini üretirken, kendi projelerine göre uygun aracı seçme imkânı sağlamaktadır. Bunların dışında, CAP¹³ teoremi avantajı, daha iyi analiz, farklı türlerde veriyi işleyebilme, optimizasyon, ölçeklendirme, servis devamlılığı, standartlaşma, yatayda genişleme ve yüksek kullanılabilirlik gibi diğer avantajlardan da bahsedilmiştir.

Kurumlar, büyük veri teknolojilerini kullanılırken karşılaştıkları en önemli sorunları sırayla nitelikli personel yetersizliği, teknik problemler ve farkındalık olarak belirtmişlerdir. Kullanılan araçların açık kaynak olması bazı kurumlar için avantaj olarak belirtilirken, bazı kurumlar için sorun oluşturmaktadır. Açık kaynaklı ürünler, çok güncellenmediği, bazı hataları olduğu ve dökümantasyon eksikliği olduğu için bu yazılımları kullanmak risk teşkil etmektedir. Lisanslı ürün olmadığı için, öğrenme güçlüğü yaşandığı, lisanslı ürünlere yatkınlıktan dolayı bu yazılımların karmaşık geldiği, farklı kaynaklardan gelen verilerin birleştirilmesinde sorunlar yaşandığı ve güvenlikle alakalı problemler olduğu kurumlar tarafından belirtilmiştir.

Northeastern Üniversitesinde Profesyonel Çalışmalar ve Analitik Programı profesörlerinden Thomas Goulding, 2020'nin en büyük analitik probleminin, büyük miktardaki veri ile çalışmak için gereken araçlar ve eğitim ile nitelikli veri analistlerinin eksikliği olacağını ifade etmektedir (DiFranza, 2019). Goulding'in belirttiği problemle kurumların karşı karşıya kaldığı sorunların örtüşmesi dikkat çekicidir. Bu konu ile ilgili şu ifadelere yer verilmiştir;

"Personel noktasında, yetişmiş eleman noktasında bir sıkıntı var ama biz hazırda eleman bulamasak da yeni mezun arkadaşları alıp burada yetiştirip, bu konuda belli bir yetkinlik sağlayarak o şekilde açığı kapattığımızı düşünüyoruz." (Kişi-11)

¹³ CAP: Consistency, Availability, Partition Tolerance

4.2. Beklentiler

Bu kategoride kurumlar, bu alanda çalışmak üzere hangi bölüm mezunlarını tercih ettiklerini ve mezunların yetenek ve teknik becerileri konusunda ne gibi özelliklerin olmasını istediklerini ifade etmektedirler.

4.2.1. Bölüm Tercihi

Büyük veri alanında çalışmayı düşünen mezunlara, sektörün özellikle hangi bölümden mezun olan kişileri tercih ettiklerini öğrenmeye yönelik sorulan sorularda, görüşmecilerin cevapları doğrultusunda iki ana kategoride bölüm tercihleri belirttikleri görülmüştür. Bu kategorilerden ilki yazılım geliştirme olarak belirtilmekte, katılımcıların çoğu bu alan için bilgisayar mühendisi tercih ettiklerini belirtmiştir.

Bir katılımcı, ekiplerinde elektronik mühendisi de yer aldığını ancak bu kişilerin de üniversitede eğitim alırken bilgisayar alt branşını tercih etmiş ve bilgisayar mühendisliği dersleri almış kişiler olduğunu belirtmiştir. Bir başka katılımcı yazılım geliştiriciler için bilgisayar mühendisliğinin kesinlikle birinci tercihleri olduğunu ifade ederken şunları belirtmektedir;

“Bir bilgisayar mühendisi daha sistemli, düzenli çalışıyor. Kişiden kişiye değişir ama bir eğitim çerçevesinde hareket ediyor. Ötekisi bir şekilde kendini bu işe adanmış, mesela çok tutucu olabiliyor, kendi yöntemlerinin çok ideal olduğunu, çok iyi olduğunu zannediyor. Ekip çalışmasına yatkınlık zor oluyor çünkü kendi başına çalışarak kendini geliştirmiş, onun için ekip içerisinde olmaktan çok hoşnut olmuyor. Veya mesela temel metodolojileri atlamış, bir bilgisayar mühendisinin sahip olması gereken şeyleri atlamış, doğrudan kod yazmaya başlamış. Öyle olunca mesela software engineering dersi almamış, database dersi almamış, bir normalization vs. gibi şeyleri çok bilmiyor.” (Kişi-2)

Katılımcıların üzerinde durdukları diğer bir husus, büyük veri alanında çalışmayı düşünen kişilerin işe alım sürecinde özel bir bölüm tercihlerinin olmamasıdır. Ancak bu belirtilirken, mühendisliklerden bilgisayar, elektrik elektronik, endüstri gibi bölümler ve matematik, istatistik, programcılık gibi bölümler

belirtilerek bunların arasından özel bir tercihimiz yok şeklinde ifade edilmektedir. Katılımcılardan biri bu konuda şu ifadelere yer vermektedir;

“Matematikçi de var biz de İstatistikçi de var, Bilgisayar mühendisi de var, Programcı da var. Tamamen kişinin özel yeteneğine bağlı, biraz sevmesi gerekiyor, kendini geliştirebiliyor olması gerekiyor.” (Kişi-1)

Büyük veri alanında sektörün tercih ettiği bölümlerden ikinci kategoriye, veri bilimi, veri madenciliği ve makine öğrenmesi oluşturmaktadır. Görüşmelere katılan kurumların %60'ı istatistik, %50'si matematik ve %50'si bilgisayar mühendisliği bölümlerini tercih etmektedir. Bu bölümlerin yanında, yönetim bilişim sistemleri, endüstri, elektrik-elektronik ve matematik mühendisliği bölümlerini tercih eden katılımcılar da mevcuttur (Tablo 4).

Tablo 4. Kurumların Veri Bilimi Alanında Tercih Ettikleri Bölümler

	Kamu (1)	Özel (8)	Lab (1)	Toplam (10)
Matematik Mühendisi	0	1	0	1
Elektrik-Elektronik Mühendisi	0	1	0	1
Endüstri Mühendisi	0	2	0	2
Yönetim Bilişim Sistemleri	0	1	1	2
Bilgisayar Mühendisi	1	3	1	5
Matematik	1	3	1	5
İstatistik	1	5	0	6
Bölüm Tercihi Yok	1	4	1	6

Görüşmecilerden birinin özellikle istatistik bölümü üzerinde durdukları, şu ifadelerden anlaşılmaktadır;

“Şöyle, buradaki ekibe bakarak öyle, illa bilgisayar mühendisliği olsun gibi bir kriterimiz yok, istatistikçi çok var.” (Kişi-10)

Veri bilimi alanında da özellikle bir bölüm tercihleri olmadığını belirten çok sayıda katılımcı mevcuttur. Özel bir bölüm tercihi olmayan kurumlar, istatistik, matematik, bilgisayar mühendisliği, elektrik-elektronik, endüstri ve matematik mühendisliği ve yönetim bilişim sistemleri gibi bölümleri kastetmektedir. Katılımcılardan biri bu konuyu şu şekilde ifade etmektedir;

“Veri bilimi yeni bir şey olduğu için, siz istatistikten mezun birini alayım deseniz de bu alana ilgi duyuyor olması lazım, bunda çalışması, kendini geliştirecek birtakım şeyler yapmış olması lazım, bazı projelerde çalışmış olması lazım. Biz, bu alanda kendisinin yetkin olduğunu düşünen kişilerle görüşüyoruz. Zaten o belli ediyor, geçmişinden, tecrübesinden, kullandığı jargondan, her şeyden anlaşılıyor.” (Kişi-2)

Görüşülen kurumların, herhangi bir bölüm tercih etmemesi, kurumların, hangi bölümden mezun olduğundan ziyade teknik becerileri ön planda tuttıklarını göstermektedir.

4.2.2. Mezunlardan Beklentiler

Kurumların, büyük veri alanında çalışmak üzere üniversiteden yeni mezun kişilerde aradıkları özellikler, yetenek ve teknik beceri olarak iki kategoride incelenmiştir. Mezunlardan beklenen yetenekler içerisinde en çok ifade edilen farkındalık ve analitik düşünme yeteneği olmuştur. Farkındalık, konuya ilgili ve alakalı olma ve teknolojileri takip etme gibi ifadelerle belirtilmektedir. Farkındalık beklentisi üzerine görüşmecilerden biri şu ifadelerle yer vermiştir;

“Yeni mezun alacaksam eğer her şeyden önce konuya ilgiye bakarım. Daha sonra yaptığı benzer çalışmalar var mı, tez safhasında veya yüksek yapıyorsa çalışmalarına bakarım. Tabi ki bu konuda ilgisine bakıyoruz öncelikle, genelde nasıl ifade edeyim farkındalık bekliyoruz domaine dair.” (Kişi-11)

Katılımcılar tarafından analitik düşünme yeteneği; problem çözme yeteneği, sonuç odaklı olma ve stratejik düşünme gibi ifadelerle belirtilmektedir. Görüşmecilerden biri, beklentilerini şu şekilde ifade etmiştir;

“Analitik düşünme yeteneği olması lazım, sonuç odaklı olması lazım. Stratejik düşünebilen ve yaptığı işin ne gibi faydalar çıkarabileceğine bakabilen bir arkadaş olması lazım.” (Kişi-4)

Bunların yanında, takım çalışmasına yatkınlık, öğrenme yeteneğinin olması ve hevesli olma, kendi gelişimine vakit ayırma ve iletişim becerilerinin olması gibi kriterler bazı kurumlar tarafından ifade edilmiştir. Bulut Bilişim ve Büyük Veri Çalıştayı Raporunda, veri işleme ve analizinde aşılması gereken sorun olarak üçüncü sırada farkındalık olarak belirtilmiştir (TÜBİTAK, 2016). Çalıştayda kurumların

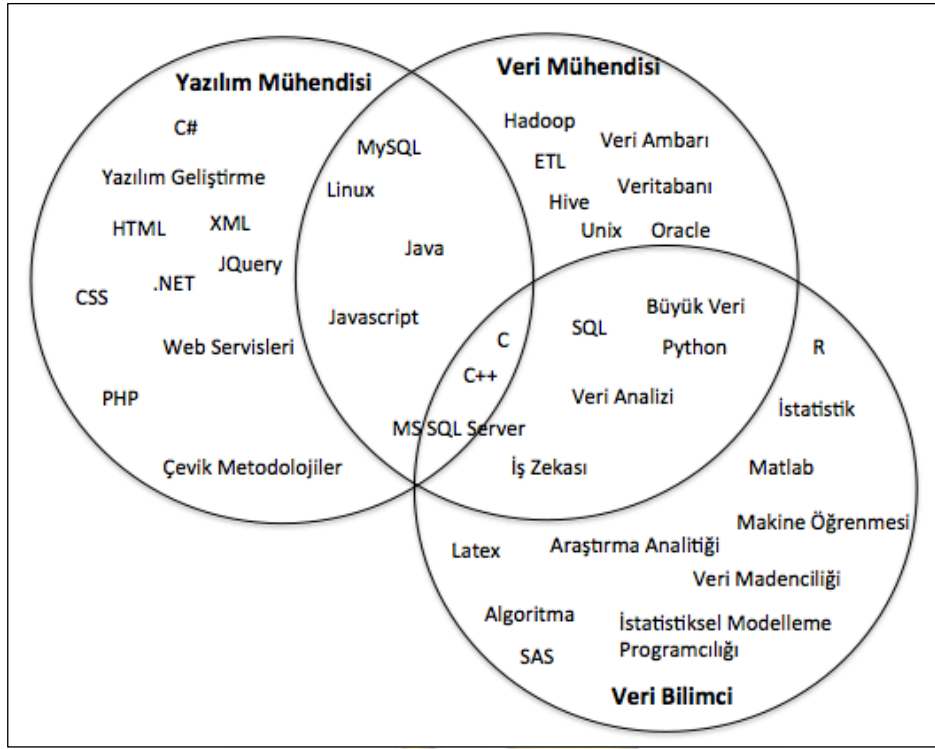
karşılaştığı sorunlardan olan farkındalık, görüşülen kurumlar arasında sektörün yeni mezunlardan en çok beklediği özellik olması dikkat çekicidir.

Görüşülen kurumlardan teknik yetenek beklentileri olarak ilk sırada belirtilenler, veri tabanı yani SQL ve temel programlama yetenekleri gelmektedir. Ardından en çok beklenen yetenekler içinde Linux ve dosya sistemi gelmektedir. Az sayıda da olsa ifade edilen diğer teknik yetenekler, Hadoop ve araçları, iş zekası, ETL araçları, veri analizi ve temel bilgisayar bilimleri yetenekleridir. Bu konuda bir görüşmecinin ifadeleri şu şekildedir;

“Temel programlama yeteneği, temel bilgisayar bilimleri bilgisi, Linux sistemleriyle ilgili temel yetenekler gerekiyor. Java, Python bilgileri önemli, Scala bilgisi önemli. Veri tabanları hakkında en genel bir bilgiye sahip olması önemli.” (Kişi-9)

Amazon şirketinden Guy Lebanon, büyük veriden anlamlı bilgiye ulaşmak için bilgisayar ve yazılım mühendisliği; makine öğrenmesi, istatistik ve optimizasyon; ürün anlayışı ve deneyimden oluşan üç becerinin gerektiğini belirtmektedir. Bilgisayar ve yazılım mühendisliğinin, C++ veya Java gibi programlama dilleri, veri toplama, paralel ve dağıtık hesaplama ile veri işleme, kod dokümantasyonu, birim ve entegrasyon testleri gibi yazılım mühendisliği uygulama bilgisi, yazılım hizmetleri arasındaki iletişim ve verimlilik gibi konuları içerdiğini belirtmektedir. Makine öğrenmesi, doğrusal olmayan optimizasyon, pratik yöntemler ve yaygın makine öğrenme sorunları gibi konuları kapsamaktadır. Ürün anlayışı, iş hedeflerini ölçmek için çevrimiçi değerlendirme süreci, model eğitimi, ürün modifikasyonu ve geçmiş kullanımı ile ilgili kararları içermektedir (Mellody, 2014). Lebanon’ın belirttiği teknik yetenekler, görüşmecilerin mezunlardan beklediği teknik becerilerle aynı doğrultudadır. Şekil 3’de yazılım mühendisliği, veri mühendisliği ve veri bilimcilerde olması gereken teknik yetenekler gösterilmiştir. Görüşülen kurumlardan alınan cevaplar doğrultusunda, bu üç pozisyonun ayırımının tam olarak yapılamadığı ve teknik yetenekler noktasında da net ayrımlar yapılamadığı görülmüştür. Ancak görüşmecilerin ifadeleri doğrultusunda, veri bilimcilerinin temel programlama dillerinden herhangi birini biliyor olması belirtilirken, Swansstrom (2016)’un veri

bilimciye olması gereken teknik yetenekler içerisinde de C, C++ gibi temel programlama dillerinin olduğu görülmektedir (Şekil 3).



Şekil 3. Yazılım Mühendisliği, Veri Mühendisliği ve Veri Bilimcinin Teknik Yetenekleri (Swanstrom, 2016)

Veri mühendisliği giderek daha fazla yazılım mühendisliği disiplini içinde bir alt küme olarak ortaya çıkmaktadır. Yazılım mühendisliği, veri mühendisliği ve veri bilimci arasında önemli farklar bulunmaktadır. Yazılım mühendisi, uygulama ve sistem geliştirme, tasarım, kod yazma ve inceleme, test etme gibi süreçlerde yer almaktadır. Yazılım mühendisliği bu üç rol içinde en eski olanıdır. Veri mühendisi, yazılım mühendisleri tarafından oluşturulan çeşitli uygulama ve sistemlerden gelen verileri çeken, birleştiren ve depolayan sistemleri kurmaktadır. Diğer bir ifadeyle, veri bilimcilerin analiz edeceği ve raporlayacağı büyük veri altyapısını hazırlamaktadır. Veri mühendislerinin %40'ı daha önce yazılım mühendisi olarak çalıştığı ifade edilmektedir. Veri bilimcisi ise veriler üzerinde analiz yapmaktadır. Bu analiz, müşteri davranışını daha iyi anlamaya çalışan bir ekip için bir defalık analiz olabilirken, yazılım ve veri mühendisleri tarafından kullanılmak üzere bir makine öğrenmesi algoritması da olabilmektedir (Swanstrom, 2016). Görüşülen kurumların ifadeleri

doğrultusunda, yazılım mühendisinin, veri mühendisi ve veri bilimciden henüz çok ayrılmadığı görülmektedir. Özellikle veri mühendisi, kurumlar tarafından pozisyon olarak pek kullanılmamaktadır. Burada belirtilen bu üç rolün, kurumlarda ayrımının yapılmadığı görülmektedir.

4.3. Öneriler

Bu kategoride, kurumların büyük veri alanında çalışmayı düşünen öğrencilere ve bu alanla ilgili bölümleri olan üniversitelere tavsiyeleri yer almaktadır.

4.3.1. Öğrencilere Öneriler

Görüşmecilerin verilerinden elde edilen öneriler arasında farkındalığa sahip olma, okulda öğretilenler dışında ekstra çaba gösterme, bu alanda yarı zamanlı çalışma veya staj yapma, öğrenme devamlılığına sahip olma ve öğrenmeye açık olma, çevrimiçi eğitimlere katılma, seminer, kurs ve etkinliklere katılma, kendini tanıma, ne istediğini bilme ve derslerinde başarılı olma yer almaktadır.

Kurumlar tarafından %90 oranla en çok belirtilen öneri, alt kategorileri ile beraber farkındalığa sahip olma olarak belirtilmiştir. Farkındalığa sahip olma kategorisi, yeni teknolojileri takip etme, temel programlama bilgisine sahip olma, dağıtık sistemler konusunda bilgi sahibi olma, büyük veri çözümleri sunan kurumları takip etme, konuyla ilgili çalışmalar yapma, veritabanı ve Sql konusunda bilgi sahibi olma gibi alt kategorilerden oluşmaktadır. Yeni teknolojileri takip etme konusunda çoğu katılımcı, teknolojilerin çok hızlı değiştiğini ve alternatiflerinin çıktığını, bu hızla yetişebilmek için sürekli yeni çıkan teknolojilerden haberdar olmaları gerektiğini ifade etmişlerdir. Bu konuda bir görüşmecinin şu ifadeleri dikkat çekicidir;

"Kendinizi teknolojide güncel tutmaz iseniz, bildikleriniz yetmiyor, sürekli üstüne üstüne koymanız gerekiyor. Onlarca teknoloji var, birkaç sene sonra, 5-10 tane daha başka ürün oluyor olacak, yeni ürünler girmiş olacak. O zaman sürekli takip etmesi gerekecek." (Kişi-3)

Temel programlama bilgisi olarak ifade edilen, bir tanesinin iyi bilinmesinin yeterli olduğu ve diğerlerinin kolayca öğrenebileceği şeklindedir. Bunun yanında araçların geçici olabileceği ve işin temel mantığı olan dağıtık sistemlerin ve temel veri

tabanı konularının iyi kavranması gerektiği ortaya konulmuştur. Büyük veri çözümleri sunan kurumları ve bunların ne gibi sorunlarla uğraştıklarını takip ederek, konuyla ilgili çalışmalar yaparak farkındalığı arttırmaları gerektiği vurgulanmaktadır.

İkinci en çok belirtilen öneri, %60 oranla okulda öğretilenler dışında ekstra çaba gösterilmesidir. Bu konuda katılımcılarımız, öğrencilerin sadece okulun verdikleri ile yetinmemeleri gerektiğini, boş zamanlarında ekstra bir merak göstererek çalışmalar yapmalarını, ilgi alanlarına göre çevrimiçi kaynaklardan kendi kendilerini geliştirmelerini ifade etmişlerdir. Görüşmecilerin, öğrencilerin okul dışında ekstra çaba göstermesi konusunda çeşitli ifadeler yer vermeleri, üniversitelerde neden güncel konu ve teknolojilerin müfredatlarda yer bulamaması sorusunu akıllara getirmektedir. Bu konuda bir katılımcı şu ifadeleri kullanmıştır;

“Bu alanla ilgili temel yeteneklerini geliştirsinler çünkü hala benim bildiğim kadarıyla üniversitelerde büyük veri eğitimi yok. Dolayısıyla alternatif kaynaklardan işte online veya ilgili kurslardan öğrenmek durumundalar. Youtube ve Udemy bu konuda güzel kaynaklar. Bu alanda kendi yeteneklerini geliştirsinler ve kendi yaptıkları böyle küçük de olsa projelerini, kodlarını LinkedIn gibi sosyal medyada paylaşsınlar. Başkalarının kendilerini tanımalarına izin versinler.” (Kişi-9)

4.3.2. Üniversitelere öneriler

Büyük veri alanına personel yetiştiren özellikle bilgisayar mühendisliği, istatistik ve matematik bölümleri, yönetim bilişim sistemleri bölümü gibi bölümleri olan üniversitelere görüşmecilerin önerileri, müfredat yenileme, uzun dönem staj ve yüksek lisans programları açmak yönündedir.

Bu kategoride katılımcıların %80'i üniversitelere müfredat yenileme önerisinde bulunmaktadır. Müfredat yenileme önerisinin alt kategorileri, yeni ders açma, NoSql konusunun veritabanı dersine eklenmesi, özel sektörden fikir alınması, müfredatların piyasa taleplerine uygun olması ve temel teknolojiler konusunda güncelin takip edilmesi olarak ifade edilmiştir. Yapılan görüşmelerde, müfredat yenileme önerisi içerisinde en çok ifade edilen konu, yeni ders açılması yönünde olmuştur. Bu konu ile ilgili katılımcılardan biri şu ifadeler yer vermiştir;

“Büyük veri ile ilgili bir bölüm olmasa bile en azından bilgisayar mühendisliği, yazılım mühendisliği ve yönetim bilişim sistemleri gibi bilgisayar bilimlerine dokunan bölümlerde büyük veri dersleri konabilir.” (Kişi-9)

Bilgisayar bilimleri ve bilişim sistemleri gibi bölümlerden mezun olanlar, okudukları okulun müfredatlarında büyük veri analitiği dersi varsa veri analizi alanına katkıda bulunabilmektedir. Firmalar büyük veri analitiğine yatırım yapsalar da, verinin potansiyelini elde etmek için yeterli veri bilimci veya uzmanı yoktur. İyi bir müfredat tasarımı ile bilgisayar bilimleri ve bilişim sistemleri bölümlerinde okuyan öğrenciler, büyük veri analitiği alanına katkıda bulunabilir (Lawler ve Molluzzo, 2015). Lawler ve Molluzzo'nun ifade ettikleri gibi, görüşmeciler de büyük veri analitiği ile alakalı olarak ilgili bölümlerde yeni derslerin açılması ile öğrencilerin bu alanda daha nitelikli eğitim alabileceklerini ifade etmişlerdir.

Kurumların üzerinde durduğu diğer konu yeni açılan derslerle, verinin mühendisliğiyle baştan sona bir akış olarak görme konusunda belli başlı program ve araçların tanıtılmasıdır. Açık kaynak kodlu bu ürünlerin sürekli yenilerinin çıkması ile ilgili olarak hepsinin tek tek öğretilmesinden ziyade, işin temel mantığının verilmesi üzerinde durulmuştur. Dağıtık sistemler, Hadoop ekosistemi ile Spark, Hive gibi büyük veri araçları, NoSql veritabanları, MapReduce fonksiyonları, sanallaştırma, ACID-CAP gibi teoremler ve ölçeklenebilirlik gibi konular da kurumların dikkat çektikleri konular olmuştur.

İkinci öneri olarak, kurumların üzerinde durduğu konu staj olmuştur. Özellikle uzun dönem staja vurgu yapılmış, stajların nitelikli olması gerektiği ve karşılıklı fayda sağlanabilmesi için dönem stajı olmasa bile staj sürelerinin uzatılmasının iyi olacağı belirtilmiştir. Kısa dönem stajlarının şirkete yük olduğu, karşılıklı fayda sağlanamadığı ve stajyerin tam birşeyler öğrenmeye başladığı zaman staj süresinin bittiği ifade edilmiştir. Son öneri olarak, yüksek lisans programlarında büyük veri alanında daha çok program açılmasının faydalı olacağı belirtilmiştir.

Rensselaer Politeknik Enstitüsünden Peter Fox, eğitimcilerin veri bilimi yöntem ve tekniklerini öğretirken, belirli teknolojileri öğretmekten ziyade temel kavramları ve teorileri öğretmelerini tavsiye etmektedir. Lisans ve yüksek lisans düzeyinde veri

bilimi dersi veren Fox, tecrübelerinden başlangıçtan itibaren disiplinlerarası olmanın, teknik ve becerileri aynı anda geliştirmenin zor olduğunu buyüzden becerilerin sürekli olarak güçlendirilmesi gerektiğini, teknoloji değil, yöntem ve ilke öğretmenin ve işbirliği yapmanın önemine dikkat çekmektedir (Melody, 2014). Fox'un teknolojiden ziyade yöntemin öğretilmesi gerektiği konusunda görüşmecilerin de aynı tavsiyelere yer verdikleri görülmüştür.

5. TARTIŞMA VE SONUÇ

Araştırmada görüşmeler, Ankara'da büyük veri çözümleri sunan kamu ve özel sektörden küçük ve orta ölçekli kurumların büyük veri proje ekibi liderleriyle yapılmıştır. Büyük veri alanında çözüm üreten kurum sayısının azlığı ve kurumların iş yoğunluğu sebebiyle aranılan ölçütlerde katılımcı bulmak bu araştırmanın kısıtlılıklarından biridir. Küçük ölçekli işletmelerde genellikle genel müdürle görüşme sağlanmıştır. Bir kurumda, büyük veri projeleri ve diğer projeler birlikte yürütüldüğünden dolayı proje lideri yerine büyük veri projelerinde yer alan bir kişi ile görüşme yapılmıştır.

Bu çalışma ile büyük veri çözümleri sunan kurumların bu alanla ilgili mevcut durumları, beklentileri ve önerileri ortaya konulmuştur. Bu doğrultuda, büyük veri projelerinde çalışanların pozisyonu belirlenirken veri bilimci, veri mühendisi veya veri analisti gibi ayrımların kurumlar tarafından henüz net bir şekilde yapılamadığı ve genellikle pozisyonu yazılım mühendisi olarak tercih ettikleri sonucuna varılmıştır. Bu pozisyon ayrımlarının daha net yapılabilmesi için, veri bilimi gibi disiplinler arası alanlarda lisans programlarının açılması ve ilgili bölümlerde müfredatların bu teknolojileri ve konuları daha kapsamlı içerecek şekilde hazırlanması, kurumların pozisyonları belirlerken ve işe alımlarda daha net karar vermelerine yardımcı olacağı görülmektedir. Görüşülen kurumların kullandıkları araçlar (Apache Hadoop, Spark, Hive, Kafka, Cassandra), dünya genelinde kullanılan araçlarla (HDFS, Apache Hadoop, HBase, Cassandra, Hive, Oracle Big Data ve R) (Bhadani ve Jothimani, 2016) benzer olduğu sonucuna varılmıştır.

Kurumların bu alanda çalışmayı düşünen kişileri en çok bilgisayar mühendisliği, ardından matematik, istatistik, yönetim bilişim sistemleri, endüstri

mühendisliği gibi bölümlerden tercih ettikleri ortaya çıkmıştır. Ancak çoğu kurumun özel bir bölüm tercihi olmadığını belirtmeleri, belirtilen bölümlerde müfredatın belirli bir standartta olmadığı ve kişilerin bireysel çaba ve başarılarının daha çok dikkate alındığı sonucuna varılmıştır. Kurumlar, mezunlardan en çok alanla ilgili farkındalık beklemektedir, teknik olarak ise veri tabanı ve temel programlama bilgisinin olmasını beklemektedir. Burada, analitik yeteneğe sahip olmak ve sürekli değişen her yeni teknolojinin öğrenilmesinden ziyade, temel altyapının lisans eğitiminde alınmasının önemli olduğu sonucuna varılmıştır.

Kurumların önerileri doğrultusunda, öğrenciler için en çok ifade edilen farkındalığa sahip olma, üniversiteler için ise müfredat yenileme olarak belirtilmiştir. Görüşmecilerin büyük veri alanında öğrencilerden farkındalıktan sonra, okulda öğretilenler dışında ekstra çaba göstermelerini beklemeleri, üniversitelerde güncel konu ve teknolojilerin müfredatlarda yeteri kadar yer bulamadığını göstermektedir. Türkiye’de YÖK Atlas aramalarına göre veri bilimi ya da büyük veri temelli bir lisans programı yoktur. Ankara ilinde, 5 devlet, 7 özel, toplam 12 üniversitede bilgisayar mühendisliği programlarında, sadece 3 üniversitede (Hacettepe, TOBB ve Gazi Üniversitesi) veri bilimi ile ilgili zorunlu dersler bulunmaktadır. Teknik seçmeli olarak da sadece 2 üniversitede veri bilimi ve büyük veri ile ilgili dersler bulunmaktadır. Hacettepe ve TOBB Ekonomi ve Teknoloji Üniversitelerinde 2019-2020 eğitim öğretim yılı için ilk defa lisans düzeyinde Yapay Zekâ Mühendisliği bölümü açılmıştır ve bu bölümlerde büyük veri ile alakalı dersler müfredata eklenmiştir (YÖK Atlas, 2020). Hacettepe Üniversitesi Yapay Zekâ Mühendisliği bölümünde 4. yarıyılıda *Elements of Data Science* (Veri Biliminin Unsurları) ve *Data Science Lab.* (Veri Bilimi Laboratuvarı) ve TOBB Üniversitesi Yapay Zekâ Mühendisliği bölümünde 2. yarıyılıda *Veri Bilimine Giriş* dersi bulunmaktadır. Ankara ilinde, Hacettepe ve TOBB dışında veri bilimi ile alakalı zorunlu ders olarak sadece Gazi Üniversitesi bilgisayar mühendisliğinde (5. yarıyılıda *Introduction to Data Science* (Veri Bilimine Giriş) dersi bulunmaktadır. Teknik seçmeli olarak Bilkent ve Çankaya Üniversiteleri bilgisayar mühendisliği bölümlerinde *Veri Bilimine Giriş* dersi bulunmaktadır. Ankara ilinde yönetim bilişim sistemleri bölümü olan üç üniversitenin içinde yalnızca Başkent Üniversitesi’nde seçmeli *Python Programlama ve Veri Bilimi* dersi bulunmaktadır. Ankara ilindeki

üniversitelerden yola çıkılarak, üniversitelerde veri bilimi veya büyük veri ile ilgili derslerin birçok üniversitede olmadığı görülmektedir.

Gelecek çalışmalarda farklı sektörler bazında kıyaslamalarla veya ulusal düzeyde daha büyük bir örnekleme daha derin ve kapsamlı içgörülere ulaşılabilir. Bu çalışmada ortaya çıkan bulguların yıllara göre takip edilmesi, üniversitelerin ve YÖK'ün bu alandaki atılımlarının yıllara göre karşılaştırılması gelecek çalışmalarda gerçekleştirilebilir. Ayrıca üniversite yöneticileriyle yapılacak görüşmeler burada tartışılan ve elde edilen bulgularla karşılaştırılarak büyük veri alanında ortaya çıkan veya çıkabilecek üniversite-sektör arasındaki eşgüdüm sorunları ele alınabilir.

KAYNAKÇA

- Analytics Center & Kariyer.net. (2016). *Türkiye Analitik Yetenek ve Yetkinlik Araştırması*. İstanbul: Smartcom.
- Bhadani, A. K., & Jothimani, D. (2016). Big data: challenges, opportunities, and realities. In *Effective Big Data management and opportunities for implementation* (pp. 1-24). IGI Global.
- Corbin, J., & Strauss A. (2015). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications. Fourth Edition.
- Cukier, K. (2010). *Data, data everywhere*. The Economist. URL: <https://www.economist.com/special-report/2010/02/27/data-data-everywhere> Son Erişim Tarihi: 08.05.2019
- DataScience. (2020). *Data Science Colleges and Universities*. URL: <http://datascience.community/colleges> Son Erişim Tarihi: 14.03.2020
- Davis, K., & Patterson, D. (2012). *Ethics of Big Data: Balancing risk and innovation*. O'Reilly Media.
- Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013, May). Addressing big data issues in scientific data infrastructure. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 48-55). IEEE.
- DiFranza, A. (2019, 11 29). *The Biggest Data Analytics Challenge Of 2020*. URL: <https://www.northeastern.edu/graduate/blog/data-analytics-challenges/> Son Erişim Tarihi: 30.05.2020
- Dontha, R. (2017). Who came up with the name Big Data. *Data Science Central*, 13.
- Gahi, Y., Guennoun, M., & Mouftah, H. T. (2016, June). Big data analytics: Security and privacy challenges. In *2016 IEEE Symposium on Computers and Communication (ISCC)* (pp. 952-957). IEEE.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: strategies for qualitative research [pdf]. *Zugriff unter http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser_1967.pdf*.
- Gökşen, Y., & Aşan H. (2015). *Veri büyüklüklerinin veritabanı yönetim sistemlerinde meydana getirdiği değişim: NOSQL*. *International Journal of Informatics Technologies*, 8(3), 125.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information systems*, 47, 98-115.
- Jyothirmayee, A. V. N. S., Reddy, D. G., & Akbar, K. (2014). Understanding Big Data & DV2 law. *International Journal of Emerging Technology and Advanced Engineering*, 4(7).

- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.
- Lawler, J., & Molluzzo, J. C. (2015). A proposed concentration curriculum design for big data analytics for information systems students. *Information Systems Education Journal*, 13(1), 45.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Miller, S., & Hughes, D. (2017). The quant crunch: How the demand for data science skills is disrupting the job market. *Burning Glass Technologies*.
- Marzullo, K. (2016). *Administration Issues Strategic Plan For Big Data Research and Development*. URL: <https://obamawhitehouse.archives.gov/blog/2016/05/23/administration-issues-strategic-plan-big-data-research-and-development> Son Erişim Tarihi: 08.06.2019
- Mayring, P. (2011). *Nitel Sosyal Araştırmaya Giriş*. Ankara: BilgeSu Yayıncılık.
- Mellody, M. (2014). *Training Students to Extract Value from Big Data: Summary of a Workshop*. National Academies Press. 500 Fifth Street NW, Washington, DC 20001.
- National Science Foundation [NSF], (2019, 2 13). *Big Data Science and Engineering Program*. URL: <https://www.nsf.gov/pubs/2019/nsf19039/nsf19039.jsp> Son Erişim Tarihi: 27.06.2020
- Merriam, S. B. (2013). *Nitel araştırma desen ve uygulama için bir rehber* (S. Turan, Çev.) Ankara: Nobel Yayın Dağıtım.
- Rydning, D. R. J. G. J. (2018). The digitization of the world from edge to core. *Framingham: International Data Corporation*.
- Singh, D. S., & Singh, G. (2017). Big Data-A Review. *International Research Journal of Engineering and Technology*, 4(04), 2395-0056.
- Strateji Ve Bütçe Başkanlığı. (2018). *Stratejik Plan (2019-2023)*. Strateji Ve Bütçe Başkanlığı.
- Swanstrom, R. (2016, 11 28). Data Scientists, Data Engineers, Software Engineers: The Difference According to LinkedIn. *DataScience 101*.
- Trifu, M. R., & Ivan, M. L. (2014). Big Data: present and future. *Database Systems Journal*, 5(1), 32-41.
- TÜBİTAK. (2016). *Bulut Bilişim Ve Büyük Veri Çalıştayı 2015 Faaliyet Raporu*. Kocaeli: TÜBİTAK Bilgem.
- Warden, P. (2011). *Big Data Glossary*. Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo : O'Reilly Media.
- Wormer, P. V. (2014, 11 11). *A sense of urgency: Executives rush to adopt Big Data analytics*. URL: <http://info.totaltraxinc.com/blog/a-sense-of-urgency-executives-rush-to-adopt-big-data-analytics> Son Erişim Tarihi: 15.02.2021
- Yıldırım, A., & Şimşek, H. (2013). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. Seçkin Yayıncılık.
- YÖK Atlas. (2020). *Yapay Zeka Mühendisliği Programı Bulunan Tüm Üniversiteler*. URL: <https://yokatlas.yok.gov.tr/lisans-bolum.php?b=554009> Son Erişim Tarihi: 10.11.2020

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

