



**BÜYÜK VERİ EĞİTİMİ İHTİYACININ FARKINDALIĞINA YÖNELİK
ANKARA İLİNDE BİR ARAŞTIRMA**

Zeynep AYTAÇ

DOKTORA TEZİ

YÖNETİM BİLİŞİM SİSTEMLERİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

ŞUBAT 2021

Zeynep AYTAÇ tarafından hazırlanan “BÜYÜK VERİ EĞİTİMİ İHTİYACININ FARKINDALIĞINA YÖNELİK ANKARA İLİNDE BİR ARAŞTIRMA” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Yönetim Bilişim Sistemleri Ana Bilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Hasan Şakir BİLGE

Elektrik-Elektronik Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. Hadi GÖKÇEN

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Metehan TOLON

İşletme Bölümü Ana Bilim Dalı, Ankara Hacı Bayram Veli Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Prof. Dr. Zekai ÖZTÜRK

Sağlık Yönetimi Bölümü Ana Bilim Dalı, Ankara Hacı Bayram Veli Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Hakan ÇERÇİOĞLU

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 09/02/2021

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Aslıhan TÜFEKÇİ

Bilişim Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Zeynep AYTAÇ

09/02/2021

BÜYÜK VERİ EĞİTİMİ İHTİYACININ FARKINDALIĞINA YÖNELİK ANKARA İLİNDE BİR ARAŞTIRMA

(Doktora Tezi)

Zeynep AYTAÇ

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Şubat 2021

ÖZET

Kamu ve özel kurumlarda büyük veri ve araçlarının kullanımı ve büyük veri analizinden anlayan nitelikli personel ihtiyacı gün geçtikçe artmaktadır. Ancak Türkiye’de nitelikli personel ihtiyacını karşılayan üniversitelerde büyük veri eğitimi ihtiyacının farkındalığına yönelik bir çalışmaya rastlanmamıştır. Bu bağlamda, bu çalışma büyük veri farkındalığının tespit edilmesi doğrultusunda dört alt çalışmadan oluşmaktadır. İlk çalışmanın amacı, kamu ve özel kurumlarda büyük veri ve araçlarının kullanımını inceleyerek, kurumların bu alanda ihtiyaç duyduğu nitelikli personel açığını kapatmak için öğrencilerden ve üniversitelerden beklentilerini belirlemektir. Bu çalışmada nitel bir araştırma süreci tasarlanmış ve veriler kamu ve özel kurumların yöneticileri ile yapılan mülakatlar ile toplanmıştır. İkinci çalışmanın amacı, büyük veri araçlarını öğrenirken öğrencilerin deneyimlerini değerlendirmek ve karşılaştıkları zorlukları belirlemektir. Bu çalışmada da, nitel metodoloji izlenerek veriler, bir günlük büyük veri eğitimine katılan lisans öğrencilerinden, katılımlı gözlem tekniği ve katılımcıların öz-değerlendirmeleri ile toplanmıştır. Üçüncü çalışmanın amacı, farklı sektörlerin nitelikli personel ihtiyacını karşılayan üniversitelerde öğretim elemanlarının büyük veri farkındalık ve beklentilerinin analiz edilmesidir. Nicel olarak kurgulanan bu çalışmada veriler, anket yoluyla toplanmıştır. Dördüncü çalışmanın amacı, öğrencilerin büyük veri yeterlilik, farkındalık ve beklentilerini ölçmektir. Ayrıca bu çalışmada büyük veri farkındalık, yeterlilik ve beklenti ölçeği geliştirilmiştir. Nicel olarak kurgulanan bu çalışmada da veriler, anket yoluyla toplanmıştır. Üniversitelerde büyük veri alanında atılması gereken adımlar ve eğitim müfredatında yapılması gereken değişiklikler tartışılıp, öneri olarak sunulmuştur. Ayrıca elde edilen bulgular, On Birinci Kalkınma Planında Yükseköğretim Kurulu’nun sorumlu kuruluş olduğu politika tedbirleri ile birlikte değerlendirilerek yapıcı eleştiriler geliştirilmiştir.

Bilim Kodu : 92414

Anahtar Kelimeler : Büyük veri, farkındalık, yükseköğretim, eğitim

Sayfa Adedi : 155

Danışman : Prof. Dr. Hasan Şakir BİLGE

A RESEARCH IN ANKARA ON THE AWARENESS OF BIG DATA EDUCATION NECESSITY

(Ph. D. Thesis)

Zeynep AYTAÇ

GAZİ UNIVERSITY

INSTITUTE OF INFORMATICS

February 2021

ABSTRACT

The use of big data and tools in public and private institutions and the need for qualified personnel who understand big data analysis are increasing day by day. However, there is no comprehensive study on the awareness of big data education necessity in universities that meet the need for qualified personnel in Turkey. In this context, this study is composed of four sub-studies intended to detect big data awareness. The first study aims to examine the use of big data and tools in the public and private institutions and determine the industry's expectations from higher education and students for qualified personnel in this field. In the first study, a qualitative research process was designed, and data were collected through interviews with managers of public and private institutions. The second study evaluates students' experiences and identifies the difficulties they face while learning big data tools. In this study, the qualitative methodology was followed. The data were collected from undergraduate students participating in one-day big data training, participant observation, and self-assessments of participation. The third study aims to analyze the big data awareness and expectations of faculty members in universities that meet different sectors' qualified staff needs. In this study, data were collected through a questionnaire. The fourth study aims to measure students' big data competence, awareness, and expectations in higher education. In addition, a big data awareness, competence and expectations scale was developed in this study. In this quantitatively constructed study, the data were collected through a questionnaire too. It is integrated into the big data subject curriculum in higher education institutions in Turkey on evolutions to be made on issues discussed. It is presented these evolutions as suggestions. The findings are also evaluated in the Eleventh Development Plan with the policy measures that Higher Education Council is the responsible institution and is developed constructive criticisms.

Science Code : 92414
Key Words : Big data, awareness, higher education, education
Page Number : 155
Supervisor : Prof. Dr. Hasan Şakir BİLGE

TEŞEKKÜR

Tez çalışmam boyunca ilgi ve katkılarıyla beni yönlendiren değerli danışmanım Prof. Dr. Hasan Şakir BİLGE'ye kıymetli desteğinden dolayı çok teşekkür ederim. Tez izleme komitemde yer alan kıymetli hocalarım Prof. Dr. Hadi GÖKÇEN ve Doç. Dr. Metehan TOLON'a yönlendirmeleri ve değerli fikirlerinden dolayı teşekkürlerimi sunarım.

Doktora eğitimim süresince her zaman yanımda olduğunu hissettiğim değerli eşim Dr. Araştırma Görevlisi Muhammed Bilgehan AYTAÇ'a ve hayata gözlerini açtığı günden beri varlığı ile bana mutluluk veren biricik kızım Fatma Ece AYTAÇ'a gösterdikleri sabırdan ve verdikleri manevi destekten dolayı teşekkür ederim.

Tüm hayatım boyunca destek ve dualarıyla yanımda olan sevgili anneme, babama ve kardeşlerime şükranlarımı sunarım.

Gazi Üniversitesi'nde gerçekleştirdiğimiz bir günlük eğitimde, katılımcılara eğitim vererek değerli vaktini ayıran Bilgisayar Mühendisi Erkan ŞİRİN'e, ayrıca görüşme gerçekleştirdiğim kurumlarda değerli vakitlerini ayırarak desteklerini esirgemeyen kıymetli Genel Müdür ve Proje Müdürlerine çok teşekkür ederim. Eğitime katılan değerli öğrencilere ve ankete katılan değerli öğretim elemanları ve öğrencilere de teşekkürü bir borç bilirim.

İÇİNDEKİLER

Sayfa

ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ	x
ŞEKİLLERİN LİSTESİ	xii
SİMGELER VE KISALTMALAR	xiii
1. GİRİŞ.....	1
2. LİTERATÜR.....	5
2.1. Büyük Veri	5
2.1.1. Büyük veri tarihi ve tanımı.....	5
2.1.2. Büyük veri karakteristikleri.....	8
2.1.3. Büyük veri türleri	13
2.1.4. Büyük veri kaynakları	14
2.1.5. Büyük veri değer zinciri	15
2.1.6. Büyük veri altyapısı ve Hadoop ekosistemi	17
2.1.7. NoSql veri tabanları	23
2.1.8. Büyük veri teknolojileri	25
2.2. Büyük Veri Kullanım Alanları	27
2.3. Büyük Veri Pazar Hacmi.....	33
2.4. Büyük Veri ve İnsan Kaynağı İhtiyacı	36
2.5. Kamuda Büyük Veri.....	38
2.5.1. Türkiye’de büyük veri ile ilgili planlar	40
2.5.2. Dünyada büyük veri ile ilgili planlar.....	42
2.6. Büyük Veri ve Eğitim.....	45
2.6.1. Yükseköğretimde büyük veri	46
2.6.2. Yükseköğretimde büyük veri eğitimi	50
3. TEZ KAPSAMINDAKİ ÇALIŞMALARIMIZ.....	55
3.1. Çalışmalara Genel Bakış	55

	Sayfa
3.1.1. Çalışmaların amacı ve kapsamı.....	55
3.1.2. Çalışmaların önemi	58
3.2. Çalışma 1 – Kamu ve Özel Kurumların Büyük Veri Alanında Beklentilerinin Belirlenmesi.....	59
3.2.1. Amaç	59
3.2.2. Araştırma deseni.....	59
3.2.3. Örneklem.....	60
3.2.4. Veri toplama.....	62
3.2.5. Araştırmanın kısıtlılıkları	62
3.2.6. Verilerin analizi.....	63
3.2.7. Güvenilirlik ve geçerlilik	65
3.2.8. Bulgular.....	67
3.3. Çalışma 2 - Büyük Veri Eğitiminde Öğrencilerin Deneyimleri ve Karşılaştıkları Zorluklar	84
3.3.1. Amaç	84
3.3.2. Araştırma deseni.....	84
3.3.3. Örneklem.....	85
3.3.4. Veri toplama.....	86
3.3.5. Araştırmanın kısıtlılıkları	87
3.3.6. Verilerin analizi.....	88
3.3.7. Güvenilirlik ve geçerlilik	89
3.3.8. Bulgular.....	91
3.4. Çalışma 3 – Öğretim Elemanlarının Büyük Veri Beklentilerinin Değerlendirilmesi	97
3.4.1. Amaç	97
3.4.2. Örneklem.....	97
3.4.3. Veri toplama.....	98
3.4.4. Araştırmanın kısıtlılıkları	99
3.4.5. Verilerin analizi.....	99
3.4.6. Bulgular.....	99

Sayfa

3.5. Çalışma 4 – Lisans Öğrencilerinin Büyük Veri Yeterlilik, Farkındalık ve Beklentilerinin Ölçülmesi.....	104
3.5.1. Amaç	104
3.5.2. Örneklem.....	105
3.5.3. Veri toplama	106
3.5.4. Araştırmanın kısıtlılıkları	107
3.5.5. Verilerin analizi.....	107
3.5.6. Güvenilirlik ve geçerlilik	108
3.5.7. Bulgular.....	111
4. GENEL BULGULAR VE MÜFREDAT ÖNERİSİ.....	119
4.1. Yapılan Çalışmalardan Elde Edilen Genel Bulgular	119
4.2. Müfredat Önerisi	121
4.3. On Birinci Kalkınma Planında YÖK’ün Sorumlu Kuruluş Olduğu Politika Tebirleri ve Geliştirilen Büyük Veri İle İlgili Öneriler	125
5. SONUÇ VE ÖNERİLER	129
KAYNAKLAR	135
EKLER	147
EK-1. Gazi Üniversitesi Ölçme Değerlendirme Etik Alt Çalışma Grubu kararı.....	148
EK-1. (devam) Gazi Üniversitesi Ölçme Değerlendirme Etik Alt Çalışma Grubu kararı.....	149
EK-2. Öğretim elemanları aşına olduğu ve kullandığı büyük veri yazılımları	150
EK-3. Öğrencilerin büyük veri farkındalık, yeterlilik ve beklenti ölçeği anketi.....	151
EK-3. (devam) Öğrencilerin büyük veri farkındalık, yeterlilik ve beklenti ölçeği anketi	152
EK-4. Öğrencilerin aşına oldukları büyük veri yazılımları	153
ÖZGEÇMİŞ.....	154

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Çeşitli kaynaklarda büyük veri tanımları	8
Çizelge 2.2. Geleneksel veri ve büyük veri arasındaki farklılıklar	9
Çizelge 2.3. Büyük verinin 5V karakteristik özellikleri.....	9
Çizelge 2.4. Farklı veri kaynakları	14
Çizelge 2.5. Büyük veri değer zinciri.....	15
Çizelge 2.6. Büyük veri yetenekleri ve öncelikli teknolojiler.....	27
Çizelge 2.7. Türkiye’de büyük veri çalışmalarına veri sağlayabilecek kamu projeleri	38
Çizelge 3.1. Araştırılan çalışmaların detayları	57
Çizelge 3.2. Görüşülen kurum ve görüşmecilerin profili	61
Çizelge 3.3. Görüşmelerden sonra oluşturulan kod sayısı ile toplam kod sayısı	64
Çizelge 3.4. Görüşülen kurumların detaylı bilgileri.....	68
Çizelge 3.5. Görüşmecilerden elde edilen veriler doğrultusunda belirlenen kategoriler	69
Çizelge 3.6. Kurumların veri bilimi alanında tercih ettikleri bölümler.....	76
Çizelge 3.7. Katılımcıların demografik özellikleri.....	91
Çizelge 3.8. Araştırmanın kavramsal çerçevesinde belirlenen temalar	92
Çizelge 3.9. Öğretim elemanlarının tanımlayıcı istatistikleri	98
Çizelge 3.10. Öğretim elemanlarının büyük veri yazılımlarına aşinalıkla ilgili tanımlayıcı istatistikleri.....	100
Çizelge 3.11. Büyük veri alanına yönelik akademik ilgi	101
Çizelge 3.12. Büyük veri konusunun bahsedildiği dersler	101
Çizelge 3.13. Öğretim elemanlarının tek örneklem t-testi analizi.....	103
Çizelge 3.14. Öğrencilerin tanımlayıcı istatistikleri.....	106
Çizelge 3.15. Büyük veri eğitim ve beklenti ölçeği güvenilirlik ve geçerlilik analizi	109
Çizelge 3.16. Büyük veri kavramı ve yazılımlarına aşinalıkla ilgili tanımlayıcı istatistikler.....	111
Çizelge 3.17. Büyük veri eğitim ve beklenti (tek örneklem t-testi)	112
Çizelge 3.18. Bölümlere göre eğitim ve beklenti (bağımsız örneklem t-testi)	114
Çizelge 3.19. Büyük veri farkındalık ve yeterlilik (tek örneklem t-testi)	116
Çizelge 3.20. Bölümlere göre farkındalık ve yeterlilik (bağımsız örneklem t-testi).....	117

Çizelge**Sayfa**

Çizelge 4.1. Yapılan çalışmalardan elde edilen bulguların özeti	120
Çizelge 4.2. Veri bilimi müfredat önerisi.....	124
Çizelge 4.3. 2019-2023 Stratejik Planındaki büyük veri ile ilgili hedef ve ihtiyaçlar	125
Çizelge 4.4. On Birinci Kalkınma Planı'nda YÖK'ün sorumlu kuruluş olduğu politika tedbirleri ve büyük veri ile ilgili geliştirilen öneriler.....	126



ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Yayımlanmış toplam makale sayısı (1996-2015)	7
Şekil 2.2. Küresel veri miktarının yıllık büyülüğü.....	10
Şekil 2.3. Hadoop ekosistemi	18
Şekil 2.4. Map-Reduce fonksiyonları şematik yapısı.....	20
Şekil 2.5. En popüler NoSql veritabanı sıralaması.....	25
Şekil 2.6. Sektörlere göre büyük veri benimsenmesi	28
Şekil 2.7. Dünya çapında büyük veri pazarı gelir tahmini (milyar ABD doları), 2011-2027	35
Şekil 2.8. Dünya çapında Hadoop ve büyük veri pazar hacmi, 2017-2022	35
Şekil 3.1. Çalışmanın kapsamı	56
Şekil 3.2. Çalışma verisi üzerinden nitel araştırmalarda koddan kategoriye akış modeli	63
Şekil 3.3. Görüşmelerin gerçekleştirilme süreci	66
Şekil 3.4. Kurumların büyük veri alanında çalışanlara verdikleri pozisyon isimleri	71
Şekil 3.5. Kurumların büyük veri alanı için tercih ettikleri bölümler	75
Şekil 3.6. Kurumların mezunlardan beklentileri	77
Şekil 3.7. Yazılım mühendisliği, veri mühendisliği ve veri bilimcinin teknik yetenekleri.....	79
Şekil 3.8. Kurumların öğrencilere önerileri	80
Şekil 3.9. Kurumların üniversitelere önerileri.....	82
Şekil 3.10. Eğitimin nitel izleme akış modeli	85

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
5V	Hacim, Hız, Çeşitlilik, Gerçeklik, Değer (Volume, Velocity, Variety, Veracity, Value)
ABD	Amerika Birleşik Devletleri
ACID	Bölünmezlik, Tutarlılık, Yalıtım, Dayanıklılık (Atomicity, Consistency, Isolation, Durability)
ACM	Bilgi İşlem Makineleri Derneği (Association for Computing Machinery)
B3LAB	Bulut Bilişim ve Büyük Veri Laboratuvarı
BARC	Bhabha Atom Araştırmaları Merkezi (Bhabha Atomic Research Center)
BDA	Büyük Veri ve Analitiği (Big Data and Analytics)
BVYS	Büyük Veri Yönetim Sistemleri
CAP	Tutarlılık, Ulaşılabilirlik, Bölünebilme Toleransı (Consistency, Availability, Partition Tolerance)
CERN	Avrupa Nükleer Araştırma Örgütü (European Organization for Nuclear Research)
CIO	Bilişim Kurulu Başkanı (Chief Information Officer)
ETL	Ayıklama, Dönüştürme, Yükleme (Extract, Transform, Load)
IDC	Uluslararası Veri Kurumu (International Data Corporation)
İTÜ	İstanbul Teknik Üniversitesi
IEEE	Elektrik ve Elektronik Mühendisleri Enstitüsü (Institute of Electrical and Electronics Engineers)
IoT	Nesnelerin İnterneti (Internet of Things)
MEB	Milli Eğitim Bakanlığı
MHRS	Merkezi Hekim Randevu Sistemi

Kısaltmalar**MOOC****NSA****NSF****ÖSYM****SGK****STEM****TBD****TÜBİTAK****YBS****YÖK****XML****Açıklamalar**

Kitlesel Açık Çevrimiçi Kurslar

(Massive Open Online Courses)

Ulusal Güvenlik Ajansı (National Security Agency)

Ulusal Bilim Vakfı (National Science Foundation)

Ölçme, Seçme ve Yerleştirme Merkezi

Sosyal Güvenlik Kurumu

Bilim, Teknoloji, Mühendislik, Matematik

(Science, Technology, Engineering, Mathematics)

Türkiye Bilişim Derneği

Türkiye Bilimsel ve Teknolojik Araştırma Kurumu

Yönetim Bilişim Sistemleri

Yükseköğretim Kurulu

Genişletilebilir İşaretleme Dili

(Extensible Markup Language)

1. GİRİŞ

Büyük veri, bilişim ve teknoloji dünyasında özellikle son yıllarda çok fazla adından söz edilen ve ilgi duyulan bir araştırma alanı olmuştur. Teknolojinin hızla ilerlemesi, internetin gelişmesi ve sosyal medya kullanımının artış göstermesi gibi birçok faktörle beraber verinin büyüklüğü de hızla artmaktadır. Verinin büyüklüğünün çok hızlı bir şekilde artmasının nedenlerinin başında internet ve sosyal medya üzerinden erişilebilen kullanıcıların ürettiği veri gelmektedir. Bunun dışında sağlık uygulamalarından gelen veriler, sensör verileri, sunuculardan alınan log kayıtları da farklı kaynaklardan gelen verilere örnek olarak verilmektedir. Uluslararası Veri Kurumu (IDC)¹, dünyadaki verilerin her iki yılda bir, ikiye katlandığını iddia etmektedir. IDC istatistiklerine göre, 2018'den 2025'ye kadar ulaşılacak veri miktarının 33 zetabayt² 175 zetabayta kadar artacağı öngörülmektedir (Rydning, 2018). Verinin eksponansiyel (üstel) olarak büyümesi iş dünyasında dolayısıyla da veri tabanı sistemlerinde, veriye erişme, saklama ve 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 veri tabanı sistemlerinin farklı kaynaklardan gelen, çok hızlı büyüyen ve farklı formatlarda olan bu veriler üzerinde işlem yapma kapasitesi yeterli olamamaktadır. IBM³'de yapılan bir çalışmaya göre günümüz iş liderlerinin yarısından fazlası, yapması gereken işleri kavrama ve öngörü erişimlerinin olmadığı kanaatinde olduklarını belirtmişlerdir (Zikopoulos, Eaton, deRoos, Deutsch ve Lapis, 2012). Bu eksiklikler veriye alternatif yöntemlerle yaklaşmayı gerektirdiğinden büyük veri kavramı ve sistemleri ortaya çıkmıştır.

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

1 IDC: International Data Corporation

2 1 Zetabayt = 1,024⁷ bayt

3 1911 yılında ABD'de kurulan uluslararası teknoloji şirketi

4 1990 yılında kurulan McKinsey şirketinin işletme ve araştırma enstitüsüdür.

(Manyika ve diğ erleri, 2011). Obama Y netimi, b y k veri teknolojilerini geliřtirmek, b y k veri uygulamalarını g stermek ve yeni nesil veri bilimcilerini eēitmek i in 2012 yılında ‘B y k Veri Arařtırma ve Geliřtirme Giriřimi’ni bařlatmıř, aynı zamanda Beyaz Saray’a ilk kez Veri Bilimi Y neticisi iře almıřtır (Marzullo, 2016). Amerika Birleřik Devletleri’nde  zellikle yayınlanan raporlar ve h k metin bu alana verdiēi  nem sonucunda, eēitim alanında b y k veri ile ilgili geliřmeler bařlamıřtır. ABD’deki kolejler ve  niversiteler tarafından y r t len federal destekli arařtırmaların bir kısmını finanse eden Amerika Birleřik Devletleri Ulusal Bilim Vakfı, 2012 yılında kendi bařına bir disiplin olarak kurulan veri bilimine, b y k veri uygulamaları programını d hil ederek bu alana katkıda bulunmuřtur (National Science Foundation [NSF], 2019).

B y k verinin deēeri genellikle daha iyi kararlar alma yeteneēinden kaynaklanır (The Economist, 2012). Firmalar, s re lerin,  r nlerin ve hizmetlerin sonu larını optimize etmek i in b y k veri analiziyle a ık a ilgilenmektedirler (Kim, Trimi ve Chung, 2014). H k met ve saēlık end strisi de, s re lerin ve hizmetlerin sonu larını optimize etmek i in verimlilik, iyileřtirme ve maliyet kontrol  alanlarında b y k veri analizinde giriřimlerde bulunmaktadır (Liyakasa, 2013). řirketler b y k veri analizine yatırım yapıyor olsa da verilerin potansiyelini ortaya  ıkarmak i in yeterli veri bilimcisi veya uzmanına sahip deēillerdir. Bilgisayar bilimleri ve bilgi sistemleri b l mlerinden mezun olanlar, okul m fredatında b y k veri analizi dersi varsa ve hesaplamalı matematiksel ve istatistiksel becerilere sahiplerse veri analizi alanına katkıda bulunabilirler (May, 2013; Hulme, 2014). Bilgisayar Bilimleri, Bilgisayar/Yazılım M hendisliēi ve Bilgi Sistemleri b l mlerinde, disiplinlerin hızla deēiřen doēasından  t r , m fredat ve eēitim programlarının g ncel kalabilme konusunda endiře duyulmaktadır. Teknolojilerin, sekt r n ve ilgili mesleklerin dinamik doēası gereēi, bu alanlardaki eēitim m fredatları baskı altına alınmaktadır. Bununla birlikte, resmi m fredatın statik yapısı hızlı deēiřim g stermezken, teknoloji, ara lar ve uygulamalar geniřlemeye devam etmektedir. B y k veri de son on yılda  ok hızlı oranlarda b y yen teknolojileri i ermektedir (Okur ve B y kke eci, 2014).

B y k veri kamu ve  zel kurumlarda yaygın bir řekilde kullanılmakta ve b y k veri analizinden anlayan nitelikli personel ihtiyacı da artmaktadır. 2015 yılında ABD’deki Temel Veri Bilimi ve Analitiēi iřleri i in 2 milyon 350 binden fazla iř ilanı vardı ve 2020

yılına kadar 2 milyon 720 bine yakın iş ilanı olacağı tahmin edildi. Veri Bilimi Analitiği alanında çalışan sayısına olan arzın, talebin arkasında tehlikeli bir şekilde arttığına dair endişeler artmaktadır. Ağ ve Bilgi Teknolojileri Araştırma ve Geliştirme (NITRD)⁵ programında, daha geniş işgücü için hem derin analitik yetenek hem de analitik kapasite için ABD’de ulusal ortamın iyileştirilmesi gerektiği belirtilmektedir. Eğitimciler ve eğitim sağlayıcıları, öğrencileri bugünün ve yarının analitikle ilgili pozisyonlarına hazırlayan programlarla analitik becerilere yönelik artan talebe proaktif bir şekilde yanıt vermek zorundadır (Miller ve Hughes, 2017). Ancak Türkiye’de ulusal düzeyde büyük veri ve veri bilimi alanlarında eğitim sektöründe analitik yeteneğe sahip işgücünün iyileştirilmesine yönelik adımlara çok rastlanılmamaktadır.

Bu çalışmada, kamu ve özel kurumların büyük veri ve yazılımlarını benimseme düzeyleri, yükseköğretim kurumları ve öğrencilerden beklentileriyle, büyük veri alanındaki akademik personel ve öğrencilerin farkındalık ve beklentilerinin tespit edilmesi amaçlanmaktadır. Bu doğrultuda, dört alt çalışmadan oluşan geniş kapsamlı bir yapı tasarlanmıştır.

İlk çalışmanın amacı, Türkiye’de kamu ve özel kurumlarda büyük veri yazılımlarının kullanımını inceleyerek, bilgi teknolojileri sektörünün beklentilerinin ne yönde olduğunu belirlemektir. Bu doğrultuda Türkiye’de en çok tercih edilen büyük veri araçlarını ortaya koymak ve sektörün bu alanda ihtiyaç duyduğu nitelikli personel açığını kapatmak için öğrencilerden ve üniversitelerden beklentilerini belirlemektir. Bu çalışmada nitel bir araştırma süreci tasarlanmıştır ve veriler kamu ve özel kurumların yöneticileri ile yapılan mülakatlar ile toplanmıştır. İkinci çalışmanın amacı, büyük veri yazılımlarını öğrenirken öğrencilerin deneyimlerini değerlendirmek ve karşılaştıkları zorlukları belirlemektir. Bu çalışmada da, nitel metodoloji izlenerek veriler, bir günlük büyük veri eğitimine katılan lisans öğrencilerinden, katılımlı gözlem tekniği ve katılımcıların öz-değerlendirmeleri ile toplanmıştır. Üçüncü çalışmanın amacı, farklı sektörlerin nitelikli personel ihtiyacını karşılayan üniversitelerde öğretim elemanlarının büyük veri farkındalık ve beklentilerinin analiz edilmesidir. Nicel olarak kurgulanan bu çalışmada veriler, anket yoluyla toplanmıştır. Dördüncü çalışmanın amacı, üniversitelerde öğrencilerin büyük veri yeterlilik, farkındalık ve beklentilerini ölçmektir. Nicel olarak kurgulanan bu çalışmada da veriler, anket yoluyla toplanmıştır.

⁵ NITRD: Ağ ve Bilgi Teknolojileri Araştırma ve Geliştirme Programı, ulusun bilgisayar, ağ ve yazılımda ileri bilgi teknolojileri üzerinde federal olarak finanse edilen çalışma kaynağıdır.

Bir sonraki bölümde büyük veri alanı ile ilgili literatür taraması yer almaktadır. Ayrıca büyük verinin kullanım alanları, pazar hacmi, büyük veri alanında insan kaynağı ihtiyacı, kamuda büyük veri çalışmaları ve eğitimde büyük veri konularına yer verilmiştir. Üçüncü bölümde dört temel çalışma gerçekleştirilmiş ve elde edilen bulgular literatürdeki bulgularla tartışılarak değerlendirilmiştir. Dördüncü bölümde, gerçekleştirilen dört ayrı çalışmadan elde edilen genel bulgulara, geliştirilen örnek bir Veri Bilimi lisans müfredatına ve On Birinci Kalkınma Planında YÖK'ün sorumlu kuruluş olduğu politika tedbirleri ve geliştirilen büyük veri ile ilgili önerilere yer verilmiştir. Sonuç ve öneriler bölümünde, elde edilen sonuçlar ve Türkiye'de üniversitelerde, büyük veri alanının eğitim müfredatına entegre edilmesi hususunda yapılması gereken değişiklikler ve yenilikler tartışılıp, öneri olarak sunulmuştur.

2. LİTERATÜR

2.1. Büyük Veri

Bu bölümde; büyük veri tarihi ve tanımı, büyük verinin karakteristikleri, büyük veri türleri, büyük veri kaynakları, büyük veri değer zinciri, büyük veri altyapısı (mimarisi) ve büyük veri araçları ile ilgili genel bilgiler verilmektedir.

2.1.1. Büyük veri tarihi ve tanımı

Büyük veri ifadesinin kavramsallaşması, bilgi patlaması (information explosion) ifadesinin ilk defa 1941 yılında Oxford İngilizce Sözlüğü tarafından kullanılmasıyla başlamıştır. 1944 yılında Fremont Rider'ın, Amerikan üniversite kütüphanelerinin 16 yılda bir iki katına çıktığını ileri sürdüğü Bilimsel Araştırma Kütüphanesinin Geleceği adlı makalesindeki ifadelerle göre, 2040 yılında Yale Üniversitesinin kütüphanesinde 6.000 milden⁶ fazla rafı işgal edecek 200.000.000 cilt olacağını ve bunlar için 6000'den fazla çalışanın olması gerektiğini ileri sürmüştür (Press, 2013). Bilimsel bilginin büyümesi, bilimsel bildiri ve dergi sayısındaki büyümeye bakılarak listelenmiş ve yeni dergilerin sayısının lineer olmaktan ziyade üstel olarak arttığı, 15 yılda bir iki katına çıktığı ve her yarım yüzyılda on kat arttığı sonucuna varılmıştır (Solla-Price, 1975). Amerika Birleşik Devletleri'ndeki gelişmeler dışında, Japonya'nın Posta ve Haberleşme Bakanlığı, Japonya'da dolaşan bilgi hacmini takip etmek için Bilgi Akışı Sayımı projesini yürütmeye başlamış ve 1975 yılında yapılan sayıma göre, bilgi üretiminin, bilgi tüketiminden çok daha hızlı arttığını tespit etmiştir (Hilbert, 2012). Bu yıllarda, bilgiyi üretme, iletme ve bilgiye erişme kolaylığı arttıkça, bilgiyi depolama ihtiyacı da artmıştır. 1981 yılında Macar Merkez İstatistik Ofisi, ülkenin bilgi endüstrilerini hesaba katarak, bilgi hacmini bit⁷ cinsinden ölçmek için bir araştırma projesi başlatmıştır (Hilbert, 2012). Bilginin hızlı bir şekilde artmasıyla, depolamanın da hacmi hızla artmış ve yeni depolama alanları ortaya çıkmıştır. Dijital veri depolama, kâğıt arşivleri kullanmaktan çok daha uygun maliyetli hale gelmiştir. 100.000 sayfayı fiziksel olarak saklamak için ortalama on adet dosya dolabı gerekirken, benzer miktardaki bilgi yalnızca bir DVD'ye sığdırılabilir hale gelmiştir (Morris ve Truskowski, 2003). 'Büyük Veri' ifadesi ilk kez Nasa araştırmacıları Michael Cox ve David Ellsworth

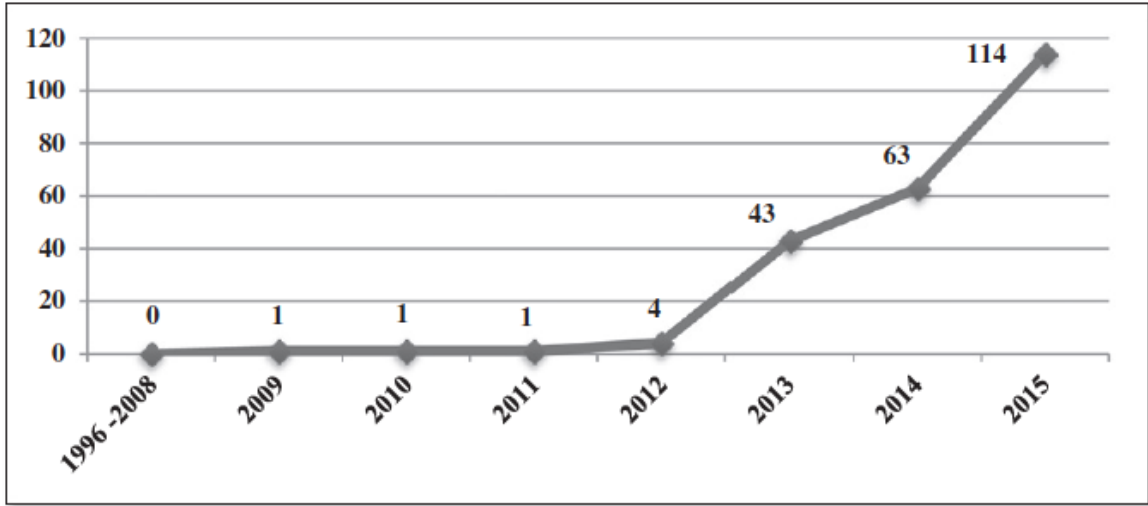
⁶ 1 mil = 1,6 km

⁷ Bilgisayar bilimlerinde 0 ve 1'i temsil eder. 8 bit bir araya gelerek 1 baytı oluşturur.

tarafından, 1997 yılında o yılların en yaygın sorununu tarif ederken, süper bilgisayarların büyük miktarlarda veri üretmelerinden bahsederek kullanılmıştır (Friedman, 2012). Cox ve Ellsworth (1997), yayınladıkları makalede şu ifadelerle yer vermişlerdir; “Veri setleri genellikle oldukça büyüktür ve ana bellek, yerel disk ve remote diskin kapasitelerini gösterir. Buna büyük veri sorunu diyoruz.” Lyman ve Varian (2000), bilgisayar depolama açısından dünyada yıllık olarak oluşturulan ve dört fiziksel ortamda (kâğıt, film, optik, manyetik) depolanan yeni ve orijinal bilgilerin toplam miktarını ölçmek için ilk defa kapsamlı bir çalışma yapmıştır. Web 2.0⁸ ile birlikte sosyal medyanın ve akıllı telefonların kullanılmaya başlanması ile üretilen veri miktarı da hızla artmaya başlamıştır. Yaygın olarak erişilen ve çok sayıda kişinin bağımsız olarak çok çeşitli veriler üretmesine izin veren uygulamalar da oluşturulmaya başlanmıştır. Kullanıcıların ürettikleri verilerin kaynağı olan sosyal ağlardan Facebook 2004 yılında, Youtube 2005 yılında, Twitter 2006 yılında ve Instagram 2010 yılında kurulmuştur (Devakunchari, 2014). 2007 yılının Mart ayında IDC’nin, “Genişleyen Dijital Evren: 2010 Yılına Kadar Dünya Çapında Bilgi Büyümesi Tahmini” başlıklı bildirisi, her yıl oluşturulan ve çoğaltılan dijital veri miktarını tahmin eden ilk çalışma olmuştur (Press, 2013). IDC’nin verilerine göre, 2006 yılında bütün dünyada 161 eksabayt⁹ veri oluşturulmuştur (Gantz ve diğerleri, 2007). Aynı çalışmanın 2010 ve 2012 yıllarındaki yayınlarına göre, yıllık oluşturulan dijital veri miktarının öngördükleri miktarı aştığını, 2010’da 1227 eksabayta ve 2012’de 2837 eksabayta ulaştığını belirtmişlerdir (Gantz ve Reinsel, 2010, 2012). Gandomi ve Haider (2015), ProQuest araştırma kütüphanesinde büyük veri terimini içeren belgelerin sıklık oranlarının yıllara oranla değişimini gösterirken, büyük veri teriminin 2011 yılında yaygınlaştığını ifade etmektedir. İşletme ve Yönetim, Bilgisayar Bilimi, Karar Bilimi ve Sosyal Bilimler alanlarında, büyük veri ve analitiği konusunda 1996’dan 2015 yılına kadar yayınlanan makalelerin yıllara oranla artışı incelenmiştir. Şekil 2.1’de, 2012 yılından sonra büyük veri ve büyük veri analitiği araştırma alanlarında, makalelerin sayısında ciddi bir artış olduğu gösterilmektedir (Sivarajah, Kamal, Irani ve Weerakkody, 2017).

⁸ Web 2.0: 2004’de ortaya çıkan terim, internet kullanıcılarının kendi içeriklerini ürettikleri interaktif bir yapıdadır. Sosyal ağ siteleri bu teknoloji içerisinde kullanılmaya başlanmıştır.

⁹ 1 Eksabayt = 1,024⁶ bayt



Şekil 2.1. Yayımlanmış toplam makale sayısı (1996-2015) (Sivarajah ve diğerleri, 2017)

Büyük veri, geleneksel veri tabanı teknolojileriyle saklanması, işlenmesi ve analiz edilmesi zor olan veri hacmindeki artışı ifade etmek için kullanılan bir terimdir (Hashem ve diğerleri, 2015). Büyük veri kelime olarak üretilmiş çok miktarda veri anlamındadır. McKinsey Global Enstitüsü raporunda büyük veri, “Boyutu, geleneksel veri tabanı yazılım araçlarının kaydettiği, yönettiği ve analiz ettiği yeteneklerin ötesinde bir boyuta sahip olan veri setleri” olarak tanımlanmıştır (Manyika ve diğerleri, 2011). Gartner¹⁰ analistlerinden Doug Laney, 2001’de yapılan araştırma raporuna göre büyük veriyi üç boyutuyla (3V: Volume, Velocity, Variety) ele alan ilk isim olmuştur (Laney, 2001). Daha sonra 2012’de, Gartner tanımını yüksek hacimli, çok çeşitli ve yüksek hızlı olarak güncellemiştir (Laney, 2012). 2005 yılında O’Reilly Media’dan ¹¹ Roger Mougallas, 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 ise “SQL gibi geleneksel veri tabanı protokolleri tarafından analiz edilemeyecek kadar büyük boyutlarda olan veri” olarak tanımlamışlardır (Davis ve Patterson, 2012:4). Çizelge 2.1’de farklı kaynaklarda belirtilen büyük veri tanımları listelenmiştir.

¹⁰ 1979 yılında kurulan araştırma ve danışmanlık şirketi

¹¹ Tim O’Reilly tarafından kurulmuş olan bu şirket, kitap yayınlamakta, teknik konferanslar düzenlemekte ve çevrimiçi öğrenme platformu sağlamaktadır.

Çizelge 2.1. Çeşitli kaynaklarda büyük veri tanımları (De Mauro, Greco ve Grimaldi, 2015)

Kaynak	Tanım
(Beyer ve Laney, 2012)	Gelişmiş görüş ve karar alma için yenilikçi, düşük maliyetli bilgi işleme biçimleri talep eden yüksek hacim, hız ve çeşitlilikte bilgi varlıklarıdır.
(Dijcks, 2012)	Büyük veriyi tanımlayan dört özellik hacim, hız, çeşitlilik ve değerdir.
(Intel, 2012)	Karmaşık, yapılandırılmamış veya büyük miktarda verilerdir.
(Suthaharan, 2013)	Üç veri özelliği kullanılarak tanımlanabilir: nicelik, süreklilik ve karmaşıklık
(Schroeck ve diğerleri, 2012)	Büyük veri, kuruluşların günümüzün dijitalleşen pazarda rekabet avantajını kazanması için bir fırsat yaratan hacim, çeşitlilik, hız ve gerçeklik kombinasyonudur.
(NIST Big Data Public Working Group, 2014)	Verimli depolama, manipülasyon ve analiz için ölçeklenebilir bir mimari gerektiren hacim, hız ve çeşitlilik özelliklerinde büyük veri kümeleridir.
(Ward ve Barker, 2013)	NoSql, Map-Reduce ve Makine Öğrenmesi tekniklerinin ve daha fazlasının kullanıldığı büyük ve karmaşık veri kümelerinin analizi ve depolanmasıdır.
(Microsoft, 2013)	Büyük ve karmaşık bilgi kümelerine makine öğrenmesi ve yapay zeka kullanılarak bilgi işlem uygulama sürecidir.
(Dumbill, 2013)	Geleneksel veritabanı sistemlerinin işlem kapasitesini aşan verilerdir.
(Fisher ve diğerleri, 2012)	Kolayca idare edilemeyen ve işlenemeyen verilerdir.
(Shneiderman, 2008)	Bir ekrana sığmayacak kadar büyük veri kümesidir.
(Manyika ve diğerleri, 2011)	Boyutu, tipik veritabanı yazılımı araçlarının tutma, saklama, yönetme ve analiz etme yeteneğinin ötesinde olan veri kümeleridir.
(Chen ve diğerleri, 2012)	İleri ve benzersiz veri depolama, yönetim, analiz ve görselleştirme teknolojileri gerektirecek kadar büyük ve karmaşık uygulamalardaki veri kümeleri ve analitik teknikleridir.
(Boyd ve Crawford 2012)	Teknoloji, analiz ve mitoloji etkileşimine dayanan kültürel, teknolojik ve bilimsel bir olgudur.
(Mayer-Schönberger ve Cukier, 2013)	Toplumu nasıl anladığımızı ve örgütlediğimizi, dönüştüren bilgileri analiz etme yöntemimizde üç önemli değişiklik meydana getiren olgu: 1.Daha fazla veri, 2. Messier (eksik) veri, 3. Nedenselliğin ötesinde korelasyon

2.1.2. Büyük veri karakteristikleri

Geleneksel araçlar ve yöntemler, sürekli büyüyen verileri yönetmek ve analiz etmek için uygun bulunmamaktadır. Yeni yöntem ve araç kullanımı gereksinimleri, verinin değişmesinden kaynaklanmaktadır. Büyük verinin karakteristiklerine bakmadan önce, geleneksel veriler ve büyük veri arasındaki farklılıkları ortaya koymak gereklidir. Bu

farklılıklar, büyük veri karakteristiklerinin belirlenmesinde önemli rol oynamaktadır ve Çizelge 2.2’de sıralanmıştır.

Çizelge 2.2. Geleneksel veri ve büyük veri arasındaki farklılıklar (Rajendran, Asbern, Kumar, Rajesh ve Abhilash, 2016)

No	Geleneksel Veri	Büyük Veri
1	Yapısal Veri	Yapısal olmayan veya Yarı-yapısal Veri
2	Verilerin boyutu çok küçük	Boyut, geleneksel veri boyutundan çok büyük
3	Veriler merkezidir	Veriler dağınıktır
4	Veriler üzerinde çalışmak ve işlemek kolaydır	Verileri işlemek zordur
5	Normal sistem yapılandırması, verileri işlemek için yeterlidir	Verileri işlemek için yüksek sistem yapılandırması gerekmektedir
6	Geleneksel bir veri tabanı aracı yeterlidir	Özel araçlar gerekmektedir
7	Verileri işlemek için normal fonksiyonlar yeterlidir	Verileri işlemek için özel fonksiyonlar gerekmektedir

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 bileşenlerinden oluşan 5V modelidir (Gahi, Guennoun ve Mouftah, 2016), Çizelge 2.3’de büyük verinin 5V karakteristik özellikleri gösterilmektedir.

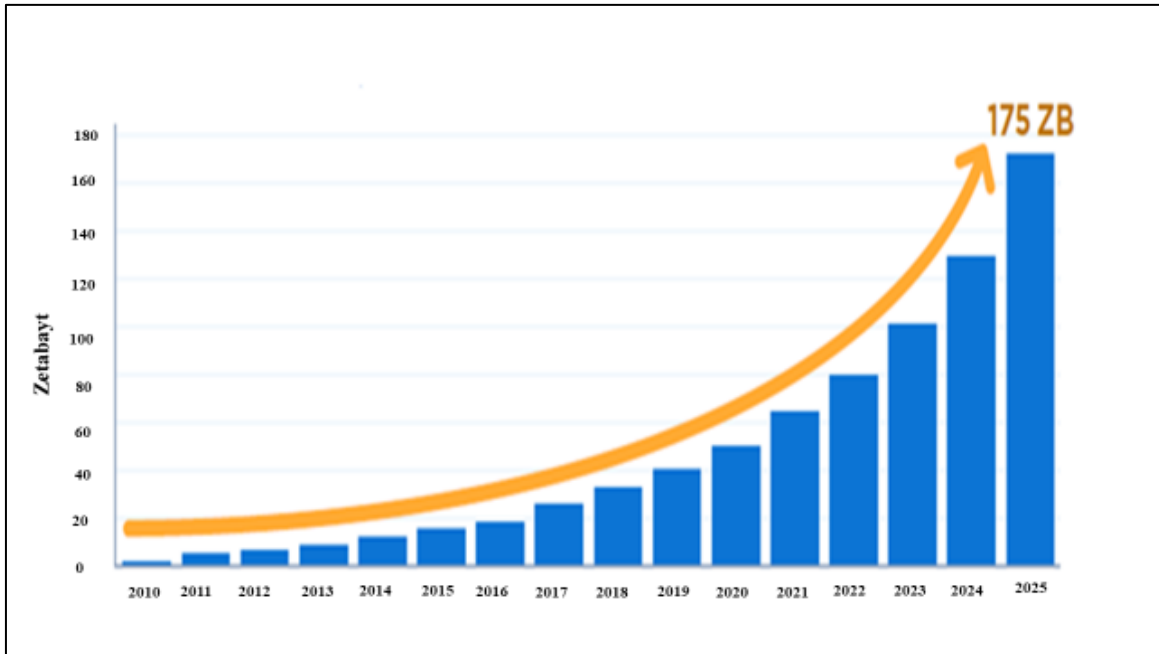
Çizelge 2.3. Büyük verinin 5V karakteristik özellikleri (Demchenko, Grosso, deLaat ve Membrey, 2013)

Veri Hacmi (Volume)	Veri Hızı (Velocity)	Veri Çeşitliliği (Variety)	Veri Gerçekliği (Veracity)	Veri Değeri (Value)
Terabaytlar Kayıtlar İşlemler Tablolar Dosyalar	Batch Gerçek Zamanlı Yakın Zamanlı Süreçler Akışlar	Yapısal Yarı Yapısal Yapısal Olmayan Çok Faktörlü Olasılıksal	Güvenilirlik Geçerlilik Kaynak Saygınlık Ulaşılabilirlik Hesaplanabilirlik	İstatistiksel Olgular Korelasyonlar Varsayımlar

Hacim (Volume): Hacim, şu anda kullanılan tüm geleneksel teknolojilere ve araçlara özel gereksinimler gerektiren büyük verinin en önemli ve ayırt edici özelliğidir. 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.

IDC, dünyadaki verilerin her iki yılda bir ikiye katlandığını iddia etmiştir. IDC istatistiklerine göre, 2018’den 2025’ye kadar ulaşılacak veri miktarının 33 zetabaytdan 175 zetabayta kadar artacağı öngörülmektedir (Rydning, 2018). Şekil 2.2’de küresel veri miktarının artışı yıllara oranla gösterilmektedir. Bu kadar büyük miktarda veriye sahip olmanın zorlukları, veriyi depolamak, analiz etmek, devamlılığını sağlamak ve kullanmak olarak belirtilmektedir (Trifu ve Ivan, 2014).



Şekil 2.2. Küresel veri miktarının yıllık büyüklüğü (Rydning, 2018)

Hız (Velocity): Büyük verinin işlenebildiği süreyi ifade etmektedir. Sensör verileri ve çoklu işlemlerden gelen veriler de dâhil olmak üzere büyük veriler, gerçek zamanlı, gerçek zamanlıya yakın, toplu olarak veya akışlar halinde, yüksek hızda işlenmesi gerekmektedir. Bazı faaliyetler çok önemlidir ve hızlı yanıt verilmesine ihtiyaç duyulmaktadır, bu nedenle veriyi hızlı işlemenin, verimliliği en üst düzeye çıkardığı görülmektedir. Mobil uygulamalar üzerinden akan veriler, günlük müşteri talepleri için gerçek zamanlı ve kişiselleştirilmiş şekilde kullanılabilir. Akıllı telefonların yükselen kullanımı göz önüne alındığında, perakendeciler yakında gerçek zamanlı analitik gerektiren yüzbinlerce

akan veriyle uğraşmak zorunda kalacaklardır. Örneğin Walmart¹² saatte bir milyondan fazla işlem gerçekleştirmektedir (Cukier, 2010). Geleneksel veri yönetimi sistemleri de, büyük veri akışlarını anında işleyememekte ve yeni veri analizi teknikleri ortaya çıkmaktadır (Gandomi ve Haider, 2015). 2019 yılında dünya genelinde internette 1 dakikada, Google’da 3.8 milyon arama yapıldığı, Facebook’a 1 milyon giriş yapıldığı, Youtube’da 4.5 milyon video izlendiği, Netflix’de 694 bin saat içerik izlendiği, 188 milyon e-mail gönderildiği ve 996 bin 956 dolar harcandığı belirtilmektedir (Desjardins, 2019).

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. Web’de veri alışverişi için metinsel bir dil olan Genişletilebilir İşaretleme Dili (XML), yarı yapılandırılmış verilerin tipik bir örneğidir. XML belgeleri, makine tarafından okunabilir, kullanıcı tanımlı veri etiketleri içerir. Yüksek çeşitlilik, kuruluşlar için yeni bir özellik değildir, dahili kaynaklardan (sensör verileri) ve harici kaynaklardan (sosyal medya verileri) gelen yapılandırılmamış verileri biriktirmektedir. Yenilikçi olan ise organizasyonların iş süreçlerindeki verileri kullanmasını sağlayan yeni veri yönetimi teknolojileri ve analitiklerinin ortaya çıkmasıdır (Gandomi ve Haider, 2015). Çeşitlilik, büyük veri ve bilgilerin karmaşıklığı ve bu verilerin arkasındaki anlamsal modeller ile ilgilidir. Veri çeşitliliği, özellikle yukarı ve aşağı ölçeklendirme, veri formatına dinamik olarak adapte olması gereken veri depolama ve veri tabanı tasarımına yeni gereksinimler getirmektedir.

Gerçeklik (Veracity): IBM, bu özelliği veri kaynaklarında bulunan güvenilirliği temsil eden dördüncü V olarak kullanmıştır. Örneğin, sosyal medyadaki müşteri duyguları 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). Büyük verinin gerçeklik boyutunun iki yönü vardır: veri tutarlılığı ve veri güvenilirliği. Veri tutarlılığı veya kesinliği, istatistiksel güvenilirlikle

¹² Dünya çapında şubeleri olan bir Amerikan perakende satış yapan mağazalar zinciridir.

tanımlanabilmektedir. Veri güvenilirliği ise, veri kaynağı, toplama ve işleme yöntemleri dâhil olmak üzere bir dizi faktörle tanımlanabilmektedir. Veri gerçekliği tamamen, büyük veri altyapısından erişilebilen güvenlik altyapısına dayanmaktadır. Veri gerçekliğinin sağlanması için aşağıdaki hususların tanımlanması ve ele alınması gerekmektedir (Demchenko ve diğerleri, 2013):

- Veri bütünlüğü ve bağlantılı veriler (örneğin, karmaşık hiyerarşik veri, dağıtık veri)
- Veri orijinallığı ve güvenilir veri kaynağı
- Veri ve kaynağının tanımlanması
- Bilgisayar ve depolama platformu güvenilirliği
- Kullanılabilirlik ve zamanlama
- Hesap verebilirlik (mesuliyet) ve itibar

Veri hacminin ve veri çeşitliliğinin çok olması nedeniyle eldeki temizlenmemiş veriden faydalı ve doğru veriyi tanımlamak oldukça zordur, bu yüzden gerçeklik, büyük veri ile ulaşılabilecek özelliklerin en zor olanıdır. Eldeki temizlenmemiş veri, çığ gibi büyüyen hatalara, doğru olmayan sonuçlara neden olduğundan ve büyük verinin hız özelliğine olan olumsuz etkilerinden dolayı büyük bir sorun olarak görülmektedir. Eğer iyi bir veri temizleme ekibi yoksa büyük verinin temel amacı bozulmakta ve bütün bilgiler faydasız ve çok pahalı bir büyük veri çevresine dönüşmektedir (Trifu ve Ivan, 2014).

Değer (Value): Oracle, değeri büyük verilerin tanımlayıcı bir özelliği olarak tanıtmıştır. Oracle’ın 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.

Büyük verinin en önemli özelliklerinden birisi olarak ortaya çıkmaktadır ve toplanan verinin doğruluğuna karşılık gelmektedir (Kaisler, Armour, Espinosa ve Money, 2013).

Büyük verinin işlenmesi sonucunda ortaya çıkan değer, tüm veri yığınının anlamlı hale getirmektedir. Diğer tüm büyük veri özellikleri veri yığınındaki değeri ortaya çıkarmaya hizmet etmektedir.

2.1.3. Büyük veri türleri

Veri büyüklüğünün çok hızlı bir şekilde artmasının nedenlerinin başında internet ve sosyal medya üzerinden erişilebilen kullanıcıların ürettiği veri gelmektedir. We are Social ve Hootsuite'in son yayınlanan Global Dijital 2019 raporuna göre, Ocak 2018'den bu yana her gün bir milyondan fazla kişi çevrimiçi olmaktadır. Raporda, dünya genelinde 2019'da 4.39 milyar internet kullanıcısının olduğu ve bu rakamın Ocak 2018'e kıyasla 366 milyon (yüzde 9 oranında) artış gösterdiği belirtilmektedir. Aynı zamanda, 2019'da 3.48 milyar sosyal medya kullanıcısının olduğu ve geçen yıldan bu yana 288 milyon (yüzde 9 oranında) artış gösterdiği belirtilmektedir (Kemp, 2019).

Büyük veri, geleneksel veri tabanı teknikleri kullanılarak işlenmesi mümkün olmayan büyük miktarda heterojen veriyi tanımlayan yeni bir kavramdır. Veriler, yapılandırılmış veri, yarı yapılandırılmış veri ve yapılandırılmamış veri olmak üzere türüne göre üç kategoride incelenmektedir (Gahi ve diğerleri, 2016).

Yapılandırılmış veri: Yapılandırılmış veri, modellenmesi, veri girişi yapılması, saklanması, sorgulanması, işlenmesi ve görselleştirilmesi kolay olan her türlü veridir. Genellikle ilişkisel veri tabanlarında veya elektronik tablolarda yönetilen belirli tür ve boyutlarda önceden tanımlanmış alanlar olarak sunulmaktadır. Yüksek performanslı işlemler gerektirmediğinden faydalı bilgilerin elde edilmesi diğer veri türlerine göre daha kolaydır.

Yarı yapılandırılmış veri: Yarı yapılandırılmış veri, etiketler ve işaretçiler yardımıyla belirli öğeleri tanımlamak ve verilerdeki farklı alanların hiyerarşik bir gösterimini tanımlamak için kullanılan bir tür meta-model içermektedir. En çok bilinen yarı yapılandırılmış veri örnekleri XML (Extensible Markup Language) ve JSON (JavaScript Object Notation)'dır.

Yapılandırılmamış veri: Yapılandırılmamış veri, önceden tanımlanmış bir format olmadan sunulan ve saklanan bir kayıt türüdür. Genellikle metinler, e-postalar, resim, ses ve video

dosyalarından oluşur. Bu tür verileri kayıt altına almanın zor olması sebebiyle, NoSql gibi yeni veri işleme mekanizmaları kullanılmaktadır.

2.1.4. Büyük veri kaynakları

Yüksek çözünürlüklü videolar, ses kayıtları, filmler, fotoğraflar, tıbbi kayıtlar ve görüntüler, sunuculardan alınan log kayıtları, sismik görüntüler, coğrafi haritalar, iklim bilimi ve hava durumu kayıtları gibi farklı kaynaklardan gelen veriler, büyük veri kaynaklarına örnek olarak verilmektedir. Günümüzde, Nesnelerin İnterneti (IoT)¹³ bağlı olarak geliştirilen akıllı şehirlerin tarımı, endüstrisi ve bakımı da yeni büyük veri kaynağı haline gelmiştir. Çizelge 2.4’de farklı kaynaklardan gelen veriler sektör bazında özetlenmektedir. Ayrıca üretilen verinin türü ve kullanım amacı da yer almaktadır (Bhadani ve Jothimani, 2016).

Çizelge 2.4. Farklı veri kaynakları (Bhadani ve Jothimani, 2016)

Sektör	Üretilen Veri	Kullanım Amacı
Astronomi	Yıldızların, uyduların hareketleri	Asteroitlerin ve uyduların faaliyetlerini izlemek
Finans	Video, ses, twitter ve haber raporu yoluyla haber içeriği	Ticari kararlar vermek
Sağlık Hizmetleri	Elektronik tıbbi kayıtlar ve görüntüler	Kısa vadede halk sağlığını izlemeye ve uzun vadede epidemiyolojik araştırma programlarına yardımcı olmak
Nesnelerin İnterneti	Sensör verileri	Akıllı şehirlerdeki çeşitli faaliyetleri izlemek
Yaşam Bilimleri	Gen dizileri	Genetik çeşitlilik ve potansiyel tedavi etkinliğini analiz etmek
Medya/Eğlence	Kullanıcı medya görüntüleme davranışı ve içerikler	Daha fazla izleyici yakalamak
Sosyal Medya	Blog mesajları, tweetler, sosyal ağ siteleri, günlük log ayrıntıları	Müşteri davranış örüntülerini analiz etmek
Telekomünikasyon	Çağrı Detay Kayıtları (Call Detail Records-CDR)	Müşteri kaybını yönetmek
Taşımacılık, Lojistik, Perakende, Kamu Hizmetleri	Filo alıcı vericileri, RFID etiket okuyucuları ve akıllı sayaçlardan üretilen sensör verileri	İşlemleri optimize etmek
Video Gözetimi	CCTV’den IPTV kameralarına ve kayıt sistemlerine yapılan kayıtlar	Hizmet geliştirme ve güvenlik için davranış modellerini analiz etmek

¹³ IoT (Internet of Things): teknolojisi akıllı cihazların birbiriyle iletişime geçmesi, haberleşmesidir. Küçük ev aletlerinden akıllı şehirlere kadar bu teknoloji kullanılmaktadır. Burada oluşan veriler büyük veri olarak karşımıza çıkmaktadır.

2.1.5. Büyük veri değer zinciri

Porter tarafından sunulan değer zinciri modeli, bir firmanın müşterilerine bir ürün veya hizmet sunma aşamasında değer katmak için gerçekleştirdiği bir faaliyet grubunu ifade etmektedir. Benzer bir şekilde veri değer zinciri, mevcut verilerden değer yaratmak için bir dizi faaliyetle ilgilenen çerçeveyi ifade etmektedir. Büyük veri değer zincirinin aşamaları Çizelge 2.5'te gösterilmektedir: Veri Oluşturma, Veri Toplama, Veri İletimi, Veri Önişleme, Veri Depolama, Veri Analizi ve Karar Verme (Bhadani ve Jothimani, 2016).

Çizelge 2.5. Büyük veri değer zinciri (Bhadani ve Jothimani, 2016)

Veri Oluşturma	Veri Toplama	Veri İletimi	Veri Önişleme	Veri Depolama	Veri Analizi
İş Verisi E-Ticaret Sosyal Medya Bilimsel Veri Sağlık Verisi Finans Verisi Telekom Verisi	Log Dosyaları Sensörler Web Tarayıcı Ağ İzleme Sistemleri	İnter-DCN İletimi İntra-DCN İletimi	Birleştirme Temizleme Gereksiz Verilerin Giderilmesi Özellik Çıkarma	GFS HDFS Dynamo BigTable Cassandra CouchDB Dyrad	Veri ve Text Madenciliği Çok Değişkenli İstatistiksel Analiz Sosyal Medya Analitiği Ağ Analitiği Multimedya Analitiği

Veri Oluşturma: Büyük veri değer zincirinin ilk ve en önemli adımı veri üretimidir. Facebook ve Twitter gibi sosyal medyada üretilen veriler, bloglar, arama kayıt detayları gibi çok çeşitli kaynaklardan üretilen veriler örnek olarak verilebilir.

Veri Toplama: Farklı kaynaklardan gelen verileri toplamak için çeşitli metotlar kullanılmaktadır. Veriler toplanırken log dosyaları, sensörler, web tarayıcıları ve ağ izleme sistemlerinden faydalanılmaktadır. Örneğin, Twitter ve Facebook gibi sosyal medya sitelerinden, müşterilerin görüşlerinden veya şikâyetlerinden (şirketin Facebook sayfasında paylaşılan görüşler) elde edilen veriler ile Telekom'daki müşteri kaybı tahmin edilmektedir.

Veri İletimi: Veriler toplandıktan sonra, analiz için bir veri depolama ve işleme altyapısına aktarılmaktadır. İki aşamada gerçekleştirilebilir; Dinamikler Arası Devre Ağı (Inter-DCN) ve Dinamikler İçi Devre Ağı (Intra-DCN). Inter-DCN iletimi, verilerin veri kaynağından veri merkezine aktarılmasını sağlarken, Intra-DCN veri merkezi içindeki aktarımına yardımcı olmaktadır.

Veri Önışleme: Çeřitli veri kaynaklarından toplanan bazı veriler, işlenmemiř haliyle, gereksiz, hatalı veya tutarsız olabilmektedir. Dolayısıyla bu aşamada veriler, analiz için gereken veri kalitesini arttırmak için önceden işlenmektedir. Bu aynı zamanda analizin doğruluğunu arttırmaya ve depolama masraflarını azaltmaya yardımcı olmaktadır. Veriler, birleřtirme, temizleme ve gereksiz verilerin giderilmesi gibi adımlar yardımıyla önceden işlenebilmektedir.

Veri Depolama: Verilere güçlü erişim sağlanması için güvenilir depolama alanı gerekmektedir. Dağıtık depolama sistemleri veriler için, tutarlılık (Consistency), kullanılabilirlik (Availability) ve bölüm toleransı (Partition Tolerance) gibi faktörleri dikkate almaları gerekmektedir. Brewer (2000) tarafından önerilen CAP¹⁴ teorisine göre, dağıtık depolama sistemleri, aynı anda iki gereksinimi karşılayabilmektedir. Büyük veri depolama mekanizması alanında hala önemli arařtırmalar devam etmektedir. Bu konuda ilerleme kaydeden sistemlere, Google Dosya Sistemi (GFS), Hadoop Dağıtık Dosya Sistemi (HDFS), BigTable, Cassandra, CouchDB ve Dyrad örnek olarak verilebilir.

Veri Analizi: Veri toplandıktan, dönüřtürüldükten ve depolandıktan sonra veri analizi gerçekteřirilmelidir. Toplanan ve dönüřtürölen verilere dayanarak, belirli bir sorun için bir dizi ölçüm tanımlanmalıdır. Geliřtirilecek analizin zamanında yapılmasına baėlı olarak uygun mimari seřilmelidir. Verilerin sürekli deėiřmeye devam ettiėi alanda gerçekte zamanlı analizler kullanılmakta ve işlemlerin hızlı bir řekilde yapılmasına ihtiyaç duyulmaktadır. Mevcut mimariler arasında bellek tabanlı hesaplamalar ve paralel işleme sistemleri bulunmaktadır. Veri analizi için en önemli adımlardan biri uygun tekniklerin seřilmesidir. Kümeleme analizi, regresyon analizi ve veri madenciliėi algoritmaları gibi birkaç geleneksel veri analizi teknikleri, bazı büyük veri analizleri için hala kullanılmaktadır. Ayrıntıların çoklu ölçeklerde ve dakika bazında detayların denetlenmesi için veri görselleřtirilmesi de veri analizinde önemli bir yere sahiptir.

Karar Verme: Analizlere ve görselleřtirilmiř sonuçlara dayanarak karar vericiler, olumsuz bir davranıřı nasıl deėiřtireceklerine ve olumlu bir davranıřı nasıl ödüllendireceklerine

¹⁴ CAP: Consistency, Availability, Partition Tolerance

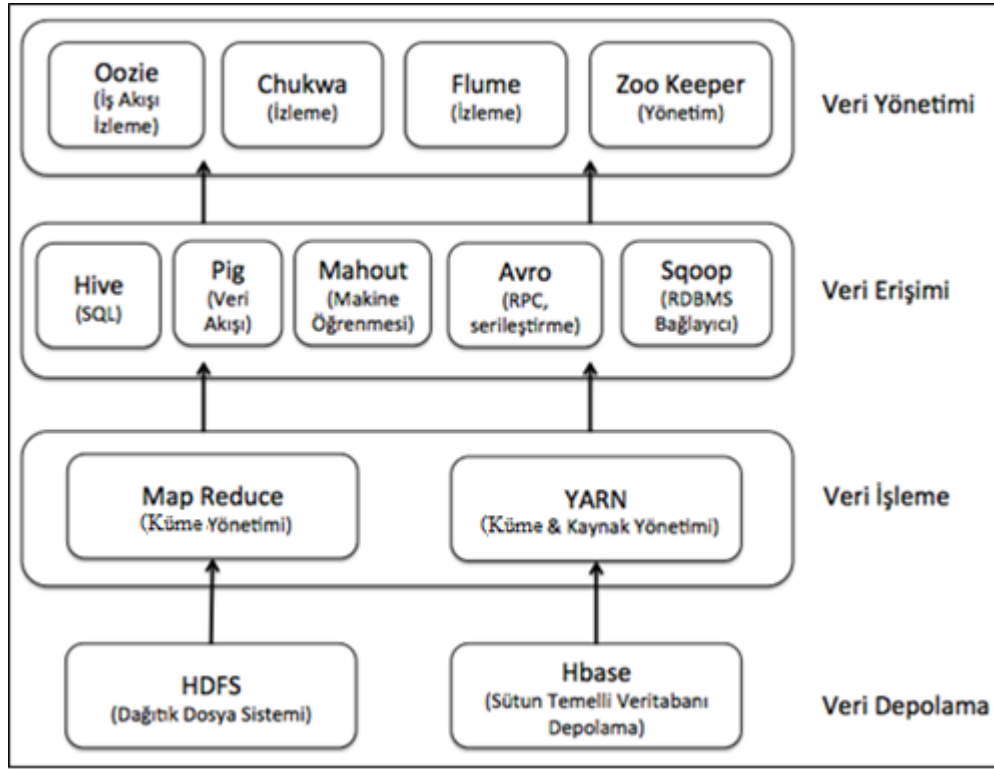
karar verebilmektedir. Büyük veri, gerekli eylemleri planlamaya ve bilinçli kararlar alınmasına yardımcı olmaktadır.

2.1.6. Büyük veri altyapısı ve Hadoop ekosistemi

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 Eşle-İndirge (Map-Reduce) 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 Eşle-İndirge (Map-Reduce) olmak üzere iki temel bileşeni vardır (Jyothirmayee, Reddy ve Akbar, 2014).

Hadoop, güvenilir ve paylaşılabilen depolama ve analiz sistemi sağlamaktadır. Map-Reduce temel olarak, anahtarlar ve değerlerle hesaplama yapan bir programlama modeli sağlayarak disk okuma ve yazma sorununu azaltmaktadır. Depolama işlemi, Hadoop Dağıtık Dosya Sistemi tarafından, analiz işlemi ise Eşle-İndirge tarafından gerçekleştirilmektedir (Katal, Wazid ve Goudar, 2013).

Dağıtık dosya sistemleri, Eşle-İndirge hesaplama sistemi ve dağıtık kilitleme servisleri dâhil olmak üzere büyük veri sistemlerinin birçoğu Google öncülüğünde gerçekleştirilmiştir. Açık kaynak kodlu yazılım geliştiren topluluklar da yıllar içerisinde Hadoop, HBase, MongoDB, Cassandra, RabbitMQ ve daha birçok projelerle bu alana katkıda bulunmuşlardır (Marz ve Warren, 2014). Büyük veri mimarisinin çözümleri veri ambarı gibi diğer veri saklama çözümlerinden farklılık göstermektedir. Büyük veri altyapısında, doğası gereği açık kaynak kodlu yazılımlar ve araçlar kullanılmaktadır. Facebook, Twitter, Google, Yahoo ve LinkedIn gibi büyük şirketler açık kaynak kodlu projelere katkıda bulunmaktadır. En iyi bilinen büyük veri araçları günümüzde açık kaynak kodlu olarak kullanılmaktadır. En çok kullanılan ve en çok bilinen araçların başında Hadoop gelmektedir. Büyük veri altyapısının önemli rollerinden birini de NoSql veri tabanları oluşturmaktadır (Trifu ve Ivan, 2014).



Şekil 2.3. Hadoop ekosistemi (Savvycom, 2016)

Hadoop ekosisteminde, veri depolama için HDFS ve Hbase, veri işleme için Map-Reduce ve Yarn, veri erişimi için, Hive, Pig, Mahout, Avro ve Sqoop, veri yönetimi için ise Oozie, Chukwa, Flume ve Zoo Keeper araçları kullanılmaktadır (Bkz. Şekil 2.3).

Veri depolama araçları

HDFS - Hadoop Dağıtık Dosya Sistemi

HDFS, verinin güvenli ve hızlı performans sağlamasını amaçlayan, veriyi bir kümenin birkaç düğümü üzerinden hızla çoğaltan bir Hadoop bileşenidir (Jyothirmayee ve diğerleri, 2014). Bir dosya sistemi eğer birden fazla makine üzerindeki verileri bir ağ üzerinden depolanması ve yönetilmesini sağlayabiliyorsa, o dosya sistemi bir dağıtık dosya sistemidir. Ağ üzerinde oluşabilecek sorunların karmaşıklığı sebebi ile dağıtık dosya sistemleri daha karmaşık ve zordur. Hadoop yapısının temelinde, dosya sistemi olarak HDFS kullanılmaktadır. HDFS, büyük boyuttaki dosyaları depolamak için sıradan sunucu kümelerini kullanmak üzere tasarlanmıştır (White, 2012). Hadoop Dağıtık Dosya Sistemi, kümelerdeki verilerin daha küçük parçalara (blok) bölünmesine ve küme boyunca

dağıtılmasına yardımcı olmaktadır. HDFS, üç parçadan oluşmaktadır; Namenode, Secondary Namenode ve Datanode. Bunlardan Namenode, HDFS üzerindeki verilerin yönetimini yapmaktan sorumludur. İkincil olan Secondary Namenode, Namenode'un yedeği olarak çalışmaktadır. Üçüncü parçayı oluşturan Datanode ise gerçek verilerin saklandığı bölümdür. Hadoop sisteminde Namenode ve Secondary Namenode birer adet bulunur ve ana makine (master) üzerinde çalışmaktadır. DataNode bir veya daha fazla sayıda olabilir ve diğer istemci makineler (slave) üzerinde çalışmaktadır (Apache Hadoop, 2016).

HBase

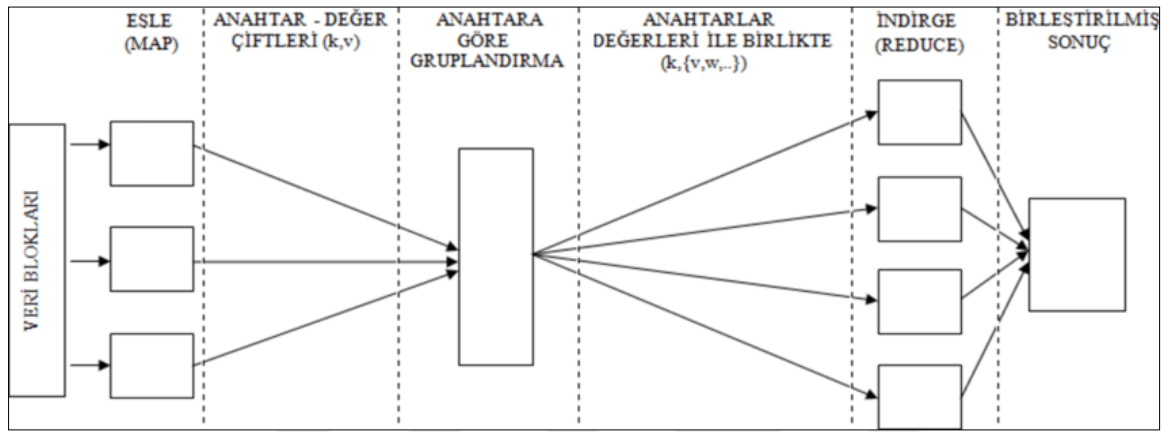
HBase, ilişkisel olmayan, dağıtık, ölçeklenebilen bir Hadoop veri tabanıdır (Apache Hbase, 2016). HBase, büyük veri üzerinde rastgele ve gerçek zamanlı hızlı veri okuma ve yazma amacıyla kullanılabilen, HDFS üzerinde çalışan, ilişkisel olmayan, sütun tabanlı bir NoSql veri tabanıdır (Fidan, 2015). Büyük veriler ve HDFS üzerinde çalışan ilişkisel veri tabanı dışında bir veri tabanına ihtiyaç duyulduğu zaman kullanılabilir. Ölçeklenebilirliği sağlamaktadır. Bu veri tabanı üzerinde tablo boyutları çok büyük olabilmektedir. Bu veri tabanının en önemli farklılığı hızlı erişim sağlamasıdır.

Veri işleme araçları

Eşle-İndirge (Map-Reduce)

Eşle-İndirge, dağıtık mimaride tutulan verilerin paralel olarak çalışıp işlenmesini ve analiz edilmesini sağlayan bir yazılım altyapısı (framework)'dır (Apache Hadoop, 2013). Sistem, 1960'lı yıllarda geliştirilen fonksiyonel programlamadaki eşleme (map) ve indirgeme (reduce) fonksiyonlarından esinlenmiştir. Veriler işlenirken temel olarak bu iki fonksiyon kullanılmakla beraber, her bir aşamada ek olarak bazı fonksiyonlar da mevcuttur. Her ne kadar ilk ortaya çıkma zamanı tartışılırsa da, 2004 yılında Jeffrey Dean ve Sanjay Ghemawat tarafından yayınlanan "MapReduce: Simplified Data Processing on Large Clusters" adlı makale ile tanınmaya başlanmıştır. Bu makalede, Google'ın büyük verileri nasıl parçalara ayırdığı ve işlediği anlatılmaktadır (Miner ve Shook, 2013). Bu makaleden sonra Nutch isimli açık kaynak kodlu bir proje ile Map-Reduce uyarlama çalışmaları yapılmıştır. Zamanla yaygınlaşmış ve özellikle Yahoo'nun yatırımları ile gelişmiş ve zamanla ayrılarak Apache tarafından geliştirilmeye başlanmıştır. Apache Hadoop'un açık

kaynak kodlu olmasının avantajlarından biri, geliştiricilerin sürekli sistemi geliştirerek farklı projeler ile Eşle-İndirge yöntemi üzerinde çalışma yapmalarıdır. Bu projelere Pig, Hive ve Hbase örnek olarak verilebilir. Eşle fonksiyonu, Şekil 2.4'te gösterildiği üzere büyük verileri, küçük, yönetilebilir, alt verilere bölen, daha sonra onları düğüm adı verilen çok sayıda sunucuya dağıtan bir eşleme fonksiyonudur. İndirge fonksiyonu ile anahtarlara göre işlenmiş veri bloklarını gruplandırarak indirgeme işlemi gerçekleştirilir.



Şekil 2.4. Eşle-İndirge (Map-Reduce) fonksiyonları şematik yapısı (Aktan, 2018)

Eşle-İndirge fonksiyonları kullanılan bir kelime sayma örneğinde, Hadoop sunucu kümesi düğümlerden oluşmakta ve girdi düğümlere bölünmektedir. Her düğüm kendi içinde anahtar kelimeye göre eşleme yapmakta ve adet bilgisini değer olarak atamaktadır. Eşleme işleminden sonra anahtarlara göre gruplandırma yapıp indirgeme aşamasına geçilmektedir. Bu aşamada, her gruptaki anahtara karşılık gelen değerler toplanmakta ve sonuçlandırılmaktadır. Böylece Eşle-İndirge fonksiyonları ile hangi kelimeden kaç adet olduğu sonucuna varılmaktadır (Raj, 2014).

YARN (Yet Another Resource Negotiator)

YARN, Hadoop 2.0'daki Map-Reduce'un en son sürümüdür. YARN, bir işletim sisteminin rolünü yerine getirmektedir. Görevi, iş yüklerini yönetmek ve izlemek, birden fazla istemciye hizmet verebildiğinden ve güvenlik kontrolleri yapabildiğinden emin olmaktır. Ayrıca YARN, Map-Reduce tarafından desteklenmeyen yeni işleme modelini desteklemektedir (Savvycom, 2016).

Veri erişimi araçları

Hive

Sql kodları ile Hadoop mimarisi üzerinde sorgulama işlemleri yapmayı sağlayan bir teknolojidir. Ayrıca veri ambarı uygulamaları geliştirmek için de kullanılabilir. Apache Hive, dağıtılmış depolama yazılımı, Hadoop'a SQL ve depolama yetenekleri eklemek için 2008 yılında Facebook tarafından geliştirilmiştir.

Pig

Apache Pig, Yahoo araştırması sonucu başlamış bir Hadoop teknolojisidir. SQL operasyonlarına imkân sağlamaktadır. En büyük artılarından biri Java programlama dili bilmeyenlere veriyi işleme ve çıktı alma imkânı sunmasıdır. Pig, Apache Hadoop üzerinde prosedürel bir veri akışı yazmayı sağlayan bir veri işleme platformudur. Veri kullanıcılarına, Hadoop'un güçlü, dağıtık ve esnek yapısına ileri seviye Java kodları yazmadan daha üst bir katmandan erişim imkânı sağlayan bir yapıdır. Pig, Hadoop üzerindeki büyük verileri geliştirmek için geliştirilmiş olup Pig Latin olarak isimlendirilen yüksek seviyeli bir dile sahip veri işleme aracıdır (Apache Pig, 2018).

Mahout

Apache Mahout, işletim sistemi bağımsız olarak çalışabilen ve Hadoop üzerinde çalışan büyük veri kümeleri üzerinde veri madenciliği algoritmalarını çalıştırmak amacıyla kullanılan bir makine öğrenme kütüphanesidir (Apache Mahout, 2016). Üzerinde bulunan hazır algoritmalar ile büyük veri probleminin çözümünde ileri veri madenciliği algoritmalarının da kullanılmasına imkân sağlamaktadır. Kümeleme, sınıflandırma, önerme vb. alanlarda uyarlanmış birçok makine öğrenme metodunu barındırmaktadır (Fidan, 2015).

Avro

Avro, verileri kompakt bir ikili veri formatında serileştiren ve zengin veri yapıları ve kalıcı verilerin depolanması için saklama dosyası sağlayan bir veri serileştirme sistemidir. Avro, veri okumak ve yazmak için şemaları, veri türleri ve protokolleri tanımlamak için JSON (Java Script Open Notation) kullanmaktadır. Hadoop düğümleri arasında ve istemci ve

servisler arasında iletişim kurmak için kablo formatını kullanmaktadır (Narasimhan ve Bhuvaneshwari, 2014).

Sqoop

Apache Sqoop, ilişkisel veri tabanları gibi, Apache Hadoop ve yapılandırılmış veri depoları arasında verimli bir şekilde toplu veri aktarımı için tasarlanmış bir araçtır. Sqoop (Sql+Hadoop=Sqoop), ilişkisel veri tabanları ile Hadoop arasında paralel veri aktarımı sağlayan bir komut satırı arayüz uygulamasıdır. Sqoop, harici veri tabanlarından HDFS, HBase veya Hive'e veri aktardığı gibi, ilişkisel veri depolarına HDFS, HBase veya Hive'den veri çıkarmak için de kullanılmaktadır. Hadoop ve ilişkisel veri tabanları arasındaki verilerle ilgili güncellemeleri aktarmak için bir kaç kez çalışabilen, kaydedilmiş işleri ve basit SQL sorguları kullanmaktadır (Narasimhan ve Bhuvaneshwari, 2014).

Veri yönetimi araçları

Oozie

Oozie, Hadoop için bir iş akışı zamanlayıcısıdır. Hadoop ve Map-Reduce, Pig, Sqoop, Hive gibi uygulamalar arasında iş akışları oluşturma ve koordinasyon işlerini yönetme sürecini kolaylaştırmaktadır. Oozie'nin temel sorumlulukları şunlardır: ilk olarak yürütülecek bir dizi eylem tanımlamak; ikincisi, bu eylemler için tetikleyiciler yerleştirmek (Savvycom, 2016).

Chukwa

Chukwa, HDFS ve Map-Reduce üzerine inşa edilmiş bir başka sistemdir. Amacı dinamik ve güçlü bir veri toplama sistemi sağlamaktır. Chukwa, toplanan verilerden en iyi şekilde yararlanmak için sonuçları izleme, analiz etme ve sunma yeteneğine sahiptir (Savvycom, 2016).

Flume

Flume, büyük miktarda log verilerinin toplanması ve taşınması için verimli bir şekilde

dağıtılmış, güvenilir bir araçtır. Kullanıcıların günlük verilerini en iyi şekilde kullanmasına yardımcı olmaktadır. Yüksek hacimli ve gerçek zamanlı web günlüklerini toplayarak birden fazla kaynaktan veri akışı sağlamaktadır (Narasimhan ve Bhuvaneshwari, 2014).

Zoo Keeper

ZooKeeper, Hadoop çerçevesinde operasyon hizmetleri sunmaktadır. Kayıt defteri adı verilen yapılandırma bilgilerini korumak için dağıtılmış uygulamalar tarafından kullanılan ve veri senkronizasyonu sağlayan merkezi bir hizmettir. ZooKeeper mimarisi yedekli hizmetler yoluyla yüksek kullanılabilirliği desteklemektedir. Çeşitli dağıtım işlemlerinin, paylaşılan bir hiyerarşik ad aracılığıyla kendi aralarında koordinasyon yapmasına olanak tanımaktadır (Narasimhan ve Bhuvaneshwari, 2014).

2.1.7. NoSql veri tabanları

‘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 İlişkisel Veri Tabanı Yönetim Sistemleri’nden (RDBMS)¹⁵ farklı bir yapıdadır. NoSql’in tasarımcısı Carlo Strozzi, ilişkisel veri tabanı modelinden farklı yapıda olduğundan dolayı, ilişkisel olmayan anlamında ‘NoREL’¹⁶ veya benzer bir ad vermek gerektiğini ileri sürmektedir (Gökşen ve Aşan, 2015).

Gaurav Vaish’in “Getting Started with NoSql” kitabında, neden NoSql kullanmamız gerektiği açıklanmıştır. Vaish’e göre, çoğu NoSql veri tabanları, zaman içinde geliştirilmeye devam etme imkânı sunan şablonsuz veri gösterimi sunmaktadır. Birden çok tablo arasında verileri birleştirmek ve karmaşık sorgularla uğraşmak zorunluluğu kalmamaktadır. Bu da geliştirme zamanından tasarruf yaratmaktadır. Aygıtlar üzerinden çok daha hızlı veri dağıtımı gerçekleştirilmektedir. Sadece tek bir planı gerçekleştirmek yerine esnek yapısından dolayı ölçeklenebilirlik imkânı sağlamaktadır. NoSql veri tabanları için modellenen içeriklerin mümkün olduğu çok çeşitli depolama türleri vardır. En temel NoSql veri tabanı türleri şunlardır (Vaish, 2013):

¹⁵ RDBMS: Relational Database Management Systems

¹⁶ NoREL: No RELation

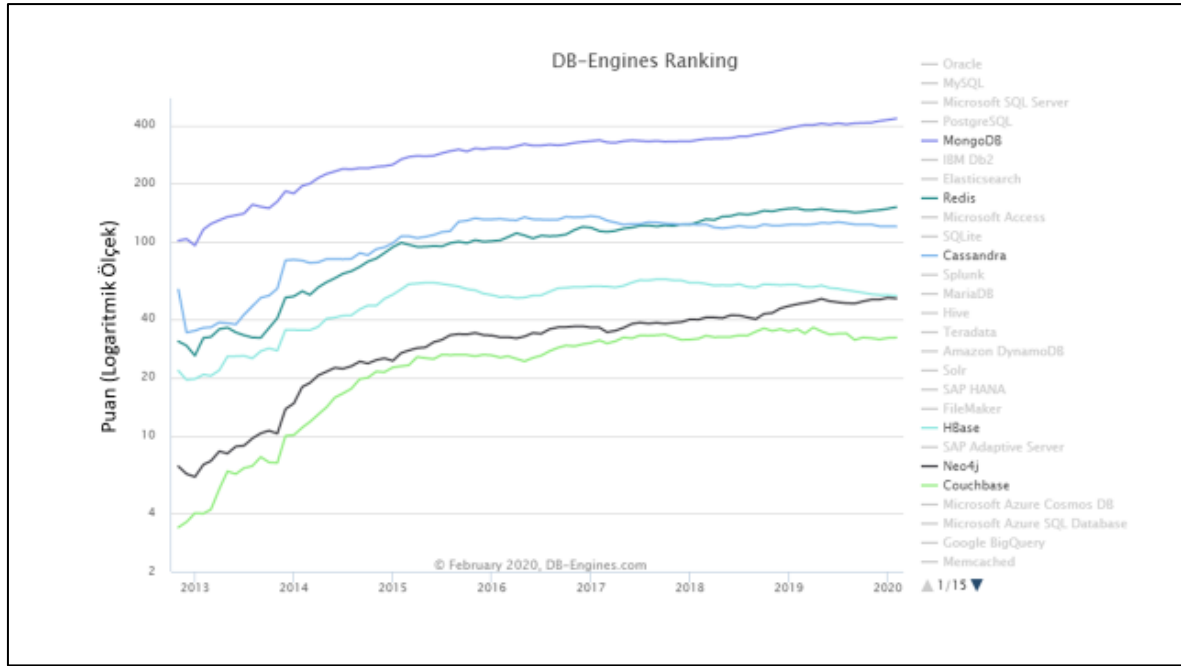
Sütun Temelli (Column-based Store): Sütun temelli NoSql veri tabanlarında, veriler, veri satırları yerine veri sütunlarında gruplanan hücrelerde saklanmakta ve sütunlar mantıksal olarak gruplanmaktadır. HBase, Apache Cassandra ve Google Big Table sütun temelli veri tabanları örnek olarak verilmektedir.

Doküman Temelli (Document-based Store): Doküman temelli NoSql veri tabanlarında, bir doküman veri tabanı yarı-yapılandırılmış verilerin eklenmesini, alınmasını ve değiştirilmesini sağlamaktadır. Bu kategori altında mevcut olan veri tabanlarının çoğu XML, JSON, BSON kullanır. HTTP veya Apache Thrift protokolleri üzerinden veri erişimine sahiptir. CouchDB ve MongoDB bu tür veri tabanlarına örnek olarak verilmektedir.

Anahtar-Değer (Key-Value Store): Anahtar-değer NoSql veri tabanları, doküman temelli NoSql veri tabanlarına benzer özelliktedir, veriler üzerinde bir değer bir anahtara karşı depolanmasına olanak tanımaktadır. Doküman temelli veri tabanlarından farklı olarak, yeni bir doküman eklendiği zaman, anahtar-değer veri tabanında, anahtarın belirtilmesi gerekmektedir. Redis, Riak ve Amazon Dynamo'nun geliştirdiği Voldemort, anahtar-değer NoSql veri tabanlarına örnek olarak verilmektedir.

Graf Temelli (Graph-based): Graf temelli veri tabanları, ilişkilerin graflar olarak temsil edildiği özel bir NoSql veritabanı kategorisini temsil etmektedir. Bir grafta iki düğüm arasında, iki düğümün paylaştığı çoklu ilişkileri temsil eden çoklu bağlantılar olabilmektedir. Neo4j ve ilk kez Twitter'ın kullandığı FlockDB, graf temelli NoSql veri tabanlarına örnek olarak verilmektedir.

DB-Engines'in aylık güncellediği en popüler veri tabanı sıralamasında en çok tercih edilen NoSql veritabanları Şekil 2.5'te gösterilmektedir. Aşağıda öne çıkan NoSql veri tabanlarından MongoDB ve Couchbase döküman temelli, Redis anahtar-değer temelli, Cassandra ve HBase sütun temelli, Neo4j graf temelli olarak en çok tercih edilen veri tabanları olmuştur.



Şekil 2.5. En popüler NoSql veritabanı sıralaması (DB-Engines, 2020)

Büyük veri analizinde, hem ilişkisel veri tabanları hem de NoSql veri tabanları ayrı ayrı kullanılabilirken, ikisinin beraber kullanımı da yaygınlaşmaktadır. 2019 yılında DeveloperWeek¹⁷ konferansına katılan yazılım geliştirici, mühendis, yazılım mimarı ve BT liderlerinin verdikleri cevaplara göre SQL kullanımı %60.48, NoSql kullanımı %39.52 olduğu belirtilmiştir. Kullanıcıların %44.3'ü birden çok veri tabanı kullandıklarını ve %75.6'sı SQL ve NoSql veri tabanlarını birlikte kullandıklarını belirtmişlerdir. En çok kullanılan NoSql veritabanları %24.6 oranıyla MongoDB, %17.4 oranı ile Redis ve %3 oranı ile Cassandra gelmektedir (ScaleGrid, 2019).

2.1.8. Büyük veri teknolojileri

Büyük veriden elde edilen değeri alabilmek yani anlamlı bilgiye ulaşabilmek için, veri analizi üzerine yeni teknikler ve teknolojiler geliştirilmesi gereklidir. Şimdiye kadar teknoloji ve yazılım geliştiriciler büyük veriyi tutabilmek, iyileştirmek, analiz etmek ve görselleştirmek için çok çeşitli teknikler ve teknolojiler geliştirmişlerdir. Bu teknikler ve teknolojiler, bilgisayar bilimi, ekonomi, matematik, istatistik ve diğer uzmanlıklar dâhil olmak üzere birçok disiplini kapsamaktadır. Büyük verinin değerli bilgilerini keşfedebilmek için disiplinlerarası yöntemlere ihtiyaç vardır. Mevcut araçlar, toplu işlem

¹⁷ DeveloperWeek: Her yıl San Francisco'da 70'den fazla ülkeden 8 binden fazla yazılım geliştirici, mühendis ve yöneticinin bir araya geldiği konferans ve etkinlik serisi.

araçları, akış işleme araçları ve etkileşimli analiz araçları olmak üzere üç sınıfa bölünebilmektedir. Toplu işlem araçlarının çoğu Mahout ve Dyrad gibi Apache Hadoop altyapısına dayanmaktadır. Akış işleme araçları, gerçek zamanlı analitik işleme için akan veri uygulamalarını kapsamaktadır. Storm ve S4, büyük ölçekli akan veri analitiği için iyi platformlardır. Etkileşimli analiz, verileri etkileşimli bir ortamda işlemekte ve kullanıcıların kendi veri analizlerini yapmalarını sağlamaktadır. Kullanıcı doğrudan bilgisayara bağlıdır ve bu nedenle gerçek zamanlı olarak etkileşime girebilmektedir. Veriler, tablo veya grafikler üzerinden veya her ikisi üzerinden aynı anda incelenip, karşılaştırılabilmekte ve analiz edilebilmektedir. Google'ın Dremel'i ve Apache Drill, etkileşimli analize dayalı büyük veri platformlarıdır (Chen ve Zhang, 2014).

Büyük veri altyapısında, doğası gereği açık kaynak kodlu yazılımlar ve araçlar kullanılmaktadır. Facebook, Twitter, Google, Yahoo ve LinkedIn gibi büyük şirketler açık kaynak kodlu projelere katkıda bulunmaktadır. En çok kullanılan ve en çok bilinen araçların başında açık kaynak kodlu Apache Hadoop gelmektedir (Jyothirmayee ve diğerleri, 2014). Veri tabanı araçlarından olan Cassandra ise, Facebook tarafından geliştirilmiş ancak Apache kuruluşu tarafından satın alınmıştır. Cassandra açık kaynak kodlu, dağıtık bir veri tabanı yönetim sistemidir. Netflix, Twitter, Cisco gibi birçok büyük şirket büyük ve aktif veri setleri için Cassandra'yı kullanmaktadır (Apache Cassandra, 2016). Cassandra, Netflix hizmetlerinin akışında arka plan veri tabanı olarak kullanılmaktadır (Jyothirmayee ve diğerleri, 2014). Çizelge 2.6'da büyük verinin depolama ve yönetim, veri tabanı, işleme, veri entegrasyonu ve istatistiksel analiz yeteneklerine göre kullanılan teknolojiler ve özellikleri belirtilmektedir.

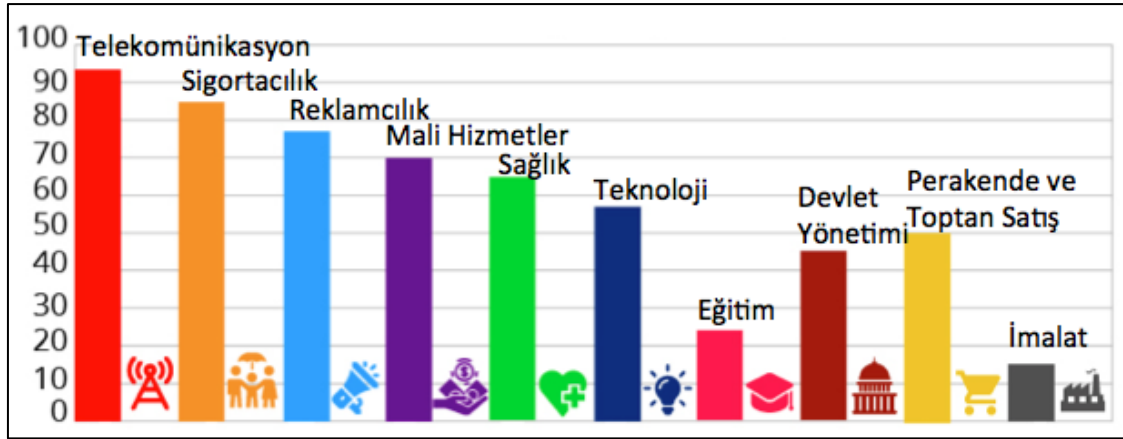
Çizelge 2.6. Büyük veri yetenekleri ve öncelikli teknolojiler (Bhadani ve Jothimani, 2016)

Büyük Veri Yeteneği	Öncelikli Teknoloji	Özellikleri
Depolama ve Yönetim Yeteneği	Hadoop Distributed File System (HDFS)	Açık kaynaklı dağıtık dosya sistemi, yüksek performanslı donanımda çalışır, yüksek oranda ölçeklenebilir depolama ve otomatik veri çoğaltma
Veri Tabanı Yeteneği	Oracle NoSql	Dinamik ve esnek şema tasarımı, yüksek ölçeklenebilir çoklu düğüm, çoklu veri merkezi, hataya dayanıklı, ACID işlemleri, yüksek performanslı anahtar-değer veri tabanı
	Apache HBase	Bölge sunucuları arasında otomatik yük devretme desteği, tabloların otomatik ve yapılandırılabilir şekilde paylaşılması
	Apache Cassandra	Her düğüm için hataya dayanıklılık özelliği, log güncellemeleri ve önbellekleme performansı ile sütun indeksleme
	Apache Hive	Map-Reduce ile sorgu çalıştırma, SQL benzeri HiveQL'in kullanılması, HDFS veya Apache HBase'den kolay ETL işlemi
İşleme Yeteneği	Map-Reduce	Veri iş yüklerinin binlerce düğüme dağıtılmasını sağlar, problemi daha küçük alt-problemlere böler
	Apache Hadoop	Yüksek seviyede özelleştirilebilir altyapı, yüksek seviyede ölçeklenebilir paralel seri işleme, hataya dayanıklı
Veri Entegrasyonu Yeteneği	Oracle big data connectors, Oracle data integrator	Map-Reduce sonuçlarını RDBMS, Hadoop ve diğer kaynaklara aktarır, grafiksel bir kullanıcı arayüzü içerir
İstatistiksel Analiz Yeteneği	R and Oracle R Enterprise	İstatistiksel analiz için programlama dili kullanılmasını destekler

2.2. Büyük Veri Kullanım Alanları

Büyük veri en çok, müşterilerin davranışlarını ve tercihlerini daha iyi anlamak için kullanılmaktadır. Şirketler, müşteri profillerinin eksik bilgilerini tamamlayabilmek için geleneksel veri setlerini, sosyal medya verileri, web tarayıcı kayıtları, metin analizleri ve sensör verileriyle genişletmeyi hedeflemektedirler. Müşterilerinden herhangi birinin ne zaman bebek bekleyeceğini doğru tahmin edebilen Amerika Birleşik Devletleri'nin en büyük perakende şirketlerinden biri olan Target, büyük veriyi ne kadar iyi kullandığını göstermektedir. Büyük veri kullanarak, telekomünikasyon şirketleri artık müşteri kaybını; perakendeciler, hangi ürünlerin satılacağını; araba sigorta şirketleri müşterilerinin gerçekte ne kadar iyi araba kullandıklarını daha iyi tahmin edebilmektedir (Marr, 2016).

Dresney Danışma Servisleri tarafından yürütülen bir anketin sonucu olan aşağıdaki grafik, sektörler arasındaki büyük veri benimsenme oranını açıklamaktadır. Açıkça görülüyor ki, dünyada her sektörün ana akımlarında büyük veri kabullenilmektedir (Şekil 2.6).



Şekil 2.6. Sektörlere göre büyük veri benimsenmesi (Agrawal, 2019)

Ülkemizde, kamu kurumlarında sağlık, eğitim, savunma, istihdam, ekonomi, enerji, ulaştırma, sosyal güvenlik, vergi, bütçe gibi birçok alanda büyük veri kullanılmaktadır. Kamu kurumlarında gerçekleştirilmiş büyük veri uygulamaları veya büyük veri çalışmalarına girdi sağlayabilecek projeler bulunmaktadır. T.C. Başbakanlık İdareyi Geliştirme Başkanlığı (KAYSİS), Enerji ve Tabii Kaynaklar Bakanlığı (Enerji Tahmin), Sosyal Güvenlik Kurumu (e-Bildirge Sistemi, MEDULA¹⁸, Aylık Tahsis, ALO 170, Veri Ambarı), Milli Eğitim Bakanlığı (MEBBİS¹⁹, e-Okul, FATİH, e-YAYGIN, ALO 147) ve Sağlık Bakanlığı (e-Nabız, Sağlık.NET, MHRS²⁰, Aşı Takip) gibi kurumların çalışmaları örnek olarak verilmektedir (Türkiye Bilişim Derneği [TBD], 2016). Büyük veri uygulamaları birçok sektörde şirketler tarafından önemli kararların alınmasında kullanılmaktadır. Bu sektörlerden bazıları aşağıda belirtilmiştir.

Finans ve Bankacılık: Büyük veri algoritmaları, sosyal medya ağlarından ve haber sitelerinden gelen veriler ile hisse senedi alım satım işlemlerinin büyük çoğunluğuna ve bu alım satım kararlarının saniyeler içinde verilmesine olanak sağlamaktadır. Bilgisayarlar, ticaret fırsatlarını araştıran ve özelleştirilebilir koşullar için pazar araştırması yapan algoritmalar ile programlanmaktadır. Bu programlar, müşterinin ihtiyaçları ve istekleri doğrultusunda, insan etkileşimli veya insan etkileşimi olmadan çalışmak üzere tasarlanmaktadır. Bankacılık alanında da büyük veri analitiği ile veri setlerinden artık daha fazla kazanç elde edilmektedir. Amerikan bankalarından biri olan “Bank of America”,

¹⁸ Çevrimiçi Reçete Onay Sistemi

¹⁹ Milli Eğitim Bakanlığı’nın bilişim sistemlerinin sunduğu elektronik hizmetlerin bütünü

²⁰ MHRS: Merkezi Hekim Randevu Sistemi

büyük veri analitiği kullanarak çağrı merkezinde en iyi performans gösteren çalışanlarının, birlikte mola veren çalışanlar olduğunu fark etti. Bunun üzerine grup olarak ara verme politikası oluşturdu ve performansın %23 arttığını tespit etti (Marr, 2016). Veri tabanlarında tutulan kayıtlar sayesinde büyük veri kullanımıyla bankalar, nakit hareketlerinin detaylarını görebilmekte, müşteri davranışlarını daha iyi anlayabilmekte, hırsızlık ve çeşitli sorunları öngörüp önleyebilmektedir (Naik ve Joshi, 2017).

Telekomünikasyon: Akıllı telefonların ve diğer mobil cihazların hızlı bir şekilde yaygınlaşmasıyla, iletişim hizmeti sağlayıcılarının ağları arasında dolaşan çeşitli veri hacminden dolayı işlemleri hızlı yapmaları, kayıt almaları ve çıkarım yapmaları gerekmektedir. Telekomünikasyon şirketleri, çağrı detay kayıtlarından, cep telefonu kullanımından, ağ ekipmanından, sunucu kayıtlarından, faturalandırmadan, müşterileri ve ağları hakkında birçok bilgi sunan sosyal ağlardan büyük miktarda veri toplamaktadır. Büyük veri analizi, telekomünikasyon şirketlerinin ağ hizmetlerini veya kullanımlarını optimize ederek, müşteri deneyimini geliştirerek ve güvenliği artırarak karlılığı arttırmalarına yardımcı olmaktadır. Büyük veri, telekomünikasyon sektörünün en genel iş kategorilerinde etkin bir şekilde kullanılmaktadır. Bunlar; müşteri kazanımı, müşteriye elde tutma, ağ hizmetleri optimizasyonu ve güvenlidir (McDonald, 2017).

Sağlık: Klinik, halk sağlığı ve davranışsal verilerin oluşturduğu veriler, tedavi kalitesini arttıran, aynı zamanda maliyeti düşüren güçlü bir tedavi sistemi geliştirilmesine yardımcı olmaktadır (Manyika ve diğerleri, 2011). Veri analistleri, verileri analiz etmek için farklı kaynaklardan gelen verileri kullanmaktadır. Bu veriler, elektronik hasta kayıtlarından, tıbbi görüntülerin bulunduğu klinik karar destek sistemlerinden, doktorun yazılı notları ve reçetesinden, eczane ve laboratuvar kayıtlarından, makine tarafından oluşturulan sensör verilerden oluşmaktadır. Büyük veri teknikleri, prematüre bebek ünitelerinde, her bebeğin kalp atışlarını ve nefes alışkanlıklarını kaydederek ve analiz ederek, herhangi fiziksel bir belirti ortaya çıkmadan 24 saat önce enfeksiyonları tahmin edebilen algoritmalar geliştirilerek kullanılmaktadır. Büyük veri analitiği aynı zamanda hastalık salgınlarının gelişimlerinin izlenmesine ve tahmin edilmesine olanak sağlamaktadır. İnsanların, ‘Bugün kendimi hasta hissediyorum’ veya ‘Hastayım, evde yatıyorum’ gibi sadece söylediklerini takip ederek, tıbbi kayıtlardan elde edilen verilerin sosyal medya analitiğiyle bütünleştirilmesi ile grip salgınlarının gerçek zamanlı olarak izlenmesine olanak tanımaktadır (Marr, 2016).

Kentleşme: Büyük veri, şehirlerin hava durumu verileri ve gerçek zamanlı trafik bilgilerine dayanarak trafik akışlarının optimize edilmesini sağlamaktadır. Birçok şehirde şu anda ulaşım altyapısı ve hizmet süreçlerinin birleştiği akıllı şehirlere dönüşmek amacıyla büyük veri analizleri yürütülmektedir. Bu altyapıyla, bir otobüs, gecikmiş bir tren için bekleyebilmekte; trafik sinyalleri, trafik yoğunluğu için tahminlerde bulunabilmekte ve trafik sıkışıklığını en aza indirmek için çalışabilmektedir. Kaliforniya'nın Long Beach şehri, yasa dışı sulamayı gerçek zamanlı olarak tespit etmek için akıllı su sayaçları kullanmakta ve bazı ev sahiplerinin su kullanımlarını %80'e kadar azaltmalarına yardımcı olmaktadır. Veniam adında bir teknoloji şirketi, Portekiz'in Porto şehrinin her yerine mobil Wifi noktaları oluşturmak için 600'den fazla şehir içi otobüs ve taksileri Wifi vericileri ile donatarak kamuya açık internet hizmeti sunmaktadır. Şehir muazzam miktarda veri toplamakta ve bunu Wifi maliyetini dengelemek için kullanmaktadır. Örneğin Porto şehrinde, sensörler çöp konteynırları dolu olduğunda şehrin atık yönetimi departmanına söylediğinden, tamamen dolu olmayan konteynırları boşaltmak için zaman harcanmadığından dolayı zamanı ve çalışma saatlerini daha verimli kullanabilmektedir (Marr, 2016).

Perakende: Perakendeciler stoklarını sosyal medya verilerinden, web arama trendlerinden ve hava tahminlerinden kaynaklanan verilere göre optimize edebilmektedirler (Marr, 2016). Dünya çapındaki iş verilerinin hacmi her 1.2 yılda bir ikiye katlanmaktadır. Walmart'ın dünya genelinde 6000 mağazasında günde yaklaşık 267 milyon işlem bulunmaktadır. Perakendede daha yüksek rekabet gücü arayışı için, Walmart kısa süre önce 4 petabayt veriyi, satış noktalarının her birinden alım kaydını takip edebilecek bir veri ambarı kurmak için Hewlett-Packard²¹ şirketi ile işbirliği yapmıştır.

Güvenlik: Yasaların uygulanmasını sağlamak ve güvenliği arttırmak için büyük veri teknolojilerinden yararlanılmaktadır. ABD'deki Ulusal Güvenlik Ajansı (NSA), terör saldırılarını önlemek için; polis güçleri, suçluların faaliyetlerini tahmin etmek ve yakalamak için veya siber saldırıları tespit etmek ve önlemek için; kredi kartı şirketleri, sahte işlemleri tespit etmek için büyük veri teknolojilerini kullanmaktadırlar (Marr, 2016).

²¹ ABD merkezli, uluslararası bilgi teknolojileri şirkettir.

Eğitim, Araştırma-Geliştirme: Büyük veri birçok eğitim kurumu tarafından hem öğretmenler, hem öğrenciler hem de yöneticiler tarafından kaliteli bir eğitim ortamının devamlılığını sağlamak için kullanılmaktadır. Örneğin öğrencilerin sisteme ne zaman giriş yaptıkları, gezindikleri web sayfaları, sayfalarda ne kadar süre harcadıkları ve belirli bir zaman içindeki faaliyetleri gibi öğrencilerin ürettiği veriler üzerinden genel bir örüntü ortaya çıkarılmasında büyük veriden yararlanılmaktadır (Naik ve Joshi, 2017). Siemens ve Long (2011), yükseköğretimin geleceğini şekillendirilmesinde ve geniş veri dizisinin verimli bir şekilde kullanılmasında büyük verinin kapsamlı bir çerçeve sunduğunu belirtmektedir. Yükseköğretim bağlamında büyük veri, gelecekteki performansı öngörmek ve akademik programlama, araştırma, öğretim ve öğrenim ile ilgili potansiyel konuları belirlemek için kurumsal performansın değerlendirilmesinde geniş bir yelpazedeki idari ve operasyonel veri toplama süreçlerinin yorumlanmasında yer almaktadır (Picciano, 2012). Bilim ve araştırma teknikleri şu anda büyük verilerin getirdiği yeni imkânlarla dönüştürülmektedir. Örneğin CERN, dünyanın en büyük ve en güçlü parçacık hızlandırıcısına sahip olan nükleer fizik laboratuvarı, veri merkezlerinde 30 petabayt veriyi analiz etmek için 65.000 işlemciye sahiptir. Bununla birlikte, verileri analiz etmek için dünya çapında 150 veri merkezine dağıtılan binlerce bilgisayarın hesaplama gücünü kullanmaktadır (Marr, 2016).

İletişim, Medya, Eğlence: Medya ve eğlence sektöründeki televizyon, haber, iletişim ve oyun endüstrisi de elde edilen büyük veri sayesinde yeni iş modelleriyle karşı karşıya kalmaya başlamıştır. Teknolojik gelişmelerle beraber müşteriler, istediği zaman, istediği yerde, istediği cihazdan içerik arama ve erişme imkânına sahip olabilmektedir. Eğlence sektöründeki şirketlerde, bu içerikleri müşteriye sunabilmek için yeni, gelişmiş ve çok kanallı bir yapı ortaya koyma baskısı oluşturmıştır. Bu internet altyapılı yeni medya, müşterilerin faaliyetleri ve eğilimlerinin anlaşılması ile oluşturulmaktadır. Bu yapı sayesinde müşteri, içeriğini dilediği zaman, dilediği yerde ve dilediği cihazdan istekleri doğrultusunda içeriği kendi zevkine göre uyarlama imkânına sahip olmaktadır. Medya ve eğlence kuruluşları daha çok müşteri kazanmak, içeriğin verimliliğini ölçmek, eşzamanlı olarak detaylı müşteri profilleri oluşturmak için büyük veri kaynaklarını etkili bir şekilde kullanmaktadır. Örneğin, Wimbledon ²² turnuvasında TV, cep telefonu ve web kullanıcılarının yorumlarını ve tenis maçlarının detaylarını iletmek için büyük veri

²² Tenis sporunun en eski ve en prestijli turnuvasıdır.

kullanılmaktadır. Birçok teknoloji devi de isteğe bağlı müzik servisleri sağlayarak, müşterilerine bilinçli müzik önerisi sunmak amacıyla büyük veri kullanmaktadır. Amazon Prime, tek bir satış noktasından, isteğe bağlı müzik, video ve Kindle kitapları sunmak için bu teknolojiden yararlanmaktadır (Naik ve Joshi, 2017). Netflix²³ de bilgisayar, televizyon ve diğer akıllı cihazlardan kullanıcıların bilgileri ve ürettiği veriler sayesinde, büyük veriyi doğru bir şekilde kullanarak ve kullanıcının zevkleri hakkında önemli bilgiler edinerek kullanıcı odaklı üretim gerçekleştirmektedir.

Tarım: Biyoteknoloji firmaları tedarik verimliliğini arttırmak için sensör verileri kullanmaktadır. Bitkilerin farklı durumlardaki değişikliklere nasıl tepki verdiğini anlamak üzere bitki incelemesi gerçekleştirmektedirler. Sıcaklık, su seviyesi gibi sürekli toplanan farklı bilgilerle toprağın verimi ve kalitesini artırma işlemi gerçekleştirilmektedir. Büyük veri teknolojileri ile bu bilgiler, ideal çevreyle ilgili durumu tespit etmek için doğru şartların bir araya nasıl getirilebileceği konusunda yardımcı olmaktadır (Memon, Soomro, Jumanı ve Kartio, 2017).

Enerji: Sensörlerin sürekli uygulanması, kablosuz iletim, ağ iletişimi ve bulut bilişim teknolojilerinin uygulanmasıyla enerji sektöründe büyük miktarlarda veri giderek daha fazla toplanmaktadır. Enerji sektöründe, büyük miktarda enerji üretim ve tüketim verileri üretilmekte ve bilgi teknolojilerinin artması ile enerji sistemleri dijitalleşmektedir. Temiz enerji üretimi, verimli enerji iletimi, dinamik enerji dağıtımı ve rasyonel elektrik tüketimi için yenilikçi depolama çözümleri sunan ve dağıtık üretim kaynakları içeren akıllı enerji sistemleri önerilmektedir. Akıllı enerji sistemi, internetin ve enerji sistemi boyunca yayılan çeşitli akıllı cihazların ve sensörlerin birleşmesi olarak ifade edilebilmektedir. Akıllı şebekede, temel veri kaynağı, akıllı şebekenin temelini oluşturan teknolojilerden biri olan gelişmiş ölçüm altyapısıdır. Gelişmiş ölçüm altyapıları, son kullanıcı tarafında çok sayıda akıllı sayaç ve diğer ölçüm terminalleri kullanmaktadır. Enerji sektöründe üretilen büyük veri, akıllı sayaçlardan gelen veriler olabildiği gibi, hava durumu verisi ve coğrafi bilgi sistemi gibi diğer kaynaklardan gelen verileri de kapsamaktadır. Bu veriler sayesinde güç üretimi ve enerji verimliliği arttırılabilmektedir (Zhou, Fu ve Yang, 2016).

²³ İnternet bağlantılı dizi, film ve daha fazlasını içeren geniş arşivli bir yayın hizmetidir.

2.3. Büyük Veri Pazar Hacmi

Büyük veri pazarının dünya çapında ulaşacağı değer tahminleri çeşitli firmalar tarafından yapılmaktadır. Büyük veri, büyük veri ve iş analitiği veya analitik teknolojiler ve büyük veri olarak pazarı bölümlendirdiklerinden dolayı farklı şirketlerden farklı pazar hacmi rakamları gösterdiği görülmektedir. Bu şirketlerden IDC'nin "Dünya Çapında Büyük Veri ve Analitik Harcamaları Yarıyıl Rehberi" ne göre, 'Büyük Veri ve Analitiği' pazarı dünya çapında, 2018-2022 yılları boyunca %13,2 olan büyüme oranını koruyacağı ve 2022 yılına kadar pazarın 274,3 milyar dolar olacağı tahmin edilmektedir. 2019 yılında Büyük Veri ve İş Analitiği (BDA) çözümleri pazarının, dünya çapında 2018'e göre %12 artışla 189,1 milyar dolara ulaşması beklenmektedir (IDC, 2019).

General Electric ve Accenture²⁴, yayınladıkları raporda şirketlerin %80-90 oranında ilk üç önceliklerinden bir tanesinin büyük veri teknolojileri olduğu belirtilmektedir (General Electric ve Accenture, 2014:5). IDC'nin 2015 yılında yayınladığı rapora göre 2020 yılına gelindiğine, Büyük Veri ve Analitik (BDA)²⁵ girişimleri üzerinde potansiyel olarak en büyük etkiye sahip beklenen eğilimler belirtilmektedir (Vesset ve diğerleri, 2015):

- Bulut tabanlı BDA teknolojisine yapılan harcamalar, şirket içi çözümler için yapılan harcamalardan 4,5 kat daha hızlı büyüyecek; açık kaynak teknolojisi bu yeni mimarinin çekirdeğini temsil edecektir.
- Tüm iş analitiği yazılımlarının %50'si bilişsel bilgi işlem işlevselliği üzerine kurulu kuralcı analizler içerecektir.
- Nitelikli personel yetersizliği, veri bilimcilerden mimar ve veri yönetimi uzmanlarına kadar genişleyecek ve devam edecektir.
- Veri tabanlarının %90'ı (ilişkisel ve ilişkisel olmayan) bellek için optimize edilmiş teknolojiye dayanacaktır.
- Dağıtık mikro analitik ve veri manipülasyonu tüm büyük veri ve analitik dağıtımlarının bir parçası olacaktır.
- Self servis görsel keşif ve veri hazırlama pazarına yapılan harcama, benzer işlevsellik için geleneksel bilgi teknolojileri kontrol araçlarında 2,5 kat daha hızlı büyüyecektir.

²⁴ İrlanda merkezli danışmanlık, dijital teknoloji ve operasyon hizmetleri sunan uluslararası bir şirkettir.

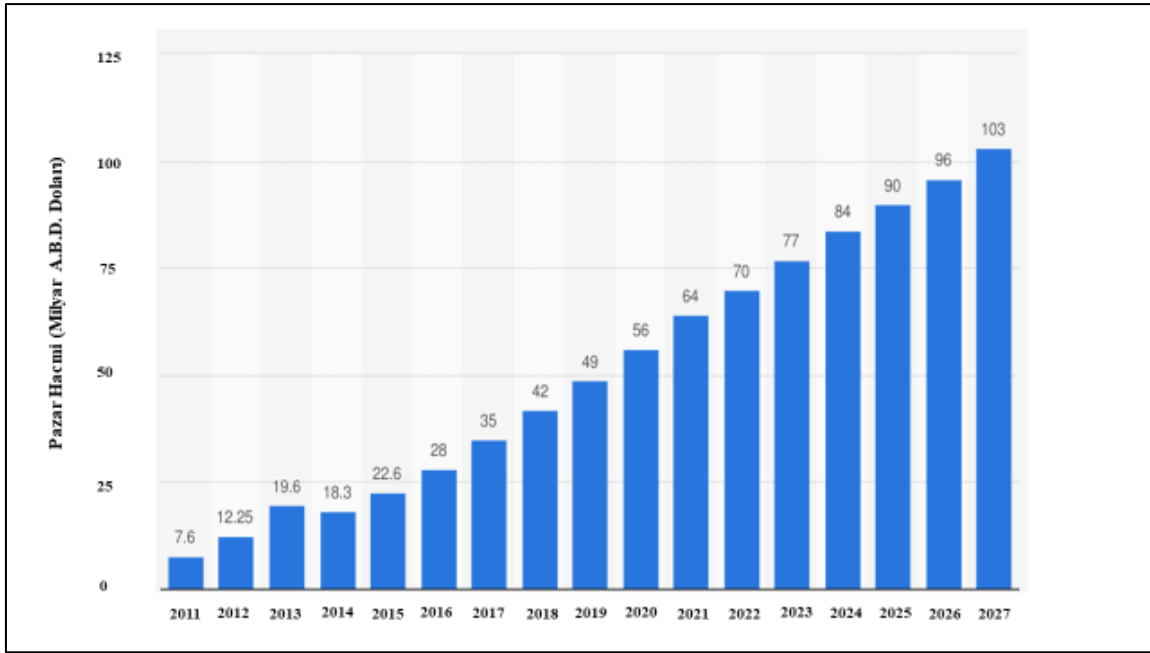
²⁵ BDA (Big Data and Analytics): Büyük Veri ve Analitiği

- Veri üzerinden para kazanma çabaları, pazarın kendi verilerinin tüketimini 100 kat veya daha fazla artıran dijital dönüşüm girişimlerini sürdüren işletmelere yol açacaktır.
- Harekete geçirilebilir zekâyı elde etmek için analiz etmeye değer yüksek değerli veriler ikiye katlanacaktır.
- Karar vericilere iletilen bilgilerin %60'ı, mevcut (2015) seviyenin iki katına çıkarak daima harekete geçirilebilir sayılacaktır.
- İlgili tüm verileri analiz edebilen ve harekete geçirilebilir bilgiler sağlayabilen kuruluşlar, daha az analitik yönelimli meslektaşlarına kıyasla üretkenlik avantajları açısından 430 milyar dolar daha fazla kazanacaklardır.

Statista²⁶, küresel büyük veri pazarının 2027 yılına kadar, 2018'de beklenen pazar büyüklüğünün iki katından fazla artarak 103 milyar dolara yükseleceğini tahmin etmektedir. Bu büyüme, yıllık ortalama %10.48'lik bir büyümeye denk gelmektedir. 2027 yılına kadar %45'lik payla, yazılım segmenti büyük veri pazarı segmentine dönüşecektir. Statista'nın yayınladığı Şekil 2.7'deki grafikte dünya çapında 2011- 2027 yılları arasında büyük veri pazarı gelir tahmini gösterilmektedir (Liu, 2019). NewVantage Partners²⁷'a göre, büyük veri, şirketlere en çok katma değeri, maliyetlerin azaltılması konusunda sağlamakta olduğunu belirtmiştir.

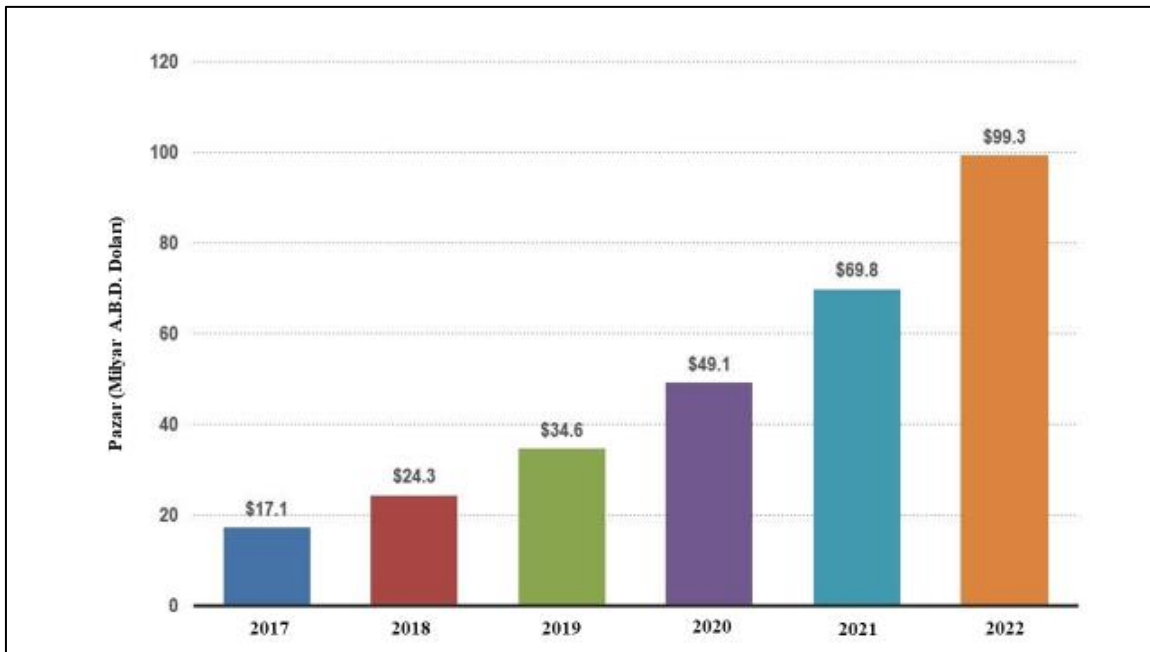
²⁶ Almanya'da kurulmuş, ekonomi ve resmi istatistiklerden araştırmalar yayınlayan bir çevrimiçi portaldır.

²⁷ Boston, ABD'de 2001 yılında kurulmuş, büyük veri strateji danışmanlığı yapan bir şirkettir.



Şekil 2.7. Dünya çapında büyük veri pazarı gelir tahmini (milyar ABD doları), 2011-2027 (Statista, 2018)

Statista'nın verilerine göre, Hadoop ve Büyük Veri pazarının 2017-2022 yılları arasında yıllık %28.5 büyüyerek, 17.1 milyar dolardan 99.31 milyar dolara çıkması beklenmektedir (Şekil 2.8). En fazla büyümenin 2021 ile 2022 yılları arasında olacağı düşünülmekte, pazarın bir yıl içinde 30 milyar dolar büyüyeceği öngörülmektedir (Columbus, 2018).



Şekil 2.8. Dünya çapında Hadoop ve büyük veri pazar hacmi, 2017-2022 (Statista, 2018)

IDC'nin verilerine göre, Türkiye'de Büyük Veri ve Analitik yatırımları 2018 yılında 247 milyon dolar olduğu ve 2023 yılına kadar yıllık %16 büyüme oranıyla 520 milyon dolara ulaşacağı tahmin edilmektedir. Türkiye'de, kurumların %55'i şirket içi büyük veri işleme konusunda çalışmalarını sürdürdüğü ve şirketlerin %50'si geçmiş verileri analiz etmek için çok sayıda veri kaynağını bir araya getirmekte zorlandığını belirtmektedir (DigitalAge, 2019). 2022 yılında, Büyük Veri ve Analitiği'nin Türkiye'de en çok kullanılacağı üç sektörün telekom, bankacılık ve merkezi yönetim (kamu) olacağı öngörülmektedir. IDC'nin tahminleri göre 2022 yılında Türkiye'de Büyük Veri ve Analitiği ile alakalı en çok kullanılacak olan yazılımlar, İlişkisel Veri Ambarı Yönetimi, Son-Kullanıcı Sorgulama, Raporlama, Analiz ve İleri Öngörülü Yazılım Analitiği olacağı belirtilmektedir (IDC, 2018).

2.4. Büyük Veri ve İnsan Kaynağı İhtiyacı

Kurumlar daha önce hiç karşılaşmadıkları çeşitlilikte ve büyüklükte bir veri potansiyeli ile karşı karşıya gelmiş durumdadır. Büyük veriden fayda sağlamak için öncelikle bu çok büyük hacimli bilgi yığınının ehlileştirilmesi ve mantıksal analizlere imkân sağlayan bir formata dönüştürülmesi gerekmektedir (Davenport, Patil ve Cohen, 2013). Büyük veri pazar hacmi ve önümüzdeki yıllardaki pazar hacmi tahminleri göz önünde bulundurulduğunda, bu alanda insan kaynağına yani alanında uzman teknik personele olan ihtiyacın ciddi oranlarda artacağı öngörülmektedir.

Şirketler, gelişmiş analitik ekipleri oluşturmak için işe alım yaparken belirli becerileri daha fazla talep etmektedirler. İleri analitik yeteneğe sahip olan bir ekibin sekiz kilit rolden faydalanması gerekmektedir. Bu roller, veri mühendisi, veri bilimci, veri mimarı, makine öğrenmesi mühendisi, ürün yöneticisi, veri analisti, yazılımcı ve sistem yönetici mühendisi ve kullanıcı arayüz geliştiricisinden oluşmaktadır. 2018 yılında dünya genelinde ileri analitik yeteneğin 2020 yılına kadar iki katına çıkması beklenmektedir. İleri analitik eğitimi alan yeteneklerin büyümesi ABD'den sonra, özellikle en çok iki eğilimden dolayı Hindistan'da gerçekleşecektir. Bu eğilimlerden ilki; eğitim programı veri ve analitik

becerileri vurgulayan STEM²⁸ lisans ve lisansüstü mezunlarının artan sayıda iş gücüne katılmaları, ikincisi; Hindistan'ın bilişim teknolojilerinde özellikle programlama ve sistem entegrasyonu alanlarındaki mevcut derin ekosistemidir (Brahm, Sheth, Sinha ve Dai, 2019:4).

McKinsey Global Institute²⁹, 2018 yılına kadar Amerika Birleşik Devletleri'nin büyük veri alanında derin analitik yetenek gerektiren pozisyonlara olan talebin, öngörülen arzdan %50 - %60 (140 -190 bin) oranında daha fazla olacağını öngörmektedir (Manyika ve diğerleri, 2011). Capgemini³⁰ ve Informatica³¹ (2016)'nın yayınladığı raporda oluşturulan ankete, yarısı Amerika'da diğer yarısı İngiltere, Almanya, Fransa, İtalya ve Hollanda'da yer alan, personel sayısı 1000'in üzerinde ve ortalama personel sayısı 23000 olan şirketlerde çalışan üst düzey yöneticiler (%85 Bilgi Teknolojileri Yönetimi, %15 Veri Yönetimi) katılmıştır. Ankete göre büyük veriyi kullanarak anlamlı iş çözümleri üretirken karşılaşılan en önemli zorluklardan, nitelikli personel eksikliği (%32) 4. sırada yer almıştır. BARC Enstitüsü'nün ³² “Avrupada Büyük Veri Araştırması” adlı çalışmasında Almanya, Avusturya, İsviçre, Fransa ve İngiltere'de yer alan küçük, orta ve büyük ölçekli şirketlerde çalışan 274 üst düzey yöneticiye anket soruları yöneltilmiştir. Ankette yer alan ‘Büyük veri teknolojilerini kullanırken hangi problemlerle karşılaşıyorsunuz?’ sorusuna verilen cevaplara göre, %46 oranla yetersiz teknik uzmanlık ilk sırada, %44 oranla yetersiz analitik uzmanlık ikinci sırada yer almaktadır (Bange, Grosser ve Janoschek, 2013:26).

Analytics Center ve Kariyer.net, Türkiye’de analitik yetenek ve yetkinlikler üzerine yaptığı araştırmada, Kariyer.net’in veri tabanında yer alan 70 bine yakın firma içerisinde 2008 yılından itibaren 621 firma 2657 ilanında veri analitiği ve iş zekası gibi alanlara yer verdiğini açıklamaktadır. Aynı araştırmada, Türkiye’deki uzmanlıkların dağılımı, %47 veri tabanı, %34 iş zekası, %10 veri analizi, %6 veri ambarı, %2 veri madenciliği, %1 veri bilimi uzmanı olarak belirtilmektedir. Türkiye’de mevcut olarak veri odaklı şirketler %33, tam olarak veri odaklı olmayan şirketler %67 olarak belirtilmektedir. Analitik ve büyük veri alanında karşılaşılan zorluklar; teknik engeller; veri güvenliği, performans,

²⁸ STEM: Fen (Science), Teknoloji (Technology), Mühendislik (Engineering), Matematik (Mathematics) gibi dört önemli disiplinin bir araya getirilmesiyle oluşturulan bir öğretim modelidir.

²⁹ 1962 yılında Chicago Üniversitesi profesörü James O. McKinsey tarafından kurulan bir yönetim ve danışmanlık şirkettir.

³⁰ Danışmanlık, teknoloji ve profesyonel hizmetler sunan çok uluslu bir Fransız şirkettir.

³¹ 1993 yılında ABD’de kurulmuş bir yazılım geliştirme şirkettir.

³² BARC (Bhabha Atomic Research Center): Hindistanın önde gelen nükleer araştırma merkezidir.

ölçeklenebilirlik, ulaşılabilirlik, maliyet; organizasyonel engeller; yetenek ve yetkinlikler; kültürel engeller; veri odaklı kültür olarak belirtilmektedir. Veri odaklı olmanın önündeki engel olarak birinci sırada organizasyon yapısı ve ikinci sırada çalışan yetkinlik ve bilgi seviyesi yer almaktadır (Smartcon, 2016).

2.5. Kamuda Büyük Veri

Türkiye Bilişim Derneği'nin 2016 yılında yayınladığı, “Büyük Veri Uygulamaları Çalışma Grubu Raporu” na göre, Türkiye’de aktif olarak büyük veri üzerine kurgulanmış bir kamu projesi bulunmadığı belirtilmiştir (TBD, 2016). Ancak TBD (2016) tarafından yayımlanan raporda, ülkemizde büyük veri çalışmalarına girdi sağlayabilecek belli başlı kamu kurum projeleri belirtilmiştir (Çizelge 2.7).

Çizelge 2.7. Türkiye’de büyük veri çalışmalarına veri sağlayabilecek kamu projeleri (TBD, 2016)

Kurum	Proje
Sosyal Güvenlik Kurumu (SGK)	e-Bildirge Sistemi
	MEDULA
	Aylık Tahsis ve Diğer Uygulamalar
	ALO 170 Projesi
	Veri Ambarı Projesi
Milli Eğitim Bakanlığı (MEB)	MEBBİS
	e-Okul
	Fatih Projesi
	e-Yangın
	ALO 147 Projesi
Sağlık Bakanlığı	E-Nabız
	USBS Projesi (Sağlık.Net)
	MHRS Projesi

Kamu Bilişim Merkezi Yöneticileri Birliği Kamu Bilişim Platformu tarafından hazırlanan ‘Büyük Veri Uygulamaları’ raporunda, Türkiye’de kamu kurumlarındaki ‘büyük veri’ resminin çekilmesi, kurumlarda mevcut ve potansiyel büyük veri uygulamalarının incelenmesi ve nihai olarak kamu kurumlarına büyük veri konusunda bir stratejik yol haritasının sunulması hedeflenmiştir. Raporda, kamuda büyük veri kullanım alanlarına, büyük veri içerikli projelere ve uygulamalara değinilmiş, kamuda büyük veri algısı anket sorularıyla incelenmiştir. Ankete, ulaştırma, belediye, eğitim, ekonomi ve sağlık kurumları başta olmak üzere 18 kamu kurumu katılmıştır. Raporda yer alan ‘Kamu Kurumlarında Büyük Veri Anketi’ne göre, kurumlardan alınan geri dönüşler doğrultusunda, ‘Büyük veri

projelerinde çalışan sayınız kaçtır?’ sorusuna verilen cevaplara göre, %48,3’ü büyük veri konusunda henüz personel istihdamı yapmadıklarını, %34,5’i 1-5 arası çalışan ile, %6,9’u 6-10 arası çalışan ile, %10,3’ü 10’un üzerinde çalışan ile büyük veri projelerini geliştirdiklerini belirtmişlerdir. Ayrıca ankete katılan kurumların yarısında büyük veri analitiği konusunda yetkin çalışan olmadığı belirtilmektedir (TBD, 2016).

TÜBİTAK BİLGEM’in ‘Bulut Bilişim ve Büyük Veri Çalıştayı’na, kamu kurumları düzeyinde büyük veri açısından mevcut durumun tespit edilmesine yönelik ankete davetli 132 kurumdan toplam 70 kurum katılım sağlamıştır. Anket sonuçlarına göre, ankete katılan kurumlarda veri işleme ve analizinde aşılması gereken sorun olarak ilk sırada nitelikli personel (55 kişi), ikinci sırada farklı kaynaklardaki verilerin birleştirilmesi (54 kişi) ve üçüncü sırada farkındalık (44 kişi) olarak belirlenmiştir (TÜBİTAK, 2016:27). Bu çalıştayda tahminleme, öngörü ve karar destek odaklı kamu ihtiyaçlarının büyük veri analitiği ile karşılanmasına yönelik kullanım senaryolarının belirlenmesi hedeflenmektedir. Bu doğrultuda oluşturulan ihtiyaçlar listesinde, akademik ihtiyaçlar şu şekilde belirtilmiştir:

1. Çocukların ilkokuldan itibaren mevcut yetenekleri doğrultusunda doğru alanlara yönlendirilmesi.
2. Eğitim kurumlarındaki sosyal, kültürel, sportif ve sanatsal etkinlikler ile kurumların akademik başarı düzeyleri arasındaki ilişkinin tespit edilmesi.
3. Üniversite okumaması gereken bireylerin tespit edilmesi gerekir.
4. Sınavlarda öğrencilerin test konularındaki başarı oranlarına göre milli eğitim müfredatının gözden geçirilmesi. Üniversite giriş sınavlarında, Türkiye'deki illerin başarı durumlarına göre MEB'in eğitim alanında (orta öğretim) strateji oluşturmaya ihtiyaç var.
5. Eğitim kurumlarının varlıkları ve performanslarına yönelik bölgesel haritalar oluşturulması.
6. Yurtdışında çalışan üniversite mezunlarının envanterinin oluşturulması.
7. Eğitimde sayısal içeriğin nitelik ve nicelik olarak artırılması (FATİH Projesi bağlamında ele alınmalı).
8. Üniversitelerin kontenjan ve kayıt miktarlarının geleceğe dönük değerlendirilmesi.
9. Öğrencinin yetenek, bilgi ve ilgisinin anlık ölçülüp sınavsız bir eğitim sistemi oluşturulması.

2.5.1. Türkiye’de büyük veri ile ilgili planlar

Bilgi ve iletişim teknolojilerindeki gelişmeler, kamu ve özel kurumlarda iş geliştirme ve gerçekleştirme süreçlerinde değişikliklere sebep olmuştur. İnternetin gelişimi ve veri tabanlarındaki yeniliklerle beraber kamuda ‘e-devlet’ olarak adlandırılan yenilikler ortaya çıkmıştır (Hughes, 2012). Bu teknolojik yenilikler, hükümetin bakanlıklar düzeyinde yayınladıkları stratejik eylem planları ve kalkınma planlarında yer almaktadır.

Türkiye Cumhuriyeti Cumhurbaşkanlığı, Strateji Ve Bütçe Başkanlığı Stratejik Planında (2019-2023), teknolojik olarak tespit edilen sorunlardan bir tanesi, ‘Büyük veri ve ekonomi araştırmalarında kullanılan teknolojilerin çeşitlenmesi ve gelişmesi’ olarak belirtilmiş ve çözüm olarak ‘Kurumlar arasında veri paylaşımı altyapısının güçlendirilmesi’ olarak belirlenmiştir. Aynı zamanda, Amaç 4’de belirtilen “Ekonomik ve sosyal kalkınma alanlarında karar süreçlerine destek sağlayacak kısa, orta ve uzun vadeli tahmin, analiz ve araştırma çalışmalarını etkin bir şekilde yürütmek” amacının ikinci hedefinde, “Hedef 4.2: Büyük veri temelli politika ve etki analizi çalışmaları yapılacaktır.” ifadesi yer almaktadır. Bu hedef kapsamında, “Büyük veri analizi konusunda tecrübeli uzman/veri bilimci kapasitesinin artırılması ve altyapının güçlendirilmesi” ve “Kamu politikalarının belirlenmesinde büyük veriye dayalı analizlerin kullanılması” ihtiyaç olarak belirtilmiştir. Yine bu hedef kapsamında, “Dünyada gelişen büyük veri kullanımının Türkiye’de henüz yaygınlaşmamış olması” ve “Büyük veri çalışmalarının güçlü donanım/bilgi teknoloji altyapısı gerektirmesi” tespit edilmiştir. Bu hedef kapsamında belirlenen stratejiler, “Büyük veriye dayalı kamu politikaları belirlenecektir.” ve “Büyük veriye dayalı analiz ve modelleme konusunda teknik kapasite geliştirilecektir.” olarak belirlenmiştir (Strateji Ve Bütçe Başkanlığı, 2018:60). Cumhurbaşkanlığı Strateji Ve Bütçe Başkanlığı’nın yayınladığı On Birinci Kalkınma Planında (2019-2023);

- 77. maddesinde, “Kamu yönetiminin karar alma ve uygulama kapasitesinin artırılmasında büyük veri uygulamaları önem kazanmaktadır. Kamunun ürettiği verilerin yanı sıra, mobil uygulamalar ve sosyal medya aracılığıyla oluşturulan veriler; sağlık, trafik, asayiş gibi kamu hizmetlerine ilişkin mevcut ve muhtemel sorunların tespit ve çözümüne imkân sağlamaktadır.”

- 244.6. maddesinde, “Tarım ürünlerinde arz ve rekolte tahminlerinin sağlıklı yapılabilmesine imkân veren bir erken uyarı sistemi kurularak, gıda ürünlerinde fiyat dalgalanmasının büyük veri ve ileri analitik yöntemler kullanılarak yakından takip edileceği Ürün Gözetim Mekanizması hayata geçirilecektir.”
- 355. maddesinde, “Ülkemizde Milli Teknoloji Hamlesinin gerçekleştirilmesine yönelik olarak yapay zekâ, nesnelerin interneti, artırılmış gerçeklik, büyük veri, siber güvenlik, enerji depolama, ileri malzeme, robotik, mikro/nano/opto-elektronik, biyoteknoloji, kuantum, sensör teknolojileri ve katmanlı imalat teknolojilerine ilişkin gelişim yol haritalarının hazırlanması, gerekli altyapının tesis edilmesi, ihtiyaç duyulan nitelikli insan kaynağının yetiştirilmesi ve toplumsal yönelimin bu alanlara odaklanması sağlanacaktır.”
- 428.2. maddesinde, “Turizm sektöründe ürün geliştirme, tanıtım ve pazarlamada kullanılmasına yönelik büyük verinin toplanması ve analizi çalışması yürütülecektir.”
- 809.2. maddesinde, “Kamu hizmetlerinin iyileştirilmesinde büyük veri, bulut bilişim, mobil platformlar, nesnelerin interneti, yapay zekâ, blok zincir gibi yeni teknolojilerden faydalanılabilmesi için süreç ve teknolojik altyapı iyileştirmeleri yapılacaktır.” ifadeleri yer almaktadır (Strateji ve Bütçe Başkanlığı, 2019).

Kalkınma planları ve stratejik plarlarda büyük veriye verilen önem dikkat çekicidir. Özellikle, Strateji Ve Bütçe Başkanlığı'nın yayınladığı son stratejik planda, büyük veri analizi konusunda tecrübeli uzman/veri bilimci kapasitesinin artırılması ve altyapının güçlendirilmesi bir ihtiyaç olarak belirtilirken, Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) Başkanlığı ve Milli Eğitim Bakanlığının (MEB) en son yayınlanan stratejik planlarında büyük veri ile ilgili herhangi bir ifadeye yer verilmemiştir (ÖSYM, 2018; MEB, 2018). Yükseköğretim Kurulu tarafından 100/2000 YÖK Doktora Bursu programı ile 2019-2020 eğitim öğretim yılı bahar dönemi çağrısında ülkemizde doktoralı insan kaynağına olan ihtiyacın karşılanması için yurtiçinde devlet üniversitelerinde 100 tematik alanda YÖK doktora bursu verileceği açıklanmıştır. Bu alanlar içinde Fen ve Mühendislik Bilimleri alt alanlarda, Veri Bilimi ve Bulut Bilişim alanına Büyük Veri Teknolojileri de eklenmiştir (YÖK, 2019). Ayrıca YÖK, 2020 yılında belirlenen öncelikli alanlarda araştırma görevlisi ve öğretim üyesi istihdamının gerçekleştirileceği “YÖK-Gelecek

Projesini” başlatmıştır. Bu alanlar içerisinde “Büyük Veri ve Veri Analitiği” de yer almaktadır (YÖK, 2020).

2.5.2. Dünyada büyük veri ile ilgili planlar

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 (NSF, 2019). Ulusal Bilim Vakfı, Bütünleştirici Faaliyetler Ofisi Başkanı Suzanne Iacono, ABD Federal Hükümetin üç nedenden dolayı büyük veriyle ilgilendiğini ifade etmektedir;

1. Ticareti ve ekonomiyi canlandırmak
2. Keşif hızını arttırmak ve yeni faaliyetler sağlamak
3. Eğitim, sağlık ve kamu güvenliği konularındaki ulusal zorlukların üstesinden gelmek

2012 yılında, hükümetin büyük veri Ar-Ge’sini geliştirmek için 200 milyon dolar tahsis edilmiştir. Büyük veri Ar-Ge çalışmalarına yönelik çerçevede işgücü ve eğitime yeni yaklaşımlar geliştirmek de yer almaktadır. 2012 ve 2013 yılında, Ulusal Bilim Vakfı ve Ulusal Sağlık Enstitüsü’nün büyük veri üzerine 45 proje ödüllendirdiği belirtilmektedir. Bu projelerin yarısı veri toplama ve yönetimi ile dörtte biri sağlık ve biyoinformatik ile ilgili olurken geri kalanı, sosyal ağlar, fizik ve mühendislik, algoritmalar ve siber altyapı ile ilgili olmuştur (National Academies, 2014).

2011 yılında, ABD’de Federal Büyük Veri Ar&Ge faaliyetlerini koordine etmek amacıyla gerçek zamanlı, çeşitli ve büyük miktardaki verilerin etkili analizi ve bilgi keşfi için Büyük Veri Kurumlararası Çalışma Grubu (BD IWG)³³ 17 kurumun katılımıyla kurulmuştur (NITRD, 2020). Obama Yönetimi, büyük veri teknolojilerini geliştirmek, büyük veri uygulamalarını göstermek ve yeni nesil veri bilimcilerini eğitmek için 2012 yılında Büyük Veri Araştırma ve Geliştirme Girişimi’ni başlatmıştır. Obama yönetimi aynı zamanda,

³³ BG IWG: Büyük Veri Kurumlararası Çalışma Grubu (Big Data Interagency Working Group)

büyük verinin gizlilik ve medeni özgürlükler gibi temel değerleri artıracak şekilde kullanılmasını sağlamak için Beyaz Saray’a ilk kez Veri Bilimi Yöneticisi işe almıştır. Yönetim, 2016 yılında, Federal Büyük Veri Ar-Ge planlarının geliştirilmesi veya genişletilmesi için rehberlik sağlayan Federal Büyük Veri Araştırma ve Geliştirme Stratejik Planı yayınlamıştır. 2012 Büyük Veri Araştırma ve Geliştirme Girişimi, zengin büyük veri kaynaklarından faydalanabilmek için geliştirilen bu plan için önemli bir dönüm noktasıdır (Marzullo, 2016). Federal Büyük Veri Araştırma ve Geliştirme Stratejik Planı, büyük veri araştırma ve geliştirme için önemli görülen temel alanları temsil eden aşağıdaki yedi strateji etrafında inşa edilmiştir;

1. Büyük veri temelleri, teknikleri ve teknolojilerini kullanarak yeni nesil yetenekler oluşturma
2. Verinin güvenilirliğini anlamak ve keşfetmek, daha iyi kararlar vermek, yeni keşifler sağlamak ve güvenli bir şekilde harekete geçmek için Ar-Ge’yi destekleme
3. Temsilci misyonlarını destekleyerek büyük veri inovasyonu sağlayan siber altyapı araştırmasını oluşturmak ve geliştirmek
4. Veri yönetimi ve paylaşımını destekleyen politikalar yoluyla verinin değerini artırma
5. Büyük veri toplama, paylaşma, güvenlik, gizlilik ve etikle alakalı olarak kullanımına yönelik anlama
6. Daha geniş işgücüne yönelik hem derin analitik yetenek hem de analitik kapasite talebinin artması için büyük veri eğitimi ve öğretiminin ulusal olarak geliştirilmesi
7. Ulusal Büyük Veri inovasyon ekosisteminde bağlantılar sağlama ve geliştirme

ABD’de gerçekleşen bu girişimlerden sonra, ABD ve dünyanın diğer bölgelerinde yer alan üniversitelerin Bilgisayar Bilimleri alanında en çok takip ettikleri müfredatları yayınlayan ACM³⁴ ve IEEE³⁵ kuruluşları 2020 yılı bilişim müfredatına, Siber Güvenlik ve Veri Bilimi programlarını eklemiştir. Veri bilimi lisans programında bilişimin temelleri, veri gereksinimi ve yönetimi, veri erişimi ve depolanması, veri gizliliği, güvenliği ve

³⁴ ACM (Association for Computing Machinery): Bilgisayar bilimleri alanlarıyla uğraşan profesyonellere ve öğrencilere kaynak sağlayan, mesleki ve akademik yayınları ile toplantılar gerçekleştiren bir mesleki kuruluştur.

³⁵ IEEE (Institute of Electrical and Electronics Engineers): Elektrik, elektronik, bilgisayar, otomasyon, telekomünikasyon ve diğer birçok alanda, mühendislik teori ve uygulamalarının gelişimi için çalışan kar amacı olmayan, dünyanın önde gelen teknik organizasyonudur.

bütünlüğü, makine öğrenmesi, veri madenciliği, büyük veri, veri analizi ve sunumu gibi alanlar yer almaktadır. Veri bilimi müfredatında ayrıca matematik, ayrık yapılar, olasılık teorisi, istatistik ve lineer cebir gibi dersler de yer almaktadır (ACM-IEEE, 2020).

2009 yılında ABD, İngiltere, Kanada ve Yeni Zelanda gibi çeşitli ülkelerin hükümetleri, açık veriyi yaygınlaştırmak için girişimlerini duyurmuşlardır. Federal Hükümet tarafından oluşturulan yüksek değerli, makinada okunabilen veri kümelerine halk erişimini sağlamak amacıyla, 2009 Mayıs ayında Amerika Birleşik Devletleri Federal Bilişim Yönetici (CIO) tarafından başlatılan, şeffaf bir devlet için data.gov internet sitesi kurulmuştur. Site tarım, üretim, eğitim, enerji, finans, sağlık, yerel yönetimler, iklim, halk güvenliği, bilim ve araştırma gibi konularda veri setleri sunan veri bilimciler, teknoloji girişimcileri ve geliştiriciler için zengin bir kaynak sunmaktadır (Data.gov, 2019). 2011 yılında, Kanada Hükümeti açık verilerin makinede okunabilen formatlarda dağıtımını sağlamak için data.gc.ca internet sitesi adıyla açık veri portalını başlatmıştır. 2013 yılında Kanada Açık Veri'ye ek olarak Açık Bilgi ve Açık Diyalog portallarını da başlatmıştır (Open.canada.ca, 2019). İngiltere ve Yeni Zelanda 2010 yılında, insanların açık hükümet verilerini bulmalarına ve kullanmalarına yardımcı olmak ve hükümet yayıncılarına veri korumalarında destek olmak için kendi internet sitelerini kullanıma sunmuştur. Ekonomi, çevre, sağlık, eğitim, ulaşım, hükümet ve harcamaları, şehir ve kentler, savunma, suç ve adalet gibi konularda çok zengin veri setleri bulunmaktadır (Data.gov.uk, 2019; Data.govt.nz, 2019). 2013 yılında, İrlanda İş, Girişim ve Yenilik Bakanı tarafından İrlanda'da büyük veri farkındalığını ve gelişmelerini ortaya koyan Endüstri/Devlet Ortak İş Gücü bildirisini yayınlanmıştır. Avrupa şirketlerinin verinin verimli kullanımı ile üretkenliğin ve yeniliğin artmasını sağlamak amacıyla başta Almanya, İspanya ve İrlanda olmak üzere 2014 yılında kamu özel ortaklığı ile Büyük Veri Değeri Derneği (Big Data Value Association) kurulmuştur (Cavanillas, Curry ve Wahlster, 2016). Güney Kore'de, 2011 yılında Bilgi İletişim Teknolojileri konseyi tarafından büyük veri altyapı ve teknolojilerini geliştirmek için Büyük Veri Çalışma Grubu (Big Data Task Force) oluşturulmuştur. Japonya Eğitim, Kültür, Spor, Bilim ve Teknoloji Bakanlığı 2005-2011 yılları arasında üniversiteler ve araştırma enstitüleri işbirliği ile büyük veri teknolojilerinden daha çok faydalanmak için Bilgi Patlaması Devri İçin Yeni Bilgi Teknolojileri Altyapısı projesini yürütmüştür (TBD, 2016).

2.6. Büyük Veri ve Eğitim

Teknoloji, kurumların yeni zorluklarla yüzleşme yeteneklerini güçlendirerek, amacını ve değerini değiştirmektedir. Eğitim kurumları da, büyük veri ve analitiği kullanımı ile birlikte karar alma mekanizmalarını geliştirmektedir. Yükseköğretimde çok fazla miktarda verinin mevcut olduğu ancak bu verilerin verimli bir şekilde kullanılmadığı görülmektedir. McKinsey şirketinin raporunda, yükseköğretimin veri toplama, bilgi teknolojileri yoğunluğu ve veriye dayalı karar vermede endüstrinin tüm sektörleri arasında daha zayıf bir bağlantı olduğu dikkat çekiliyor. Kütüphane kullanımı, öğrenme yönetim sistemleri ve Twitter gibi sosyal etkileşimler yoluyla elde edilen bu dijital veriler, kurumsal süreçlerin başarısı ile ilgili girdilere yol açmaktadır. Büyük veri ve analitiği, yükseköğretime, karar verme ve kaynak yönetimi, risk analizi ile öğrenci başarısının artırılması, kurumsal büyümenin artırılması, yenilikçi modeller geliştirilmesi gibi değerler katabilmektedir (Tulasi, 2013).

Büyük veri ve analitiğinin yükseköğretime katacağı değerler şu şekilde sıralanmaktadır (Siemens ve Long, 2011):

1. İdari karar alma ve örgütsel kaynak tahsisini geliştirebilmektedir.
2. Risk altındaki öğrencileri tanımlayabilmekte ve öğrencilerin başarıya ulaşmalarına yardımcı olmak için müdahale sağlayabilmektedir. Eğitimciler, Moodle gibi Öğrenme Yönetim Sistemlerinde okunan mesajlarının, gönderilen tartışma mesajlarının ve ödevleri tamamlamanın analizi ile okulu bırakma riski olan öğrenciler belirlenebilmektedir.
3. Şeffaf veriler ve analizler yoluyla kurumun başarıları ve karşılaştıkları zorluklar ile ilgili olarak ortak bir anlayış oluşturabilmektedir.
4. Akademik modeller ve pedagojik yaklaşımların yanı sıra üniversite sistemini de yenileyerek değişiklik yapabilmektedir.
5. Sosyal ağlar ile teknik ve bilgi ağlarının birleşimi yoluyla karmaşık konuların anlaşılmasına yardımcı olabilmektedir: yani algoritmalar, veri ve risk altındaki zorlukları tanıyabilmekte ve bunlarla ilgili bilgi sağlayabilmektedir.
6. Yöneticilerin, karmaşık bir disiplindeki çeşitli unsurların (örneğin, maliyetlerin düşürülmesi) etkisini araştırmak için senaryoları ve deneyleri analiz ederek bütünsel karar verme sürecine geçmelerine yardımcı olabilmektedir.
7. Güncel bilgi sağlayarak ve zorluklara hızlı yanıt vererek kurumsal üretkenliği ve

etkinliđi artırabilmektedir.

8. Kurumsal yöneticiler tarafından yürütölen zor (örneğin, patentler, arařtırmalar) ve kolay (örneğin, itibar, profil, öđretim kalitesi) fakölte faaliyetlerini belirlemelerine yardımcı olabilmektedir.
9. Öđrencilere kendi öđrenme alışkanlıkları hakkında bilgi verebilmekte ve gelişim için önerilerde bulunabilmektedir. Örneđin, Maryland Üniversitesinin ‘Etkinliđimi Kontrol Et’ aracı gibi öđrenmeye yönelik analizler, öđrencilerin kendi etkinliklerini, sınıf arkadaşlarının etkinliklerinin anonim bir özetini ile karşılařtırmasına olanak sağlayabilmektedir.

Kuruluřların büyük veri analitiđinden anlayan nitelikli personele olan ihtiyacı sonucunda, řirketlerin dörtte biri bu alanda eğitim programları hazırladıklarını belirtmektedirler. Örneđin IBM’in Veri Bilimi Akademisi, sekiz günlük bir eğitim kampı ve ardından bireysel çevrimiçi öđrenme sağlayan ve IBM’in veri bilimcilerinin de eğitime dâhil olduđu bir projedir. Airbnb’nin Veri Üniversitesi, veri farkındalıđı, toplama, görselleřtirme ve ölçekleme alanında 30’dan fazla ders sunmaktadır ve üniversitenin ilk altı ayında 500’den fazla çalışan yer almıřtır. Diđer eğitim seçenekleri arasında ücretsiz MOOC³⁶’lar (Kitleysel Çevrimiçi Açık Dersler) ve ücretli eğitim programları bulunmaktadır (Brahm ve diđerleri, 2019:5)

2.6.1. Yükseköđretimde büyük veri

Scopus veri tabanında, büyük veri ve büyük veri analizi üzerine özellikle Yönetim, Bilgisayar Bilimleri, Karar Bilimleri ve Sosyal Bilimler alanlarında büyük veri zorluklarını, büyük veri teknolojilerini, tekniklerini, yöntemlerini ve yaklařımlarını konu alan, 1996 ve 2015 yılları arasında yayınlanmış 227 makale üzerinde sistematik bir literatür taraması gerçekleştirilmiřtir. Bu arařtırmada, yazarlar özellikle 2 sorunun cevabını aramaya yönelik çalışmalar yapmıřlardır (Sivarajah ve diđerleri, 2017). Bunlar:

1. Kurumlar tarafından büyük veri kullanımında teorikleřtirilen / öđngörölen / karşı karşıya kalınan farklı zorlukları nelerdir?

³⁶ MOOC (Massive Open Online Courses): İnternet üzerinden sınırsız katılım ve açık erişimle hizmet veren çevrimiçi dersler

2. Karşılaşılan zorlukların üstesinden gelmek için teorikleştirilen / öngörülen / istihdam edilen farklı büyük veri yöntem türleri nelerdir?

Büyük veri ve büyük veri analizi konularında yayınlanan makalelerin 1996'dan 2015 yılına kadar dünyada 42 farklı lokasyonda yayınlanan makale sayıları belirtilmiştir. Akademik olarak en çok katkısı yapan 241 kişiyle Çin, ardından 145 kişiyle Amerika Birleşik Devletleri olmuştur. Türkiye 2 kişiyle büyük veri ve büyük veri analizleri konusunda daha az yayın yapan bölgeler arasında yer almaktadır (Sivarajah ve diğerleri, 2017).

Türkiye'de YÖK tez veri tabanında büyük veri kavramı, 2012 yılına kadar tez özetinde ve tez başlığında düzenli bir artış göstermezken, 2012 yılından sonra her yıl büyük veri kavramının tez içeriklerinde düzenli bir şekilde artış gösterdiği görülmektedir. Türkiye'de YÖK tez veri tabanlarında büyük veri kavramına dizinlerde ilk defa 2014 yılında rastlanmıştır (Bayrakçı ve Albayrak, 2019).

Yükseköğretimin geleceğini şekillendirilmesinde ve geniş veri dizisinin verimli bir şekilde kullanılmasında büyük verinin kapsamlı bir çerçeve sunduğunu belirtilmektedir (Siemens ve Long, 2011). Aynı zamanda, yükseköğretimde büyük veri uygulamaları, teknolojik ve analitik gelişmelerin ilerlemesine yönelik katalizör görevi gördüğü kaydedilmektedir (Wagner ve Ice, 2012). Yükseköğretim bağlamında büyük veri, gelecekteki performansını öngörmek ve akademik programlama, araştırma, öğretim ve öğrenim ile ilgili potansiyel konuları belirlemek için kurumsal performansın değerlendirilmesinde, idari ve operasyonel veri toplama süreçlerinin yorumlanmasında yer almaktadır (Picciano, 2012). Daniel ve Butson (2013), mevcut literatürü düzenlemek ve araştırmaya yönelik yaklaşımların çerçevesini belirlemek için, yükseköğretimde büyük veriyi dört bileşen üzerinde tanımlayarak kavramsal bir çerçeve önermektedir.

1. Kurumsal Analitik: Kurumsal analitik, kurumsal düzeyde bütün departmanlarda ve bölümlerde iyileştirme yapma konusunda etkili kararlar almak için analiz edilebilecek çeşitli operasyonel verilere işaret etmektedir. Değerlendirme politikası analizleri, öğretim analitiği ve yapısal analitiği içermektedir.
2. Bilgi Teknolojileri Analitiği: Bilgi teknolojileri analitiği, teknolojinin geliştirilmesi veya uygulanması, veri standartları, araçlar, süreçler, örgütsel sinerji ve politikaların geliştirilmesi için gerekli olan kullanım ve performans

verilerini kapsamaktadır. Çeşitli sistemlerden gelen verileri (öğrenci bilgileri, öğrenim yönetimi ve mezun sistemleri) ve sınıf dışındaki öğrenme deneyimlerini yöneten sistemleri birleştirmeyi amaçlamaktadır.

3. Akademik/Program Analitiği: Akademik analitiği, yönetim, araştırma, kaynak dağıtımı ve yönetimini etkileyen yükseköğretimdeki bütün aktiviteleri kapsamaktadır (Tulasi, 2013). Belirli bir programda olup bitenler ve performans zorluklarına nasıl ulaşılabileceği konusunda genel bilgiler sağlamaktadır.
4. Öğrenme Analitiği: Öğrenme analitiği, öğrenme ve öğrenme ortamlarını optimize etmek ve anlamak amacıyla öğrencilerle ilgili ölçme, toplama, analiz ve raporlama verileri ile ilgilidir (Siemens ve Long, 2011). Öğrenme analitiği daha çok, bir kurumun öğretim ve öğrenim düzeyinde gerçekleştirilir ve öğrencinin başarısının geliştirilmesi ile büyük ölçüde ilgilidir (Jones, 2012).

Büyük veri kavramı, çevrimiçi öğrenme sistemlerinden toplanan verinin potansiyeli ile yükseköğretim paydaşlarının veri özelleştirilmesi ve kişiselleştirilmesi gibi konularda yeni fırsatlar sunmaktadır. Örneğin, öğrenciler, eğitmen tarafından sağlanan materyallerden çevrimiçi öğrenme gerçekleştirilebilmekte, öğrencilerin sistem üzerinden yaptıkları hareketlerle toplanan veriler sayesinde, eğitmen öğrencilerin durumuna göre ders yapısını değiştirerek kaliteyi arttırabilmektedir. Büyük veri analizi sayesinde eğitimde yapılandırılmış ve yapılandırılmamış veri kaynakları bulunmaktadır. Yapılandırılmış veri kaynakları, öğrenci kayıtları, finansal kayıtlar, e-kütüphane kayıtları, web tıklama davranışları gibi kayıtlardan oluşurken, yapılandırılmamış veri kaynakları sosyal ağ etkileşimli kaynaklar olabilmektedir. Büyük veri analitiği ile performans ve kalite kontrolünü sağlamak aynı zamanda öğrencilere daha iyi hizmet verebilmek için, öğrencilerin verilerinden örüntü çıkarılarak, hizmetin kişiselleştirilmesi ve modüllerin özelleştirilmesi imkânı sunmaktadır (Huda, Anshari, Almunawar ve Shahrill, 2016).

Büyük veri ve veri bilimi alanına 2010 yılından bu yana sağlık, telekomunikasyon, finans, bankacılık, kentleşme, teknoloji, eğitim gibi alanlarda, ABD ve Avrupa ülkelerinin farkındalıkları ve buna bağlı olarak 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 görülmektedir. ABD dışında veri bilimi lisans programı olan üniversiteler İngiltere, Yeni Zelanda, Güney Kore, İrlanda, Hong-Kong, Fransa, İspanya ve Almanya'da yer almaktadır (DataScience, 2019). Washington Üniversitesi'nden Bill Howe, MOOC'ların (Kitleli Çevrimiçi Açık Kurslar) büyük bir nüfusa büyük veri bilimi tekniklerini öğretmenin bir yolu olduğunu ancak, veri bilimi katılımcılarının öğrenciler değil, bilgisayar bilimleri uzmanları olma eğiliminde olduğunu belirtmektedir (National Academies, 2014). Bu doğrultuda bilgisayar bilimleri uzmanlarının veri bilimi alanında aldıkları eğitim konusunda yetersizlik yaşadıkları ortaya çıkmaktadır. Ancak veri bilimi bölümünün hangi fakültede ve departmanda yer aldığı konusunda bir standart oluşturulması gerekmektedir. Amazon Şirketinden Guy Lebanon, veri bilimi programlarının farklı konsantrasyonlarla farklı departmanlarda olmasını tutarsız bulduklarını belirtmektedir. Sonuç olarak, kurumlar bir mezunun hangi teknik beceri ve yeteneklere sahip olduğundan emin olamadıklarını ve bu konuda bir standart geliştirilmesinin faydalı olacağını belirtmiştir (National Academies, 2014).

Veri bilimi alanında kariyer yapmak isteyenlere programlar sunan 365 DataScience'ın verilerine göre, veri bilimcilerin mezun oldukları alanlar yedi akademik birime bölünmektedir. 2019 yılında, bu listenin en başında %22'lik payıyla Bilgisayar Bilimleri en üst sırada yer alırken, %12'lik oranla Veri Bilimi ve Analizi bölümleri dördüncü sırada yer almaktadır. Ancak 2020 yılında, Veri Bilimi ve Analizi %20.72'lik oranla ilk sıraya çıkmıştır. Ardından Bilgisayar Bilimleri (%18.3), İstatistik ve Matematik (%16.3), Ekonomi ve Sosyal Bilimler (%12.4), Mühendislikler (%10.8), Doğa Bilimleri (%10.6) ve Diğer Bölümler (%11) gelmektedir. Veri bilimcilerin %56.2'sini yüksek lisans, %26.6'sını doktora, %12.8'ini lisans mezunları oluşturmaktadır (365 DataScience, 2020).

Veri mühendisliği giderek daha fazla yazılım mühendisliği disiplini içinde bir alt küme olarak ortaya çıkmaktadır. Yazılım mühendisliği, veri mühendisliği ve veri bilimi 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'ının 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).

365 DataScience'ın verilerine göre veri bilimcilerin %74'ü Python, %56'sı R programlama dillerini, %51'i SQL sorgulama dilini, ve temel programlama dillerinden %21'i Matlab, %17'si Java, %15'i C/C++, ve %11'i SAS'i kullanmaktadır. Ayrıca toplamda veri bilimcilerinin %90'ı Python veya R dilini kullanmaktadır (365 DataScience, 2020).

Pace Üniversitesi, veri bilimi alanında 2015 yılında, 'Büyük Veri Analitiği Kavramları' dersini, lisans düzeyinde müfredatlarına eklemeye karar vermiştir. Veri bilimi alanı, lisans öğrencilerinin işletme, matematik, istatistik, grup çalışmaları ve üst düzey teknoloji becerilerini geliştirmeleri için tasarlanmıştır (Lawler ve Molluzzo, 2015).

2.6.2. Yükseköğretimde büyük veri eğitimi

Eğitim programları, bireyleri çalışacakları sektörün ihtiyaçlarına göre daha geniş anlamda hazırlamak zorundadır. Ancak, teknoloji, araçlar ve uygulamalar genişlemeye devam ederken, çoğunlukla resmi müfredatın statik niteliği bu ihtiyaçlara cevap verememektedir. Büyük veri ve ilgili teknolojileri içeren bu önemli alan, son on yılda çok hızlı bir şekilde gelişmektedir. ACM ve IEEE'nin 2005 yılında yayınladığı bilişim müfredatında, Bilgisayar Mühendisliği, Bilgisayar Bilimleri, Bilgi Sistemleri, Bilgi Teknolojileri ve Yazılım Mühendisliği olmak üzere beş temel lisans programı yer almaktadır (ACM ve IEEE, 2005:7). 20 ülkeden 50 uzmanın değerlendirmesiyle hazırlanan 2020 yılı bilişim müfredatında, bu beş alana ek olarak Siber Güvenlik ve Veri Bilimi programları eklenmiştir. Veri bilimi lisans programında bilişimin temelleri, veri gereksinimi ve yönetimi, veri erişimi ve depolanması, veri gizliliği, güvenliği ve bütünlüğü, makine öğrenmesi, veri madenciliği, büyük veri, veri analizi ve sunumu gibi alanlar yer almaktadır. Veri bilimi müfredatında ayrıca matematik, ayrık yapılar, olasılık teorisi, istatistik ve lineer cebir gibi dersler de yer almaktadır (ACM ve IEEE, 2020:30). 2013-2016 yılları arasında yayınlanan öneri belgelerinde, lisans müfredatı, dağıtık paralel işleme algoritmaları ve depolama yapıları gibi konuları bir ölçüde içeren dersler olsa da, büyük

veri varlığı olan dersler bulunmamaktadır (ACM ve IEEE, 2013:14, 2016:14). Büyük veri teknolojilerinde eğitim ve öğretim sağlamaya yönelik erken girişimler özellikle yüksek lisans düzeyinde veri analizi programları ve belirli uygulama türlerindeki sertifika kursları şeklinde ortaya çıkmıştır (Okur ve Büyükkeçeci, 2014).

Büyük Veri Yönetim Sistemleri (BVYS), bilimsel veri analizi, sosyal medya, veri madenciliği, öneri sistemleri ve web servis kayıtları üzerindeki korelasyon analizi gibi pek çok uygulama senaryosunda başarıyla kullanılmıştır. Dean ve Ghemawat (2004), BVYS ile ilgili Map-Reduce'ün temellerinin atıldığı makaleyi 2004 yılında yayınlamışlardır. BVYS'nin alanı nispeten yeni olsa da 2004'den bu yana hızla gelişme kaydetmekte ve düzinelerce açık kaynaklı, ticari ürün ortaya çıkmaktadır. Eğiticilerin karşılaştıkları ilk zorluklardan biri de bu farklı teknoloji ve ürünleri, bilgisayar bilimleri bölümlerinin müfredatına dahil edebilmektir. BVYS'deki temel üniteler için aşağıdaki alanlar önerilmektedir (Silva, Dietrich, Reed ve Tsosie, 2014):

- Map-Reduce: Büyük veri işlemek için temel programlama modellerinden biri olarak kabul edilmektedir. Makine kümelerinin ölçeklenebilir ve başarısızlığa dayanıklı olarak çalıştırılan yüksek oranda dağıtılmış programlar oluşturulmasını sağlamaktadır. Map-Reduce uygulamalarına bazı örnekler şunlardır: Apache Hadoop, Google Map-Reduce, Microsoft Dryad.
- NoSql: NoSql, geleneksel ilişkisel veri tabanlarından daha yüksek ölçeklenebilirlik ve kullanılabilirlik sağlamak üzere tasarlanmış veri depolarını içermektedir. Birçok NoSql veri tabanı sistemi, SQL sorgu dilini desteklememektedir ve basitleştirilmiş bir işlem ve tutarlılık modeli uygulamaktadır. NoSql veri tabanlarına bazı örnekler şunlardır: Apache Cassandra, Google Big Table, MongoDB ve Apache HBase.
- NewSql: NewSql, NoSql veri tabanı sistemlerinin ölçeklenebilirlik ve kullanılabilirlik açısından aynı düzeyde olmalarını hedeflemektedir ancak yine de geleneksel ilişkisel veritabanlarının ACID³⁷ özelliklerini ve SQL sorgu dilini korumaktadır. NewSql sistemlerine bazı örnekler şunlardır: VoltDB, NuoDB, Google Spanner ve Clustrix.

³⁷ ACID: Atomicity, Consistency, Isolation, Durability

Okur ve Büyükkeçeci (2014), büyük veriyi ilgilendiren uygulamaların, karar verme ve iş zekası içerdiğini dolayısıyla büyük verinin temel bileşenlerinin depolama, işleme ve veri analitiği olduğunu belirtmektedirler. Columbia Üniversitesinde, Büyük Veri Analitiği ve İleri Büyük Veri Analitiği derslerinin temel konuları, başlangıç seviyesinde büyük veri analitiği, platformlar, veri depolama ve veri işleme konularına odaklanılmıştır. Daha sonra, öneri, kümeleme ve sınıflandırma gibi farklı büyük veri analitiği algoritmalarına değinilmiştir. Son olarak, veri görselleştirme ve grafik hesaplama gibi konular yer almaktadır (Columbia Üniversitesi, 2020). Daniel (2017), büyük veri analitiğine odaklanan bir derse şu konuları önermektedir:

- Dağıtık ve bulut tabanlı veri yönetimi, veri temizleme ve veri entegrasyonu
- Meta verileri kullanma
- Yapılandırılmamış verilerin toplanması ve çıkarılması
- Olasılıksal ve öngörücü modelleme
- Örüntü tanıma
- Veri, metin ve görüntü madenciliği
- Ağ analizleri (Sosyal ilişkiler, yapısal sonuçlar, bilgi akışı)
- Anlamsal ağ ve ontolojiler
- Duygu analizi

Yükseköğretimde büyük veri alanına verilen önem küresel olarak artmakta, büyük veri alanının da içinde bulunduğu Veri Bilimi lisans programlarının sayısı her geçen yıl artmaktadır. Sadece Veri Bilimi olarak değil, Büyük Veri ve İşletme Analitiği (Suffolk Üniversitesi), Büyük Veri Yönetimi ve Uygulamaları (SUSTech), İstatistik ve Veri Bilimi (Colorado Eyalet Üniversitesi), Matematik ve Veri Bilimi (Güney Dakota Eyalet Üniversitesi), Moleküler Biyoloji ve Büyük Veri (Arizona Üniversitesi), İşletme Yönetimi ve Veri Analitiği (Kansas Eyalet Üniversitesi), Veri Bilimi ve İktisat (Colorado Eyalet Üniversitesi), Muhasebe ve Veri Analitiği (St. Mary's Üniversitesi) gibi çok çeşitli alanlarda büyük veri konusunu kapsayan lisans programları açılmaktadır (BachelorsPortal, 2021). Türkiye'de YÖK Atlas aramalarına göre büyük veri alanı ile ilgili herhangi lisans programı bulunmamaktadır. Ancak İstanbul Teknik Üniversitesi'nde 2020-2021 eğitim öğretim yılında ilk öğrencilerini kabul eden Yapay Zeka ve Veri Mühendisliği bölümü

açılmıştır. Bu doğrultuda bu çalışma büyük veri eğitimi ihtiyacının farkındalığına yönelik gerçekleştirilmiş dört ayrı çalışmadan oluşmaktadır. Bir sonraki bölümde bu çalışmaların genel amacından bahsedildikten sonra, her bir çalışmanın ayrı ayrı amaç, araştırma deseni, örneklem, veri toplama, araştırmanın kısıtlılıkları, verilerin analizi, güvenilirlik, geçerlilik ve bulgularına yer verilmiştir.





3. TEZ KAPSAMINDAKİ ÇALIŞMALARIMIZ

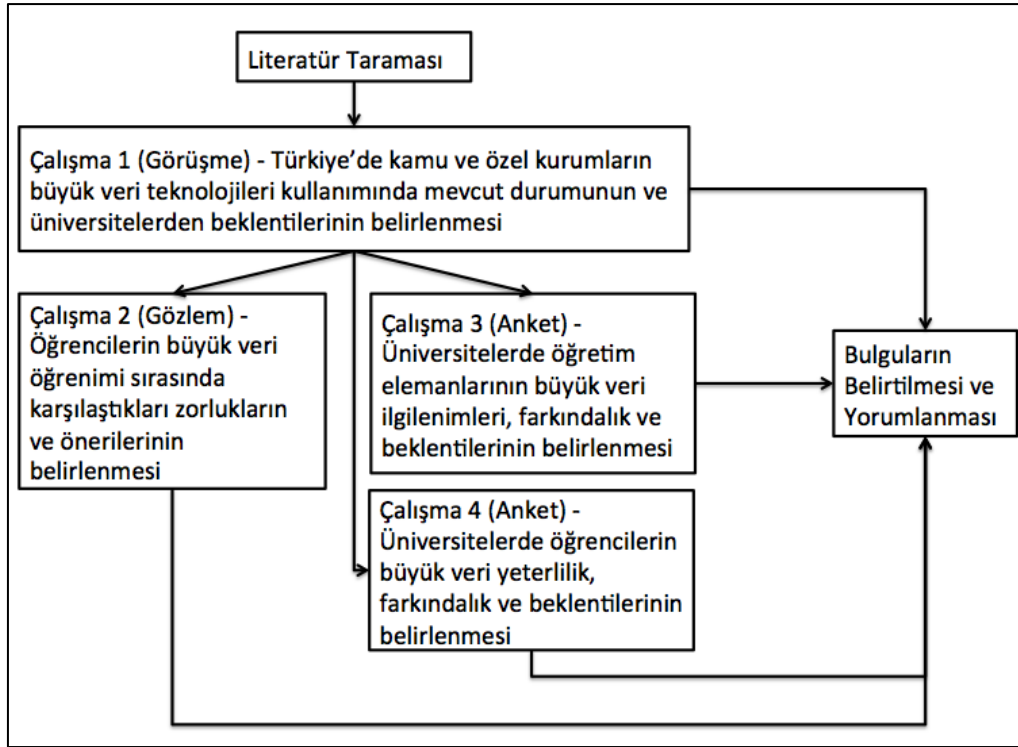
3.1. Çalışmalara Genel Bakış

Daha önceki bölümlerde değinildiği gibi, büyük veri kamu ve özel kurumlarda yaygın bir şekilde kullanılmakta ve büyük veri analizi alanında nitelikli personel ihtiyacı da gün geçtikçe artmaktadır. Ancak Türkiye’de büyük veri eğitimi ihtiyacının farkındalığına yönelik kapsamlı bir çalışmaya rastlanmamıştır.

3.1.1. Çalışmaların amacı ve kapsamı

Bu çalışmada, hem büyük veri çözümleri sunan kurumların mevcut durumunu ve beklentilerini, hem de öğretim elemanları ve öğrencilerin beklentilerini ve farkındalığını belirleyerek büyük veri eğitimi ihtiyacının farkındalığının belirlenmesi amaçlanmaktadır. İlk çalışmada, kamu ve özel kurumların büyük veri teknolojilerini benimseme düzeylerinin belirlenmesi amaçlanmaktadır. Bu doğrultuda görüşülen kurumların en çok tercih ettikleri büyük veri araçlarını ortaya koymak, büyük veri alanında çalışanların pozisyonlarını, çalışan sayılarını ve yeterliliğini belirlemektir. Aynı zamanda kurumların bu alanda ihtiyaç duyduğu nitelikli personel açığını kapatacak olan üniversitelerden ve öğrencilerden beklentileri belirlenmektedir. Ayrıca öğretim elemanlarının büyük veri ilgilenimleri, farkındalık ve beklentilerinin ve öğrencilerin büyük veri yeterlilik, farkındalık ve beklentilerinin tespit edilmesi ve büyük veri konusuna ilgili olan öğrencilerin büyük veri öğrenimi sırasında karşılaştıkları zorlukların ve deneyimlerin belirlenmesi amaçlanmaktadır.

Bu doğrultuda dört alt çalışma tasarlanmıştır ve çalışmanın kapsamı Şekil 3.1’de gösterilmektedir. Çalışma 1’de gerçekleştirilen görüşmeler ile belirlenen mevcut durum ve beklentiler doğrultusunda Çalışma 2, Çalışma 3 ve Çalışma 4 gerçekleştirilmiştir. Her bir çalışmanın bulguları ayrı ayrı yorumlanmıştır. Gerçekleştirilen dört çalışmanın genel bulgularına ise bir sonraki bölümde yer verilmiştir.



Şekil 3.1. Çalışmanın kapsamı

İlk çalışmanın amacı, kamu ve özel kurumların büyük veri teknolojilerinin kullanımını inceleyerek, büyük veri çözümleri sunan kurumların mevcut durumlarının ve beklentilerinin ne yönde olduğunu belirlemektir. Bu doğrultuda Türkiye’de en çok tercih edilen büyük veri araçlarını ortaya koymak ve kurumların bu alanda ihtiyaç duyduğu nitelikli personel açığını kapatmak için öğrencilerden ve üniversitelerden beklentilerini belirlemektir. Bu çalışmada nitel bir araştırma süreci tasarlanmıştır ve veriler kamu ve özel kurumların yöneticileri ile yapılan mülakatlar ile toplanmıştır.

İkinci çalışmanın amacı, büyük veri teknolojilerini öğrenirken öğrencilerin deneyimlerini değerlendirmek, karşılaştıkları zorlukları ve önerilerini belirlemektir. Bu çalışmada da, nitel metodoloji izlenerek veriler, bir günlük büyük veri eğitime katılan lisans öğrencilerinden, katılımlı gözlem tekniği ve katılımcıların öz-değerlendirmeleri ile toplanmıştır.

Üçüncü çalışmanın amacı, farklı sektörlerin nitelikli personel ihtiyacını karşılayan üniversitelerde öğretim elemanlarının büyük veri ilgilenimleri, farkındalık ve

beklentilerinin analiz edilmesidir. Nicel olarak kurgulanan bu çalışmada veriler, anket yoluyla toplanmıştır.

Dördüncü çalışmanın amacı, Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümleri lisans öğrencilerinin büyük veri yeterlilik, farkındalık ve beklentilerini ölçmektir. Nicel olarak kurgulanan bu çalışmada da veriler, anket yoluyla toplanmıştır. Bu doğrultuda büyük veri farkındalık ve beklenti ölçeği ile büyük veri farkındalık ve yeterlilik ölçeği, öğrencilerin büyük veri konusundaki farkındalığını, yeterliliğini ve beklentilerini ölçmeye yönelik olarak ilk defa bu çalışma kapsamında geliştirilmiştir. Bu çalışmada aynı zamanda, Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümlerinde okuyan öğrencilerin büyük veri konusundaki farkındalıkları arasında anlamlı bir fark olup olmadığı da incelenmiştir. Çalışma 1’de görüşülen kurumların beklentileri doğrultusunda, öğrencilerden bekledikleri teknik becerilerin kazanılacağı derslerin daha çok bu iki bölümde olması dolayısıyla bu karşılaştırma yapılmıştır. Gerçekleştirilen dört ayrı çalışmanın araştırma soruları, metodolojileri, veri toplama yöntemleri ve örnekleme Çizelge 3.1’de gösterilmektedir.

Çizelge 3.1. Araştırılan çalışmaların detayları

	Araştırma Sorusu	Metodoloji	Veri Toplama Yöntemi	Örneklem
Çalışma 1	Türkiye’de kamu ve özel kurumların büyük veri teknolojileri kullanımında mevcut durumu ve üniversitelerden beklentileri nelerdir?	Nitel	Görüşme	Ankara ilinde büyük veri çözümleri sunan kamu ve özel kurumlardan büyük veri proje yöneticileri (n=10)
Çalışma 2	Öğrencilerin büyük veri öğrenimi sırasında karşılaştıkları zorluklar ve önerileri nelerdir?	Nitel	Katılımlı Gözlem	Gazi Üniversitesi Bilgisayar Mühendisliği 3. ve 4. sınıf lisans öğrencileri (n=11)
Çalışma 3	Üniversitelerde öğretim elemanlarının büyük veri ilgilenimleri, farkındalık ve beklentileri nelerdir?	Nicel	Anket	Ankara ilindeki üniversitelerin bilgisayar mühendisliği ve ilgili diğer bölümlerin öğretim elemanları (n=153)
Çalışma 4	Üniversitelerde öğrencilerin büyük veri yeterlilik, farkındalık ve beklentileri nelerdir?	Nicel	Anket	Ankara Yıldırım Beyazıt Üniversitesi Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri lisans öğrencileri (3. ve 4. sınıf) (n=179)

Yapılan çalışmalar, kamu ve özel kurumların, öğretim elemanlarının ve öğrencilerin büyük veri alanında farkındalık ve beklentilerinin belirlenmesiyle, üniversitelere katkıda

bulunmayı hedeflemektedir. Türkiye’de üniversitelerde, büyük veri analitiğinin eğitim müfredatına entegre edilmesi hususunda yapılması gereken değişiklikler ve yenilikler tartışılıp, öneri olarak sunulması hedeflenmektedir. Ayrıca elde edilen bulgular, On Birinci Kalkınma Planında Yükseköğretim Kurulu’nun sorumlu kuruluş olduğu politika tedbirleri ile birlikte değerlendirilerek yapıcı eleştiriler geliştirilmesi planlanmaktadır.

Tüm araştırmaların etik açıdan değerlendirilmesi Gazi Üniversitesi Ölçme Değerlendirme Etik Alt Çalışma Grubu tarafından yapılan toplantıda görüşülmüş ve etik açıdan bir sakınca bulunmadığına oy birliği ile karar verilmiştir (Bkz. EK-1).

3.1.2. Çalışmaların önemi

Strateji ve Bütçe Başkanlığı Stratejik Planında (2019-2023), Hedef 4.2 kapsamında, “Büyük veri analizi konusunda tecrübeli uzman/veri bilimci kapasitesinin artırılması ve altyapının güçlendirilmesi” ihtiyaç olarak belirlenmiş ve “Dünyada gelişen büyük veri kullanımının Türkiye’de henüz yaygınlaşmamış olması” ve “Büyük veri çalışmalarının güçlü donanım/bilgi teknoloji altyapısı gerektirmesi” tespit edilmiştir (Strateji Ve Bütçe Başkanlığı, 2018:60). Stratejik planlarda, Türkiye’de büyük veri analizi alanında nitelikli personele ihtiyaç duyulmasına ve henüz Türkiye’de büyük veri kullanımının yaygınlaşmamış olmasına dikkat çekilmiştir. 2011 yılından bu yana özellikle ABD büyük veri alanında ortaya koyduğu politikalar ve yatırımlar sonucunda, ACM ve IEEE kuruluşları 2020 yılı bilişim müfredatına, Veri Bilimi programını eklemiştir. ABD’de 60’a yakın üniversitede veri bilimi lisans programının olduğu görülmektedir. ABD dışında veri bilimi lisans programı olan üniversiteler İngiltere, Yeni Zelanda, Güney Kore, İrlanda, Hong-Kong, Fransa, İspanya ve Almanya’da yer almaktadır (DataScience, 2019). Türkiye’de bu alanda ihtiyaçlar belirlenirken, veri bilimci kapasitesinin arttırılmasına katkı sağlayacak olan üniversitelerde yeni bölümler açılmasının ve müfredatların güncellenmesinin bu alana önemli bir katkı sağlayacağı düşünülmektedir.

Bu çalışma, çok yönlü bir araştırma metodu kullanılarak, Türkiye’de eğitim sektöründe, büyük veri alanında atılması gereken adımlara ışık tutmak için bir yol haritası olma

niteliğindedir. Kamu ve özel kurumların, öğretim elemanları ve öğrencilerin bu konudaki görüşleri dikkate alınarak geliştirilen öneriler, bu çalışmanın önemini ortaya koymaktadır.

3.2. Çalışma 1 – Kamu ve Özel Kurumların Büyük Veri Alanında Beklentilerinin Belirlenmesi

3.2.1. Amaç

Bu çalışmanın amacı, kamu ve özel kurumlarda büyük veri teknolojilerinin kullanımını incelemek ve kurumların beklentilerinin ne yönde olduğunu belirlemektir. Bu doğrultuda en çok tercih edilen büyük veri araçlarını ortaya koymak ve kurumların bu alanda ihtiyaç duyduğu nitelikli personel açığını kapatacak olan üniversitelerden ve öğrencilerden beklentilerini belirlemektir. 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 Mühendisliği ve ilgili bölümlerde okuyan öğrencilerden ve üniversitelerden büyük veri alanıyla ilgili beklentileri nelerdir?

3.2.2. Araştırma deseni

Bu çalışmada nitel araştırma yöntemi kullanılmıştır. Mayring (2011) nitel araştırmayı, nicel araştırmanın bir alternatifi olarak görülmemesi gerektiğini ve nitel araştırmanın temel noktalarından birinin genelleştirme süreci olduğunu vurgulamaktadır. Araştırmalardaki sonuçların genelleştirme sürecinin otomatik bir süreç olarak değil, genelleme dayanaklarının her bir durum için adım adım değerlendirilmesi gerektiğini belirtmektedir. Bu çalışmada, Mayring'in anlayışı dikkate alınarak sonuçlar tümevarımsal olarak adım adım genelleştirilmiştir. Nitel araştırmalar çoğu zaman değişime açık ve esneklik ile yeniden düzenlemeye açık olma anlayışına sahiptir (Yıldırım ve Şimşek, 2013: 53). Doğal ortama duyarlılık, araştırmacının katılımcı rolü, bütüncül yaklaşım, algıların ortaya konması, araştırma deseninde esneklik, tümevarıma dayalı analiz nitel araştırmaların altı temel özelliğini oluşturmaktadır (Yıldırım, 1999). Bu çalışmada, belirtilen özellikler

dikkate alınarak görüşmeler gerçekleştirilmiştir. Nitel araştırmalarda, araştırmacıya çalışmanın bütün aşamalarında kesin çizgilerle bir yönlendirme olmasa bile, esnek bir yaklaşım sağlayarak, araştırma aşamalarının tutarlı olmasına katkıda bulunan bir araştırma deseni belirlenmelidir. Nitel araştırma deseni, araştırmanın ne tür bir yaklaşım ile oluşturulması gerektiği ve araştırmanın aşamalarının bu yaklaşım çerçevesinde tutarlı olması için rehberlik eden bir strateji olduğu belirtilmektedir. Nicel araştırmada desen, araştırma sorularından, veri toplama yöntemi, veri analizi ve raporlaştırma aşamalarına kadar araştırmacıya her aşamada yol göstermektedir ancak nitel araştırmalarda araştırılan olay, olgu veya duruma göre değişebilen bir süreç söz konusudur (Yıldırım ve Şimşek, 2013). Bu yaklaşımda, veri toplama ve analiz birlikte yürütülmekte, veriler toplandıktan sonra analizle ortaya çıkan kavramlar ve süreçler sonraki veri toplama aşamalarına dahil edilmektedir. Bazı araştırmalarda ilk aşamada belirlenen görüşme soruları ile veri toplama işleminin sonuna doğru oluşturulan sorular arasında farklılıklar olabilmektedir (Yıldırım ve Şimşek, 2013). Bu çalışmada, nitel analizlerin esneklik özelliği çerçevesinde, ilk yapılan görüşmeler sonrasında, verilerin analizi sonucunda, diğer görüşmelere bazı sorular eklenmiştir.

3.2.3. Örneklem

Kurumlarla yapılan görüşmelerin örnekleme, olgu ve olayların keşfedilmesinde ve açıklanmasında yararlı olan amaçlı örnekleme seçimi gerçekleştirilmiştir. Amaçlı örneklemede, katılımcıları belirlerken hangi seçim kriterlerinin önemli olduğuna karar verilmesi önemlidir (Merriam, 2013). Bu çalışmada, kurumlarla yapılan mülakatlarda örneklem seçimindeki kriter, kurumun büyük veri çözümleri sunuyor olmasıdır. LeCompte ve diğerleri (1993), amaçlı örnekleme yerine ölçüte dayalı veya ölçüt tabanlı örnekleme terimini kullanmaktadır. Temel olarak kullanılan amaçlı örnekleme yöntemleri şu şekilde listelenebilir: tipik durum örnekleme, aşırı ve aykırı durum örnekleme, maksimum çeşitlilik örnekleme, benzeşik örnekleme, kritik durum örnekleme, kartopu veya zincir örnekleme, doğrulayıcı ve yanlışlayıcı örnekleme, ölçüt örnekleme ve kolay ulaşılabilir durum örnekleme (Patton, 1987). Bu çalışma için Patton tarafından geliştirilmiş olan örneklem sınıflandırmalarından kartopu örnekleme yöntemi tercih edilmiştir. Kartopu örnekleminin diğer adı zincir örneklemedir. Bu örnekleme yönteminde, çalışmada yer alan

kriterlere kolayca uyan katılımcılara yer verilmektedir ve araştırmacının problemine ilişkin, zengin bilgi kaynağı olabilecek kişi veya durumların belirlenmesinde etkilidir. Çalışma süresi ilerledikçe elde edilen isimler veya durumların kartopu gibi büyüyerek belirli isimlerin öne çıkması ile görüşülmesi gereken kişi sayısı ortaya çıkmaktadır (Yıldırım ve Şimşek, 2013: 53). Kartopu örneklemede temel olarak izlenen yol, istenilen özelliklerde bir kaç kişi veya durum belirlenerek, ilk görüşmelere başlandıktan sonra görüşülen kişilerin yönlendirmeleriyle zincirleme bir liste belirlemektir (Berg ve Lune, 2015:67). Kartopu örnekleme yöntemi sonucunda on kurumun genel müdürü ya da büyük veri proje müdürleri ile görüşmeler gerçekleştirilmiştir. Önceki bölümde anlatılan sektörlere göre büyük veri benimsenmesi (Şekil 2.8) dikkate alınarak, farklı sektörlere hizmet veren kurumlar tercih edilmiştir.

Bu aşamada araştırmalar, Ankara ili genelinde büyük veri çözümleri sunan kamudan 2 ve özel sektörden 8 kurumla nitel araştırma yöntemlerinde kullanılan görüşmeler (mülakatlar) yapılarak gerçekleştirilmiştir. Nitel araştırmada örneklem sayısı, toplanan verilere, sorulan sorulara, analizlere ve problemin durumuna göre değişmektedir. Temel kriter verinin doygunluğa ulaşmasıdır (Merriam, 2013: 79). Görüşülecek kişi sayısı nitel araştırma metodolojisine uygun olacak şekilde önceden belirlenmemiş, verinin doyuma ulaştığı aşamada görüşmeler yeterli bulunup durdurulmuştur. Çizelge 3.2’de görüşmecilerin demografik özellikleri; cinsiyeti, görevi ve tecrübesi ve kurumların özellikleri; kurum türü ve kuruluş yılı gösterilmiştir.

Çizelge 3.2. Görüşülen kurum ve görüşmecilerin profili

Kurum	Kurum Türü	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	Lab Yöneticisi	5

3.2.4. Veri toplama

Bu arařtırmada veri toplama derinlemesine mülakatlar üzerinden gerekleřtirilmiřtir. Mülakatlar, yapılarına gre tam yapılandırılmıř, yarı yapılandırılmıř ve yapılandırılmamıř olmak zere  bařlık altında toplanmaktadır (Merriam, 2013: 87). Bu arařtırma kapsamında grüşmeler yarı yapılandırılmıř olarak gerekleřtirilmiřtir. Yarı yapılandırılmıř grüşmelerin zellikleri řu řekilde belirtilmektedir (Merriam, 2013: 87);

- Grüşme klavuzu yarı yapılandırılmıř sorularını ierir.
- Sorular esnektir.
- Genellikle her grüşmeciden spesifik veriler toplanır.
- Grüşmenin byk kısmı cevaplanması istenen sorular veya konulardan oluřur.
- nceden belirlenmiř soru ile ilgili ayrıntılar yoktur.

Gerekleřtirilen yarı yapılandırılmıř grüşmelerde, katılımcıya belirli bir esneklik tanınmıřtır ancak katılımcının tamamen esnek olduėu yapılandırılmamıř grüşme trnden de kaınılmıřtır. Grüşme sresince nceden belirlenen sorular zerinde durularak, yukarıda deėinilen arařtırma sorularına yanıt alınabilecek noktalara odaklanılmıř ve katılımcıların bu ereve ierisinde kalmalarına dikkat edilmiřtir. Yz yze yapılan grüşmeler katılımcıların kendi tercihleriyle kendi ofislerinde gerekleřtirilmiřtir. Katılımcıların izni ile ses kaydı alınmıřtır.

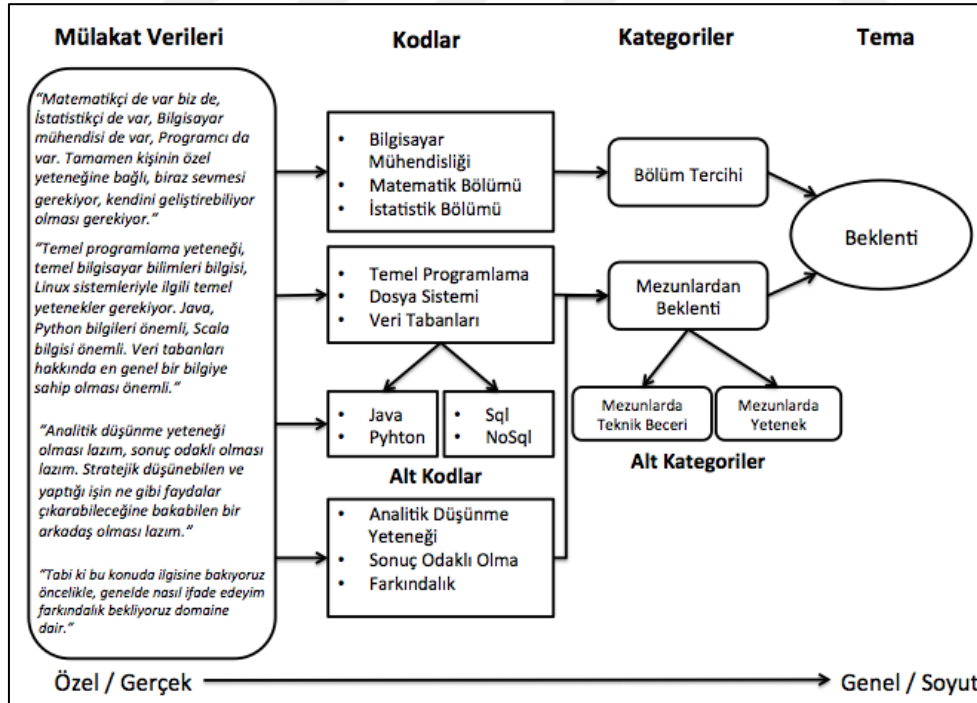
3.2.5. Arařtırmanın kısıtlılıkları

Arařtırmada grüşmeler, Ankara’da byk veri zmleri sunan kamu ve zel sektrden kk ve orta lekli kurumların byk veri proje ekibi liderleriyle yapılmıřtır. Byk veri alanında zm reten kurum sayısının azlıėı ve kurumların iř yoėunluėu sebebiyle aranılan kriterlerde katılımcı bulmak bu arařtırmanın kısıtlılıklarından biridir. Kk lekli iřletmelerde genellikle genel mdrle grüşme saėlanmıřtır ancak bir kurumda, byk veri projeleri ve diėer projeler birlikte yrtldėnden dolayı proje lideri yerine byk veri projelerinde yer alan kiři ile grüşme yapılmıřtır.

3.2.6. Verilerin analizi

Araştırmada, görüşmecilerden toplanan verilerin analizi için en uygun teknik, olgu-merkezli kuramlaştırma veya diğer adıyla gömülü teori olarak belirlenmiştir. Nitel araştırmalarda veri analizi, verinin anlamını ortaya çıkarma yani insanların ne söylediğini, araştırmacının ne gördüğünü ve okuduğunu birleştirme, indirgeme ve yorumlama sürecidir. Aynı zamanda nitel veri analizi, araştırma sorusunu cevaplamak için kullanılan bir süreçtir (Merriam, 2013: 167, 168).

Araştırmada, toplanan mülakat verilerinin analizi için öncelikle ses kayıtlarının transkripsiyon (deşifre) işlemleri yapılmıştır. Görüşme dökümleri dikkatli bir şekilde incelendikten sonra NVivo 12 programı kullanılarak kodlama işlemine tabi tutulmuştur. Nitel araştırmada kodlama süreci, elde edilen verileri bölümlere ayırmayı, incelemeyi, karşılaştırmayı, kavramlaştırmayı ve ilişkilendirmeyi gerektirir (Yıldırım ve Şimşek, 2013: 259). Kodlama benzer olarak kodlanmış verilerin kategoriler şeklinde gruplanmasını sağlamaktadır (Saldana, 2013: 3, 9).



Şekil 3.2. Çalışma verisi üzerinden nitel araştırmalarda koddan kategoriye akış modeli

Şekil 3.2’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 akış modeli örnek olarak gösterilmiştir. Kodlar ve kategoriler oluşturulurken alt kodlar ve alt kategoriler belirlenmiştir. Özelden genele, somut gerçekliklerden soyut genellemelere giden bu yaklaşıma tümevarımsal yaklaşım denilmektedir. Bu yaklaşımda, kategori ve tema arasında kesin bir geçişten bahsetmek pek mümkün değildir.

Kategoriler ve temalar, Glaser ve Strauss (1967) tarafından ortaya atılan olgu-merkezli kuramlaştırma yöntemi kullanılarak oluşturulmuştur. Bu teknikte araştırmacı veri toplama aşamasında, yorumlama için çerçeve oluşturabilmekte, veri toplama ve yorumlama eş zamanlı yapılabilir. Belirli saptamalara ulaşıldığı zaman artık veri toplama bırakılabilmektedir. Oluşturulması gereken kategori sayısı araştırmayla ilgili bazı durumlara bağlı olarak değişmekte ancak kategori sayısının az olması daha yüksek seviyede soyutlanma sağlamak ve bulguların paylaşılmasını kolaylaştırmaktadır (Merriam, 2013: 107, 179). Bu bağlamda araştırmada temel kategorilerin oluşturulması amaçlanmıştır ve alt kategorilerin sayısının sınırlı tutulmasına özen gösterilmiştir. Elde edilen bulguların yorumlanması aşamasında yapılan açıklamaları destekleyen ve örnekleyen, katılımcılardan birebir alıntılar yapılmıştır. NVivo 12 programı ile ilk görüşmede 38 kod oluşturulmuş ve sekizinci görüşmeden sonra yeni kod oluşmadığı için örneklemin yeterli olduğu düşünülerek görüşmeler tamamlanmış ve toplam 78 kod oluşturulmuştur (Çizelge 3.3). Görüşmelerden elde edilen bazı veriler üzerinde araştırılan konuyla ilgisiz olduğu için kod oluşturma gereği görülmemiştir.

Çizelge 3.3. Görüşmelerden sonra oluşturulan kod sayısı ile toplam kod sayısı

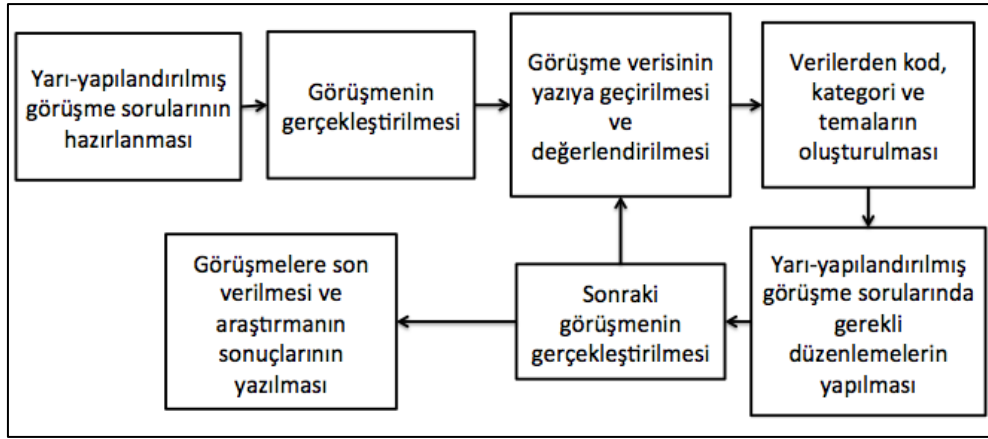
Görüşülen Kurum	Yeni Kodlar	Oluşturulan	Toplam Kod Sayısı
Kurum-1	38		38
Kurum-2	14		52
Kurum-3	5		57
Kurum-4	4		61
Kurum-5	9		70
Kurum-6	5		75
Kurum-7	1		76
Kurum-8	2		78
Kurum-9	0		78
Kurum-10	0		78

Görüşme yapılan kurumlardan toplanan verilerden elde edilen bulguların ifade edilmesi aşamasında, nitel verilerin sayısal analizi yöntemi de kullanılmaktadır. Bu yöntemdeki amaç, istatistiksel yöntemlere başvurarak genellemeler yapmak veya değişkenler arasında ilişki aramak değildir. Nitel verinin sayısallaştırılmasında dört temel amaç vardır. Bunlar; sayısallaştırmanın güvenilirliği arttırması, yanlılığı azaltması, verinin analizi sonucunda ortaya çıkan tema ve kategoriler arasında karşılaştırma yapılmasına olanak sağlaması ve yapılan küçük ölçekli araştırmanın daha sonra geniş bir örnekleme ulaşılarak tekrar sınanmasına olanak vermesidir. Temel olarak verinin sayısallaştırılması, basit yüzde hesapları ve sözcük sıklık hesapları olmak üzere iki yöntemle yapılmaktadır (Yıldırım ve Şimşek, 2013: 274, 275).

3.2.7. 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 baz alınarak mülakat sürecinin güvenilirliğine ilişkin bilgilere yer verilmiştir.

Nitel araştırmalarda kullanılan teknik ve ölçüm araçlarının, nicel araştırmalarda olduğu gibi standart olması mümkün değildir. Mayring izlenen yol ve tekniklerin açık bir şekilde belirtilmesi gerektiğine dikkat çekmiştir. Birinci ölçüt doğrultusunda, araştırmanın bütün süreçleri açık bir şekilde ortaya konması hedeflenmiştir. Yukarıda örneklem, veri toplama ve veri analizi süreci ayrıntılı bir şekilde aktarılmıştır. Araştırma esnasında gerçekleştirilen görüşmelere ilişkin ayrıntılı bir tasvir yapılarak takip edilen yöntemle biraz daha ı ışık tutulmaya çalışılmıştır (Şekil 3.3).



Şekil 3.3. Görüşmelerin gerçekleştirilme süreci

İkinci ölçüte göre, araştırmada elde edilen bulgular argümantatif (tartışmalı) olarak ifade edilmeli, yorumlamalar öngörülere uygun olmalıdır. Negatif örnekler veya olumsuzlayan açıklamalar ve yorumlamalar değerlendirilmelidir. Bu ölçütte, katılımcıların büyük veri beklentileri ve önerileri analiz edilmiş ve olumsuz örnekler ve açıklamalar göz önünde bulundurulmuş, kesin hükümlerden kaçınılmış ve yorumlamalarda açıklama noktalarına bağlantılar kurulmuştur.

Üçüncü ölçüt dikkate alınarak, analiz basamakları kurallı bir şekilde gerçekleştirilmiş, materyal anlamlı parçalara ayrılmış ve bu birimler arasında sistematik olarak ilerlemiştir. Ancak kurallara uygun ilerlemek, her durumda ön-belirlemelere sıkı sıkıya bir bağlılık anlamına gelmediği vurgulanmaktadır. Şekilde gösterildiği gibi veri analizi aşamasında belirli bir sistematik uygulanmıştır. İlk gerçekleştirilen mülakatın betimsel görüntüsü incelenmiş ve yeni mülakatlar için gerekli olan düzenlemelere karar verilmiştir. Yapılan mülakatlar gömülü teori yaklaşımına göre önceki mülakatlarla ve literatürle karşılaştırılarak değerlendirilmiştir. Mülakatların tamamlanmasından sonra veriler NVivo 12 programı ile kodlama işlemine tabi tutulmuş, ilgili kodlar bir araya getirilerek kategoriler oluşturulmuştur. Bu kategorilerden temalar oluşturulmuş yani çekirdek kategori yaratılmıştır.

Bir sonraki ölçüt olan olguya yakınlıkta, katılımcılarla yakınlık kurarak onların dünyasıyla bağ kurmak önemli bir kriterdir. Katılımcıların doğal paylaşımlarını ne kadar aktarabildiği önemli bir doğrulama ölçütüdür. Bu ölçüt çerçevesinde, katılımcılarla yapılan

görüşmelerde, katılımcıların kendilerini rahat hissettikleri yerde ve şekilde görüşmelerin yapılması sağlanmıştır. Yüz yüze yapılan görüşmeler katılımcıların kendi istekleri doğrultusunda kendi ofislerinde gerçekleştirilmiş. Görüşme esnasında katılımcıların izniyle ses kaydı alınmış ancak sistematik bir soru cevap takibinden uzak durulmuştur.

Beşinci ölçüt doğrultusunda, sonuçlar ve yorumların geçerliliği katılımcılarla tekrar bağlantı kurularak birlikte tartışılabilir olmalıdır. Aynı zamanda, katılımcıların öznel anlamlandırmaları, önyargıları ve ideolojileri dikkate alınmalıdır. Bu kapsamda katılımcılara, özellikle eklemek istedikleri bir durum veya rahatsız oldukları bir hususun olup olmadığı sorulmuştur. Projelerinde kamu verisi kullanan ve savunma sistemleri çözümleri sunan bazı kurumların görüşmecileri, kişisel veya gizli sayılabilecek veriler konusunda hassas olduklarını belirtmişler ve cevaplarını bu hassasiyetle yanıtladıklarını ifade etmişlerdir. Bir katılımcının isteği üzerine ses kaydının transkripsiyon işlemi yapıldıktan sonra ses döküm dosyasının kendisine gönderilmesini talep etmiş ve bu talebi yerine getirilmiştir.

Altıncı doğrulama ölçütü olan sacayağında araştırma kalitesinin, analiz çeşitlenmesi ve farklı veri kaynakları kullanımı ile arttırılması önerilmektedir. Bu ölçütte amaç, kullanılan araçlar arasında tam bir örtüşme aramak değil farklı açılardan sonuçların karşılaştırılarak güçlü ve zayıf noktaların gösterilmesidir. Bu araştırmada veri çeşitliliğini yakalamak için, kamu ve özel kurumlar dışında TÜBİTAK'ın Bulut Bilişim ve Büyük Veri Laboratuvarı (B3Lab) yöneticisi ile görüşülmüş ve bu görüşmede yarı-yapılandırılmış sorular haricinde ekstra sorulara yer verilmiştir. Aynı zamanda görüşme yapılan kurumların çalışan sayıları, kuruluş tarihleri, hizmetleri, analiz yaptıkları veri miktarı ve türü ve kullanılan yöntemler dikkate alındığında bu kriterler doğrultusunda yapılan farklı sorgulamalarla çeşitlilik sağlanmıştır.

3.2.8. Bulgular

Bu çalışmada, bulgular ifade edilirken basit yüzde hesapları, sözcük sıklık hesapları kullanılmış ve yapılan açıklamaları destekleyen katılımcılardan birebir alıntılar yapılmıştır. Çizelge 3.4'te görüşülen kurumla ilgili detaylı bilgiler yer almaktadır. Görüşülen kurumların hizmet ettiği sektör, veri analizi sırasında kullanılan yöntemler, veri miktarı ve

türü, kullanılan sunucu türü gösterilmektedir. Ayrıca kurumların toplam çalışan sayısı ve büyük veri projelerinde çalışan sayısının oranına da yer verilmiştir.

Çizelge 3.4. 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 Zekası	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 Zekası	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 Zekası, ETL	20/8 (%40)
Kurum-10	Kamu	Terabyte	Yapısal-Yapısal Olmayan	Sunucu ve Bulut Sunucu	Bulut Bilişim, Büyük Veri, Makine Öğrenmesi, Yapay Zeka	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 türleri hem yapısal hem de yapısal olmayan veri türüdür. Kurumların bir kaçının büyük veri ekipleri, normal sunucuların yanında bulut sunucular ile de çalışmaktadır. Kurumlar, veri madenciliği, veri ambarı, iş zekası, ETL,³⁸ makine öğrenmesi, bulut bilişim, metin madenciliği, görüntü işleme ve büyük veri teknikleri gibi veri analizi yöntemlerinden bir kaçını bir arada kullanmaktadırlar.

İlk olarak, görüşme yapılan kurumlardan toplanan verilere göre elde edilen bulgular, kurumların büyük veri analitiği mevcut durumu, üniversitelerin ilgili bölümlerinden ve öğrencilerden beklentileri ve öğrencilere ve üniversitelere öneriler 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 (Çizelge 3.5).

Çizelge 3.5. Görüşmecilerden elde edilen veriler doğrultusunda belirlenen kategoriler

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

³⁸ 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.

Büyük veri mevcut durum

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

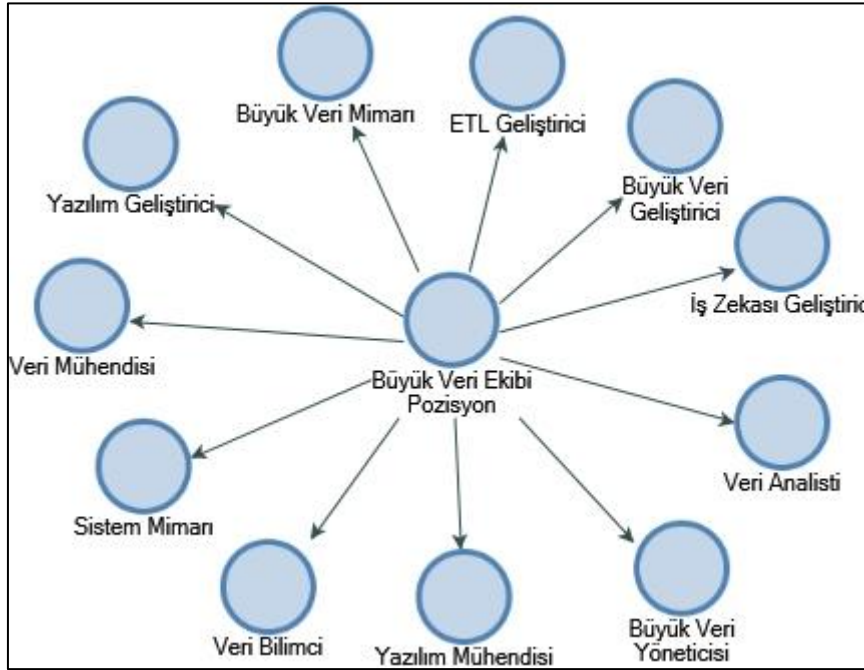
Büyük veri projelerinde çalışanların pozisyonları

Görüşülen kurumların verdikleri cevaplar doğrultusunda, çok çeşitli pozisyon isimleri kullanıldığı görülmekte 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 zekası 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 yazılım mühendisi veya yazılım geliştirici olarak belirtilmektedir. Ancak ilana ek olarak hangi büyük veri araçlarını kullanıyorlarsa o araçlarda tecrübeli olma, daha önce bu araçlarla çalışma gibi kriterler de belirtilmektedir. 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. Kurumların büyük veri alanında çalışan kişilerin hangi pozisyon isimlerini verdikleri Şekil 3.4’te gösterilmektedir.



Şekil 3.4. Kurumların büyük veri alanında çalışanlara verdikleri pozisyon isimleri

ABD İşgücü İstatistikleri Bürosu³⁹, 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. 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 ünvanlarının veri bilimi ve analitiği pozisyonlarının çoğunda tutarlı olmadığı belirtilmektedir (Miller ve Hughes, 2017). Miller ve Hughes (2017)’in belirttiği gibi bu çalışmada görüşülen kurumlarda büyük veri analitiği alanında iş ünvanlarının net bir şekilde ortaya konmadığı ortaya çıkmaktadır. Görüşülen kurumlarda büyük veri alanında çalışanların farklı teknik beceri ve yeteneklere sahip olmalarından pozisyon ismi olarak genel de yazılım mühendisi ve benzer isimler tercih edilmiştir.

³⁹ ABD’de işgücü piyasasının temel metriklerini izlemekle görevli federal kurum

Büyük veri projelerinde çalışan sayısı ve yeterliliği

Görüşülen kurumlarda, toplam çalışan sayısının, büyük veri projelerinde çalışan sayısına oranı Çizelge 3.4'te gösterilmiştir. Bu oranlar, kurumun hizmet ettiği sektör, toplam çalışan sayısı, gerçekleştirilen ürün ve projeleri ile alakalı olarak değişmektedir. Büyük veri projelerinde çalışan sayısını yeterli gören kurumlar %50 oranındadır. Bu sayının yeterli görülmesi, kurumlardan bazılarının ifadesiyle, proje tabanlı çalıştıklarından dolayı bu projelere göre yeterli oldukları şeklindedir. Kurumlardan bir tanesi, ürünlerin yıllar içerisinde geliştiğini ve şuan 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 görmemektedirler. B3Lab'ın bu oranı yeterli bulmaması, özünde bulut bilişim ve büyük veri teknolojileri geliştirmek üzere 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)

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. Çizelge 2.6'da yer alan büyük veri analitiği çözümlerinde kullanılan öncelikli teknolojiler sıralamasına bakıldığında görüşülen kurumların kullandıkları araçların hemen hemen hepsi bu listede yer almaktadır.

Kurumların neredeyse hepsi, büyük veri araçlarını kullanırken en önemli avantaj olarak, performans ve daha hızlı analiz olduğunu belirtmekte ve görüşmecilerden biri bu avantajı şu şekilde ifade etmektedirler;

“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. 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, CAP 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 kullanırken 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 sorunlar 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 ifadeler yer verilmiştir;

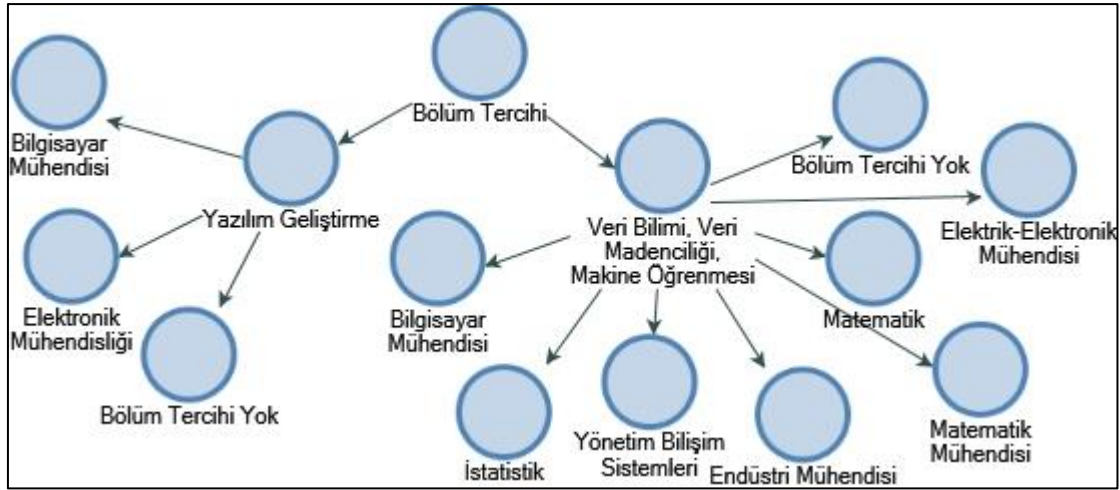
“Personel noktasında, yetişmiş eleman noktasında bir sıkıntı var ama biz hazırda eleman bulamasak da yeni mezun arkadaşları alıp burada yetiştirip, bu konuda belli bir yetkinlik sağlayarak o şekilde açığı kapattığımızı düşünüyoruz.” (Kişi-11)

Beklentiler

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

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 (Şekil 3.5). Bu kategorilerden ilki yazılım geliştirme olarak belirtilmekte, katılımcıların çoğu bu alan için bilgisayar mühendisi tercih ettiklerini belirtmektedir.



Şekil 3.5. Kurumların büyük veri alanı için tercih ettikleri bölümler

Bir katılımcı, ekiplerinde elektronik mühendisi de yer aldığını ancak bu kişilerin de üniversitede eğitim alırken bilgisayar alanı tercih etmiş ve bilgisayar mühendisliği dersleri almış kişiler olduğunu belirtmiştir. Bir başka katılımcı yazılım geliştiriciler için bilgisayar mühendisliğinin kesinlikle birinci tercihleri olduğunu ifade ederken, özellikle bilgisayar mühendislerinin sistemli ve düzenli çalıştığını belirterek şu ifadelerle yer vermiştir;

“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, 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 ifadelerle 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 kategoriyi, 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 (Çizelge 3.6).

Çizelge 3.6. 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;

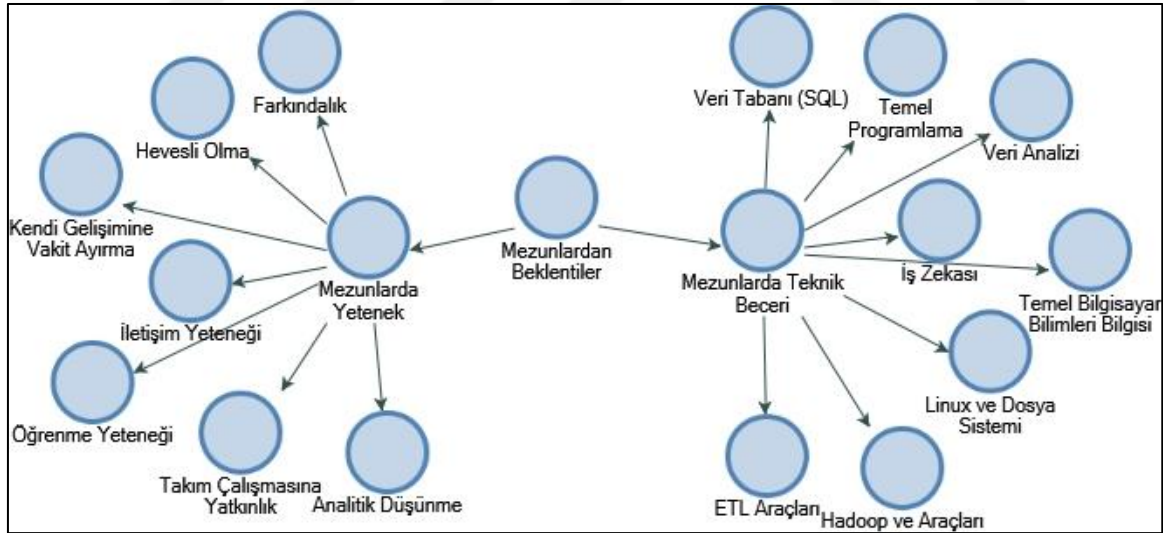
“Veri bilimi yeni bir şey olduğu için, siz istatistikten mezun birini alayım deseniz de bu alana ilgi duyuyor olması lazım, bunda çalışması, kendini geliştirecek bir

takım 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.

Mezunlardan beklentiler

Kurumların, büyük veri alanında çalışmak üzere üniversiteden yeni mezun kişilerde aradıkları özellikler, yetenek ve teknik beceri olarak iki kategoride incelenmektedir. Mezunlardan beklentiler kategorisinin alt kategorileri Şekil 3.6’da gösterilmektedir. 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.



Şekil 3.6. Kurumların mezunlardan beklentileri

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)

Analitik düşünme, katılımcılar tarafından, problem çözme yeteneği, sonuç odaklı olma ve stratejik düşünme gibi ifadelerle belirtilmektedir. Görüşmecilerden biri, beklentilerini şu şekilde ifadeler 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)

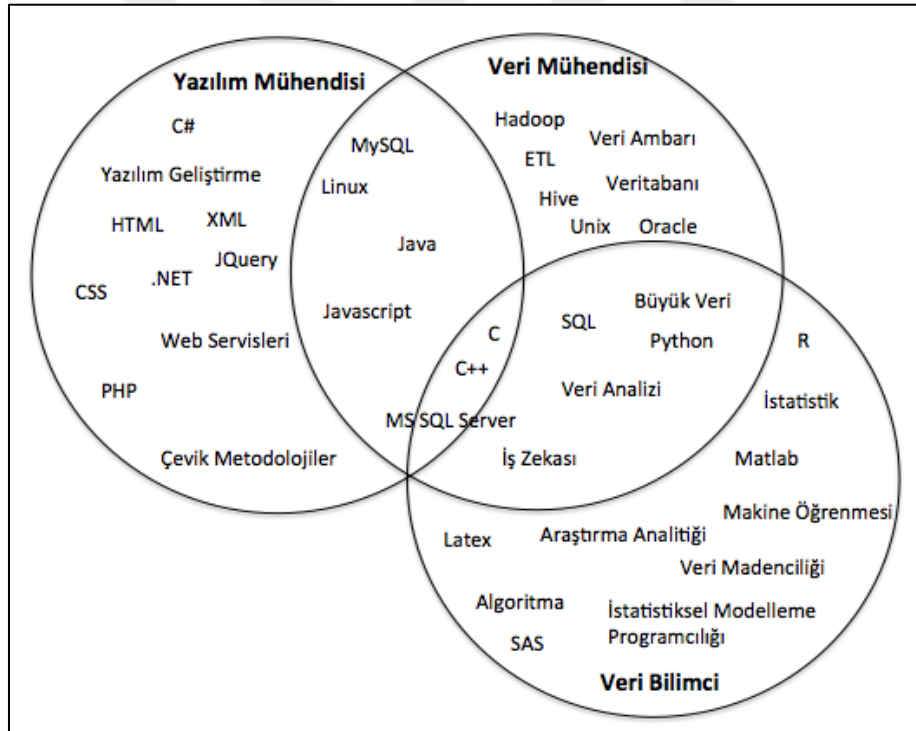
Bunların yanında, takım çalışmasına yatkınlık, öğrenme yeteneğinin olması ve hevesli olma, kendi gelişimine vakit ayırma ve iletişim becerilerinin olması gibi kriterler bazı kurumlar tarafından ifade edilmiştir. Bulut Bilişim ve Büyük Veri Çalıştayı Raporunda, veri işleme ve analizinde aşılması gereken 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 sıralarda 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 dökümantasyonu, 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 (National Academies, 2014). Lebanon'ın belirttiği teknik yetenekler, görüşmecilerin mezunlardan beklediği teknik becerilerle aynı doğrultudadır. Şekil 3.7'de yazılım mühendisliği, veri mühendisliği ve veri bilimcilerde olması gereken teknik yetenekler gösterilmiştir. Görüşülen kurumlardan alınan cevaplar doğrultusunda, bu üç pozisyonun ayırımı tam olarak yapılamadığı için, teknik yetenekler noktasında da net ayrımlar yapılamamıştır. Ancak genel olarak kurumların hepsi Şekil 3.7'de gösterilen teknik yetenekleri görüşmelerde belirtmişlerdir.



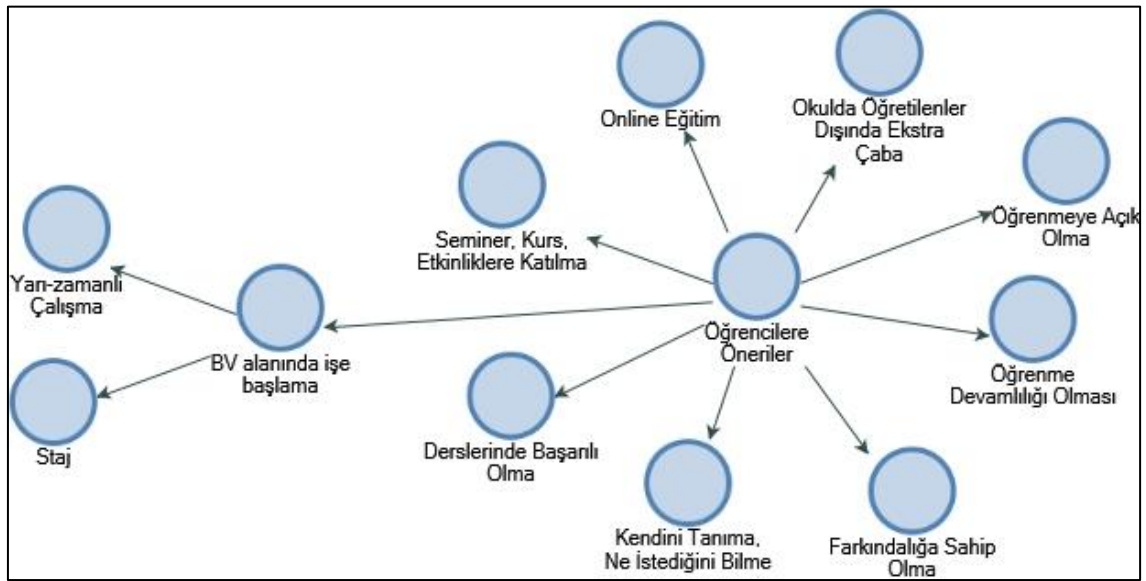
Şekil 3.7. Yazılım mühendisliği, veri mühendisliği ve veri bilimcinin teknik yetenekleri (Swanstrom, 2016)

Öneriler

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

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. Şekil 3.8’de görüşülen kurumların öğrencilere önerilerinin kategorileri ve alt kategorileri yer almaktadır.



Şekil 3.8. Kurumların öğrencilere önerileri

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 firma, teknolojilerin çok hızlı değiştiği ve alternatiflerinin çıktığı, bu hıza yetişebilmek için sürekli yeni çıkan teknolojilerden haberdar olmaları gerektiğini ifade etmektedirler. Bu konuda bir görüşmecinin şu ifadeleri dikkat çekicidir;

“Kendinizi teknolojide güncel tutmaz iseniz, bildikleriniz yetmiyor, sürekli üstüne üstüne koymanız gerekiyor. Onlarca teknoloji var, bir kaç 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 konulmaktadır. Büyük veri çözümleri sunan firmaları ve bunların ne gibi sorunlarla uğraştıklarını takip ederek, konuyla ilgili çalışmalar yaparak farkındalığı arttırmaları gerektiği vurgulanmaktadır.

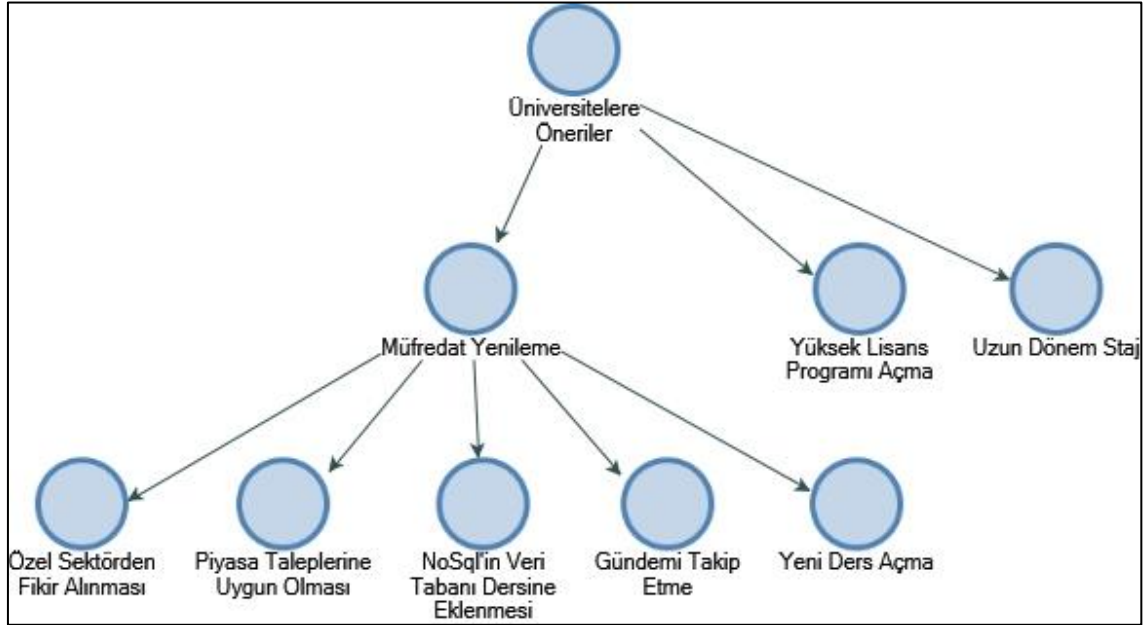
İkinci en çok belirtilen öneri, %60 oranla okulda öğretilenler dışında ekstra çaba gösterilmesidir. Bu konuda katılımcılarımız, öğrencilerin sadece okulun verdikleri ile yetinmemeleri gerektiğini, boş zamanlarında ekstra bir merak göstererek çalışmalar yapmalarını, ilgi alanlarına göre çevrimiçi kaynaklardan kendi kendilerini geliştirmelerini ifade etmektedirler. Görüşmecilerin, öğrencilerin okul dışında ekstra çaba göstermesi konusunda çeşitli ifadelerle 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)

Üniversitelere öneriler

Büyük veri alanına personel yetiştiren özellikle Bilgisayar Mühendisliği, İstatistik, Matematik ve Yönetim Bilişim Sistemleri bölümü gibi bölümleri olan üniversitelere sektörün önerileri, müfredat yenileme, uzun dönem staj programları oluşturma ve yüksek

lisans programları açmak yönündedir. Şekil 3.9'da görüşme yapılan kurumların üniversitelere olan önerilerinin alt kategorileri görülmektedir.



Şekil 3.9. Kurumların üniversitelere önerileri

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 bazıları şu ifadelere yer vermişlerdir;

“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' (2015) nın ifade ettikleri gibi, görüşmeciler de büyük veri analitiği ile alakalı olarak ilgili bölümlerde yeni derslerin açılması ile öğrencilerin bu alanda daha nitelikli eğitim almaları önemsemektedir.

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, Map-Reduce 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 öğretmeye, 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 bu yüzden becerilerin sürekli olarak güçlendirilmesi gerektiğini, teknoloji değil, yöntem ve ilke öğretmenin ve işbirliği yapmanın önemine dikkat çekmektedir (National Academies, 2014). Fox'un teknolojiden ziyade yöntemin öğretilmesi gerektiği konusunda görüşmeciler de aynı tavsiyelere yer vermişlerdir.

3.3. Çalışma 2 - Büyük Veri Eğitiminde Öğrencilerin Deneyimleri ve Karşılaştıkları Zorluklar

3.3.1. Amaç

Bu çalışmanın amacı, öğrencilerin büyük veri öğrenimi sırasında karşılaştıkları zorlukların ve deneyimlerin belirlenmesidir. Ayrıca, öğrencilerin büyük veri alanı ile ilgili okudukları üniversite ve bölümlerinden beklentilerinin belirlenmesi hedeflenmiştir. Bu çalışmada aşağıdaki araştırma sorularına cevap aranmaktadır.

- Lisans öğrencilerinin büyük veri öğrenimi sırasında karşılaştıkları zorluklar nelerdir?
- Lisans öğrencilerinin büyük veri alanı ile ilgili üniversite ve bölümlerinden beklentileri nelerdir?

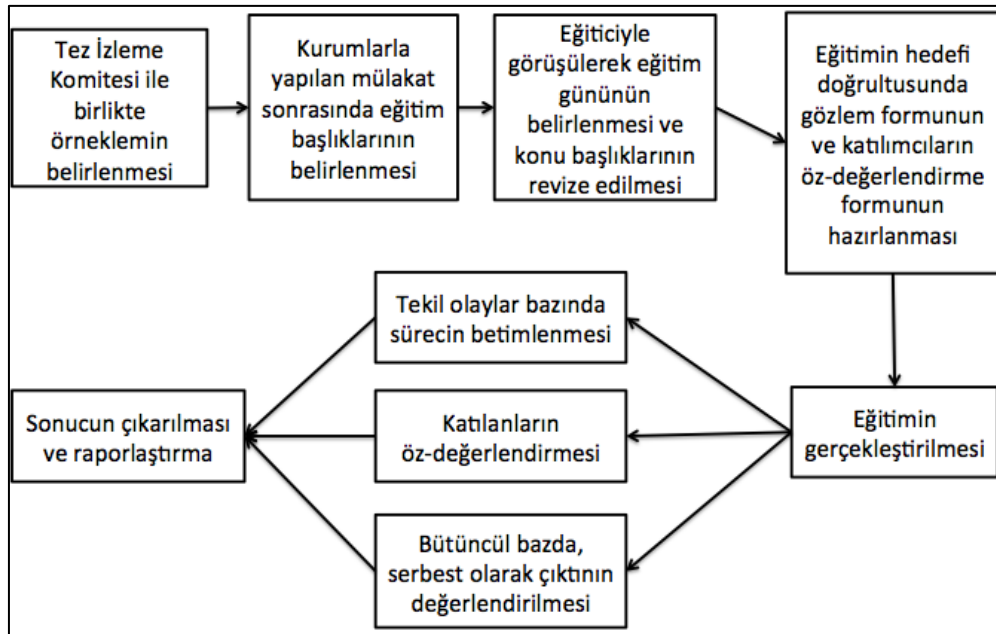
3.3.2. Araştırma deseni

Bu çalışmada nitel araştırma yöntemi kullanılmıştır. Yıldırım (1999), özellikle eğitim araştırmalarında problemi derinlemesine araştırmak isteyen araştırmacılar için nicel yöntemlerin yetersiz kaldığını, nitel araştırma yöntemlerinin eğitim araştırmalarında önemli bir yere sahip olduğunu belirtmektedir. Nitel araştırma yöntemlerinin eğitim çalışmalarında neden önemli olduğunu gösteren üç tür araştırma problemi ortaya konulmaktadır. Bunlardan ilki, kişilerin algılarını ortaya koymak ve neden belli bir şekilde davrandığını anlamak için nitel araştırmalara ihtiyaç olmasıdır. İkincisi, araştırmaya temel teşkil edecek bir teorik yapının olmadığı durumlarda nitel yöntemlerin kullanılması uygun görülmektedir. Aynı zamanda nitel desenlerin oluşturulmasında belirli bir esneklik sağladığı için araştırmacı, önceden hiç araştırılmamış problemlerin araştırılmasında katkı sağlayabilmektedir. Son olarak, nicel yöntemlere karşı olumsuz bir tutum içinde olan kişilerin araştırmaya dâhil edilmesi hususunda, görüşme veya gözlem gibi nitel yöntemler daha uygun olabilmektedir (Yıldırım, 1999).

Nitel izleme çalışmasında, araştırmacı herhangi bir müdahalede bulunmadan uygulamada değişikliklerin nasıl bir etkisinin olduğunu sınınamaya çalışmaktadır. Eğitim bilimlerinde eğitim programları, nitel izleme çalışmasının konusuna girmektedir (Wottawa ve Thierau, 1990). Bu çalışma, belirtilen desenler içerisinde en çok nitel izleme çalışmasına uygun bulunmuştur. Nitel yönelimli izleme çalışmasında aşağıdaki hususlara dikkat çekilmiştir (Mayring, 2011):

- Çıktı değerlendirmesinin yanında derinleştirilmiş, tekil-olayları içeren bir betimleme de eklenmelidir.
- Dikkate değerse gözlemlenen süreçlerden yeni değerlendirme ölçütleri belirlenebilmektedir.
- Katımlı gözlem ve alan araştırmasına daha yakındır.
- Katılımcıların öz-değerlendirmeleri, izleme çalışmasına dâhil edilmelidir.
- Pratikte olanların etkilerinin ifade edildiği bütüncül bir sonuç çıkarımı yapılmalıdır.

Bu çalışmada gerçekleştirilen bir günlük eğitimin nitel izleme çalışması Şekil 3.10'da gösterilmektedir:



Şekil 3.10. Eğitimin nitel izleme akış modeli

3.3.3. Örneklem

Öğrencilerle yapılan eğitimin örneklem seçiminde amaçlı örnekleme seçimi gerçekleştirilmiştir. Bu çalışma için Patton tarafından geliştirilmiş olan örneklem sınıflandırmalarından kritik durum örnekleme tercih edilmiştir. Kritik durum örneklemede, kritik durumun varlığının göstergesi “bu grup belirli problemlerle karşılaşılıyorsa, diğer gruplar da bu problemle karşı karşıya kalır” veya “bu grup belirli problemlerle karşılaşmıyorsa, diğer gruplar da bu problemle karşılaşmaz” şeklinde ifade edilmektedir (Patton, 1987). Bu çalışmanın örneklem seçimindeki kriterler, bilgisayar mühendisliğinde lisans düzeyinde eğitim alıyor olması ve büyük veri konusuna ilgili olması olarak belirlenmiştir. Eğitim boyunca daha iyi gözlem yapabilmek için kısıtlı sayıda öğrenci eğitime dâhil edilmiştir. Bu doğrultuda, Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği bölümünde okuyan ve büyük veri alanına ilgili olan 3. ve 4. sınıf öğrencilerinden 11 kişinin katılımıyla eğitim gerçekleştirilmiştir. Kritik örneklem seçimi yaklaşımı doğrultusunda, katılımcıların hepsinin büyük veriye ilgili olmaları, yarıdan fazlasının büyük veri uygulama deneyimi olması ve üçüncü ve dördüncü sınıf öğrencileri olması dolayısıyla eğitim boyunca büyük veri teknolojilerini kullanırken belirli sorunlarla karşılaşılıyorsa, büyük veri deneyimi ve yeterliliği olmayan bilgisayar mühendisliği ve ilgili diğer bölüm öğrencileri bu teknolojileri kullanırken bu sorunlarla karşılaşabilecekleri öngörülmektedir.

3.3.4. Veri toplama

Veriler, bir günlük büyük veri eğitimine katılan lisans öğrencilerinden, nitel yöntemlerden katılmalı gözlem tekniği izlenerek ve katılımcıların öz-değerlendirmeleri ile toplanmıştır. Katılmalı gözlem olarak da adlandırılmaktadır (Mayring, 2011). Katılmalı gözlem, bir gözlem cetveliyle uygun gözlemler işaretlenerek gerçekleştirilen tamamen standartlaştırılmış bir gözlem değildir. Katılmalı gözlem, açık uçlu olmalıdır ve bir gözlem klavuzu oluşturmak yararlıdır. Bunun için yarı-standartlaştırılmış klavuz uygun görülmektedir. Bu doğrultuda, eğitim boyunca katılımcıların deneyimleri ve karşılaştıkları zorluklar gözlem notu olarak kaydedilmiş, aynı zamanda açık uçlu sorular ile katılımcıların öz-değerlendirmeleri kayda alınmıştır. Diğer veri toplama yöntemleriyle karşılaştırıldığında, gözlemin dikkat çekici bazı artıları vardır. Bunlar, sözel olmayan davranış, doğal çevre ve uzun süreli analiz olarak belirtilmektedir. Çalışılan konuyla

alakalı olarak katılımcıların bütün yönleriyle ve derinlemesine incelenmesine olanak sağlamaktadır (Yıldırım ve Şimşek, 2013: 203). Bu çalışmada gözlem, katılımcıların kendi üniversitelerinde ve dersliklerinde gerçekleştirilmiş olup, katılımcıların eğitim süresince özellikle uygulamalı oturumda yaşadıkları sorunların ve eğitmeniye yönelttikleri soruların derinlemesine irdelenmesi sağlanmıştır. Nitel araştırmada yapılan gözlem boyunca her şeyin gözlenmesi mümkün değildir ancak araştırmacı araştırmının hangi kapsamda gözleneceğini içeren, önceden hazırlanmış bir gözlem formu hazırlaması faydalı olmaktadır (Yıldırım ve Şimşek, 2013: 205). Bu çalışmada, katılımcıların bu eğitimde teorik, kurulum ve uygulama aşamalarında karşılaştıkları zorlukların ve eğitmenin bu alanla alakalı sorduğu sorulara katılımcıların verdikleri cevapların not alınması şeklinde gerçekleştirilmiştir.

8 Mart 2019 tarihinde, büyük veri yöneticisi olarak görev yapan, çevrimiçi ve kurumlara büyük veri eğitimleri veren eğitmenimiz tarafından iki oturumdan oluşan bir günlük eğitim 11 bilgisayar mühendisliği lisans öğrencisinin katılımıyla gerçekleştirilmiştir. Birinci oturumda, büyük veri teorisi, Hadoop genel tanıtımı ve bir veri setinin tablo formatında Hadoop'a yüklenmesi gibi konular üzerinde durulmuş olup, ikinci oturumda öğrencilerin kendi bilgisayarlarında Hadoop bileşenlerinden olan Spark uygulamaları gerçekleştirilerek soruları yanıtlanmıştır. Eğitimin temel başlıkları aşağıda belirtilmektedir;

- Büyük Veriye Giriş (Teorik)
- Hadoop'a Giriş (Teorik)
- Büyük Veri Ortamının Kurulması
- Örnek Bir Veri Setinin Hadoop'a Yüklenmesi
- Hadoop Dağıtık Dosya Sistemi (HDFS) Teori ve Uygulama
- Spark Uygulamaları

3.3.5. Araştırmanın kısıtlılıkları

Araştırma Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği bölümünde üçüncü ve dördüncü sınıfta okuyan 11 öğrenci ile kısıtlanmaktadır. Ayrıca katılımcıların büyük veri alanına ilgili olmaları veya daha önceden büyük veri alanı ile ilgili ders almış olmaları veya büyük veri araçlarından herhangi biriyle deneyim sahibi olmaları da araştırmanın kısıtlılıklarındandır.

3.3.6. Verilerin analizi

Nitel arařtırmalarda veri analizi, verinin anlamını ortaya ıkarma yani insanların ne sylediğini, arařtırmacının ne grdüğünü ve okuduğunu birleřtirme, indirmeye ve yorumlama sürecidir. Aynı zamanda nitel veri analizi, arařtırma sorusunu cevaplamak için kullanılan bir süreçtir (Merriam, 2013: 167, 168). Nitel veri analizinde genellikle üç temel yöntem önerilmektedir. İlk yöntemde, elde edilen verinin özgün haline mümkün olduđu kadar bağılı kalınması beklenmektedir. Gerekli görüldüğü takdirde katılımcıların ifadelerinden doğrudan alıntı yapılarak, betimsel bir şekilde veriler sunulmalıdır. Gözlem notlarının özgün bir şekilde sunulması örnek olarak gösterilebilir. İkinci yöntemde, betimsel bir yaklaşıma ek olarak sistematik bir şekilde bazı açıklayıcı sonuçlara ulaşmak amacıyla belirlenen kod, kategori ve temalar ile verilerin sunulması söz konusudur. Son yöntemde ise, birinci ve ikinci yaklaşım temel alınarak veri analizi sürecine arařtırmacının yorumlarının dahil edilmesi yöntemidir. Bu üç yöntemin, birbirinden tamamen bağımsız olamayacağı kabul edilmektedir (Creswell, 2002; Miles ve Huberman, 1994; Wolcott, 1994). Bu alıřmada kullanılan gözlem notları ve katılımcılara yneltlen açık uçlu sorular ile toplanan verilerin analizi betimsel analiz yaklaşımla gerekleřtirilmiřtir. Betimsel analizde, toplanan verilerin daha nceden belirlenen temalara göre zetlenmesi ve yorumlanması esas alınmıřtır. Arařtırma soruları dahilinde oluřturulan temalara göre ya da gözlem ve görüřme ařamalarında belirtilen sorular dikkate alınarak veriler ifade edilebilir (Yıldırım ve řimřek, 2013). Aynı zamanda betimsel analiz, nitel veriler üzerinde detaylı ayrıřtırma gerektirmeyen verilerin iřlenmesinde, katılımcıların demografik özellikleri ile farklı niteliklerinin tasvir edilmesinde kullanılmaktadır (Miles ve Huberman, 1994). Nitel arařtırmalarda betimsel analiz dört ařamadan oluřmaktadır. (1) Betimsel analiz için bir çereve oluřturma (2) Tematik çereveye göre verilerin iřlenmesi (3) Bulguların tanımlanması (4) Bulguların yorumlanması (Miles ve Huberman, 1994; Creswell, 2002; Yıldırım ve řimřek, 2013; Marshall ve Rossman, 2014).

Betimsel Analiz için Bir Çereve Oluřturma: Arařtırmanın kavramsal çerevesinden, arařtırma sorularından ya da gözlem ve görüřmelerde yer alan sorulardan yola ıkılarak veri analizi için bir çereve oluřturulmalıdır. Bu alıřmanın kavramsal çerevesi ierisinde, bilgisayar mühendisliğinde okuyan lisans ğrencileri ile yapılan bir günlük büyük veri

eğitimi sırasında öğrencilerin karşılaştıkları zorlukları ve deneyimleri ve öğrencilerin büyük veri alanı ile ilgili okudukları üniversite ve bölümlerinden beklentilerini belirlemek amaçlandığı için temalar bu doğrultuda oluşturulmuştur. Ayrıca, kavramsal çerçevede belirlenen araştırma soruları doğrultusunda, açık uçlu sorular katılımcılara yöneltilmiş ve gözlem notları bu çerçevede değerlendirilmiştir.

Tematik çerçeveye göre verilerin işlenmesi: Bu aşama, oluşturulan çerçeveye göre toplanan verilerin okunması ve düzenlenmesi aşamasıdır. Elde edilen verilerin anlamlı bir şekilde bir araya getirilmesi gerekmektedir. Verilerden bazıları oluşturulan temalara göre dışarıda kalabilir ve sonuç aşamasında kullanılacak alıntılar bu aşamada belirlenebilir. Bu doğrultuda, çalışmada tutulan gözlem notları ve katılımcıların öz-değerlendirme formları, temalar doğrultusunda gruplanarak düzenlenmiştir. Çalışmada alınan bazı gözlem notları detayları ve değerlendirme sorularından bazıları temalara dâhil edilmemiştir. Bu veriler katılımcıların eğitimden önceki büyük veri deneyimleri ve bu alanda eğitim geçmişlerini öğrenmek amaçlı katılımcıların derinlemesine büyük veri tecrübe düzeylerini anlamaya yönelik sorulmuştur.

Bulguların tanımlanması: Bu aşamada düzenlenen verilerin açık ve anlaşılır bir dille tanımlanmasına dikkat edilmelidir. Gerekli görülen yerlerde doğrudan alıntı yapılarak veriler desteklenebilir ancak gereksiz tekrarlardan kaçınılmasına dikkat edilmelidir. Oluşturulan temalar ışığında düzenlenen veriler, anlaşılır ve okunabilir bir dille ifade edilmiştir.

Bulguların yorumlanması: Tanımlanan bulguların açıklanması, ilişkilendirilmesi ve anlamlandırılması bu aşamada yapılmaktadır. Bulgular arasında neden sonuç ilişkilerinin ifade edilmesi ve farklı olgular arasında karşılaştırma yapılması, yorumlamanın daha nitelikli olmasını desteklemektedir. Bu çalışmada, katılımcılara eğitimden önce büyük veri ile alakalı bir araç kullanıp kullanmadıkları sorularak buna bağlı olarak yaşanan sorunlar üzerinde durularak, neden sonuç ilişkisi içerisinde bulgular değerlendirilmiştir.

3.3.7. Güvenilirlik ve geçerlilik

Bu çalışmada, Mayring'in doğrulama ölçütleri baz alınarak gözlem sürecinin güvenilirliğine ilişkin bilgilere yer verilmiştir. Mayring'in birinci doğrulama ölçütü, yol ve

tekniklerin belgelenmesidir. Bu ölçütte, kullanılan yöntemlerin doğru ve açık bir şekilde aktarılması önemlidir. Buna bağlı olarak bütün süreçlerin detaylı ve kapsayıcı şekilde ortaya konması hedeflenmiştir. Çalışmanın bu aşamasında nitel araştırma desenlerinden (planlarından), nitel izleme çalışması gerçekleştirilmiştir. Nitel yönelimli izleme çalışması, pratik uygulamalara bilimsel olarak yaklaşmakta ve onların etkilerini değerlendirmektedir. Birinci doğrulama ölçütü doğrultusunda gerçekleştirilen gözlem sürecinin nitel izleme akış modeli gösterilmiştir (Şekil 3.10). İkinci ölçüt doğrultusunda, katılımcıların alternatif açıklamaları göz önünde bulundurulmuş ve olumsuzlayan açıklamalar dikkate alınmıştır. Kesin yargılardan kaçınılmış ve sonuçlar araştırmanın kapsamı çerçevesinde tartışılmıştır. Üçüncü ölçüt, araştırma süresince kullanılan materyalin sistemli bir şekilde işlenebilmesi için araştırmanın belirli kurallar dâhilinde gerçekleştirilmesidir. Bu aşamada, eğitim sürecinin basamakları önceden belirlenmiş ve sistematik bir ilerleme sağlanmıştır. Örneklem seçiminden, veri toplama, veri analizi ve bulguların yorumlanmasına kadar izlenen yöntemler önceki bölümlerde aktarılmıştır. Bir sonraki ölçüt olguya yakınlık olarak belirtilmiştir. Yapılan eğitim, katılımcıların kendi üniversiteleri ve dersliklerinde, formallikten uzak bir şekilde gerçekleştirilmiştir. Bütün oturumlarda ve aralarda araştırmacı ve eğitimciyle konu ile ilgili sorular ve sorunlar üzerine konuşularak yakınlık kurulmuştur. Beşinci doğrulama ölçütü iletişimsel geçerlemedir. Katılımcılarla sonuçlar paylaşılabilmekte ve onlardan yeniden onaylama alınabilmektedir. Yorumlamalarda kişilerin mitleri, önyargıları, ideolojileri de dikkate alınmalıdır. Katılımcılara yöneltilen değerlendirme formunda gerekli görüldüğü takdirde araştırmanın devamına katkı yapmak isteyenler ile görüşme yapılması hususunda sorular sorulmuştur. Bu doğrultuda katılımcılardan bazıları, eğitim sonrasında sorulan sorulara kendi istekleriyle katılım sağlamamışlardır. Son doğrulama ölçütü sacayağında, maksimum çeşitlemenin sağlanması, değişik veri kaynaklarının kullanımı, değişik yorumlar, yaklaşım ve metotların kullanımının araştırmanın kalitesini arttıracakları belirtilmektedir. Bu hususta araştırmada eğitim boyunca alınan gözlem notları yanı sıra, eğitim sonunda katılımcıların öz-değerlendirmeleri karşılaştırılarak, veriler farklı perspektiflerden incelenmiştir.

3.3.8. Bulgular

Bir günlük büyük veri eğitimi, Gazi Üniversitesi Bilgisayar Mühendisliği lisans öğrencilerinden 11 kişinin katılımıyla gerçekleşmiştir. Eğitime katılan öğrencilerin tamamı büyük veri ekosistemi olan Apache Hadoop ve Spark gibi araçları daha önce duyduğunu belirtmiştir.

Çizelge 3.7’de belirtildiği gibi katılımcıların sekizi erkek, üçü kadındır ve yaşları 21 ile 27 arasındadır. Katılımcıların sekizi üçüncü sınıf öğrencisi, üçü dördüncü sınıf öğrencisidir. Katılımcılardan yedisi daha önce büyük veri araçlarından bir veya bir kaçının kurulumunu gerçekleştirerek, uygulamalarda bulunmuştur. Katılımcıların ikisi büyük veri uygulama deneyimi yaşamamış, ikisi de uygulama deneyimi olup olmadığını belirtmemiştir.

Çizelge 3.7. Katılımcıların demografik özellikleri

Ad	Cinsiyet	Yaş	Sınıf	Not Ortalaması	Büyük Veri Uygulama Deneyimi
Kişi-1	Erkek	23	4	2.14	Yok
Kişi-2	Erkek	23	3	2.64	Var
Kişi-3	Erkek	22	3	3.47	Belirtilmemiş
Kişi-4	Kadın	22	3	3.51	Var
Kişi-5	Kadın	21	3	3.49	Var
Kişi-6	Erkek	26	4	3.36	Var
Kişi-7	Erkek	22	3	2.80	Belirtilmemiş
Kişi-8	Erkek	21	3	3.47	Var
Kişi-9	Erkek	27	4	2.10	Var
Kişi-10	Kadın	21	3	3.02	Var
Kişi-11	Erkek	21	3	2.80	Yok

Eğitim, teorik konular ve pratik uygulamaları içeren iki oturumdan oluşmaktadır. Birinci oturumda, büyük veri teorisi, Hadoop genel tanıtımı ve bir veri setinin tablo formatında Hadoop’a yüklenmesi gibi konular üzerinden teorik bilgiler verilmiştir. İkinci oturuma 5 kişi katılmış olup, öğrencilerin kendi bilgisayarlarında Hadoop bileşenlerinden olan Spark kurulumu ve uygulamaları gerçekleştirilerek soruları yanıtlanmıştır. Eğitmen, internet üzerinden örnek bir veri seti indirerek, bütün öğrencilerin eş zamanlı olarak kurulum yapmasını, kodları yazmasını ve örnekleri gerçekleştirmesini sağlamıştır. Eğitim boyunca alınan gözlem notlarına ve katılımcıların kendi öz bildirimlerine bağlı olarak, katılımcıların büyük veri deneyimleri, büyük veri araçları kullanırken karşılaşılan sorunlar ve üniversite ve bölümden beklentilerine yer verilmiştir. Araştırmanın kavramsal çerçevesinde belirlenen temalar Çizelge 3.8’de belirtilmiştir.

Çizelge 3.8. Araştırmanın kavramsal çerçevesinde belirlenen temalar

Büyük Veri Deneyimi	Yeterlilik Düzeyleri
	Lisans Eğitiminde Ders
	Çevrimiçi veya Yüz yüze Eğitim
Karşılaşılan Sorunlar	Büyük Veri Araçlarının Kurulumu Esnasında
	Büyük Veri Araçlarının Kullanımı Esnasında
	Büyük Veri Araçlarının Öğrenirken
Beklentiler	Eğitim Aldığı Üniversiteden
	Eğitim Aldığı Bölümden

Bu doğrultuda, ilk tema olan büyük veri deneyimi, öğrencilerin büyük veri alanıyla ilgili yeterlilik düzeyleri, lisans eğitimi boyunca büyük veri ile alakalı ders alıp almadıkları, lisans eğitimi dışında çevrimiçi veya yüz yüze eğitime katılıp katılmadıkları kategorilerinden oluşmaktadır. İkinci temada, öğrencilerin büyük veri araçlarını kurarken, kullanırken ve öğrenirken karşılaştıkları sorunlar yer almaktadır. Son temada ise katılımcıların eğitim aldıkları üniversite ve bölümden beklentilerine yer verilmiştir.

Büyük veri deneyimi

Büyük veri deneyimi teması içerisinde büyük veri yeterlilik düzeyi kategorisi yer almaktadır. Yeterlilik düzeyini açıklayan ifadeler şu şekilde oluşturulmuştur; (1) büyük veri ya da büyük veri analitiği kavramlarından ve büyük veri araçlarından haberdar olmaları (2) büyük veri tanımını oluşturan 5V karakteristiklerini bilmeleri (3) büyük veri araçlarını kullanarak çeşitli uygulamalar yapmaları (4) kullandıkları araçlar.

Eğitime katılan öğrencilerin hepsi büyük veri ve büyük veri analitiği kavramlarını daha önce duyduklarını belirtmektedir. Katılımcıların çoğu, veri türünün çeşitlenmesi ve veri miktarının artmasıyla büyük veri çözümlerinin ortaya çıktığı konusunda hemfikirdir. Büyük veri teknolojilerini kullanmanın kurumlara çok çeşitli avantajlar sağladığı hususunda katılımcıların %91'inin aynı fikirde olduğu görülmektedir. Ancak büyük verinin tanımını oluşturan 5V karakteristiklerinin neler olduğunu ve büyük veri ile tam olarak neler yapılabileceğini bilenlerin oranı %73'tür. Katılımcıların hepsi Apache Hadoop ekosistemini daha önce duyduklarını ve çoğunun Apache araçlarından bazılarının kurulumlarını yaparak deneme fırsatı bulduklarını belirtmektedir. Apache Hadoop'dan

sonra, katılımcılar tarafından en çok bilinen araçlar sırayla Spark, HDFS, Hive, Yarn, Cassandra ve HBase olduğu ortaya çıkmaktadır. Ayrıca, Cloudera, Microsoft Hadoop, Oracle Big Data, Hortonworks gibi büyük veri yönetim platformlarından haberdar oldukları da katılımcıların öz değerlendirmeleri ve gözlem notlarından ortaya çıkmaktadır. Katılımcıların yarısının, NoSql veritabanları hakkında, bir kaçının Map-Reduce fonksiyonları hakkında, yalnızca bir kişinin yatay ve dikey ölçeklendirme hakkında fikir sahibi olduğu gözlem notlarından ortaya çıkmaktadır. Bunun yanı sıra, eğitim sırasında eğitimcinin sorduğu yapısal, yarı-yapısal ve yapısal olmayan veri türleri hakkında kimsenin bilgi sahibi olmadığı ortaya çıkmaktadır. Eğitime katılanların iki tanesi, büyük veri araçlarından herhangi birini kurup deneme ihtiyacının olmadığını belirtmiş, geri kalan katılımcıların hepsi başta Spark olmak üzere Kafka, Nifi, Sqoop, Hive, Flink gibi araçları kurarak deneme fırsatı bulduklarını belirtmişlerdir. Katılımcılardan biri büyük veri deneyimleriyle ilgili şu ifadelerle yer vermiştir.

“Kafka, Nifi, Sqoop, Hive ve Hadoop gibi büyük veri araçlarından bazılarının kurulumunu gerçekleştirerek, deneme fırsatım olmuştur.” (Kişi-6)

Katılımcıların bir kaçı Map-Reduce (Eşle-İndirge) fonksiyonlarını daha önce duyduğu ve bir katılımcının Map-Reduce mantığını kullanarak çalışmalar yaptığı gözlem notları ve öz-değerlendirmelerinden anlaşılmaktadır.

İkinci kategoride eğitime katılan öğrencilerin lisans eğitimi boyunca büyük veri ile alakalı ders alıp almadıklarının belirlenmesidir. Katılımcıların yarıdan fazlası, lisans eğitimi boyunca büyük veri ile ilgili herhangi bir ders almadıklarını belirtmektedir. Katılımcılardan bir kaçı seçmeli ‘Veri Bilimine Giriş’ dersi aldıklarını ancak dersin giriş seviyesinde olduğunu ve uygulamaya yönelik olmadığını belirtmektedirler. Bu konuda katılımcılardan biri şu ifadelerle yer vermiştir.

“Veri bilimine giriş dersinde büyük verinin çok az konusu geçmişti. Bunun dışında uygulama veya direk teorik anlamda bir ders almadım.” (Kişi-8)

Katılımcılardan bir tanesi bu derste büyük veri alanına ilgisinin oluştuğunu ve bu sayede bitirme projesini bu alanda yapmaya karar verdiğini ifade etmektedir. Katılımcıların çoğu lisans dönemi boyunca büyük veri ile alakalı yeterli eğitim alamadığını düşünmektedir.

Katılımcıların hepsi, bu alana ilgili olduklarını ve bu alanla ilgili eğitim, seminer ve konferanslara katılmak ve üniversiteden mezun olmadan önce bu alanla ilgili ders almak istediğini belirtmektedir.

Son kategoride, lisans eğitimi dışında çevrimiçi veya yüzyüze büyük veri ile ilgili eğitimlere katılıp katılmadıkları belirtilmektedir. Katılımcıların çoğu kendilerini geliştirmek için çevrimiçi ya da yüzyüze alanında uzman kişiler tarafından eğitim aldıklarını belirtmektedirler. Online eğitimler için dijital olarak ders içerikleri sunan Udemy'nin tercih edildiği dikkat çekicidir. Katılımcılarda biri bu konuda şu ifadeler yer vermiştir.

“Udemy üzerinden Spark ve Flink eğitimlerine katıldım.” (Kişi-5)

Karşılaşılan sorunlar

Bu kategoride, büyük veri araçlarının kurulumu ve kullanımı sırasında ve öğrenme sürecinde katılımcıların karşılaştıkları sorunlara yer verilmektedir. Büyük veri araçlarından herhangi birinin kurulumunu yapan katılımcıların hepsi araçların kurulumu sırasında çeşitli sorunlarla karşılaştıklarını belirtmektedirler. Bu sorunlar şu şekilde ifade edilmektedir;

- Kurulum aşamasının vakit alması ve diğer araçlarla uyumlu çalıştırabilme
- Sistem gereksinimleri ve sürüm hataları
- Diğer platformlarla bağlantı hataları
- Yanlış veya eksik kurulum gerçekleştirme
- Hadoop kurulumunda entegrasyon hataları
- Veri tabanı kurarken erişim sorunları ve seri olarak SQL çalıştırma hataları

Büyük veri araçlarının kullanımı sırasında karşılaşılan sorunlar daha çok örnek uygulama ve materyal eksikliğinden kaynaklanmaktadır. Katılımcıların belirttikleri sorunlar şu şekildedir;

- Örnek kod ve uygulamaların yetersiz olması
- Araçlarla ilgili kaynak ve materyal eksikliği

- Bilgi eksikliği
- Verinin entegre edilmesi ve kaydedilmesiyle ilgili sorunlar

Büyük veri analitiğini ve araçlarını öğrenirken karşılaşılan sorunlar yine kaynak ve eğitim eksikliği üzerinde yoğunlaşmaktadır. Katılımcıların belirttikleri sorunlar aşağıdaki gibidir;

- Materyal, eğitim ve örnek uygulama eksikliği
- Lisans eğitimi dahil çevrimiçi ve yüz yüze eğitimlerin yetersizliği
- Açık olarak paylaşılabilen veri ihtiyacı
- Donanım (laboratuvar) ihtiyacı: Donanımsal olarak kişisel bilgisayarların büyük veri ortamına yeterli olmaması
- Büyük veri mantığını öğrenme konusunda sıkıntılar
- Gerçek veriler üzerinde çalışamama

Eğitim süresince, katılımcıların sordukları sorular ve eğitmenin yönelttiği sorular karşısında katılımcıların verdikleri cevaplar sonucunda, genel olarak katılımcıların büyük veri ile ilgili olarak temel bazı kavramlarda bilgi eksikliği yaşadıkları ve pratikte de çok çeşitli sorunlarla karşılaştıkları gözlemlenmiştir. Katılımcıların kendi öz değerlendirmeleri ile lisans eğitimi boyunca bu alanda yeterli eğitim alamadıkları ifade edilmektedir. Kendilerini bu alanda geliştirebilmek için üniversite dışında çevrimiçi veya yüz yüze eğitimlere katıldıklarını ve karşılaştıkları sorunları genellikle kendi çabalarıyla internetten araştırma yaparak çözmeye çalıştıklarını belirtmektedirler. Bir kaç katılımcı sorunların üstesinden gelmeye çalışırken arkadaşlarına ve hocalarına danıştıklarını ifade etmiştir.

Beklentiler

Bu bölümde, katılımcıların büyük veri alanıyla ilgili, okumakta oldukları üniversite ve bölümden beklentilerine yer verilmiştir. Katılımcıların çoğu, büyük veri alanı ile ilgili üniversitelerde bölüm açılması gerektiğini düşünmektedir. Kaliforniya Berkeley Üniversitesi, 2014 yılında, veri bilimi eğitimine artan ilgiye yanıt vermek için 30'dan fazla ders müfredatı ile lisans düzeyinde veri bilimi eğitimi programı başlatmıştır. Program yöneticisi Cathryn Carson, veri bilimi anadalının disiplinler arası bir program olduğunu, bu programı öğrencilerin ihtiyaçlarına göre yaptıklarını ve öğrencileri gerçek dünyadaki işlere hazırlayan bir anadal başlatmanın heyecan verici olduğunu ifade etmiştir (Harcourt, 2018).

BachelorsPortal⁴⁰ üzerinden ‘Veri Bilimi ve Büyük Veri’ disiplini ile arama yapıldığında, sadece Amerika Birleşik Devletleri’nde örgün eğitim olarak lisans programı sayısı 100 olarak gösterilmektedir (BachelorsPortal, 2020). Bu doğrultuda eğitime katılan öğrencilerin de çoğu büyük veri ile ilgili üniversitelerde bölüm açılması konusunda aynı görüştedir. Katılımcıların hepsi öz değerlendirmelerinde mezun olmadan önce büyük veri ile ilgili bir ders almak istediklerini, bilgisayar mühendisliği ve ilgili bölümlerde büyük veri konusunun müfredatta daha fazla yer edinmesi gerektiğini ifade etmektedir. Okur ve Büyükkeçeci (2014), bilgisayar bilimleri ile alakalı bölümlerde müfredatların, özenle hazırlanmış temel ve seçmeli derslerle dolu olduğunu ancak yeni derslerin eklenmesinin zor olduğunu belirtmektedir. Çoğu temel dersin içeriği ve öğrenme çıktıları 1960’lı 1970’li yıllara dayanan gelişmelere dayanmaktadır. En belirgin örnekler; Veri Yapıları ve Algoritmalar, Veri Yönetimi, Dosya Organizasyonu ve Veri Tabanı Sistemleri dersleridir. Bilişim teknolojileri sektörü tarafından benimsenen büyük veri gelişmeleri için müfredatta yer açılması zor görünmekte ancak büyük veri ile alakalı bazı temel konuların mevcut temel ve seçmeli derslere entegre edilmesi gerekmektedir (Okur ve Büyükkeçeci, 2014). Müfredatta daha fazla büyük veri ile ilgili ders açılması konusunda katılımcılar da aynı fikirdedirler. Bu konuda katılımcıların 10’u bilgisayar mühendisliği ve ilgili bölümlerde büyük veri dersinin zorunlu ders olarak verilmesi gerektiğini, 8’i seçmeli ders olarak verilmesi gerektiğini belirtmiştir. Aynı zamanda, katılımcıların çoğu mezun olduktan sonra büyük veri ile ilgili bir bölümde yüksek lisans yapmak istediğini belirtmiştir. Katılımcıların beklentileri şu şekildedir;

- Yeni teknolojilere uygun seçmeli ders sayısının artırılması
- Laboratuvarlarda uygulamalı derslerin ve proje temelli derslerin daha çok olması
- Donanımsal kaynakların öğrencilere sunulması
- Dersleri veren hocaların sektörden alanında uzman kişiler tarafından verilmesi
- Alanında uzman kişiler tarafından eğitim ve kursların verilmesi
- Büyük veri ile ilgili seminerlerin sayısının artırılması
- Kurumlarla açık veriye erişme konusunda anlaşmalar yapılması
- Bitirme projelerinin bu alanlarda daha çok verilmesi

⁴⁰ BachelorsPortal: Uluslararası, üniversitelerde lisans programı arama portalı

3.4. Çalışma 3 – Öğretim Elemanlarının Büyük Veri Beklentilerinin Değerlendirilmesi

3.4.1. Amaç

Bu çalışmanın amacı, öğretim elemanlarının büyük veri alanı ile ilgili farkındalık ve eğitimle alakalı beklentilerinin değerlendirilerek belirlenmesidir. Bu doğrultuda aşağıdaki soruların cevapları aranmaktadır.

- Öğretim elemanlarının aşına oldukları ve kullandıkları büyük veri yazılımları nelerdir?
- Öğretim elemanlarının büyük veri alanına yönelik akademik ilgileri ne yöndedir?
- Öğretim elemanlarının büyük veri alanı ile ilgili eğitimle alakalı beklentileri nelerdir?

3.4.2. Örneklem

Araştırmanın evreni olarak Ankara ilinde çizelge 3.9’da belirtilen üniversitelerin Bilgisayar Mühendisliği, Matematik, İstatistik ve Yönetim Bilişim Sistemleri bölümlerinde ders veren öğretim elemanları olarak belirlenmiştir. Bu bölümlerin tercih edilmesi, birinci çalışmada büyük veri çözümleri sunan kurumlarla yapılan görüşmelerden sağlanan bilgiler neticesinde olmuştur. Toplam 153 üniversite öğretim elemanı araştırmaya veri sağlamıştır. Başkent, Ostim Teknik ve Ankara Yıldırım Beyazıt üniversitelerinin Yönetim Bilişim Sistemleri bölümündeki öğretim elemanları araştırmanın evrenine dâhil edilmiştir. Bilgisayar Mühendisliği, Matematik ve İstatistik bölümlerinde ise anket sorularına cevap veren öğretim elemanlarının bölümlerindeki toplam sayı evrene dâhil edilmiştir. Bu doğrultuda üniversitelerin belirlenen bölümlerindeki öğretim elemanlarının sayısı her üniversitenin resmi web sayfasından alınarak, araştırmanın evreni 248 olarak belirlenmiştir. Krejcie ve Morgan (1970), örneklem büyüklüğü belirleme tablosunda 0,05 örnekleme hatası (d) ile 250 kişilik bir evren için, 152 kişilik bir örneklem olması gerektiğini belirtmektedir. Bu doğrultuda, 248 evren büyüklüğünde 153 öğretim elemanının katıldığı örneklem, araştırma yapmak için uygun olduğu gözlemlenmektedir. Katılımcılara dair çeşitli tanımlayıcı ve demografik istatistikler Çizelge 3.9’da gösterilmiştir.

Çizelge 3.9. Öğretim elemanlarının tanımlayıcı istatistikleri

		N	%
Cinsiyet	Erkek	110	71,9
	Kadın	43	28,1
Üniversite	Ankara Üniversitesi	17	11,1
	Ankara Yıldırım Beyazıt Üniversitesi	7	4,6
	Atılım Üniversitesi	15	9,8
	Başkent Üniversitesi	17	11,1
	Bilkent Üniversitesi	6	4
	Çankaya Üniversitesi	10	6,5
	Gazi Üniversitesi	23	15
	Hacettepe Üniversitesi	17	11,1
	Orta Doğu Teknik Üniversitesi	19	12,4
	Ostim Teknik Üniversitesi	11	7,2
	TOBB Ekonomi Ve Teknoloji Üniversitesi	7	4,6
	Belirtilmemiş	4	2,6
Unvan	Prof.Dr.	57	37,3
	Doç. Dr.	22	14,4
	Dr. Öğr. Üyesi	46	30,1
	Dr. Öğr. Gör.	1	0,7
	Öğr. Gör.	13	8,5
	Araş. Gör.	12	7,8
	Belirtilmemiş	2	1,4
Bölüm	Bilgisayar Mühendisliği	64	41,8
	İstatistik Bölümü	34	22,2
	Matematik Bölümü	36	23,5
	Yönetim Bilişim Sistemleri	15	9,8
	Belirtilmemiş	4	2,6
Toplam		153	100,0

Katılımcıların %71,9'u erkek, %28,1'i kadındır. En çok katılım sırayla Gazi, Orta Doğu Teknik, Hacettepe, Ankara ve Başkent üniversitelerinden olmuştur. Bu üniversitelerin katılımları %10'un üzerindedir. Atılım Üniversitesi'nden de %10'a yakın katılım sağlanmıştır. Bölüm olarak en çok katılım %41,8'lik oranla Bilgisayar Mühendisliği'nden olmuştur. Matematik bölümünden %23,5, İstatistik bölümünden %22,2, Yönetim Bilişim Sistemleri bölümünden %9,8 oranında katılım sağlanmıştır. Katılımcıların %2,6'sı üniversite belirtmemiştir. Katılımcıların %37,3'ü Profesör, %14,4'ü Doçent, %30,1'i Dr. Öğretim Üyesi, %8,5'i Öğretim Görevlisi, %7,8'si Araştırma Görevlisi ve %0,7'si de Dr. Öğretim Görevlisi olarak belirlenmiştir. Katılımcıların %1,4'ü unvan belirtmemiştir.

3.4.3. Veri toplama

Bu çalışmada veriler, nicel araştırma yöntemlerinin veri toplama yöntemi olan anket yoluyla elde edilmiştir. Anket üç bölümden oluşmaktadır. Birinci bölüm, öğretim

elemanlarının büyük veri alanına olan akademik ilgileri ve büyük veri yazılımlarına olan aşinalık ve kullanımlarıyla ilgili dokuz sorudan oluşmaktadır. İkinci bölümde, öğretim elemanlarının büyük veri konusundaki beklentilerini anlamaya ve Türkiye'deki mevcut durumu değerlendirmelerine yönelik ifadelerin yer aldığı 21 maddelik bir soru formu hazırlanmıştır. Üçüncü bölümde ise, katılımcıların, cinsiyet, unvan, bölüm ve üniversite bilgilerinin yer aldığı demografik sorular yer almaktadır. Anket soruları çevrimiçi olarak tasarlanmıştır ve ilgili üniversitelerdeki öğretim elemanlarına e-posta yoluyla ulaştırılmıştır. Söz konusu form için herhangi bir ölçek geliştirme çalışması örneklemin az olması sebebiyle gerçekleştirilmemiştir.

3.4.4. Araştırmanın kısıtlılıkları

Bu araştırma, büyük veri alanı ile ilgili olabilecek bölümlerde eğitim veren öğretim elemanlarını içermektedir. Bu doğrultuda, çalışmaya katılım Ankara ili içerisindeki 11 üniversitenin Bilgisayar Mühendisliği, Matematik, İstatistik ve Yönetim Bilişim Sistemleri bölümlerinde ders veren öğretim elemanları ile kısıtlanmaktadır. Yönetim Bilişim Sistemleri bölümü, Başkent, Ostim Teknik ve Ankara Yıldırım Beyazıt üniversitelerinde yer alan öğretim elemanları ile kısıtlanmaktadır. Bilgisayar Mühendisliği, Matematik ve İstatistik bölümlerinde ise anket sorularına cevap veren öğretim elemanlarının bölümleri ile kısıtlanmaktadır. Öğretim elemanlarının yoğunluğu nedeniyle örnekleme ulaşmada zorluklar yaşanmıştır.

3.4.5. Verilerin analizi

Bu çalışmada, anket verilerinin analizi için SPSS 20.0 yazılımı ile veri analizi yöntemlerinden t-testi analizi kullanılmıştır. Ortalama farkı olarak da bilinen bu yöntemde, temel olarak iki ortalama karşılaştırılır ve istatistiki olarak bir fark olup olmadığı ölçülür. Tek örnek t-testi, bağımlı iki grup ortalamasının testi ve bağımsız iki grup t-testi olmak üzere üç t-testi yöntemi vardır (Kurtuluş, 2010). Burada tek örnek t-testi yöntemi kullanılmıştır. Bu yöntemde, bir tek örneğe ait olan ortalamanın beklenen bir ortalamadan veya belirlenen bir değerden farkının ölçülmesi ile gerçekleştirilir.

3.4.6. Bulgular

Kurtuluş (2010), Cronbach's alpha değerinin en az 0,70 olmasının uygun olduğunu belirtmektedir. Bu çalışmada güvenilirlik analizi yapıldığında, Cronbach's alpha değeri 0,865 olarak hesaplanmıştır. Çizelge 3.10 incelendiğinde, öğretim elemanları tarafından en çok aşına olunan büyük veri yazılımları Oracle Big Data, Apache Hadoop, Microsoft Hadoop ve Spark, en çok kullanılan yazılımlar ise sırayla Apache Hadoop, Oracle Big Data, Spark ve HDFS olmuştur. Ancak kullanılan yazılım sayılarının oldukça düşük olduğu görülmektedir. Katılımcıların aşına oldukları ve kullandıkları büyük veri yazılımlarının detaylı tablosu Ek-2'de paylaşılmıştır.

Çizelge 3.10. Öğretim elemanlarının büyük veri yazılımlarına aşinalıkla ilgili tanımlayıcı istatistikleri

	Ortalama (μ)	Standart Sapma
Aşına olunan büyük veri yazılım sayısı	2,73	3,414
En az bir kere kullanılan büyük veri yazılım sayısı	1,08	2,113
n		
En fazla tanınan yazılımlar	Oracle Big Data	76
	Apache Hadoop	67
	Microsoft Hadoop	46
	Spark	30
En fazla kullanılan yazılımlar	Apache Hadoop	19
	Oracle Big Data	11
	Spark	10
	HDFS	8
	Hiçbirini kullanmadım	55

Çizelge 3.11'de akademisyenlerin büyük veriyle ilgili akademik ilgilenimleri, tamamlamakta oldukları veya halen üzerinde çalıştıkları akademik çalışmalar üzerinden gösterilmektedir. Akademisyenlerin %15,5'inin büyük veri ile ilgili yayınlanmış bilimsel makalesinin olduğu ve %17'sinin devam etmekte olan yayınının olduğu görülmektedir. Akademisyenlerin %7,8'inin tamamlanmış projesi varken, %14,4'ünün halen devam etmekte olan projesinin olduğu görülmektedir. Bu yüzdelerle bakıldığında bilimsel ayın ve proje geliştirme konusunda öğretim elemanlarının büyük veri alanına olan ilgilerinin arttığı gözlemlenmektedir. Akademisyenlerin büyük veri konusu ile ilgili yürütülmüş olan lisansüstü tez danışmanlığının oranı %13'7 iken, yürütmekte olunan tez danışmanlığı oranı %8,5 olduğu görülmektedir. Öğretim elemanlarının büyük veri alanı ile ilgili yürütmekte

oldukları tez danışmanlıklarında, tamamlanmış tez danışmanlıklarına göre azalma olduğu görülmektedir.

Çizelge 3.11. Büyük veri alanına yönelik akademik ilgi

İfade	Evet		Hayır	
	n	%	n	%
1. Büyük Veri konusu ile ilgili yayınlanmış bilimsel yayınınız var mı?	24	15,5	129	84,3
2. Büyük Veri konusu ile ilgili üzerinde çalışmakta olduğunuz (halen devam etmekte olan) bilimsel yayınınız var mı?	26	17	127	83
3. Büyük Veri konusu ile ilgili tamamlanmış projeniz var mı?	12	7,8	141	92,2
4. Büyük Veri konusu ile ilgili üzerinde çalışmakta olduğunuz (halen devam etmekte olan) projeniz var mı?	22	14,4	131	85,6
5. Büyük Veri konusu ile ilgili yürütmüş olduğunuz lisansüstü tez danışmanlığınız var mı?	21	13,7	132	86,3
6. Büyük Veri konusu ile ilgili yürütmekte olduğunuz lisansüstü tez danışmanlığınız var mı?	13	8,5	140	91,5

Bu çalışmada merak edilen sorulardan biri de öğretim elemanlarının mevcut lisans müfredatında büyük veri konusundan bahsedip bahsetmediği ve eğer bahsediyorsa bunun hangi derslerde olduğudur. Çizelge 3.12’de, katılımcıların %7,8’i büyük veri ile ilgili bir ders yürüttüğünü belirtmektedir. Büyük veri ile ilgili yürütülen dersler içerisinde, Büyük Veri, Büyük Veri Programlama, Büyük Veri ve Kişisel Veri Güvenliği, Dağıtık Veri İşleme ve Analiz ve Veri Madenciliği gibi derslerin bulunduğu görülmektedir. Katılımcıların %38’i yürüttükleri derslerde büyük veri konusundan bahsettiklerini belirtmektedir. Katılımcıların verdikleri cevaplar doğrultusunda büyük veri konusundan bahsedilen derslerin çok çeşitli olduğu ve farklı bölümleri ilgilendirdiği görülmektedir. Büyük veri konusundan bahsedilen derslerin oranının yüksek olması ve çok çeşitli derslerde bahsedilmesi, ihtiyaç duyulan bir ders olduğu sonucuna varılabilmektedir.

Çizelge 3.12. Büyük veri konusunun bahsedildiği dersler

	Evet		Hayır	
	n	%	n	%
Lisans düzeyinde büyük veri ile ilgili her hangi bir ders yürüttünüz mü?	12	7,8	141	92,2
Yürütmüş olduğunuz dersin adını belirtiniz.				
Bilgi Sistemleri	İstatistik ve Akıllı Yaklaşımlar			
Büyük Veri	Veri Madenciliği			
Büyük Veri Programlama	Veri Yoğunluğu Uygulamaları			
Büyük Veri ve Kişisel Veri Güvenliği	Yönetim Bilişim Sistemleri			
Dağıtık Veri İşleme ve Analiz	Zaman Serisi Analizi			
	n	%	n	%

Çizelge 3.12. (devam) Büyük veri konusunun bahsedildiği dersler

Lisans düzeyinde yürüttüğünüz derslerin herhangi birinde büyük veri konusundan bahsediyor musunuz?	58	37,9	95	62,1
Lisans düzeyinde Büyük Veri konusundan hangi derste bahsediyorsunuz?				
Algoritma	Oyunlaştırma			
Bilişim Sistemleri Giriş	Örüntü Tanıma			
Bilişim Teknolojileri	Proje yönetimi			
Biyoinformatik Algoritmaları	Üretim Planlama ve Stok Kontrolü			
Bulut Hesaplama	Veri Bilimi			
Büyük Veri	Veri Madenciliği			
Eğitimde Yeni Teknolojiler	Veri Tabanı			
Görüntü İşleme	Veri Yapıları			
İleri Programlama	Veri Yoğunluk Uygulamaları			
İstatistik	Yalın Üretim			
Kestirim (Forecasting) Yöntemleri	Yazılım Mühendisliği			
Makine Öğrenmesi ve Veri Yönetimi	Yönetim Bilişim Sistemleri			
Nitel Araştırma Yöntemleri				

Anket formunda kullanılan ifadeler, bu ifadelerin ortalama ve standart sapma değerleri tek örneklem t-testi sonuçlarıyla birlikte Çizelge 3.13’de paylaşılmıştır. T-testi için referans değeri ölçeğin orta değeri olan 3 (Fikrim Yok) olarak alınmıştır.

ACM ve IEEE (2020), bilişim müfredatına, Veri Bilimi programını eklenmiştir ve lisans programında Büyük Veri dersleri de bu alanda yer almaktadır. Aynı doğrultuda katılımcılar, büyük veri konusunun lisans müfredatında daha fazla yer alması gerektiğini (3,830) ve müfredatta mevcut bir derse konularının eklenmesi gerektiğini (3,895) düşünmektedir (Ortalama değer (3,830) ve (3,895) ölçeğin orta noktasından (3) istatistiki olarak anlamlı bir şekilde ($p=,000$) yüksektir). Ankette verilen ifadelerden ortalamanın üzerindeki en büyük fark, yani katılımcıların en çok hem fikir olduğu ifade, bilgisayar mühendisliği ve ilgili bölümlerde büyük veri üzerine seçmeli ders açılması ($\mu=4,248$) ve Türkiye’de büyük veri çözümlerinin önümüzdeki yıllarda daha çok kullanılacağını düşünülmesi ($\mu=4,209$) olduğu görülmektedir.

Çizelge 3.13. Öğretim elemanlarının tek örneklem t-testi analizi

İfadeler (N=153)	Ortalama (μ)	Standart Sapma	t	df	p	Ortalama Farkı
Büyük Veri ile ilgili konular lisans müfredatında daha fazla yer almalıdır.	3,830	,8568	,8568	,8568	,000	,8301
Lisans düzeyinde Büyük Veri konusunun, müfredatta mevcut bir derse konularının eklenmesi gerektiğini düşünüyorum.	3,895	,8596	,8596	,8596	,000	,8954
Lisans ders müfredatlarında, teknolojik gelişmelere göre güncelleme yapıldığını düşünüyorum.	3,111	1,1154	1,1154	1,1154	,220	,1111
Lisans ders müfredatlarında, sektörün ihtiyaçları göz önünde bulundurularak güncelleme yapıldığını düşünüyorum.	2,928	1,0767	1,0767	1,0767	,410	-,0719
Bilgisayar Mühendisliği ve ilgili bölümlerde Büyük Veri üzerine ders açılmalıdır. [Zorunlu Ders]	3,353	1,1780	1,1780	1,1780	,000	,3529
Bilgisayar Mühendisliği ve ilgili bölümlerde Büyük Veri üzerine ders açılmalıdır. [Seçmeli Ders]	4,248	,8451	,8451	,8451	,000	1,2484
Büyük Veri ile ilgili üniversitelerde lisans programları açılmalıdır.	2,562	1,1575	1,1575	1,1575	,000	-,4379
Büyük Veri alanında lisansüstü daha çok program açılmalıdır.	3,556	1,0569	1,0569	1,0569	,000	,5556
Büyük Veri ile ilgili etkinliklere katılmak isterim. [Eğitim]	3,647	1,2111	1,2111	1,2111	,000	,6471
Büyük Veri ile ilgili etkinliklere katılmak isterim. [Seminer]	3,869	1,0176	1,0176	1,0176	,000	,8693
Büyük Veri ile ilgili etkinliklere katılmak isterim. [Konferans]	3,739	1,0989	1,0989	1,0989	,000	,7386
Büyük Veri ile ilgili etkinliklere katılmak isterim. [Online Eğitim]	3,732	1,1921	1,1921	1,1921	,000	,7320
Büyük Veri ile ilgili etkinliklere katılmak isterim. [Çalıştay]	3,641	1,1732	1,1732	1,1732	,000	,6405
Sosyal medyada Büyük Veri ile ilgili paylaşımlar ilgimi çeker.	3,595	1,0224	1,0224	1,0224	,000	,5948
Türkiye’de büyük veri çözümlerinin önümüzdeki yıllarda daha çok kullanılacağını düşünüyorum.	4,209	,7403	,7403	,7403	,000	1,2092
Türkiye’de Büyük Veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Kamu Kurumlarında]	3,484	1,0007	1,0007	1,0007	,000	,4837
Türkiye’de Büyük Veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Özel Kurumlarda]	3,144	1,0222	1,0222	1,0222	,084	,1438
Türkiye’de Büyük Veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Yükseköğretim Kurumlarında]	3,333	1,1239	1,1239	1,1239	,000	,3333
Görev yapmakta olduğum kurumda, yöneticiler Büyük Veri alanına gerekli önemi vermektedir.	3,098	,9785	,9785	,9785	,217	,0980
Yurt dışında bir üniversitede, Büyük Veri ile ilgili bir pozisyonda görev yapmak isterim. [Kısa Süreli Pozisyon]	3,131	1,2123	1,2123	1,2123	,184	,1307
Yurt dışında bir üniversitede, Büyük Veri ile ilgili bir pozisyonda görev yapmak isterim. [Uzun Süreli Pozisyon]	2,680	1,2012	1,2012	1,2012	,001	-,3203

1= Kesinlikle Katılmıyorum... 5= Kesinlikle Katılıyorum

Ayrıca bilgisayar mühendisliği ve ilgili bölümlerde büyük veri üzerine zorunlu ders açılması ($\mu=3,353$) konusuna da katılmakta oldukları görülmektedir. Katılımcılar, büyük

veri ile ilgili üniversitelerde daha çok lisansüstü program açılması gerektiğini düşünmekte ($\mu=3,641$), ancak lisans programı açılması konusuna katılmamaktadır ($\mu=2,563$). Katılımcıların ortalamasının üzerinde bir değerle, büyük veri ile ilgili eğitim, seminer, konferans, çevrimiçi eğitim ve çalıştay gibi etkinliklere katılma isteği göstermekte ve sosyal medyada bu alanla ilgili paylaşımlar ilgilerini çekmektedir ($p=,000$). Ayrıca öğretim elemanları, Türkiye’de büyük veri konusuna kamu kurumlarında ve yükseköğretim kurumlarında yeterince ilgi gösterilmediğini düşünmektedir.

3.5. Çalışma 4 – Lisans Öğrencilerinin Büyük Veri Yeterlilik, Farkındalık ve Beklentilerinin Ölçülmesi

3.5.1. Amaç

Bu çalışmanın amacı, öğrencilerin büyük veri alanı ile ilgili farkındalık, yeterlilik ve eğitimle alakalı beklentilerinin değerlendirilerek belirlenmesidir. Bu doğrultuda büyük veri kariyer ve eğitim beklentisi ölçeği ile büyük veri farkındalık ve yeterlilik ölçeği, öğrencilerin büyük veri konusundaki farkındalığını, yeterliliğini ve beklentilerini ölçmeye yönelik olarak ilk defa bu çalışma kapsamında geliştirilmiştir. Bu kapsamda aşağıdaki soruların cevapları aranmaktadır.

- Öğrencilerin aşına oldukları büyük veri kavramları ve yazılımları nelerdir?
- Öğrencilerin büyük veri alanına yönelik farkındalık, yeterlilik ve beklentileri nelerdir?

Bu çalışmada aynı zamanda, Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümlerinde okuyan öğrencilerin büyük veri konusundaki farkındalıkları arasında anlamlı bir fark olup olmadığı da incelenmiştir. Çalışma 1’de görüşülen kurumların beklentileri doğrultusunda, öğrencilerden bekledikleri teknik becerilerin kazanılacağı derslerin daha çok bu iki bölümde olması dolayısıyla bu karşılaştırma yapılmıştır. Bu ölçek dâhilinde bağımsız örneklem t-testi uygulanarak ankete katılan öğrencilerin bölümlerine göre faktörler arasında anlamlı bir farklılık olup olmadığı test edilmiştir. Bu bağlamda, H_0 ve H_1 hipotezi aşağıdaki gibi kurulmuştur:

H_0 : Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri öğrencilerinin büyük veri konusundaki farkındalıkları anlamlı bir şekilde farklı değildir.

H_1 : Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri öğrencilerinin büyük veri konusundaki farkındalıkları anlamlı bir şekilde farklıdır.

3.5.2. Örneklem

Bu çalışmanın evrenini, Ankara Yıldırım Beyazıt Üniversitesi Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümlerinde okuyan üçüncü ve dördüncü sınıf lisans öğrencileri oluşturmaktadır. Ankara Yıldırım Beyazıt Üniversitesi'nin tercih edilme nedeni, kurumların büyük veri alanında en çok tercih ettikleri bölümleri içeren Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümlerinin ikisinin de bu üniversitede bulunmasıdır. Görüşülen kurumlardan alınan bilgiler doğrultusunda, büyük veri alanında işe almak istedikleri kişilerde bulunması gereken özellikler göz önünde bulundurulduğunda, Türkiye'de yer alan üniversitelerin en çok Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümlerinde bu özellik ve yeteneklerin kazanılabileceği görülmektedir. Bu doğrultuda, Ankara ilinde bu iki lisans programı olan Ankara Yıldırım Beyazıt Üniversitesi tercih edilmiştir. Yükseköğretim Lisans Atlası kaynağından alınan verilere göre 2016 ve 2017 yıllarında Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümüne toplam 240 kayıtlı öğrenci bulunmaktadır (YÖK Atlas, 2020). Dolayısıyla araştırmanın evreni, 240 kişi olarak belirlenmiştir. Krejcie ve Morgan (1970), örnek büyüklüğü belirleme tablosunda 0,05 örnekleme hatası (d) ile 240 kişilik bir evren için, 148 kişilik bir örneklem olması gerektiğini belirtmektedir. Bu doğrultuda, 240 evren büyüklüğünde 179 öğrencinin katıldığı bu örneklem, araştırma yapmak için uygun olduğu gözlemlenmektedir. Katılımcılara dair çeşitli tanımlayıcı ve demografik istatistikler Çizelge 3.14'de gösterilmiştir.

Çizelge 3.14. Öğrencilerin tanımlayıcı istatistikleri

	Ortalama (μ)	Standart Sapma	
Yaş	22,011	2,4723	
		n	%
Cinsiyet	Erkek	116	64,8
	Kadın	63	35,2
Bölüm	Yönetim Bilişim Sistemleri	103	57,5
	Bilgisayar Mühendisliği	76	42,5
Sınıf	3	82	45,9
	4	95	53
	4+	2	1,1
Not Ortalaması (4'lük Sistem)	0 ile 1,5 arasında	5	2,8
	1,5 ile 2 arasında	3	1,7
	2 ile 2,5 arasında	45	25,1
	2,5 ile 3 arasında	62	34,6
	3 ile 3,5 arasında	45	25,1
	3,5 ile 4 arasında	19	10,6
Toplam		179	100

Katılımcıların %65'i erkek, %35'i kadın, %43'ü Bilgisayar Mühendisliği öğrencisi, %57'si Yönetim Bilişim Sistemleri öğrencisidir. Katılımcıların bölümleri arasındaki dağılımın dengeli olduğu söylenebilir. Katılımcıların %46'sı üçüncü sınıf, %54'ü dördüncü sınıf öğrencisidir.

3.5.3. Veri toplama

Bu çalışmada veriler, nicel araştırma yöntemlerinin veri toplama yöntemi olan anket yoluyla elde edilmiştir. Anket üç bölümden oluşmaktadır. Birinci bölüm, öğrencilerin büyük veri kavramına ve yazılımlara aşinalığıyla ilgili sorulardan oluşmaktadır. İkinci bölümde, öğrencilerin büyük veri konusundaki farkındalıklarını, yeterliliklerini ve beklentilerini anlamaya yönelik sorular yer almaktadır. Üçüncü bölümde, katılımcıların, yaş, cinsiyet, bölüm, sınıf, not ortalaması bilgilerinin yer aldığı demografik sorular yer almaktadır. Anket soruları öğretim üyelerinin izni ile derslerden önce öğrencilerin katılımıyla gerçekleştirilmiştir. Katılımcılar tamamen gönüllülük esasına göre anketi doldurmuş ve 179 üniversite öğrencisi araştırmaya veri sağlamıştır. Ölçek geliştirme temel olarak sekiz aşamayı kapsamaktadır. Bunlar; neyin ölçülmek istendiğinin belirlenmesi, madde havuzunun oluşturulması, ölçme aracının formatına karar verilmesi, maddelerin uzmanlar tarafından gözden geçirilmesi, madde geçerliliğinin sağlanması, ölçeğin uygulanması, maddelerin değerlendirilmesi ve ölçeğin son halinin verilmesidir (DeVellis, 2003). İlk olarak ölçülmek istenen olgu veya durumlar açıkça belirlenmelidir. Bu

çalışmada, öğrencilerin büyük veri farkındalık, beklenti ve yeterliliklerinin ölçülmesi hedeflenmektedir. Ölçülmek istenen değişken ve ilişkili değişkenler detaylı olarak belirtilmiştir. İkinci olarak, madde havuzu oluşturulmalıdır. Çalışma 1’de kurumlarla yapılan görüşmeler, Çalışma 2’de öğrencilerle yapılan bir günlük eğitimin gözlem notları ve katılımcıların değerlendirmelerinden toplanan veriler ve ilgili literatürde yer alan çalışmalardan yararlanılarak 40 soruluk madde havuzu oluşturulmuştur. Bu çalışmada 5’li likert ölçeği kullanılmıştır. 40 soruluk madde havuzu alanında uzman, 5 Bilgisayar Mühendisliği Anabilim Dalı öğretim üyesi, 2 bilgisayar mühendisi özel sektör çalışanı ve 1 Eğitimde Ölçme ve Değerlendirme Anabilim Dalı öğretim üyesi tarafından gözden geçirilmiştir ve çeşitli düzenlemeler yapılarak ölçeğin uygulanması adımına geçilmiştir. Bu doğrultuda araştırmacının 40 maddeden oluşan madde havuzu, ölçeğin güvenilirliğinin istenilen seviyede olabilmesi için 33 maddeden oluşan son halini almıştır. Geliştirilen anketin son hali Ek-3’te paylaşılmıştır.

3.5.4. Araştırmanın kısıtlılıkları

Kurtuluş (2010), araştırma amacının ihtiyaç duyulan veriyi elde edecek kadar geniş, ancak bilgiye ulaşmayı zorlaştırmamak için de mümkün olduğu kadar kısıtlı tutulması gerektiğini belirtmektedir. Bu doğrultuda, araştırma Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümlerinin ikisinin de bulunduğu Ankara Yıldırım Beyazıt Üniversitesi Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri bölümü üçüncü ve dördüncü sınıf öğrencileri ile kısıtlanmaktadır.

3.5.5. Verilerin analizi

Bu çalışmada anket verilerinin analizi için SPSS 20.0 yazılımı kullanılmıştır. Ölçeğin geçerliliğini belirleyebilmek için keşfedici faktör analizi kullanılmıştır. Ölçekte kullanılan ifadelerin ortalaması ve bu ortalamanın ölçeğin orta noktasından uzaklığının anlamlı olup olmadığı tek örneklem t-testi ile analiz edilmiştir. Ayrıca bağımsız örneklem t-testi uygulanarak bölümlere göre faktörler arasında anlamlı bir farklılık olup olmadığı test edilmiştir.

3.5.6. Güvenilirlik ve geçerlilik

Ölçeğin güvenilirliğini ifade eden Cronbach's Alpha değeri 0,853 olarak bulunmuştur. Ölçeğin geçerliliğini belirleyebilmek için keşfedici faktör analizinden faydalanılmıştır. Gerçekleştirilen keşfedici faktör analizinin sonucunda ortaya çıkan alt boyutlar ve bunlara ait değerler Çizelge 3.15'de paylaşıldığı gibidir. Ölçekte kullanılan 22 maddenin 6 boyuta dağıldığı görülmektedir. Bu alt boyutlar; kariyer ve lisans dışı eğitim, lisans eğitimi, yurtdışı ve staj, bilgi alışverişi, Türkiye'de büyük veriye ilgi ve Türkiye'de fırsatlardır. Ölçeğin açıkladığı varyans %57,948 düzeyindedir. KMO değeri analizde yer alan örneklem büyüklüğünün yeterli olduğunu göstermektedir (,825). Bu değer 0,5'ten büyük olması ve mümkün mertebe 1'e yakın olması önerilmektedir (Çokluk, Şekercioğlu & Büyüköztürk, 2012:190-205). Barlett testi ise değişkenler arasındaki korelasyonun yeterliliğini göstermektedir ve istatistiki olarak anlamlı olması veri setinin faktör analizine elverişli olup olmadığını göstermektedir (Hair, Black, Babin ve Anderson, 1998). Barlett testi sonucunda uygulanan faktör analizinin bütün olarak anlamlı olduğu $p=0,000$ değeri ile doğrulanmıştır.

Çizelge 3.15. Büyük veri eğitim ve beklenti ölçeği güvenilirlik ve geçerlilik analizi

Büyük Veri Farkındalık ve Beklenti Ölçeği $\alpha = 0.853$ Toplam açıklanan varyans= %66,048 $KMO = ,825$; <i>Bartlett testi: 1751,551; $p = .000$</i>	Ortalama (μ)	Standart Sapma	Faktör Yükleri
Alt boyut 1: Kariyer ve Lisans Dışı Eğitim <i>Öz değer=6,839; $\alpha = ,858$</i>	3,676	,86793	
Mezun olduktan sonra Büyük Veri Analisti olarak çalışmak isterim.	3,553	1,0606	,830
Mezun olduktan sonra Büyük Veri analizi yapılan bir şirkette çalışmak isterim.	3,821	1,0339	,809
Büyük Veri ile ilgili bir bölümde yurt dışında lisansüstü eğitim almak isterim.	3,575	1,1892	,702
Büyük Veri ile ilgili bir bölümde lisansüstü eğitim almak isterim.	3,497	1,0933	,686
Büyük veri ile ilgili eğitim, seminer ve konferanslara katılmak isterim.	3,933	1,0472	,623
Alt boyut 2: Lisans Eğitimi <i>Öz değer=2,372; $\alpha = ,784$</i>	4,0570	,75795	
Bilgisayar mühendisliği ve ilgili bölümlerde Büyük Veri üzerine ders açılmalıdır. [Seçmeli Ders]	4,056	1,0899	,754
Lisans düzeyinde Büyük Veri konusunun müfredatta mevcut bir derse konularının eklenmesi gerektiğini düşünüyorum.	4,128	1,0001	,692
Büyük Veri ile ilgili konular lisans müfredatında daha fazla yer almalıdır.	4,196	,8551	,675
Mezun olmadan önce Büyük Veri ile ilgili ders almak isterim.	4,123	1,0203	,507
Bilgisayar mühendisliği ve ilgili bölümlerde Büyük Veri üzerine ders açılmalıdır. [Zorunlu Ders]	3,782	1,1816	,486
Alt boyut 3: Yurtdışı ve Staj <i>Öz değer=1,641; $\alpha = ,761$</i>	3,9106	,71822	
Yurt dışında Büyük Veri ile ilgili daha nitelikli eğitim alabilirim.	3,989	,9118	,809
Yurt dışında Büyük Veri alanında daha çok iş fırsatı vardır.	3,832	,9801	,705
Büyük Veri alanında stajların, kariyerime katkısı olacağını düşünüyorum [Kısa Dönem Staj]	3,844	,9529	,577
Büyük Veri alanında stajların, kariyerime katkısı olacağını düşünüyorum [Uzun Dönem Staj]	3,978	,9177	,562

Çizelge 3.15. (devam) Büyük veri eğitim ve beklenti ölçeği güvenilirlik ve geçerlilik analizi

	Ortalama (μ)	Standart Sapma	Faktör Yükleri
Alt boyut 4: Bilgi Alışverişi <i>Öz değer=1,362; α =,790</i>	3,4749	,86177	
Büyük veri konusunda bilgi alışverişi yapmaya istekliyimdir. [Arkadaşlarımla]	3,553	,9951	,861
Büyük veri konusunda bilgi alışverişi yapmaya istekliyimdir. [Sosyal Ağlarda]	3,346	1,1281	,755
Büyük veri konusunda bilgi alışverişi yapmaya istekliyimdir. [Hocalarımla]	3,525	,9500	,733
Alt boyut 5: Türkiye’de Büyük Veriye İlgisi <i>Öz değer=1,221; α =,639</i>	3,4786	,81497	
Türkiye’de büyük veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Yükseköğretim Kurumlarında]	3,520	1,0512	,780
Türkiye’de büyük veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Kamu Kurumlarında]	3,709	1,1190	,719
Türkiye’de büyük veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Özel Kurumlarda]	3,207	1,0368	,655
Alt boyut 6: Türkiye’de Fırsatlar <i>Öz değer=1,095; α =,662</i>	2,4972	,88447	
Türkiye’de Büyük Veri alanında yeterli eğitim fırsatı vardır.	2,547	,9897	,846
Türkiye’de Büyük Veri alanında istihdamın yeterli olduğunu düşünüyorum.	2,447	1,0553	,787
<i>1= Kesinlikle Katılmıyorum... 5= Kesinlikle Katılıyorum</i> <i>N=179</i>			

Alt boyutlar incelendiğinde ilk alt boyutun büyük veri alanındaki mezuniyet sonrası kariyer ve lisans eğitimi dışındaki ek eğitim faaliyetlerine katılma isteğiyle ilgili maddelerden oluştuğu görülmektedir. İkinci alt boyutta, lisans eğitiminde öğrencilerin beklentilerini içeren maddeler olduğu görülmektedir. Üçüncü alt boyutun, yurtdışında eğitim, kariyer ve yurtiçi ve yurt dışı staj fırsatlarıyla ilgili maddelerden oluştuğu görülmektedir. Dördüncü alt boyutta, öğrencilerin arkadaşlarıyla, öğretim elemanlarıyla ve sosyal ağlarda bilgi alışverişine olan istekleriyle ilgili maddeler yer almaktadır. Beşinci alt

boyut, Türkiye’de farklı kurumlarda büyük veriye olan ilgilerin değerlendirilmesiyle oluşan maddeleri kapsamaktadır. Sonuncu alt boyutta ise Türkiye’deki eğitim ve iş fırsatları ile ilgili maddeler bulunmaktadır.

3.5.7. Bulgular

Çizelge 3.16’da katılımcıların büyük veri kavramına ve yazılımlarına aşinalığıyla ilgili tanımlayıcı istatistikler sunulmuştur. Öğrencilerin çok büyük bir kısmının (%87,2) daha önce büyük veri kavramıyla karşılaştığı görülmektedir. Ankete katılan öğrenciler tarafından en çok tanınan teknolojiler, öğretim elemanları tarafından en çok tanınan teknolojilerle benzerdir. Bunlar; Oracle Big Data, Apache Hadoop ve Microsoft Hadoop’tur. Katılımcılar tarafından Hadoop araçlarından en çok bilinenler sırayla Spark, Hive ve Kafka’dır. Büyük veri yazılım ve platformlarının tanınma sayılarıyla ilgili detaylar Ek-4’deki tabloda paylaşılmıştır.

Çizelge 3.16. Büyük veri kavramı ve yazılımlarına aşinalıkla ilgili tanımlayıcı istatistikler

	Ortalama (μ)	Standart Sapma
Aşına olunan büyük veri yazılım sayısı	3,492	2,8012
		n
Büyük Veri ya da Büyük Veri Analitiği kavramlarını daha önce duyduğunuz mu?	Evet	156
	Hayır	23
Toplam		179
		%
		87,2
		12,8
		100
En fazla tanınan yazılımlar	Oracle Big Data	103
	Apache Hadoop	92
	Microsoft Hadoop	81

Büyük veri eğitim ve beklenti ölçeği, öğrencilerin büyük veri konusundaki eğitim beklentilerini ölçmeye yönelik olarak ilk defa bu çalışma kapsamında geliştirilmiştir. Ölçek 22 maddeden oluşmaktadır. Ölçeğin güvenilirliğini ifade eden Cronbach’s Alpha değeri 0,853 olarak bulunmuştur. Bu değer önerilen en düşük değerden ($\alpha > 0.7$) (Hair ve diğerleri, 1998) yüksekte olduğu için ölçek yeterince güvenilir kabul edilebilir. Ölçekte kullanılan ifadelerin ortalamasına bakmak ve bu ortalamanın ölçeğin orta noktasından yani 5’li likert ölçeğinin orta değeri olan 3 (Fikrim Yok) değerinden uzaklığının anlamlı olup olmadığı tek örneklem t-testi ile analiz edilmiştir. Çizelge 3.17’de bu analizin sonuçları paylaşılmıştır.

Çizelge 3.17. Büyük veri eğitim ve beklenti (tek örneklem t-testi)

İfadeler (N=179)	Ortalama (μ)	Standart Sapma	t	df	p	Ortalama farkı
Mezun olduktan sonra Büyük Veri Analisti olarak çalışmak isterim.	3,553	1,0606	6,976	178	,000	,5531
Mezun olduktan sonra Büyük Veri analizi yapılan bir şirkette çalışmak isterim.	3,821	1,0339	10,627	178	,000	,8212
Büyük Veri ile ilgili bir bölümde yurt dışında lisansüstü eğitim almak isterim.	3,575	1,1892	6,474	178	,000	,5754
Büyük Veri ile ilgili bir bölümde lisansüstü eğitim almak isterim.	3,497	1,0933	6,085	178	,000	,4972
Büyük veri ile ilgili eğitim, seminer ve konferanslara katılmak isterim.	3,933	1,0472	11,920	178	,000	,9330
Bilgisayar mühendisliği ve ilgili bölümlerde Büyük Veri üzerine ders açılmalıdır. [Zorunlu Ders]	3,782	1,1816	8,856	178	,000	,7821
Bilgisayar mühendisliği ve ilgili bölümlerde Büyük Veri üzerine ders açılmalıdır. [Seçmeli Ders]	4,056	1,0899	12,961	178	,000	1,0559
Lisans düzeyinde Büyük Veri konusunun müfredatta mevcut bir derse konularının eklenmesi gerektiğini düşünüyorum.	4,128	1,0001	15,096	178	,000	1,1285
Büyük Veri ile ilgili konular lisans müfredatında daha fazla yer almalıdır.	4,196	,8551	18,705	178	,000	1,1955
Mezun olmadan önce Büyük Veri ile ilgili ders almak isterim.	4,123	1,0203	14,725	178	,000	1,1229
Yurt dışında Büyük Veri ile ilgili daha nitelikli eğitim alabilirim.	3,989	,9118	14,510	178	,000	,9888
Yurt dışında Büyük Veri alanında daha çok iş fırsatı vardır.	3,832	,9801	11,363	178	,000	,8324
Büyük Veri alanında stajların, kariyerime katkısı olacağını düşünüyorum [Kısa Dönem Staj]	3,844	,9529	11,844	178	,000	,8436
Büyük Veri alanında stajların, kariyerime katkısı olacağını düşünüyorum [Uzun Dönem Staj]	3,978	,9177	14,253	178	,000	,9777
Büyük veri konusunda bilgi alışverişi yapmaya istekliyimdir. [Hocalarımla]	3,525	,9500	7,396	178	,000	,5251
Büyük veri konusunda bilgi alışverişi yapmaya istekliyimdir. [Arkadaşlarımla]	3,553	,9951	7,436	178	,000	,5531
Büyük veri konusunda bilgi alışverişi yapmaya istekliyimdir. [Sosyal Ağlarda]	3,346	1,1281	4,108	178	,000	,3464
Türkiye’de büyük veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Yükseköğretim Kurumlarında]	3,520	1,0512	6,613	178	,000	,5196
Türkiye’de büyük veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Kamu Kurumlarında]	3,709	1,1190	8,483	178	,000	,7095
Türkiye’de büyük veri konusuna yeterince ilgi gösterilmediğini düşünüyorum. [Özel Kurumlarda]	3,207	1,0368	2,667	178	,008	,2067

Çizelge 3.17. (devam) Büyük veri eğitim ve beklenti (tek örneklem t-testi)

İfadeler (N=179)	Ortalama (μ)	Standart Sapma	t	df	p	Ortalama farkı
Türkiye'de Büyük Veri alanında yeterli eğitim fırsatı vardır.	2,547	,9897	-6,117	178	,000	-,4525
Türkiye'de Büyük Veri alanında istihdamın yeterli olduğunu düşünüyorum.	2,447	1,0553	-7,012	178	,000	-,5531
<i>I= Kesinlikle Katılmıyorum... 5= Kesinlikle Katılıyorum</i>						
<i>n=179</i>						

Öğretim elemanları ile aynı doğrultuda öğrenciler, büyük veri konusunun lisans müfredatında daha fazla yer alması gerektiğini düşünmektedir (Ortalama değer [4,196] ölçeğin orta noktasından [3] istatistiki olarak anlamlı bir şekilde [$p=,000$] yüksektir). Katılımcıların en çok hem fikir olduğu ifadeler, bilgisayar mühendisliği ve ilgili bölümlerde büyük veri üzerine seçmeli ders açılması ($\mu=4,056$), zorunlu ders açılması ($\mu=3,782$), büyük veri konusunun müfredatta mevcut bir derse konularının eklenmesi gerektiği ($\mu=4,128$) ve mezun olmadan önce büyük veri ile ilgili ders almak istedikleri ($\mu=4,123$) olduğu görülmektedir ($p=,000$). Genel olarak, katılımcıların mevcut müfredatlarında büyük veri alanının daha çok yer alması gerektiği konusu 'katılıyorum' ve 'kesinlikle katılıyorum' aralığında çıkmaktadır. Katılımcılar aynı zamanda, Türkiye'de büyük veri alanında yeterli eğitim fırsatı olmadığını, yurt dışında büyük veri ile ilgili daha nitelikli eğitim alabileceğini ve büyük veri ile ilgili eğitim, seminer ve konferanslara katılmak istediklerini belirtmişlerdir. Katılımcılar büyük veri alanında uzun ve kısa dönem stajların kariyerlerine katkısı olacağını düşünmektedir.

Örneklem kısmında değinildiği üzere iki farklı bölüm (Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri) öğrencilerinden veri toplanmıştır. Çalışmalara genel bakış kısmında değinilen H1'in testi için bağımsız örneklem t-testi uygulanarak bölümlere göre faktörler arasında anlamlı bir farklılık olup olmadığı test edilmiştir. Bir başka deyişle H1 her bir alt boyut için ayrı ayrı test edilmiştir. Söz konusu test sonuçları Çizelge 3.18'de gösterilmiştir. H1 hipotezi, kariyer ve lisans dışı eğitim, lisans eğitimi, yurt dışı ve staj, bilgi alışverişi alt boyutları için doğrulanmışken, Türkiye'de farkındalık ve Türkiye'de fırsatlar alt boyutları için doğrulanmamıştır. YBS öğrencilerinin bilgisayar mühendisliği öğrencilerine kıyasla; Kariyer ve lisans dışı eğitim alt boyutu incelendiğinde, daha fazla büyük veri alanında kariyer yapma ve kendini geliştirmeye istekli oldukları görülmektedir.

($t[177]= 4,923$; $p=,000$). Lisans eğitimi alt boyutu incelendiğinde, büyük veri üzerinde lisans müfredatında daha fazla yer açılmasını düşündükleri ve mezun olmadan önce daha çok büyük veri dersi almak istekleri görülmektedir ($t[131,206]= 4,101$; $p=,000$). Yurt dışında daha nitelikli eğitim alabilecekleri ve iş fırsatlarının daha fazla olduğu, stajların kariyerlerine daha fazla katkı sağladığını düşünmektedirler ($t[177]= 4,538$; $p=,000$). Büyük veri alanı ile ilgili arkadaşlarıyla, hocalarıyla ve sosyal ağlarda daha çok bilgi alışverişi yapmaya istekli oldukları görülmektedir ($t[177]= 3,580$; $p=,000$). Ancak her iki grubunda Türkiye'deki büyük veri farkındalığının düşük olduğu ($t[177]= ,316$; $p=,753$) ve Türkiye'deki fırsatların az olduğu ($t[175,507]= -,831$; $p=,407$) konusunda görüşlerinin farklı olmadığı görülmektedir.

Çizelge 3.18. Bölümlere göre eğitim ve beklenti (bağımsız örneklem t-testi)

Alt Boyutlar	Bölüm	n	Ortalama (μ)	Standart Sapma	t	df	p	Ortalama Farkı
Kariyer ve Lisans Dışı Eğitim	Yönetim bilişim Sistemleri	103	3,9340	,72943	4,923	177	,000	,60766
	Bilgisayar Mühendisliği	76	3,3263	,92136				
Lisans Eğitimi	Yönetim bilişim Sistemleri	103	4,2563	,61956	4,101	131,206	,000	,46947
	Bilgisayar Mühendisliği	76	3,7868	,84433				
Yurtdışı ve Staj	Yönetim bilişim Sistemleri	103	4,1092	,68481	4,538	177	,000	,46778
	Bilgisayar Mühendisliği	76	3,6414	,67742				
Bilgi Alışverişi	Yönetim bilişim Sistemleri	103	3,6667	,85048	3,580	177	,000	,45175
	Bilgisayar Mühendisliği	76	3,2149	,81232				
Türkiye'de Büyük Veriye İlgisi	Yönetim bilişim Sistemleri	103	3,4951	,83479	,316	177	,753	,03901
	Bilgisayar Mühendisliği	76	3,4561	,79226				
Türkiye'de Fırsatlar	Yönetim bilişim Sistemleri	103	2,4515	,95874	-,831	175,507	,407	-,10775
	Bilgisayar Mühendisliği	76	2,5592	,77446				

Büyük veri farkındalık ve beklenti ölçeği dışında, büyük veri farkındalık ve yeterlilik ölçeği ise yine bu çalışma kapsamında geliştirilmiş olup öğrencilerin büyük veri konusunda en temel düzeyde bilgilerinin ölçülmesi amaçlanmıştır. Türk Dil Kurumu

sözlüğünde, farkındalık/farkında olmak, “görülmesi veya bilinmesi gereken şeylerden haberi bulunmak, kavranması gereken bir şeye dikkat etmek” olarak tanımlanmıştır (TDK, 2019). Bu tanımdan yola çıkarak, kişilerin bir konu hakkında farkındalığının olması, beraberinde o konu ile ilgili bilinmesi gereken şeylerden en temel düzeyde bilgi sahibi olmasının gerektiğini de ifade etmektedir. Dolayısıyla, farkındalıkla beraber en temel seviyede büyük veri ile ilgili yeterlilik ölçütleri bu ölçekte yer almaktadır. Orijinal ölçek 12 madde olarak geliştirilmiştir ancak ölçeğin güvenilirliğinin istenilen seviyeye çıkabilmesi için ($\alpha > 0.7$; Hair ve diğerleri, 1998) bir madde (Büyük veri analitiği ile ilgili yeteri kadar bilgi ve teknik beceriye sahip olduğumu düşünmüyorum) çıkartılmıştır.

Ölçek geliştirilirken Çalışma 1’de kurumlarla yapılan görüşmeler, Çalışma 2’de öğrencilerle yapılan bir günlük eğitimin gözlem notları ve katılımcıların değerlendirmelerinden toplanan veriler ve ilgili literatürde yer alan çalışmalardan yararlanılmıştır. Ölçeğin 11 maddeli son halinin Cronbach’s Alpha değeri 0,715 olarak hesaplanmıştır. Ölçekte kullanılan ifadeler ve bu ifadelere ait değerler Çizelge 3.19’da paylaşılmıştır. Aynı tabloda ayrıca tek örneklem t-testine göre ortalama değerlerin ölçeğin orta noktasından (3=Fikrim Yok) uzaklığı da paylaşılmıştır.

Çizelge 3.19. Büyük veri farkındalık ve yeterlilik (tek örneklem t-testi)

	Ortalama (μ)	Standart Sapma	t	df	p	Ortalama farkı
Ölçeğin ortalaması: 3,4733						
Standart Sapma: ,50558						
1. Büyük veri analitiği ile tam olarak neler yapılabileceğini biliyorum.	3,1844	1,08854	2,266	178	,025	,18436
2. Veri türünün çeşitlenmesi ve veri miktarının artmasıyla büyük veri çözümleri ortaya çıkmıştır.	3,9553	,92906	13,757	178	,000	,95531
3. Büyük verilere, bulut çözümlerinde depolama ve erişme imkânı sağlanmaktadır.	3,8324	,95686	11,639	178	,000	,83240
4. Büyük veri teknolojilerini kullanmak şirketlere çok çeşitli avantajlar sağlamaktadır.	4,2570	,82849	20,299	178	,000	1,25698
5. Büyük veri tanımını oluşturan 5V' nin ne olduğunu biliyorum.	2,6257	1,33223	-3,759	178	,000	-,37430
6. NoSql kavramının ne anlama geldiğini bilmiyorum.	3,1676	1,31323	1,707	178	,089	,16760
7. Map Reduce (Eşle-İndirge) kavramını daha önce duydum.	2,7318	1,37213	-2,615	178	,010	-,26816
8. Sosyal medya, akıllı cihazlar ve internet kullanımı büyük veri miktarını arttırmaktadır.	4,3073	,90597	19,305	178	,000	1,30726
9. Farklı kaynaklardan, çeşitli formatlarda gelen verinin işlenmesi, büyük veri analitiğinin zorluklarından biridir.	3,8156	,97415	11,202	178	,000	,81564
10. Büyük verileri işlemek için dağıtık sistemlerden yararlanılır.	3,4022	,90888	5,921	178	,000	,40223
11. Dağıtık bir dosya sisteminin nasıl çalıştığı hakkında fikrim var.	2,9274	1,08114	-,899	178	,370	-,07263
<i>1= Kesinlikle Katılmıyorum... 5= Kesinlikle Katılıyorum</i>						
<i>N=179</i>						

Katılımcılar, veri türünün çeşitlenmesi ve veri miktarının artmasıyla büyük veri çözümlerinin ortaya çıktığı ($\mu=3,9553$), büyük verilere, bulut çözümlerinde depolama ve erişme imkânı sağlandığı ($\mu=3,8324$), büyük veri teknolojilerini kullanmanın şirketlere çok çeşitli avantajlar sağladığı ($\mu=4,2570$), sosyal medya, akıllı cihazlar ve internet kullanımının büyük veri miktarını arttırdığı ($\mu=4,3073$) ve farklı kaynaklardan, çeşitli formatlarda gelen verinin işlenmesinin, büyük veri analitiğinin zorluklarından biri olduğu ($\mu=3,8156$) konularında farkındalığa sahip oldukları görülmektedir ($p=,000$). Ancak büyük veri tanımını oluşturan 5V'nin ne olduğu, NoSql kavramının ne anlama geldiği, Eşle-İndirge kavramının ne olduğu ve dağıtık bir dosya sisteminin nasıl çalıştığı konularında yeterlilik sahibi olmadıkları görülmektedir.

Büyük veri yeterliliğinin incelenen iki bölüm için farklı olup olmadığı bağımsız örneklem t-testiyle gerçekleştirilmiş olup elde edilen sonuçlar Çizelge 3.20’de paylaşılmıştır.

Çizelge 3.20. Bölümlere göre farkındalık ve yeterlilik (bağımsız örneklem t-testi)

Boyut	Bölüm	n	Ortalama (μ)	Standart Sapma	t	df	p	Ortalama Farkı
Büyük Veri Yeterlilik	Yönetim Bilişim Sistemleri	103	3,6169	,43722	4,544	142,933	,000	,33824
	Bilgisayar Mühendisliği	76	3,2787	,52912				

Belirtilen iki bölüm arasında büyük veri yeterliliği olarak belirgin bir fark olduğu görülmektedir (ortalama farkı=0,33824). Yönetim Bilişim Sistemleri öğrencilerinin, Bilgisayar Mühendisliği öğrencilerine göre yeterliliklerinin anlamlı bir şekilde yüksek olduğu görülmektedir ($t[142,933]=4,544$; $p=,000$).

Gerçekleştirilen dört ayrı çalışmanın genel bulgularına, geliştirilen müfredat önerisine ve On Birinci Kalkınma Planında YÖK’ün sorumlu kuruluş olduğu politika tedbirleri ve büyük veri ile ilgili geliştirilen önerilere bir sonraki bölümde değinilmiştir.

4. GENEL BULGULAR VE MÜFREDAT ÖNERİSİ

Bu bölümde gerçekleştirilen dört ayrı çalışmanın genel bulgularına, bu doğrultuda geliştirilen müfredat önerisine ve On Birinci Kalkınma Planında YÖK'ün sorumlu kuruluş olduğu politika tedbirleri ve büyük veri ile ilgili geliştirilen önerilere yer verilmiştir.

4.1. Yapılan Çalışmalardan Elde Edilen Genel Bulgular

Büyük veri çözümleri sunan kurumlarla yapılan görüşmelerden, Bilgisayar Mühendisliği öğrencilerine bir günlük büyük veri eğitimi sırasında alınan gözlem notları ve katılımcıların öz-değerlendirmelerinden, öğretim elemanlarının büyük veri beklentilerinin değerlendirilmesiyle ve öğrencilerin büyük veri farkındalık, yeterlilik ve beklentilerinin ölçülmesiyle elde edilen bulgulara yer verilmiştir. Gerçekleştirilen dört çalışmadan elde edilen özet bulgular Çizelge 4.1'de gösterilmiştir.

Bu özet bulgulara bakıldığında, kurumların kullandıkları araçlar ile öğretim elemanları ve öğrencilerin aşına oldukları araçlar birbiriyle örtüşmektedir. Kurumların, büyük veri alanında çalışanlara verdikleri pozisyon ismi genel olarak yazılım mühendisi olduğu görülmektedir. Kurumların %50'si büyük veri alanında çalışan sayısını yeterli görmemektedir. Buna karşılık, öğretim elemanlarının büyük veri alanına olan akademik ilgilerinin ve büyük veri ile ilgili ders verme oranlarının düşük olduğu görülmektedir. Öğrencilerin ise, büyük veri yazılım kullanımı ve öğrenimi sırasında karşılaştıkları sorunlar şunlardan oluşmaktadır; materyal, eğitim ve örnek uygulama eksikliği, lisans eğitimi dâhil çevrimiçi ve yüz yüze eğitimlerin yetersizliği, açık olarak paylaşılabilen veri ihtiyacı, donanım (laboratuvar) ihtiyacı, gerçek veriler üzerinde çalışamama.

Çizelge 4.1. Yapılan çalışmalardan elde edilen bulguların özeti

	Kurumlar	Öğretim Elemanları	Öğrenciler
Mevcut Durum	Kullanılan Araçlar: Apache Hadoop, Spark, Cassandra, Hive, Kafka, Elastic Search, HBase, HDFS, Impala, Kudu ve MongoDB Tercih edilen pozisyon adı: Yazılım Mühendisi Büyük Veri Alanında Çalışan Sayısı Yeterliliği: Kurumların %50'si yeterli görmemekte.	Kullanılan Araçlar: Apache Hadoop, Oracle Big Data, MS Hadoop, Spark ve HDFS Akademik İlgi: %15,5'inin büyük veri ile ilgili yayınlanmış, %17'sinin devam etmekte olan bilimsel makalesi mevcut. Ders Verme: %7,8'i büyük veri ile ilgili bir ders yürütüyor.	Kullanılan Araçlar: Oracle Big Data, Apache Hadoop, MS Hadoop, Spark, Hive ve Kafka Karşılaşılan Sorunlar: Materyal, eğitim ve örnek uygulama eksikliği, Lisans eğitimi dâhil çevrimiçi ve yüz yüze eğitimlerin yetersizliği, Açık olarak paylaşılabilen veri ihtiyacı, Donanım (laboratuvar) ihtiyacı, Gerçek veriler üzerinde çalışamama
Beklenti	Bölüm: Bilgisayar Mühendisi, Matematik, İstatistik, Yönetim Bilişim Sistemleri Öğrencilerden: Farkındalık, analitik düşünme yeteneği, veri tabanı (SQL), temel programlama, dosya sistemi (Linux)	Büyük veri konusunun lisans müfredatında daha fazla yer alması Büyük veri ile ilgili eğitim, seminer, konferans, çevrimiçi eğitim ve çalıştay gibi etkinliklere katılma Türkiye'de büyük veri çözümlerinin önümüzdeki yıllarda daha çok kullanılacağı	Laboratuvarında uygulamalı derslerin daha çok olması, Donanımsal kaynakların öğrencilere sunulması, Dersleri veren hocaların sektörden alanında uzman kişiler tarafından verilmesi, Kurumlarla açık veriye erişme konusunda anlaşmalar yapılması, Mezun olmadan önce büyük veri ile ilgili ders almak
Öneri	Öğrencilere: Farkındalığa sahip olma, Okulda öğretilenler dışında ekstra çaba gösterme, Yarı zamanlı çalışma (staj), Öğrenmeye açık olma, Çevrimiçi eğitimlere katılma, Seminer, kurs ve etkinliklere katılma Üniversitelere: Müfredat yenileme, Uzun dönem staj ve Yüksek lisans programları açma	Büyük veri üzerine seçmeli ve zorunlu dersler açılması Büyük veri konusunun müfredatta mevcut bir derse konularının eklenmesi Büyük veri ile ilgili daha çok lisansüstü program açılması	Büyük veri üzerine seçmeli ve zorunlu dersler açılması Büyük veri ile ilgili lisans programı açılması Büyük veri ile ilgili daha çok lisansüstü program açılması Uzun ve kısa dönem stajların arttırılması

Kurumlar, büyük veri alanında en çok sırayla Bilgisayar Mühendisliği, İstatistik, Matematik, Yönetim Bilişim Sistemleri ve Endüstri Mühendisliği bölümlerinden mezun olanları istihdam etmektedir. Kurumlar, öğrencilerden farkındalık ve analitik düşünme yeteneği gibi özellikler ve veri tabanı (SQL), temel programlama, dosya sistemi (Linux) gibi konularda bilgi sahibi olmalarını beklemektedir. Öğrencilere, farkındalığa sahip olma, okulda öğretilenler dışında ekstra çaba gösterme, yarı zamanlı çalışma (staj), öğrenmeye açık olma, çevrimiçi eğitimlere katılma ve seminer, kurs ve etkinliklere katılma gibi

önerilerde bulunmaktadır. Üniversitelere ise, müfredat yenileme, uzun dönem staj ve yüksek lisans programları açma önerilerinde bulunmuşlardır.

Öğretim elemanları, büyük veri konusunun lisans müfredatında daha fazla yer alması gerektiğini ve Türkiye’de büyük veri çözümlerinin önümüzdeki yıllarda daha çok kullanılacağını beklemektedir. Ayrıca büyük veri ile ilgili eğitim, seminer, konferans, çevrimiçi eğitim ve çalıştay gibi etkinliklere katılmak istedikleri ortaya çıkmıştır. Büyük veri üzerine seçmeli ve zorunlu dersler açılmasını, büyük veri konusunun müfredatta mevcut bir derse konularının eklenmesini ve büyük veri ile ilgili daha çok lisansüstü program açılmasını önermektedir.

Öğrenciler, laboratuvarında uygulamalı derslerin daha çok olmasını, donanımsal kaynakların öğrencilere sunulması, dersleri veren hocaların sektörden alanında uzman kişiler tarafından verilmesini, kurumlarla açık veriye erişme konusunda anlaşmalar yapılmasını, mezun olmadan önce büyük veri ile ilgili ders almak istediklerini konusunda beklentileri olduğu görülmektedir. Ayrıca, büyük veri üzerine seçmeli ve zorunlu dersler açılmasını, büyük veri alanı ile ilgili lisans programı açılmasını, büyük veri alanı ile ilgili daha çok lisansüstü program açılmasını, uzun ve kısa dönem stajların arttırılmasını önermektedir.

Genel bulgular ve literatür araştırmaları sonucunda, Türkiye’de ilk defa örnek teşkil edebilecek, büyük veri konularını da kapsayan Veri Bilimi alanında bir lisans programı müfredat önerisi hazırlanmıştır.

4.2. Müfredat Önerisi

Türkiye’de büyük veri ile ilgili doğrudan bir bölüm yoktur ancak veri bilimi ile ilgili tek lisans bölümü 2020-2021 eğitim öğretim yılında ilk öğrencilerini kabul eden İstanbul Teknik Üniversitesi (İTÜ)’nde yer almaktadır. İTÜ, Yapay Zeka ve Veri Mühendisliği bölümünü açarak, Türkiye’nin yapay zeka ve veri bilimini birleştiren ilk bölüm olduklarını belirtmiştir (İstanbul Teknik Üniversitesi, 2020). Büyük veri ile ilgili dersler içeren Yapay Zeka Mühendisliği bölümü ise ilk olarak, Hacettepe ve TOBB Ekonomi ve Teknoloji üniversitelerinde 2019-2020 eğitim öğretim yılında lisans düzeyinde açılmıştır. Ardından 2020-2021 eğitim öğretim yılında Bahçeşehir Üniversitesi de Yapay Zeka Mühendisliği

bölümünü açarak, toplamda Türkiye’de sadece üç üniversitede bu bölüm bulunmaktadır (YÖK Atlas, 2020).

Ankara ilinde, 5 devlet, 7 özel, toplam 12 üniversitede Bilgisayar Mühendisliği, Matematik, İstatistik ve Yönetim Bilişim Sistemleri bölümlerinde, sadece üç üniversitenin Bilgisayar Mühendisliği bölümlerinde (Hacettepe, TOBB ve Gazi Üniversitesi) büyük veri ile ilgili zorunlu dersler bulunmaktadır⁴¹. Hacettepe Üniversitesi Yapay Zeka Mühendisliği bölümünde, 4. yarıyıda Veri Biliminin Temelleri (Elements of Data Science) ve Veri Bilimi Laboratuvarı (Data Science Lab.) dersleri ve TOBB Üniversitesi Yapay Zeka Mühendisliği bölümünde, 2. yarıyıda Veri Bilimine Giriş dersi bulunmaktadır. Gazi Üniversitesi Bilgisayar Mühendisliği bölümünde, 5. yarıyıda Veri Bilimine Giriş (Introduction to Data Science) dersi bulunmaktadır. Teknik seçmeli olarak Bilkent ve Çankaya Üniversite’lerinin Bilgisayar Mühendisliği bölümlerinde Veri Bilimine Giriş dersi bulunmaktadır. Ankara ilindeki üniversitelerin Matematik bölümlerinde, yalnızca Bilkent Üniversitesi’nde 4. yılda seçmeli Veri Biliminin Matematiksel Temelleri (Mathematical Foundations of Data Science) dersi bulunmaktadır. İstatistik bölümlerinde ise Ankara Üniversitesinde Büyük Veri Analizi, Hacettepe Üniversitesinde ise Veri Bilimi seçmeli dersleri bulunmaktadır. Ankara ilinde Yönetim Bilişim Sistemleri bölümünde, yalnızca Başkent Üniversitesi’nde seçmeli Python Programlama ve Veri Bilimi dersi bulunmaktadır. Ankara ilindeki üniversitelerden yola çıkılarak, üniversitelerde büyük veri ile ilgili derslerin zorunlu veya seçmeli olarak birçok üniversitede olmadığı görülmektedir. Bu doğrultuda aşağıdaki üniversitelerin Veri Bilimi lisans programları dikkate alınarak Türkiye’de örnek teşkil etmesi açısından bir Veri Bilimi bölümü müfredat önerisi hazırlanmıştır (Yale Üniversitesi, 2019) (San Francisco Üniversitesi, 2020) (Worcester Polytechnic Enstitüsü, 2020) (Brigham Young Üniversitesi, 2020) (Warwick Üniversitesi, 2020) (Massey Üniversitesi, 2020) (Harbour Space Üniversitesi, 2020) (İstanbul Teknik Üniversitesi, 2020).

- Yale Üniversitesi (ABD) - Sosyal Bilimler Fakültesi- İstatistik ve Veri Bilimi Bölümü
- San Francisco Üniversitesi (ABD) - Fen Edebiyat Fakültesi - Veri Bilimi Bölümü

⁴¹ Ankara ilindeki üniversitelerin resmi web sayfalarında yer alan ders müfredatlarından edinilmiştir.

- Worcester Polytechnic Enstitüsü (ABD) - Fen Edebiyat Fakültesi- Veri Bilimi Bölümü
- Brigham Young Üniversitesi (ABD) -Fen ve Mühendislik Fakültesi-Veri Bilimi Bölümü
- Warwick Üniversitesi (İngiltere) – İstatistik: Veri Bilimi Bölümü
- Massey Üniversitesi (Yeni Zelanda) Fen Bilimleri Fakültesi -Veri Bilimi Bölümü
- Harbor.Space Üniversitesi (İspanya) -Veri Bilimi Bölümü
- İstanbul Teknik Üniversitesi (Türkiye) – Yapay Zekâ ve Veri Mühendisliği

Büyük veri yerine Veri Bilimi lisans programı önerisi hazırlanmasının sebebi, yurt dışındaki örneklere bakıldığında veri bilimi lisans programının, büyük veri konusunu da içine kapsayan daha genel bir bölüm olarak açılmasıdır. Veri Bilimi bölümünün yaygınlaşmasından sonra büyük veri özelinde bölümler açıldığı görülmüştür. Türkiye’deki mevcut lisans programları incelendiğinde öncelikle büyük veri konularını da kapsayan Veri Bilimi alanında lisans programlarının oluşturulması ardından büyük veri lisans programları ve alt programları oluşturulmasının daha gerçekleştirilebilir olacağı öngörülmüştür.

ABD’de Veri Bilimi bölümü, Bilgisayar Bilimleri, İstatistik ve Matematik bölümlerini kapsayan disiplinler arası bir bölüm olduğu için ağırlıklı olarak Fen veya Mühendislik Fakültelerinde açılması tercih edilmiştir. Ancak Sosyal Bilimler ve İşletme fakültelerinde de Veri Bilimi bölümü görülmektedir. Veri Bilimi bölümünün bulunduğu fakülteye bağlı olarak seçmeli ders alternatifleri değişmektedir. Bunun yanında sağlık ve işletme gibi alanlarda veri bilimi derslerinin seçenekleri üniversitenin potansiyeline göre çeşitlenmektedir (DataScience, 2019). Bazı üniversitelerde Veri Bilimi bölümü, Bilgisayar Bilimleri bölümünün alt programı olarak oluşturulmuş ve seçmeli derslerle desteklenmiştir. Bunun bir örneği Hacettepe Üniversitesinde görülmektedir. Bilgisayar Mühendisliği bölümünün altında Yapay Zeka Mühendisliği bölümü, temel ve seçmeli derslerle yeni bir bölüm olarak açılmıştır. ABD’deki Bilgisayar Bilimleri bölümü müfredatlar açısından incelendiğinde, Türkiye’deki Bilgisayar Mühendisliği bölümüne karşılık gelmektedir. Türkiye’deki genel yapıya bakıldığında Veri Bilimi bölümünün, Bilgisayar Mühendisliği bölümünün de bulunduğu Mühendislik Fakültelerinde, İstatistik ve Matematik bölümlerinden seçmeli derslerle desteklenen bir yapıda olması önerilebilir. Bu doğrultuda, yapılan çalışmalar, literatür incelemeleri ve dört ABD, bir İngiltere, bir Yeni Zelanda, bir

İspanya’da yer alan üniversitelerin Veri Bilimi bölümlerinin müfredatları ve Türkiye’de İstanbul Teknik Üniversitesinin Yapay Zeka ve Veri Mühendisliği bölümünün müfredatı dikkate alınarak, örnek bir Veri Bilimi bölümü müfredat önerisi hazırlanmıştır.

Çizelge 4.2. Veri bilimi müfredat önerisi

1.Yıl	
1. Yarıyıl	2. Yarıyıl
Veri Bilimine Giriş	R ile Veri Bilimi
Matematik I	Matematik II
Olasılık ve İstatistik	Veri Bilimi için İstatistik
Programlamaya Giriş (Python)	Nesne Tabanlı Programlama
Türk Dili I	Türk Dili II
Yabancı Dil I	Yabancı Dil II
2.Yıl	
3. Yarıyıl	4. Yarıyıl
Lineer Cebir	Ayrık Matematik
Veri Tabanına Giriş	Veri Tabanı Yönetim Sistemleri
İşletim Sistemleri	Veri Bilimi ve Etik
Veri Yapıları ve Algoritmalar	Veri Keşfi ve Analizi
Atatürk İlk. ve İnk. Tar. I	Atatürk İlk. ve İnk. Tar. II
Seçmeli Ders	Seçmeli Ders
3.Yıl	
5. Yarıyıl	6. Yarıyıl
Veri Madenciliği	Makine Öğrenmesi
Veri Görselleştirme	Veri Gizliliği ve Güvenliği
Büyük Veri Analitiği	Uygulamalı Veri Bilimi Metotları
Seçmeli Ders	Seçmeli Ders
Seçmeli Ders	Seçmeli Ders
4.Yıl	
7. Yarıyıl	8. Yarıyıl
Bitirme Projesi I	Bitirme Projesi II
Büyük Veri ve Map-Reduce	Paralel ve Dağıtık Hesaplama
Seçmeli Ders	Seçmeli Ders
Seçmeli Ders	Seçmeli Ders
Seçmeli Ders	Seçmeli Ders
Seçmeli Dersler	
Yapay Zeka	Uygulamalı İstatistik
Görüntü İşleme	Kariyer Planlama
Doğal Dil İşleme	Doğrusal Modeller
Derin Öğrenme	Optimizasyon Teknikleri
Bulanık Mantık	R/Pyhton ile Zaman Serileri
Metin Madenciliği	Veri Bilimi için Hesaplamalı Araçlar
Mantıksal Programlama	Veri Bilimi için Sinyal İşleme
Biyoenformatiğe Giriş	İleri İstatistik
Sağlık Alanında Yapay Öğrenme	Hesaplama ve Optimizasyon
Web Tasarımı ve Geliştirme	Bilgisayar Ağları

Geliştirilen müfredat önerisi haricinde, On Birinci Kalkınma Planında YÖK'ün sorumlu kuruluş olduğu politika tedbirleri incelenerek bu tedbirlere yönelik büyük veri ile ilgili geliştirilen önerilere yer verilmiştir.

4.3. On Birinci Kalkınma Planında YÖK'ün Sorumlu Kuruluş Olduğu Politika Tedbirleri ve Geliştirilen Büyük Veri İle İlgili Öneriler

Strateji Ve Bütçe Başkanlığı'nın yayınladığı son stratejik planda, büyük veri analizi konusunda tecrübeli uzman/veri bilimci kapasitesinin artırılması ve altyapının güçlendirilmesi bir ihtiyaç olarak belirtilirken üniversitelerin müfredatlarındaki eksiklikler ve bu alanla ilgili herhangi bir bölüm olmaması dikkat çekicidir.

Çizelge 4.3. 2019-2023 Stratejik Planındaki büyük veri ile ilgili hedef ve ihtiyaçlar (Strateji ve Bütçe Başkanlığı, 2018)

Hedef 4.2: 'Büyük veri temelli politika ve etki analizi çalışmaları yapılacaktır.'	
İhtiyaç 1: 'Büyük veri analizi konusunda tecrübeli uzman/veri bilimci kapasitesinin artırılması ve altyapının güçlendirilmesi'	İhtiyaç 2: 'Kamu politikalarının belirlenmesinde büyük veriye dayalı analizlerin kullanılması'

2019-2023 Stratejik Planında belirtilen hedef ve ihtiyaçlar dikkate alındığında, On Birinci Kalkınma Planı Kurumsal Sorumluluklar Tablosunda sorumlu kuruluş olarak Yükseköğretim Kurulunun belirlendiği, politika tedbirlerine yönelik büyük veri ve veri bilimi ile ilgili atılacak adımlar doğrultusunda geliştirilen öneriler Çizelge 4.4'te yer almaktadır.

332.4. politika tedbiri "Öncelikli sektörlerdeki firmaların Ar-Ge ve yenilik süreçlerinde yer alan insan kaynağı kapasitesini geliştirmeye yönelik olarak üniversite ve sanayi işbirliğinde lisansüstü programlar oluşturulacaktır." kapsamında, büyük veri ve veri bilimi alanının, Ar-Ge ve yenilik süreçlerinde yer alan insan kaynağı kapasitesini geliştirmeye yönelik alan olarak belirlenmesi önerilmiştir.

Çizelge 4.4. On Birinci Kalkınma Planı'nda YÖK'ün sorumlu kuruluş olduğu politika tedbirleri ve büyük veri ile ilgili geliştirilen öneriler

Politika Tedbirleri	Eylem Açıklaması	İşbirliği Yapılacak Kuruluş	Yapılacak İşlemler	Geliştirilen Öneriler
332.4.	Öncelikli sektörlerdeki firmaların Ar-Ge ve yenilik süreçlerinde yer alan insan kaynağı kapasitesini geliştirmeye yönelik olarak üniversite ve sanayi işbirliğinde lisansüstü programlar oluşturulacaktır.	TOBB, Üniversiteler, TÜBİTAK, Cumhurbaşkanlığı İnsan Kaynakları Ofisi	İlgili üniversite ve programların belirlenmesi	Büyük veri ve veri bilimi alanının, Ar-Ge ve yenilik süreçlerinde yer alan insan kaynağı kapasitesini geliştirmeye yönelik alan olarak belirlenmesi
332.5.	Öncelikli sektörlerle yönelik önlisans, lisans ve lisansüstü programların sayısı ve çeşitliliği artırılacak ve bu alandaki Ar-Ge faaliyetlerine üniversiteler nezdinde özel önem verilecektir.	Üniversiteler, TOBB, TÜBİTAK, Strateji ve Bütçe Başkanlığı, Sanayi ve Teknoloji Bakanlığı	1) Öncelikli sektörlerle yönelik önlisans, lisans ve lisansüstü programlar belirlenmesi 2) Öncelikli sektörlerin yerel sorunlarının ve gelişme alanlarının belirlenmesi ve üniversitelerin Ar-Ge faaliyetlerinin yönlendirilmesi	Büyük veri ve veri bilimi alanında, bilgi teknolojileri sektöründe oluşan personel ihtiyacı dolayısıyla, bu alanda önlisans, lisans ve lisansüstü programlarının açılarak, Ar-Ge faaliyetlerinin yönlendirilmesi
349.3.	Üniversite ve sanayi işbirliğinde, öncelikli sektörler başta olmak üzere sanayinin ihtiyacına yönelik lisansüstü programlar oluşturulacak, bu programları açan üniversiteler teşvik edilecektir	TÜBİTAK	1) Sanayinin ihtiyacına yönelik lisansüstü programların oluşturulması (YÖK) 2) TÜBİTAK 2244 Sanayi Doktora Programı kapsamında sanayinin ihtiyaç duyduğu ihtisas alanlarına yönelik bilgiler YÖK ile paylaşılacaktır. (TÜBİTAK)	Büyük veri ve veri bilimi alanlarında üniversite ile kamu ve özel kurumlar arasında yapılan anlaşmalar ile lisansüstü programlar ve eğitimler oluşturulması
356.2.	Bu yetenek ve yetkinliklerin karşılanmasına yönelik kritik teknolojilerde ihtisaslaşmış disiplinler arası lisans ve lisansüstü programlar açılacaktır.	Üniversiteler	1) Kritik teknoloji alanlarında ihtiyaç duyulan yetenek ve yetkinlikler belirlenecektir. (YÖK, TÜBİTAK) 2) Gerekli yetkinliklerin karşılanmasına yönelik üniversiteler tarafından disiplinlerarası programlar açılacaktır. (YÖK)	Analitik yetenek gerektiren disiplinler arası büyük veri ve veri bilimi alanında lisans ve lisansüstü programlarının açılması

Çizelge 4.5. (devam) On Birinci Kalkınma Planı'nda YÖK'ün sorumlu kuruluş olduğu politika tedbirleri ve büyük veri ile ilgili geliştirilen öneriler

Politika Tedbirleri	Eylem Açıklaması	İşbirliği Yapılacak Kuruluş	Yapılacak İşlemler	Geliştirilen Öneriler
561.5	Yükseköğretim kurumlarının kontenjanları, sektörel ve bölgesel beceri ihtiyaçları, üniversitelerin kapasiteleri, arz-talep dengesi ve mevcuttaki programların asgari doluluk oranları dikkate alınarak belirlenecek; eğitim-istihdam bağlantısı güçlendirilecektir.	Strateji ve Bütçe Başkanlığı, Cumhurbaşkanlığı İnsan Kaynakları Ofisi, Hazine ve Maliye Bakanlığı, TÜBİTAK, Sanayi ve Teknoloji Bakanlığı, Üniversiteler Yükseköğretim Programları Danışma Kurulu	Yükseköğretim kurumlarının kontenjanlarının Yükseköğretim Programları Danışma Kurulu önerileri doğrultusunda belirlenmesi ve arz talep dengesinin sağlanması	Sektörün derin analitik yeteneğe sahip becerilere olan talebe, büyük veri ve veri bilimi lisans ve lisansüstü programlarının açılması ile arz edilmesi
561.7	Mezunların kariyer süreçlerinin takibi ve mezun-üniversite işbirliğinin güçlendirilmesi için mezun izleme sistemi kurulacaktır.	Üniversiteler, Cumhurbaşkanlığı İnsan Kaynakları Ofisi	Tüm üniversitelerde mezun izleme sisteminin kurulmasının teşvik edilmesi	Mezun izleme sistemi üzerinden elde edilen verilerin analizini gerçekleştirecek olan yeteneklere sahip bölümlerin açılması
563.1	Etkili tanıtım çalışmalarıyla yükseköğretim sistemine uluslararası erişim kolaylaştırılacaktır.	Üniversiteler	1) "Study in Turkey" markası altında hizmet veren internet sitesinin yeni web teknolojileriyle yeniden tasarlanarak daha etkileşimli bir yapıda hizmet vermesinin sağlanması 2) Stratejik planın uygulanması 3) Uluslararası fuarlara katılım ve tanıtım etkinliklerinin raporlanması ve bu raporlara göre gerekli tedbirlerin alınması	Başta ABD olmak üzere Veri Bilimi ve Büyük Veri lisans programlarına olan talebin artması ile Türkiye'de uluslararasılaşmayı sağlayacak bölümlerin açılması

5. SONUÇ VE ÖNERİLER

Bu bölümde, gerçekleştirilen dört çalışmanın bulgularına dayalı olarak ulaşılan sonuçlara, önerilere ve gelecek çalışmalara yönelik önerilere yer verilmiştir. Birinci çalışmada büyük veri çözümleri sunan kurumlarla görüşmelerden elde edilen sonuçlara, ikinci çalışmada Bilgisayar Mühendisliği öğrencilerine bir günlük büyük veri eğitimi sırasında alınan gözlem notları ve katılımcıların öz-değerlendirmelerinden elde edilen sonuçlara, üçüncü çalışmada öğretim elemanlarının büyük veri beklentilerinin değerlendirilmesiyle elde edilen sonuçlara ve dördüncü çalışmada öğrencilerin büyük veri farkındalık, yeterlilik ve beklentilerinin ölçülmesiyle elde edilen sonuçlara yer verilmiştir.

Birinci çalışmada elde edilen bulgular neticesinde, büyük veri çözümleri sunan kurumların mevcut durumlarıyla alakalı olarak büyük veri projelerinde çalışanların hangi pozisyonda çalıştıkları, büyük veri projelerinde çalışan sayılarının yeterliliği ve büyük veri projelerinde kullanılan araçlar ve teknolojiler belirlenmiştir. 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 veya yazılım geliştirici olarak tercih ettikleri sonucuna varılmıştır. Kurumların yarısı büyük veri projelerinde çalışan sayılarını yeterli bulduklarını belirtmiş ancak bunun nedeni kurumların proje bazlı hizmet vermeleri ve çalışan sayısına uygun oranda yeni proje aldıkları sonucu da ortaya çıkmıştır. Kurumların kullandıkları teknolojiler, literatür araştırmalarında belirtilen dünya genelinde kullanılan teknolojilerle benzer olduğu ortaya çıkmıştır. Açık kaynaklı yazılımların tercih edildiği, en çok Hadoop ve ekosisteminin kullanıldığı, Spark, Hive, Kafka, Cassandra gibi araçların da en çok tercih edilen Hadoop araçları olduğu sonucuna varılmıştır.

Kurumların bu alanda çalışmayı düşünen kişileri en çok Bilgisayar Mühendisliği, ardından Matematik, İstatistik, Yönetim Bilişim Sistemleri ve Endüstri Mühendisliği bölümlerinden tercih ettikleri ortaya çıkmıştır. Kurumlar mezunlardan en çok alanla ilgili farkındalık beklemektedir, teknik olarak ise veri tabanı ve temel programlama bilgisinin olmasını beklemektedir. 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. Dördüncü çalışmanın temelini oluşturan öğrencilerin büyük veri farkındalığı, aynı zamanda kurumların mezunlarda en çok aradığı özellik olmuştur.

İkinci çalışmada, büyük veri alanına ilgili olan öğrencilerin büyük veri deneyimleri ve karşılaştıkları zorluklar belirlenmiştir. Katılımcıların çoğu büyük veri araçlarından bazılarının kurulumunu yapıp, deneme fırsatı bulmuştur ve katılımcılar tarafından, Hadoop ekosisteminde en çok bilinen araçlar sırayla Spark, HDFS, Hive, Yarn, Cassandra ve HBase olduğu ortaya çıkmıştır. Bu araçlar, birinci çalışmada belirlenen, kurumların en çok kullandığı büyük veri yazılımları ile örtüşmektedir. Katılımcıların büyük veri araçlarını kullanırken en çok karşılaşılan zorluklar, kaynak, eğitim ve örnek uygulama yetersizliği olduğu sonucuna varılmıştır. Donanım (laboratuvar) ve açık erişimli verilerin yetersizliği (gerçek veride çalışamama) de diğer karşılaşılan zorluklar olmuştur. Katılımcıların beklentileri, seçmeli ders sayısının artırılması, bu alanda bölüm açılması, uygulamalı derslerin ve proje temelli derslerin daha çok olması, donanımsal kaynakların öğrencilere sunulması, dersleri veren hocaların sektörden alanında uzman kişiler tarafından verilmesi, kurumlarla açık veriye erişme konusunda anlaşmalar yapılması gibi konuları kapsamaktadır.

Üçüncü çalışmada, öğretim elemanlarının büyük veri alanı ile ilgili farkındalık ve eğitimle alakalı beklentileri değerlendirilmiştir. Öğretim elemanları tarafından en çok aşına olunan büyük veri yazılımları Oracle Big Data, Apache Hadoop ve Microsoft Hadoop, en çok kullanılan yazılımlar sırayla Apache Hadoop, Oracle Big Data, HDFS ve Spark olmuştur. Kullanılan yazılım sayılarının oldukça düşük olduğu sonucuna varılmıştır. Büyük veri alanına olan akademik ilginin arttığı, halen üzerinde çalışılmakta olunan proje sayısının, tamamlanmış proje sayısının iki katından fazla olduğu sonucuna varılmıştır. Katılımcılar arasında büyük veri ile ilgili derslerin yürütülme oranının (%8) çok az olduğu anlaşılmıştır. Öğretim elemanları, Türkiye’de büyük veri çözümlerinin önümüzdeki yıllarda daha çok kullanılacağı ve Bilgisayar Mühendisliği ve ilgili bölümlerde büyük veri üzerine daha çok seçmeli ders açılması gerektiği sonucuna varmıştır. Katılımcılar, büyük veri ile ilgili üniversitelerde daha çok lisansüstü program açılması gerektiğini düşünmektedir.

Son çalışmada, öğrencilerin büyük veri alanı ile ilgili farkındalık, yeterlilik ve eğitimle alakalı beklentileri belirlenmiştir. Bu doğrultuda öğrencilerin büyük veri farkındalık ve yeterlilik ölçeği ile büyük veri eğitim ve beklenti ölçeği, öğrencilerin büyük veri konusundaki yeterliliğini ve beklentilerini farkındalık ile beraber ölçmeye yönelik olarak ilk defa bu çalışma kapsamında geliştirilmiştir. Öğrencilerin çok büyük bir kısmının daha önce büyük veri kavramıyla karşılaştığı görülmektedir. Ankete katılan öğrenciler tarafından aşına olunan yazılımlar, öğretim elemanları tarafından en çok tanınan yazılımlar ile benzerdir. Öğretim elemanları ile aynı doğrultuda öğrenciler de, büyük veri konusunun lisans müfredatında daha fazla yer alması gerektiğini düşünmektedir. Katılımcılar, Bilgisayar Mühendisliği ve ilgili bölümlerde büyük veri üzerine seçmeli ve zorunlu derslerin açılmasını istemektedir. Katılımcıların mezun olmadan önce büyük veri ile ilgili ders almak istedikleri sonucuna ulaşılmıştır. Öğrencilerin büyük veri eğitim ve beklentileri ölçeğinde kullanılan maddelerin altı alt boyuta dağıldığı görülmektedir. Bu alt boyutlar; kariyer ve lisans dışı eğitim, lisans eğitimi, yurtdışı ve staj, bilgi alışverişi, Türkiye’de büyük veriye ilgi ve Türkiye’de fırsatlardır. Yönetim Bilişim Sistemleri öğrencilerinin Bilgisayar Mühendisliği öğrencilerine kıyasla; kariyer ve lisans dışı eğitim, lisans eğitimi, yurt dışı ve staj, bilgi alışverişi alt boyutlarında farkındalık ve beklentilerinin anlamlı bir şekilde yüksek olduğu sonucuna varılmıştır. Ancak her iki bölümde de Türkiye’deki büyük veriye olan ilginin düşük olduğu ve Türkiye’deki fırsatların az olduğu konusunda görüşlerinin farklı olmadığı görülmektedir. Yönetim Bilişim Sistemleri öğrencilerinin, Bilgisayar Mühendisliği öğrencilerine göre farkındalık ve yeterliliklerinin anlamlı bir şekilde yüksek olduğu sonucuna ulaşılmıştır.

Yapılan dört ayrı çalışma, literatür araştırmaları, stratejik planlardaki ihtiyaçlar ve On Birinci Kalkınma Planında YÖK’ün sorumlu kuruluş olduğu büyük veri ile ilgili politika tedbirleri dikkate alındığında şu öneriler ortaya çıkmaktadır.

1. Analitik yetenek ve yetkinliklerin karşılanmasına yönelik kritik teknolojilerde ihtisaslaşmış disiplinler arası lisans ve lisansüstü programların açılması
 - 1.1 Büyük veri ve veri bilimi alanlarında lisansüstü programların açılması ve arttırılması
 - 1.2 Büyük veri ve veri bilimi alanlarında lisans programlarının açılması
2. Sektörün nitelikli personel ihtiyacını karşılayacak alanlarda ve donanımda, üniversitelerde öğrencilere imkânlar sunmak

- 2.1 Büyük veri ve veri bilimi arařtırmaları için arařtırma merkezleri ve laboratuvarlar kurulması
- 2.2 Öğrencilerin gelişimlerine destek olan her türlü donanım, yazılım ve diğer (laboratuvar-açık veri erişimi) gereksinimlerinin karşılanması
- 2.3 Veri bilimi alanında, eğitim, seminer, konferans ve çalıştayların artırılması
3. Bilişim alanında lisans programları ve müfredatların güncellenmesi
 - 3.1 Büyük veri/veri bilimi alanlarının bilgisayar bilimleri ve ilgili bölümlerin lisans müfredatında daha fazla yer alması
 - 3.2 Büyük veri ile ilgili seçmeli ve zorunlu derslerin açılması ve artırılması
 - 3.3 Büyük veri ile ilgili temel konuların mevcut müfredattaki derslere eklenmesi
4. Eğitim-öğretim kalitesini arttırmak, uluslararasılaşmayı yaygınlaştırmak ve sektörün beklentilerini dikkate almak
 - 4.1 Uluslararası rekabete açık arařtırma geliştirme çalışmaları için veri bilimi/büyük veri laboratuvar ve arařtırma merkezlerinin fiziki alt yapılarının iyileştirilmesi
 - 4.2 Uluslararası düzeyde başarılı olan üniversitelerin veri bilimi/büyük veri alanlarında eğitim-öğretim programlarının takip edilmesi
 - 4.3 Üniversite-sanayi işbirliğini güçlendirecek, çıktısı ekonomik değere dönüştürülebilir büyük veri girişim projelerinin sayısının artırılması
5. Üniversite ve sanayi işbirliğinde, öncelikli sektörler başta olmak üzere sanayinin ihtiyacına yönelik lisansüstü programlar açılması
 - 5.1 Üniversiteler ile özel ve kamu kurumları arasında yapılan anlaşmalar ile büyük veri/veri bilimi alanlarında lisansüstü programların açılması
 - 5.2 Üniversiteler ile özel ve kamu kurumları arasında yapılan anlaşmalar ile büyük veri/veri bilimi alanlarında eğitim ve staj programların oluşturulması

Yapılan çalışmalarda, arařtırmaların kısıtlılıklarında belirtildiği gibi örnekleme bağı olarak, ileride yapılacak çalışmalarda bu kısıtlar göz önünde bulundurularak farklı örneklemeler üzerinde çalışılabilir. Büyük veri çözümleri sunan kurumlarla yapılan görüşmeler için, İstanbul, İzmir gibi özel kurumların yoğun olduğu iller tercih edilebilir, örneklem sektör bazında (telekomünikasyon, sağlık, eğitim, iletişim vb.) alt gruplara ayrılabilir. Büyük veri ve veri bilimi alanı disiplinler arası bir alan olduğu için, sadece

bilgisayar bilimleri ve ilgili bölümlerde değil, işletme, sağlık ve diğer alanlarda da öğrencilerin mezun olmadan önce yetenek ve beceri sahibi olmak istedikleri bir alan olabileceği için, bu alanlarda benzer çalışmalar yapılabilir.



KAYNAKLAR

- ACM and IEEE. (2005). Computing Curricula 2005, The Overview Report. *The Association for Computing Machinery (ACM) and The Computer Society (IEEE-CS)*, 1-56.
- ACM and IEEE. (2013). Computer Science Curricula 2013, Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. *ACM and IEEE CS*, 1-514.
- ACM and IEEE. (2016). Computer Engineering Curricula 2016, Curriculum Guidelines for Undergraduate Degree Programs in Computer Engineering, *ACM and IEEE CS*, 1-149.
- ACM and IEEE. (2020). Computing Curricula 2020, A Computing Curricula Series Report Version 44. *ACM and IEEE CS*, 1-207.
- Aktan, E. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu. *Ankara Üniversitesi Bil-Bem Bilgi Yönetimi Dergisi*, 1(1), 1-22.
- Bange, C., Grosser, T., and Janoschek, N. (2013). Big Data Survey Europe, Usage, Technology and Budgets in European Best-Practice Companies. *BARC Institute*, Wuerzburg, Germany, 1-46.
- Bayrakçı, S., ve Albayrak, M. A. Büyük Verinin Akademik Çalışmalarda Kullanımı Üzerine Mukayeseli Bir Veri Tabanı Araştırması. *Academic Journal of Information Technology-e: Bilişim Teknolojileri Online Dergisi*, 10(36), 73-94.
- Berg, B. L., and Lune, H. (2019). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri (9. Baskıdan Çeviri)*. (Çev. Asım Arı), Konya: Eğitim Yayınevi, (Eserin orijinali 2016'da yayımlandı), 67.
- Brahm, C., Sheth, A., Sinha, V., and Dai, J. (2019). Solving the New Equation for Advanced Analytics Talent. *Bain And Company*, 1-9.
- Columbus, L. (2018). 10 Charts That Will Change Your Perspective Of Big Data's Growth. (2018, May 23), *Forbes*.
- Cox, M., and Ellsworth, D. (1997, October). Application-controlled demand paging for out-of-core visualization. *Proceedings. Visualization'97 (Cat. No. 97CB36155)*, Phoenix, AZ, USA, 235-244.
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative*. Upper Saddle River, NJ: Prentice Hall, 146-166.
- Cukier, K. (2010). Data, data everywhere. (2010, Feb 27), *The Economist*.

- Çokluk, Ö., Şekercioğlu, G., ve Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (Vol. 2). Ankara: Pegem Akademi, 190-205.
- Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920.
- Daniel, B. K., and Butson, R. (Nov 29-Dec 1, 2013). *Technology Enhanced Analytics (TEA) in Higher Education*. International Association for the Development of the Information Society, Paper presented at the International Conference on Educational Technologies (ICEduTech), Kuala Lumpur, Malaysia, 89-96.
- Davenport, T., and Patil, D., Cohen T. (2013). Veri Bilimcilik: 21.Yüzyılın En Çekici Mesleği. (Ekim 2012), *Harvard Business Review Türkiye*.
- Davis, K., and Patterson, D. (2012). *Ethics of Big Data: Balancing risk and innovation*. O'Reilly Media., 4.
- Dean, J., and Ghemawat, S. (2004). *MapReduce: Simplified Data Processing on Large Clusters*. Paper presented at Sixth Symposium on Operating System Design and Implementation, San Francisco, CA., 137-150.
- De Mauro, A., Greco, M., and Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. *AIP conference proceedings, American Institute of Physics*, 1644(1), 97-104.
- Demchenko, Y., Grosso, P., De Laat, C., and Membrey, P. (2013, May 20-24). *Addressing big data issues in scientific data infrastructure*. Paper presented at 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 48-55.
- Desjardins, J. (2019). This is what Happens in a minute on internet. (2019, Mar 15), *World Economic Forum and Visual Capitalist*.
- Devakunchari, R. (2014). Analysis on big data over the years. *International Journal of Scientific and Research Publications*, 4(1), 1.
- DeVellis, D. R. F. (2003). *Scale Development: Theory and Applications Second Edition*, Sage Publications, 1-113.
- Dontha, R. (2017). Who came up with the name Big Data. *Data Science Central*, 13.
- Fidan, G. (2015). *Büyük Veri de Türkiye'den Uygulama Örnekleri*. Argedor Bilişim Teknolojileri, 1-24.

- Friedman, U. (2012). Big Data: A Short History. (2012, Oct 8), *Foreign Policy*.
- Gahi, Y., Guennoun, M., and Mouftah, H. T. (2016, June). *Big data analytics: Security and privacy challenges*. Paper presented at 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, Italy, 952-957.
- Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Gantz, J. F., Reinsel, D., Chute, C., Schlichting, W., McArthur, J., Minton, S., ... and Manfrediz, A. (2007, Mar). The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010. *An IDC White Paper Sponsored by EMC*, 1-20.
- Gantz, J., and Reinsel, D. (2010, May). The digital universe decade, Are you ready? *IDC iView: IDC Analyze the future* 1-16.
- Gantz, J., and Reinsel, D. (2012, Dec). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 1-16.
- General Electric and Accenture, (2014). Industrial Internet Insights Report for 2015. *General Electric and Accenture*, 1-36.
- Glaser, B. G., and Strauss, A. L. (1967). The discovery of grounded theory: strategies for qualitative research. *Zugriff unter* http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser_1967.pdf.
- Gökşen, Y., ve Aşan H. (2015). Veri büyüklüklerinin veritabanı yönetim sistemlerinde meydana getirdiği değişim: NOSQL. *International Journal of Informatics Technologies*, 8(3), 125.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. (1998). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice hall, 5(3), 207-219.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, 47, 98-115.
- Hilbert, M. (2012). Info Capacity| Introduction—How to Measure “How Much Information”?. *International Journal of Communication*, 6, 14.
- Huda, M., Anshari, M., Almunawar, M. N., Shahrill, M., Tan, A., Jaidin, J. H., ... and Masri, M. (2016). Innovative teaching in higher education: the big data

approach. *The Turkish Online Journal of Educational Technology*, 15(Special Issue), 1210-1216.

Hughes, O. E. (2012). *Public management and administration: An introduction*. Macmillan International Higher Education, 1-384.

Hulme, G. (2014, May). Start class: How crunching numbers can drive decisions. CSO.

İnternet: 365 DataScience. *How to Become a Data Scientist in 2020 – Top Skills, Education, and Experience*. URL: <https://365datascience.com/become-data-scientist-2020/> Son Erişim Tarihi: 29.02.2020

İnternet: Agrawal, V. (2019). *Big data in a nutshell*. URL: <https://jaxenter.com/big-data-nutshell-159112.html> Son Erişim Tarihi: 07.08.2019

İnternet: Apache Cassandra. (2016). *What is Cassandra?*. URL: <http://cassandra.apache.org> Son Erişim Tarihi: 24.12.2017

İnternet: Apache Hadoop. (2013, 08 04). *MapReduce Tutorial*. URL: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html Son Erişim Tarihi: 24.12.2017

İnternet: Apache Hbase. (2016, 12 07). *Welcome to apache HBase*. URL: <http://hbase.apache.org> Son Erişim Tarihi: 24.12.2017

İnternet: Apache Mahout. (2016). *Apache Mahout*. URL: <http://mahout.apache.org> Son Erişim Tarihi: 24.12.2017

İnternet: Apache Pig. (2018). *Welcome to Apache Pig*. URL: <https://pig.apache.org/> Son Erişim Tarihi: 24.12.2019

İnternet: BachelorsPortal. (2020). *Bachelor Degrees*. URL: <https://www.bachelorsportal.com/search/#q=di-282|lv-bachelor|tc-EUR|de-fulltime|mh-face2face|ci-82> Son Erişim Tarihi: 12.03.2020

İnternet: BachelorsPortal. (20201). *Bachelor Degrees*. URL: <https://www.bachelorsportal.com/search/#q=di-282|lv-bachelor> Son Erişim Tarihi: 12.01.2021

İnternet: Brigham Young Üniversitesi (2020). *Statistics: Data Science. Program Requirements*. URL: <https://catalog.byu.edu/physical-and-mathematical-sciences/statistics/statistics-data-science-bs> Son Erişim Tarihi: 14.03.2020

İnternet: Columbia Üniversitesi. (2020). *Big Data Analytics and Advanced Big Data Analytics*. URL: <https://www.ee.columbia.edu/~cylin/course/bigdata/> Son Erişim Tarihi: 14.03.2020

İnternet: Data.gov. (2019). *About Data.gov*. URL: <https://www.data.gov/about> Son Erişim Tarihi: 13.06.2019

İnternet: Data.gov.uk. (2019). *Find Open Data*. URL: <https://data.gov.uk/about> Son Erişim Tarihi: 13.06.2019

İnternet: Data.govt.nz. (2019). *Data.govt.nz*. URL: <https://data.govt.nz/about/about-data-govt-nz/> Son Erişim Tarihi: 27.06.2020

İnternet: DataScience. (2019). *Data Science Colleges and Universities*. URL: <http://datascience.community/colleges> Son Erişim Tarihi: 14.03.2020

İnternet: DB-Engines. (2020). *DB-Engines Ranking- Trend Popularity* URL: https://db-engines.com/en/ranking_trend Son Erişim Tarihi: 06.05.2020

İnternet: 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

İnternet: DigitalAge. (2019). *Türkiye’de Büyük Veri pazarı, 2023’te 520 milyon dolara ulaşacak*. URL: <https://digitalage.com.tr/turkiyede-buyuk-veri-pazari-2023te-520-milyon-dolara-ulasacak/> Son Erişim Tarihi: 06.01.2020

İnternet: Harbour Space Üniversitesi. (2020). *Data Science Programme Curriculum*. URL: <https://harbour.space/data-science?curriculum-Bachelor> Son Erişim Tarihi: 14.03.2020

İnternet: Harcourt, R. (2018). *Celebrating Berkley’s First Data Science Majors*. URL: <https://data.berkeley.edu/news/celebrating-berkeley-first-data-science-majors> Son Erişim Tarihi: 25.06.2020

İnternet: IDC. (2019, 4 4). URL: <https://www.idc.com/getdoc.jsp?containerId=prUS44998419> Son Erişim Tarihi: 23.06.2019

İnternet: İstanbul Teknik Üniversitesi. (2020). *Yapay Zeka ve Veri Mühendisliği Ders Planı*. URL: <http://www.sis.itu.edu.tr/tr/dersplan/plan/YZVE/000000.html> Son Erişim Tarihi: 25.06.2020

İnternet: İstanbul Teknik Üniversitesi. (2020). *Hakkımızda*. URL: <https://yapayzeka.itu.edu.tr/hakkimizda> Son Erişim Tarihi: 25.06.2020

İnternet: Liu, S. (2019). *Big data market size revenue forecast worldwide from 2011 to 2027*. URL: <https://www.statista.com/statistics/254266/global-big-data-market-forecast/> Son Erişim Tarihi: 11.10.2019

İnternet: Massey Üniversitesi (2020). Bachelor of Information Sciences (Data Science)/Couses and Planning. URL: https://www.massey.ac.nz/massey/learning/programme-course/programme.cfm?prog_id=93068&major_code=PDTSC&study_year=2021# Son Erişim Tarihi: 14.03.2020

İnternet: McDonald, C. (2017). *Big Data Opportunities for Telecommunications*. URL: <https://mapr.com/blog/big-data-opportunities-telecommunications/> Son Erişim Tarihi: 17.10.2019

İnternet: National Science Foundation. (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

İnternet: NITRD. (2020). *Big Data Interagency working Group*. URL: <https://www.nitrd.gov/nitrdgroups/index.php?title=Big-Data> Son Erişim Tarihi: 05.07.2020

İnternet: Open.canada.ca. (2019). *Open Data 101*. URL: <https://open.canada.ca/en/open-data-principles#toc99> Son Erişim Tarihi: 13.06.2019

İnternet: San Francisco Üniversitesi. (2020). *Data Science Major*. URL: https://catalog.usfca.edu/preview_program.php?catoid=22&poid=13141 Son Erişim Tarihi: 14.03.2020

İnternet: Savvycom. (2016, 01 21). *What you need to know about Hadoop and Its Ecosystem*. URL: <https://savvycomsoftware.com/what-you-need-to-know-about-hadoop-and-its-ecosystem/> Son Erişim Tarihi: 27.06.2018

İnternet: ScaleGrid. (2019, 03 04). *2019 Database Trends*. URL: <https://scalegrid.io/blog/2019-database-trends-sql-vs-nosql-top-databases-single-vs-multiple-database-use/> Son Erişim Tarihi: 27.06.2020

İnternet: Smartcon. (2016). *Türkiye Analitik Yetenek ve Yetkinlik Araştırması*. İstanbul: Smartcon. URL: <https://docplayer.biz.tr/30423591-Istanbul-turkiye-analitik-yetenek-ve-yetkinlik-arastirmasi.html>

İnternet:Türk Dil Kurumu. (2019). *Türk Dil Kurumu Sözlükleri* URL: <https://sozluk.gov.tr> Son Erişim Tarihi: 27.12.2019

İnternet: Warwick Üniversitesi. (2020). *Data Science Handbook Contents*. URL: <https://warwick.ac.uk/fac/sci/statistics/currentstudents/handbooks/datascience/courses/> Son Erişim Tarihi: 14.03.2020

İnternet: Worcester Polytechnic Enstitüsü. (2020). *Data Science: Degrees and Courses*. URL: <https://www.wpi.edu/academics/departments/data-science/courses> Son Erişim Tarihi: 14.03.2020

İnternet: 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

İnternet: Yale Üniversitesi. (2019, 12 20). *Department of Statistics and Data Science: Course List*. URL: <http://www.stat.yale.edu/Courses/2019-20/> Son Erişim Tarihi: 14.03.2020

İnternet: YÖK. (2019). *100/2000 YÖK Doktora Bursu*. URL: <https://yuzukibinbursu.yok.gov.tr/Sayfalar/HaberDuyuru/100-2000-yok-doktora-bursu-2019-2020-guz.aspx> Son Erişim Tarihi: 14.01.2020

İnternet: YÖK. (2020). *YÖK Gelecek Projesi*. URL: <https://www.yok.gov.tr/Sayfalar/Haberler/2020/yok-gelecek-projesi-nin-ikinci-asamasina-iliskin-beyanat.aspx> Son Erişim Tarihi: 14.01.2020

İnternet: 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

İnternet: YÖK Atlas. (2020). *YÖK Lisans Atlası*. URL: <https://yokatlas.yok.gov.tr/lisans-anasayfa.php> Son Erişim Tarihi: 10.11.2020

Jones, S. J. (2012). Technology review: the possibilities of learning analytics to improve learner-centered decision-making. *Community College Enterprise*, 18(1), 89-93.

Jyothirmayee, A. V. N. S., Reddy, D. G., and Akbar, K. (2014). Understanding Big Data & DV2 law. *International Journal of Emerging Technology and Advanced Engineering*, 4(7), 688-695.

Kaisler, S., Armour, F., Espinosa, J. A., and Money, W. (2013, January). *Big data: Issues and challenges moving forward*. Paper presented at 2013 46th Hawaii International Conference on System Sciences, Wailea, Maui, HI, USA, 995-1004.

Kemp, S. (2019). *Global Digital Report 2019-We Are Social*. We Are Social, London, 1-70.

- Kim, G. H., Trimi, S., and Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, 57(3), 78-85.
- Krejcie, R. V., and Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, 30(3), 607-610.
- Kurtuluş, K. (2010). *Araştırma Yöntemleri*. Türkmen Kitabevi. İstanbul, 181.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.
- Laney, D. (2012). Deja VVVu: others claiming Gartner's construct for big data. *Gartner Blog*, 14.
- Lawler, J., and Molluzzo, J. C. (2015). A proposed concentration curriculum design for big data analytics for information systems students. *Information Systems Education Journal*, 13(1), 45.
- LeCompte, M. D., and Preissle, J. (1993). *Ethnography and Qualitative Design in Educational Research*. San Diego, CA: Academic. Language, 425.
- Liyakasa, K. (2013, 6). Executive Demand for Data Analysis Grows. *Customer Relationship Management*, 20-21.
- Lyman, P., and Varian, H. R. (2000). Reprint: How Much Information?. *Journal of Electronic Publishing*, 6(2), 1-7.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. 1-143.
- María Cavanillas, J., Curry, E., and Wahlster, W. (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer Nature, 1-303.
- Marr, B. (2016). *Big Data in Practice, How 45 Successful Companies Used Big Data Analytics To Deliver Extraordinary Results*. Wiley, 1-308.
- Marshall, C., and Rossman, G. B. (2014). *Designing qualitative research*. Sage publications, 1-352.
- Marzullo, K. (2016). Administration Issues Strategic Plan For Big Data Research and Developement. (May 23, 2016), *The White House President Barack Obama*.

- Marz, N., and Warren, J. (2015). *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co, 1-28.
- Miller, S., and Hughes, D. (2017). The quant crunch: How the demand for data science skills is disrupting the job market. *Burning Glass Technologies*.
- Mayring, P. (2011). *Nitel Sosyal Araştırmaya Giriş. Nitel Düşünce İçin Bir Rehber*.(Çev: A, Gümüş ve S. Durgun). Ankara: Bilgesu Yayınevi.
- MEB. (2018). *Milli Eğitim Bakanlığı 2019-2023 Stratejik Plan Hazırlık Programı*. Milli Eğitim Bakanlığı, 1-12.
- Memon, M. A., Soomro, S., Jumani, A. K., and Kartio, M. A. (2017). Big data analytics and its applications. *Annals of Emerging Technologies in Computing*, 1(1), 45-54.
- Merriam, S. B. (2013). *Nitel araştırma desen ve uygulama için bir rehber* (Çev. S. Turan) Ankara: Nobel Yayın Dağıtım.
- Miles, M. B., and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Miner, D., and Shook, A. (2012). *MapReduce design patterns: building effective algorithms and analytics for Hadoop and other systems*. O'Reilly Media, Inc.
- Morris, R. J., and Truskowski, B. J. (2003). The evolution of storage systems. *IBM systems Journal*, 42(2), 205-217.
- Naik, K., and Joshi, A. (2017, February). *Role of Big Data in various sectors*. Paper presented at 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 117-122.
- Narasimhan, R., and Bhuvaneshwari, T. (2014). Big data—a brief study. *International Journal of Scientific and Engineering Research*, 5(9), 350-353.
- Okur, M. C., and Büyükkeçeci, M. (2014, May). *Big data challenges in information engineering curriculum*. Paper presented at 2014 25th EAEEIE Annual Conference, Institute of Electrical and Electronics Engineers, 1-4.
- ÖSYM. (2018). Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı 2019 - 2023 Stratejik Planı. ÖSYM Başkanlığı, 1-62.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation* (No. 4). Sage.
- Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of asynchronous learning networks*, 16(3), 9-20.

- Press, G. (2013). A Very Short History Of Big Data. (2013, May 09), *Forbes*.
- Raj, P. (Ed.). (2014). *Handbook of research on cloud infrastructures for big data analytics*. IGI Global, 1-566.
- Rajendran, P. K., Asbern, A., Kumar, K. M., Rajesh, M., and Abhilash, R. (2016). Implementation and analysis of MapReduce on biomedical big data. *Indian Journal of Science and Technology*, 9(31), 1-6.
- Rydning, D. R. J. G. J. (2018). The digitization of the world from edge to core. *Framingham: International Data Corporation*.
- Siemens, G., and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.
- Silva, Y. N., Dietrich, S. W., Reed, J. M., and Tsosie, L. M. (2014, March). *Integrating big data into the computing curricula*. Paper presented at Proceedings of the 45th ACM technical symposium on Computer science education, 139-144.
- Singh, D. S., and Singh, G. (2017). Big Data-A Review. *International Research Journal of Engineering and Technology*, 4(04), 2395-0056.
- Sivarajah, U., Kamal, M. M., Irani, Z., and Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- Solla-Price, D. J. (1975). *Science since Babylon: enlarged edition*, 1-232.
- Strateji Ve Bütçe Başkanlığı. (2018). Stratejik Plan (2019-2023). *Strateji Ve Bütçe Başkanlığı*, 1-76.
- Strateji Ve Bütçe Başkanlığı. (2019). On Birinci Kalkınma Planı (2019-2023). *Strateji Ve Bütçe Başkanlığı*, 1-209.
- Swanstrom, R. (2016, 11 28). Data Scientists, Data Engineers, Software Engineers: The Difference According to LinkedIn. *DataScience 101*.
- TBD. (2016). *Büyük Veri Uygulamaları Çalışma Grubu Raporu*. Türkiye Bilişim Derneği.
- The Economist. (2012). The Deciding Factor: Big Data & Decision Making. *Capgemini*, 1-24.

- Trifu, M. R., and Ivan, M. L. (2014). Big Data: present and future. *Database Systems Journal*, 5(1), 32-41.
- Tulasi, B. (2013). Significance of big data and analytics in higher education. *International Journal of Computer Applications*, 68(14).
- TÜBİTAK. (2016). *Bulut Bilişim ve Büyük Veri Çalıştayı 2015 Faliyet Raporu*. Kocaeli: TÜBİTAK Bilgem, 26-27.
- Vaish, G. (2013). *Getting started with NoSQL*. Packt Publishing Ltd, 1-142.
- Vesset, D., Nadkarni, A., McDonough, B., Bond, S., Li, Q., Olofson, C., . . . Carnelley, P. (2015). IDC FutureScape: Worldwide Big Data and Analytics 2016 Predictions. *International Data Corporation*.
- Wagner, E., and Ice, P. (2012). Data changes everything: Delivering on the promise of learning analytics in higher education. *Educause Review*, 47(4), 32.
- Warden, P. (2011). *Big Data Glossary*. O'Reilly Media, 1-60.
- White, T. (2012). *Hadoop: The definitive guide*. O'Reilly Media, Inc, 1-629.
- Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Sage, 1-440.
- Wottawa, H., and Thierau, H. (1990). *Lehrbuch Evaluation*. Bern:Huber, 1-176.
- Yıldırım, A. (1999). *Nitel Araştırma Yöntemlerinin Temel Özellikleri ve Eğitim Araştırmalarındaki Yeri ve Önemi*. Eğitim Ve Bilim, 23(112), 7-17.
- Yıldırım, A., ve Şimşek, H. (2013). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. Seçkin Yayıncılık, 1-446.
- YÖK. (2015). Yükseköğretim Kurulu Stratejik Planı 2016-2020. *Yükseköğretim Kurulu*, Ankara, 1-47.
- Zikopoulos, P., and Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 1-166.
- Zhou, K., Fu, C., and Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, 215-225.





EKLER

EK-1. Gazi Üniversitesi Ölçme Değerlendirme Etik Alt Çalışma Grubu kararı

Evrak Tarih ve Sayısı: 13.01.2020-E.5392



**T.C.
GAZİ ÜNİVERSİTESİ
Ölçme Değerlendirme Etik Alt Çalışma Grubu**



Sayı : 91610558-302.08.01-
Konu : Bilimsel ve Eğitim Amaçlı

BİLİŞİM ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 28.11.2019 tarih ve E.150460 sayılı yazı

İlgi yazınız ile göndermiş olduğunuz, Enstitünüz Yönetim Bilişim Sistemleri Anabilim Dalı **Doktora Öğrencisi Zeynep AYTAÇ'ın, Doç.Dr.Hasan Şakir BİLGE'nin** danışmanlığında yürüttüğü **"Üniversitelerde Büyük Veri Eğitimi İçin Stratejik Plan Geliştirilmesi"** adlı tez çalışması ile ilgili konu Kurulumuzun **07.01.2020** tarih ve **01** sayılı toplantısında görüşülmüş olup,

İlgilinin çalışmasının, yapılması planlanan yerlerden izin alınması koşuluyla yapılmasında etik açıdan bir sakınca bulunmadığına oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi ve gereğini rica ederim.

**e-imzalıdır
Prof. Dr. Mehtap ÇAKAN
Kurul Başkanı**

Araştırma Kod No: 2020-03

Ek: 1 Liste

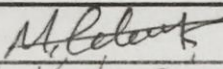
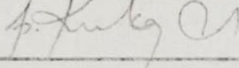
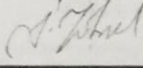
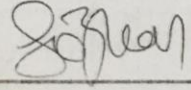
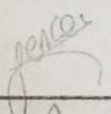
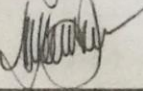
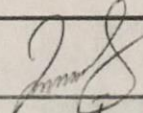
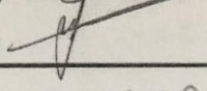


Ankara
Tel 0 (312) 202 20 57 - 0 (312) 2... Faks 0 (312) 202 38 76
İnternet Adresi : <http://etikkomisyon.gazi.edu.tr/>

Bilgi için Burak Çıtrak
Birim Evrak Sorumlusu

Bu belge 5070 sayılı Elektronik İmza Kanununun 5. Maddesi gereğince güvenli elektronik imza ile imzalanmıştır.

EK-1. (devam) Gazi Üniversitesi Ölçme Değerlendirme Etik Alt Çalışma Grubu kararı

GAZİ ÜNİVERSİTESİ ÖLÇME DEĞERLENDİRME ETİK ALT ÇALIŞMA GRUBU KATILIM LİSTESİ	
TOPLANTI TARİHİ : 07/01/2020	TOPLANTI SAYISI : 01
ADI-SOYADI	İMZA
Prof. Dr. Mehtap ÇAKAN Başkan	
Prof.Dr.İsmail KARAKAYA Başkan Yrd.	
Prof.Dr.Galip YÜKSEL	KATILAMADI
Prof.Dr.İsmet YÜKSEL	
Prof.Dr.Seçil ÖZKAN	
Prof.Dr.Cevriye Temel GENCER	
Prof.Dr. C. Haluk BODUR	
Prof.Dr.İbrahim DOĞAN	KATILAMADI
Prof.Dr.Aymelek GÖNENÇ	
Doç.Dr.Nihan KAFA	İZİNLI
Doç.Dr.Zehra GÖÇMEN BAYKARA	
Doç.Dr.İlyas OKUR	
Doç.Dr.Necdet KARASU	KATILAMADI

EK-2. Öğretim elemanları aşına olduğu ve kullandığı büyük veri yazılımları

Yazılım	Aşinalık	Kullanılma	Yazılım	Aşinalık	Kullanılma
	Kişi (n)	Kişi (n)		Kişi (n)	Kişi (n)
Oracle Big Data	67	11	Pig	4	3
Microsoft Hadoop	46	6	Flink	8	2
Apache Hadoop	67	19	Hortonworks	9	5
Amazon EMR	25	3	Ooize	3	2
Cloodera	22	5	Solr	0	0
Kafka	12	4	Sqoop	4	1
HDSF	20	8	Zoo Keeper	8	3
Cassandra	23	4	HBase	15	5
Yarn	5	3	Hive	12	6
Storm	10	1	Chukwa	1	0
Spark	30	10	Mahout	6	1
Avro	2	2	Flume	2	1
				N=153	N=153

EK-3. Öğrencilerin büyük veri farkındalık, yeterlilik ve beklenti ölçeği anketi

Bu anket formu Gazi Üniversitesi Bilişim Enstitüsü'nde yürütülmekte olan doktora tez çalışması için yapılmaktadır. Bu çalışmanın genel amacı, üniversitede eğitim görmekte olan öğrencilerin “Büyük Veri” konusundaki farkındalık, eğilim ve beklentilerini saptamaktır. Anketten elde edilecek olan veriler çalışma sonucunda analiz edilecek ve bu doğrultuda bilimsel çıkarımlar yapılacaktır. Maddelere içtenlikle ve doğru cevaplar vermeniz çalışmanın doğru sonuçlar vermesi açısından çok yararlı olacaktır. Değerli katılımınız ve bilime katkınız için teşekkür ederiz.

Öğr. Gör. Zeynep AYTAÇ
Prof. Dr. Hasan Şakir BİLGE

Büyük Veri ya da Büyük Veri Analitiği kavramlarını daha önce duyduunuz mu? Evet ☐ Hayır ☐

Aşağıdaki Büyük Veri Teknolojilerinden hangilerini duyduunuz?

1-Apache Hadoop ☐	2-Microsoft Hadoop ☐	3-Cloudera ☐	4-Hortonworks ☐	5-Apache Hadoop ☐	6-Oracle Big Data ☐	7-HDFS ☐	8-HBase ☐
9-Cassandra ☐	10-Spark ☐	11-Hive ☐	12-Kafka ☐	13-Flink ☐	14-Pig ☐	15-Storm ☐	16-Yarn ☐
17-Sqoop ☐	18-Mahout ☐	19-Avro ☐	20-Oozie ☐	21-Chukwa ☐	22-Flume ☐	23-ZooKeeper ☐	24-Diğer ☐

Aşağıdaki ifadeler büyük veriye dair farkındalığı ve kariyer beklentilerini ölçmektedir. Lütfen bu ifadelere katılım düzeyinizi belirtiniz.

		1-Kesinlikle Katılmıyorum	2-Katılmıyorum	3-Fikrim Yok	4-Katılıyorum	5-Kesinlikle Katılıyorum
	Büyük Veri-Farkındalık ve Yeterlilik					
A1	Büyük veri analitiği ile tam olarak neler yapılabileceğini biliyorum.	1	2	3	4	5
A2	Veri türünün çeşitlenmesi ve veri miktarının artmasıyla büyük veri çözümleri ortaya çıkmıştır.	1	2	3	4	5
A3	Büyük verilere, bulut çözümlerinde depolama ve erişme imkânı sağlanmaktadır.	1	2	3	4	5
A4	Büyük veri teknolojilerini kullanmak şirketlere çok çeşitli avantajlar sağlamaktadır.	1	2	3	4	5
A5	Büyük veri tanımını oluşturan 5V' nin ne olduğunu biliyorum.	1	2	3	4	5
A6	NoSql kavramının ne anlama geldiğini bilmiyorum.	1	2	3	4	5
A7	Map Reduce (Eşle-İndirge) kavramını daha önce duydum.	1	2	3	4	5
A8	Sosyal medya, akıllı cihazlar ve internet kullanımı büyük veri miktarını artırmaktadır.	1	2	3	4	5
A9	Farklı kaynaklardan, çeşitli formatlarda gelen verinin işlenmesi, büyük veri analitiğinin zorluklarından biridir.	1	2	3	4	5
A10	Büyük verileri işlemek için dağıtık sistemlerden yararlanılır.	1	2	3	4	5
A11	Dağıtık bir dosya sisteminin nasıl çalıştığı hakkında fikrim var.	1	2	3	4	5

EK-3. (devam) Öğrencilerin büyük veri farkındalık, yeterlilik ve beklenti ölçeği anketi

	Büyük Veri Eğitim ve Beklentiler					
B1	Mezun olduktan sonra Büyük Veri Analisti olarak çalışmak isterim.	1	2	3	4	5
B2	Mezun olduktan sonra Büyük Veri analizi yapılan bir şirkette çalışmak isterim.	1	2	3	4	5
B3	Büyük veri ile ilgili bir bölümde yurt dışında lisansüstü eğitim almak isterim.	1	2	3	4	5
B4	Büyük veri ile ilgili bir bölümde lisansüstü eğitim almak isterim.	1	2	3	4	5
B5	Büyük veri ile ilgili eğitim, seminer ve konferanslara katılmak isterim.	1	2	3	4	5
B6	Bilgisayar Mühendisliği ve ilgili bölümlerde Büyük Veri üzerine zorunlu ders açılmalıdır.	1	2	3	4	5
B7	Bilgisayar Mühendisliği ve ilgili bölümlerde Büyük Veri üzerine seçmeli ders açılmalıdır.	1	2	3	4	5
B8	Lisans düzeyinde Büyük Veri konusunun müfredatta mevcut bir derse konularının eklenmesi gerektiğini düşünüyorum.	1	2	3	4	5
B9	Büyük veri ile ilgili konular lisans müfredatında daha fazla yer almalıdır.	1	2	3	4	5
B10	Mezun olmadan önce Büyük Veri ile ilgili ders almak isterim.	1	2	3	4	5
B11	Yurt dışında Büyük Veri ile ilgili daha nitelikli eğitim alabilirim.	1	2	3	4	5
B12	Yurt dışında Büyük Veri alanında daha çok iş fırsatı vardır.	1	2	3	4	5
B13	Büyük veri alanında kısa dönem stajların kariyerime katkısı olacağını düşünüyorum.	1	2	3	4	5
B14	Büyük veri alanında uzun dönem stajların kariyerime katkısı olacağını düşünüyorum.	1	2	3	4	5
B15	Büyük veri konusunda hocalarımla bilgi alışverişi yapmaya istekliyimdir.	1	2	3	4	5
B16	Büyük veri konusunda arkadaşlarımla bilgi alışverişi yapmaya istekliyimdir.	1	2	3	4	5
B17	Büyük veri konusunda sosyal ağlarda bilgi alışverişi yapmaya istekliyimdir.	1	2	3	4	5
B18	Türkiye’de büyük veri alanına Yükseköğretim kurumlarında yeterince ilgi gösterilmediğini düşünüyorum.	1	2	3	4	5
B19	Türkiye’de büyük veri alanına kamu kurumlarında yeterince ilgi gösterilmediğini düşünüyorum.	1	2	3	4	5
B20	Türkiye’de büyük veri alanına özel kurumlarda yeterince ilgi gösterilmediğini düşünüyorum.	1	2	3	4	5
B21	Türkiye’de büyük veri alanında yeterli eğitim fırsatı vardır.	1	2	3	4	5
B22	Türkiye’de Büyük Veri alanında istihdamın yeterli olduğunu düşünüyorum.	1	2	3	4	5

Cinsiyetiniz	Erkek ☐	Kadın ☐	Yaşınız: _____	Not Ortalamanız:
Sınıfınız: _____	Ağırlıklı not ortalamanız: _____			
Üniversiteniz: _____	Okumakta olduğunuz bölüm: _____			
<i>Gerekli görüldüğü takdirde katılımcılar arasında araştırmanın devamına katkı yapmak isteyenler ile büyük veri hakkında derinlemesine görüşme yapılacaktır. Görüşmelerin süresi yaklaşık olarak 30-40 dakikadır.</i>				
Araştırmanın devamına katkı yapmak ister misiniz? Yanıtınız evet ise lütfen size ulaşabileceğimiz iletişim bilgilerinizi paylaşınız.			Evet ☐ E-mail: _____ Cep: _____	Hayır ☐

EK-4. Öğrencilerin aşına oldukları büyük veri yazılımları

Yazılım	Kişi (n)	Yazılım	Kişi (n)
Oracle Big Data	103	Pig	15
Microsoft Hadoop	81	Flink	7
Apache Hadoop	92	Hortonworks	5
Amazon EMR	67	Ooize	2
Cloudera	32	Solr	1
Kafka	23	Sqoop	2
HDSF	19	Zoo Keeper	13
Cassandra	17	HBase	18
Yarn	18	Hive	23
Storm	14	Chukwa	4
Spark	59	Mahout	6
Avro	6	Flume	8
N=179			

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : AYTAÇ, Zeynep
 Uyuğu : T.C.
 Doğum yeri :
 Medeni hali :
 Telefon :
 E-mail :

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Doktora	Gazi Üniversitesi / Yönetim Bilişim Sistemleri	Devam ediyor
Yüksek Lisans	Boston Üniversitesi / Computer Information Systems	2012
Lisans	İstanbul Üniversitesi / Bilgisayar Mühendisliği	2008
Lise	Etimesgut Anadolu Lisesi	2004

İş Deneyimi

Yıl	Yer	Görev
2013-Halen	Aksaray Üniversitesi	Öğretim Görevlisi
2012-2013	Milli Eğitim Bakanlığı	Bilgisayar Mühendisi
2008-2009	Bilişim AŞ.	Bilgisayar Mühendisi

Yabancı Dil

İngilizce

Yayınlar

Aytaç, Z., ve Bilge, H.Ş., (2020). Türkiye’de Kamu Ve Özel Kurumların Büyük Veri Alanında Mevcut Durum Ve Beklentilerinin Değerlendirilmesi Üzerine Nitel Bir Araştırma, *Business & Management Studies: An International Journal*, 8(5): 4646-4679 doi: <https://doi.org/10.15295/bmij.v8i5.1635>

Aytaç, Z., Bilge, H. Ş., (2020). Big Data Analytics In Higher Education: A Systematic Review. *Journal of Internet Applications and Management*, 11 (2) , 81-99. Retrieved from <https://dergipark.org.tr/tr/pub/iuyd/issue/58749/685228>

Kitap Bölümü

Aytaç Z., Adem K., (2020). Büyük Veri., H. Bakırtaş ve S. Çavuş (Editörler). *Yapay Zeka Disiplinleri Dönüştürüyor. Değişime Hazır mıyız?* Ekin Yayınevi, s. 23-50.

Bildiriler

Aytaç and Bilge, Big Data Analytics in Higher Education and Learning Environments: A Review, International Instructional Technologies In Engineering Education Conference (IITEEC), March 29, 2019, Ege University, İzmir, Turkey

Aytaç and Aytaç, *Smes’ Social Media Marketing Adoption; Attitude Toward Technology, Mimetic Motives And Performance*, 8th International Conference on Social Science and Humanities (ICSSH), Dec 17-18, 2016, Dubai.

Poster Sunumu

Duman (Aytaç) and Khan, *Protein-Protein Docking using Monte Carlo and Mean Field Algorithms*, The Ninth Annual USF CS Night, Dec 8, 2010, San Francisco, CA



GAZİLİ OLMAK AYRICALIKTIR.