



**T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



YÜKSEK LİSANS TEZİ

**STRATEJİK BÜYÜK VERİ YÖNETİMİNİN YATIRIMLAR
ÜZERİNDEKİ ETKİLERİ**

Ümit DÜLGER

Mühendislik Bilimleri Anabilim Dalı

Mühendislik Bilimleri Programı

Danışman

Prof. Dr. Ş. İsmail KIRBAŞLAR

II. Danışman

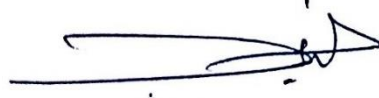
Yrd. Doç. Dr. Gamze SART

Aralık, 2015

İSTANBUL

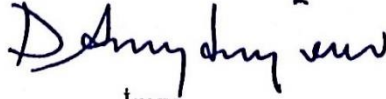
Bu çalışma 07/12/2015 tarihinde aşağıdaki jüri tarafından Mühendislik Bilimleri Anabilim Dalı Mühendislik Bilimleri programında Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Jürisi:



İmza

Prof. Dr. Ş. İsmail KIRBAŞLAR (Danışman)
İstanbul Üniversitesi
Mühendislik Fakültesi



İmza

Prof. Dr. Davut AYDOĞAN
İstanbul Üniversitesi
Mühendislik Fakültesi



İmza

Doç. Dr. Lütfullah Sevgili
İstanbul Üniversitesi
Mühendislik Fakültesi



İmza

Doç. Dr. Hasan USLU
Beykent Üniversitesi
Mühendislik Mimarlık Fakültesi



İmza

Yrd. Doç. Dr. Niyazi KILIÇ
İstanbul Üniversitesi
Mühendislik Fakültesi

ÖNSÖZ

İnternetin ve bilişim araçlarının hızlı gelişimi, yaygınlaşması ve hayatımızın her alanına girmesiyle, klasik verinin aksine çok fazla kaynaktan, çok fazla dağınık ve çok büyük miktarda “Veri Üretimi” gerçekleşmektedir.

Makineler ve kullanılan teknolojik aygıtlar vasıtasıyla insanlar tarafından üretilen bu veri, kullanıcıların coğrafi konumları, alışkanlıkları, harcamaları, sevdikleri, beğendikleri, izledikleri şeyler ve bunun gibi birçok konu hakkında bilgi sağlamaktadır.

Uluslararası alanda “Big Data” olarak adlandırılan ve bilişim çağının yeni bir olgusu olduğu düşünülen “Büyük Veri” ile yapılan analizlerin, endüstriyel alanda yapılabilecek yatırımlar, problem çözümleri, süreç iyileştirmeleri, müşteri memnuniyet analizleri, satış politikaları ve genel şirket stratejileri oluşturmada kullanılması hedeflenmektedir.

“Büyük Veri” analizi başta e-ticaret, finansal hizmetler, kamu hizmetleri, eğitim, sağlık olmak üzere birçok alanda uygulanma imkânı bulmaktadır. En net şekilde, çok fazla özelleştirilmiş teklifler ve iletişim biçimleri şeklinde e-ticaret ile karşımıza çıkmaktadır.

Bir işletmenin “Stratejik Verileri” o işletmeyi daha iyi anlamayı sağlayacak, bununla beraber işletmenin özellikle iş geliştirme, pazarlama ve zamanlama hususundaki kararları ile ilgili kritik bilgiler sunacaktır. Doğru kararlar işletmelerin rekabet gücünü belirleyen en önemli unsurdur. Bu yüzden “Büyük Veri” artık kaçınılmaz bir olgu olarak karşımızda durmaktadır.

Bu çalışma, dünyada “yeni”, Türkiye’de “daha yeni” olan Stratejik “Büyük Veri” ve onun yatırımlar üzerindeki etkilerini araştırmayı amaçlamıştır.

Öncelikle çalışmamda bana yol gösteren, anlayış ve desteğini esirgemeyen tez danışmanım çok değerli hocam Prof. Dr. Ş. İsmail KIRBAŞLAR’a, yapmış olduğu yardımlardan, verdiği değerli bilgilerden ve bana gösterdiği vizyondan dolayı ikinci danışmanım Yrd. Doç. Dr. Gamze SART’a en içten dileklerle teşekkürlerimi sunarım.

Hayattaki azmi ve dürüstlüğüyle bana rehber olan babam Mehmet DÜLGER ve fedakâr insan annem Sevim DÜLGER’e, ayrıca çalışmam sırasında tüm sıkıntılarıma katlanan ve benden manevi desteklerini esirgemeyen sevgili eşim Zeynep DÜLGER ve canım oğlum Mehmet Akif DÜLGER’e de çok teşekkür ederim.

Aralık, 2015

Ümit DÜLGER

İÇİNDEKİLER

Sayfa No

| | |
|--|----------|
| ÖNSÖZ..... | i |
| İÇİNDEKİLER | ii |
| ŞEKİL LİSTESİ..... | v |
| TABLO LİSTESİ | vii |
| SİMGE VE KISALTMA LİSTESİ | viii |
| ÖZET..... | x |
| SUMMARY | xii |
| 1. GİRİŞ..... | 1 |
| 2. GENEL KISIMLAR | 9 |
| 2.1. BÜYÜK VERİ | 10 |
| 2.1.1. Büyük Veri Kavramı | 10 |
| 2.1.2. Büyük Veri Kaynakları..... | 14 |
| 2.1.2.1. Kamusal Veri: <i>Public Data</i> | 14 |
| 2.1.2.2. Özel Veri: <i>Private Data</i> | 14 |
| 2.1.2.3. Çevresel Veri: <i>Data Exhaust</i> | 14 |
| 2.1.2.4. Topluluk Verisi: <i>Community Data</i> | 15 |
| 2.1.2.5. Kişisel Veri: <i>Self-Quantification Data</i> | 15 |
| 2.1.3. Veri Paylaşımı, Gizlilik ve Etiketler | 15 |
| 2.1.4. Büyük Veri Teknolojisi | 19 |
| 2.1.4.1. Büyük Veri Teknolojisine Giriş..... | 19 |
| 2.1.4.2. Büyük Veri Analiz Yazılımı: <i>Hadoop</i> | 21 |
| 2.1.5. Bazı Özel Büyük Veri Yazılımları ve Şirketleri..... | 25 |
| 2.1.6. Büyük Veri Bileşenleri | 27 |
| 2.1.6.1. Çeşitlilik: <i>Variety</i> | 28 |
| 2.1.6.2. Hız: <i>Velocity</i> | 28 |
| 2.1.6.3. Veri Büyüklüğü: <i>Volume</i> | 28 |
| 2.1.6.4. Doğrulama: <i>Verification</i> | 29 |
| 2.1.6.5. Değer: <i>Value</i> | 30 |
| 2.1.7. Büyük Veri Analizindeki Zorluklar..... | 31 |
| 2.1.7.1. Kapsamlı Veri Kaynakları ve Kötü Veri Kalitesi..... | 32 |
| 2.1.7.2. Büyük Verinin Oldukça Verimli Depolanması..... | 33 |

| | |
|---|-----------|
| 2.1.7.3. Veriyi Verimli Bir Şekilde İşleme..... | 33 |
| 2.1.7.4. Yiğın Veri Madenciliği..... | 34 |
| 2.1.8. Büyük Veri Analizi Teknolojisi | 34 |
| 2.1.8.1. Toplama ve Önışleme..... | 34 |
| 2.1.8.2. Depolama..... | 35 |
| 2.1.8.3. İşleme..... | 37 |
| 2.1.8.4. Veri Madenciliği | 38 |
| 2.1.9. Büyük Verinin Avantajları | 41 |
| 2.1.10. Büyük Verinin Zorlukları | 41 |
| 2.1.11. Bazı Büyük Veri Uygulama Alanları | 42 |
| 2.1.11.1. Finansal Servisler | 45 |
| 2.1.11.2. Pazarlama ve Perakendecilik | 49 |
| 2.1.11.3. Kamu ve Hükümetler | 49 |
| 2.1.11.4. Eğitim..... | 50 |
| 2.1.11.5. Sağlık | 52 |
| 2.1.11.6. Tarım..... | 56 |
| 2.1.11.7. Enerji | 57 |
| 2.1.11.8. Haberleşme ve Ulaşım | 58 |
| 2.1.12. Örnek Vakalar: Büyük Veri Yönetimi Çalışması..... | 59 |
| 2.1.12.1. Örnek Vaka 1: Müşteriler için Benzersiz Deneyimler Oluşturacak Hızlı Analitik Modelleme ve Simülasyon Kullanımı | 60 |
| 2.1.12.2. Örnek Vaka 2: Yeni Pazarlar Gelişirken Rasyonelleştirilmiş Teklif Aralığıyla Müşterileri Eşleştirme..... | 61 |
| 2.1.12.3. Örnek Vaka 3: Analitik Fırsat Olarak Büyük Müşteri Çeşitliliğinin Davranışlarının Değerlendirilmesi..... | 62 |
| 2.2. BÜYÜK VERİNİN GELİŞİMİ VE VERİ YÖNETİMİ | 63 |
| 2.3. BÜYÜK VERİNİN ZORLUKLARI VE FIRSATLARI | 65 |
| 2.4. BÜYÜK VERİYE UYUM STRATEJİLERİ | 69 |
| 2.5. BÜYÜK VERİNİN İNSAN UNSURU | 70 |
| 2.6. VERİ TIPLERİ VE VERİNİN GERÇEK DEĞERİ | 72 |
| 2.7. YENİ MODELLER VE BÜYÜK VERİ..... | 76 |
| 3. MALZEME VE YÖNTEM | 79 |
| 3.1. YÖNTEM..... | 79 |
| 3.1.1. Giriş | 79 |
| 3.1.1.1. Analitik Araştırmaya Karşın Betimsel Araştırma..... | 79 |

| | |
|---|------------|
| 3.1.1.2. Temel Araştırmaya Karşı Uygulamalı Araştırma..... | 80 |
| 3.1.1.3. Ampirik (DeneySEL) Araştırmaya Karşın Kavramsal Araştırma | 81 |
| 3.1.1.4. Diđer Bazı Araştırma Türleri | 81 |
| 3.1.1.5. Nicel (Sayısal) Araştırmaya Karşı Nitel (Vasfi) Araştırma..... | 82 |
| 3.1.2. Araştırmanın Yöntemi | 87 |
| 3.2. VERİ TOPLAMA | 90 |
| 3.3. VERİ ANALİZİ VE RAPORLAMA | 91 |
| 3.4. ARAŞTIRMANIN KISITLARI..... | 91 |
| 3.5. ETİK HUSUSLAR..... | 92 |
| 4. BULGULAR | 93 |
| 4.1. ÖRNEKLEM..... | 93 |
| 4.2. GÖRÜŞMELER..... | 94 |
| 4.3. NVIVO İLE GÖRÜŞMELERİN İNCELENMESİ..... | 96 |
| 4.3.1. Bilgisayar Destekli Nitel Veri Analizi Hakkında | 96 |
| 4.3.2. NVivo ile Görüşmelerin Deđerlendirilme Süreci..... | 98 |
| 4.4. GÖRÜŞME ANALİZ TABLOSU | 104 |
| 5. TARTIŞMA VE SONUÇ | 108 |
| 5.1. BÜYÜK VERİNİN KULLANIMI | 108 |
| 5.2. BÜYÜK VERİ HAKKINDAKİ GÖRÜŞLER VE TAHMİNLER..... | 111 |
| 5.3. SONUÇ | 117 |
| KAYNAKLAR | 129 |
| EKLER..... | 135 |
| EK 1. “Stratejik Büyük Verinin Yatırımlar Üzerindeki Etkileri” Tezi Araştırması Görüşme Rehberi..... | 135 |
| ÖZGEÇMİŞ..... | 138 |

ŞEKİL LİSTESİ

| | Sayfa No |
|---|-----------------|
| Şekil 1.1: Veri Büyüklüğü – Bayt birimleri..... | 1 |
| Şekil 1.2: İnternette 1 dakika içerisinde gerçekleşenler. | 3 |
| Şekil 2.1: Büyük Veri Yönetim sistemleri..... | 25 |
| Şekil 2.2: Büyük Veri bileşenleri. | 27 |
| Şekil 2.3: Küresel Verinin hacminin tahmini. | 29 |
| Şekil 2.4: Büyük Verinin en önemli beş bileşeni. | 30 |
| Şekil 2.5: Bir Büyük Veri senaryosu. | 32 |
| Şekil 2.6: Büyük Veri Mimarisi. | 35 |
| Şekil 2.7: Arama motoru mimarisi. | 39 |
| Şekil 2.8: Akıllı öneri sistemi mimarisi..... | 40 |
| Şekil 2.9: Perakende alanında Büyük Veri kullanım ve projeleri..... | 49 |
| Şekil 2.10: Öğrenci davranışları analizi yaklaşımı. | 52 |
| Şekil 2.11: Taşıma sistemleri akıllı servisler entegrasyonu..... | 58 |
| Şekil 2.12: Büyük Verinin Fırsat Sunduğu Alanlar..... | 60 |
| Şekil 2.13: Büyük veri fırsatları: 560 işletmenin %50'den fazlası büyük verinin operasyonel verimliliği arttırmada faydalı olabileceğini düşünür vb. | 67 |
| Şekil 2.14: Bilgi keşif süreci. | 68 |
| Şekil 4.1: CAQDAS (Computer Aided Qualitative Data Analysis) ve NVivo. | 97 |
| Şekil 4.2: NVivo programına görüşmelerin yüklenmesi. | 98 |
| Şekil 4.3: Nodes ve Tree Nodes kodlama görünümü. | 99 |
| Şekil 4.4: Nodes ve Tree Nodes kodlamalarının incelenmesi. | 100 |
| Şekil 4.5: Analizin genel çerçeve modellemesi. | 101 |
| Şekil 4.6: Analizin detay modellemesi | 101 |
| Şekil 4.7: NVivo kelime sorgulamaları. | 102 |
| Şekil 4.8: NVivo'da grafik desteği örnek gösterimi. | 103 |

| | |
|--|-----|
| Şekil 5.1: Büyük Veri sektörünüzde kullanılıyor mu? | 108 |
| Şekil 5.2: Büyük Veri kurumunuzda kullanılıyor mu?..... | 111 |
| Şekil 5.3: İş dünyasında Büyük Veri Yönetimi yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir?..... | 113 |
| Şekil 5.4: Büyük Veri ve Veri Yönetiminin sınırının olacağını tahmin edebiliyor musunuz? | 114 |
| Şekil 5.5: Sizce Büyük Veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir? | 116 |
| Şekil 5.6: Türkiye’de Büyük Veri için altyapısal yeterlilik var mıdır? | 118 |
| Şekil 5.7: Büyük Veri Yönetimine olan ihtiyaç, size göre altyapısal mı yoksa yönetsel bir konu mudur?..... | 119 |
| Şekil 5.8: Büyük Verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz? | 120 |
| Şekil 5.9: Sizce Büyük Veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır? | 121 |
| Şekil 5.10: Türkiye’de Büyük Veriden faydalanma seviyeleri ile ilgili neler söylemek istersiniz? | 121 |
| Şekil 5.11: Sizce Türkiye’de Google, Facebook, Amazon vb. gibi Büyük Veriden ciddi anlamda faydalanan benzeri firmalar var mıdır? | 122 |
| Şekil 5.12: Büyük Verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz? | 122 |
| Şekil 5.13: Büyük Verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz? | 123 |
| Şekil 5.14: Büyük Veri yatırımlarınız üzerinde ne kadar etkili olacaktır? | 124 |
| Şekil 5.15: Büyük veri ile ilgili ABD Beyaz Saray anketi. | 126 |

TABLO LİSTESİ

| | Sayfa No |
|---|-----------------|
| Tablo 2.1: Büyük veri işleme gereksinimleri..... | 37 |
| Tablo 2.2: Sayısallaştırılmış-Kişisel hareket ve sosyal medyanın Büyük Veri ile sağlık alanında kullanılmasına örnekler..... | 54 |
| Tablo 3.1: Nicel arařtırmaların avantajları ve sınırlılıkları..... | 83 |
| Tablo 3.2: Nitel arařtırmaların avantajları ve sınırlılıkları..... | 83 |
| Tablo 3.3: Nicel ve Nitel arařtırma arasındaki farklar..... | 84 |
| Tablo 4.1: Örneklem tablosu..... | 93 |
| Tablo 4.2: Görüşme analiz tablosu..... | 104 |

SİMGE VE KISALTMA LİSTESİ

Simgeler

Açıklama

| | |
|------------|------------------------------------|
| EB | : Eksabayt, 10^{18} |
| EUR | : Euro, Avrupa Birliği Para Birimi |
| GB | : Gigabayt, 10^9 |
| KB | : Kilobayt, 10^3 |
| M | : Milyon |
| MB | : Megabayt, 10^6 |
| PB | : Petabayt, 10^{15} |
| TB | : Terabayt, 10^{12} |
| USD | : Amerikan Doları, ABD Para Birimi |
| YB | : Yottabayt, 10^{24} |
| ZB | : Zettabayt, 10^{21} |

Kısaltmalar

Açıklama

| | |
|---------------|--|
| AB | : Avrupa Birliği |
| ABD | : Amerika Birleşik Devletleri |
| API | : Application Programming Interface – Uygulama Programlama Arayüzü |
| BI | : Business Intelligence – İş Zekası |
| BLOB | : Binary Large Object – İkili Büyük Nesnelere |
| BSP | : Bulk Synchronous Parallel – Paralel Yığın Senkronizasyon |
| BT | : Bilişim Teknolojileri |
| CAP | : Consistency, Availability, Partition Tolerance – Tutarlılık, Ulaşılabilirlik, Bölünebilirlik Toleransı |
| CAQDAS | : Computer Aided Qualitative Data Analysis – Bilgisayar Destekli Nitel Veri Analizi |
| CEO | : Chief Executive Officer – İcra Kurulu Başkanı |
| CCTV | : Close Circuit Television – Kapalı Devre Televizyon Sistemi |
| CPU | : Central Process Unit – Merkezi İşlem Birimi |
| CRM | : Customer Relationship Management – Müşteri ilişkileri Yönetimi |
| DAG | : Directed Acyclic Graph – Döngüsüz Grafik Diyagramı |
| DI | : Digital Inventory – Dijital Envanter |
| EMR | : Electromagnetic Radiation – Elektromanyetik Radyasyon |
| ER | : Entity Relationship – Varlık İlişki |
| ERP | : Enterprise Resource Planning – Kurumsal Kaynak Planlaması |
| FTC | : Federal Trade Commission – Federal Ticaret Komisyonu |
| GFS | : Google File System – Google Dosya Sistemi |
| GPS | : Global Positioning System – Küresel Konumlama Sistemi |
| GSM | : Global System for Mobile Communication - Küresel Mobil İletişim Sistemi |
| HANA | : High Performance Analytic Appliance – Yüksek Performanslı Analitik Araç |
| HDFS | : Hadoop Distributed File System – Hadoop Dağıtık Dosya Sistemi |
| HPA | : High-Performance Analytics – Yüksek Performanslı Analiz |
| IP | : Internet Protocol Address – İnternet Protokol Adresi |

| | |
|-----------------|--|
| I/O | : Input/Output – Giriş/Çıkış |
| IDC | : International Data Corporation – Uluslararası Veri Kurumu |
| IT | : Information Technology – Bilgi Teknolojileri |
| MIKE 2.0 | : Method for an Integrated Knowledge Environment – Entegre Bilgi Çevresi Yönetimi |
| NFS | : Network File System – Ağ Dosya Sistemi |
| NIH | : The National Institute of Health – Ulusal Sağlık Enstitüsü |
| NoSQL | : Not Only SQL – Yalnızca SQL değil – İlişkisel Veritabanı Yönetim Sistemi |
| ODS | : The Operational Data Store – Operasyonel Veri Depolama |
| PI | : Physical Inventory – Fiziksel Envanter |
| POSIX | : Portable Operating System Interface for Unix – UNIX için Taşınabilir İşletim Sistemi Arabirimi |
| RAID | : Redundant Array of Independent Disks – Bağımsız Disklerin Artıklıklı Dizisi |
| RAM | : Random Access Memory - Rastgele Erişimli Bellek |
| RFID | : Radio Frequency Identification – Radyo Frekansı ile Tanımlama |
| ROI | : Return On Investment – Yatırı Getirisi |
| SLA | : Service Level Aggrement – Hizmet Seviye Anlaşması |
| SQL | : Structured Query Language – Yapılandırılmış Sorgu Dili |
| SSD | : Solid State Drive - Katı Hal Sürücü |
| UN | : The United Nations – Birleşmiş Milletler |
| USPTO | : United States Patent and Trademark Office – ABD Patent Ofisi |
| VP | : Vice President – Başkan Yardımcısı |

ÖZET

YÜKSEK LİSANS TEZİ

STRATEJİK BÜYÜK VERİ YÖNETİMİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ

Ümit DÜLGER

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Mühendislik Bilimleri Anabilim Dalı

Danışman: Prof. Dr. Ş. İsmail KIRBAŞLAR

II. Danışman: Yrd. Doç. Dr. Gamze SART

Kurumların kendi veri ambarlarında –tüm veri deposu – ve veri silolarında – diğer departmanlarından bağımsız bir departmanın kontrolü altındaki alt veri depoları – yapılandırılmış şekilde yerel veri birikmektedir. Yerel veri ambarındaki veriler ile stratejik kararlar hedeflenirken, veri silolarındaki veriler ile acil bir amaca yönelik ya da taktiksel kararlar hedeflenmektedir. Fakat bu yerel veriler, içinde bulunduğumuz çağ itibariyle kurumlar ve işletmeler için doğru ve stratejik kararlar vermeye yardımcı olma hususunda yetersiz kalmaktadır.

Bugün bu yerel verilerden başka, milyonlarca insan tarafından kullanılan bilgisayar, GPS aygıtları, cep telefonları, tıbbi aygıtların milyarlarca etkileşimleri ve bloglar, sosyal medyada paylaşılan veriler, muhtelif sensörlerden toplanan veriler, elektronik postalar, fotoğraflar, videolar, web günlükleri gibi platformlar tarafından yapılandırılmamış ya da yarı-yapılandırılmış “Veri Seli” oluşmaktadır. Ağırlıklı olarak bu etkileşimlerin birçoğu şu ana kadar ihtiyaçları ve alışkanlıkları yetersiz bir şekilde anlaşılan insanlar tarafından kullanılan mobil aygıtlar aracılığıyla oluşmaktadır.

Bu doğrultuda artışı devam eden ancak yapısal veri olarak dikkate alınmayan verinin analiz edilerek yapısal veri birikimi gibi değerlendirilmesi gündeme geldi. Söz konusu verinin kurum ve kuruluşlardaki örtük bilginin, açık bilgi ile birleştirilmesiyle elde edilen yapısal bilgiyi daha da güçlendirecek bir içerik ve zenginliğe erişirmesi beklenebilir. Böyle bir uygulamanın sonucu ise kurum ve kuruluşların alacakları stratejik kararların hedef kitleyi kapsama ve etkileme alanını genişletebileceği olasılığını da beraberinde getirecektir.

Bu doğrultuda, modern çağda veri hacmi gitgide büyürken, işletmeler ve diğer kuruluşlar geleneksel veri yöntemlerinden çok daha geniş bir şekilde tanımlanan “**Büyük Veri**” ile

başa çıkmak zorunda kalmaktadırlar. Büyük Veri birçok kurum ve şirket için zorluk anlamına geldiği gibi, aynı zamanda fırsat anlamına da gelmektedir. Diğer yandan pek çok endüstri trendi geleneksel veri yönetimi, iş zekâsı platform ve araçları üzerinde baskı meydana getirmektedir.

Kar amacı güden ya da gütmeyen birçok kurum ve kuruluş yatırımlar, problem çözümleri, süreç iyileştirmeleri, müşteri memnuniyeti çalışmaları, satış politikaları ve genel kurum stratejileri oluşturmada “Büyük Veri” analizlerinden faydalanmaları gerektiğini öngörmektedirler.

Bununla birlikte, araştırmacılar ve politikacılar da düşük gelirli nüfusun faydası için krizlerin öngörülmesi ve önlenmesi, muhtelif hizmetler sağlanması, ihtiyaçların tanımlanması için kullanılabilen, işlemeye uygun bilgi içerisindeki veri sağanaklarının potansiyel zorluğunu fark etmeye başlıyor.

Birlikte kararlaştırılmış uyumlu aksiyon ile veriyi oluşturan bireylere ve topluluklara yardım eden hükümetler, kalkınma kuruluşları ve şirketler için bu veriden faydalanmak gereklidir.

Retention Science veri analizi firmasının CEO’su Jerry Jao’nun müşterilerinden birisi olan ve internet üzerinde online bisiklet, motosiklet parçaları satışı yapan Bikeberry.com firması ile yaptıkları çalışma için verdiği rakamlar dikkat çekicidir: Bikeberry.com için öncelikle web tarayıcılar, giriş sayıları, geçmiş alışverişler ve daha başka şekillerde büyük bir müşteri verisi elde edildi. Elde edilen veriye göre müşterilere 5 ayrı teklif oluşturuldu. Bunlar ücretsiz kargo, %5, %10, %15 ve %30 indirim şeklinde tanımlandı. Oluşturulan yazılım ile müşterilerin özelliklerine ve geçmiş alışverişlerine bakılarak uygun özelleştirilmiş teklifler gönderildi. Sonuçta, Bikeberry.com’un satışları %133 ve kullanıcı bağlılığı yaklaşık %200 arttı. Ayrıca Bikeberry.com oluşturduğu teklif seviyelerinden dolayı tüm müşterilerine aynı yüksek indirim yapmadığından dolayı kar marjını da yükseltmiş oldu (<http://www.bigdatalandscape.com/news/why-big-data-is-a-must-in-ecommerce>, 2015).

Bu çalışmada “**Stratejik Büyük Veri Yönetiminin Yatırımlar Üzerindeki Etkileri**” ile ilgili olgusal ve nitel bir araştırma yürütülmüştür. Konuyla ilgili farklı kuruluşlardan bireyler ile yarı-yapılandırılmış görüşmeler yapılmış, bu kişilerin “Büyük Veri” ile ilgili görüşleri, deneyimleri, önyargıları hakkında bilgi edinilmiştir.

Genel olarak Büyük Veri ile ilgili gerçekçi bir yaklaşımın olduğu tespit edilmiş ve bazı görüş ayrılıklarının varlığı dikkat çekmiştir.

Aralık 2015, 151 Sayfa.

Anahtar kelimeler: Büyük Veri, Yatırım, Yönetim, Bulut Bilişim, Sosyal Medya

SUMMARY

M.Sc. THESIS

**THE EFFECTS OF STRATEGIC BIG DATA MANAGEMENT ON
INVESTMENTS**

Ümit DÜLGER

İstanbul University

Institute of Graduate Studies in Science and Engineering

Department of Engineering Sciences

Supervisor: Prof. Dr. Ş. İsmail KIRBAŞLAR

Co-Supervisor: Asst. Prof. Dr. Gamze SART

Corporations collect structured local data in their data storages – whole data storage – and data silos – sub data storages which is under control of independent department –. When strategic decisions targeting with data from local data storage, urgent propovise or tactical decisions targeting with data. But these local datas are getting insufficient about helping to make right and strategic choices for corporations and businesses in the age that we live in.

Today, structured or semi-structured “Data Torrent” consisting without local data from the platforms like blogs, videos, photographs, e-mails, datas collected from various sensors, data shared in social media, billions of interactions of medical devices, cellphones, GPS devices, millions of computers used by people. Mainly these interactions occuring by mobile devices which can not understood enough by users with their needs and habits.

Accordingly, evaluating structural data such as the accumulation of data analyzed came up which is continuing growth in this direciton but not considered as structured data. Said data can be expected that is bring to wealth and content that strenghten obtained structural information with gathered from open and hidden data in institutions and businesses. The result of such an applications would bring the possibility of expand area of influence and target audiance of institutions and businesses who makes strategic decisions.

In this direction, while data capacity growing more and more in the modern age, businesses and other corporations must deal with “**Big Data**” which is way too larger than conventional data methods. Big Data is mean difficulty for most of corporations and businesses, but on the other hand, it means oppurtunity in the same time. On the other

hand, it creates pressure for business intelligent platforms, tools and traditional data administrations of plenty industry.

Many profit or nonprofit corporations and institutions investments, problem solutions, process restorations, customer satisfaction studies, sales policies and for create general institution strategies, should profit by “Big Data” analysis.

However, researchers and politicians starting to realize the difficulty of potential data torrents in data that are suitable for processing that can predict and prevent crisis for benefit of low income population, provide various services for using define needs.

For development corporations and businesses, governments who aid to individuals and communities which can create compatible action and data together.

The CEO of Retention Science data analysis company, Jerry Jao makes business and give some interesting numbers to a company named Bikeberry.com which sells motorbike and bike spare part via internet: Firstly, web browsers, number of entries, past shoppings and with other different types of data gathered. 5 different proposal were created with the gathered data. These were defined as free shipping and %5, %10, %15 and %30 discount. Customized proposals sent to customers as looking their features and past shoppings with a created software. At the end, Bikeberrycom’s sales of increased %133, customer loyalty increased about %200. Also Bikeberry.com increased its profit margin because of did not making same discount to their all customers (<http://www.bigdatalandscape.com/news/why-big-data-is-a-must-in-ecommerce>, 2015).

In this study, a factual and qualitative research conducted about “**The Effects of Strategic Big Data Management on Investments**”. Data gathered of prejudice, experiences, thoughts about “Big Data” while making semi-structured interviews with people from different institutions.

Generally realistic attitude confirmed about Big Data and some divergence existence drew attention.

December 2015, 151 Pages.

Keywords: Big Data, Investment, Management, Cloud Computing, Social Media

1. GİRİŞ

Teknolojinin yaygınlaşması, herkes tarafından ulaşılabilir olması ve kullanım biçiminin dönüşmesi, üretilen ve işlenen verinin büyüklüğünün de devasa olarak artmasını beraberinde getirdi. Nasıl ki kişisel hayatlarımızda iki üç disket veya birkaç CD'ye sığan bilgilerimiz artık 1 TB'lık yüksek kapasiteli disklerle zor sığar hale geldiyse, kurumlar da bu devasa hız ve çeşitlilikte artan verilerini nereye sığdıracaklarını bilemiyorlar. Büyük miktarda veri ile uğraşmak bilişim dünyasının yabancı olduğu bir konu değil fakat “Büyük Veri” konusu veri ambarı, ODS (The Operational Data Store: Operasyonel Veri Depolama), iş analizi ve veri analizi başlıklarından oldukça farklı görünmektedir.

Dünyada mobil abone sayısı 5 milyara internet kullanıcısı sayısı da 2 milyara ulaştı. Siber dünya tüm öngörülerin çok daha ötesinde bütün boyutları aşan, sürekli ve inanılmaz hızda büyüyen sayısal bir bilgi patlamasının etkisi altına girmiş bulunuyor. 2000 yılında New Mexico eyaletindeki gözlemevindeki teleskopun çalışmaya başladığı ilk birkaç hafta içinde topladığı veri astronomi tarihinde o zamana dek toplanan tüm veriden daha fazla oldu. 2013 yılı içinde ise bu merkezin 10 yıllık arşivinde 140 TB tutarında bilginin toplandığı tahmin ediliyor. Bu nicelikteki bilginin Şili'de 2016 yılında devreye girecek yeni bir gözlemevinde her beş gün içinde toplanacak bilgi kadar olacağı öngörülüyor.

| Bayt Birimleri | | | | | | |
|-----------------|--------|-----------|----------|--------------|--------|----------|
| Yaygın örnek | | | | Binari örnek | | |
| Ad | Sembol | Ondalık | İkilik | Ad | Sembol | İkilik |
| kilobayt | KB | 10^3 | 2^{10} | kibibayt | KiB | 2^{10} |
| megabayt | MB | 10^6 | 2^{20} | mebibayt | MiB | 2^{20} |
| gigabayt | GB | 10^9 | 2^{30} | gibibyte | GiB | 2^{30} |
| terabayt | TB | 10^{12} | 2^{40} | tebibayt | TiB | 2^{40} |
| petabayt | PB | 10^{15} | 2^{50} | pebibayt | PiB | 2^{50} |
| eksabayt | EB | 10^{18} | 2^{60} | eksbibayt | EiB | 2^{60} |
| zettabayt | ZB | 10^{21} | 2^{70} | zebibayt | ZiB | 2^{70} |
| yottabayt | YB | 10^{24} | 2^{80} | yobibayt | YiB | 2^{80} |

Şekil 1.1: Veri Büyüklüğü – Bayt birimleri.

(Kaynak: <https://tr.wikipedia.org/wiki/Bayt>)

Bilgisayar Bilimleri Kurumu'nun (Computer Sciences Corporation) yayınladığı bir rapora göre 2020 yılında veri boyutunun günümüze göre %4300 artış göstereceği öngörülmektedir. Günümüzde orta ölçekli bir kurum bile 1 TB'lık veriyi kısa bir zamanda üretebilmektedir ve bu veri birçok kaynaktan gelerek yüksek çeşitlilikte olabilmektedir. Bu nedenle "Büyük Veri" kavramı ortaya çıkmıştır. Büyük veri teknolojik ilerlemelerin neticesi olarak ortaya çıkan bu yüksek hacimli, hızda ve çeşitlilikteki verileri ve bunların saklanması, içindeki bilgilere erişimi, işlenmesi, analiz edilmesi gibi konuları ifade etmekte olup, günümüz dünyasında teknoloji şirketlerinin gerek performans ve yönetim zorluğundan ötürü gerekse de rekabet avantajı yaratmak için gündem maddelerinin başında gelmektedir.

Yüksek hacimli, çeşitlilikteki ve hızdaki verilerin geleneksel veri tabanı yönetim sistemleri ile yönetilmesi, işlenmesi ve bilgi çıkarımı mümkün olmamaktadır. Bu yüzden sunucu kümelerinde paralel olarak işleyen yazılımlar gibi farklı algoritmalar, teknikler ve teknolojiler gerekmektedir. Yüksek hızda üretilen veri beraberinde verinin aynı hızda alınması, anlaşılması ve işlenmesi problemlerini de getirmektedir. Büyük veri kavramı yüksek hacimdeki ve hızdaki veri ile ilgili konuları içermekle beraber aynı zamanda çeşitlilik içeren video, ses, metin gibi yapısal olmayan veya akışkan verilerin saklanması, yönetilmesini ve üzerinden bilgiye erişimi de ifade etmektedir.

Örneğin; internet üzerinde her 1 dakikada ne yapıldığı, bir enformasyon yönetimi stratejisi geliştirilebilmesi açısından büyük önem arz etmektedir. Intel'in 2013 yılında yaptığı bir araştırmanın bulguları (Lepi, 2014) göstermektedir ki, internet üzerinde 1 dakika içerisinde 639.800 GB'lık veri trafiği gerçekleşmektedir. Bu veri akışı içinde Facebook ortalama 277.000 kullanıcı girişi ve 6 milyon kullanıcı, Google 2 milyon kavram araması, Youtube 1,3 milyon video izlemesine sahiptir. Bununla birlikte Twitter 100.000 twit – 2014 yılı itibarıyla bu rakam dakikada 340.000 twite çıkmıştır –, Flickr 3000 fotoğraf yüklemesi ve LinkedIn 100'ün üzerinde yeni kullanıcı hesabı açılmasıyla ticari işletmelerin veri akışında ne denli büyük pay sahibi olduklarını ortaya koymaktadır. Gönderilen 204 milyon elektronik posta ki bunun % 75'i spam mesajlardır (Tellan, 2014), 133 zararlı yazılım bulaşması, 20 kimlik hırsızlığı ise internete erişmenin bedelinin kullanıcı adı, parola, şifre, kredi kartı bilgileri ile vatandaşlık ve sosyal güvenlik numaralarının kaybedilmesi pahasına olduğunu açığa çıkarmaktadır.



Şekil 1.2: İnternette 1 dakika içerisinde gerçekleşenler.

(Kaynak: <http://scoop.intel.com/what-happens-in-an-internet-minute/>, 2012)

Pek çok ülkede 8400 mağazası ve 2 milyondan fazla çalışanı olan dünyanın en büyük perakende mağazalar zinciri Wal-Mart'ta bir saat içindeki müşteri işlemleri 1 milyonu geçiyor. Bu işlemler sonucu veri tabanlarına ulaşan bilgi Amerikan Kongre Kütüphanesi'ndeki kitapların 167 katıdır. Dünyada internet aramalarının yarısının yapıldığı Google saniyede 35 bin kez tıklanmayla bilgi sağlıyor. Bütün bu gelişmeler sayısal bilgi çoğalmasının sonsuza dek uzandığı görünümünü veriyor.

Daha farklı bir örnekte, Formula 1 yarışına katılan bir araba üzerindeki 150 sensörden 20 GB veri üretir ki, bu veriler aracın bileşenlerinin teknik performansını analiz etmeye yardımcı olabilir. Fakat sürücü reaksiyonları, pit stop gecikmeleri, mürettebat ve şoför arasındaki iletişim de bütün performansa katkı sağlar (Munford, 2014).

Bilgi patlaması aynı zamanda birçok yeni sorunu da beraberinde getiriyor. Bütün bu bilgiyi toplayan, işleyen ve paylaşan algılayıcıların, bilgisayarların, mobil telefonların

sayılarının artmasına karşın var olan depolama kapasitesi toplanan bilginin çok altında kalıyor. Küresel bilgi hacmi, depolama kapasitesini 2007 yılından beri aşmış bulunuyor.

BT dünyasında şimdi bu durum karşısında bilgi yönetimi de çok büyük önem kazanıyor. Son yıllarda Oracle, IBM, Microsoft ve SAP'ın 25 milyar dolar tutarında yaptıkları yatırımlarla veri yönetiminde uzmanlaşmış yazılım şirketlerini satın aldıkları görülüyor. Sadece veri yönetimi yazılımları endüstrisinin değeri 100 milyar doları aşarken yıllık büyüme hızı da yazılım endüstrisinin iki katı olarak yüzde 10'a yaklaşıyor. Büyük şirketlerde bilgi yönetim bölümlerinin başkanları çok önemli bir konuma gelirken, toplanan veriden gerekli bilgiyi çıkarma işi de o denli önemli bir işlevlik kazanıyor.

Evet, günümüzde sadece insanın değil insan, çevre, makine ve benzeri işlem gören her bir sürecin bilgilerine ulaşılmakta ve bu bilgiler kayıtlarda tutulmaktadır. Bu kapsamda da, insan-insan, insan-makine, makine-makine ara yüzleri arasında sürekli işlem gören bir bilgi oluşumu ve akışı söz konusu olabilmektedir.

Özellikle bilişim sektöründe Pazar araştırmaları yapan IDC (International Data Corporation, ABD) firmasının tespitine göre "Dünyadaki verinin yüzde 1'inden azı analiz ediliyor" başlığı altında şu bilgilere yer verilmektedir; Bulut bilişim ve getirileri hızla artarken, piyasaların bu gelişime ayak uyduramadığı görülüyor. Buluta dayalı BT çalışanlarının 2015 yılına kadar yıllık yüzde 26'lık bir büyüme göstermesi bekleniyor. IDC'nin raporuna göre, bu durum da dünya çapında 7 milyon yeni bulut bağlantılı iş imkânı doğmasını ifade ediyor. Buna rağmen BT çalışanı istihdamına bakıldığında 2012 yılında 1,7 milyon bulut bağlantılı pozisyonun boş kaldığı ortaya çıktı. Sebep olaraksa eğitim ve sertifikasyonun zayıf olması gösterildi.

Bilgisayar, akıllı telefon gibi cihazların yaygınlaşması, Türkiye gibi gelişmekte olan ülkelerdeki internet erişiminin artması ve CCTV kameraları ya da akıllı algılayıcılar gibi makinelerin oluşturduğu verinin çoğalması dijital dünyanın 2,8 ZB'a kadar ulaşmasını sağladı. Raporlara göre 2020 yılında dünyada 40 ZB'ın üzerinde veri birikmesi bekleniyor. Buradaki en belirleyici etkenin 2005'te tüm verilerin yüzde 11'ini oluşturan makinelerin, 2020'de yüzde 40'ını oluşturacak olması görülüyor.

Bu değerli veri yığınlarının analiz edilememesi ya da sınıflandırılmaması önemli bir sorunu ortaya çıkartırken, eğer yapılabilseydi 2012 yılındaki verilerin yüzde 23'ü (643

EB) büyük veri için faydalı olacaktır. Şu anda yalnızca verilerin yüzde 3'ü sınıflandırılırken, yüzde 0,5'i analiz edilebilmiştir.

Bilgi patlamasının en büyük nedeni olarak sürekli gelişen teknoloji gösterilmektedir. Sayısal araçların bir yandan kapasiteleri ve uygulamaları artarken diğer yandan fiyatları düşmekte ve bu araçları kullanıp bilgiye ulaşmak isteyenlerin sayısı da görülmemiş bir hızla büyümektedir.

Küresel bilgisayar kullanımının büyümesi de bilgiye ulaşma isteğini körüklemektedir. Bilginin büyük bir bölümü paylaşılırken iletişim donanımları üreticisi Cisco'ya göre ise, 2013 yılında internet trafiği yılda 667 EB'a ulaşacaktır ve veri miktarının da ağların taşıyabileceği kapasitenin çok daha üstünde olacağı tahmin edilmektedir (Cisco, <http://www.cisco.com/web/TR/news/press/archive/2009/070209.html>, 2009). Bugün geldiğimizde noktada ise, Cisco'nun son raporunda 2015 küresel internet trafiği yılda 966 EB'a ulaşacak ve 1 ZB seviyesine yaklaşacaktır. 2010'da ayda yaklaşık 20,2 exabyte olan küresel IP trafiğinin 2015'te ayda 80.5'e çıkması; 2015'teki küresel ortalama IP trafiğinin saniyede 245 terabyte'a (200 milyon insanın her gün aynı anda 1.2 Mbps hızında HD film izlemesine eşdeğer) ulaşması bekleniyor (Cisco, <http://www.cisco.com/web/TR/news/press/archive/2011/020611.html>, 2011).

Bulut bilgi işlem ve açık kaynak yazılımı, verinin kullanımında en verimli sonuçların alındığı iki teknoloji olarak ortaya çıkıyorlar. Ayrıca bedelsiz olarak sağlanan programlama dili Hadoop ile evvelce süper bilgisayarların gerektiği çok büyük veri analizlerini şimdi PC'ler ile yapma olanağı işlem giderlerinde ve zamanda büyük tasarruf sağlıyor. Örneğin kısa bir süre önce New York Times 1851 ile 1922 yılları arasında arşivlerinde toplanan 400 binden fazla malzemenin indirme ve taranması işlemini bulut işlem ve Hadoop kullanarak mevcut PC'ler ile 36 saat içinde gerçekleştirdi. Öte yandan kredi kartı şirketi Visa iki yıllık 36 TB ve 73 milyar kaydın geleneksel yöntemlerle geçmişte 30 gün süren işlemini Hadoop ile 13 dakikada tamamladı.

Google'ın eski CEO'su Eric Schmidt'e göre insan var olduğundan itibaren 2003 yılına kadar sadece 5 EB'lık bir belleğe sığacak kadar bilgi üretebildi. Şu anda 2,4 milyar insan internete erişebiliyor. Bu sayı yıl bitmeden 2,7 milyara ulaşacak. O da neredeyse her iki kişiden biri anlamına geliyor.

2011 yılında oluşan toplam veri miktarının 1,8 milyar TB tutarında olduğu tahmin edilmektedir. Aslında dünyadaki tüm verilerin %90'ı bugün en az 2 yıllıktır. Açıkçası teknoloji, makine üretimi ve insan üretiminden (sosyal medya gibi) gelen veriden daha fazla potansiyel kaynaklar oluşturdukça, veri miktarı üstel katsayı oranlarında büyümektedir. Son yıllarda, “Büyük Veri” kavramı popüler ve iş odaklı yayınların yanı sıra akademik çalışmalarda da ön plana çıkmaktadır.

Bu olguya yönelik belirleyici bakış açılarını gösteren çeşitli tanımlar çıktığı için, Büyük Veri başlangıçta Doug Laney (VP, Information Innovation and Strategy, Gartner) tarafından önemli ölçüde daha yüksek “Hacim, Hız ve Çeşitlilik” e sahip veri olarak tanımlanmıştır. Örneğin, klasik veri yönetim araçları ile daha fazla yönetilemeyen, artan ve daha büyük miktardaki veriden dolayı (yaratıcı çözümleri gerektiğinden ya da yeni programlar), “teknoloji sağlayıcıları” ve işletmeler için “dijital çözüm tedarikçileri” Büyük Veriyi tanımlaması mümkün olmaktadır.

Elbette, bu bakış açısı Büyük Veriyi, kuruluşlar ve şirketlerin gelecekte uğraşması gerekecek olan bir “problem” ya da bir “engel” olarak çerçevelemektedir. Ancak, Büyük Veri verimliliğin yanı sıra firma gelirlerini artırma potansiyeline çok fazla sahiptir, hatta bazı yorumcular Büyük Veriyi “devrimsel gelişme” olarak adlandırmıştır. Örneğin, Kaliforniya Üniversitesi Bilgisayar Bilimcisi “verinin endüstriyel devrimi” adını vermiş ve Microsoft Araştırma ve Strateji Başkanı, gelecekte tüm ekonomiler “veri etrafında şekillenme” yeteneğine sahip olacak ve bu “toplumsal ve hatta makroekonomik seviyede büyük değişimler” oluşturacağını ifade etmiştir (Economist 2010). Bazı uzmanlar da bir taraftan onun öneminden bahsederken, diğer taraftan “aşırı-aldatıcı” olabilmesine karşı uyarılmaktadırlar.

Örneğin, araştırmacılar Boyd ve Crawford (2011, p. 1 - 3) şu şekilde ifade etmişlerdir: O ilgisiz ve anlamsız bir şekilde bir araya gelmiş veriler arasındaki bağlantıyı bir anda görebilme uygulamasını teşvik eden veri türüdür: Aslında var olmayan kalıpları görme, sırf bu yüzden büyük miktarlardaki veriler tüm yönlere yayılan bağlantılar sunabilirler. Aynı zamanda, Büyük Verinin kendisi birçok yönden, zayıf bir terim olsa bile, Büyük Veri Çağının başladığını kabul etmektedirler.

Toplanabilen verinin evrimi birçok işletme için daha fazla zorluk demek olsa da, aynı zamanda büyük fırsatlar da getirmektedir. Elbette, başlangıçta sadece üretim hattına odaklanan işletmeler ve dolayısıyla bir kitle karmaşık veri analizini ihtiyaç duymadı, fakat işletmeler ve küresel ekonominin boyutu büyüdükçe, karmaşık veri onunla anlaşılır bir şekilde büyüdü. Veri Yönetimi böylelikle birçok dalgalanma içerisinde büyüdü ve öncekilerin üzerinde yapılan ve eskisinin yerine geçen ardı ardına veri yönetim teknikleri gelişti. Böylece, yeni Büyük Veri Analizi Teknolojileri ve Teknikleri Dalgası, önceki teknoloji ve tekniklerin her biriyle bütünleşmiş ve birleşmiş oldu. Bu nedenle Büyük Veri basitçe bir “yazılım büyütecisi” ile değil, bütünsel bir perspektiften incelenmelidir.

Bir başka deyişle, Büyük Veri basit teknik bir konu değil, ayrıca bir yönetim konusudur; yalnızca teknik yönetim gerektirmeyen yeni veri dünyası ve sınırlandırılmış, fakat potansiyel kullanımı ve gücünün nasıl dizginleneceği üzerine kararlara yoğunlaşmıştır. Potansiyel müşterilerin nasıl kazanılacağı üzerine daha çok çalışılmış kararları oluşturmada şirketlere imkân vereceği için, birçok eleştirmen pazarlama sektöründe Büyük Verinin potansiyeli hakkında yazmıştır.

Büyük Veri ile ilgili akademik literatür arttıkça, Büyük Veri hakkında popüler basında yayınlanan makalelerin (IT profesyonelleri ve işletme yöneticilerine yönelik hedeflenmiş veya genel okuyucuya yönelik) sayısı da artmaktadır. Bu makalelerin çoğu araştırmacı Boyd ve Crawford tarafından tanımlandığı gibi, Büyük Veri için birkaç “gizli gerçekleri” veya “kışkırtmaları” göstermeyi ihmal eder. Örneğin: “Daha Büyük Veri daima en iyi veri değildir” veya “Tüm veriler eşit değildir” savı (2011, p. 7 – 8).

Bilhassa işletme yöneticileri, Büyük Veri ile uğraşırken bu karmaşık gerçekleri göz ardı edebilirler. Bundan dolayı, bu çalışma daha öncesine göre gitgide artan daha Büyük ölçeklerde tanımlanan Büyük Veriyle, işletme yöneticilerinin, müdürlerinin, IT profesyonellerinin nasıl adapte olacağını, üstesinden geleceğini, yüzleşeceğini göstermeyi de amaçlamaktadır. Görüşülen kişiler bu araştırma için örnek bir temsilci olarak seçildi. Yarı-yapılandırılmış röportajlar, Büyük Verinin kurumların yapısını veya iç operasyonlarını ne derece değiştirdiği ve Büyük Verinin varlığına nasıl adaptasyon sağlandığının sorulduğu bu kişiler ile yürütüldü.

Olgusal bir yaklaşım kullanan bir çalışma olarak, Büyük Veri ile ilgili katılımcıların deneyimleri, yargı ve görüşleri anlamındaki amaçlar araştırmanın ana bir bileşenini biçimlendirir. Bu çalışma literatür taraması olan genel kısımlar, nitel olgusal yaklaşımın kullanım nedenini açıklayan metodoloji bölümü, bulgular ve sonuç bölümü olarak dört ana bölüme ayrılmıştır.

2. GENEL KISIMLAR

1980’lerde ürünün çok daha önemli olduğu zamanlarda, şirketlerin asıl amacı belirli bir ürünü üretmek ve müşteriye ulaşımını sağlamaktı. Bu yıllarda ERP sistemlerinin ön planda olduğunu görüyoruz. ERP (Enterprise Resource Planning – Kurumsal Kaynak Planlaması) sistemlerinin geliştirilmesindeki asıl amaçlardan bazıları; müşteri, dağıtım merkezi, tedarikçiler ve üretimi bir platformda toplamaktı. Bir zamanlar çok popüler olan bu sistemde doygunluğa ulaşıldığında insanlar şu soruyu sormaya başladılar: “Benim için doğru müşteri kimdir?”

CRM sistemlerinin doğuşu da bu soruyla başladı denilebilir. CRM’nin (Customer Relationship Management – Müşteri ilişkileri Yönetimi) asıl ilgilendiği nokta “Doğru ürün ya da hizmeti, doğru müşteriye, doğru fiyatla, doğru kanalda, doğru yerde ve zamanda sunmaktır”. Yani artık ürüne göre müşteri değil, müşteriye göre ürün devri başlamıştır. Son 10 yıldır yükselen değer olan bu yöntem giderek önemini arttırmaktadır. Veri miktarı oluşumundaki hızlı artışın temel nedeni ise sosyal ağ etkileşimlerinin büyüyen hacmi, lokasyon duyarlı cihazların artması ve fiziksel dünya hakkında bilgi yakalayan ve ileten "akıllı sensörlerin" sayısındaki artış olarak özetlenebilir. Tabi ki bunlara video ve medya kaynakları da eklenebilir.

İçinde bulunduğumuz “Büyük Veri” çağında yatırımcılar, teknoloji girişimcileri, medya ve danışmanlık şirketleri Büyük Veri konusuna odaklanarak, yeni fırsatlar yakalıyor. Bulut barındırma çözümlerinin basitleşip ucuzlayarak genele yayılması, veri işleme konusundaki ekonomik dengeleri temelden değiştirdi. Yakın geleceğin en önemli teknoloji piyasasının Büyük Veri üzerinde oluşması ve önümüzdeki yıllarda bu pazarın 50 milyar doları aşması bekleniyor. Dünya çapında yıllık veri hacmindeki büyüme %59 ve büyümenin artarak devam etmesi bekleniyor. Bu büyümenin merkezinde hem geleneksel hem de yeni veri kaynakları yatıyor. IDC firması dijital kayıtların 2014 sonunda 1.2M ZB’a ulaşacağını, önümüzdeki on sene içinde de 44 katına çıkacağını tahmin ediyor. Cisco’nun raporuna göre, tabletler, mobil telefonlar, ağa bağlı cihazlar ve akıllı makinelerin sayısı 2015’te 15 milyara ulaşacaktır. 2010’da PC’ler tüketici internet trafiğinin yüzde 97’sini oluştururken, bu oran 2015’te yüzde 87’ye düşecek, bu çaprazlamaya karşın küresel mobil internet veri trafiği ise 2010-2015 arasında 26 katına

çıkacaktır (<http://www.cisco.com/web/TR/news/press/archive/2011/020611.html>,2011). Büyümenin asıl kaynağı yapısal olmayan verilerden gelmektedir. Yapısal olmayan verilerin yaklaşık %80'nin değersiz olduğuna dair önyargı ise gerek arama motorlarının gerekse de e-ticaret yapan kurumların tıklama verisini takip ederek ulaştıkları başarı sonrasında çürütülmüş görünüyor. Asıl gereksinim ise yapısal ve yapısal olmayan verinin saklanması, beraberce analiz edilerek, veri madenciliği işlemlerine tabi tutulması olarak ortaya çıkmaktadır.

İşte bunlardan dolayı çağımızda büyük önem kazanan “Büyük Veri” ile ilgili literatür bilgisinin inceleneceği bu bölümde, genel olarak veri kavramına bakış, Büyük Veri nedir, gelişimi ve veri yönetimi, zorlukları ve sunduğu fırsatlar, kuruluşların Büyük Veriye uyum sağlaması, insan unsuru, verinin değeri ve yeni modeller gibi konulara değinilecektir.

Büyük Veri ile ilgili kavramlar ve tanımlamalar olabildiğince netleştirilecek ve bu yeni olgunun tam olarak ne anlam ifade ettiği ortaya konulacaktır.

2.1. BÜYÜK VERİ

2.1.1. Büyük Veri Kavramı

Geçmişten günümüze kadar gelen süre içerisinde bilgi günden güne artmıştır, hatta son yıllarda katlanarak günümüze gelmiştir. Bunun sonucunda "Bilgi Çöplüğü" diye tabir ettiğimiz olgu gün yüzüne çıkmıştır. Birçok yazılım şirketi bu konuyla ilgili Ar-Ge çalışmaları yaptıktan sonra “Big Data (Büyük Veri)” olgusu karşımıza çıkmıştır.

“Big Data” yani “Büyük Veri” terimi, karmaşıklığı ve boyutundan dolayı geleneksel veri yönetim teknikleri ile yönetilememesi ve işlenememesinden dolayı, çok büyük miktarda veriyi tanımlamak için, bilgisayar dünyasına ilk olarak 2005 yılında O'Reilly Media'dan Roger Magoulas tarafından tanıtıldı.

Bir Araştırma ve Bilimsel Konu olarak Büyük Verinin Evrimi üzerine bir çalışma şunu gösterdi ki, “Big Data” üzerine 1970'lerde başlayan araştırmalar mevcuttu. Fakat 2008 içerisinde yayınlarda görünmeye başladı. Günümüzde Büyük Veri konsepti, birçok alanlardaki uygulamalarda farklı şekilde ele alınmaktadır.

Veri, kişisel bilgi parçaları tarih boyunca toplanmış ve kullanılmıştır. Son günlerde değişen şey ise veri toplama, depolama ve analiz etmede belirgin bir şekilde artan yeteneklere sahip dijital teknolojilerdeki ilerlemelerdir.

1880 yılında ABD’de farklı yaş, cinsiyet, ev hane sayısı, etnik köken, doğum tarihi, medeni durum, meslek, sağlık durumu, eğitim ve doğum yerini içeren demografik bilgilerin toplandığı 50 milyon insanın nüfusu sayımı gerçekleştirildi. Tüm bu bilgiler bir elde toplandı, incelendi ve devlet arşivlerine, kütüphaneler ve üniversitelere depolanmak için gönderildi. İlk toplamadan sonra düzgün bir şekilde tablolara dökülmesi 7-8 yıl sürdü.

1890 yılında, Sayım Bürosu bir yıl içinde tablo haline getirilebilen makine-okumalı zımba kartlarını benimseyerek veri toplama metotlarını yayınladı. En son dönemde ABD sayımı 2010 yılında yapıldı ve sayım bürosu halk ile anket yapmak için coğrafi bilgi sistemleri, sosyal medya, videolar, akıllı karakter tanıma sistemleri ve gelişmiş veri işleme yazılımları dâhil yeni gelişen bir dizi teknolojiler kullandı.

Bugün, Büyük Veri tek veri depolarının (veri tabanları ya da veri ambarları) çok ötesinde veri setlerini ifade etmek için kullanılmaktadır. Geleneksel veri yönetim ve işleme araçları tarafından işlenmek için çok büyük ve çok karmaşıktır. Büyük Veri işlemsel veriler, sosyal medya, kurumsal içerik, sensörler ve mobil cihazlar gibi bilgileri kapsayabilirler.

MIKE 2.0 (Veri Yönetimi İçin Açık Kaynak Standardı)’na göre, Büyük Veri birbiriyle etkileşimli, çok büyük, karmaşık ve bağımsız veri setleri ve boyutu tarafından tanımlanmaktadır. Ek olarak, Büyük Verinin önemli bir yanı, geleneksel veri yönetimi ile ele alınamaması gerçeğidir.

O’Reilly Strata Konferansındaki Ed Dumbill’e göre, Büyük Veri “geleneksel veritabanı sistemlerinin işlem kapasitesini aşan veri” olarak açıklanabilir. Veri çok büyüktür, çok hızlıdır, sizin veritabanı mimarisinin yapılandırmasına sığdırılmaz. Bu veriden değer elde etmek için, onu işlemek için alternatif yollar seçmeliyiz.

Daha basit bir tanımlamayla, Büyük Veriyi işletme süreçleri için belirli metotlar ve teknikler çok büyük, çok karmaşık, yapılandırılmamış, organize edilmiş, depolanmış ve işlenmiş farklı veri setlerinden oluşmaktadır ifadesiyle dikkate alabiliriz.

Büyük Verinin dünyada birçok dolaşan tanımı vardır, fakat her şirket kendisi önem verdiği şekilde tanımlamaktan yanadır. Her lider şirket rekabet avantajı getirmek için bir kavram tanımlamak zorundadır.

Büyük Veri; toplumsal medya paylaşımları, ağ günlükleri, bloglar, fotoğraf, video, günlük dosyaları vb. gibi değişik kaynaklardan toparlanan tüm verinin, anlamlı ve işlenebilir biçime dönüştürülmüş biçimine denir.

Pennsylvania Üniversitesi Ekonometri profesörlerinden Francis X. Diebold'un özet çalışmasında (2012) ifade ettiği üzere 'Büyük Veri' kavramı, 1990'ların ortalarından itibaren üretim yönetimi, bilişim, istatistik, ekonometri gibi farklı alanlarda analiz edilen veri miktarının büyüklüğüne dikkat çekmek için kullanılmaya başlamıştır. Kavramın akademik literatüre girişi ise S. M. Weiss ve N. Indurkha'nın "Predictive Data Mining: A Practical Guide" (1998) başlıklı bilişim sistemleri ve F. X. Diebold'un "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting" (2000) ekonometri konulu çalışmaları sonrasında mümkün olmuştur. Akademik literatürde 'büyük veri' kavramının çağrıştırdığı "algılayıcılardan ve bilimsel araçlardan büyük hacimde, yüksek çeşitlilikte ve hızla gelen verilerin toplanması, saklanması, temizlenmesi, görselleştirilmesi, analiz edilmesi ve anlamlandırılması" eylemidir (Gürsakal, 2013, s. 21).

En basit tanımıyla ise Büyük Veri (Big Data), gerek insan gerekse makineler tarafından sayısal olarak kodlanmış her türden kurumsal veri ile internet ve sosyal medya paylaşımları aracılığıyla ortaya çıkan kişisel verilerin anlamlı ve işlenebilir biçime dönüştürülmesi durumudur. Bu kapsamda halen yapılandırılmamış enformasyon olarak duran ve değersiz (çöplük) olarak algılanan veriler arasından, sosyal medyanın kamuya açık API'leri aracılığıyla, sayısal platformların kullanıcılarından elde ettikleri tüm bilgilere ulaşılabilir. Uygulama Programlama Ara yüzü anlamına gelen API (Application Programming Interface), herhangi bir uygulamanın belli işlevlerini diğer uygulamaların da kullanabilmesi için oluşturulmuş bir modüldür. Örneğin Google

AdWords Uygulama Programlama Ara yüzü (API), teknik konularda birikimi olan büyük reklam verenler ile üçüncü tarafları temsil eden geliştiriciler için tasarlanmıştır. Bunlar arasında, ajanslar, arama motoru pazarlamacıları ve birden çok müşteri hesabını veya büyük kampanyaları yöneten diğer çevrimiçi pazarlama uzmanları yer almaktadır. Google AdWords API'sı, geliştiricilere, doğrudan Google AdWords sunucusuyla etkileşim kuracak uygulamalar oluşturma olanağı vermekte; reklam verenler ve üçüncü taraflar da bu uygulama sayesinde büyük veya karmaşık AdWords hesaplarını ve kampanyalarını daha etkin ve yaratıcı biçimde yönetebilmektedir.

Olageldiği gibi, ilişkisel veri tabanlarında tutulan yapısal verinin dışında kalan, son dönemlere dek çok da kullanılmayan, yapısal olmayan veri yığınıdır. Artık yıkılmış olan yaygın bilişimci inanışına göre, yapısal olmayan veri, değersizdi, ama büyük veri bize bir şey gösterdi. O da günümüzdeki bilgi çöplüğü diye adlandırılan olgudan muazzam derecede önemli, kullanılabilir, yararlı yani çöplükten hazine çıkmasına neden olan yegâne sistemdir.

Büyük Veri; web sunucularının günlükleri, internet istatistikleri, sosyal medya yayınları, bloglar, mikrobloglar, iklim algılayıcıları ve benzer sensörlerden gelen bilgiler, GSM operatörlerinden elde edilen arama kayıtları gibi büyük sayıda bilgidir oluşuyor.

Büyük Veri, doğru analiz metotları ile yorumlandığında şirketlerin stratejik kararlarını doğru bir biçimde almalarına, risklerini daha iyi yönetmelerine ve inovasyon yapmalarına imkân sağlayabiliyor.

Şirketlerin çoğu, halen geleneksel (konvansiyonel) veri ambarı ve veri madenciliği yöntemleriyle elde ettikleri verilerden yola çıkarak, karar almaya devam ediyorlar. Ancak, tüketici eğilimlerini dinamik şekilde öngörebilmek, büyük veriyi analiz edebilmekten ve bu analizlere göre hareket edebilmekten geçiyor.

Büyük Veri, geleneksel veri tabanı araçları ve algoritmaları ile işlenmesi zor olan bu devasa verinin oluşturulması, saklanması, akışı, analiz edilmesi gibi birçok konuyu içeren bir terim olarak karşımıza çıkmaktadır.

Aynı zamanda Büyük Veri kendi içinde bir kavram ve bir konu olmaktan çıkarak, veri bilimi, veri entegrasyonu, analiz ve bilgi inovasyonu gibi mevcut kullanımlara bir ek değer olmaya doğru ilerlemektedir.

Veriler klasik veri tabanlarının kaldıramayacağı büyüklükte olduğu gibi verinin büyüme hızı da bir bilgisayar veya bir veri depolama ünitesini aşmaktadır. 2012 rakamları ile dünyada günlük 2,5 Kentilyon (10^{19}) bayt veri üretilmektedir. Bu çapta büyük veriyi işleme, transfer etme gibi işlerin tümüne “Büyük Veri (Big Data)” adı verilmektedir.

İşte bu nedenlerle birçok alanda Büyük Veri bir ihtiyaç olarak karşımıza çıkmaktadır.

2.1.2. Büyük Veri Kaynakları

Büyük Veri pürüzlü verinin farklı tiplerinden oluşmaktadır. Bunları şu şekilde inceleyebiliriz.

2.1.2.1. Kamusal Veri: Public Data

Geniş çaplı iş ve yönetim uygulamaları için potansiyel olarak kullanılabilen ve devletler, devlet kurumları ve yerel topluluklar tarafından tutulan veridir. Bireysel gizliliğimizi korumak için belirli kısıtlar altında erişilen bu verilere örnek olarak ulaşım, enerji kullanımı ve sağlık sayılabilir.

2.1.2.2. Özel Veri: Private Data

Özel şirketler, kar amacı gütmeyen kuruluşlar ve bireyler tarafından tutulan verilerdir. Örneğin, müşteri işlemleri, örgütsel tedarik zincirleri tarafından kullanılan radyo frekans tanımlama etiketleri, şirket mal ve kaynaklarının hareketi, web sitesi tarama hareketleri, mobil telefon kullanımı bunlardan bir kaçıdır.

2.1.2.3. Çevresel Veri: Data Exhaust

Bu veri tipi pasif bir şekilde toplanan ortama ait çevresel veriyi anlatır. Veri toplanan için sıfır değerinde ya da bir merkezi olmayan veridir. Bu veriler farklı amaç için toplanmış fakat yeni değerli kaynak oluşturmak için başka veri kaynaklarıyla yeniden kombine edilebilirler. Bireyler yeni teknolojiler benimsediği ve kullandığı zaman (ör. mobil telefonlar), bu cihazların günlük faaliyetleri ile ortamsal veri üretirler. İnsanlar günlük yaşamlarına devam ederken pasif bir şekilde farkında olmadan veri yayıyor da olabilirler (mesela, alışveriş yaparken, sağlık hizmetine erişirken, başkalarıyla etkileşim

halindeyken). Bu veri türü için bir başka kaynak insanların ihtiyaçlarını, arzularını ve niyetlerini anlamaya yarayan bilgi arama davranışıdır. İnternet aramaları, telefon görüşmeleri ya da özel çağrı merkezleri bu türden veri üretmektedir.

2.1.2.4. *Topluluk Verisi: Community Data*

Sosyal eğilimleri yakalamak için dinamik ağların içerisindeki yapılandırılmamış verinin – özellikle metin – özüdür. Tipik bir topluluk verisi müşterilerin ürün incelemelerini, oylama butonlarını, twitter bildirimlerini ve bunun gibi başka şekilleri içermektedir. Bu veriler sosyal yapıdaki kalıplardan damıtılmış bir sonuç çıkarmak anlamına gelebilir (Kennedy, 2008).

2.1.2.5. *Kişisel Veri: Self-Quantification Data*

Kişi eylem ve davranışlarının belirlenmesi yoluyla kişiler tarafından ortaya çıkan veri türüdür. Örneğin, bu verinin bir şekli, egzersiz ve hareket izleyen bilekliklerle sağlanan veridir, bu veri bir mobil uygulamaya yüklenir ve kişi takip edilmiş olur. Psikolojide, kişinin durumuyla ilgili bilgi edinilmesi o kişinin tercihlerini ifade eder. Örneğin, ışıkları uzun süre açık tutmak yerine tasarruf amacıyla kapatan bir kişi enerji tasarruflu ampul alabilecektir. Bu öz-miktar verisi psikoloji ve davranış arasındaki bağlantıya köprü oluşturmaya yardım eder. Farklı alanlardaki sosyal bilimlerin uzmanları psikoloji, pazarlama ya da kamu politikası gibi bu alanlardaki araştırmalar için bu veri türünden yararlanırlar.

2.1.3. **Veri Paylaşımı, Gizlilik ve Etikler**

Toplumsal farkındalığımız artarak, yeni bilgi toplama ve analiz şekillerinin olduğu bir zaman diliminde yaşıyoruz. Büyük Veri Devrimi insanlarla ilgili olan birçok veriyi kaydediyor. Hepimizin kişisel bilgi gizliliği ile ilgili var olan yönetim kuralları bulunmakta, fakat yeni veriden elde edilen yeni bilgi akışları, yeni kullanım şekilleri ve yeni kararlar bu kuralları eksik hale getirmektedir. Bu yeni çağın reddedilemez faydalarını kurban etmeden bu yeni araçların toplumsal maliyetlerini düzenleyen yeni kurallara ihtiyacımız vardır.

Fakat bu yeni kuralları şekillendirmede bize hangi değerler rehberlik edecektir? Gizlilik, şeffaflık ve kimlik değerleri belki de bu konudaki başlangıç noktası olacaktır.

Genelde “gizlilik” başlığı altında kişisel gizlilik problemlerini düşünürüz. Fakat Büyük Veri Devriminin “gizliliğin ölümü”nü işaret etmesi gerekmez. Aksine, şu anki bilgi akışı yönetim kurallarının yerine daha korunan sırlar düşünürüz ve tanımlarız. Bu durum Büyük Verinin gerçek gizlilik konusunu odak haline getirmektedir. Bizim veri akışını düzenleyen kurallara ihtiyacımız vardır ve bunun anlamı şudur; kişisel verinin toplanması gizlilik tartışmalarının sonu değil başlangıcı olmalıdır.

Son dönemde teknoloji şirketlerinden bazıları güven restorasyonu için çağrıda bulundular. 9 Ekim 2013'te Başkan Barack Obama ve Kongreye açık mektup yayınlayan Apple, Google, Microsoft, Facebook, Yahoo, LinkedIn, Twitter ve AOL gözetim ve düzenleme reformu için bir çağrıda bulundu. Microsoft Genel Danışmanı Brad Smith şunu yazmıştır; “İnsanlar güvenmedikleri teknolojiyi kullanmak istemeyecektir. Hükümetler bu güven riskini oluşturdular ve onu tamir etmek için yardımcı olmalıdırlar”. Yahoo'nun CEO'su Marissa Mayer de “Hükümetlerin gözetim aktiviteleriyle ilgili son günlerdeki ilişkiler kullanıcılarımızın güvenini sarstı, ABD Hükümeti için dünya çapında halkların güvenini yeniden sağlamak için tam zamanıdır” dedi. Evet, bu teknoloji sağlayıcıları reform için çağrıda bulunmaktadırlar. Çünkü eğer müşterilerinin kendileri ile paylaştıkları özel bilgilerin artık güvenilirliği olmaz ise, müşteriler onların hizmetlerine karşı olan güvenlerini yitireceklerdir ve bu da firmaları korkutmaktadır.

Gizlilik gibi şeffaflık konusu da güveni besleyen bir unsurdur. Hükümetlerin bilgi şeffaflığı devlet kademeleri, özgür bir basın ve bireysel haklar arasında bir denge kurar ve kontrol eder, anayasal denetim içinde kritik bir rol oynar. Finansal raporların şeffaflığı özellikle para ve hisse senetleriyle ilgili yatırımcıların isteğini artırır. Örneğin Avrupa Birliği Veri Koruma Yönergesi hâlihazırda şeffaflığın korunmasını sağlamaktadır. Ira Rubenstein, Doc Searls ve arkadaşları da şöyle bir geleceği açıklar; ilave şeffaflık korumaları müşterilerin kişisel verileri üzerinde kontrole izin veren yeni iş modellerini destekleyecek veri taşınabilirliğine izin verecektir.

Şeffaflık doğal şekilde açıklık ve gizlilik arasında bir gerilim içerir. Bu gerilim paradokslara neden olabilir. Devlet ve şirketlerin hassas şeffaflık sırları ulusal güvenlik ve ticaret sırları gibi konulara önemli zararlar verebilir. Çok az şeffaflık güven eksikliğine ve beklenmedik sorunlara yol açabilir. Şeffaflık da bir risk taşımaktadır ki, dikkatsiz ifşalar zararlı gizlilik ve gizlilik ihlali gibi beklenmedik sonuçlara neden olacaktır.

Büyük Veri kimlikle ilgili olarak da çok daha derin bir şekilde düşünmemizi gerektirir. Kimlik de gizlilik gibi tanımlanması zor fakat aynı derecede korunması hayati bir unsurdur. Kimlik birçok anlama gelebilir, özgün bir kişiye özgün bir ismi çağrıştırabilir. Gerçekten de bütün kimlik yönetimi ve koruma endüstrilerinde bu kimlik koruma türü vardır. Kimlik ayrıca hukuksal bir kanıt olarak, bir şeyin başka birisi ya da bir şeyle aynı olup olmadığı anlamına da gelebilir. Filozoflar uzun dönem bunu tartıştı ve bu tarzdaki kimliği tanımlamaya uğraştılar. Tüketiciler olarak, kimliklerimiz artan bir şekilde Büyük Veri ve şirketlerinde kontrolünde şekil almaya başlıyor. Birçok bakımdan ise, bu kontrole bizim ihtiyacımız var ve bunu istiyoruz.

Evet, bugün Büyük Veri ve diğer yeni bilgi teknolojileri ile ilgili etikleri düşünmeye ve garanti altına almaya ihtiyacımız vardır. Bu teknolojiler “doğal” ve önceden kararlaştırılmış değildir; insan seçimlerinin ürünüdür ve insani değerleri etkileyeceklerdir. Sahip olmak istediğimiz toplum türünü şekillendiren bu teknolojilerden emin olmak zorundayız. Peki, bunu nasıl yapmalıyız? Hukuksal olarak, buna yeni hukuk kuralları oluşturarak adım atabiliriz. Aslında hali hazırda zaten birçok veri işleme yönetimi kuralı bulunmaktadır. Fakat bunlar bizi yeterli derece korumasa bile hala amacına uygundur. Bununla ilgili olarak hukuksal konular uzmanlarınca tartışılmaya ve yeni alternatifler ortaya atılmaya devam etmektedir.

Şunu diyebiliriz ki, Licklider’in ön görüşüyle “insanlık tarihinin en heyecan verici ve en yaratıcı entelektüel” döneminde yaşıyor olabiliriz. Diğer yeni bilgi teknolojileri gibi, Büyük Veri insanlık için keşif ve inovasyon çağında şaşırtıcı imkânlar sunar. Savunma konusunda yapılacaklar için Büyük Veri kullanımında devlet yetkililerine izin vermemiz gerekmektedir. Büyük Veriden daha iyi faydalanmak için şirketlerle bilgilerimizi paylaşmak isteriz. İşte bu gibi sebeplerle geniş çaplı bir şekilde Büyük Veri ile ilgili düşünmemiz gerekmektedir. Ve Büyük Verinin yeni imkânlarını desteklemek için oluşan, gizlilik, şeffaflık ve kimlik gibi önemli sosyal değerleri korumak için gizlilik değerlerini, normlarını ve hukuki korumaları geliştirmeliyiz.

Büyük Veri gizlilik, güvenilirlik ve kimlik için kesin bir tehdittir, fakat bu hukukun öldüğü anlamına gelmez. Toplumsal bilgi akışını yöneten kurallar hem ana esastır hem de bir şekilde kaçınılmazdır. Bilgi Devrimi boyutu ne tür kurallar istediğimizle ilgili daha

çok hayal gücü ve daha geniş bir şekilde düşünmemiz anlamına gelir. Toplumsal bilgi akışındaki kişisel bilgiyi garanti eden kuralları geliştirmeye ihtiyacımız vardır.

Geçerli bilgi teknolojileri altyapılarında, ağ bağlantısı gibi servis sağlayıcılar sağladıkları hizmetin niteliği ve kalitesini tanımlayan SLA (Service Level Agreement: Hizmet Seviye Anlaşması) ile bağlıdır. SLA sorumluluğu kısıtlamak için önemlidir. Ağ bağlantıları ve web hizmetlerinin üstel olarak genişlemesi, sözleşme ile uyumluluğun izlenmesi ve onaylanması açısından SLA uygulama otomasyonundaki teknolojik gelişmeler önemliydi. Buna karşılık, Büyük Veri paylaşım anlaşmaları gayri resmi, yetersiz yapılmış el yordamıyla uygulanan ve soyutlanmış işlemlere bağlı kalır (Koutroumpis ve Leiponen, 2013). Bu durum özellikle kamu kaynaklarıyla entegrasyonu için özel veriye erişemeyen sosyal bilim ve yönetim araştırmaları adına önemli bir engeldir. Haklar konusundaki çalışmasında Sart; katılımcılar araştırma sonuçlarının yayımlanmasını isterken hak sahipliğinin de korunmasını talep etmektedirler. Burada katılımcıların karşı karşıya kaldıkları önemli bir durumda ortaya çıkmaktadır. Pek çok araştırmacı akademik kariyerlerinde ilerleyebilmeleri için gerekli yayını yapmak zorunda kalırken pek çok fikirlerinin korunamaması nedeniyle patent ve lisanslama haklarını kaybetmektedirler (Sart, 2013).

Veri paylaşım anlaşmaları aynı zamanda, açık veri, erişim kontrolü, hak yönetimi ve veri kullanım kontrolü için anonimleştirme içererek, veri koruma ve gizliliği için mekanizmalar ile bağlı olması gerekmektedir. Farklı kaynaklardan elde edilen veri sonucuyla bireysel kimliklerden atfedilen konular dikkatli bir şekilde göz önünde bulundurulacak ve açık bir şekilde kabul edilip izin verilmiş olacaktır. Bunu yapmak muhtemelen bize veri paylaşımı ve veri hakları için bir model düzenlemeye izin verecek ve bu gelecekte evrensel yararlar sağlar ve büyük veri işbirlikleri tanımlayabilir. Örneğin fikri mülkiyet hakları ile ilgili yapılan bir araştırmada; katılımcıların % 91'i elde edilen gelirin ne şekilde paylaşılacağına ilişkin net, şeffaf ve anlaşılabilir düzenlemelerin olmasının araştırmalardaki motivasyonu arttıracığını ve çok daha fazla katılımcının bu tür aktivitelerde yer alacağını vurgulamaktadır (Sart, 2013).

Büyük Verinin temel etiklerini geliştirerek, potansiyel zararlarına yenilmeden yeni teknolojinin faydalarından yararlanmaya çaba göstermeliyiz. Büyük Veri etikleri herkes için gereklidir.

2.1.4. Büyük Veri Teknolojisi

2.1.4.1. *Büyük Veri Teknolojisine Giriş*

Bugünkü anlamıyla Büyük Veri kavramının kökleri 2004 yılına dayanmaktadır, yani 2004 yılından beri dünyanın en büyük teknoloji devlerinin yatırım yaptığı ve kullandığı bir kavramdır. Büyük Veri yeni bir olgu, yeni bir kavramdır ve bünyesinde açık kaynak kodlu yazılımlar barındırır. Bu yazılımların çekirdeğini Hadoop ismini verdiğimiz bir dizi yazılımlar grubu kapsar. Bu yazılımlardan bazılarını şöyle listeleyebiliriz; Hadoop Core, HDFS (Hadoop Distributed File System – Hadoop Dağıtık Dosya Sistemi), Hive (Data Warehouse – Veri Ambarı), HBase, ZooKeeper, Oozie, Mahout, Sqoop, Cloudera Manager. Büyük Veri bun ve benzeri araçlarla analiz edilerek kullanılabilir sonuçlar elde edilmeye çalışılır.

“Büyük Veri” terimi geleneksel veri sistemleri için çok büyük olan veri setlerinin toplanması, yönetimi ve analizini mümkün kılan teknolojileri açıklamak için kullanılmaktadır (Dumbill 2012). Son zamanlarda iş dünyası yeni karar alma olanaklarına imkân tanıyan Büyük Veri araçlarını tartışmaktadır (Provost and Fawcett 2013). Var olan veri tabanı sistemlerinin sınırlarını ele almak için, Büyük Veri teknolojileri çok büyük ölçekte paralel hesaplama yaklaşımlarını kullanmaktadır. Dağıtık veri işleme uzun bir geçmişe sahiptir (Provost ve Kolluri 1999).

Son yıllarda veri toplama oranı ve ölçeği veri toplama teknolojilerindeki yenilikler ile gelirlerini arttırdı. Yeni veri teknolojilerini en son dönem dalgasının kaynakları 2004 yılında dağıtılmış veri işlemeyi desteklemek için büyük veri algoritmaları kullanmaya başlayan Google çalışanlarının izlerini taşıyabilir. Büyük Veri analizleri için yaygın bir şekilde kullanılan yazılım platformu Apache Hadoop Map/Reduce iskeletinden çıkartılmış, Java programlama dilinde uygulanmıştır ve özgür bir şekilde açık kaynak altında dağıtılmaktadır. Bu açık kaynak projesi Hadoop kümelenme ara yüzünün, iletişiminin ve işlem akışının farklı bölümlerini ele alan Cassandra, Pig, Hive, HDFS gibi birçok alt projeye sahiptir. Büyük Veri altyapısı bilgisayar kümeleri üzerindeki veri çevresi ve bu yazılımın uygulamasını gerektirmektedir. Çünkü Büyük Veri işlemlerini desteklemek için gereken hem donanım hem yazılım şirketler için mümkündür. Buradaki en masraflı kısım Büyük Veri sistemlerini uygularken kurulum, bakım ve işi kolaylaştıran uzman edimimidir.

Büyük miktarda verinin analiz edilmesi için oluşturulan raporların çalışma süresi çok vakit almaktadır. Bu sorunun ortadan kalkması için yazılım ve donanım üreticileri bir araya gelerek çözümler üretmeye başladılar. SAP'nin vermiş olduğu bir örnek bize bir fikir verebilir. Elimizde 2 milyar müşteri hareket verisi olsun. Biz bu veri içinde zamanında ödeme yapmamış olan müşterileri sorgulayan bir rapor oluşturalım. Bu rapor normalde 70 küsür dakika sürerken SAP'nin HANA çözümü ile 13 saniyeye inerek 350 kat hızlanabiliyor. Bu, ölçeklenebilir platformun büyük miktarda veriyi gerçek zamanlı şekilde analiz etmesini sağlıyor.

Eğer programcı bilgiyi hard disk değil de RAM üzerinden alıp işleyebilirse bu hız sağlanabiliyor. SSD'ler hard disk dünyasına yeni bir boyut açtı. Ama büyük veriler için disk teknolojisi yetersiz kalmaktadır. Siz veriyi diskten alıp ram üzerinde tutmanız ve yeri geldiğinde oradan veriyi alıp işlemeniz gerekiyor. Örnek vermek gerekirse Google'ın veri indekslerini bildiğimiz veri tabanı yönetim sistemi ile tutmaya kalkması pek mümkün değildir. Büyük Veri ile veri tabanı sistemleri ve uygulama geliştirme dilleri değişmektedir. Hadoop ve NoSQL veri tabanları günümüzde öne çıkmaktadır.

Yazılım ve donanım şirketleri artık beraber çözümler üretmeye başladı. Bugün Büyük Veri konusuna yazılımsal olarak en fazla çalışan firmalar arasında IBM, donanımsal olarak ise EMC görünmektedir. Büyük şirketlerin Büyük Veri için yapmış oldukları çözümlerden de kısaca bahsederseniz: SAP HP, Fujitsu ve IBM ile beraber geliştirdiği HANA çözümü tanıttı. SAP'nin In-Memory teknolojisi olan HANA (High Performance Analytic Appliance) çözümü ile veri ana bellekte tutuluyor ve alınan raporlar hızlı ve çevrimiçi bir şekilde olabiliyor.

Microsoft HP ile oluşturdukları kutu çözümleri ile özellikle SQL sunucuyu hızlandırmayı düşünüyor. Oracle EMC ile geliştirdikleri çözüm ile özellikle sanal makine üzerinde çalışan SAP sistemlerini hızlandırabiliyor. Ayrıca SUN ile işbirliği yapıyor. EMC bu alanda en fazla yatırım yapan şirketlerdendir. Isilon, OneFs ve Scale-Out Nas çözümlerini örnek olarak gösterebiliriz.

IBM ise bu alanda en öncü şirkettir. Netezza ve IBM DB2 Analytics Accelerator ürünlerini bu kapsamda düşünebiliriz. IBM bu alanda müşterilerine en üst seviyede büyük veri çözümleri sunan firmalardan birisidir.

Ayrıca, Thomson Reuters firması da Büyük Veri alanında çalışmalar yapmaktadır. Hatta 4 Eylül 2015 tarihli blog yazısında şu soruyu sormaktadır: “How will we feed 9 billion people in 2050?” yani “2050 yılında 9 milyar insanı nasıl besleyeceğiz?”. Ve bu konuda <http://9billionbowls.com/> adındaki sitede bir rapor hazırlamıştır. Bunu bir problem olarak ifade etmiş ve cevabı için “...The Answer is Data” yani “Cevap Veridir” diyerek Büyük Veri konusundaki projesini ortaya koymuştur.

Artık veri hacimlerini kontrol altına alan yeni teknolojiler ile –yüksek hızlı veriden faydalanarak kuralcı analizler ve analitik teknolojiler ile kademe kademe genişleyen çeşitli veri tiplerinin analiz edilmesi aracılığıyla– Büyük Veri yeni bir standart olmaktadır. Ya da başka bir deyişle, Büyük Veri her şeyin ayrılmaz bir parçası haline gelecektir.

2.1.4.2. *Büyük Veri Analiz Yazılımı: Hadoop*

Apache Hadoop hızlı gelişen bir büyük veri işleme platformudur. Cloudera'nın Mimarı ve Apache Yazılım Vakfının Başkanı Doug Cutting tarafından geliştirilen Apache Hadoop web'in patlamasından oluşan veri için gereklilikten ortaya çıktı ve veriyi ele alan geleneksel sistemlerin çok ötesinde büyüdü. Hadoop başlangıçta Google tarafından yayınlanan veri çığını ele alan taslak makalelerden esinlenmişti ve o zamandan beri fiilen yüzlerce terabayt hatta petabayt veriyi depolama, işleme ve analiz için standart halini aldı.

Apache Hadoop %100 açık kaynaktır ve temel olarak veri depolama ve işlemenin yeni şekli olarak öncülük etmiştir. Veriyi depolamak ve işlemek için pahalı, kişiye özel farklı sistemlere güvenmek yerine, Hadoop ucuz bir şekilde çok büyük devasa miktarda verinin dağıtılmış paralel işlenmesine izin verir. Büyük Veri analiz sunucuları hem depolama hem işleme üzerine, sınırsız ölçekler sunabilir.

Her gün çok daha fazla verinin oluştuğu hiper-bağlı dünyamızda, Hadoop'un çığır açan avantajları şu anlama gelmektedir: İşletmeler ve organizasyonlar işe yaramaz kabul edilen verideki değeri şimdi bulabilirler.

Hadoop farklı sistemlerden her tipteki veriyi işleyebilir; yapılandırılmış, yapılandırılmamış, günlük dosyaları, resimler, müzikler, iletişim kayıtları, e-postalar. Farklı veri tipleri bağlantısız sistemlerde saklansalar bile, daha önceden bir şemaya ihtiyaç duymadan Hadoop kümeleri içerisinde bunları depolamak mümkündür.

Tüm veri kullanılabilir yapılarak, Hadoop düzenlenmemiş ilişkiler saptamak için destek sağlar ve yanıtları açığa çıkarır.

Ayrıca, Hadoop'un miras sistemler üzerindeki maliyet avantajı veri ekonomilerini yeniden tanımlıyor. Eski sistemler, belirli bir iş yükleri için iyi iken, Büyük Verinin ihtiyaçları için tasarlanmamıştı ve bugünün çok büyük veri setleri için çok pahalıydılar.

Hadoop, sıradan sunuculardan oluşan küme (cluster) üzerinde büyük verileri işlemek amaçlı uygulamaları çalıştıran ve Hadoop Distributed File System (HDFS) olarak adlandırılan bir dağıtık dosya sistemi ile Hadoop MapReduce – büyük veri setleri ile yapılacak işlemlerin birden fazla iş birimine dağıtılmasını sağlayan yöntem – özelliklerini bir araya getiren Java ile geliştirilmiş açık kaynaklı bir kütüphanedir.

Hadoop projesi büyük verileri işleme konusundaki diğer projelere bir çatı görevi görüyor. Hadoop projesinin altında Avro (veri dizileştirme sistemi), Cassandra (yüksek erişilebilir, ölçeklenebilir NoSQL veritabanı), HBase (Hadoop üzerinde çalışan, büyük veriler için ölçeklenebilir, dağıtık NoSQL veritabanı), Hive (büyük veriler üzerinde iş zekası sistemi), Mahout (ölçeklenebilir yapay öğrenme ve veri madenciliği kütüphanesi), Pig (paralel hesaplamalar için yüksek düzeyli bir veri akışı dil ve yürütme kütüphanesi), ZooKeeper (dağıtık uygulamalar için yüksek ölçekli koordinasyon uygulaması) projeleri geliştiriliyor.

Hadoop Pig, Hive ve Zookeeper gibi bir Apache projeleri eko sistemi tarafından desteklenmektedir. Bu Hadoop'un değerini arttırmakta ve kullanılabilirliğini genişletmektedir. Fiyat-etki, ölçeklenebilirlik ve elverişli mimarisinden dolayı, Hadoop göze çarpan özelliklerinin etkisiyle, büyük ölçüde bilgi işlem dinamiklerini ve ekonomisini değiştirmektedir.

Daha yalın bir dille anlatmak gerekirse, Apache Hadoop, HDFS ve MapReduce bileşenlerinden oluşan açık kaynak kodlu bir yazılımdır.

- **MapReduce:** Bir kümedeki (cluster) düğümlere iş atayan ve anlayan iskelettir. 2004 yılında Google tarafından tanımlanmıştır ve binlerce düğüm arasındaki iş yüklerini dağıtmayı mümkün kılmıştır. “Daha küçük alt problemler içerisindeki problemi ayırmak” stratejisi üzerine temellendirilmiştir ve SQL yoluyla ve SQL-

tabanlı BI (iş zekâsı) araçları içerisinde ortaya çıkabilir. MapReduce operasyonel olarak Büyük Veri setleri ile yapılacak işlemlerin birden fazla iş birimine dağıtılmasını sağlayan yöntemdir. Buradaki büyük veri setleri birkaç bilgisayarın bile işleyemeyeceği kadar verinin toplandığı setlerdir. Bu setler üzerindeki işlemler serisi çeşitli birimlere dağıtmakta ve sonra çıktıları birleştirilip sonuç üretilmektedir. Adında da anlaşılacağı üzere iki katmandan oluşmaktadır: Map ve Reduce. Map katmanı; İsim-değer çiftlerini girdi olarak alır, yapılacak işlemi gerçekleştirir. Ürettiği sonuç listesini ise girdideki isim ile birlikte çıktı olarak verir. Reduce katmanı ise Map katmanındaki sonuç listelerini toplar ve tek sonuca indirgeme işlemi yapar. Birden fazla iş birimi Map işiyle uğraşırken, bir sonraki katman sonuçları toplayıp tek sonuç haline getirir.

- **Hadoop Dağıtık Dosya Sistemi (HDFS):** Veri depolama için bir Hadoop kümesi içindeki tüm düğümleri yayan bir Apache açık kaynak dağıtık dosya sistemidir. Büyük bir dosya sistemi oluşturmak için birçok yerel sistem düğümleri üzerindeki dosya sistemlerini birbirlerine bağlar. HDFS düğümlerin başarısız olacağını varsayar, bu yüzden birden fazla düğüm üzerinden veri çoğaltarak güvenilirliği sağlar.

HDFS sayesinde sıradan sunucuların diskleri bir araya gelerek büyük, tek bir sanal disk oluştururlar. Bu sayede çok büyük boyutta birçok dosya bu dosya sisteminde saklanabilir. Bu dosyalar bloklar halinde (varsayılan 64MB) birden fazla ve farklı sunucu üzerine (varsayılan 3 kopya) dağıtılarak RAID – diskler arasında veri kopyalama veya paylaşımı için birden fazla sabit diski kullanarak yapılan veri depolama tasarısı – benzeri bir yapıyla yedeklenir. Bu sayede veri kaybı önlenmiş olur. Ayrıca HDFS çok büyük boyutlu dosyalar üzerinde okuma işlemi (streaming) imkânı sağlar, ancak rastlantısal erişim (random access) özelliği bulunmaz.

HDFS NameNode ve DataNode süreçlerinden (process) oluşmaktadır:

NameNode ana (master) süreç olarak blokların sunucular üzerindeki dağılımından, yaratılmasından, silinmesinden, bir blokta sorun meydana geldiğinde yeniden oluşturulmasından ve her türlü dosya erişiminden sorumludur. Kısacası HDFS üzerindeki tüm dosyalar hakkındaki bilgiler (metadata) NameNode tarafından saklanır ve yönetilir. Her kümede yalnızca bir adet NameNode olabilir.

DataNode ise, işlevi blokları saklamak olan işçi (slave) süreçtir. Her DataNode kendi yerel diskindeki veriden sorumludur. Ayrıca diğer DataNode'lardaki verilerin yedeklerini de barındırır. DataNode'lar küme içerisinde birden fazla olabilir.

Hadoop MapReduce ise HDFS üzerindeki büyük dosyaları verileri işleyebilmek amacıyla kullanılan bir yöntemdir. İstedığınız verileri filtrelemek için kullanılan Map fonksiyonu ve bu verilerden sonuç elde etmenizi sağlayan Reduce fonksiyonlarından oluşan program yazıldıktan sonra Hadoop üzerinde çalıştırılır. Hadoop Map ve Reduce'lerden oluşan iş parçacıklarını küme üzerinde dağıtarak aynı anda işlenmesini ve bu işler sonucunda oluşan verilerin tekrar bir araya getirilmesinden sorumludur. Hadoop'un gücü işlenen dosyaların her zaman ilgili düğümün (node) yerel diskten okunması ile ağ trafiğini meşgul etmemesinden ve birden fazla işi aynı anda işleyerek doğrusal olarak ölçeklenmesinden geliyor diyebiliriz.

MapReduce, JobTracker ve TaskTracker süreçlerinden oluşur. JobTracker yazılan MapReduce programının küme üzerinde dağıtılarak çalıştırılmasından sorumludur. Ayrıca dağıtılan iş parçacıklarının çalışması sırasında oluşabilecek herhangi bir problemde o iş parçacığının sonlandırılması ya da yeniden başlatılması da JobTracker'ın sorumluluğundadır. TaskTracker, DataNode'ların bulunduğu sunucularda çalışır ve JobTracker'dan tamamlanmak üzere iş parçacığı talep eder. JobTracker, NameNode'un yardımıyla DataNode'un yerel diskindeki veriye göre en uygun Map işini TaskTracker'a verir. Bu şekilde verilen iş parçacıkları tamamlanır ve sonuç çıktısı yine HDFS üzerinde bir dosya olarak yazılarak program sonlanır.

Hadoop şu anda Yahoo, Amazon, eBay, Facebook, LinkedIn gibi birçok lider firmada büyük verileri analiz etmek amacıyla kullanılıyor.

IBM'de veri analizi yaparken entegre sistem ile birçok imkân sunmaktadır. IBM Big Data Platformu olarak sunduğu hizmetler temel olarak Hadoop sistemi, akışkan veri işleme ve veri ambarlarından oluşmaktadır.

| Company | System |
|------------|---|
| IBM | Apache Hadoop, InfoSphere |
| Cloudera | CDH, Cloudera Standard, Cloudera Enterprise |
| Oracle | Oracle Big Data Appliance |
| Google | BigTable |
| Yahoo! | Sherpa |
| Amazon | SimpleDB |
| Microsoft | Dryad |
| Facebook | Apache Cassandra |
| Hypertable | HyperTable |
| ASF | Apache CouchDB |

Şekil 2.1: Büyük Veri Yönetim sistemleri.

(Kaynak: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7293298, 2015)

<http://wiki.apache.org/hadoop/PoweredBy> sitesinden Hadoop kullanan tüm kuruluşlara ve verdikleri hizmetlere ulaşmak mümkündür. Hadoop projesini geliştiren, bu konuda eğitim ve danışmanlık hizmetleri sunan Hortonworks, Cloudera, MapR gibi firmalar mevcuttur.

2.1.5. Bazı Özel Büyük Veri Yazılımları ve Şirketleri

Big Data yani Büyük Veri konusu üzerine son yıllarda sıkça konuşulur oldu. Hiç kuşku yok ki yakın gelecekte de verinin ve verinin analizine duyulan ihtiyacın artmasıyla birlikte Büyük Veri konuşulmaya devam edecek gibi görünüyor.

Kendi başına bir uzmanlık alanı haline gelen Büyük Veri konusuyla uğraşan çoğu proje ve şirket, belirli noktalarda hazır çözümler kullanmayı tercih ediyor. Bunlardan bazı örnekler şunlardır.

- **Flurry:** Twitter'ın günlük tweet trafiğinin yaklaşık 3 katı kadar veriyi, kendisi ile entegre olan mobil uygulamalardan toplayan Flurry; Android, iOS, Windows Phone ve Blackberry üzerinde çalışabiliyor. 230.000 uygulamayı bünyesinde

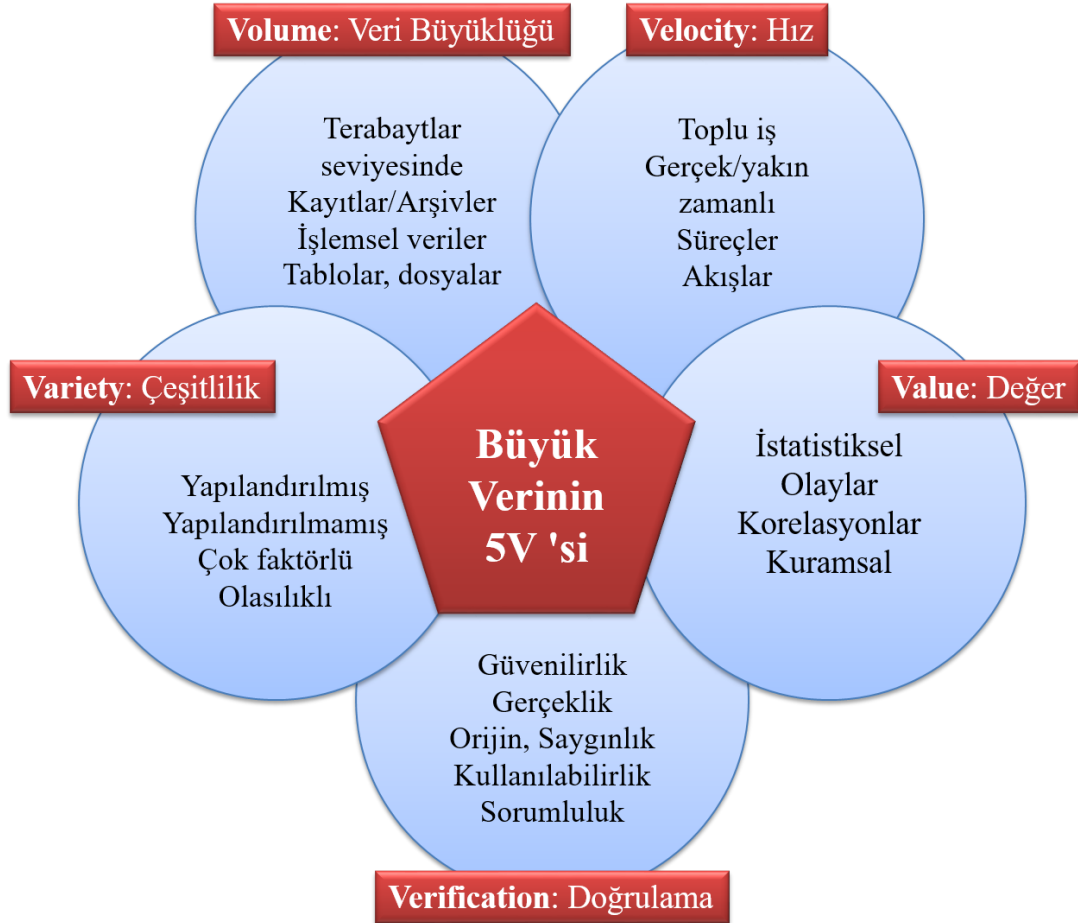
barındıran Flurry, her ay 700 milyon mobil cihaza ulaşıyor. Sahip olduğu bu veriler ve pazarın geleceği ile yıldızı parlayan şirket, aldığı yatırımlar ile birlikte yıllık 100 milyon dolar kar eden bir yapıya büründü.

- **Continuity:** Facebook (Jonathan Gray) ve Yahoo'da (Todd Papaioannou) büyük veri üzerine çalışan 2 mühendisin kurucusu olduğu bir startup olan, içerisinde Andreessen Horowitz'in de bulunduğu bir gruptan 10 milyon dolar yatırım alan California merkezli şirket, büyük veri konusunda yazılımcıların kullanabileceği bir PaaS (Platform as a Service) olmayı hedefliyor. Çoğu ürünü henüz beta aşamasında olan Continuity, Hadoop ve Hbase gibi konularda kendisi ispatlamış 2 kurucu ortağı ve aldığı yatırımlarla dikkat çekiyor.
- **Infochimps:** Texas/ABD merkezli şirket, kullandığı teknolojiler ve bulut çözümleriyle Continuity'ye oldukça benziyor diyebiliriz. Kısa bir süre önce Infochimps Enterprise Cloud isimli ürününü yayına alan şirketin, Cloudera ve Rackspace gibi araçlarla da çeşitli ortaklıkları bulunuyor.
- **Kontagent:** Aralarında EA Sports, BBC ve Warner Bros gibi isimler bulunan müşteri kitlesine, sosyal ve mobil internet dünyalarına özel analiz araçları sunuyor. Sistem üzerinde her bir saniyede 10.000 adet sosyal tetikleme gerçekleşiyor. Ayrıca şirket kSuit aracı ile müşterilerinin kullanıcı panellerine mobil cihazlarından (iOS) da ulaşmasına olanak veriyor.
- **Precog:** Hem bir Startup Weekend, hem de TechStars girişimidir. Kendilerini bu yukarıdaki diğer şirketlerden ayıran ise yazılımcılar için geliştirdikleri Labcoat IDE isimli platformdur. Bu platform sayesinde yazılımcılar, Quirrel sorgu dilini kullanarak özelleştirme yapabiliyorlar. Şirketin iddiasına göre söz konusu platformu yazılımcı olmayan kişilerin bile kullanabileceği yönünde.
- **StatsMix:** Oldukça sade ve net bir büyük veri analiz aracıdır. Colorado/ABD merkezli şirket, aynı zamanda bir TechStars üyesidir. Servis üzerinde toplanan birçok veriye ek olarak diğer istediğiniz özellikleri de kendiniz tanımlayarak takip yapabiliyorsunuz. StatsMix'in gelişmiş kontrol paneli üzerinde bu verilerin görselleşmiş sonuçlarına ulaşmanız, derseniz düzenli raporlar almanız ise mümkün oluyor.

2.1.6. Büyük Veri Bileşenleri

Büyük veri platformları, sayısal ağlar üzerinde farklı, ayrık, aykırı olanları kategorize etmeye çalışırken diğer yandan da sanal ortamlara daha çok verinin eklenmesini sağlayarak maliyetleri düşürmeye çalışmaktadır. Bu kapsamda büyük verinin 5V olarak adlandırılan unsurları önem kazanmıştır (Gürsaka, 2013; Ege, 2013; Wikipedia, 2012).

Yani şunu diyebiliriz ki, Büyük Veri platformunun oluşumunda beş bileşen vardır. İngilizce ifadelerinin baş harfleri kullanılarak kısaca 5V denilmektedir. Bunlar; Variety (Çeşitlilik), Velocity (Hız), Volume (Veri Büyüklüğü), Verification (Doğrulama) ve Value (Değer) 'dir. Bu bileşenleri şu şekilde inceleyebiliriz.



Şekil 2.2: Büyük Veri bileşenleri.

(Kaynak: <http://adressilvaa.tumblr.com/post/87206443764/big-data-refers-to-5vs-volume>)

2.1.6.1. *Çeşitlilik: Variety*

Veri yapılandırılmış, yapılandırılmamış ya da yarı-yapılandırılmış olabilir ve bu üç veri tipinin hepsi sık ve yoğun bir şekilde birbiriyle yer değiştirebilmektedir. Yapılandırılmış veriler veri tabanlarında depolanan büyük verinin sadece %20'lik bir kısmını oluşturmaktadır. Kullanıcılar tarafından internet üzerinde, sosyal ağlarda, fiziksel algılama cihazlarındaki veriler dinamiktir ve yapılandırılmamıştır.

Üretilen ve veri tabanlarında depolanan verinin yüzde 80'i yapısal (yapılandırılmış) değildir ve her yeni üretilen teknoloji, farklı formatlarda veri üretebilmektedir. Telefonlardan, tabletlerden, bütünleşik devrelerden gelen türlü çeşitlilikte “Veri Tipi” ile uğraşılması gerekmektedir. Bir de bu verilerin farklı dillerde, Non-Unicode olabileceğini düşünürseniz, bütünleşik olmaları, birbirlerine dönüşmeleri de gereklidir.

Sosyal medya, sensör verileri, CRM dosyaları, dokümanlar, imajlar, videolar vb. aklınıza gelebilecek bütün verileri, kaynakları hayal edin. Bunların tamamını ilişkisel bir veri tabanında sakladığınızı hatta veri tabanını dahi değil, bildiğimiz bir dosya sistemi üzerinde bile saklamak pek mümkün değil ve maliyetlidir. Verilerin çeşitliliği artmışsa ve bütün bu verileri işlemek, analiz etmek ve saklamak istiyorsak Büyük Veri kavramı bunun için biçilmiş kaftandır.

2.1.6.2. *Hız: Velocity*

Hız, büyük veri analizlerindeki yarı-gerçek zamanlı ve gerçek zamanlı işleme üzerindeki yüksek gereksinimleri ifade etmektedir. Geleneksel veri ambarı ve iş zekâsı için gerçek-zaman gereksinimleri daha düşüktür.

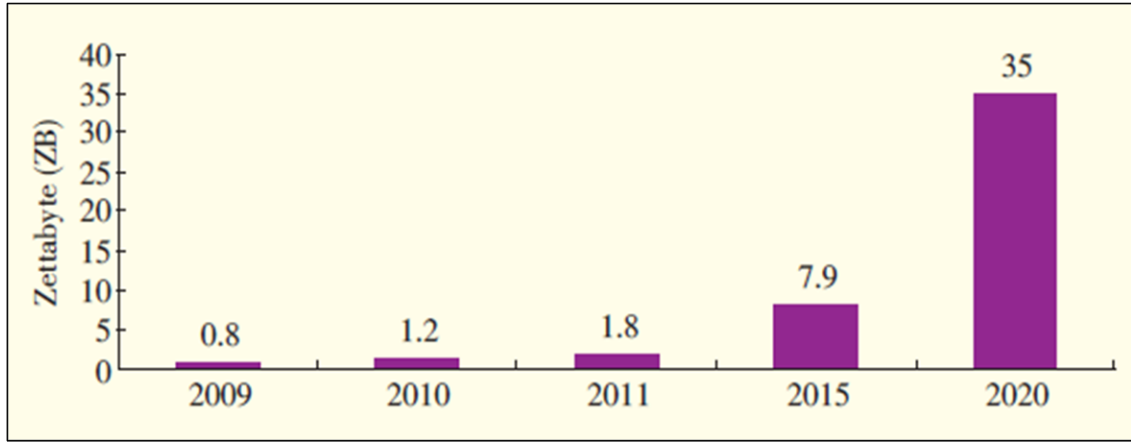
Büyük Verinin üretilme hızı çok yüksek ve gittikçe artmaktadır. Daha hızlı üreyen veri, o veriye muhtaç olan işlem sayısının ve çeşitliliğinin de aynı hızda artması sonucunu doğurmaktadır.

2.1.6.3. *Veri Büyüklüğü: Volume*

Üretilen verinin çok hızlı artmasından dolayı, dünya çapında veri hacmi çok devasa boyutlara ulaşmaktadır. Her yerdeki mobil cihazlar ve kablosuz veri sensörleri her dakika veri üretiyor ve toplu veri alışverişleri milyarlarca internet servisleri arasında her saniye oluşmaya devam ediyor. Bilimsel uygulamalar, video gözetimleri, tıbbi kayıtlar, operasyonel ticari veriler ve e-ticaret verileri büyük veri kaynaklarını oluşturmaktadır.

Uluslararası Veri Kurumu IDC, 2011’de, dünyadaki verilerin her iki yılda bir ikiye katlandığını iddia etmiştir. Yine IDC istatistiklerine göre, 2020’de ulaşılacak veri miktarının, 2009’un 44 katı olacağı öngörülmektedir.

Şu anda kullanılan, “büyük” diye adlandırdığımız kapasiteleri ve “büyük sistemleri” düşünüp, bunların 44 kat büyüklükte verilerle nasıl başa çıkacaklarını hayal etmek gerekiyor. Bu devasa artışlar depolama için önemli sorunlar oluşturmaktadır. Veri arşivleme, işleme, bütünleştirme, saklama vb. teknolojilerinin bu büyüklükte veri hacmi ile nasıl başa çıkacağına kurgulanması gerekiyor. 2010’lu yıllarda dünyadaki toplam bilişim harcamaları yılda %5 artmakta, ancak üretilen veri miktarı %40 artmaktadır.

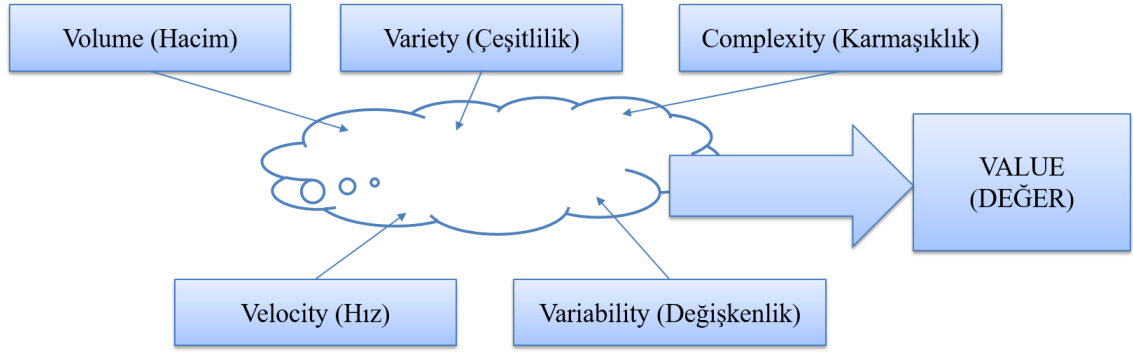


Şekil 2.3: Küresel Verinin hacminin tahmini.

(Kaynak: Uluslararası Veri Kurumu, IDC, 2011)

2.1.6.4. Doğrulama: Verification

Bu bilgi yoğunluğu içinde verinin akışı sırasında “güvenli” olması da bir diğer bileşendir. Akış sırasında, doğru katmadan, olması gerektiği güvenlik seviyesinde izlenmesi, doğru kişiler tarafından görünebilir veya gizli kalması gerekmektedir.



Şekil 2.4: Büyük Verinin en önemli beş bileşeni.

(Kaynak: <http://mis.sadievrenseker.com/2014/06/buyuk-veri-big-data/>)

2.1.6.5. *Değer: Value*

Bir işletmedeki geleneksel veri statik ve arşivsel analizler için kullanılabilir. Ancak, Büyük Veri daha çok gelecek trendleri ve önemli kararların alınmasını etkileyecek önemli bir faktördür. Bununla birlikte Büyük Veri daha düşük değer yoğunluğuna sahiptir. Örneğin bir video kayıt cihazından sürekli gözetim verileri üretilebilir, fakat bu görüntülerin sadece birkaç saniyesi aslında faydalı olabilecektir.

Bu yüzden en önemli bileşen ise büyük verinin bir değer oluşturmasıdır. Bütün yukarıdaki bileşenlerle tarif edilen Büyük Verinin, veri üretim ve işleme katmanlarından sonra kurum için bir artı değer meydana getireyor olması gereklidir.

Karar veriş süreçlerine anlık olarak etki etmesi, doğru kararı vermede hemen el altında olması gerekiyor.

Örneğin; sağlık konusunda stratejik kararlar alan bir devlet kurumu anlık olarak bölge, il, ilçe vb. detaylarda hastalık, ilaç, doktor dağılımlarını görebilmelidir. Hava Kuvvetleri bütün uçucu envanterindeki taşıtlarının anlık yerlerini ve durumlarını görebilmeli, geriye dönük bakım tarihçelerini izleyebilmelidir. Bir banka kredi vereceği kişinin, sadece demografik bilgilerini değil, yemek yeme, tatil yapma alışkanlıklarını dahi izleyebilmeli, gerekirse sosyal ağlarda ne yaptığını dahi görebilmelidir.

2.1.7. Büyük Veri Analizindeki Zorluklar

Günümüzde internet şirketlerinin servisleriyle ilgili büyük veri hızlı bir şekilde büyümektedir. Örneğin, her ay Google 100'lerce petabayt veri işler, Facebook her ay 10 petabaytı aşkın veri üretir, bir Çinli şirket olan Baidu 10'larca petabayt, Alibaba 'nın bir yan kuruluşu olan Taobao çevrimiçi ticaret için her gün 10'larca petabayt veri üretmektedir. Bilgi teknolojilerinin en son ilerlemeleri veri üretimini daha kolay bir hale getirmiştir. Örneğin, Youtube'a ortalama 1 dakikada 72 saatlik video yüklenmektedir.

Bundan dolayı veri kaynaklarının çok geniş dağılımından ve devasallığından dolayı bir entegrasyon ve toplama zorluğu ile karşı karşıyayız. İnternet araçlarının ve bulut bilişimin çok hızlı gelişmesi veri büyümesinde keskin bir artış sağlar.

Veri işleme teknolojileri uzun bir dönem boyunca sürekli olarak gelişmiştir, fakat büyük veri ani bir şekilde iki önemli değişikliğe neden oldu. Şu anda, tüm veriler kaydedilebilir. Bunun anlamı şudur; Uygulama öncesinde biriktirilmiş veri gerektiren uygulamalar daha kolay bir şekilde uygulanabilir.

Ayrıca, kısıtlı veriden, veri seline bir kayma meydana gelmiştir. Öyle ki, veri seli uygulamalar için yeni zorluklar meydana getirmiştir. Basit bir şekilde, sadece arama motorlarından elde edilen veriler artık bugünkü uygulamalar için yeterli değildir.

Bu büyük veri kütesinden kullanışlı veri elde etme ve işleme süreci gittikçe zorlaşmaktadır. Büyük Verideki teknolojik atılımlar yapılmaya devam edecek ve bu yüzden Büyük Veri önümüzdeki 2 ila 5 yıl içerisinde daha yaygın bir şekilde kullanılacaktır.



Şekil 2.5: Bir Büyük Veri senaryosu.

(Kaynak:http://www.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)

Geleneksel olarak veriler, tek bir kaynaktan gelen ve nispeten düşük bir hacme sahiptir. Depolanması, yönetimi ve analizi büyük zorluklar getirmez ve en çok işlem süreci ilişkili veri tabanları ve veri ambarları aracılığıyla yapılmaktadır. Büyük Veri alanında ise, verinin hacmi o kadar büyüktür ki, geleneksel veri işleme sistemleri bu verinin depolanması, yönetimi ve analizi ile başa çıkamaz. Geleneksel iş zekâsı yazılımları yapılandırılmamış veri analizi ve işleme için eksik araçlar ve metotlara sahiptir.

Büyük verinin analizi bazı zorlukları beraberinde getirmektedir. Bu zorluklar şu şekilde sıralanabilecektir.

2.1.7.1. *Kapsamlı Veri Kaynakları ve Kötü Veri Kalitesi*

Büyük Veri senaryoları heterojen veri kaynakları ile karakterize edilmektedir. Örneğin, işlem kayıtları, metinler, resimler ve videolar. Bu veri farklı şekillerde açıklanmakta ve veri giriş oranı saniyede yüzlerce megabaytlara hatta gigabaytlara kadar ulaşabilmektedir.

Yapısal veriyi açıklamak için kullanılan geleneksel metotlar Büyük Veriyi açıklamak için yeterli değildir. Ek olarak, Büyük Veri gürültüden kolayca etkilenebilir, kaybolabilir ya da tutarsızlaşabilir. Eksik, gürültülü ve tutarsız veriyi filtreleme ve bütünleştirme büyük veriyi verimli bir şekilde depolama ve işleme süreci için ön koşuldur.

2.1.7.2. *Büyük Verinin Oldukça Verimli Depolanması*

Büyük Verinin depolanması yalnız maliyetleri değil, aynı zamanda analiz ve işleme verimliliğini de etkiler. Büyük Veri çoğunlukla kurumsal depolama alanı ağları ve buna bağlı ağlarda ele alınmayan petabaytlarla hatta exabaytlar ile ölçülmektedir. Büyük Veri çağında hizmet ve analiz gereksinimlerini karşılamak için güvenilir, yüksek performanslı, yüksek kullanılabilirliğe sahip, düşük maliyetli çözümler geliştirmeye ihtiyaç vardır. Büyük Verinin çeşitli kaynaklardan gelmesinden dolayı, aynı veri sistemde zaten var olabilir ve bu fazlalığa neden olabilir. Algılama ve fazlalıkları eleme depolama yerini arttırır. Bu Büyük Veri depolama platformları için gerekli olan temel unsurdur.

2.1.7.3. *Veriyi Verimli Bir Şekilde İşleme*

Kurumsal veri ağırlıklı olarak ilişkisel veri tabanları ve veri ambarlarında işlenmektedir. Ancak, bu tür veri tabanları ve veri ambarları yapılandırılmamış ve yarı-yapılandırılmış büyük veriyi işlemek için yetersizdir. Büyük Veri ile yazma/okuma işlemleri çok sayıda kullanıcı için oldukça eş zamanlı olmalıdır. Büyük Verinin depolanması ve veriye erişim yüksek verimli olması gerekmektedir ve sistem son derece ölçeklenebilir olmak zorundadır.

Veri setlerinin büyüklükleri arttıkça, algoritmalar verimsiz hale gelebilir ve bölünmezlik, ilişkisel veri tabanlarının tutarlılık, yalıtkanlık, dayanıklılık özellikleri yoğun kaynaklardır. CAP teoremi dağıntık bilgisayar sistemleri için tutarlılık, ulaşılabilirlik ve bölünebilme toleransının garanti olmadığını ifade eder. Çünkü tutarlılık paralel ilişkili veri tabanlarında gereklidir, bu veri tabanları iyi derecede ölçeklenebilir ve mevcut değildir. Büyük Veri analizleri için yüksek sistem ölçeklenebilirliği en önemli gerekliliktir ve son derece genişletilebilir veri analiz teknikleri geliştirilmiş olması gereklidir.

2.1.7.4. *Yığın Veri Madenciliği*

Veri setleri boyutlarının artmasıyla, Büyük Veri için çok daha fazla makine öğrenme algoritmaları ve veri madenciliği algoritmaları ortaya çıkmıştır. Araştırmalar göstermiştir ki, daha büyük veri setleri için, makine öğrenimi daha doğrudur ve makine öğrenme algoritmaları arasında çok az farklar vardır.

Ancak, geniş veri setleri veri madenciliği algoritmaları için olduğu kadar geleneksel makine öğrenme algoritmaları için de problemdir. Aslında, en geleneksel veri madenciliği algoritmaları büyük veri setleri tarafından geçersiz hale getirilir. Bununla birlikte, veri seti boyutları petabaytlar seviyesine ulaştığında, seri algoritmalar kabul edilebilir bir süre içerisinde hesaplanmak için başarısız kalabilir. Etkili makine öğrenme ve veri madenciliği algoritmaları büyük veri setleri için geliştirilmek zorundadır.

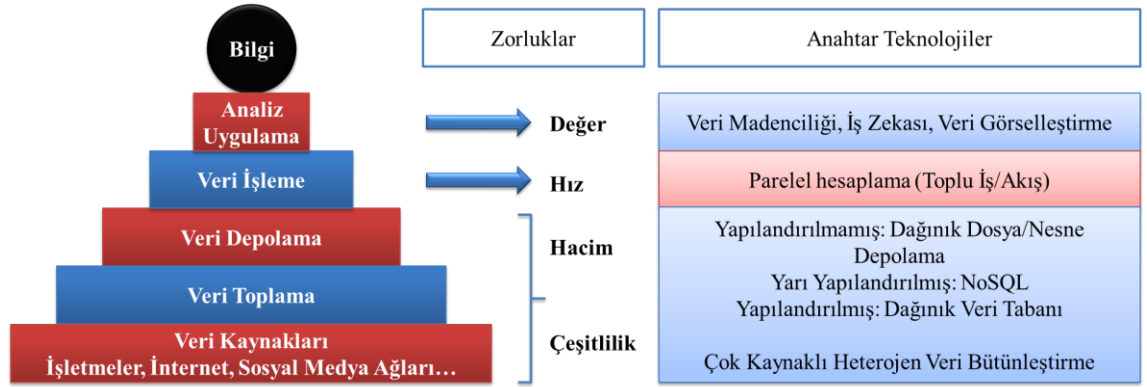
2.1.8. **Büyük Veri Analizi Teknolojisi**

Tipik bir Büyük Veri işleme sistemi toplama ve ön-işleme, depolama, analiz, madencilik ve değer uygulama aşamaları içermektedir.

Şekil 2.6 bir Büyük Veri sisteminin mimarisini göstermektedir. En alttaki veri kaynakları tabakasında işletmelerden, endüstri kuruluşlarından, internet ve internete bağlı aygıtlardan veriler gelir. Bunun bir üstündeki veri toplama tabakasında toplanan veri ön işleme sürecinden geçirilir. Bu ön işleme süreci veri temizleme ve heterojen veri işleme sürecini içerir. Sonraki veri depolama katmanında yapılandırılmış, yapılandırılmamış ve yarı-yapılandırılmış veri depolanmakta ve yönetilmektedir. Veri işleme katmanında ise, veri analiz edilir ve veri madenciliği yapılır ki, kullanıcılar yaygın telekomünikasyon ve internet servisleri gibi hizmetleri analiz eder.

2.1.8.1. *Toplama ve Ön İşleme*

Büyük miktarlardaki yığın verinin biçimini değiştirme masraflıdır ve buna veri toplamanın zorluğu eklenir. Geleneksel veri toplama araçlarının modası geçmiştir ve çoğu internet şirketi kendi büyük veri toplama sistemlerine sahiptir. Bu sistemlerin bazı örnekleri Apache Chukwa, Facebook Scribe, Cloudera Flume, and LinkedIn Kafka'dır.



Şekil 2.6: Büyük Veri Mimarisi.

(Kaynak:http://www.en.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)

Temizleme ve ayıklama teknolojileri ağlardaki zararlı, gereksiz ve kullanışsız yığın veriyi temizlemek ve analiz için kaliteli veriyi ortaya çıkartmada kullanılır. Hadoop veri temizleme, dönüştürme ve yüklemeyi iyileştirmek ve paralel verinin ön işleme sürecini geliştirmek için kullanılmaktadır. Büyük veri toplama ve ön işleme teknolojileri çok kaynaklı ve çok modlu verinin kalitesini optimize etmek için tasarlanmıştır. Bu teknolojiler çoklu veri entegrasyonu için gerekli zorunlulukları yerine getirir. Bilgi içerisindeki yüksek kaliteli veriyi dönüştürür, farklı kaynaklardan elde edilen verinin kalitesini kontrol eder ve veri analizi için temeli oluştururlar.

2.1.8.2. Depolama

Daha önce de bahsettiğimiz gibi depolanan büyük verinin %80'i yapılandırılmamış ve yarı-yapılandırılmış veriden oluşmaktadır. Şu anda, büyük veri depolama, belirli yöntemler kullanılarak ölçeklenebilir olan ve masraflı olmayan X86 sunucu küme sistemleri üzerinde yapılandırılmaktadır.

Yapılandırılmamış veri dağınık dosya sistemlerinde depolanmaktadır. Geleneksel ağ dosya sisteminin (NFS) aksine, bir dağınık dosya sistemi ölçeklenebilirliği geliştirmek için kontrol akışlarını veri akışlarından ayırır. Bir dağınık dosya sisteminde, veri seti sunucuları tüm veri setlerini ve veri depolama sunucularının küme bilgilerini yönetir. Örneğin, Google File System (GFS), Hadoop Distributed File System (HDFS), Lustre, ve Ceph sistemlerinin kapasiteleri 10 hatta 100 petabaytlara kadar çıkabilir. Bu depolama

sistemleri POSIX (UNIX için Taşınabilir İşletim Sistemi Arabirimi) ara yüzleri üzerinden dosya erişimini sağlar. Eğer bir veri seti sunucusu bu tür bir depolama sisteminde başarısız olursa, tüm dosya sistemi kullanıcılara hizmet sağlamada başarısız olur. Bir tek düğüm veri seti sunucusunun işleme ve depolama kapasitesi de oldukça sınırlıdır. Sistemdeki veri trafik miktarı artarken, veri seti sunucularının işleme kapasitesi sistem ölçeklenebilirliği için bir darboğaz halini alır.

Veri seti sunucuları bundan dolayı aktif/uyku modunda çalıştırılmaktadır. Normal koşullarda, aktif veri seti sunucusu tüm talepleri yerine getirir, tüm dağınık dosya sistemlerini yönetir ve düzenli bir şekilde senkronizasyon için uyku halindeki veri seti sunucusuna veri gönderir. Eğer aktif sunucu başarısız olursa, uyku modundaki sunucu devreye girer ve kesintiye uğramadan tüm faaliyetler devam eder. Uyku modundaki sunucuya gönderilmeyen veri kaybolabilir. Bu modu kullanmak, tek nokta hataları için çözüm olabilir, fakat sistem ölçeklenemez. Bu yüzden, yüksek derecede ölçeklenebilir veri sunucu kümesi bir dağınık dosya sisteminde hayati öneme sahiptir.

Büyük Veri çoğu yapılandırılmış düz metin veri olan yarı-yapılandırılmış veri içerir, fakat ilişkisel veri tabanlarındakinden daha esnek modellere sahiptir. Yarı-yapılandırılmış veri katı veri tabanı işlemlerini gerektirmez. Çoğunlukla basit tablo sorgusundan oluşur ve bazı durumlarda düşük tutarlılık gereksinimlerine sahiptir. Bundan dolayı, veri tabanı işlem yönetimi ağır yüklü veri tabanları içindeki bir yüküdür. NoSQL veritabanı, geniş bir alanda diğerleriyle konuşan, ilişkisel veritabanı arasında bağlantı kuran ilişkisel olmayan veri tabanıdır. NoSQL sabit bir tablo yapısı ya da bağlantılı işlemler gerektirmez, bu durum büyük veriye erişimi açısından önemli avantajlar sağlar.

NoSQL veri tabanlarında, ilişkisel veri tabanı modelleri dağınık yatay genişlemeyi ve büyük veriyi destekleyen bir özgür-şema prensibi sayesinde terkedilmiştir. Dynamo, BigTable, Cassandra, ve MongoDB'yi kapsayan birçok NoSQL veritabanı ürünleri ve açık kaynak projeleri vardır. NoSQL sistemleri henüz olgunlaşmamıştır: Birçoğu küçük ticari desteğe sahip açık kaynak projelerdir. NoSQL sistemlerinin birleşik uygulama programı ara yüzleri eksiktir ve SQL desteklemezler.

2.1.8.3. İşleme

Karışık yükler ile büyük veri setlerini işleme süreci karmaşıktır ve Tablo 2.1’de görünen belirli işleme gereksinimlerine ihtiyacı bulunmaktadır.

Tablo 2.1: Büyük veri işleme gereksinimleri.

| | |
|---------------------------------------|---|
| Yüksek Ölçeklenebilirlik | Dağınıklaştırma ölçeklenebilirliği, geniş ölçekli paralel veri işlemeyi destekleme |
| Yüksek Performans | Yığın veri sorgulamasına hızlı yanıt ve analiz gereksinimleri |
| Düşük Maliyet | Evransel donanım sunucuları üzerine yapılandırma, yüksek fiyat/performans oranı |
| Hata Toleransı | Donanım güvenilirliğinin bağımsızlığı, bazı operasyonlar (diğerlerinden ayrı olanlar) hata durumunda çalışmalıdır |
| Kullanım Kolaylığı ve Açık Ara yüzler | Geleneksel SQL ara yüzleri ile uyum, sorgulama ve çok boyutlu veri analizini destekleme |
| Geçmişe Yönelik Uyumluluk | Geleneksel BI (Business Intelligence: İş Zekâsı) araçlarını destekleme |

Yapılandırılmamış Büyük Veri temel olarak dağınık bilgisayar mimarisi kullanılarak işlenmektedir. Farklı hesaplama modelleri üzerine yapılandırılmış 3 tip dağınık bilgisayar mimarisi bulunmaktadır.

- **Map Reduce Modeli:** Bu model (giriş anahtarı, giriş değeri) → (çıkış anahtarı, çıkış değeri) olarak ifade edilmektedir. Çıkış anahtarı/değeri çiftleri işlenmektedir, giriş anahtarı/değeri çiftleri MAP ve Reduce fonksiyonları kullanılarak üretilmektedir. Bu model basit bir mantığa sahiptir ve yaygın olarak kullanılmaktadır.
- **Bulk Synchronous Parallel (BSP) Modeli:** Bu basit bir bilgisayar modeline benzeyen tekrarlamalı hesaplama modelidir. Bununla beraber farkı iletişim içerisinde yatmaktadır. Bu modelde, tüm düğümler her hesaplama turundan sonra senkronize edilmektedir. Bu tekrarlamalı senaryolar için uygundur. Örneğin, Google Pregel BSP mimarisi üzerinde yapılandırılmıştır.

- **Directed Acyclic Graph (DAG) Modeli:** Bu model karmaşık hesaplama süreçlerini ve aralarındaki ilişkileri açıklama için DAG kullanır. Microsoft, Dryad projesinde bu hesaplama modelini kullanmaktadır.

Örneğin, Yahoo S4 ve Twitter Storm gibi gerçek zamanlı akış işleme platformları büyük veri işleme gereksinimlerini yerine getirmek için oluşturulmuştur. Bu platformlar hafızadaki bir veri akışını gerçek zamanlı işler ve çok fazla veriyi tutmazlar.

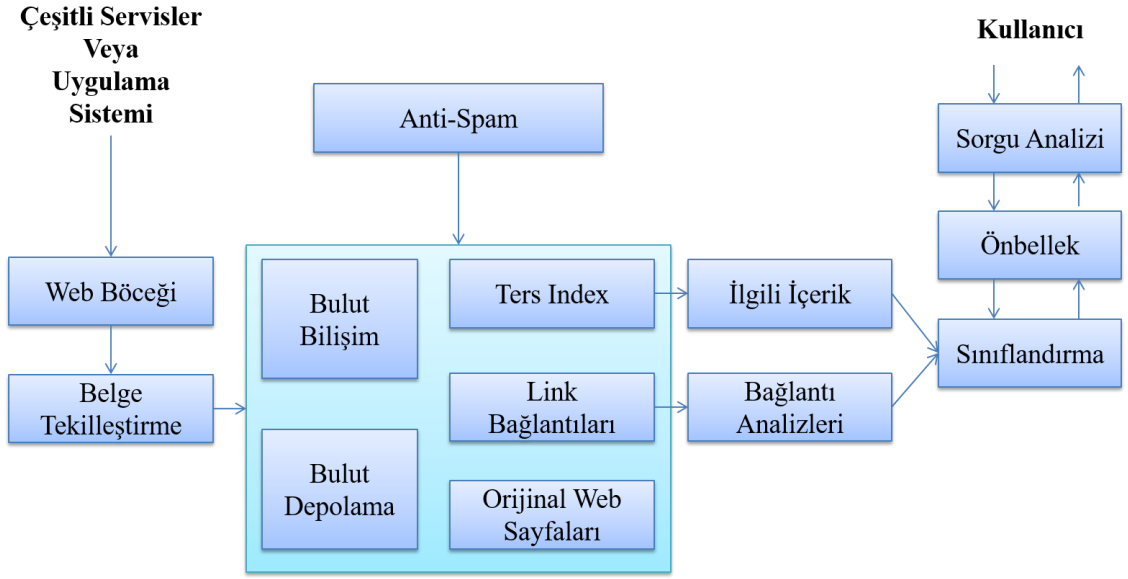
2.1.8.4. *Veri Madenciliği*

Büyük Veri madenciliği teknolojisi büyük veriden etkin bir şekilde veri ayıklamak için kullanılmaktadır. Bu teknoloji paralel veri madenciliği, arama motoru, tavsiye motoru ve sosyal ağ analizlerini kapsamaktadır.

Paralel veri madenciliği büyük ölçüde veri madenciliği algoritmaları uygulanarak büyük veri madenciliği hızını arttırmaktadır. Hadoop ve HDFS bu uygulamadan yararlanmaktadır. Veri madenciliği algoritmaları paralel sınıflandırma algoritması ve paralel kümeleme algoritmasını içermektedir.

Örneğin, arama motoru çeşitli servislerden ya da uygulama sistemlerinden veri toplayan bir bilgi erişim sistemidir. Veriyi depolar, işler ve yeniden organize eder, sorgulama fonksiyonları ve sonuçlarıyla kullanıcılara sunar. Arama motorları, veri depolama sistemlerinden elde edilen büyük veriden sonra, veri yönetimi için önemli bir araçtır. Kullanıcıların basit sorgular girmesine izin verir ve kullanışlı bilgiler sağlar.

Aşağıdaki Şekil 2.7 bir çok teknolojiyi içeren evrensel arama motoru mimarisini göstermektedir. Anahtar teknoloji web crawler (web gezgini), belge anlama, belge indeksleme, ilişki hesaplama ve kullanıcıyı anlama öğelerini içerir. Tavsiye motoru bilgi yığını içerisinde kişiselleştirilmiş hizmetleri ve içeriği sağlamak için kullanıcıya yardım eder. Bu bir arama çağından keşif çağına geçişin göstergesi olmaktadır. Tavsiyelerin kalitesi sadece model ve algoritmalara bağlı değildir, aynı zamanda ürün şekli ve hizmet modu gibi bazı teknik olmayan faktörlere de bağlıdır.

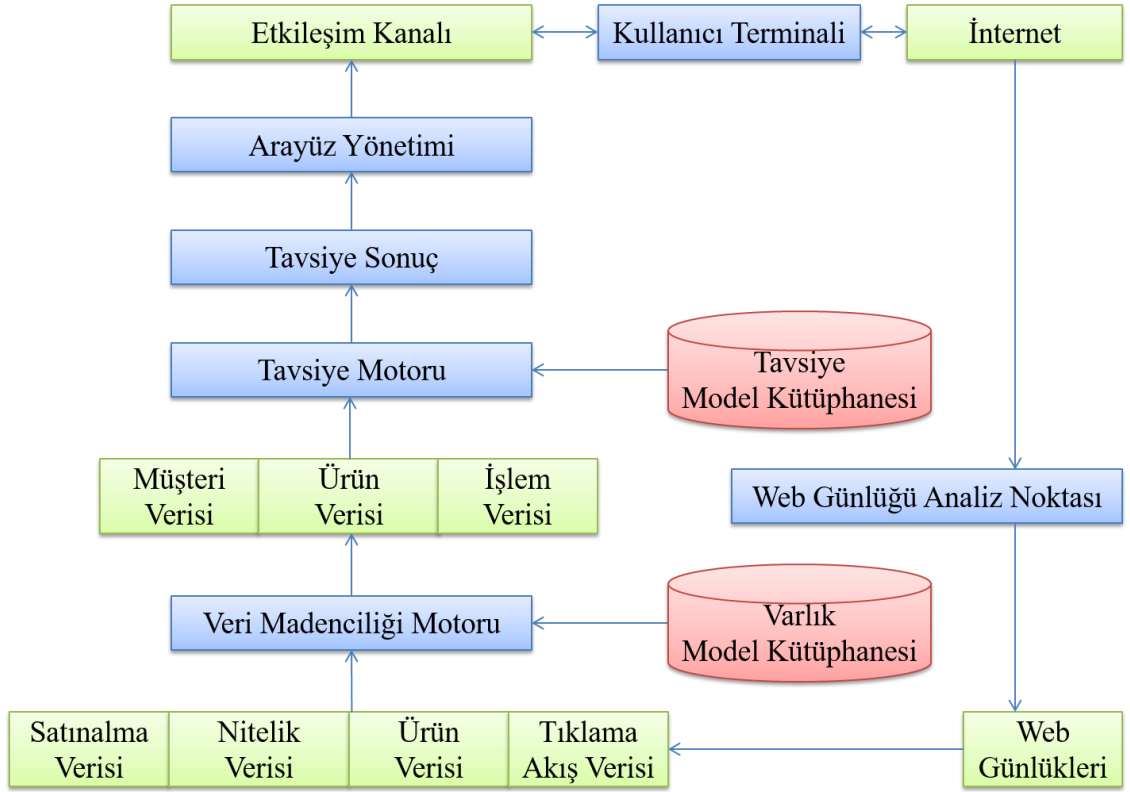


Şekil 2.7: Arama motoru mimarisi.

(Kaynak:http://wwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)

Şekil 2.8 kullanıcı işlemi veri tabanı, veri madenciliği motoru, veri ambarı, tavsiye model kütüphanesi, tavsiye motoru ve kullanıcı etkileşimi aracını içeren tipik bir tavsiye mimarisini göstermektedir. Bununla birlikte, bu alt sistemler, veri tabanları ve temel işlemler kullanıcılar kişisel hizmetler sağlar. Tavsiye motoru içerisindeki temel algoritmalar; içerik tabanlı filtreleme algoritması, ortak tabanlı filtreleme algoritması ve ilişki analiz algoritmasıdır.

Sosyal ağ analizi yeni analiz problemleri için yeni bir kavramdır. Bunlar üyeler arasındaki ilişkiler üzerinde temellendirilmiş ve interaktif veri madenciliği için metotlar ve araçlar sağlar. Bu çok kaynaklı bilgi ve fikirleri önümüze koyar. Sosyal filtreleme, pazarlama, tavsiyeler ve aramalar için önemlidir. Sosyal ağ analizi kullanıcı ilişkileri, konular, ilgi alanları, kimlik tanımlama, etki ve duygu analizi içerir. O ayrıca topluluğun keşfini sağlamaktadır.



Şekil 2.8: Akıllı öneri sistemi mimarisi.

(Kaynak:http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013).

Bulut tabanlı uygulamaların, mobil internetin ve internetin her türlü uygulamasının yaygınlaşması ve sosyal ağların yükselişi ile Büyük Veri ilgi odağı haline gelmiştir. Büyük Verinin birçok çeşitli tipleri var ve işleme metotları çok karmaşık hal almaktadır. Bu durum çok büyük zorluklar oluşturmaktadır. İlişkisel veri tabanları ve veri ambarları gibi geleneksel veri işleme metotları büyük veri işlemek için gayret göstermektedir. Dağıtılmış depolamaya sahip büyük veriye uygun sistemler Map/Reduce paralel hesaplama ve veri madenciliğidir. Fakat bunlar da henüz gelişim aşamasındadırlar.

Büyük Veri analizini geliştirmek için sistem, ölçme ve anlamlandırma üzerine araştırma yapılmalıdır. Ayrıca büyük veri depolamanın maliyeti de azaltılmalıdır ve esnek, yüksek verimli büyük veri hesaplama mimarileri ve veri madenciliği algoritmaları geliştirilmelidir.

2.1.9. Büyük Verinin Avantajları

Büyük Verinin ana önemi farklı veri tipteki büyük hacimli veri kullanımını bağlamında verimliliği arttırmak için potansiyel oluşundandır. Eğer Büyük Veri düzgün bir şekilde tanımlanır ve buna uygun olarak kullanılırsa, kurumlar işlerine daha iyi bir biçimde bakabilirler. Bundan dolayı satış, ürün geliştirme ve daha başka farklı alanlarda etkin bir şekilde faydalanabilirler.

McKinsey Global Enstitü'nün yayınladığı bir raporda (McKinsey Global Institute Report, 2011) Büyük Veri kullanımının değer arttıran bazı avantajlarından bahsedilmektedir:

- Bilgiyi daha sıklıkla şeffaf ve kullanılabilir yaparak değer katıyor.
- Şirketler daha fazla işlemsel veri üretip dijital şekilde sakladıkça, ürün envanterinden hastalık günlerine kadar her şey hakkında daha doğru ve ayrıntılı bilgi edinebiliyorlar. Böylece çeşitliği artırıyor ve performanslarını katlayabiliyorlar. Lider şirketler veri toplamayı ve analizini daha iyi yönetim kararları vermek için kontrollü deneyler yapmakta kullanıyor. Diğerleri veriyi, seyrek ve uzun vadeli yerine sık aralıklı ve şimdi-vadeli tahminler ile işlerini gerçek zamanlı yapmak için kullanıyor.
- Büyük veri müşteri segmentlerini giderek daha da daraltıyor. Böylece daha dikkatle kişiselleştirilmiş ürün ve hizmetler mümkün oluyor.
- Sofistike analizler karar vermeyi ciddi olarak kolaylaştırır.
- Yeni nesil ürün ve hizmetlerin geliştirilmesinde kullanılabilir. Hali hazırda üreticiler ürünler içine gömülen sensörlerle elde edilen veriyi, proaktif bakım (hata oluşmadan, hatta fark edilmeden önce koruyucu önlemleri almak) gibi yaratıcı satış-sonrası hizmetlerde kullanıyorlar.

Örneğin, perakendecilerde operasyonel kârın %60 dolayında artması, ABD sağlık hizmetlerinde yılda 300 milyon dolar civarında maliyet azalması gibi vurucu rakamları bu raporda ifade edilmiştir.

2.1.10. Büyük Verinin Zorlukları

Büyük Veriyi anlama çok önemlidir. Bir şirket ya da kurum stratejik kararlar vermek için sizin analizlerinize güveneceğinden dolayı, doğru ve düzgün bir şekilde analiz edilmesi

çok önemlidir. Ayrıca, bu analizlerin süresi çok önemlidir, iş dünyasındaki hızlı değişimlerden dolayı sıklıkla ve hızlı bir şekilde uygulanması gerekmektedir.

Diğer bir yönü ise, her gün yeni teknolojiler geliştirilmektedir. Büyük Verinin kuruluşlar için yeni olması göz önüne alındığında, yeni teknolojiler pazara girer girmez, onların nasıl kullanılacağını öğrenmeleri gerekir. Bu işletmeler için rekabet avantajı getirecek yönüdür.

IT uzmanları ihtiyacı da Büyük Verinin diğer zorluklarından birisidir. McKinsey'in Büyük Veri üzerine çalışmasına göre, sadece ABD'de 190.000 den fazla analitik uzman ve 1,5 milyondan fazla veri okur-yazarı müdüre ihtiyaç vardır. Bu istatistikler, Büyük Veri girişimini ele alan ve bu konuda zayıf olan şirketlerin ya uzmanlar kiralamasını ya da kendi personelini bu yeni alan üzerine eğitmek zorunda olmasını beraberinde getirmektedir.

Gizlilik ve güvenlik de Büyük Verinin zorluklarından birisidir. Büyük Veri çok karmaşık verilerden oluşmaktadır, bu yüzden bir şirket için bilgileri gizlilik seviyelerine göre sıralamak ve güvenlik uygulamak çok zordur. İlave olarak, bu günlerde birçok şirket ülkeler ve kıtalararası çaprazlama işler yapıyor ve bunların gizlilik hukuklarındaki farklılıklar dikkat çekiyor. Bir şirket Büyük Veri girişimine başladığında bunları göz önünde bulundurmaya zorundadır.

Bu sebeplerle, Büyük Veri uygulaması yapılırken tüm bu faktörlerin hepsi çok dikkate alınmalıdır.

2.1.11. Bazı Büyük Veri Uygulama Alanları

İnternetin çok fazla kişi tarafından kullanılır olması, günlük yaşamda aldığımız hizmetlerin internet üzerine taşınan uygulama yazılımları yoluyla kolayca erişilebilir olması nedeniyle oluşan yaygın kullanım, özellikle hizmet sektöründe satış sonrası müşteri memnuniyeti sağlama gereğine yönelik olarak, son kullanıcıların her türlü bilgisinin firmalarca alınıp saklanması sonucunu doğurdu. Saklanacak ayrıntılı bilgilerin, üzerinde tutulacağı sayısal ortamlarda baş gösteren "yer darlığı" yeni arayışlar için başlangıç oldu. Son birkaç yılda öne çıkan yeni yapıyı özetlersek: İşletmeler; müşterilerini daha iyi tanıyıp, onlara "bireyselleştirilmiş - kişiye özel" hizmetleri

sunabilmek için onlarla ilgili çok sayıda bireysel bilgiyi saklamak durumunda kalmaktadır.

Spesifik olarak bakarsak, örneğin; sosyal medyada süregelen reklam savaşları Yüksek Hacimli Verinin en önemli “müşterilerinden” biridir. Facebook, Twitter, Instagram ya da müzik dinleme hizmeti sunan birbirinden farklı ya da benzer mecra, reklam verenler için durmaksızın veri toplar. Kullanıcı alışkanlıklarını belirlemek, bu alışkanlıkları gerektiğinde paraya, gerektiğinde ise kısa dönemde kar sağlayacak stratejilere çevirmek asıl amaçtır. X bir sosyal mecraı kullanan bir kullanıcı, Y bir sayfaya gelmeden önce neler yapmıştır, incelediği sayfada bulunma sebebi nedir, kullanıcı ne yapılırsa reklam verenin arzu ettiği alana gelir, bu noktada istenilen ürüne yönlendirilebilir mi, nasıl yönlendirilir? İşte bunların hepsi Yüksek Hacimli Verinin işlenmesi ile ortaya çıkan sonuçların işidir.

Sermaye açısından Büyük Verinin önemi, açığa çıkardığı ticari veri havuzundan anlaşılacaktır. Günümüzde müşteri taleplerine hızla yanıt vermek, reel ve potansiyel tüketicilere karşı şeffaf bir kurum kültürü oluşturmak, yeni müşteri deneyimleri oluşturmak amacıyla fiziki ve sanal dünya arasındaki bağlara yönelik inovasyona öncülük etmek ve veri analitiği ile ilgilenmek, sıradan şirketlerin bağımsız olarak gerçekleştirebilecekleri bir işlem değildir.

IBM’in 56 ülkeden 524 şirket üst düzey yöneticisinin katılımıyla gerçekleştirdiği araştırmada, yöneticilerin % 35’i çoklu veri kaynaklarını entegre etmede zorlandıklarını, % 30’u analitik yaklaşımı nasıl uygulayacaklarından emin olmadıklarını ya da bunu yapacak yeteneklerinin bulunmadığını ve % 22’si veriye hiç erişemediğini ya da yetersiz oranda eriştiğini ifade edilmiştir (Bloomberg Businessweek Türkiye, 2014). Bu kapsamda “büyük verinin nasıl analiz edileceği ve yönetileceği?” sorusu gündeme gelmekte; yanıtı ise “çok uluslu şirketler tarafından kontrol edilen ticari veri havuzu aracılığıyla” şeklinde ifade edilmektedir. ABD Patent Ofisi (USPTO) enformasyon yönetiminin toplumsal kontrolünün temel göstergelerinden biri olarak görülmektedir. USPTO tarafından kayıt altına alınan patent sayısına baktığımızda; 2012 yılında IBM’in 6457, Samsung’un 5043, Canon’un 3173, Sony’nin 3017, Matsushita Electric (Panasonic)’in 2748, Microsoft’un 2610, Toshiba’nın 2415, General Electric’in 1650, LG Electronics’in 1617 ve Fujitsu’nun 1527 patent aldığı tespit edilmiştir (USPTO, 2014).

Aynı dönemde (2012 yılında) Türkiye’de tescil edilen yerli patent sayısının 1025 ve bu rakamın % 6,8’inin elektrik-elektronik alanında olması ise büyük veri havuzundan ulusal ölçekte ne düzeyde yararlanıldığının açık göstergesidir.

Büyük verinin sermaye açısından işlevselliğini sağlayan bir diğer konu ise yeni iş modelleri geliştirilebilmesine imkân tanınmasıdır. Büyük veri müşterilere, organizasyonlara, ürünlere, pazarlara ve diğer tüm ticari unsurlara ilişkin ayrıntılı analizlerle büyük bir ekonomik değer açığa çıkarmaktadır. Büyük veri ile öngörülebilir pazar bölümlendirilmesi yapılması, tüketici karar verme süreçlerine ilişkin karmaşık analiz tekniklerinin kullanılması ve mal ya da hizmet geliştirme/iyileştirme çalışmalarında yeni yöntemlere başvurulması mümkün hale gelmiştir (Gürsakal, 2013).

Şirketler veri madenciliği yoluyla selin önüne file germekte ve fileye takılanları analiz ederek ticari kararlar almaktadırlar. Sahip olunan veriler ile sorunların nasıl çözümlenebileceğinin öğrenilmeye başlaması, şirketleri, ‘üretim yönetimi’, ‘pazarlama’, ‘insan kaynakları’ ya da ‘muhasabe-finans’ departmanlarının elinden kurtararak ‘bilgi işlem’ birimlerine doğru yönlendirmektedir.

2012 yılının sonlarında ABD merkezli teknoloji araştırma şirketi Gartner, Büyük Verinin ilerleyişini destekleyecek biçimde 2015 yılı itibariyle küresel ölçekte tüm sektörleri kapsayacak biçimde 4,4 milyon enformasyon teknolojisi (IT) temelli istihdam gelişeceği öngörüsünde bulunmuştur. Gartner, yetişmiş personel eksikliği nedeniyle bu pozisyonların sadece 1/3’ünün dolacağını ifade etmiştir. Veri uzmanlığının, piyasa yeterli sayıda veri uzmanı yetiştirinceye ve nihai mamul üreticisi şirketler ileri teknoloji, finans ve biyoteknoloji sektörlerindeki şirketlerle rekabet edebilecek düzeye gelinceye değin, orta vadede “kıt ve değerli bir ürün” olacağını belirtmiştir (The Economist, 2013, s. 7-8).

Şunu da belirtmek gerekir ki, Büyük Veri tüm ölçekteki organizasyonlar tarafından kullanılmaktadır. Büyük çaplı firmalara ek olarak küçük işletmeler, girişimciler ve devlet ajansları Büyük Veriye arzu duyan kullanıcılarıdır. FTC (Federal Ticaret Komisyonu) Başkanı Bayan Edith Ramirez bu konuda şöyle ifade etmektedir “Büyük Veri artık birkaç dev şirketin uzmanlık alanında değildir”. Sonuç olarak, daha küçük firmalar tarafından kullanımı yeni rekabet şekli ile tanıştıracak bir potansiyele sahiptir ve yerleşmiş endüstrileri rahatsız etmektedir.

“Büyük Veri” yönetiminin nerelerde kullanıldığıyla ilgili birkaç örnek olarak şunları söyleyebiliriz:

2.1.11.1. *Finansal Servisler*

Bu alan belki de Büyük Veriden faydalanma açısından en geniş alanlardan birisidir. Mobil para hizmetlerinden toplanan veri sektörler ve bölgeler arasında harcama ve tasarruf alışkanlıkları içerisinde derin kavrayış sağlar. Dijital ödeme tarihleri, emanetler ve diğer kredi tabanlı finansal servisler için onların talep ettikleri kredi tarihlerini yapılandırmak için kişilere imkân verir.

Bankalar, müşterileriyle ilgili olarak toplayıp sakladıkları bilgiler yoluyla, kullanıcıyı tanıyan, internet şubesine o gün ne için girdiğini bilen ve buna göre ana sayfayı, menüyü en etkin hale getiren, müşterisine hatırlatmalar yapan, özelleştirilebilir ara yüzler sunan, zengin içerikli, hızlı ve kullanışlı bir 7/24 şube haline geldi.

Sosyal medya/ağlar her geçen gün insan hayatındaki yerini büyütmektedir. Twitter üzerinde 2008 Beijing olimpiyatları ile ilgili 6 milyon kullanıcı tweet atarken, 2012 Londra olimpiyatlarında kullanıcı sayısı 83 katına çıkarak 500 milyonu bulmuş, bir günde atılan tweet sayısı ise 300 binden 400 milyona çıkarak 1333 kat artış göstermiştir. 2013 yılı itibarıyla Twitter’da günlük atılan tweet sayısı 500 milyonu aşmakta, Facebook kullanıcı sayısı 1 milyarı, LinkedIn kullanıcı sayısı 250 milyonu geçmiş durumdadır. Bu rakamlar, bu kaynaklardan üretilebilecek yapısal olmayan verinin boyutu hakkında fikir vermektedir. Üretilen bu kadar veri içinden kendilerini ilgilendiren kısmı hızlı bir şekilde filtreleyip ayrıştırma, işleme ve saklama için bankalar büyük veri teknolojilerinden yararlanabilir.

Müşterilerin yaşadığı kötü bir deneyim hızla sosyal medya/ağlar üzerinden yayılabilmektedir. Ernst ve Young’un yaptığı bir ankete göre, Amerika Birleşik Devletleri’nde müşterilerin %63’ü bankacılık ürünleri ile ilgili kişisel ağlarındaki insanların fikirlerine güvenmekte ve müşterilerin %45’i aldıkları hizmetle ilgili sosyal medya üzerinden paylaşımda bulunmaktadır.

Bankalar bu yapısal olmayan verileri içindeki yapısal olan verilerle entegre ederek müşterileri ve hizmetleri ile ilgili düzenlemeler yapabilir, gerçek zamanlı aksiyonlar alabilir. Örneğin banka hakkında kötü bir tecrübe paylaşan müşteriye bunu telafi etmek

için bir teklifte bulunulabilir veya kredi ihtiyacı olduğunu belirten bir müşteriye anında uygun bir teklif sunulabilir. Yine müşteri olmayan insanların paylaşımlarına göre ihtiyaçlarına uygun ürünler önerilebilir. Bankalar aynı zamanda sosyal medya üzerinde rakipleri ve ürünleri hakkındaki düşünceleri de yakalayabilir.

Bununla birlikte müşterilerin sosyal medya/ağ erişim bilgilerini bankalar ve diğer kurumlarla paylaşmak istememesi bu noktada karşılaşılabilecek bir engel olarak göz önüne alınmalıdır, bankalar bu açıdan paylaşım teşvik çalışmaları yapabilir.

Çağrı merkezleri bankaların vazgeçilmez kanallarından biri ve müşterilere istedikleri bilginin ve bankacılık servisinin verilmesi ve müşterilerin iyi/kötü geri dönüşlerini almak için kullanılan dağıtım kanalıdır. Çağrı merkezi verimliliği için kullanılan ve yapısal verilerden elde edilen ortalama yanıt verme süresi, ortalama görüşme süresi, ortalama aktif/bekleyen çağrı sayısı birçok metrik olmasına rağmen aslında çağrı merkezinin en büyük verisini oluşturan, yapısal olmayan çağrı merkezi görüşme kayıtlarının analizi oldukça önemlidir.

Ayrıca yine bu alanda dijital ortamlarda dolandırıcılık tespit konusu çok önemlidir. Buna da en çok ihtiyaç duyan bankalar ve benzeri finans kuruluşlarıdır. Bankacılık gibi kritik bir finansal sektörde en önemli konulardan biri güvenlidir. Her banka için her an dolandırıcılık tehdidi mevcuttur. ABD Adalet Bakanlığı Tüketici Koruma Ağı'nın verdiği rakamlara göre dünya çapındaki toplam kredi kartı dolandırıcılığı 5,5 milyar USD olup her geçen gün artmaktadır. Bu yüzden bankalar dolandırıcılığı tespit etmek ve engellemek için sistemler geliştirmeye ve bunları sürekli iyileştirmeye yönelik büyük yatırımlar yapmaktadır.

Dolandırıcılık tespitinin gerçek zamanlı yapılabilmesi için çeşitli kanallardan yüksek hızda gelen büyük hacimdeki verilerin gerçek zamanlı olarak işlenerek çıkarımlar yapılması gerekmektedir. Geleneksel sistemler genellikle geçmişe yönelik bu tarz durumların tespitini yapabilmektedir. Ancak büyük veri teknolojileri ile satış noktası, sosyal medya, müşteri veri tabanları ve veri sağlayıcılardan alınan yüksek hacimli veriyi kullanarak gerçek zamanlı tespit ile anında aksiyon alınabilmesi mümkün olmaktadır.

Büyük Veri teknolojilerinin dolandırıcılığı önleme için kullanımında öncelikle bankalarda bulunan verilerin analizi, müşterilerin aktivite desenlerinin çıkarılması

önerilmektedir. Bir kullanıcının hangi periyotta hangi kanaldan ve hangi lokasyondan bankaya giriş gerçekleştirdiği, hangi işlemi yaptığı, kullanıcının tuş dinamikleri gibi bilgiler dolandırıcılık tespitinde kullanılan baş yöntemlerdir. Aktivite desenlerinin çıkarılması esnasında bankacılık kanallarının izole olarak ayrı ele alınmaması, tespit etmesi zor çapraz kanal dolandırıcılıklarını yakalamak için bütün kanallardan yapılan işlemlerin göz önüne alınması gerekmektedir. Büyük Veri ayrıca davranışsal kimlik doğrulamaya da olanak sağlamaktadır.

Dolandırıcılık tespitinde hileli aktivitelerin normal aktivitelerden ayrılmasında ayrıca mobil veriler ve sosyal ağlardaki paylaşımlardan da yararlanılabilir. Örneğin bir banka müşterisinin tatile çıktığındaki harcamalarında bankanın dolandırıcılık izleme sistemi müşterinin mobil veri ve sosyal ağlarda yapmış olabileceği tatil ile ilgili paylaşımları da göz önüne alırsa yanlış pozitif sonuçların üretilmesi önlenir. Bununla birlikte dikkate alınması gereken bir konu da müşteri gizliliği ile ilgili yasal düzenlemelere uygun bir şekilde yasal sınırlar içinde kalmaktır. Dolandırıcılık tespitinde izlenen yaklaşımlar aynı zamanda kredi riski için de izlenebilir.

Bankaların geleneksel veri ambarı modellerinde çok büyük miktarda veri, günlük yüksek hacimde veri aktarımı, saatlerce süren sorgular mevcuttur. Bu yüzden bankalar sorgu performanslarında ciddi kazançların yollarını aramaktadırlar. Ayrıca bankaların geleneksel veri ambarı yaklaşımlarında mevcut durumda cevap aradıkları sorular için bazı çıkarımlar yaparak o çıkarımlara gerek duydukları verileri saklama, şu anda cevap aramadıkları konularda kalan verileri silme eğilimi bulunmaktadır. Bu eğilime neden olarak verilerin saklanma maliyeti, korunma ve yönetim zorlukları ve veri boyutu büyüdükçe sorgulama performansının kötüleşmesi örnek verilebilir. Ancak değişebilen iş ve piyasa koşulları ileride bankalar için cevabını aradıkları yeni sorular ortaya çıkabilmekte ve bu silinen verilerden çıkarımlar yapılmasına ihtiyaç olabilmektedir. Büyük Veri teknolojilerinin sağladığı imkânlar ile minimum veri silme, büyük miktarda veri saklama ve istenilen bilgiye tüm bu verilerden erişilebilme sağlanabilmektedir.

Var olan müşterileri elinde tutmak ve yeni müşterileri cezbetmek bankaların en öncelikli hedeflerinden birisidir. Banka müşteri verilerinin kapsamlı olarak öngörülerek yapılan analizi ile sunulacak yeni ürün, servis ve hizmetler ile rekabet avantajı ve yeni gelir kapıları elde edilebilir. Bankalar her ne kadar hesap hareketleri ve bölümlenme ile

müşterileri hakkında bilgi sahibi olsalar da büyük veri teknolojileri ile birlikte müşterilerin davranış ve ihtiyaçları daha derin ve öngörü ile analiz edilerek teklif optimizasyonu ve çapraz satış gibi konularda fark yaratılabilir, gerçek zamanlı kişiselleştirilmiş ürünler ve servisler sunulabilir.

Geleneksel bankacılık sistemlerinde pazarlama ve ürün geliştirme departmanlarının müşteri verilerine erişimi kısıtlıdır. Bu yüzden örneğin yeni bir kredi kartı ürünü geliştirileceği zaman hangi müşterilerin kredi kartının olduğu, dahası bu müşterilerin başka bankalardan daha çok harcama yaptıkları başka kredi kartlarının olup olmadığı bilgisini çıkartmak oldukça zordur. Fakat birçok veri kaynağından veri alan yüksek veri içeren Hadoop kümeli veri ambarı dahilinde bunun ve hangi müşterilerin hem mobil bankacılığı kullanıp hem de çek bozdurmak için şubeye uğradığı sorusunun cevabı kolaylıkla alınabilmektedir. Ayrıca çapraz kanal aktiviteleri izlenerek müşterilerin ne kadar dijital olduğunu gösteren yaşam stili skoru modeli geliştirilebileceği bilgisi paylaşılmaktadır. Bu tarz bir veri ambarında edinilebilecek bilgilere başka bir örnek olarak bankayla çalışmayı bırakan müşterilerin aktivitelerinin analizi sırasında fatura ödeme işlemi gerçekleştirilmeyi bırakan müşterilerin genellikle yaklaşık 6 ay içinde bankayla çalışmayı bıraktığının tespit edilmesi verilebilir. Buna göre artık fatura ödeme işlemi yapmayan bir müşterinin bankayla çalışmayı bırakabileceği öngörüsü ile banka müşterisini kalmaya ikna edecek kişiye özel faaliyetlerde bulunabilir.

İş analitiğinin hızlandırılması da Büyük Verinin getirdiği çok büyük bir avantajdır. Fortune 20 şirketlerinden birisi Büyük Veri teknolojilerinden yararlanarak risk raporlama süresini 45 dakikadan 45 saniyeye indirerek daha hızlı karar verebilmeyi ve iş yapabilmeyi sağlamıştır. Ayrıca tüm bu verilerin iç süreçlerdeki detaylı analizi ile operasyonel verimlilik sağlanabilir.

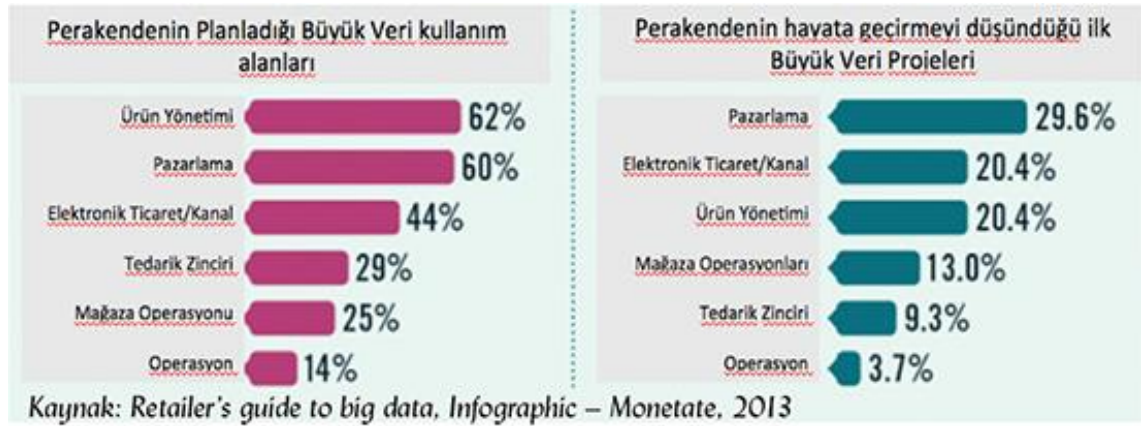
Yine bir finansal kurum içindeki sistemle ilgili tutulan birçok veri vardır. Kullanıcıların dosya sistemi erişim bilgileri, internet erişim bilgileri, veri erişim bilgileri, e-postalar, güvenlik kayıtları, uygulamaların ve donanımların ürettiği sistem günlükleri, değişiklik ve erişilebilirlik günlükleri örnek olarak verilebilir. Yasal zorunluluklardan ötürü yapılan dış denetimlerde tüm bu verilerin incelenmesi, çıkarımlarda bulunması gerekebilmektedir. Geleneksel veri tabanı araçları bu noktada zorlanmakta ve günlerce çalıştırılan sorgularla karşılaşabilmektedir, arşivlenmiş verilerin erişilebilir hale

getirilmesi sürecinde uzun süren birçok işlem yapılması gerekebilmektedir. Yine iç denetimciler tarafından sürekli kontroller gerçekleştirilmektedir. İç denetçiler tarafından veri büyüklüğü sıkıntısından ötürü örneklenmiş veriler üzerinden bilgi çıkarımına gidilebilmektedir. Büyük veri teknolojilerinden yararlanarak bu büyük miktardaki verilerin saklanması, yönetimi daha kolay ve istenilen bilgiye erişimi çok daha hızlı olmaktadır, böylece tüm veri üzerinden kaliteli denetimler yapılabilmektedir.

2.1.11.2. *Pazarlama ve Perakendecilik*

Hem klasik ticaret hem de e-ticaret alanında Büyük Veri çok etkin bir biçimde kullanılan araçlardan birisidir. Dünyanın en büyük alışveriş sitelerinden alibaba.com, amazon.com, ebay.com müşterileriyle ilgili her türlü bilgiyi satışa dönüştürmek için Büyük Veri analizlerinden faydalanmaktadırlar.

Büyük Veri pazar bölümlenmesi, müşteri değerlendirme ve çapraz satış analizleri, satış noktası veri analizleri, alışveriş sepeti analizleri, tedarik ve mağaza yerleşim optimizasyonları şeklinde kullanılarak satışları ve karlılığı arttırmaya yönelik çalışmalar yapılabilmektedir.



Şekil 2.9: Perakende alanında Büyük Veri kullanım ve projeleri.

(Kaynak: Retailer's Guide to Big Data, Infographic, 2103)

2.1.11.3. *Kamu ve Hükümetler*

Hükümetler; yurttaşlarına yönelik bilgi ve hizmetleri işleyip, saklama konusunda oluşan çok büyük ölçekli veri ile çalışmak zorundadırlar. Örneğin, RTÜK kararları gereği,

ülkemizdeki televizyon kanallarının son bir yıllık yayınlarını saklama zorunluluğu vardır. Saklanacak bilgiler, "Büyük Veri" olarak tanımladığımız türdendir.

Kamu verisinin potansiyel değeri yüksektir. Bu verinin içinden gizli ve kişisel bilgi içeren kısımlar ayıklanıp, araştırmacılar, özel şirketler ve girişimciler ile paylaşılabilirse bu veriler üzerinde çalışan katma değerli hizmetler önemli ekonomik faydalar sağlayabilir. Kamu verisinin faydaları ekonomiye katkı, çevrenin korunmasına katkı, bilimsel gelişime katkı ve katılımcı demokrasiye katkı ana başlıkları altında incelenebilir (European Commission, Aralık 2011, Open data, An engine for innovation, growth and transparent governance).

McKinsey Global Institute raporuna göre kamu büyük verisi kullanımı ABD sağlık sektöründe yılda 300 milyar dolar, AB kamu sektöründe yılda 250 milyar avroluk değer yaratabilecek potansiyele sahiptir (McKinsey Global Institute, 2011, Big data: The next frontier for innovation, competition, and productivity).

AB tarafından hazırlatılan başka bir çalışmaya göre ise daha iyi veri altyapısı ve veriye erişim engellerinin azaltılmasıyla kamu sektörü verisinin AB ekonomisine doğrudan ve dolaylı etkisinin toplamı 2008 itibariyle 200 milyar avro seviyesinde tahmin edilmiştir (Graham Vickery, Information Economics, Paris, 2011, Review Of Recent Studies On Psi Re- Use and Related Market Developments).

Ayrıca, günümüzde e-devlet kavramı oldukça kritiktir. E-devlet uzmanlarının en önemli hedefi bilgiye eş zamanlı olarak ulaşmak ve daha iyi hizmet vermektir. E-devlet uygulaması gerçekleştirilen ülkelerde kamu kuruluşları ziyaretçilerin sayfalarını nasıl kullandığı, ihtiyaç duyulan formlara kolayca ulaşıp ulaşılamadığı, web sayfa tasarımının nasıl en iyi kullanılabilir hale getirilebileceği, hangi sayfaların hangi sıra ile ziyaret edildiğinin anlaşılması, geçmişteki ziyaretçi davranışlarına göre kurumun web sayfasını vatandaşın ihtiyacına daha iyi yanıt verecek şekilde nasıl yeniden düzenlenmesi gerektiği sorularına çözüm aramaktadırlar. Bu anlamda Büyük Veriden faydalanabileceklerdir.

2.1.11.4. *Eğitim*

Mobil katma değerli servislerin kullanımından elden edilen veri, kritik bilgiyi yaymak için daha fazla odaklama ve zamanlama girişimlerine izin vererek, bilgi boşluklarını ve eğitim ihtiyaçlarının anlayarak kamu sektörünün gelişmesi için kullanılabilir.

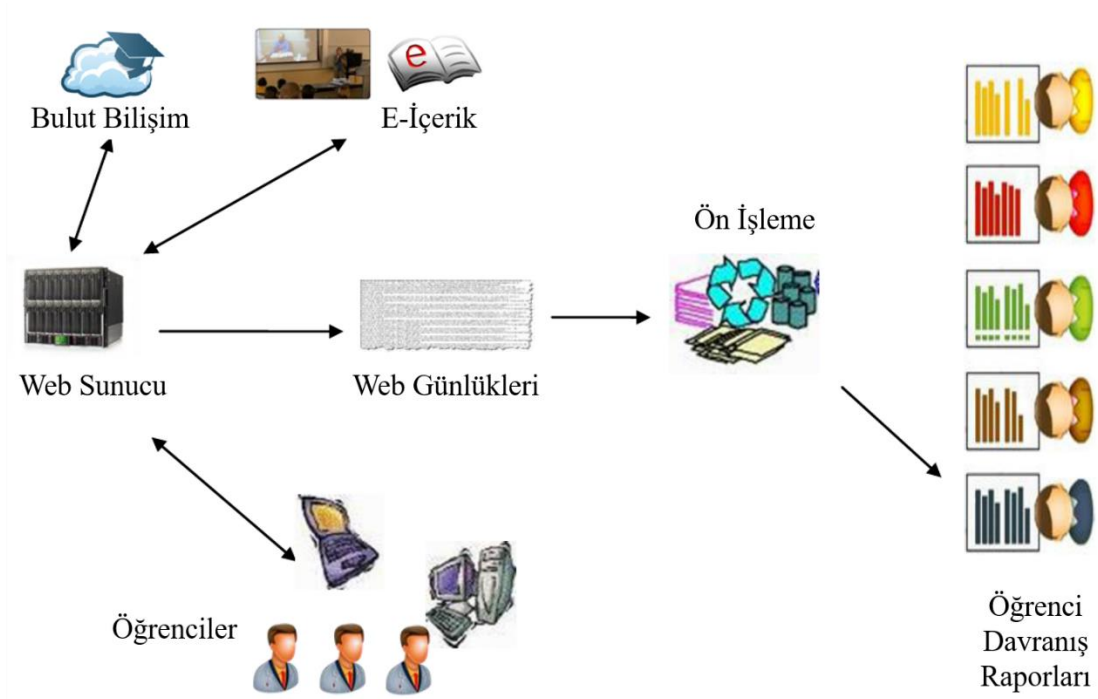
Eđitim sektöründe, verilerin işlenmesi ve buna ihtiyaç gün geçtikçe artmaktadır. İnternet teknolojisinin gelişimindeki artış eğitim kurumlarındaki “tebeşir ve konuşma” sistemini deđiştirdi. İnternet ve bilişim teknolojilerinin etkisiyle web tabanlı uygulamalar hayata geçirildi.

Modern ve saygın teknik yükseköğretim kurumlarının çođu şu anda kâğıtsız çevrimiçi eğitim sistemi ile çalışmak için gerekli adaptasyonu sağlamış durumdadır. Bu teknolojilere adaptasyon, yeni nesil eğitim sistemini hayata geçirmede daha iyi iletişim ve daha gelişmiş yetenekler sunarak, baskıyı azaltmakta ve zaman kazandırmaktadır.

Kısaca şu faktörler yükseköğrenim kurumlarında Büyük Veri neslini oluşturmaktadır:

- **Akademik Trend:** Yükseköğrenim kurumlarının çođu farklı yönetsel iş amaçları için çeşitli özelleştirilmiş araçlar uygulamaya koymaktadırlar. Örneđin, öğrenci kayıt kabul, ücret toplama, öğretmenler ve diđer çalışanlar için personel takibi, rapor sistemi vb. iç deđerleme, devam kayıtları, not tabloları, bilgisayar laboratuvarı iş yükleri ve planlama gibi günlük öğrenci aktiviteleri büyük miktarda veri üretmektedir.
- **Performans İzleme:** Kurumların sağlıklı istatistiksel veriler elde edebilmeleri için bütün mezun ettikleri öğrencilerinin verilerini de korumaları gerekmektedir. Mezun olan öğrenci raporları, devam eden öğrenciler ile ilgili öngörude bulunmakta fayda sağlamaktadır. Öngörü analizlerinin çıktılarından faydalanarak, öğrencinin yetenekleri belirlenir ve üzerinde durulur, eksik yanları tespit edilir ve tamamlanır. Böylece, öğrenciler rüyalarındaki işi elde etmek için en iyi donanıma yükseltilirler.
- **Sanal Sınıflar:** Akademik eğitimdeki yüksek rekabetin daha da artması ile öğrenciler yeteneklerini daha etkin bir biçimde güncelleyebilecekleri çeşitli mekanizmalara adapte oluyorlar. Gelişmiş eğitim yazılımlarıyla, uzman eğitimcilerin rehberlikleri sanal toplantılar ile kolayca sağlanmaktadır. Çeşitli tanınmış kurumlar şu anda zaten bulut tabanlı sanal sınıflara adaptasyonu sağlamış durumdadır. Bu sınıflarda öğrencilerine en iyi eğitim paketleri ve çeşitli dijital içerikler sunabilmektedirler. Bu dijital içerikler genellikle PowerPoint sunumlar ve ses dosyaları halindedir. Bazı durumlarda sanal sınıflar öğrencinin sistemde onaylanması ile kayda izin vererek çevrimdışı de görünebilmektedir. Bu

nedenle, bu tip sanal sınıflar muazzam veri üretirler ki, geleneksel sunucular da barınamazlar ve bulut tabanlı depolama sistemlerine ihtiyaç duymaktadırlar.



Şekil 2.10: Öğrenci davranışları analizi yaklaşımı.

(Kaynak: <http://research.ijcaonline.org/icacthpa2014/number4/icacthpa6041.pdf>, 2014)

2.1.11.5. Sağlık

Mobil aygıtlardan toplanan veri sağlık çalışanlarınca elde edilmiş, bireyler tarafından sunulmuş ya da serbest veri şeklindeki analizle, salgınları durdurma ya da sağlık eğilimlerini anlamada önemli bir araç olabilir. Bireysel elektronik sağlık kayıtları bağlamında toplandığında, bu veri sadece bireylerin bakımının sürekliliğini geliştirmez, aynı zamanda tedaviler ve çıktılar ile büyük veri setleri oluşturmak için kullanılabilir ve bunları etkin verim ve maliyetle karşılaştırabilir.

Hastaneler; hastalarına yönelik etkili, bireysel, kişiselleştirilmiş, tıbbi hizmetler sunabilmek için, bireysel bazdaki verileri kendi sayısal ortamlarında depolamaktadırlar.

İlaç sanayiinde; örneğin "kanser araştırmaları" için oluşturulan büyük genomik veri tabanları, araştırmacıların sürekli erişimine açık olmak durumundadır.

Aslında şunu söyleyebiliriz: Araştırmacılar halen sağlık alanında Büyük Verinin sınırlarını ve tanımlarını tartışırken, Büyük Verinin sağlık alanındaki faydaları şu ana kadar üç ana alanda kendini göstermektedir. Bunlar, hastalıkları önlemek, hastalıklar için risk faktörlerini değişebilir olarak tanımlamak ve farklı sağlık davranış değişimleri için müdahaleler tasarlamak olarak ifade edilebilir. Dünya çapında kurumlar Büyük Veri hareketini fark ediyor ve veri destekli karar alma ve bilgiyi keşfetmek için yeni girişimleri uygulamaya koyuyor. Örneğin, Ulusal Sağlık Enstitüsü (The National Institute of Health - NIH) geniş çaplı biyomedikal veri analizlerini kolaylaştırmak için paylaşılan sayısal bir çevre olan (veri standartları, ontolojiler, veri katalogları, sanallaştırılmış bulut bilişim vb.) Altyapı Plus Programı ve Enformasyon için Büyük Veri kuruyor. Özellikle, NIH Tıp Kütüphanesi etkileyici bir zengin veri seti kaynağına ev sahipliği yapmaktadır. Bunlar öncelikle biyomedikal veri ve NIH'nin finanse ettiği araştırmacılar tarafından gelen bilgi paylaşım sistemlerinden oluşmaktadır. İlaveten, Birleşmiş Milletler (The United Nations - UN) Küresel Nabız Projesini başlatmaktadır. Bu proje ayrıca, bireylerden ve kurumlardan veriye katkı sağlanmasını isteyerek “veri hayırseverliğini” de savunmaktadır.

Sağlıkta Büyük Veri akışı üç kategori içerisinde özetlenebilir:

- **Geleneksel Medikal Veri:** Öncelikle sağlık kurumları orijinelidir. Örneğin; EMR, kişisel ve aile sağlığı tarihçesi, ilaç tedavisi geçmişi, laboratuvar raporları, patoloji sonuçları. Bu verilerin analizinden hedeflenen; hastalık çıktılarını ve kişilerin risk faktörleri daha iyi anlamak, sağlık sistemi masraflarını düşürmek ve etkisini arttırmaktır.
- **Omik Veri:** Gen bilimi, protein bilimi, mikrobiyotik, metabolomiks vb. gibi biyolojik ve moleküler alanların genel adına olarak tanımlanan “Omics”; büyük veri setleri oluşturmaktadır. Bu verilerin analizi ile amaçlanan da; hastalıkların mekanizmalarını anlamak ve “en doğru ilaç” gibi tıbbi tedavilerin kişiselleştirilmesini hızlandırmaktır. 2013 Stanford Biyotıpta Büyük Veri konferansında Alice Whittmore şu noktaya değindiği gibi, genomik ile test ve haritalama, örneğin, göğüs kanser gelişimi yüksek risk içindeki kadınları işaret edebilir. Bu bilgiler tıp insanlarına, diğer düşük riskli kadınlar için önceden,

potansiyel tehlikelere erken müdahale, geniş çaplı ihtiyaçları azaltma ve öncelikli tedaviyi sağlamaya imkân vermektedir.

- **Kişisel Veri:** İnternet kullanımı, sosyal medya, mobil uygulamalar, sensörler, giyilebilir aygıtlar, diğer teknolojik ya da teknolojik olmayan bilgi işlem cihazları kişilerin davranışları ve sağlıkla ilgili işaretleri kendi yaşamlarında ve hareket tarzlarından alarak daha iyi bilgi sağlar ve kişilerin yaşam sağlığını arttırmaya katkıda bulunurlar.

Spesifik birkaç örnek vermek gerekirse; Ontario Üniversitesi her gün, erken doğan bebeklerden (prematüre) yaklaşık 100 milyon adet veri toplayarak, analizini en hızlı şekilde gerçekleştiriyor. Bunun sonucunda, hasta muayenesi sırasında erken teşhis edilen değişimler, bir hastalık durumuyla ilişkilendirilebiliyor. Asya Sağlık Bürosu, hasta görüntüleme verilerini Hadoop üzerinde tutup analiz ederek radyoloji ve patoloji uzmanlarının hem daha hızlı hem de daha az hata yaparak teşhis koymalarını sağlamıştır.

Aşağıdaki tabloda Büyük Verinin sağlık alanında nasıl kullanıldığı ve örneklerini özet olarak görmekteyiz.

Tablo 2.2: Sayısallaştırılmış-Kişisel hareket ve sosyal medyanın Büyük Veri ile sağlık alanında kullanılmasına örnekler.

| Veri Tipi | Sağlık Alanında Nasıl Kullanıldığı |
|--|--|
| Sayısallaştırılmış-Kişisel Veri (aygıtlar, kendi raporları ve sensörler vasıtasıyla) | Kişisel ya da grupların proaktif davranışlarının izlenmesi / Biyolojik, fiziksel, davranışsal ya da çevresel potansiyel risk faktörleri üzerinde daha zengin ve detaylı bilgi sağlar / Standart sorgular kullanarak şu anda mümkün olandan daha uzun takip sürelerinde veri toplanmasına izin verir |
| | Örnekler |
| | Gıda tüketimi / Beslenme bilgisi / Beklenmedik kas hareketleri / Kahve tüketimi, sosyal etkileşim ve ruh hali / Fikir izleme süreci / Bir soluk alma cihazı sensörü ile astım ilaçlarının kontrolü ve yardım kullanımı / Şeker hastalarında kan glikoz seviyesini izleme / Psikolojik, mental ve bilişsel durumlar ve tedavileri / Fiziksel aktiviteleri izleme / Diyet / Uyku kalitesi / İlaç bağımlılığı |

Tablo 2.2 (devam): Sayısallaştırılmış-Kişisel hareket ve sosyal medyanın Büyük Veri ile sağlık alanında kullanılmasına örnekler.

| Veri Tipi | Sağlık Alanında Nasıl Kullanıldığı |
|---|--|
| Konum tabanlı bilgiler | Global konum bilgilendirici servislerden bilgi sağlama / Çevresel ve sosyal belirleyicilerden bilgi sağlanması / Konum civarındaki salgın hastalıkların izlenmesi |
| | Örnekler |
| | Hava örnekleri / Kirlilik seviyeleri / Alerjenler / Trafik modelleri / Su kalitesi / Çevre yürünebilirliği / Taze meyve ve sebze erişim (ör. süpermarketler gibi.) / Sağlık Haritası |
| Veri Tipi | Sağlık Alanında Nasıl Kullanıldığı |
| Twitter (Not: 2011 deki bir çalışmaya göre; İngilizce tweetlerin %8,5'i hastalıkla ve %16,6'sı da sağlıkla ilgilidir) | Gerçek zamanlı hastalık yayılmasını değerlendirme / Duygu ve ruh hallerini değerlendirme / Tıbbi asistan ihtiyacı olan ve izin veren kişilere acil hizmetini kolaylaştırma / Kriz haritalamayı kolaylaştırma / Acil olmayan sağlık üzerindeki araştırmaları kolaylaştırma |
| | Örnekler |
| | Tıbbi yanlıgıları ölçmek (beyin sarsıntısı gibi) / Antibiyotik kullanımı gibi kötü tedavilerin yaygınlaşması / Kalp durma ve yeniden harekete geçme trendleri / Rahim ağzı ve göğüs kanseri taraması / Doğum sonrası depresyon / Grip salgını (hastalık aktivitesi ve kamu endişesi) / 2010 Haiti kolera salgını / Boston maraton patlamasındaki acil durumlar |
| Veri Tipi | Sağlık Alanında Nasıl Kullanıldığı |
| Sağlıkla ilgili sosyal ağ siteleri | Hastalar ve tüketiciler arasındaki tavsiyeler ve kişisel sağlık paylaşımlarını kolaylaştırma / Kalabalık yoluyla bulaşıcı hastalıkların yaygınlaşmasının izlenmesi |
| | Örnekler |
| | PatientsLikeMe.com / Katılımcıların raporlarının toplandığı hastalık gözetim siteleri / Bulaşıcı salgın hastalıklar hakkında bilgi yayılması, haritalanması ve analiz edilmesini destekleyen resmi olmayan veri kaynakları (ör. Flu Near You, HealthMap, GermTracker, Sickweather) |

Tablo 2.2 (devam): Sayısallaştırılmış-Kişisel hareket ve sosyal medyanın Büyük Veri ile sağlık alanında kullanılmasına örnekler.

| Veri Tipi | Sağlık Alanında Nasıl Kullanıldığı |
|---|--|
| Diğer sosyal ağ siteleri (ör. Çevrimiçi tartışma, Facebook) | Hastaların endişelerini tartışmak için sosyal medyayı nasıl kullandıklarının izlenmesi / Sokaktaki insanın ne söylediğinin farkındalığını sağlar |
| | Örnekler |
| | Yan etkiler ve ilgili ilaç bağımlılık davranışları (ör. ilaç değiştirme ve kesme) |
| Veri Tipi | Sağlık Alanında Nasıl Kullanıldığı |
| Arama sorguları ve web günlükleri | Nüfus seviyesi sağlık davranışlarının yaygınlaşması için daha yüksek öngörülede bulunmak / Arama kelimesi seçimi güvenilir küratör sağlık içeriğine varış için kritik olabilir / Web günlüklerinden tıklama akışlarının yönü kişisel özellikler hakkında bilgilendirici olmaktadır, ör. ruh sağlığı ve beslenme tercihleri |
| | Örnekler |
| | Google ve Yahoo arama sorguları hastalık salgınlarının öngörülmesi için kullanılmaktadır, örneğin; Grip (Google 2013) / Deng Humması / Ruh sağlığı, depresyon ve intihar mevsimselliği / Kene ısırıkları / Sigara içmenin önlenmesi ve elektronik sigara kullanımı |

Bu bağlamda Türkiye’de de SGK, toplamış olduğu büyük miktardaki yapılandırılmış ve yapılandırılmamış veriler üzerinde çeşitli analizler yaparak verimliliği artırmak, kayıp-kaçak oranlarını düşürmek ve hizmet kalitesini yükseltmek için büyük veri konusunda çalışmalara başlamıştır (Özbilgin, 2015).

2.1.11.6. *Tarım*

Tarım ürünleri, girdi satın almaları ve sübvansiyonlar (devlet yardımları) için mobil ödemeler öngörülen gıda üretim trendleri ve teşvik tedbirlerine hükümetlere daha fazla yardımcı olabilir. Bu bilgi uygun ürün depolama kullanılabilirliğini sağlamak, israfı ve bozulmayı azaltmak ve çiftçilerin hangi tip finansal hizmetlere ihtiyacı olduğunu bilgisini

daha iyi sağlamak için kullanılabilir. Mobil kullanım şekilleri sıkıntılı bölgeleri belirlemede hükümetlere ve kalkınma örgütlerine yardımcı da olabilir. Erken teşhis tarımsal üretimin daha da azalmasını ve ailelerin topraklarından ayrılmasını engelleyebilir.

2.1.11.7. *Enerji*

Enerji firmaları, akıllı şebeke ve sayaçlar kullanarak, abonelerinin bireysel kullanımlarıyla ilgili oluşan verileri saklayıp işlemek durumundalar.

İşte çağımızda bunlar gibi birçok uygulama alanında kendine yer bulan, fakat işlenmesi, saklanması ve irdelenmesi önceki nesil veri tabanlarından farklı yöntemler gerektiren büyük ölçekli verilere, kısaca “Büyük Veriye” odaklanmak durumundayız.

Bunlarla birlikte şunu da göz ardı etmemek gerekir, Uydu/Harita Dizgeleri (GPS), Akıllı Gezgin Telefonlar (GSM), çok yüksek çözünürlüklü fotoğraflar çekebilen yeni nesil kameralarla üretilen, ses ve görüntü bilgileri; saklama ortamlarının sınırlarını zorlayıp, verimliliklerini düşürüyor. Her türlü gezgin aygıtlar üzerinde çalışabilen, internet tabanlı yazılım ve uygulamalarca üretilenlerle birlikte, Facebook, Twitter, Google Plus, Instagram vb. gibi toplumsal medya ortamlarında kullanıcılarca üretilen bilgilerin saklanması gereği, bilişim konusu ile ilgili girişimcileri de, Büyük Veri alanına yönlendirmektedir. Bu da yeni bir “Büyük Veri” sektörü oluşturmaktadır.

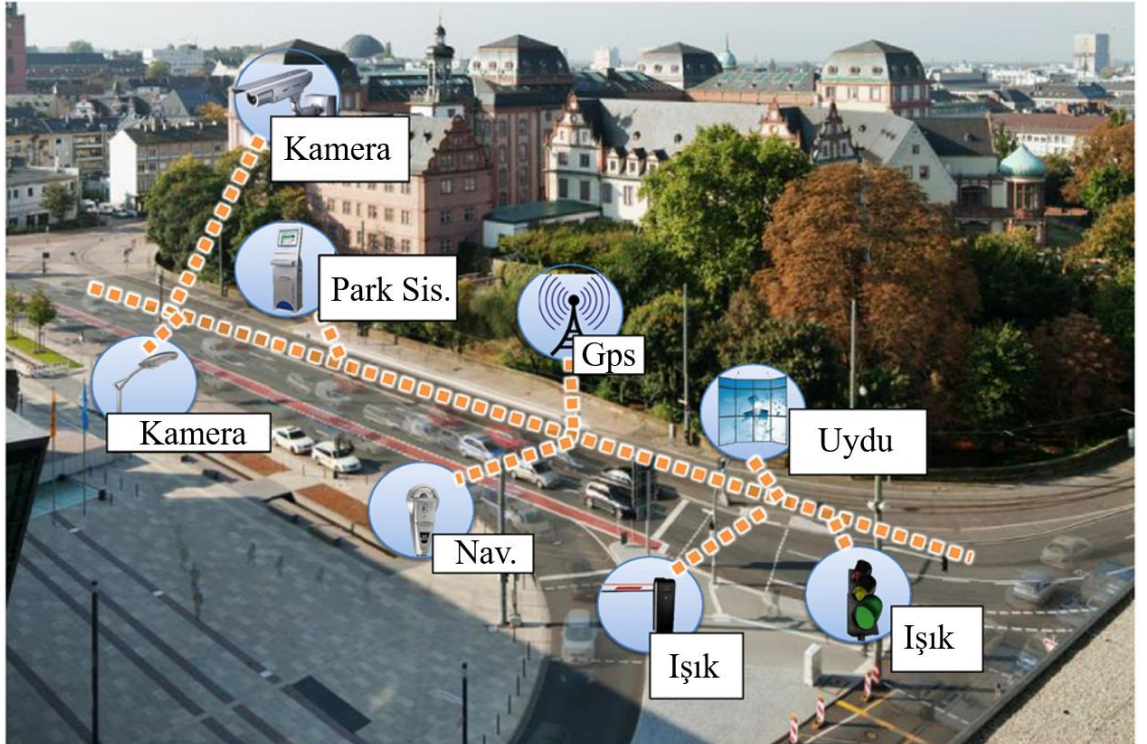
İnternet üzerindeki üretici ve tüketicilerin veri üretimini hızlandırması, başta servis sağlayıcı firmalara olmak üzere, büyüyen bilgiyi harmanlayıp, anlamlı biçime dönüştürerek kullanıma yeniden sunma görevi yüklüyor.

Örneğin, Twitter’da her 24 saatte neredeyse 100 milyon tweet atılıyor. Twitter kullanıcıları, lokasyon bildiriyor, hashtag’lerle yorumlarda bulunuyor. Tüm bu veri birikimi akıllı bir yazılımla düzenlenerek, işletmeler için hedef kitleler hakkında bir veriyi anlamlandırabilir. Sonrasında bunları kendi verileriyle birleştirerek satış sürecini yeniden yorumlayabilir. Bu şirketler için belki de ciddi bir rekabet ve zorluk alanı, aynı zamanda da doğru müşteriye odaklanabilmek için bir fırsat olabilir.

2.1.11.8. *Haberleşme ve Ulaşım*

Telekom sektöründe en önemli sorun müşteri kaybıdır. Kuruluşlar hangi müşterilerini kaybedebileceklerini önceden belirleyebildikleri takdirde bu müşterilerini elde tutma amaçlı stratejiler geliştirebilir, düşük maliyetli ve etkili kampanyalar düzenleyebilirler. Kaybetme olasılığı olmayan bir müşteriye kalıcılığını sağlama amaçlı bir mesaj göndermek hem müşterinin kendisine verilmek istenen mesajın ne olduğunu algılamasını zorlaştıracak hem de maliyetleri artıracaktır. Örneğin, Amerika'nın en büyük kablosuz iletişim sağlayıcısı olan Verizon kaybetme olasılığı yüksek olan müşterilerini ve müşteri kaybına neden olan faktörleri belirleme amaçlı bir çalışma yapmıştır.

Öte yandan, bugün navigasyon cihazlarının geniş kullanım, trafik sistemlerinin çevrimiçi merkezlere bağlanabilmesi, insanların konum hareketlerinin değerlendirilmesi ile ulaşım konusu da Büyük Verinin etkin kullanılabileceği alanlardandır. Şekil 2.11'de en ayrıntılı uygulamalardan olan trafik ve taşımacılıkla ilgili bir çalışma görülmektedir, bunun gibi diğer tüm alanlarda da benzer uygulamalar yapılmaktadır.



Şekil 2.11: Taşıma sistemleri akıllı servisler entegrasyonu.

(Kaynak: <http://www.sciencedirect.com/science/article/pii/S0167739X13001593>)

2.1.12. Örnek Vakalar: Büyük Veri Yönetimi Çalışması

Büyük verinin zorlukları her ölçekteki kurumu etkiler, ama bunun yanında yeni fırsatlarda oluşturur. Temel bir varlık gibi görünen bilgi aracılığıyla, tüm ölçekteki kurumlar en önemli satış, pazarlama ve müşteri hizmetleri girişimini ve iş büyütme hareketini desteklerler.

Var olan müşteri verisinin bütünüyle hacim, çeşitlilik ve hızı, pazarlama, satış ve müşteri desteğinin yeni anlamlarını araştırırken, bilgi stratejisi ve inovasyon açısından yöneticiler için çok büyük bir umuttur.

Artık bundan sonra yeni dönemde, işletme analistleri ve veri bilimcileri müşteriler için benzersiz deneyimler oluşturduğu zaman fırsatlara kavuşmak için hızlı analitik modelleme ve simülasyon kullanmalıdırlar. Pazarlamacılar yeni teklifler ve pazarlar geliştirirken, tekliflerin rasyonel sınıflandırılması ile müşterilerini eşleştirmek için yeni ve gizli bilgi kaynakları kullanmalıdır.

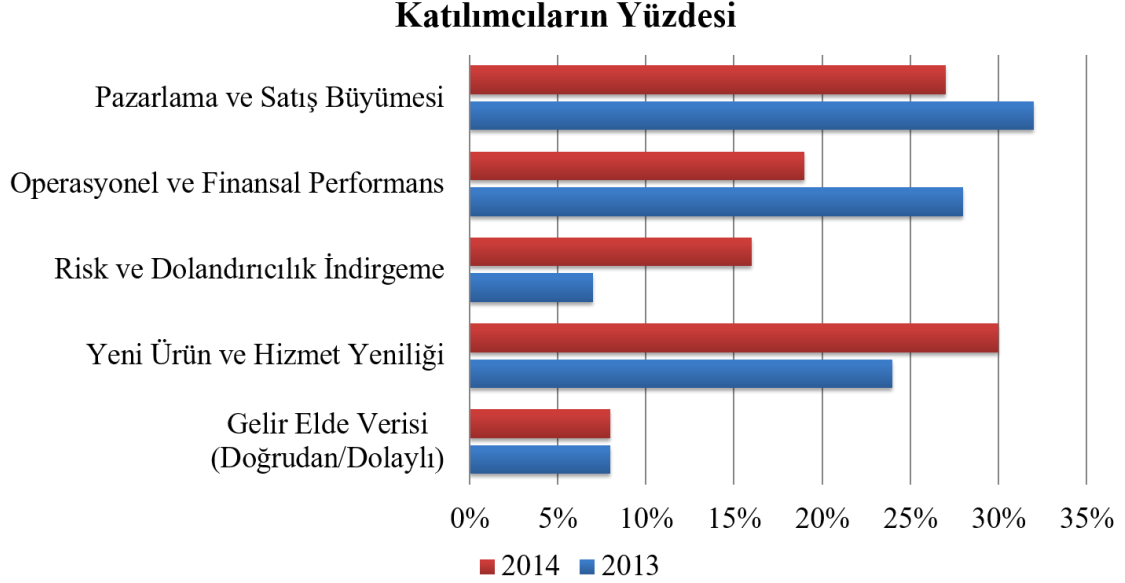
Değerli bir şirket varlığı olarak görünen bilgi aracılığıyla, hemen hemen her kurum Büyük Veri kullanımı için yeni yaklaşımlar, CRM ve müşteri deneyim girişimleriyle yeni rekabet avantajları bulmak için analizler keşfedebilirler.

Tüm coğrafyalarda ve tüm endüstrilerde kurumların en önemli satış, pazarlama ve müşteri hizmeti girişimlerini desteklemek için analizler ve Büyük Veri kullanımı için yenilikçi yaklaşımlar kullanan, bütün ölçeklerdeki şirketlerden birçok örnek vardır. Bu yeniliklerin birçoğu geleneksel veri deposundan ve iş zekâsı girişimlerinden gelmez, bunun yerine Büyük Veri ile ilgili fırsatçı deneyimlerden doğmaktadır.

Aşağıdaki şekil işletme ve bilişim uzmanlarının Büyük Verinin hangi alanlar için en büyük fırsatlar sunduğunu algıladıklarını göstermektedir. Doğrudan müşteri odaklı aktiviteler için Büyük Veri kullanımı örneğin, satış ve pazarlama alanında yüksek kalmasına rağmen, yeni ürün ve hizmet gelişimi ve dolaylı müşteri odaklı aktiviteler için artan bir şekilde tercih edilmektedir.

Burada “Büyük Veri” analizi yaparak problemlerine çözümler üretmiş 3 örnek vakayı inceleyeceğiz. Bu 3 vaka çalışmasında müşterilerin bakış açılarını sezebilmek için analitik çalışmaları destekleyen ve bunu yapmaya çalışırken karşılaşılan zorlukların

üstesinden gelen kurumlara göz atılacaktır. Bu örnekler farklı kurum türlerinden gelmesine rağmen, her biri değerli genel dersler sunmaktadır.



Şekil 2.12: Büyük Verinin Fırsat Sunduğu Alanlar.

(Kaynak: Gartner Büyük Veri web seminerinde veri, <https://www.gartner.com/doc/2703617/customer-analytics-art-possible-big>, Mart 2013 ve 2014).

2.1.12.1. *Örnek Vaka 1: Müşteriler için Benzersiz Deneyimler Oluşturacak Hızlı Analitik Modelleme ve Simülasyon Kullanımı*

New York'taki Story adlı bir perakende mağazası, sıradan bir teknoloji parçası kullanarak müşteri deneyimlerine benzersiz bir bakış açısı kazandı. Prism Skylabs adlı bir teknoloji ortağı ile çalışan Story, kendisinin satış, pazarlama ve müşteri destek deneyimlerini geliştirmek için zaten mağazada bulunan güvenlik kameralarından birçok perakendecinin bir masraf olarak gördüğü teknoloji ile elde ettiği bilgiyi faydalı hale getirmenin bir şeklini buldu.

Problem : Story 3 ila 8 hafta içerisinde, mağaza düzenini tamamen değiştirebildiği yeni bir perakende konsepti oluşturmak için yola çıktı. Bu fikir gelişen trendler, konular ve modalara adaptasyon için her bir kaç haftada bir değişen bir bakış açısıyla, yeni bir

mağaza modeli olmalıydı. Malların düzeni için yer planındaki her şey yeni bir “Story” i yansıtmak için tasarlanmış olacaktı.

Çözüm : Prism ile çalışan Story mağazasında müşteri deneyiminin içerisindeki bakışı algılamak için, ısı haritaları oluşturan güvenlik kameraları kullanıldı. Bu haritalar mağazanın düzenine bağlı olarak, mağazayı ve ürünleri incelemek için gelen müşterilerin nasıl hareket ettiklerini analiz etmeye başladı. Bu müşterilerin neleri beğendiğine bağlı olarak, personelin ve ürünlerin hareket etmesine izin vermektedir.

Sonuçlar : Analitik şekilde optimize edilen bu yaklaşım Story’ye müşterilerine cazip bir ortam kurabilmeyi mümkün kıldı. Müşteriler tipik olarak Story’de 40 dakika harcamaktalar ve 3 haftalık tatil süresi boyunca, müşterilerinin %30’unun alım yapmak için mağazaya geri döndüğü görüldü.

Çok aşılmaz engelleri olmasa bile, Story’nin yaptığı gibi, birçok perakendeci mağaza katlarını henüz bu şekilde analiz etmemiştir. Story Büyük Veri ve analizinin büyük ya da küçük geleneksel perakende dünyası şirketlerine nasıl yardım ettiğini gösterdi.

2.1.12.2. *Örnek Vaka 2: Yeni Pazarlar Gelişirken Rasyonelleştirilmiş Teklif Aralığıyla Müşterileri Eşleştirme*

Satış ve Pazarlama için Büyük Veri ve analizlerin en ortak kullanımlarından birisi şirketin teklifleri ile ilgili müşterileri eşleştirmektir. Bunu akıllıca yapmak için, kurumlar ürün ve hizmet tekliflerinin karakteristikleri ve müşterilerle ilgili karmaşık bilgiye ihtiyaç duyarlar. Trendler, müşteri lezzetleri, ihtiyaçlar ve durumlar zamanla değişebilir, bu yüzden veri taze olmayı ve trendler tatmin edici olmayı gerektirir. Analizler çok fazla varsayımlar yapmamalıdır ve sadece iyi bir eşleştirme olup olmadığını değil, tercihen kabul edilen tekliflerin olasılığını öngörmelidir.

Şirketler tekliflerini rasyonelleştirmelidir. Bugün, birçok şirket üretim, parça envanteri, kullanılabilirlik, uzun süreli satış öngörülleri, teslimat ve hizmet ile ilişkisindeki konulara neden olan, sonsuz varyasyonlar, konfigürasyonlar ve herhangi belirli bir ürün için seçenekleri kolayca üretebilir. Bazı şirketler bu konuları adreslemek için ürün konfigürasyonlarını rasyonelleştirecek/optimize edecek çözümlere dönmüştür.

ABD Washington Tacoma'daki 29 akrelik (~117.363 m²) bir hayvanat bahçesi olan Point Defiance Zoo & Aquarium (PDZA) satış ve pazarlama girişimlerini desteklemek için bilgisayar analizleri kullanmıştır.

Problem : Tam zamanlı 80 çalışanıyla küçük bir kurum olan PDZA gelirlerini arttırmak istedi.

Çözüm : PDZA Büyük Veri analizleri için IBM'in bir iş ortağı olan BrightStar Partners ile çalışmayı seçmiştir. Bu teknolojiyle, önceki ve gelecekteki üyeler ve ziyaretçiler için, daha akıllıca hedeflenmiş kampanyalar yapma yeteneği veren, hayvanat bahçesinin yıllık 600.000 ziyaretçisinin örnek kalıplarını ve davranışlarını inceledi. Örneğin, PDZA'nın üyelikleri sona eren insanlar için e-posta tabanlı bir tanıtıcı kampanya amaçladığını tanıttı. Kişiselleştirilmiş e-postaların hızlı oluşum ve yürütülmesi sayesinde, tanıtımı hayvanat bahçesinin geleneksel doğrudan e-posta kampanyalarının iki katına çıkarmayı başararak, satışa %6 katkı yaptı. PDZA'nın müşteri analizleri doğru zamanda doğru tanıtım türleriyle müşteri sağlanmasına imkân veren anlayış sağladı.

Sonuçlar : Yeni analitik teknolojileri kabul ettiğinden beri, PDZA çevrimiçi bilet satışlarında, 2012'de %700'ü aşan ve 2013'de de ilaveten %50 daha büyüdü.

2.1.12.3. *Örnek Vaka 3: Analitik Fırsat Olarak Büyük Müşteri Çeşitliliğinin Davranışlarının Değerlendirilmesi*

Kurumlar yalnızca bir müşteri iletişim listesine ihtiyaç duymazlar, aynı zamanda tüm kullanılabilir müşteri verisinden yararlanabilmeye ihtiyaç duyarlar.

Büyük Veride Gartner'in "3V"si (Volume, Velocity, Variety: Hacim, Hız, Çeşitlilik) ile veri kaynaklarındaki çeşitlilik teknoloji şirketlerinin müşterileri tarafından hem en büyük zorluk hem de en büyük fırsat olarak görülmüştür. Firmaların kendi ERP, CRM ve diğer iç uygulamalarında kullanılan veri, mobil, sosyal ve web kanalları gibi dış kaynaklarla birlikte müşterilere daha iyi bir anlayış verebilir ve daha etkili bir şekilde müşteri ilişkileri yapılandırmaya yardım edebilir.

ABD'deki en büyük çiftlik işletmecisi kooperatiflerinden biri olan Southern States Kooperatifi, doğrudan e-posta için bir öngörü modeli yaklaşımını benimsedi ve sonuçlar

aldı. Karşılık oranı %34'e kadar artarken, katalogların sayısının %63'e kadar azalabildiğini keşfetti.

Problem : Southern States 1200 perakende satış noktası ve 23 eyalet üzerinde 300.000 çiftlik olmak üzere binlerce tarım ürününe sahiptir. Çoklu sistemler, karmaşık ve rekabetçi analitik platformlarda veri ile mücadele etmekte ve bunu yapabilecek birçok personele ihtiyaç duymaktadır.

2012 yılında, Southern States'in doğrudan e-posta kampanyası önceki yıllardaki gibi aynı şekilde yürütüldü. Sonuçlar %3 oranında borç ödemesi ve %39 pazarlama ROI (Return On Investment) yatırım getirisiyle, önceki yıllardaki kampanyalar çizgisindeydi. Sonuçta, Southern States'e bu durumu artan karlar üretiminden daha fazlaya mal oldu ve kooperatif bunun yeterince iyi olmadığını farkındaydı. Southern States müşterilerinin ve onlara nasıl hizmet edeceğinin tam bir resmine ihtiyaç duyduğuna karar verdi.

Çözüm : Kooperatifin katalog ve promosyonlarının kime postalanması gerektiğini belirleyen gelişmiş bir yaklaşımı yoktu, fakat bir analitik tedarikçi olan Alteryx ile ortaklık yoluyla, pazarlama aktivitelerini sürekli bir şekilde optimize edebildi. Alteryx herhangi bir kod yazmadan veya veri verme ya da alma olmadan gelişmiş optimizasyon ve mekânsal analiz sağladı.

Sonuçlar : 2013 yılında, Southern States bu kez öngörücü modelleme yaklaşımını kullanarak fakat aynı doğrudan posta kampanyası yürüttü. Sonuçlar %10 oranında borç ödemesi ve %59 ROI (Return On Investment) yatırım getirisiyle dramatik bir şekilde farklıydı. Yalnız geçen çeyrek boyunca, Southern States %189 ortalama bir pazarlama ROI (Return On Investment) yatırım getirisiyle, sekiz doğrudan posta kampanyası (hepsi modellenmiş) gerçekleştirdi.

Doğrudan posta kampanyalarının başarısını takiben, Southern States daha iyi işletme kararları almasına yardımcı olması için kurum çapında analiz uygulamalarını yaydı.

2.2. BÜYÜK VERİNİN GELİŞİMİ VE VERİ YÖNETİMİ

Büyük Veri mevcut ve geleceğin araştırma sınırlarından birisi oldu. Gartner "Top 10 Strategic Technology Trends For 2013" ve "Top 10 Critical Tech Trends For The Next Five Years" listeleri yayınladı ve Büyük Veri her ikisine de girdi.

Judith Hurwitz, Alan Nugent, Dr. Fern Halper ve Marcia Kaufman'ın ifadelerine göre, Büyük Veri Analizi, "Her veri yönetim dalgası gereklilikten ortaya çıkmıştır ve özgün bir problemin çözümüdür". İlk dalgada veri, plansız "düz tablolar"da depolanmıştı ve şirketler veriden değer çıkartmak için detaylandırılmış program içeren "kaba kuvvet" kullanmak zorundaydı. Bu Öğe-İlişki (Entity-Relationship, ER) modeli tarafından desteklenen veri üzerinde dayatılan yapı olan ilişkisel sisteme yol açmıştır. Bu soyutlama düzeyini arttırarak verinin kullanılabilirliğini arttırdı. İlişkisel veri tabanları için pazar hala günümüze kadar güçlü kalmıştır. Verinin hacmi arttıkça, veri ambarları gelişti; onların hız ve erişilebilirlik programlarına yanıt vermek için veri pazarları geliştirildi. Bunlar elle işlenen veriler için oldukça yeterliydi, fakat yapılandırılmamış ya da yarı-yapılandırılmış veri artarken ikili büyük nesnelere (BLOB) ve nesne veri tabanı yönetim sistemleri hızla çoğaldı. Bu durum adreslenebilir parça setleri için "yapılandırılmamış veri ile başa çıkmak için birleşik bir yaklaşımla" işletmelere ve analistlere sağlanan, depolanmış olan yapılandırılmamış veriye izin verdi (Hurwitz et. al p. 13). Bu gelişmeler dikkate değerdir, çünkü geçmişin aksine, bugün elde edilebilir birçok veri yapılandırılmamış şekildedir.

Büyük Veri analiz tekniklerinin gelişiminin son safhası sosyal medya ve web 2,0, artan veri sanallaştırma ve bulut programlama ile internetin yükselişinde yatmaktadır. Özellikle son gelişme, artan erişim hızı ve maliyet verimliliğinin yanı sıra, bir fiziksel konumda tüm verileri depolamak için ihtiyaç duyulan şirketlerin kurulmasıdır. Aynı zamanda, insan tarafından üretilen veri miktarı yüksek oranlarda artarak geliştikçe, mantıklı bir şekilde veriyi yönetecek şirketlere duyulan ihtiyaç da önem kazanmaktadır. Ayrıca "Modern Büyük Veri" girişimlerinin birçok teknolojik ihtiyacı vardır ve bu ihtiyaçların hepsi önceki veri yönetim dalgaları tarafından yapılan gelişmelere kısmen dayanmaktadır.

Dyche etkin "Büyük Veri" girişimleri için altı ana basamak tanımlar. Bunlar; veri toplama, ham veri setlerinin kullanılabilir hale getirilmesi için bir makine ya da bir bilgisayar tarafından işlenmesi, veri yönetimi (sınıflandırma, tanımlama ve heterojen kaynaklardan gelen verinin açıklanması), toplanabilen verideki oranların ölçülmesi, sağlama yani "işlemin orijinal gerekliliği ile veri kullanılmasındaki sonucu" ve hemen bu veri için depolama çözümlerini adreslemedir.

Ek olarak, Büyük Veri girişimleri için teknik gereksinimler “geniş veri işlem basamaklarıyla başa çıkmak için çok yüksek çıktıyı” desteklerken, “basit sorguları, kısa zamanda yürütme ve veri yakalama için öngörülen kısa sürede teslim etmek” ve “çok yüksek işlem hacimlerin üstesinden gelmek” için gerekli altyapı kabiliyetlerini içermelidir. Bununla beraber “orijinal depolama yerinde işleyebilmeli ve sürdürmelidir” (Oracle 2012). Ayrıca, güvenlik endişeleri ve potansiyel hassas veri toplanan çok miktarda veriye özgüdür, bu yüzden veri için “güvenli çerçeve” sağlanması da fevkalade bir endişe olmalıdır (Shuster 2012).

Daha önce belirtildiği gibi, bugün yeni teknolojiler Büyük Veri girişimleri için ihtiyaç duyulan altyapıyı oluşturmak için önceki ilerlemelerden gelişti. Örneğin, Amerikan çokuluslu yazılım şirketi Oracle, 2012 yılında Büyük Veri gereksinimleri ile başa çıkabilen ürünlerinden bazılarının nasıl olduğunu gösteren bir tanıtım belgesi yayınladı. NoSQL veri tabanları sık sık geniş veri miktarları edinmek ve depolamak için kullanılmaktadır. Hadoop Dağıtılmış Dosya Sistemi özgün depolama kümesinde kalan veriye izin verirken, geniş hacimli verinin işlenmesi ve organizasyonuna müsaade eden bir teknolojidir (çok geniş miktardaki verinin hareketi hem zaman tüketimi hem de zorluk olduğu için, Büyük Veri girişimleri mümkün olduğu kadar ondan kaçınmaya uğraşır). Bu yazılım ve benzer diğer birçok programlar 1970-1990’ların programlama ve yönetim teknolojisindeki daha önce bahsedilen ilerlemelerden inşa edilmiştir.

2.3. BÜYÜK VERİNİN ZORLUKLARI VE FIRSATLARI

Açıksası, Büyük Verinin ilerleyişi birçok yeni fırsat getirebilir, fakat kurumlar ve şirketler için zorluklar da oluşturur. Yukarıda belirtilen teknik gereksinimlere ek olarak, şirketler de ulusal kimlik numaraları dahil (ABD’de sosyal güvenlik numaraları, Türkiye’de kimlik numaraları), tıbbi kayıtlar ve diğer potansiyel sınıflandırma gibi potansiyel “zehirli veriye” dikkat etmesi gerekmektedir ki, bunların kişi bilgileri ortaya çıkarsa (kasıtlı veya başka bir şekilde) bu durum bir şirket için yasal sorun oluşturabilir (Shuster 2012). Computing Research Association (CRA, 2011) tarafından yayınlanan bir makale Büyük Veri yönetiminin bir kaç zorluğuna işaret etti: Teknik mülahazalar, örneğin hataları minimize etme ki, bu bile veri temizleme, düzeltme çabaları ve güç tüketimi zorluklarının arkasından kalır. Zamanında erişim, güç tüketimi ve hata azaltılmasının bu teknik endişelerine ek olarak, dikkate alınması gereken sosyal-politik-

hukuki hususlar da vardır. Bu hususa örnek olarak, etik olarak toplanan Büyük Verinin ve çeşitli kullanıcıların coğrafi konumlarının gizliliğine saygı gösterilebilir.

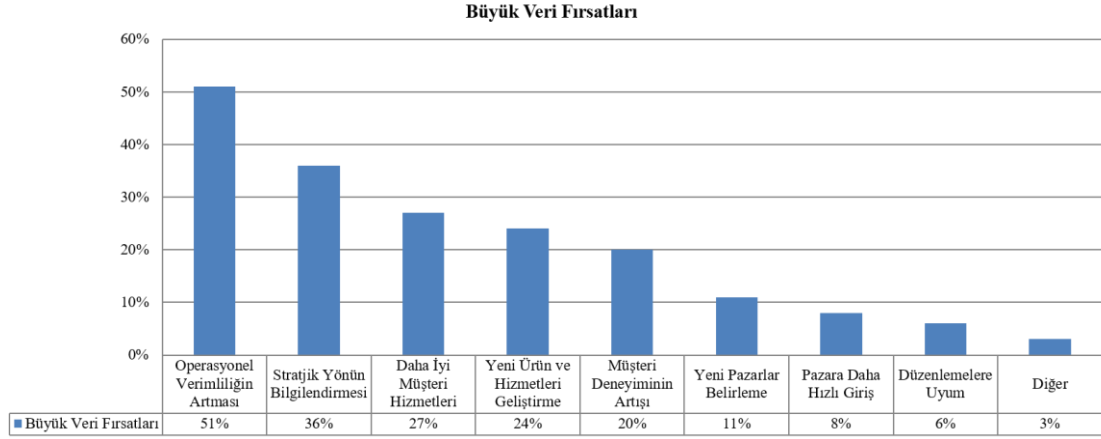
Elbette, bu zorluklarla birlikte şunu da söylemeye gerek yoktur. Büyük Veri şirketler ve işletmeler için burada hepsini adlandırmadığımız birçok potansiyel kullanıma sahiptir. Pazardaki verinin kullanım potansiyelini, standart muhasebeye göre varlıklarının toplamı 6,3 milyar USD olmasına rağmen, pazara giriş yaptığı zaman Facebook'un ilk halka arz değerinin neden 104 milyar USD olduğunu açıklar (Mayer-Schonbeger, 2012, pp. 100 – 120). Wikibon firmasından açık kaynak analisti, 5 yıl içerisinde Büyük Verinin toplam pazar değerinin 50 milyar USD'ye ulaşacağını tahmin etmektedir.

Gelecekte aşırı spekülasyonu önlemek amacıyla, şirketlerin Büyük Veri girişimlerinin gerçekçi, elde edilebilir amaç ve umutlar sağlaması gerektiğinden, Büyük Veri etkileri üzerinde artan çalışmaya daima ihtiyaç vardır.

Aynı zamanda, McKinsey Enstitüsünün raporuna göre, Büyük Verinin etkin kullanımı dönüşen ekonomiler için belli başlı faydalara sahiptir ve yeni bir büyüyen üretim dalgasını doğuruyor. Büyük Verinin ötesinde değerli bilgi avantajları sağlama bugünün işletmeleri için ana rekabetçi unsur olacak ve Büyük Verinin kabiliyetlerine sahip ilgi çekici işler yapabilen yeni rekabetçi firmalar oluşturacaktır. Araştırmacılar, politikacılar ve karar verenler alanlarındaki büyüyen gelecek dalgayı ortaya çıkartmak için Büyük Veriyi kullanma potansiyelini fark etmek zorundalar. Aşağıdaki şekilde görüldüğü gibi, operasyonel verimliliği artırma, stratejik değerlendirme, daha iyi müşteri hizmetleri gelişimi, yeni ürünler ve hizmetler tanımlama ve geliştirme, yeni müşteriler ve pazarlar bulma vb. iş bölümlerinde Büyük Veri kullanımıyla elde edilebilen birçok avantajlar vardır. Dikey eksen kaç tane işletmenin Büyük Verinin özgün amaçlarıyla ilgili onlara yardımcı olabileceğini düşündüğünü yüzde olarak gösterir.

Liberal tahminlere göre, Büyük Veri ABD'deki sağlık bakımı için yıllık 300 milyar USD, AB kamu yönetimi için 250 milyar EUR potansiyel değer üretebilir. Dünya çapındaki kişisel lokasyon verisi kullanılarak yıllık 600 milyar USD potansiyel müşteri fazlası olacak ve %60 oranında potansiyel bir artış sağlanacaktır. Sadece ABD'de Büyük Veri 140.000 ila 190.000 arasında derin analitik yetenek pozisyonları ve 1,5 milyon veri

meraklısı yöneticiler üretmektedir. Kuşkusuz, eğer doğru bir şekilde araştırılıyorsa Büyük Veri genel olarak çekici ve karlı olmaktadır.



Şekil 2.13: Büyük veri fırsatları: 560 işletmenin %50'den fazlası büyük verinin operasyonel verimliliği arttırmada faydalı olabileceğini düşünür vb.

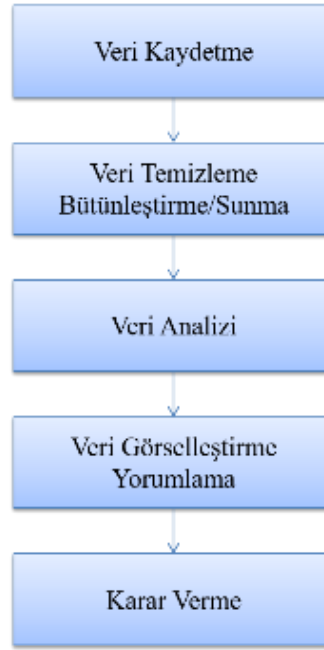
(Kaynak: McKinsey Global Institute Report,
<http://www.sciencedirect.com/science/article/pii/S0020025514000346>, 2011)

Fırsatların ardından daima zorluklar gelmektedir. Bir yandan, Büyük Veri çok cazip fırsatlar getirir. Diğer yandan da Büyük Veri problemlerini ele aldığımızda birçok zorlukla yüz yüze kalırız. Bu zorluklar veri toplama, depolama, arama, paylaşma, analiz ve görüntüleme üzerine yayılır. Eğer bu zorlukları aşamazsak, özellikle bilgi bizim kullanım kabiliyetimizi aştığı zaman, Büyük Veri bir altın değerinde olacaktır fakat onu keşfetme imkânlarına sahip olamayız. Birkaç on yıldır bilgisayar mimarisinde var olan bir engel CPU'nun güçlü fakat I/O'nun zayıf olmasıdır. Bu sistem dengesizliği Büyük Veriden faydalanma gelişimini hala kısıtlamaktadır.

CPU performansı yaklaşık her 18 ayda bir ikiye katlanıyor, aynı zamanda disk sürücülerinin performansı da aynı oranda katlanıyor, fakat disklerin rotasyonel hızları son on yılı aşkın sürede çok daha az gelişti. Bu dengesizliğin sonucu ardışık I/O hızları çok yavaş bir şekilde artarken, rastgele I/O hızları ortalama şekilde gelişti. Dahası, bilgi üstel oranlarda artıyor, fakat bilgi işleme yöntemlerinin gelişimi görece daha yavaş oluyor. Birçok Büyük Veri uygulamasında, özellikle de gerçek zamanlı analizler için, teknikler

ve teknolojiler ideal bir şekilde problemleri çözemiyor. Şu ana kadar kısmen ifade edilen genel kanı şudur; altın cevherini hakkıyla işleyebileceğimiz tamamen uygun araçlara sahip değiliz.

Tipik bir genel analiz süreci aşağıdaki şekilde gösterilmiştir.



Şekil 2.14: Bilgi keşif süreci.

(Kaynak: <http://www.sciencedirect.com/science/article/pii/S0020025514000346>)

Büyük Veri analizindeki diğer zorluklar veri tutarsızlığı ve eksikliği, ölçeklenebilirlik, güncellik ve veri güvenliği olarak değerlendirilebilir. Veri analizi için ilk basamak verinin iyi yapılandırılmış olmasıdır. Fakat veri setlerindeki dikkate değer çeşitlilik hedefe ulaşmak için bize büyük engeller çıkartmaktadır. Veri setleri genelde çok büyüktür, bazen birkaç gigabaytlara hatta daha fazlasına ulaşmaktadır, heterojen kaynaklardan gelmektedir, gerçek dünyanın veri tabanları süreksiz, tamamlanmamış ve gürültülü olmaktadır. Bu ve bu gibi nedenlerden dolayı veri temizleme ve veri dönüşümünü kapsayan veri işleme tekniklerine gürültü kaldırma ve tutarsızlıkları düzeltmek için

başvurulabilir. Veri tabanı uygulamalarından geldiğinde, farklı engeller her bir alt işlemi arttırmaktadır.

2.4. BÜYÜK VERİYE UYUM STRATEJİLERİ

Tim Suther, Bill Burkart ve Jie Cheng (2013) şu şekilde ifade etmiştir; veri ve “Büyük Veri” değerli olmasına rağmen, durum aşırı basitleştirilmiş olmamalıdır. Şirketler Büyük Veri girişimleri, içerik gizliliği, yasal ve yazılım endişeleri ile gelen ölçülebilir zorluklarla Büyük Veriden gelecek potansiyel kazançları dengelemek için birkaç “somut adım” atmaya ihtiyaç duyar.

Suther ve diğerleri “veri çok büyük değer kilidini açar” şeklinde açıklar, fakat özellikle verinin türü hayati şekilde önemli bir değişkendir. Örneğin, en önemli soru müşterilerin ne aldığı harcama alışkanlıklarıyla ilgili midir? Onlar ne istediklerini söylerler? Ne araştırırlar? Müşterilerin sosyal medya aktivitelerindeki veri? Ya da internette gezerken bıraktıkları dijital kısıntıları? Bu sorulardan birisinin sihirli “Veri Mermisi” olmasından ziyade, aslında sonuçta önemli olan bu farklı faktörlerin bir kombinasyonudur. Bu zor olsa bile “liderler müşterileri hakkında tüm uygun perspektiflerden rafine anlayışta çok boyutlu olarak düşünmelidir”. Bunun nedeni, müşteri davranışını tam olarak açıklayan bir veri yoktur, çünkü her veri noktası veya değişkeni, kendi konumunda değerli iken, ayrıca eksiklikleri de vardır (örneğin, bir müşteri kendi elleriyle deneyene kadar, yeni bir elektronik alet istediğini bilmeyebilir; müşterilerin cinsiyet bilgileri önemlidir fakat bazen yanlışdır vb.)

Bundan dolayı, Suther ve diğerleri şu şekilde ifade eder; bir şirket “kendi müşteri portföyünü büyütür” Büyük Veri girişimleriyle gerçek başarıyı sağlayabilir. Bu “çok boyutlu anlayış” başarılı bir girişimin “dayanağını oluştururken”, Suther ve diğerlerine göre birçok firmanın gözden kaçırdığı anahtar nokta şudur: Kendi var olan müşteri profiline yetersiz yatırım yaparlar.

İfade edildiği gibi, şirketler “şu anki müşteri ilişkilerinin değerini ve yatırım kararları yürütmek için onu kullanmanın değerini” hesaplamalıdır. Sonra, “tersyüz” çalışarak, amaç şu andaki var olan kendi müşterilerinin en karlı alt kümesine benzeyen potansiyel müşterileri eşleştirmek olacaktır. Bu şekilde, müşteri portföyünün değeri katlanarak artacaktır.

Son olarak, Suther ve diğeri şunu söylemiştir; özellikle veri sağlayabilmesine benzer, büyük potansiyel güç göz önüne alındığı zaman Büyük Veri ile uğraşırken ahlaki davranış hayatidir. “Bu ticari zorunlulukları aşan veri müşteriler için kullanılmalıdır, onlara başka bir şey yapmak için kullanılmamalıdır. Bu hızlı puanlara değil, güven tabanlı ilişkiler kurarak uzun dönemli optimizasyon anlamına gelir” şeklinde ifade etmişlerdir. Bu prensipleri takip etme karşılığında, Suther ve diğeri; şirketler üç önemli karşılık umabilir: Pazarlama performansında ölçülebilir gelişmeler, müşteri portföylerinin artan değeri ve daha anlaşılabilir ve korunabilir fiyat. Büyük Veri yanlış tanıtımı düzeltmek için özel potansiyele sahiptir.

ComScore tarafından yapılan bir çalışmaya göre, çevrimiçi reklam %80 ini aşan oranda istenilen hedefe ulaşmada başarısız olmaktadır. Bu durum kısmen, örneğin yaş ve cinsiyet gibi aşırı geniş bir genelleme alanına dayanmak yerine, müşterilerin bireysel alım uygulamaları ve profillerindeki nüans özelliklerinin başarısızlığından dolayı ortaya çıkabilir (Comscore 2009). Eğer şirketler Büyük Veri girişimlerini “daha akıllı hedefleme ve daha iyi yönetim” içerisine tutturmak için kullanırlarsa, Suther ve diğeri göre, pazarlama yatırım geri dönüşlerinin %15-30, müşteri brüt kar marjlarının %10-15 arttığını ve %7 iyileştirilmiş fiyatlandırma görebilirler. Çünkü, veri miktarı büyürken genelde, her yeni veri parçasının marjinal değeri azalır. Daha fazla araştırma için kullanılan bilgisayarlar ve analitik işlemler etkin bir şekilde büyüyen veri miktarını faydalı hale getirmek ve başa çıkmak için tek yöntemdir.

Daha sonra görüleceği gibi, örüntü tanıma Büyük Veri yönetimi programlarının hayati bir parçasıdır. Bununla beraber yöneticiler ve firmalar Büyük Veri yönetimini yalnızca bilgisayarlar ve uygulama programlarına bırakmazlar, halen insan analizi ve girişine ihtiyaç duydukları yerler vardır.

2.5. BÜYÜK VERİNİN İNSAN UNSURU

Daqing Zhao'nun (Dir Advanced Analytics at Macys.com) belirttiği gibi, neticede bilgisayarlar analistlerin komut ve talimatlarına uydukları için, bir bilgisayar sonuçta “sadece onu kullanan analist kadar iyidir”.

Büyük Veri kaynakları, özellikle de yapılandırılmamış veriler, tamamen büyük ölçüsünden dolayı değil, ayrıca belgesi ve sınıflandırılmamış bilgiden dolayı sıklıkla

bunaltıcı olabilir. Eđer “ham” şekilde gelirse, henüz bir etiketlemesi yoktur. Kalitesini saęlamak için veri, doęruluk ve “arılık” için kontrol edilmiş olması gereklidir. Yani, Zhao “veri sadece onları analiz etme için kullanılan çabanın miktarı kadar temizdir” der. O uzun ve zahmetli fakat sonuçta gerekli olan veri doęrulumu ve kalite hata ayıklama sürecini, gelir saęlama sürecine benzetir. İnsan unsuru da cevaplanması gereken sorular için hazır olmalıdır. Model baęımlı deęişken ve baęımsız belirleyicileri tarafından yapılandırılmış olacaktır. Bunların tümü ilgili kimseleri ile analistler tarafından bilgisayarların içerisine yerleştirilir.

Gelişen müşteri servisleri açısından “servis kalitesinin” nasıl ölçüleceęi bir örnektir. Ayrıca, bir web sitesine yapılan tıklamalar, ziyaret frekansı, işlem miktarı veya kullanım süresi de bir ölçü olabilir. Sonra gün, konum, yaş ve cinsiyet gibi bir takım deęişkenler belirleyicilerdir. Analistler (ve dolayısıyla insan kaynakları) Büyük Veri girişimlerinin hala önemli bir parçasıdır, çünkü bilgisayar algoritmaları eđer modeldeki bir takım deęişken öngörü yapabiliyor ise karar verebilir, fakat bir bilgisayar eđer kritik öngörüğü gözden kaçıyor ise kendisi üzerinde fark edemez. Dięer taraftan, eđer veri analiz modeli kullanımında herhangi bir problem varsa, bilgisayarın (baęımsız olarak) bilmesinin bir yolu yoktur. Sonuçta, Zhao der ki, “bilgisayar modelleri insanlar tarafından tasarlanmak ve yönetilmek zorundadır”.

Zhao ayrıca şuna da işaret eder ki; modeller sadece uzun bir periyod süresini aşan tekrarlanabilir kalıplar için yapılandırılabilir. Gittikçe daha fazla model inşa edilirken, bazıları çok kısa “yaşam süresine” sahiptir ve sadece bir kaç hafta içerisinde performansında gerileme olur. Bunun nedeni, öngörü modelleri yapılandırılmak için uzun bir süre alır ve Büyük Veri ile birlikte, öngörücülerin birçok sayısı ve ölçüsü olabilir. Çünkü potansiyel olarak “binlerce deęişken” olabilir. Zhao yine şunu ifade etmektedir; “Kullanıma hazır modelleme ve örnek dışı test etme benzeri teknikleri kullanan, bir nebze otomatik olarak modeller yapılandırmaya yardımcı olan modelleme metodolojisine güvenmemiz gerekir. Saęlanan veri yüksek kalitededir (çünkü daha yüksek kalite veri daha yüksek modeller saęlar) ve depolama modeli de dikkate alınması gereken önemli faktörlerdendir.

2.6. VERİ TİPLERİ VE VERİNİN GERÇEK DEĞERİ

Veri girişimleri şu anda iş dünyasında oldukça yükselen bir değer olduğundan, Omer Trajman (San Francisco Bay Area) verinin “somut ve gerçek değerinin” ne olduğunu araştırmaya çalıştı. İlk olarak şunu ifade eder: Veri patentler ya da fikri hakları gibi aynı şekilde soyut varlıklardır. Fakat çoğu firma veri depolaması ile karşılaştırıldığında onların muhasebe yöntemleri tarafından kanıtlandığı gibi, bir değer olması için, onu basit bir şekilde değerlendirir (diğer bir deyişle, veri değeri onu depolama maliyeti ile iç içe geçmiştir). Trajman “bu düşünce şekli verinin gerçek değerini gizler” der, çünkü “depolama herhangi veri içersin ya da içermesin sabit maliyete sahiptir”, verinin kendisi “ek maliyete sahip veri içerikleri olan gerçek değeri ve depolamayı” içerir (2013 p.73).

Gerçek verinin değerini daha iyi anlamak için, Trajman ilk önce veriyi sınıflandırmayı araştırmaktadır. Bir alışverişin değerini kaydetmek için kullanılan veri (örneğin satın alma) işlem verisidir. Bu olay etrafında üretilen olgu “arızı veri” olarak adlandırılmıştır. Veri sıklıkla yapılandırılmış ve yapılandırılmamış kategoriler içerisinde sınıflandırılmasına rağmen, tüm veri aslında bir yapıya sahiptir.

Hurwitz ve diğerleri (2013, p. 30) “yapılandırılmamış” veya “yapılanmamış” adının gerçeklikten geldiğini açıklığa kavuşturmuştur, o belirli bir formatı takip etmez. Makine üretimi yapılanmamış veri örneği olarak radar ya da sonar (deniz radarı) verisi, gözetleme veya trafik video ve fotoğrafları, bilimsel veri araçlar ve uydu görüntüleri tarafından toplanmaktadır. İnsan üretimi yapılanmamış veri örnekleri olarak Youtube ve Flickr benzeri web sitelerinden içerikler, metin mesajları ve konum bilgisi, şirket içi e-posta ve diğer haberleşmeler, sosyal medya verileri, mobil veriyi alıntı yaparlar. Aksine, “tanımlanmış bir boy ve şekle” sahip yapılandırılmış veri dünyadaki toplam verinin yaklaşık %20 sine tekabül etmektedir. Kaufman ve diğerleri, algılayıcı veriyi kapsar (örneğin, RFID -radyo frekansı ile tanımlama-, GPS -küresel konumlama sistemi-, tıbbi aygıtlar ve akıllı sayaç okumalar), web günlükleri verisi, satış noktası verisi (Trajman’ın işlemsel olarak adlandırdığı) ve makine üretimi bu tip verinin örnekleri olarak finansal veri, giriş verisi, insan üretimi yapılandırılmış veri örnekleri olarak web sitelerinden tıklama akış verisi ve oyunla ilişkili veri (oyuncuların yaptığı hareketler vb.).

Toplam verinin sadece %20'sinin yapılandırılmış veri olması dikkat edilmesi için önemlidir, göz ardı edilmemelidir ve pazara giren yapılandırılmamış veri miktarının büyümesi kadar önemli olabilir.

Trajman'ın analizine geri dönersek, rastlantısal veri depolanırken, nadiren, bir satın almayla ilişkili isim, adres ve kredi kartı şeklindeki kişisel profil verilerin haricindekiler düzenlenmiş ve analiz edilmiş işlemsel veri şeklidir. Bununla birlikte, genelde gözden kaçan ikinci tip rastlantısal veri “daha dinamiktir” ve sadece müşterilerin nihai satın almalarını ve satın alma yapmadan önceki tıklanan ve incelenen ürünleri değil, dijital servislerin örneklerini içermektedir (Trajman 2013, p. 70). Yıllardır çok başarılı bir internet perakendecisi olan Amazon.com sadece müşterilerin son satın almalarını ve almadan baktıkları ürünleri değil, web sitesine gelen ziyaretçiler ve müşteriler için “kişiselleştirilmiş” önerileri derleyen rastlantısal verinin bu “veriyi sömüren” modelini kullandı.

Trajman, rakiplerinin ve sanayi denkleminin yapmadığı şekilde “veri kullanarak bir rekabet avantajı elde eden şirketlerin diferansiyel değerini” dikkate alarak değer olan “verinin bir makroskobik resmini alma” girişiminde bulunmaktadır. İlk olarak, “verinin değerini” hesaplar: Verinin nasıl kullanıldığı ve yaşam döngüsündeki yeri üzerinde temellenen farklı değerlere sahip olabilirken, veri yazıldığı anda kesin bir bedele sahiptir, silindiği ve tahrip olduğu zamana kadar bile, depolama gibi bir maliyet doğurur. “Verinin içsel değerinin tanımlanmasının önemini anlama için” Trajman der ki, “bizim sadece yakalama ve depolamanın maliyetini dikkate almamız gereklidir” : 2012’de veri kaybının yaklaşık maliyeti, kayıp ya da çalıntı her müşteri kaydı için 200 USD’yi aşmıştı. Elbette eğer adres ya da kredi kartı bilgisi gibi müşterilerin hassas bilgileri kırılarak (hack) veya bir şekilde birileri tarafından çalınmış olursa, veri sızıntısı başlıca bir güvenlik riskidir, bu da ciddi tanıtım, yasal ve hatta şirket ya da kurumlar için hisse senedi problemlerine neden olabilir.

Sony Playstation ağları üzerinde bununla mücadele eden bir şirket örneğidir. PlayStation Network erişiminin kesilmesi olayında, gerçekleştirilen hacker saldırısı sebebiyle Sony'nin PlayStation Network ve Qriocity servislerine erişimin kesilmesi ve çevrimiçi hizmetlerden faydalanan yaklaşık 77 milyon PlayStation 3 ve PlayStation Portable kullanıcısının hesap bilgilerinin çalınması hadisesi yaşanmıştır. 17 ile 19 Nisan 2011

tarihleri arasında gerçekleştirilen saldırının ardından Sony, 20 Nisan'da PlayStation Network erişimini kesmek zorunda kaldı. 4 Mayıs'da ise şirket tarafından yapılan açıklamayla yaklaşık 77 milyon kullanıcının kişisel tanımlayıcı bilgilerinin çalındığı kabul edildi. Yaşanan bu olay, “en büyük veri hırsızlığı olaylarından biri” olarak tarihe geçti. 26 Nisan'da Sony, hizmetlerin bir hafta içinde eski haline döneceğini belirtse de, 14 Mayıs'ta yayınlanan PlayStation 3 güncellemesiyle birlikte kullanıcıların şifrelerini yenilenmesi istendi. Bu sırada çevrimiçi hizmetlerin açılması için yapılan çalışmalar halen sürdürülmekteydi. 23 gün süren kesintinin ardından ise hizmetler yeniden kullanıma girdi. Servislerin açılmasının ardından Sony, saldırıdan etkilenen kullanıcılar için "tekrar hoş geldiniz" adı altında çeşitli hediyeler yayınladı. Oyunlar da dâhil olmak üzere birtakım içeriklerin yer aldığı bu hediyeler, kullanıcıların belli bir süreliğine ücretsiz olarak erişimine sunuldu (https://tr.wikipedia.org/wiki/PlayStation_Network_erişiminin_kesilmesi, 2015).

Bundan dolayı, verinin potansiyel değeri özünde onun maliyetine de bağlıdır; daha değerli bir veri bir şirket içindir, dahası eğer bu veri kaybolur ya da tahrip olur ise o değerini kaybedebilir.

Bir örnek olarak, Trajman şu senaryoyu sunar: Acxiom Şirketi 2012 Mayıs ayında yıllık 1,13 milyar USD gelir oluşturdu. Bunun nedeni, dünyada diğer şirketlerde olmayan veriye erişime sahip olmasıydı. Bundan dolayı, Trajman şunu sorar, “Acxiom’un verisine sahip olmak için bir kurum ne ödeyebilir? Bu özel verilerin hepsine sahip olan bir organizasyonun” değeri ne olabilir? Masraflardan sonra, Acxiom’un 2012’de net geliri 85 milyon USD idi: Trajman şunu sormaktadır: Bunun ne kadarı, bu verinin özel olmasından dolayıydı?

Bu tekellilik faktörüne ek olarak, Trajman şöyle der; rastlantısal veri işlemsel veriden daha hızlı oranda büyümesine rağmen, rastlantısal veri üzerine parasal veri koymak daha zor bir işlemdir ve bundan dolayı bilim ve medyanın ilgisini daha fazla alıyor. Fakat Google ve Facebook gibi büyük internet şirketleri, kendi geri dönüşleri için hemen hemen net pozitif bir sonuç sağlayan depolanan rastlantısal veriyi hesaplamıştır ve kabul etmiştir. Trajmana göre, bu şundan dolayı olabilir: Bir satın almayı veya içerideki başvuru vergilerini telafi etmek için kullanıldığı zaman işlemsel veri yolunda daha çok değere sahiptir, rastlantısal veri bir rekabet avantajı kazanmak ve bir işletmenin operasyonel

etkinliğini artırmak için müşterileri korumak ve çekmek için kullanılabilir. Rastlantısal verinin bu yönü birçok Büyük Veri girişimlerinin merkezindedir.

Paul Kent, Radhika Kulkarni ve Udo Sglavo'nun Büyük Verinin değerini araştırmasında görünen şudur ki, büyük firmaların sıklıkla başa çıkmak zorunda olduğu durum, farklı ürün grupları ve çeşitli müşteri portföy grupları ve zıt kutuplarla başa çıkmak etkin bir biçimde birleştirmeye yardımcı olabilir. Büyük Veri bir işletme ekibinin üyeleri tarafından önerilen yeni programlar veya çeşitli karlılık senaryolarını hızlı ve etkin bir şekilde belirleyen algoritmaları çalıştırmak için kullanılan toplanmış veriyi kullanan yüksek performanslı analizleri (HPA: High-Performance Analytics) çalıştırmak için kullanılabilir. Seigel gibi, onlar da şuna inanırlar; Büyük Veri “kesin olmayan, göreceli bir terimdir”, ve ayrıca şunu da ima ederler; veri miktarının büyümesiyle başa çıkacak bir organizasyonun kapasitesi IT departmanının konfor bölgesini geçti. Cukier ve Schonberger'in işaret ettiği gibi, ayrıca belirtmektedirler ki, birçok şirket Büyük Veri miktarlarına sahip iken, verinin potansiyel değerine bakan şirketlerden bazıları bu veriden etkin bir şekilde nasıl faydalanacağını biliyor.

“Büyük Veri” kullanan yüksek performans analizlerinin değeri, sonuçta ölçümlerin taşıyabilme hızına dayanır ve bundan dolayı bu hız karar vermede yöneticilere yardımcı olabilir. Bu nedenle, üst yönetim üyeleri arasında karar periyodları ya da çıkmazların üstesinden gelmede özel kullanıma sahiptir. Kent ve diğerlerine göre, hatta bu bir takım arasında yaratıcılığı teşvik edebilir, çünkü “eğer bir dakika içerisinde bir yanıt almak için yüksek performanslı analizlerini kullanabilirsen, hesaplamalar vasıtasıyla daha çok boş zamanla, birçok soru sorar, birçok alternatif araştırır, birçok senaryo oluşturur ve daha iyi sonuçların peşinden koşarsın”. Kent ve arkadaşlarına göre, yüksek performanslı analizler birçok avantaja sahip “dağıtımli hesaplamaya” yakından bağlıdır. Bunlar esneklik ve maliyet avantajları (çoklu görev ve çoklu izleri koruma, farklı kullanıcı grupları, müşteriler ve uygulamaların daha kolay olması) içerir ve aynı zamanda hala firmalar hedef müşterilerine yardım ederken, bunlar altyapı maliyetlerini önemli miktarda düşürebilir.

Kent ve arkadaşlarına göre, yardıma ek olarak daha fazla özelleştirilmiş müşteri pazarlaması ve reklam kampanyaları oluşturan Büyük Veri, firmaların var olan müşteri portföyünü tam anlamıyla anlamasına da yardımcı olabilir. Örneğin, veri tabanındaki

analizler olası vadesi geçmiş bir müşteri hesabının en sonunda nasıl ödeyeceğini kararlaştırmasına yardımcı olabilir ve birileri hiç ödemeyebilir ve büyük olasılıkla suçludur. Yine, zaman burada bir şirket rekabetin üstesinden gelme konusunda farklılık yapabileceği kritik bir elemandır. Kent, Kulkarani ve Sglavo şunu belirtir: Bellekteki analizler yüksek performanslı analizlerin zirvesidir. Daha hızlı kararlar vermek, uygun tüm verilerden daha kesin cevaplar kazanmak, daha fazla güven kurmak, ölçeklenebilir güvenilir analiz altyapısı, yöneticilerin daha yaratıcı ve etkin kararlar için beyin fırtınası yapabilmelerine olanak sağlayan değerli boş zamanı şirketlere verir. Sık olarak alıntı yapılan “öngörülü analizlere” ek olarak, ayrıca hatırlanması önemli olan şuna da işaret ederler, Büyük Veri girişimleri aynı zamanda “tanımlayıcı analizler” ve “kuralcı analizler” oluşturmaya yardımcı olabilir.

2.7. YENİ MODELLER VE BÜYÜK VERİ

Daniel Conway ve Diego Klabjan “veri bir şirket için etkili bir biçimde bilgiye dönüştürüldüğü zaman, yeniliği teşvik üzerinde pozitif etkilere sahip olacaktır” demektedir. Veri “uykuda ve pasif” iken, uygun bir şekilde analiz edilip, kullanıldığı zaman, “bilgi”nin bir şekli olarak değerli bir emtia olur. Özellikle kullanımdaki (göreceli) ölçü ve maliyetinden dolayı “Büyük Veri” değerinin maksimum olması için çeşitli bileşenler içerisine uygun bir biçimde entegre edilmiş olması gerekir. Ayrıca şuna da işaret etmişlerdir: Varlıkların gelecekteki kaldıraç etkilerinin nasıl olacağı bilinmediği zaman verinin değerini ölçmek zordur. Bunun için, Büyük Veriyi anlamak için birçok farklı bileşeni açıklamaktadırlar (2013 : pp. 132).

İlk olarak, “yeni bir doğal kaynak şekli” olarak Büyük Veriyi açıklamayı IBM’den alıntı yaparlar. Doğal bir kaynakta olduğu gibi, tam anlamıyla kullanılmadan önce, çıkarılmış, birleştirilmiş ve çeşitli derecelerde işlenmiş olmaya ihtiyaç duymaktadır. Ayrıca, dünyadaki geleneksel kaynakların seçilip çıkarmasıyla, ayrıştırılarak ve kullanılarak, veri zamanla daha etkin olur (geçen on yılda önemli derece büyüyen bir konsept olarak “Büyük Verinin” olma nedeninin bir parçasıdır). Bununla birlikte, doğal kaynaklardan farklı, kısırlığı ima eden bir terim, Büyük Veri miktarı sadece büyümektedir ve kurumların zorluğu da aslında bu veriyi etkili bir biçimde nasıl kullanmalarınıdır. Büyük Veriyi incelemenin bir başka yolu dijital envanter “digital inventory (DI)” olarak ona bakmaktır. DI etrafında depolama ve taşımanın maliyeti fiziksel envanterden (physical

inventory: PI) daha az iken, Büyük DI halen altyapı ve bant genişliğindeki darboğazlara maruz kalır, bu yüzden şirketler aynı zamanda “ekonomik gerçeklerin” bu engelinden haberdar olmayı gerektiren Büyük Veri girişimlerini üstlenmektedirler (2013: p. 133).

Nihayet, Kent ve meslektaşları gibi, Conway ve Klabjan Büyük Verinin yeniliği teşvik etmede oynayabildiği rolü araştırmaktadır. Yenilik takım işbirliği yoluyla olduğu kadar bireysel kaynaklarla birlikte her ikisinin üzerinde oluşabilir ve bir karar süreci “aktiviteler seti, icra emir vb. seçilecek yolları belirleyen karar kuralları”dır. Önemli biçimde, Büyük Veri her elemanda ve bu sürecin basamaklarında bir role sahip olabilir.

Ian Thomas ayrıca şunu belirtir; çoklu karar alma süreçleri içerisinde Büyük Veri bir organizasyon için, değerini ve etkisini maksimize etmek için en iyi yoldur. Büyük Veri ile ilgili yapılan teknoloji seçimleri birçok şirket ve organizasyon için zor olabilir der, çünkü “çözümler ile pazardaki çok çeşitli satıcıları birbirleriyle karşılaştırmak kolay değildir” (Thomas 2013: p. 155). Şuna da dikkat çekmiştir; sadece teknoloji yönüne odaklanmaktan başka, organizasyonlar için hatırlanması gereken önemli nokta şudur ki, Büyük Veri sistemleri ve girişimlerinin gerçek amacı işletmenin nasıl çalışacağı ile ilgili “birçok büyük ve küçük karar verenler için, her gün veriye güvenmek için kurumlarda insanlara olanak vermektir”. Bu bütünsel yaklaşım nihayetinde bir şirket için daha faydalı olabilir.

Thomas şu şekilde devam etmektedir, bir Büyük Veri projesi değerlendirildiği zaman, bu verinin kullanıcıları karşılaşılan üç şarta sahip olmaya ihtiyaç duymaktadır: Anlamlı verileri yorumlayabilmek için doğru beceri ve yardıma sahip olması gerekir, verinin güvenilir olduğuna inanmalıdır ve gerekli detayların doğru seviyelerine kolay erişebilmesi gerekmektedir. Bu üç şartın hepsi karşılanmalıdır, eğer sadece ikisi karşılanır ise, veri israf edilmiş olacaktır. Bundan dolayı, şirketler başarılı bir Büyük Veri projesinin üç içeriği olan “İnsanlar”, “Süreç” (Thomas tarafından belirlenen “sıradan bir izleme ön çalışması, veri kalitesi sorunlarını raporlama ve standartları ve tanımları savunma ile ilgilenme” (Thomas p. 156), ve “Teknolojiden” emin olmalıdır. Aslında Büyük Veri girişimlerini doğru bir şekilde uygulamak ve yorumlamak için çok yönlü ve iyi eğitilmiş bir takım gerekir. Thomas’a göre, tam olarak bölünmez bir bütünlük için, Mühendislik ve İşletmenin her ikisini de içeren personel potansiyeli tam artan bir firmanın sahip olduğu veri sağlamdır.

Nihayetinde, Hurwitz, Nugent, Halper ve Kaufman teorik düzeylerin ötesinde gerçek bir işte Büyük Veri analizlerinin pratik bir şekilde nasıl uygulanacağını bir kaç örnekle aktarmıştır. İlk olarak, net ve açık analizler için sürecin başında şirketler ihtiyaçlarını ve amaçlarını değerlendirmeleri gerekmektedir. Örüntü tanıma birçok veri analiz programının anahtar unsurudur.

Bir örnek olarak, araba kiralama şirketi örneğini aktardılar. Şirket aynı yüksek ek maliyetlere sahip olmayan gelişen şirketlerden güçlü bir rekabet deneyimliyordu (Hurwitz et. al.,: 260). Şirket onların müşteri hizmetleri ve cevap verme yeteneklerindeki zayıf noktayı tespit etti. Online bir anket oluşturdu, şikâyetler ve önerilerle ilgili e-posta ile müşterilerini cesaretlendirdi. Geleneksel olarak bununla birlikte, bir yönetici bu e-postaları veya yanıtları ayrı ayrı okumaktaydı ve elbette ek olarak bireysel kişinin takdirine bağlı e-postaların sınıflandırması ve analizi yavaş adımlarla oluyordu. Bir metin analiz programı, hâlbuki çok daha hızlı şekilde bakabilirdi, ayrıca daha hızlıca dokunan bu coğrafi veri bir yer veya bölümdeki problemleri tanımlamaktadır. Diğer örnek, Hurwitz ve diğerleri “etkileşimin erken aşamalarında” dolandırıcılığı daha etkin bir biçimde ortadan kaldırmak için, birçok veri kaynağından gelen çaprazlama potansiyel ile dolandırıcılık iddialarına yardım edebilen bir şirket olmasını sağladı. Bundan dolayı, bu verinin yine doğru zamanda analiz edilen veri “düzeltici faaliyet için çok geç olmadan önce, müşteri memnuniyetsizliği veya potansiyel ürün kusurunu tanımlamaya” nasıl yardımcı olabileceğinin bir örneğidir (Hurwitz et. Al, p. 256).

Sonuç olarak, bu bölüm göstermiştir ki, her derde deva ilaç veya her derde deva olmak yerine, “Büyük Veri” (kesin olmayan ve hatta ilişkisel bir terim olarak bile) zorlukları olmasına rağmen, bir şirket için birçok potansiyel kullanımlara ve faydalara sahiptir.

Kurumlar bir Büyük Veri projesi açmadan önce uygun insan ve teknolojik kaynaklara sahip olmayı sağlama alma ihtiyacı duyarlar ve daha sonra mümkün olduğu kadar şirketin bir çok alanının içerisine geniş veri kaynaklarını entegre etmek için at gözlüğünü çıkarmaya çaba harcarlar.

3. MALZEME VE YÖNTEM

Bu bölümün amacı araştırmanın nasıl yapıldığını açıklamaktır. Öncelikle araştırma paradigması açıklanmış, veri raporlama ve analizine bakış açısı, araştırmanın kısıtları ve araştırmadaki etik hususlara dikkat çekilmiştir.

3.1.YÖNTEM

3.1.1. Giriş

Bu araştırmaya başlamadan önce araştırma yöntemleri ile ilgili kısa bilgi vermekte fayda vardır:

Metodoloji veya yöntem, bilimsel araştırmaların mantıksal çerçeve içinde düzenlenmesi, yürütülmesi ve sonuca bağlanmasını yansıtan strateji veya genel yaklaşımdır. Kısaca metodoloji, bilimsel araştırma mantığıdır.

Metodolojinin genel prensipleri bütün bilimlerde aynıdır. Her bilim, metodolojisinde deney, gözlem, tümevarım, tümdengelim ve tümevarım-tümdengelim (bileşik metodu) benimsemiştir.

Bilimsel araştırmalarda yöntem konusu, araştırmanın bilimsel bir temele oturtulması açısından önemlidir. Çeşitli bilim alanlarında birbirinden farklı yöntemler kullanılarak bilimsel sonuçlar elde edilebilir. Önemli olan, araştırmada kullanılacak yöntemin araştırılan konuya uygun olması, o yöntemle ilgili konunun araştırılabilir olmasıdır. Bu açıdan bakılınca bilimsel araştırmalarda kullanılan yöntemlerin çeşitli şekillerde sınıflandırıldıkları görülür.

Temel olarak yöntemleri şu şekilde sınıflandırabiliriz:

3.1.1.1. *Analitik Araştırmaya Karşın Betimsel Araştırma*

Betimsel araştırma yöntemleri, ilgilenilen ve araştırılmak istenen problemin mevcut var olan durumunu ortaya koymaya yöneliktir. Betimsel araştırma, çalışılan konunun mevcut durumuna ilişkin hipotezler test etmek için veya sorulara cevap bulmak için veriler toplamayı gerektirir. Betimsel araştırma çalışmalar için Ex post facto (olayın vukuundan sonra) terimi sıklıkla kullanılmaktadır.

Bu yöntemlerin en temel özelliği, mevcut hâlihazır durumu kendi koşulları içerisinde ve olduğu gibi çalışmaktır. Bu yöntemde araştırmacının değişkenler üzerinde hiçbir etkisi yoktur, değişkenler kontrol edilemez. Yalnızca nelerin olup olmadığını rapor edebilir.

Betimleyici veriler, genellikle gözlem, anket, görüşme veya test gibi bilgi toplama yolları ile elde edilir. Bu yöntem karşılaştırmalı ve ilişkisel metotları içeren tüm anket tipleridir.

Örneğin, alışveriş sıklığı, halkın tercihleri, öğrencilerin başarıları, yöneticilerin davranışlar vb. kim, ne kadar, ne zaman, ne yapıyor, neler dinliyor, okuyor, seyrediyor türünde yapılan araştırmalardır.

Buna karşılık, Analitik araştırma, hâlihazırda eldeki verileri kullanarak mevcut durumun veya durumun ileride alacağı boyutun tahmin edildiği araştırma türüdür. Araştırmacı zaten elinde bulunan verinin eleştirel bir değerlendirmesini yapar.

3.1.1.2. *Temel Araştırmaya Karşı Uygulamalı Araştırma*

Araştırma uygulanabilir ya da temel araştırma olabilir. Temel araştırmanın bilginin salt bilgi olarak değerlendirilmesi, bilgi dağarcığına eklenmesi prensibine karşı uygulamalı araştırma, bir kuramı uygulamak ve sınamak ya da uygulamada yaşanan sorunları çözmek üzere gerçekleştirilen araştırmalardır. Genellikle uygulamadaki bir sorunun çözümüne odaklanır ve elde edilen bulgular bu soruna ilişkin karar vermede yararlıdır. Mevcut durumun iyileştirilmesinde bilginin fiilen kullanılması, kullanılabilme yol ve yöntemlerinin geliştirilmesini amaçlar.

Bir genelleme, formülasyon, saf bilgi eldesi temel araştırmalar ile yapılmaktadır. Bazı doğal olgular, matematik temel araştırma türüdür. Toplum, sanayi, kurum, sosyal hayata uygulanması gereken bir problemi çözmeye yönelik çalışmalar da uygulamalı araştırmalar ile yapılmaktadır. Bununla birlikte insan davranışlarına yönelik temel araştırma çalışmaları olsa da bunlar, aynı zamanda bir sosyal problemin çözümüne yönelik uygulamalı araştırmaya yardımcı da olabilirler. Sosyal, ekonomik, politik eğilimleri tespit etmek bu bağlamda değerlendirilebilir.

Kısaca, temel araştırma geniş bir uygulama alanına yönelik bilgiyi bulma çalışmasıdır, buna karşı uygulamalı araştırma, mevcut bilgiyi kullanarak pratik bir probleme çözüm bulma çalışmasıdır.

3.1.1.3. *Ampirik (Deneysel) Araştırmaya Karşın Kavramsal Araştırma*

Kavramsal araştırmalar, bazı soyut fikir veya teorilerle ilgili olarak, yeni kavramlar geliştirmek ya da mevcut kavramları yeniden yorumlamak amacıyla filozof ve düşünürler tarafından kullanılır. Kavramsal araştırma yeni fikirler üstünde duran, daha soyut bir yaklaşımken deneysel araştırma daha çok gözleme ve deneye dayalı bir yaklaşımdır.

Kavramsal analiz daha çok sosyal bilimlerde kullanılan bir yöntemdir. Burada araştırmacı bir teoriyi parçalara ayırarak daha derinlemesine bir inceleme yapar. Bu yöntemde bir hipotez oluşturulur. Daha sonra “Neden?” sorusu üzerine felsefi bir yaklaşımla gidilir; soru derinlemesine irdelenir ve bir sonuca varılır. Fakat bu yöntemin tek başına kullanılmasının getirdiği bazı olumsuz noktalar da vardır, bu yüzden diğer yöntemlerle beraber kullanılması araştırmalar için daha doğru sonuçlar verir.

Ampirik araştırmalar ise, genelde sistem ya da teoriye dayanmaksızın doğrudan deneyim ya da gözleme dayanmaktadır. Bunu deneysel araştırma türü olarak nitelendirebiliriz. Araştırmacının kaynağında, ilk elden gerçeklere ulaşması önemlidir. Araştırmacı kendi hipotezini kurar, daha sonra kanıtlamak veya çürütmek için yeterli gerçek verileri elde etmeye çalışır. Bu yöntemin karakteristiğinde, araştırmacının değişkenler üzerinde kontrol sağlayabilmesi ve etkilerini inceleyebilmek için değişkenleri kasıtlı manipülasyona tabii tutabilmesinin mümkün olması vardır. Ampirik araştırmada, bazı değişkenlerin bir şekilde diğer değişkenleri etkilediğine dair kanıt bulunması uygun olmaktadır.

Ampirik çalışmalar ve deneyler yoluyla elde edilen kanıtların, hipotezler için mümkün olan en güçlü desteği sağladığı düşünülmektedir.

3.1.1.4. *Diğer Bazı Araştırma Türleri*

Diğer tüm araştırma türleri, bahsedilen yaklaşımların bir ya da daha fazla varyasyonlarıdır. Bu varyasyonlar, araştırmacının amacı, araştırmacının başarısı için gerekli zaman, çevre faktörleri ve diğer faktörlere göre oluşturulur.

Araştırma zamana göre, bir kerelik ya da uzun süreli düşünülebilir. Araştırma alanı, simülasyon, deney vb. çevreye göre değişebilir. Araştırma bir olguyu teşhis üzerine olabilir. Belli bir vakayı belli dönem ve belli aralıklarla takip etmek gerekebilir. Bazı çalışmalar küçük örnekler üzerinde ve veri toplama cihazları kullanılarak yapılabilir. Bir

araştırma keşif ya da biçimlendirilmiş olabilir. Biçimlendirilmiş araştırmaların önemli bir yapısı ve belirli bir hipotezi varken, keşif araştırmaları daha çok yeni bir hipotez geliştirme amaçlıdır.

Tarihsel araştırmalar ise, belli bir zamandaki insanların ve grupların psikolojilerini içererek olaylar, fikirler, belgeler vb. tarihi kaynakları inceler.

Araştırmalar sonuç odaklı araştırma ve karar odaklı araştırma olarak da sınıflandırılabilirler. Sonuç odaklı araştırmada da araştırmacı kendi isteğine göre problemi tasarlamada ve kavramsallaştırma da özgürdür. Karar odaklı araştırmada ise, araştırmacı her zaman bu konuda özgür değildir.

Yöneylem araştırması, karar odaklı bir araştırma örneğidir. Kısaca, belirli kısıtların olduğu bir durumda, belirli bir amaca yönelik en uygun çözümün bulunması için geliştirilmiş bir yöntem olarak tanımlanabilir. Yöneylem araştırması; bir organizasyon içinde operasyonların koordinasyonu ve yürütmesi ile ilgili dünyanın gerçek karmaşık sorunları için fikir üretmede matematiksel modelleme, istatistik ve algoritma gibi bilimsel yöntemleri kullanan disiplinler arası bir bilimdir. Organizasyonun doğası maddi değildir. Soruna bilimsel olarak en uygun çözümü sağlamak için bu bilimi kullandıktan sonraki hedef organizasyonun performansını iyileştirmek ve optimize etmektir.

3.1.1.5. *Nicel (Sayısal) Araştırmaya Karşı Nitel (Vasfi) Araştırma*

Nicel araştırmalar ölçümler üzerinden temellendirilmiştir. Nitel araştırmalar ise, içerikle ilgili olarak kaliteyi, niteliksel özellikleri içeren olayla (fenomen) ilgilidir.

Nicel araştırmalar; “Ne kadar? Ne miktarda? Ne kadar sık? Ne kadar yaygın?” vb. gibi sorulara cevaplar arar. Olgu ve olayları nesnelleştirerek gözlemlenebilir, ölçülebilir ve sayısal olarak ifade edilebilir bir şekilde ortaya koyan bir araştırma türleridir. Biyoloji, kimya, fizik, mühendislik gibi doğa bilimleri alanlarında araştırmalar gözlem ve ölçmeye dayanır. Gözlem ve ölçmelerin tekrarlanabildiği ve objektif yapıldığı araştırmalara niceliksel, sayısal (“quantitative”) araştırma denir. Nicel araştırmada amaç, bireylerin toplumsal davranışlarını gözlem, deney ve test yoluyla nesnel bir şekilde ölçmek ve sayısal verilerle açıklamaktır.

Nitel arařtırmalar, “qualitative” arařtırma olarak ifade edilir, insan ve grup davranıřlarının “niçin”ini anlamaya yönelik arařtırma yöntemleri ve psikoloji, sosyoloji, antropoloji, eđitim gibi sosyal bilim alanlarında insan ve toplum davranıřlarını inceleyen arařtırma yöntemlerini kapsar.

“Motivasyon Arařtırması” en önemli niteliksel arařtırma tiplerinden biridir ve burada insan davranıřlarının nedenleri arařtırılmaktadır. Bu şekilde bir arařtırma ile hedef için derinlemesine görüřmeler yapılarak, arka plandaki düřünce, fikir, motif ve istekleri keřfetmek amaçlanmaktadır. Tutum ve davranıřlar, istekler, düřünceler, hisler ile insan davranıřlarının altında yatan nedenler üzerinde nitel arařtırma yöntemi ile çalıřılabilmektedir.

řöyle denebilir ki, nitel arařtırma uygulaması zordur, bu nedenle bu arařtırmada deneysel psikologlardan rehberlik alınabilmesi de mümkündür.

Nicel ve nitel arařtırmaların ařađıda avantajlarını ve sınırlarını görebileceđiz:

Tablo 3.1: Nicel arařtırmaların avantajları ve sınırlılıkları.

| Avantajları | Sınırlılıkları |
|---|--|
| <ul style="list-style-type: none"> • Genelleřtirilebilir sonuçlar üretilir. • Farklı gruplar arasında karřılařtırma yapılabilir. • Kuramların dođruluk derecesi tespit edilir. • Belirli bir yapı içindeki iliřkilerin incelenmesine yarar. | <ul style="list-style-type: none"> • Mükemmel örneklem almak güçtür. • Yeteri sayıda veri toplamak güçtür. • Mükemmel ölçüm şartları her zaman sağlanamaz. • Ölçme aracı önyargıyı da yansıtır. • Model dıřındaki veriler ile ilgilenmez. |

Tablo 3.2: Nitel arařtırmaların avantajları ve sınırlılıkları.

| Avantajları | Sınırlılıkları |
|---|---|
| <ul style="list-style-type: none"> • Özel durumun “gerçekliđini” yansıtır. • Sonuçlarla kuram üretilmesi kolaydır. • Farklı faktörlerin anlaşılmasını sağlar. • Arařtırmanın sonuçlarının uygulanabilirliđi daha yüksektir. | <ul style="list-style-type: none"> • Deneklerin yařadıkları deneyimleri olduđu şekliyle ifade etmeleri zordur. • Verilerin analizinde bireylerin sahip oldukları önyargı da yer alır. |

Yapılan bu incelemeler sonucunda nicel ve nitel arařtırma yöntemleri arasında ařađıdaki farklar ortaya çıkmaktadır:

Tablo 3.3: Nicel ve Nitel araştırma arasındaki farklar.

| Nicel Araştırmalar | Nitel Araştırmalar |
|---|--|
| 1. Gerçeklik nesnelidir | 1. Gerçeklik oluşturulur |
| 2. Asıl olan yöntemdir | 2. Asıl olan çalışılan durumdur |
| 3. Değişkenler kesin sınırlarla belirlenir, bu değişkenler arasındaki ilişkiler ölçülebilir | 3. Değişkenler karmaşık ve iç-içe geçmiştir bunlar arasındaki ilişkileri ölçmek zordur |
| 4. Araştırmacı olay ve olgulara dışarıdan bakar, nesnel bir tavır geliştirir | 4. Araştırmacı olay ve olguları yakından izler, katılımcı bir tavır geliştirir |
| 5. Genelleme yapılır | 5. Derinlemesine betimleme yapılır |
| 6. Sonuçlara ilişkin tahmin kullanılır | 6. Sonuçlarda yorumlama kullanılır |
| 7. Nedensellik ilişkisini açıklama | 7. Aktörlerin perspektiflerini anlama |
| 8. Kuram ve denence ile başlar | 8. Sonunda kuram ve teori oluşturulur |
| 9. Deney, manipülasyon ve kontrol vardır | 9. Kendi bütünlüğü içinde ve doğal gelişir |
| 10. Standartize edilmiş veri toplama araçları kullanılır | 10. Araştırmacının kendisine ait veri toplama aracı kullanılır |
| 11. Parçaların analizi esastır | 11. Örüntülerin ortaya çıkarılması önemlidir |
| 12. Uzlaşma ve norm arayışı vardır | 12. Çoğulluk ve farklılık arayışı vardır |
| 13. Veriler sayısal göstergelere indirgenir | 13. Veriler derinlemesine betimlenir |
| 14. Olay ve olguların dışında, yansız ve nesnellik gerekli ve mümkündür | 14. Olay ve olgulara dahil, öznel, perspektif olan ve empati kurulur |

Nitel araştırmaların kullanışlı bir veri toplama tekniği “derinlemesine görüşmelerdir” ve değerlendirme ihtiyaçları, program geliştirme, konu tanımlama ve stratejik planlama gibi çok çeşitli amaçlar için kullanılabilir. Derinlemesine görüşmeler ucu açık sorular sormak istenilen durumlar için çok uygundur ve görece az sayıda insandan (çok sayıda insan ile yürütülen ve daha çok nicel eğilimi olan anketlere göre) derinlemesine bilgi elde edilir.

Derinlemesine, nitel görüşmeler planlama ve ölçme için kullanılan mükemmel araçlardır. Programları genişletir, çünkü bu görüşmeler açık uçlu, keşif amaçlı metotlar kullanır ve bu yöntemle röportajı yapan kişi görüştüğü kişinin hislerini ve konuya bakış açısını derinlemesine yakalayabilir, konuyla ilgili birçok soruyla şekillenen zengin arka plan içinde sonuçlanır.

Derinlemesine bir görüşmenin ana karakterlerini şu şekilde sıralayabiliriz:

- **Ucu Açık Sorular:** Sorular katılımcıların konuyu açıklayıcı yanıtlar alacak şekilde kurulmalıdır, yalnızca “Evet” ya da “Hayır” cevabı verilemeyecek şekilde yapılandırılmalıdır. Birçok açık uçlu soru katılımcıların kendi kelimeleriyle soruları cevaplamasını özgür bırakacak şekilde “Neden”, “Nasıl” ile başlar.
- **Yarı Yapılandırılmış Format:** Anahtar soruları önceden planlama önemli olmasına rağmen, mümkün olduğunda bir önceki cevapları takip eden sorularla

karşılıklı konuşma şeklinde olmalıdır. Örneğin, röportaj yapılan kişi “Seçimler yaklaşıyor” gibi bir yanıt verirse, bunun arkasından “Adaylar hakkında neler düşünüyorsunuz” şeklinde bir soru sorulabilir.

- **Anlama ve Yorumlama:** Konuşmacının söylediklerini yansıtmak için aktif dinleme yetenekleri çok önemlidir. Görüşmeci söyleneni yorumlamaya çalışmalı ve röportaj boyunca anlama ve netleştirmenin peşinde koşmalıdır.
- **Yanıtların Kaydedilmesi:** Yanıtlar tipik olarak ya bir kayıt cihazı ile ya da yazılarak kaydedilir. Yazılı notlar röportaj hakkındaki kişisel yansımalarından oluşarak, hem sözlü hem de sözlü olmayan gözlemleri içerir.

Anlaşıldığı üzere, derinlemesine görüşmeler sadece sorulan soruları değil, aynı zamanda daha derin bir anlama ve anlamlandırma için soruşturulan cevapları da sistematik olarak kaydetme ve dokümantasyonu da içerir.

Ayrıca, yetenekli nitel görüşmeci şu yeteneklere sahip olmalıdır.

- Açık fikirli olmalıdır. Yargı ve eleştiri iletişim için bariyerler kurabilir. Bu yüzden görüşme süreci boyunca tarafsızlığı sürdürmek çok önemlidir. Eğer röportaj yapılan kişi yargılama ve eleştiri algılsa, muhtemelen düşüncelerini daha az açıklıkla paylaşır. Sonuçları görüşme bittikten sonra bir dergide yazılabileceğini açık bir şekilde söylemek gereklidir.
- Esnek ve duyarlı olmalıdır. İnsan ilişkileri karmaşıktır ve insanların tepkileri nadir olarak öngörülebilir. Bu yüzden, röportajı yürüten kişi çabuk reaksiyon verebilir, zorluklardan etkilenebilir. Kendinden emin olmalı ve konunun özünden kopmamalıdır.
- Sabırlı olmalıdır. Karşısındaki kişinin rahat ve özgür bir şekilde konuşmasına imkân vermelidir.
- Dikkatli olmalıdır. İyi bir röportajcı yüz ifadeleri, vücut dili ve ses tonu gibi toplayacak şekilde dikkatli olmalıdır.
- İyi bir dinleyici olmalıdır. Aktif dinleme kolay görünmesine rağmen, öğrenmek için çok pratik yapmak gerekebilir. Bundan dolayı, röportajı yapmadan önce bir arkadaşınızla aktif dinleme pratiği yapmak önemlidir. İyi bir dinleyici şu stratejileri kullanarak aktif dinleme gerçekleştirir:

- ❖ Katılımcının ne söylediğine tam olarak kulak verme bütünüyle odaklanarak yapılır. Aktif dinleme konuşma bitene kadar ve mesajı alana kadar, dinleyicinin konuşmacıya tam dikkatini vermesini gerektirir.
- ❖ Konuşmacının söylediğini yorumlama konuşmacı için onaylama demektir. Dinleyici aslında sadece dinler ve taşıdığı mesaj aldığı mesajdır. Yorumlama ayrıca konuşmayı dağıtıcı şeylerden kurtularak konuşma üzerine bütünüyle odaklanmak için konuşmacıyı zorlama faydasını sağlar.
- ❖ Yansıtma mesajın doğasındaki duyguları konuşmacıya geri yansıtır. Ses tonu ve duygu ayarına dikkat ederek, röportajcı aktarılan mesaja daha büyük bir anlam kazandırabilir.

Kvale (1996) derinlemesine röportaj yönetiminin aşamalarını detaylandırmıştır: Temalandırma, tasarlama, röportaj yapma, yazıya dökme, analiz etme, doğrulama ve raporlama.

- **Aşama 1- Temalandırma:** Bu aşamada, görüşmenin amacını netleştirmek önemlidir. İlk soru planlanan program için röportaj kullanılıp kullanılmayacağıdır. Önce genel amaca karar verilir, daha sonra derinlemesine röportaj yoluyla toplanmak istenilen anahtar bilgiler hedeflenir.
- **Aşama 2- Tasarlama:** Ne öğrenmek istenildiğine karar verildikten sonra, görüşme yoluyla bilgiyi ortaya çıkarma şekli tasarlanır. Anahtar konuları ve bilgileri toplamak için şekillendirilen soruları içeren bir görüşme rehberi oluşturulur. Konun dağılmaması, önemli konuları ve soruları atlamamak, araştırmanın önemli noktalarına odaklanmak ve görüşmenin konu ve soru sırasını oluşturmak için görüşme rehberi tasarlanmış olmalıdır. Görüşme rehberi bir bölüm olarak görüşme tarihi, yeri, görüşülen kişinin bilgileri, ikinci bölüm görüşme soruları, son olarak da görüşme sonrası notları, yorumlar, hislerin yazıldığı 3 bölümünde oluşur.
- **Aşama 3- Röportaj:** Röportajın başlangıcında bir giriş yapmak, çalışmanın amacını anlatmak, katılımcıyı rahatlatmak çok önemlidir. Eğer bir ses kaydı yapmak planlanıyorsa, katılımcının izni alınmalı ve düzgün çalışıp çalışmadığı kontrol edilmelidir. Röportajı yapanın ana sorumluluğu konuşma boyunca görüşme rehberinin doğrultusunda dinlemek ve gözlemlemektir.

- **Aşama 4- Yazıya Dökme:** Ses kaydı ile ya da görüşme esnasında cevapların not ederek kayıt edilmesi sonrası katılımcının ifadeleri kelimesi kelimesine aktarılır. Konular sorulara, cevaplara göre bölümlere ayrılır.
- **Aşama 5- Analiz:** Katılımcının yanıtlarından oluşan konuları tanımlamak için görüşme metinlerini yeniden okunması aşamasıdır. Sorulara verilen cevaplar incelenerek çözümlenir.
- **Aşama 6- Doğrulama:** Toplanan bilgilerin doğruluğunu kontrol etme aşamasıdır. Bu yöntem “triangulation” (üçgenleştirme, nirengi) olarak adlandırılır. Bu amaca ulaşmak için kullanılan yaygın bir yöntemdir. Triangulation bir bilgi setinin yorumlamak için çoklu bakış açısını kullanmayı içermektedir. Örneğin, aile bireyleri arasındaki iletişim ile ilgili yapılan bir çalışmada katılımcılar – anne, baba ve çocuklar – röportajda kesişen aynı şeyleri söylerlerse sonuçlanır. Triangulation yönteminin daha basit bir şekli, yapılan röportaj metinlerini iki ayrı kişi okur ve analiz eder, aldıkları notları karşılaştırırlar. Eğer notlar uyumlu ise, bilgi güvenilirdir.
- **Aşama 7- Raporlama:** Son olarak, yazılı ya da sözlü rapor yoluyla iç ve dış paydaşlar ile derinlemesine röportajdan elde edilen sonuçların paylaşımı önemlidir. Bu raporlar sadece sonuçları açıklamamalı, aynı zamanda gelecek çalışmalara nasıl şekil vereceğini de açıklamalıdır. Katılımcılar kullanılan bilgileri gördüğü zaman, gelecekteki veri toplama çabalarına muhtemelen daha fazla katılacaklardır.

3.1.2. Araştırmanın Yöntemi

Yapılan bu incelemeler ve değerlendirmeler sonucu, bu çalışma nitel araştırma yöntemi üzerinde yapılandırılmış yorumsal yaklaşım tarzını izlemektedir. Nitel araştırma doğaldır ve bir olayın, bir bireyin ya da bir kurumun bütünsel anlamının elde edilmesini araştırır. Orlikowski ve Baroudi’ye göre (1991, p. 6) yorumsal çalışmalar “insanların çevreleriyle etkileşimde bulunurlarken, kendi kişisel bakış açılarını oluşturduklarını ve bunları ilişkilendirdiklerini varsayar. Bundan dolayı yorumsal araştırmacılar katılımcıların düşüncelerine erişim yoluyla olguları anlamaya çalışırlar.

Bunu yapmak ve derinlemesine bir anlayış elde etmek için arařtırmacı katılımcıların çevresiyle “gözlenen olgu, katılımcıların deneyimleri ve ifade ettikleri görüşleri” içerisinde karışmış olur (Callery 2012, p. 11).

Callery (2012, p.5) şunu da ifade etmiştir “Eğer teorinin temeli tam olarak gelişmemişse, arařtırmacı deęişkenlerden tam olarak emin deęilse ve çalışmanın keşfedilen olgusu küçükse, bu durumda nitel arařtırmalar bilgi toplamanın en uygun şeklidir.” Çünkü nitel yöntemler, kurum ve oluşumlardaki sosyal çevreler, dinamizm içerisinde insanların nasıl anladığı, deneyimlediği ve işlediği ve tam anlama üzerine başlıca değeri merkezileştirir ve yerleştirir (Tewksbury, 2009). Çalışma Büyük Veri uygulayan kişilerin deneyimlerini, algılarını ve şirketlerinin Büyük Veri konusundaki beklentilerini analiz etmesini gerektirdiğinden dolayı, bu çalışma için nitel araştırma planı uygundur.

Ek olarak, görüşmeler gece gözlüğüne benzer bir şey olarak tanımlanır; görünenin sıradan olmadığını görmek ve nadir olanı görüp incelemek (Rubin & Rubin 2005: viii). Bu bağlamda, Büyük Veri konusu için nitel yaklaşımdan çok nicel yaklaşım yönteminin daha uygun olacağı ilk bakışta görülebilir. Fakat Rubin ve referans olarak Rubin'den alıntılanlar, “röportajların kullanılmasının avantajını; normal şekilde kullanıma hazır olmayan veya erişilemeyen derinlemesine bilgi elde etmektir” şeklinde net olarak açıklar. Konuyla ilgili yapılan görüşmeler, tam anlamıyla nicel yaklaşımın uygun olmayacağı bir tarzda, Büyük Veriye – nadir görülen değerli veri – karşı tutumlarını ve düşüncelerini açığa çıkarır.

Daha önceki sayısız arařtırmacı iş alanındaki arařtırmalar için röportaj kullanmanın faydaları ve avantajlarına işaret etmiştir. Örneğin, 1971’de Daniels beş farklı ülkede görüşmeler yönetti, kuramsallaşma için bu doğrudan doğruya elden edilen bilgileri kullandı ki, şirketler birçok arařtırmacının o zamandan beri test ettiği oluşan yeni teoriyi artan bir ilke üzerinde uluslararası olarak ilerletti. Teigland (2000), uluslararası bilgi dağıtımını konusu üzerine üç farklı şirket arasında karşılaştırmalar yapmak için üst düzey yöneticiler ile görüşmeyi kullandı. Bu veri sadece sayısal temeller üzerinde elde edilemezdi. Ek olarak, bilim adamları örnek olay incelemesi olarak röportajlardan faydalanabilirler (Daniels and Cannice 2004). Daniels ve Cannice ayrıca, zengin veri toplama potansiyeline sahip olası cevaplayıcıların küçük bir popülasyon olduğu zamanda, röportajların uygun olacağını işaret etmiştir. Parkhe (1993), Boyacıgiller ve Adler (1991,

p.266) iş içerikli görüşmelerin güçlü yönüne işaret ettiği kabiliyetleri “uluslararası iş araştırmasının konusu olan dinamik inceleme, bağlam-bağımlı ve interaktif olaylardır”. Çünkü Büyük Veri girişimleri ve Büyük Verinin rolü çok yönlü bir işlemdir ve olmaya devam edecektir. Bu ülkeden ülkeye, hatta şirketten şirkete de fark edebilir.

Üzerinde durulan nokta, genel kısımda daha önce açıklanan öğeler arasındaki ilişkileri vermekteydi, bununla birlikte daha önce, görüşmelerin sonuçları ve geniş çaplı anketler belgelendirilmiştir. Amaç, potansiyel özgün araştırma problemini çözmek için daha önceleri keşfedilmeyen ve bağlanamayan, bilgi kaynakları arasındaki uygun bağlantıları ortaya çıkartmaktır (Collis and Hussey, 2009). Bu proje için öncelikli veri, anlaşılabilir ve değerli geri dönüş veren, derinlemesine yapılan yarı yapılandırılmış röportajlardan gelecektir. Katılımcıların yanıtları onların beklentileri ve varsayımlarını ölçmek, Büyük Veri girişim ve önerilerinin uygulanabilirliğini kontrol etmek için kullanılacaktır.

Lester’a göre (1999), olgusal yaklaşımın amacı “onların bir olaydaki aktörler tarafından nasıl algılandığı vasıtasıyla özel tanımlı olayları aydınlatmak” tır. Araştırmacı genellikle tümevarım, nitel yöntemler aracılığıyla derin bilgi ve algıları toplar. Örneğin, röportajlar, tartışmalar ve katılımcı gözlemi. Araştırma katılımcısının perspektifinden onu temsil eder. Olgusal araştırma için yorumsal ölçüye ek olarak, pratik teori için ilkeler gibi kullanmayı sağlayıp, bilgilendirmek için izin verir, politika ve aksiyonu destekler ve meydan okur. Streubert ve Carpenter (1999: 48) şunu ifade etmektedir, olgusal yaklaşımın en önde gelen girişimlerinden birisi “sıkı, eleştirel, olayın sistematik araştırmasını” üreten yeteneğidir.

Bu araştırmanın problemi şudur : “Büyük Veri hakkındaki genel eğilimler ve Stratejik Büyük Verinin yatırımlar üzerindeki etkileri nelerdir?” İnceleme ve bu süreçlerin yönetim tarzlarını keşfetmek yorumsal anlayış ile desteklenebilir. Dahası, yorumsalcılık bir çerçeve gibi bu araştırmaya rehberlik edecektir. Çünkü bu potansiyeli düşük güvenilirliğe rağmen zengin nitel veri ve bulgular üretmeye eğilimlidir (Collis and Hussey, 2009). Daha önce belirtildiği gibi, yorumsalcı, nitel ve olgusal yaklaşımlar şu anlamdadır: Araştırma sorusu katılımcıların kendileri tarafından kısmen cevaplanmış olacaktır. Kurumlar belki de Büyük Veri ile ilgili olarak objektif veriler sağlasa da, kurumların yetkililerinin deneyim ve beklentileri bu araştırmanın temel bir parçası olacaktır.

3.2.VERİ TOPLAMA

İlk aşama önceki teorileri, makaleleri, kitapları ve iş hayatındaki operasyonların gelişmesi için Büyük Veriyi kullanma yolları ve metotlarına ilişkin araştırmaları çalışmaktı. Bu ikincil bilginin toplanmasıyla ve bu yayınlar etrafındaki bilgi kaynağıyla yapılmıştır. Bir sonraki aşama görüşülen kişilerin kurumlarında Büyük Veriyi kullanarak operasyonel gelişimini nasıl çözümlendiği ve bu sıradaki imkânlar ve zorlukların ne olduğuydu. Başlıca veriyi toplamanın bu basamağı yüz yüze bir şekilde yarı yapılandırılmış röportajlar ile başarıldı.

Araştırmacı röportajı için rehber olacak hazırlanmış sorulara ve tartışılacak konulara sahip olduğu zaman, yarı yapılandırılmış bir röportaj yoluyla bilgi toplamak için esnek bir yöntem olacaktır. Araştırmacılar çoğu kez katılımcılara, bir grup halinde dostça münasebet ve karşılıklı interaktif şekilde farklı yollardan sorularını sorabilirler (Lindlof ve Taylor, 2002). Barriball ve While (1994) ise şuna işaret etmiştir; yarı yapılandırılmış röportajlar katılımcıların inançları, güdüleri, değerleri ve davranışlarının araştırması için çok uygundur. Çünkü bu çalışma Büyük Veri girişimiyle ilgili çeşitli insanların davranışlarını, inançlarını ve beklentilerini belirleyecektir, bu şekildeki röportajın en uygun veri toplama yöntemi olduğu saptanmıştır.

Çalışmanın amacı iş süreçlerindeki inovasyon faaliyetlerini belirleyen anahtar faktörleri keşfetmeye çalışmaktır. Derinlemesine görüşme yöntemi “algılar, tutumlar ve inançlar hakkında bilgi edinmek için en uygun yöntem olarak kabul edilmektedir.” (Frankfort-Nachmias and Nachmias, 1996; Finn, Elliot- White and Walton, 2000; Denzin and Lincoln, 1994 cited in Buhalis and Deimezi, 2004). Kullanılan yarı yapılandırılmış röportajlar, dolgun bir resim sağlayan katılımcıyı göz önünde bulundurmayı ve ortaya çıkartarak yeni alanlar takip etme yeteneği sağlar.

Bryman (2004: 321) süreç için; nitekim araştırmacı bir sorular listesine ya da oldukça özgün üzeri örtülü konulara sahiptir, fakat görüşülen kişi soruları cevaplama konusunda tam bir özgürlüğe sahiptir, demektedir. Bu yöntem seçilmiştir çünkü yazar ana konuları belirler, fakat hala görüşmeyi yaptığı kişi Büyük Veri uygulamaları hakkında konuşma özgürlüğü için izin istemektedir. Bu teknik de yaratıcı görüşmenin bazı seviyelerini gerektirir.

3.3. VERİ ANALİZİ VE RAPORLAMA

Veri analizi için, Collis ve Hussey (2009, p.21)'in prosedürü şu ana özellikler ile takip edildi: Verinin “azaltılması, yeniden yapılandırılması ve kavramsallaştırılması”. Kodlayıcı kullanım sayesinde, araştırmanın amacı açısından daha az önemli konulara atıf yapan kişilerden alınan uygun veriler seçilmiştir. Daha sonra veri faktörler grubu ve katılımcılar üzerinde yapılandırılan kategoriler içerisine uygun hale getirildi. Bunun tamamlanması üzerine, veri yakından incelenerek, karşılaştırılarak, etiketlenerek ve organize edilerek farklı yollar ile incelenmiştir.

Tüm sonuçların derlenmesinden sonra, farklı cevaplar alanlarının üzerinde yapılandırılarak kümelendirilmiştir. Eğer Büyük Veri uzmanları tarafından örnek cevaplar için belirgin farklılıklar olduysa, farklı değişkenler ve cevaplar arasındaki potansiyel ilişkiler incelenmiştir.

3.4. ARAŞTIRMANIN KISITLARI

Araştırmanın nitel yaklaşımı ve röportajların kısıtlı örneklerinden dolayı kusursuz sonuçlara sahip olmak mümkün değildir, bu bir istatistiksel temsildir ve bu çalışma değerlendirilirken dikkate alınmalıdır. Bu kısıt her alandaki en saygın ve en güncel bibliyografik kaynaklar kullanılarak kısmen aşılmıştır. Bu bulgunun sonuçları geniş bir şekilde bütün kurumlara veya şirketlere mutlaka anlamlandırılmış olmayabilir. Bununla birlikte, Büyük Veri girişimleri bir iş operasyonunun artan şekilde daha geniş bir parçası olduğundan dolayı, bu çalışmanın ölçeğini sayısal verilerin de takviye için kullanılma nedeniyle mümkün olduğunca arttırılmaya çabalamak gereklidir: Diğer veri türleriyle birleştirilen röportaj verisi için çalışan birçok nitel araştırma içerisinde kesinlikte ortaktır (Flick 2009).

Diğer kısıtlar, varsayımlar ve önerilen araçlar test edileceği için, bu çalışmanın içeriğindeki sınırlı imkânları kapsar, diğer taraftan önerilerin geniş çaplı bir uygulamasını yapmak için yeterli güveni verir. Diğer bileşenlerdeki benzer araçlar çalışan başarılı bir analiz bize kısmen de olsa önerinin fizibilitesini test etmeye izin verir.

Son olarak, nitel ve nicel araştırma yaklaşımının kombinasyonu ideal olacaktır. Bu yöntem ankete katılanların sayısının artmasıyla birlikte konunun her yönüyle idrak edilmesine imkan verecek (Buhalis et al., 2004) ve bütün sektörler için sonuçları

genelleştirmeyi mümkün kılacaktır. Sonuç olarak, araştırma ana amacı olan nitel yaklaşımla başarılı bir şekilde elde edilen, konunun derinlemesine anlaşılmasına odaklanıp devam etmiştir.

3.5. ETİK HUSUSLAR

Katılımcılar araştırmaya katılmayı isteyerek kabul etmiştir. Orb ve arkadaşları da şuna işaret etmiştir, rıza “sürekli tartışmayı gerektiren güvenin tartışması”dır ve “üst-bilgilendirme ve alt-bilgilendirme arasında makul bir denge yapan” anlamındadır (Orb et al. 2001, p.93). Orb ayrıca şu hususa da değinmektedir; araştırmanın bütünlüğünün ödün verilemez parçaları olan tarafsız kalma, gizlilik isteklerine saygı ve araştırma fonlama kaynaklarını garanti etme araştırmacıların aklında tutması gereken, hayati önem taşıyan faktörlerdir. Bryman ve Bell de aynı zamanda şu şekilde belirtmiştir; eğer araştırmacılar tüm taraflardan izin ve tam rızalarını sağlama konusunu garanti ederse, güven seviyesini inşa etmek için büyük ölçüde katkı sağlayacaktır.

Bir araştırma yöntemi olarak röportaj/görüşme göz önünde bulundurulması gereken bazı özel etik kaygılar taşır. Örneğin, (Streubert & Carpenter, 1999) şu hususa dikkat çekmiştir; araştırmacılar “bir röportajın muhtemel çıktılarını öngörmek ve hem faydalarını hem de potansiyel zararlarını tartmak zorunluluğuna” sahiptir (Orb et. al 2000). Bir başka deyişle, röportajı yapan kişi ayrıca kendi işini mümkün olduğu kadar doğru bir şekilde yapmaya çalışırken “risk değerlendirmesini” aklında tutmak zorundadır. Bu çalışmada, kurumların bazı yöneticileri Büyük Veri Girişimi hakkında olumsuz konuşmak için isteksiz olabilirler, kurumlardan bazıları gizli projeler yapan kurumlar olduğundan dolayı, eğer veri girişimde diğer girişimlerin sahip olduğu dezavantajlar ortaya çıkarsa, kötü bir yönü göstermek istemezler. Onların dürüst düşüncelerini ve duygularını sağlamak önemli bir amaçken, bu röportaj yapılan kişiyle olan güven ve samimiyetin inşasıyla ve hiçbir şekilde onlara zarar vermek için kullanılmayacağını hissettirerek sağlanmış olmalıdır.

Ayrıca, katılımcılar kendi kişisel önyargılarının ve bakış açıklarının araştırma üzerinde hiç bir önceliği olmayacağına ve de görüşlerinin eşit adil ve eşit bir tarzda dikkate alındığına ve değerlendirilmiş olduğuna ikna edilmişler.

4. BULGULAR

Bu bölümde yaptığımız görüşmelerde verilen cevaplar ve değerlendirmeleri yapılacaktır. Görüşme analiz tablosunda katılımcıların cevaplarından oransal olarak nicel bir veri elde edilecektir. Elde edilen oranlar ile yapılan görüşmelerdeki yaklaşımlar birlikte değerlendirilerek analiz edilecektir.

4.1.ÖRNEKLEM

Veri ya da Büyük Veri ile ilgilenen kurumların yöneticileri ya da operasyonel birim personeli gibi çeşitli aktörler röportaj için seçildi, işini iyi bilen uzmanlar ve belli bir ürün üretimi, müşteri memnuniyetini dikkate alan kurumlar olmasına özen gösterildi.

Tablo 4.1: Örneklem tablosu.

| Kişi | E/K | Pozisyon | Sektör | Kurum Genel Özellikleri |
|------|-----|---------------------------------------|------------------------------------|---|
| 1 | E | Takım Lideri | Çağrı Merkezi, Telekomünikasyon | ~10.000 Kişilik. 2007'den beri faaliyette. Müşteri odaklı çözümler sunan bir firma. |
| 2 | E | Proje Yöneticisi | Ar-Ge, Bilişim | ~3.500 Kişilik. Proje odaklı çalışan kamu kurumu. |
| 3 | E | IT Müdürü, Süreç ve Sistem Geliştirme | Otomotiv | ~25.000 Kişilik. Uluslararası bir otomotiv firmasını Türkiye grubu. |
| 4 | E | Büyük Veri Teknik Takım Lideri | Büyük Veri, Telekomünikasyon | 2011'de kurulmuş ve Büyük Veri üzerine çalışan firma. |
| 5 | E | İş Geliştirme Uzmanı | Ar-Ge, Bilişim | ~3.500 Kişilik. Proje odaklı çalışan kamu kurumu. |
| 6 | E | Yönetici | Bilişim | Mühendislik danışmanlık hizmeti veren Kobi. |
| 7 | E | Satınalma Süreç Sorumlusu | Ar-Ge, Bilişim | ~3.500 Kişilik. Proje odaklı çalışan kamu kurumu. |
| 8 | E | Müdür | Bilişim | Yazılım ve e-imza konularında Kobi. |

Görüşülen kişilerin seçimi, aktörlerin, bilgi alanlarının hem de içeriğin çeşitliliğini yakalamak niyetiyle tasarlanmıştır. Çeşitli insanlar grubu içerideki yönetici ve operatörlerin bakış açılarını getirmek için örnek olarak tercih edildi. Bu kişiler dışarıdan uzmanların ve müşterilerin bakış açılarıyla birlikte kendi dönemlerindeki Büyük Veri kullanımının nasıl olduğunu yönetmeye etkin bir şekilde gayret gösterirler. Bu kişiler yenilikçi uygulamaları ve net bir şekilde prosedürleri önererek, yeni ürün ve servis türleri için alışılmışın dışında, orijinal fikirler gelişmesine yardım edebilirler, var olan problemlerin için çözümü için yeni yollar ortaya koyarlar ve iki tarafa (kurum ve müşterileri) karşılıklı fayda gelişmesi için girişimde bulunurlar.

Ayrıca görüşülen kişilerin hem pozisyonları itibariyle hem de tecrübeleriyle itibariyle, kurumlarındaki ilgili oldukları projelere etki edebilecek kişilerden seçilmeleri de önemli bir husustur. Bundan dolayı yeni fikirleri ortaya koyma, takibini yapma, geliştirme süreçlerinde aktif rol oynamaktadırlar. Projelerin saha uygulama safhalarında da bizzat dâhil olmaları, araştırma konusu ile ilgili olarak pozitif ya da negatif yanları daha iyi ortaya koyabilmelerini sağlamıştır.

Yine kendi pozisyonlarındaki lokal sorumlulukları ve bilgileri dışında konu ile ilgili genel literatür bilgisine ve konunun sürekli gelişmesinin gereği olarak güncel bilgilere de sahiptirler. Bu görüşme esnasında kişilerle güncel olarak aynı hususları paylaşabilme açısından avantajlar sağlamıştır.

Bu iç-dış yaklaşımlar ve bakış açıları arasında güçlü ve sağlam kıyaslama yapmak için fırsat sağlayacaktır. Bu da bu araştırma sorusu için en uygunu olan yorumlayıcı-olgusal araştırmanın bir başka özelliğidir.

Vurgulanması gereken şudur ki, aktörlerin farklılığı Büyük Veri kullanımındaki geniş kapsamlı bir perspektif kazanmak ve ana aktörlerin görüşlerindeki farklılıkları araştırmak için önemliydi. Farklı aktörler, özgün ve farklı avantajlı noktalarının tümü, algılarını çalışmak, fikirlerin görünen çatışmalarını gözlemlemek ve ilginç bulguları sonuçlandırmak için girişimde bulunmuştur.

4.2. GÖRÜŞMELER

Bu çalışmanın tamamlayıcı bir parçası yöneticiler, operatörler ve uzmanlar ile olan röportajlar ve konuşmalar üzerine dayandırılmıştır. Yarı yapılandırılmış röportajı

yürütmek için bir standart görüşme rehberi (EK.1) yapıldı. Soruların amacı çalışma tarafından kapsanan ana konular hakkında daha fazla şey öğrenmek, karşı karşıya kaldıkları yönetsel ve teknik zorluklar ile ilgili yöneticilerin/operatörlerin düşüncelerini öğrenmeye çalışmaktır. Bu çalışmadan gelen geribildirimler, kullanabilecekleri ya da bilecekleri (kullanılan Büyük Verinin içeriğindeki) araçlar ve tekniklerdir. Diğer taraftan, uzmanlar ile gerçekleştirilen konuşmalar, belirli alanlar içerisindeki anlayışlara, bu araştırmanın uygunluğu ve oryantasyonu hakkında geribildirime, kurumların Büyük Veriden nasıl yararlanmış olabileceği üzerine tavsiyeler hakkında daha başka bilgilere erişmek için tasarlanmıştır.

Aktörlerin kimlik belirleme süreci ve son röportajların ayarlanması 4-5 hafta devam etti. Her biri yaklaşık 60-90 dakika süren röportajlar görüşme yapılan kişilerin işyerlerinde yürütüldü.

Kurumların çalışanları ile yürütülen tüm görüşmeler “Büyük Veri sektörünüzde kullanılıyor mu?” sorusuyla başladı. Bu konudaki derinlemesine tartışmadan sonra, kurumun iş operasyonlarının geliştirilmesi için Büyük Verinin nasıl kullanılacağı ile ilgili anahtar yönetim tekniklerinin verilmesi istendi. Ondan sonra, kurumlarındaki yenilikçi yetenekler için zorluklar ve kolaylıklar olarak hareket alanı tartışıldı. Son bölümde, görüşülen kişiye konuyla ilgili tamamlayıcı görüş ve potansiyel çözüm soruldu.

Görüşme yapılan kişiler özgürce konuşmakta, düşüncelerini ve deneyimlerini ifade etmekte teşvik edilmiştir. Ayrıca, araştırma (izleme), konunun daha iyi anlaşılmasını sağlamak, farklı faktörleri bağlamak, bir faktörün önemini sınamak, diğer görüşlerle karşılaştırma yapmak ve cevabı etkileyebilen potansiyel önyargıyı tanımlamak için kullanılmıştır.

Bütün röportajlar görüşülen kişi ile yakın ilişki kurulmak suretiyle kaydedildi. Görüşme esnasında notlar tutuldu, karşılıklı tartışmalar yapıldı ve bunlardan anlaşılan sonuçlar soru listesinde cevaplar olarak işaretlendi.

4.3.NVIVO İLE GÖRÜŞMELERİN İNCELENMESİ

4.3.1. Bilgisayar Destekli Nitel Veri Analizi Hakkında

Ian Dey “Bilgisayar programları analizimize yardımcı olur, ancak bizim yerimize analiz yapmaz” ifadesi bu süreci özetlemektedir. Bu bağlamda, “insan” faktörü nitel araştırmaların güçlü ve zayıf yönünü oluşturur. Güçlü yanı kişinin deneyimlerine ve anlayışına bağlıken zayıf yönü ise araştırmacının becerileriyle, eğitimiyle, çalışma alanıyla ve yaratıcılığı ile çok yakından ilgilidir. Araştırmacı, nitel çalışmanın önemli bir parçasıdır, bu nedenle çalışmanın kalitesini bireyin araştırma becerisi belirler (Patton, 1988) ifadesi nitel veri analizlerinde temel unsurun “insan” olduğuna vurgu yapmaktadır.

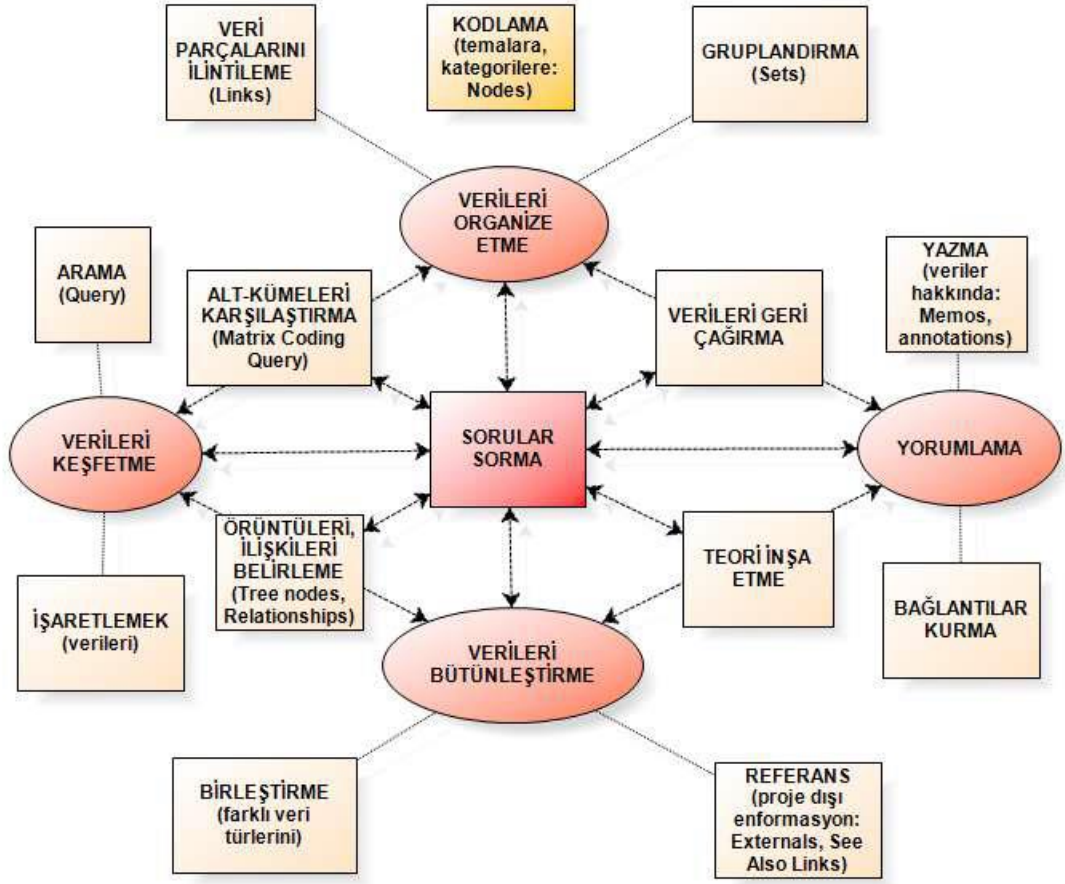
CAQDAS (Computer Aided Qualitative Data Analysis) genel adı altındaki programlar, çok farklı yazılım paketlerini kapsamaktadır. Hepsinin ortak noktası, nitel verilere yönelik nitel yaklaşıma sahip olmalarıdır (Lewins ve Silver, 2007).

Bugün gelinen noktada Fielding ve Lee'nin (2002) belirtmiş oldukları gibi, bu programları basit bir biçimde kodlama ve geri çağırma programları olarak adlandırmak yetersizdir. Bu sebeple, öne çıkan ATLAS.ti, MAX.qda ve NVivo gibi programlar, teori inşası ya da hipotez testi gibi süreçleri destekleyen çok daha gelişkin işlemlere sahiptir.

Ayrıca bu programlar artık, sadece metin formundaki verilerin analizine imkân tanımının ötesinde görsel ve işitsel (ses kaydı, resim, video gibi) verilerin analizini de mümkün kılmaktadır.

CAQDAS yazılım paketlerinin gerisinde yatan temel fikirler Lewins ve Silver (2007) tarafından şu şekilde şekillendirilmiştir (Şekil 4.1). Resimde süreçlerin NVivo'daki hangi işlemlerle karşılandığı parantez içinde gösterilmiştir.

Bu çalışmada görüşmelerin analizi NVivo ver.10 programı ile düzenlenmiş ve incelenmiştir. NVivo'nun nitel araştırmacıya sunduğu üç temel olanak; depolama (sources), kodlama (coding) ve sorgulama (query) olarak sıralanabilir.



Şekil 4.1: CAQDAS (Computer Aided Qualitative Data Analysis) ve NVivo.

(Kaynak: Lewins ve Silver, 2007)

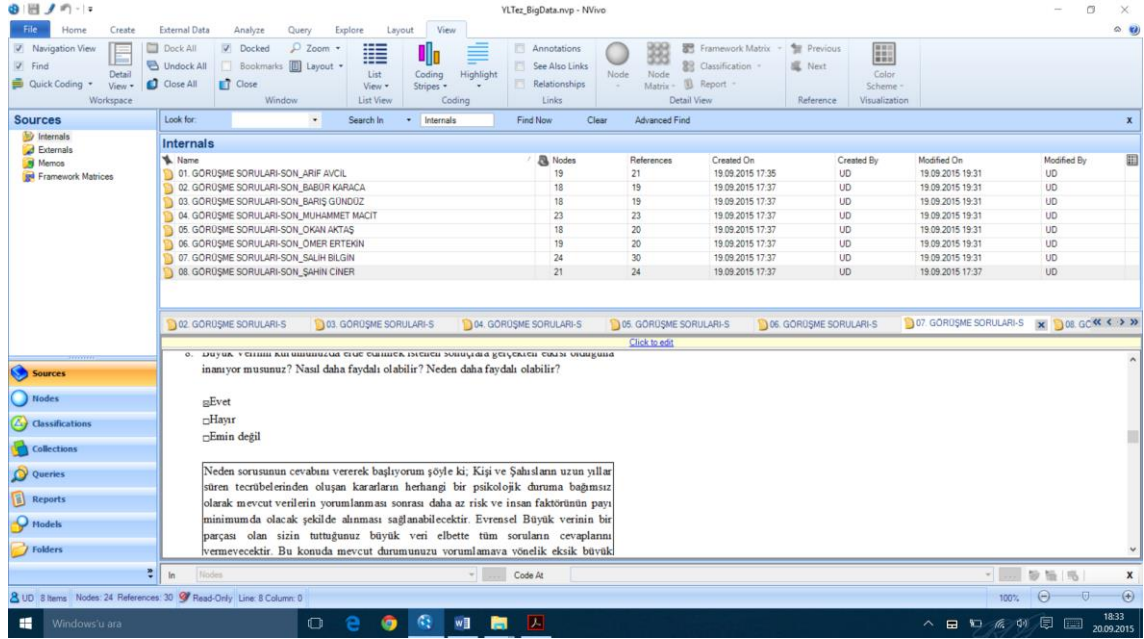
Temel olarak nitel veri analizi için kesin doğru bir yol yoktur. Verinin nasıl analiz edileceği araştırmacıya, veriye ve çalışmanın amacına bağlıdır. Coffey ve Atkinson (1996) veri analiz sürecinin kapsamlı ve sistematik olması gerektiğini, ancak bu süreci her araştırma için geçerli olabilecek standart bir süreç haline getirmenin mümkün olmayacağını belirtmektedir.

Wolcott (1994) veri analizinde üç yol önermektedir. Birinci yol, toplanan verinin özgün (orijinal) formuna mümkün olduğu kadar sadık kalarak ve gerektiğinde araştırmaya katılan bireylerin söylediklerinden doğrudan alıntı yaparak betimsel bir yaklaşımla verileri okuyucuya sunmaktır. İkinci yol ise, birinci yaklaşımı da içeren bir biçimde, bazı neden-sonuç ve açıklayıcı sonuçlara ulaşmak amacıyla “sistematik analiz” yapmaktır.

Üçüncü yaklaşım ise araştırmacı, birinci ve ikinci yaklaşımı temel alır ve buna ek olarak, veri analizi sürecine kendi yorumlarını da dâhil eder. Bu çalışmada genel olarak üçüncü yaklaşımın ağırlıklı olarak etkisi hissedilecektir.

4.3.2. NVivo ile Görüşmelerin Değerlendirilme Süreci

Görüşmeler için her bir görüşmeci ile yapılacak röportaj esnasında sorulan 16 sorudan oluşmaktadır. Bu sorularla birlikte konu ile ilgili 8 kişi ile görüşülmüş, 128 cevap alınmış, görüşmelerin Word düzenlemeleri 36 sayfadan oluşmuştur. Öncelikle yapılan görüşmeler Word dosyası formatında NVivo ver.10 programının Sources / Internals bölümüne Import edilmiştir (aktarılmıştır). Sources kısmı yaptığımız görüşmelerin dökümlerinin yer aldığı bölümlerdir.



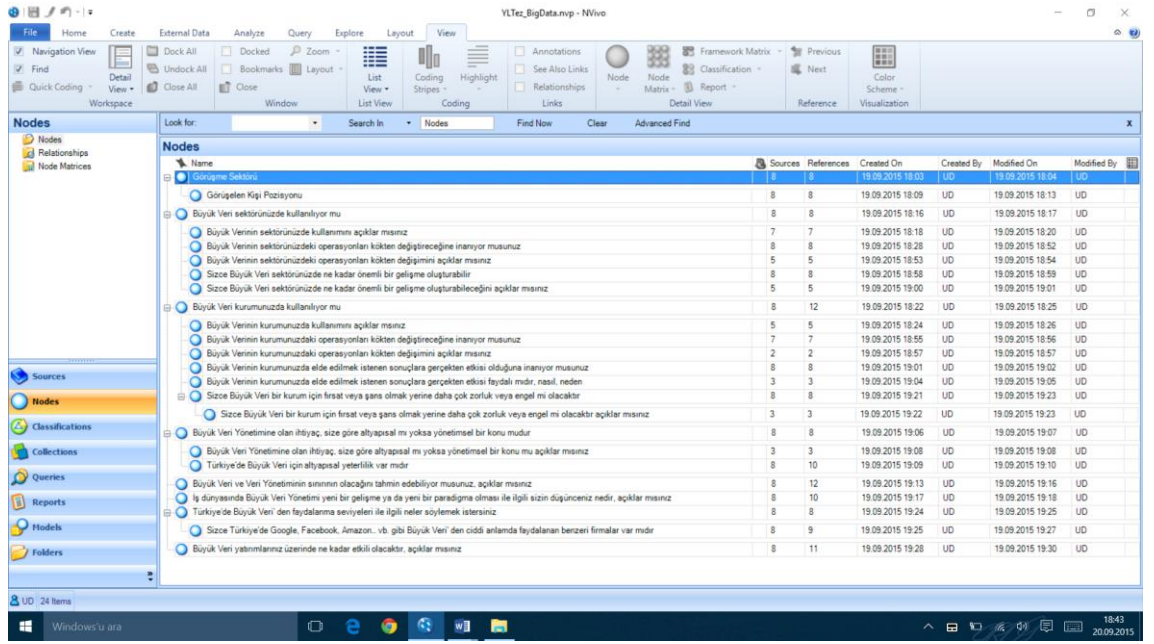
Şekil 4.2: NVivo programına görüşmelerin yüklenmesi.

(Kaynak: NVivo çıktısı, 2015)

Görüşme kaynaklarımızı programa yükledikten sonraki ikinci adım, kaynakların Nodes olarak ifade edilen kodlamalarının yapılması sürecidir. Nodes kavramı analiz sürecinin en temel noktası olarak ifade edilebilir. Nodes'lar verileri kodlarken kullanacağımız case, kod, kategori, temalar vb. yani aynı şeyi anlattığımız düşündüğümüz cümleleri toplayacağımız kelime veya kelime guruplarıdır. İki şekilde node'ları düzenlemekteyiz:

Free Nodes'lar herhangi bir kategori, kod vs. ile ilişkisi olmadığını düşündüğümüz nodes'lardır. Tree Nodes'lar ise, Nodes'ların alt-üst gibi ilişkileri olduğunu düşündüğümüzde kullandığımız Nodes'lardır. Her alt veya üst kategori eklemenizde ağaç şeklinde ilişki ağımı görebilirsiniz.

Bu şekilde Sources bölümüne yüklediğimiz görüşmelerin Nodes'lar ile kodlaması yapıldı. Aşağıdaki ekran görüntüsünde oluşturulan Nodes ve Tree Nodes'lar görünmektedir.

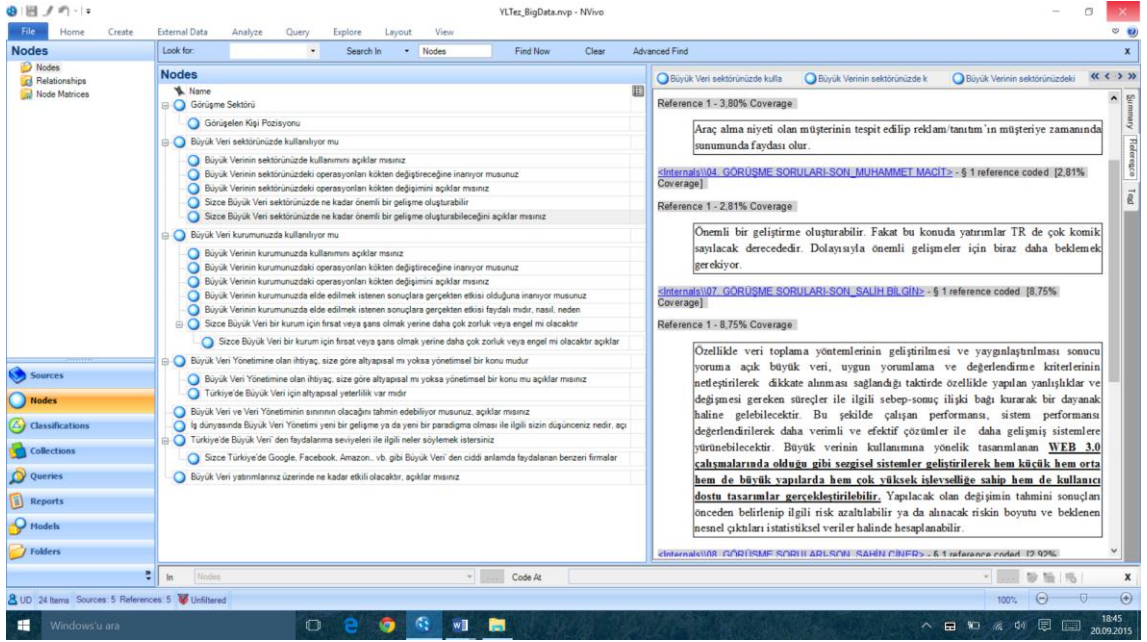


Şekil 4.3: Nodes ve Tree Nodes kodlama görünümü.

(Kaynak: NVivo çıktısı, 2015)

Her bir ilgili Nodes'lara diğer ilgili Nodes'lar ilintileştirilerek Tree Nodes'lar oluşturulmuştur. Böylece birbirine bağlı konuların ve noktaların incelenmesi daha kolay hale gelmiştir.

NVivo her bir Node ve ona bağlı Tree Nodes'ların hangi kaynaklardan geldiği ve kaç referans noktası olduğu ve bu kaynaklara ya da referanslara kolayca ulaşımı sağlamaktadır.



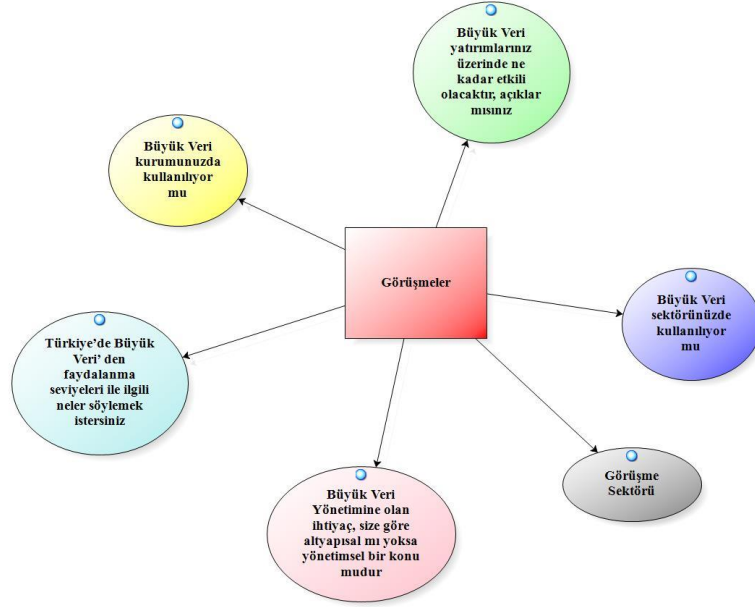
Şekil 4.4: Nodes ve Tree Nodes kodlamalarının incelenmesi.

(Kaynak: NVivo çıktısı, 2015)

Bunların yanı sıra çalışmalarda NVivo'nun bize sunduğu diğer olanaklardan yararlanılmıştır. NVivo ile birbiriyle ilişkili metinler ve konular kolayca bize rehberlik edecek şekilde görüntülenebilmektedir. Böylece elimizdeki yarı yapılandırılmış veriden daha fazla olgusal yorumlar çıkarma imkânını elde etmiş olmaktadır.

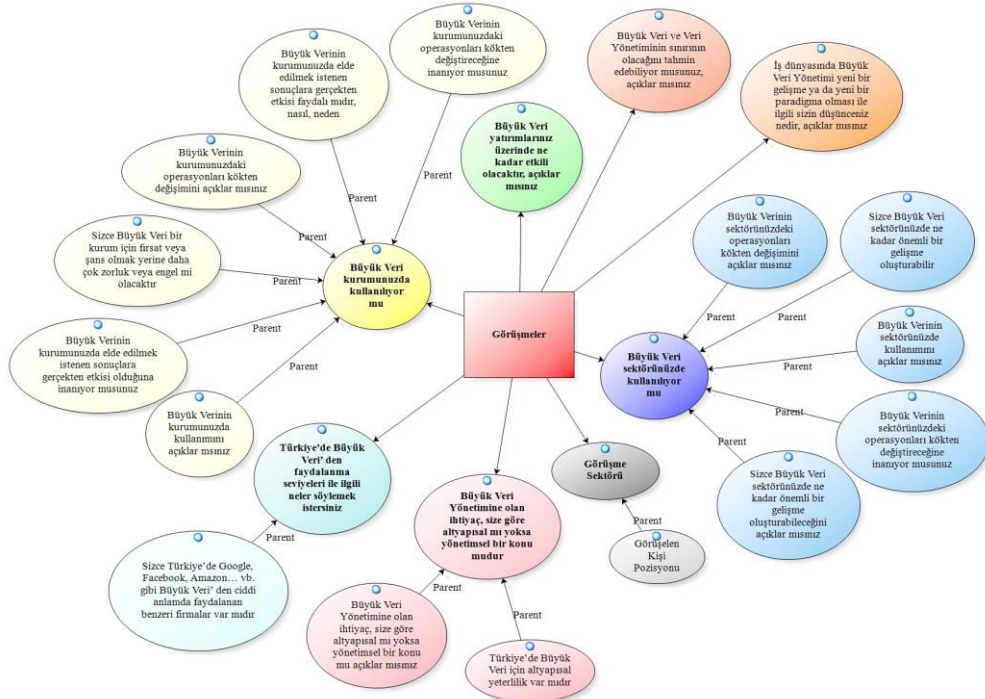
Aynı sorulara farklı kişilerin verdiği cevaplar ya da kişilerin tüm sorulara verdiği cevapları istediğimiz biçimde görme imkânını sağlamaktadır. Yine elimizdeki veriler üzerinde özel olarak ilgilendiğimiz herhangi bir kelimenin ya da kelime grubunun sorgulamasını yaparak, bu kelimenin geçtiği tüm paragrafları bir arada görme ve değerlendirmeyi daha sağlıklı bir şekilde yapabilmekteyiz.

Kodlamaların yapılmasıyla çalışma için “Model” ortaya konulmuş olmaktadır. Bu modeli NVivo yardımıyla aşağıdaki şekillerde görebilmekteyiz.



Şekil 4.5: Analizin genel çerçeve modellemesi.

(Kaynak: NVivo çıktısı, 2015)



Şekil 4.6: Analizin detay modellemesi

(Kaynak: NVivo çıktısı, 2015)

NVivo yaptığımız görüşmelerde öne çıkan kelimelerin kaynaklarını ve miktarlarını da raporlayabilmektedir. Herhangi bir şekilde sorgulama (query) yapabilmek olanağı vermektedir. Böylece en çok geçen kelimelerin kaynaklarına giderek analizimizde odaklanmayı sağlayabilmemize yardımcı olmaktadır.

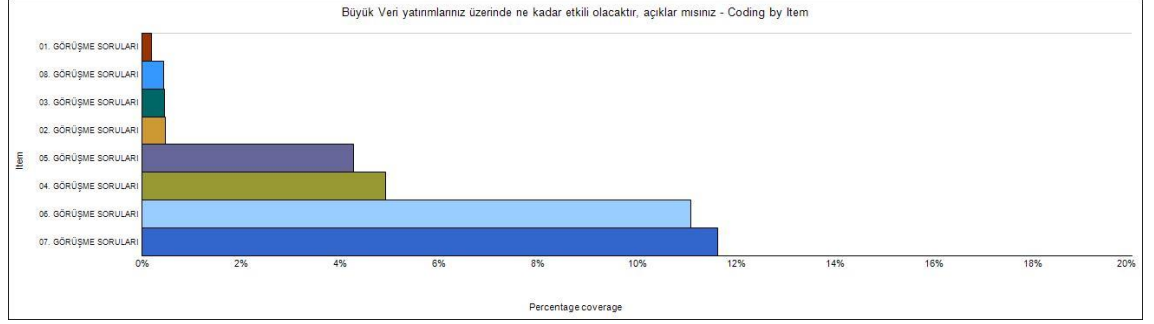
Yaptığımız çalışmada en çok geçen kelimelerin NVivo görsel hali şu şekilde olmuştur.



Şekil 4.7: NVivo kelime sorgulamaları.

(Kaynak: NVivo çıktısı, 2015)

İlaveten, NVivo ile hangi kaynakta, hangi Nodes ile ilgili daha fazla bilgi ve değerlendirme yapılmış olduğunu da göstermekte ve incelemelerimize kaynaklara yönelme açısından destek sağlamaktadır. Örneğin Şekil 4.8'deki Nodes'lardan biri ile ilgili grafik şu şekilde verilmiştir.



Şekil 4.8: NVivo'da grafik desteği örnek gösterimi.

(Kaynak: NVivo çıktısı, 2015)

Bu düzenlemeler yapıldıktan sonra yapılan kodlamalar incelenerek sonuçlar, analizler ve yorumlar yapılarak değerlendirilmeye geçilmiştir.

Bu analizleri yaparken şu hususlar göz ardı edilmemiştir:

Programdaki kodlama işlevini kullanabilmek ve bu sürecin teknik kısmına karşılık gelen adımları eksiksiz yerine getirmek bizim geçerli, tutarlı ve teorik olarak anlamlı kategoriler oluşturduğumuzun garantisi değildir. Başka bir ifadeyle, bir bilgisayar programı kullanıyor olmak tek başına analizimizi güçlendiren bir unsur değildir.

İyi bir analiz, eldeki verilerin niteliği ve kapsamına da bağlı olarak, bilgisayar programının kullanımı ile daha da güçlendirilebilir. Ancak, gerçekte analizimizi güçlü kılan şey, bir bilgisayar programının kullanılması değil, metodolojik duruşumuzun netliği ile bunu inşa ve ifşa etme becerimizdir. Yani, nasıl ki nicel veri analizi programları iyi bir analizi garantilemez, aynı biçimde nitel veri analizi programlarının kullanımı da tek başına iyi bir analizin garantisi değildir. Bu nedenle araştırmacılar, metodolojik bilgilerini güçlendirmeli ve metodolojik refleksivite – neden ve sonucun karşılıklı birbirini etkilemesi – geliştirmelidirler (Kuş, 2006).

Tablo 4.2 (devam): Görüşme analiz tablosu.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| <p>7. Sizce Büyük Veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir? Açıklar mısınız?</p> | <p>A. Çok önemli B. Önemli C. Kısmen D. Önemsiz</p> | B | C | C | B | B | C | A | C | <p>A. 12,5 B. 37,5 C. 50 D. 0</p> |
| <p>8. Büyük Verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz? Nasıl daha faydalı olabilir? Neden daha faydalı olabilir?</p> | <p>A. Evet B. Hayır C. Emin değil</p> | A | C | B | A | A | C | A | B | <p>A. 50 B. 25 C. 25</p> |
| <p>9. Büyük Veri Yönetimine olan ihtiyaç, size göre altyapısal mı yoksa yönetimsel bir konu mudur? Açıklar mısınız?</p> | <p>A. Altyapısal B. Yönetimsel C. Her ikisi</p> | C | C | C | C | C | C | C | C | <p>A. 0 B. 0 C. 100</p> |
| <p>10. Türkiye’de Büyük Veri için altyapısal yeterlilik var mıdır?</p> | <p>A. Var B. Yok C. Kısmen var</p> | C | B | C | B | B | B | C | C | <p>A. 0 B. 50 C. 50</p> |
| <p>11. Büyük Veri ve Veri Yönetiminin sınırının olacağını tahmin edebiliyor musunuz? Açıklar mısınız?</p> | <p>A. Evet B. Hayır</p> | B | B | A | B | B | B | B | B | <p>A. 12,5 B. 87,5</p> |

Tablo 4.2 (devam): Görüşme analiz tablosu.

| | | | | | | | | | | |
|--|--|---|---|---|---|---|---|---|---|---|
| <p>12. İş dünyasında Büyük Veri Yönetimi yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir? Açıklar mısınız?</p> | <p>A. Evet, Yeni bir paradigma veya olgu</p> <p>B. Hayır, Verinin genişleyen devamı</p> | A | A | A | A | A | B | B | B | <p>A. 62,5</p> <p>B. 37,5</p> |
| <p>13. Sizce Büyük Veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır? Açıklar mısınız?</p> | <p>A. Fırsat-Şans</p> <p>B. Zorluk-Engel</p> <p>C. Her ikisi</p> | C | A | C | C | A | A | A | A | <p>A. 62,5</p> <p>B. 0</p> <p>C. 37,5</p> |
| <p>14. Türkiye’de Büyük Veri’ den faydalanma seviyeleri ile ilgili neler söylemek istersiniz?</p> | <p>A. Yeterli</p> <p>B. Değil</p> <p>C. Kısmen</p> | B | B | C | B | B | C | B | B | <p>A. 0</p> <p>B. 75</p> <p>C. 25</p> |
| <p>15. Sizce Türkiye’de Google, Facebook, Amazon vb. gibi Büyük Veri’ den ciddi anlamda faydalanan benzeri firmalar var mıdır?</p> | <p>A. Evet, var</p> <p>B. Hayır, yok</p> <p>C.Önümüzdeki 10 yıl içinde öngörüyorum</p> <p>D.Önümüzdeki 20 yıl içinde öngörüyorum</p> | A | C | B | B | C | D | C | C | <p>A. 12,5</p> <p>B. 25</p> <p>C. 50</p> <p>D. 12,5</p> |

Tablo 4.2 (devam): Görüşme analiz tablosu.

| | | | | | | | | | | |
|--|---------------|---|---|---|---|---|---|---|---|---------|
| 16. Nihai olarak, yaptığınız açıklamalar ve değerlendirmeler doğrultusunda, Büyük Veri yatırımlarınız üzerinde ne kadar etkili olacaktır? Bu etkilerin nedenlerini açıklayabilir misiniz? Eğer yanıtınız olumlu ise, Büyük Verinin yatırımlar üzerinde ne tür etkileri olacaktır? Açıklar mısınız? | A. Çok etkili | B | C | C | A | B | D | C | C | A. 12,5 |
| | B. Etkili | | | | | | | | | B. 25 |
| | C. Kısmen | | | | | | | | | C. 50 |
| | D. Etkisiz | | | | | | | | | D. 12,5 |

Görüşme tablosundan da ilk bakışta anlaşılacağı üzere, Büyük Veriye uzmanlar tarafından sınırının olmadığı düşünülen yeni bir paradigma gözüyle bakılmakta ve gerçek anlamda fayda sağlanabilmesi için kurumların hem altyapısal hem yönetsel bir konu olarak ele alması ve bu şekilde yaklaşım geliştirmesi gerekmekte olduğu görülmektedir.

Teknolojik gelişmelere ve yazılım gelişmelerine doğrudan bağlı bir olgu olan Büyük Veri Yönetiminin Türkiye’deki durumunun – bu altyapıların kısmen yeterli olmasından dolayı – ağırlıklı olarak önümüzdeki yıllarda gelişimi öngörülmüştür.

Genel olarak; Büyük Veri Yönetimi kurtarıcı bir konsept olarak görünmemesine rağmen, kurumların planlamaları ve stratejik yatırımları açısından artık “kaçınılması zor bir olgu” olarak karşımızda durmaktadır.

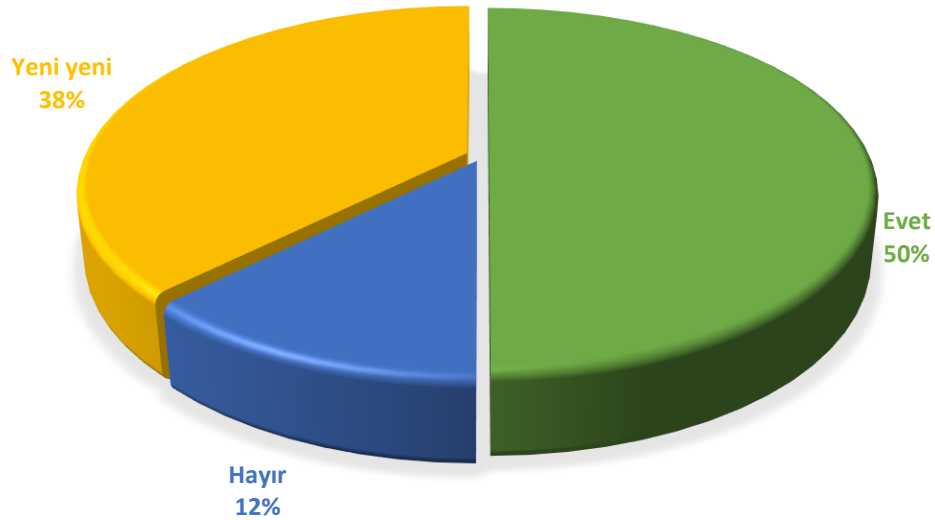
Yapılan görüşmeler, analizleri ve literatür çalışması ile elde edilen veriler “Tartışma ve Sonuç” bölümünde detaylı bir şekilde incelenerek konu ile ilgili değerlendirilmeler ortaya konulmuştur.

5. TARTIŞMA VE SONUÇ

5.1.BÜYÜK VERİNİN KULLANIMI

Genel olarak, görüşmeler kurumların hem dâhili hem harici odaklı kavramlarda “Büyük Veri” kullandığını ya da kullanılmaya çalışıldığını gösterdi. Şekil 5.1’deki grafikten de anlaşılacağı üzere %50 oranında Büyük Veri kullanımına yeni yeni başlayanlar da eklendiğinde ortalama %78 oranından Büyük Veri ile ilgilenme söz konusudur.

Örneğin, bir kurumun Büyük Veri Teknik Takım Lideri olan bir kişi sorulara şu şekilde yanıt verdi; “Artan log miktarıyla beraber sektör daha hesaplı bir şekilde bu data’ları incelemek istiyor, bu sebeple firmalar Big Data teknolojilerini kullanmayı düşünmeye başladılar.”; diğer bir deyişle, şirketler artık Büyük Veri konusunu değerlendirmeyi arzulamaktadırlar. Bununla birlikte, bir otomotiv firması yöneticisi şuna değinmiştir; “Araç alma niyeti olan müşterinin tespit edilip reklam/tanıtımın müşteriye zamanında sunumunda faydası olmaktadır”, ve bu projeler “kurumun iş modelleri üzerinde doğrudan etkiye sahip olan müşterilere ulaşmada ve onları tanımlama/karakterize etmede” hedeflendi – diğer tabirle, harice odaklanmıştır.



Şekil 5.1: Büyük Veri sektörünüzde kullanılıyor mu?

(Kaynak: Analiz çıktısı, 2015)

Birçok katılımcının cevapları gösterdi ki, kurumlar Suther'in "çok boyutluluk" olarak tanımladığı, hem kurumun var olan müşteri tabanı ve ayrıca hem de yeni müşterilere ulaşmanın anlaşılmasını güçlendirmek için Büyük Veriyi kullanıyordu.

Kurumlar ayrıca rakipler karşısında rekabet gücü kazanmak için bir çaba içerisinde Büyük Veri girişimlerini kullandı. Bir yönetici "Rekabet artacaktır." dedi ve şu konuya değindi, yapılandırılmamış veri aynı zamanda "Kurumun rakip firmalarının müşterilerini kazanmak ve kaybetme ihtimali yüksek olan müşterileri yeniden kazanmak" amacıyla kullanılıyordu. Sınıflandırılmamış sosyal medya verisi rakip firmaların müşteri profiline ve firmaların eylemlerini analiz etmek için kullanılabilir. Örneğin, Facebook üzerindeki rakip kurumun sayfasını beğenen insanların profil ve sayısı sınıflandırılmamış veridir ve veri analizleri bu müşterileri – hatta onlar Kurumun kendi müşterisi olmasa bile – derinlemesine anlamak için kullanılabilir.

Bazı katılımcılar sosyal medya sitelerinden veri toplama [Hurtwitz ve diğerlerinin "insanlar tarafından oluşturulan yapılandırılmış veri olarak tanımladığı"] kurumların farklı yönlerdeki operasyonlarında da önemli bir rol oynayabilir şeklinde düşünmektedirler. Bir yönetici bu durumu şöyle açıkladı: "Kurumu arama ve şikâyetlerini seslendirme yerine, müşteriler doğrudan doğruya sosyal medya siteleri üzerinde onlardan (Kuruma ya da birbirilerine) bahsedeler ve böylece şikâyetlerini ve önerilerini daha etkili bir şekilde yapmayı denerler. Kurum bu şikâyetleri ayrıştırır ve önemlerine derecelerine göre sıralar, dolayısıyla müşteri şikâyetlerini tam olarak saptamaktadır." Bu, tam olarak aynı olmamasına rağmen, Hurwitz ve arkadaşlarının daha fazla deneyimi olan fiziksel şubeleri tam olarak tespit etmek ve şikâyetleri daha verimli bir şekilde analiz etmek için analiz verisi kullanmaya dayalı bir araç kiralama şirketine verdikleri örnek gibidir.

Ek olarak, bir katılımcı şuna işaret etmiştir: Kurum hakkındaki yapılandırılmış büyük veri formların müşteri tarafından üretilen yaratıcı şekildeki analizlerine – doğrudan adreslenmiş olmasa bile – önem veriyor. Örneğin, herhangi bir kullanıcı açıkça Kurumdan bahseden herhangi bir Tweet'i görebilir ve Kurumdan bahseden kullanıcı tarafından üretilen diğer formlar mesela internet forum postları da halka açık olabilir. Bu mesajlar derlenebilir, analiz edilebilir ve günlük olarak kurumların müşteri hizmetinin gelişmesinde kullanılabilir.

Suther ve diğeri şü şekilde ifade etmiştir: Firma tanıtımının büyük bir yüzdesi istedikleri hedef kitleye ulaşmaz ya da sıklıkla aynı yolla benzer reklamın yanlış yönlendirmesi oluşur. Kurumun hizmet ve ürünlerinin pazarlama ve dağıtımında Büyük Veri ve analizlerinin nasıl kullanıldığı sorulduğunda, bir yönetici şü şekilde yanıtladı: Bir sektörde Büyük Verinin kullanımı ile tüm sektörlerdeki kullanımı tam olarak aynı değildir.

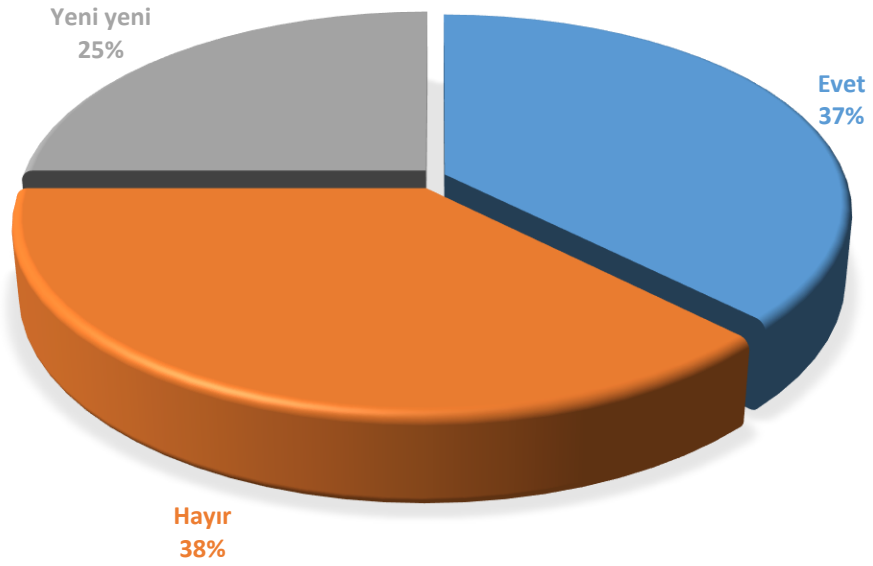
Diğeri bir katılımcı da “Büyük Verinin imkânları hakkında heyecanlıyız fakat aynı zamanda bu fırsatlar gelen zorluklarında üzerine eğilmek zorundayız” dedi.

Katılımcılar ayrıca birçok yönden bu “zorlukları” açıkladı. Bir teknik müdür şü şekilde ifade etmiştir “Türkiye’de Bulut sistemlerin kurulması mümkün olsa da bu sistemlerin yurtdışı ile fiyatta rekabet etmesi pek mümkün görünmemektedir. 5651 yasaı ile verilerin Türkiye dışına çıkarılması da mümkün olmadığı için Bulut üzerinde Büyük Veri yapabilmek olanaksız hale gelmektedir. Bu sebeple geçici süreli ve esnek kabiliyetli sunucu kümeleri kurulamamaktadır. Statik, eski sistem ile oluşturulan sunucu kümelerinin barındırmadan tutun elektriğeye kadar birçok masrafı olmaktadır.”

Gizlilik yasaları ve haklar ile ilgili olarak bir katılımcı “Büyük Verinin yasalar ve hükümet tarafından desteklenmesi ek olarak Türkiye’deki firmaların çok ciddi yatırımlar yapması lazım ki ciddi anlamda faydalanan firmalar olsun. Ama azımsanmayacak miktarda faydalanan firmalar var ve daha da artacaktır. Kişisel bilgilerin güvenliği konusunu da kapsayan bu mesele çok ortada kalmıştır” şeklinde ifade etti.

Büyük Veri ayrıca şirketin bazı limitli (fakat yaygın değil) iç yeniden yapılanmasında neticeleniyordu: Bir katılımcı “Veriye sahip olan zaten yönetmektedir. Bunların hiçbirini ciddi anlamda veri biriktirme ile çalışmamakta, kayıtlarını bir yerlerde tutup bunları bir sonraki nesle aktarmamaktadır. Beynimizde gerçekleştirilen tüm faaliyetler Büyük verinin yorumlanması olarak düşünüldüğünde asıl amaç olan deneyimlerin ya da deneyimlemenin bireye, şirkete, topluma kattığı değer birikmediği takdirde ilgili veri kişi ile yaşamakta ve ölmektedir. Bir sonraki birey teslim alamadığı veriyi yorumlama hususunda hata yapmakta bu da zarara yol açmaktadır. Önümüzdeki dönemde günümüz konuları arasında yer alan Kalite Yönetimi, Kurumsal Hafıza, Ortak Akıl, Kurumsallık gibi kavramların sübjektif kriterlerinin Büyük Veri ile yorumlanabilir verilere

dönüştürülmesinin önemini anlayacağını düşünüyorum” dedi. Konu biraz daha derinlemesine irdelendikten sonra kendisinden “kurumlarda Büyük Veri ile ilgili kişilerin yetiştirilmesi, yöneticilerin bu konulara hâkim olması ve yönetebilir hale gelmesi, bilginin diğer personele aktarılabilmesi” hususunun çok önemli olduğu fikri edinildi.



Şekil 5.2: Büyük Veri kurumunuzda kullanılıyor mu?

(Kaynak: Analiz çıktısı, 2015)

Genel olarak, katılımcıların cevapları gösterdi ki, kurumlar istenilen seviyede yüksek oranda ve etkin olmasa da çeşitli şekillerde Büyük Veriyi kullanıyordu.

5.2. BÜYÜK VERİ HAKKINDAKİ GÖRÜŞLER VE TAHMİNLER

Daha önce belirtildiği gibi, olgusal yaklaşım araştırmacıya ayrıca, bir çalışma içerisinde karışan katılımcıların subjektif deneyimlerini ve duygularını saptamaya izin verir. Bundan dolayı Büyük Veri üzerine, uzmanlar ve yöneticilerin düşünceleri, önyargıları ve genel görüşleriyle ilgili bazı sorular da sorduk.

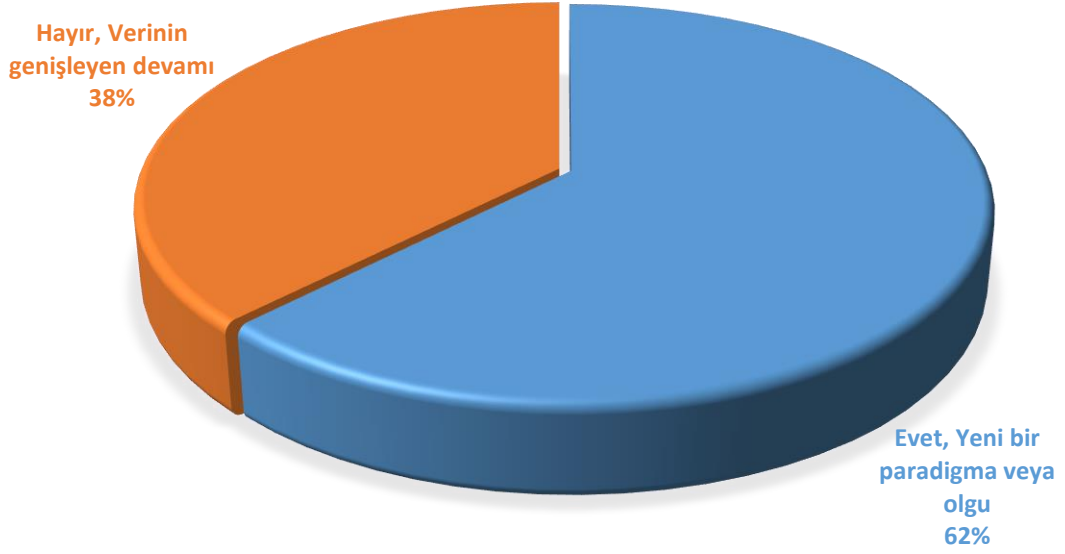
Genel olarak, katılımcılar Büyük Veriyle ilgili genel bir pozitif fakat gerçekçi görüşe sahipti. Onlar Büyük Veriyi “her derde deva bir ilaç” gibi görmedi ve birkaçı Seigel’s yankıları ya da Kent ve diğerlerinin iddiaları olan Büyük Verinin bir anlayışını ifade ettiler.

Sekiz katılımcından 5’i Büyük Verinin bir şekilde özellikle “yeni bir paradigma veya vaka” olduğunu ifade ederken, 3 katılımcı tamamen yeni bir paradigma olmasından ziyade, basit bir şekilde, ne olursa olsun durmadan büyüyen veri miktarının genişlemesi olarak ifade etmiştir.

Bununla birlikte, hiçbir katılımcı kurumlar için herhangi bir şekilde “zararlı” ya da “istenmeyen” şeklinde tarif etmemiştir; derinlemesine görüşlerini almak yerine, potansiyel faydalar ile potansiyel zorlukları dengelemek için yeni yaratıcı düşünce şekilleri gerekiyordu.

Her halükarda, hiçbir katılımcı (Büyük Veriyi özellikle farklı bir olgu olarak görmeyen birisi bile) “geçici” bir gelişme olarak nitelemedi ve genel olarak hepsi şunu kabul etmektedirler; Büyük Veri “kalmak için burada” idi ve bundan dolayı kurumların uzun dönemli odağını ve konsantrasyonunu gerektirdiğini düşünmektedirler.

Katılımcılardan bir diğeri olan bir telekomünikasyon firmasının Big Data Takım Lideri yerel teknik durumla ilgili olarak da şunu ilave etti: Büyük veri sektöre sadece yeni soluk getirecektir. Yapısı itibariyle klasik veri yönetimi sistemleriyle yer değiştirebilecek teknoloji değildir. Daha ziyade beraber kullanılması daha doğru olacaktır. Operasyonlar hemen hemen aynı kalmakla beraber elde edilen işletme değerleri artacaktır, çünkü eskiden yapılamayan raporlama ve veri işleme süreçleri çok daha kolay bir şekilde yapılabilecektir. Özellikle Türkiye’de firmaların onlarca-yüzlerce sunucu kullanarak büyük veri işleme gibi girişimleri bulunmamaktadır. Hâlbuki büyük veri gücünü paralel işleyebilme özelliğinden almaktadır.



Şekil 5.3: İş dünyasında Büyük Veri Yönetimi yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir?

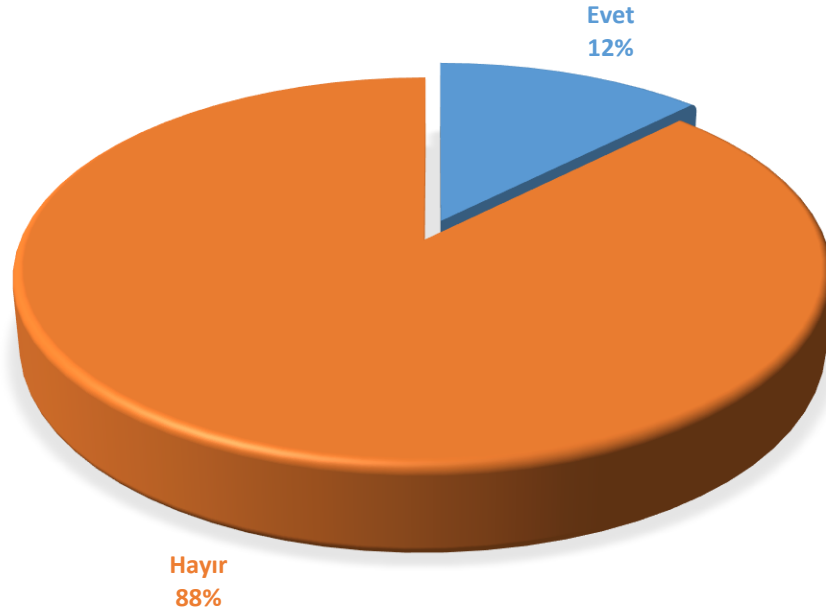
(Kaynak: Analiz çıktısı, 2015)

Şu yanıtlar da önemliydi, katılımcılardan 3’ü Büyük Veri bu sektörde “devrim yapamaz” (belki de onlar onu “yeni” bir paradigma olarak görmediği için ya da altyapısal-yönetimsel sorunlar nedeniyle), diğer 3’ü olabilir dedi ve kalan 2’si emin olamadı.

Bir katılımcı, o bir paradigmadır fakat, “Yeni bir paradigma olmasına rağmen ne yazık ki iş dünyası bu değişimi henüz anlayabilmiş değildir. Hala eski geleneksel kıstaslara göre değerlendirmeler yapılmaktadır” şeklinde ifade etmiştir.

Başka bir katılımcı ise buna karşıt olarak şunu ifade etmiştir: Şirketler kendi resmi ya da gayri resmi raporlama ihtiyaçları için tutmuş oldukları verilerin zamanla genişleyen ihtiyaçlara yeterli gelmediğini, büyüyen işletme için gereken faaliyetlerinin yönetilmesine yönelik ihtiyacın ortaya çıkması ile yeni kayıtlar tutmanın gerekliliğini anladıkları anda bu amaca yönelik yapmış oldukları tüm girişimler, firmaların veri tutma ve analiz etme kabiliyetlerinin gelişmesine sebep olur. Durumdan da görüldüğü üzere her

kararın bir sebebi, her sebebin belirli bir dönemlik farkındalığı dolayısı ile her geliřtirmenin bir ihtiyacın ortaya ıkması ile yapıldığı görölmektedir. Büyük Veri bu örnekte de olduđu gibi bir ihtiyacın giderilmesine yönelik verinin bulunmayışı sonucu iradi olarak geniřletilen veri bütününden oluşur ve geniřletildikçe Büyük Veriyi yöneten kişiye yeni karar alanları ve yöntemleri hakkında farklı ufuklar açan bir araçtır”.



Şekil 5.4: Büyük Veri ve Veri Yönetiminin sınırının olacağını tahmin edebiliyor musunuz?

(Kaynak: Analiz ıktısı, 2015)

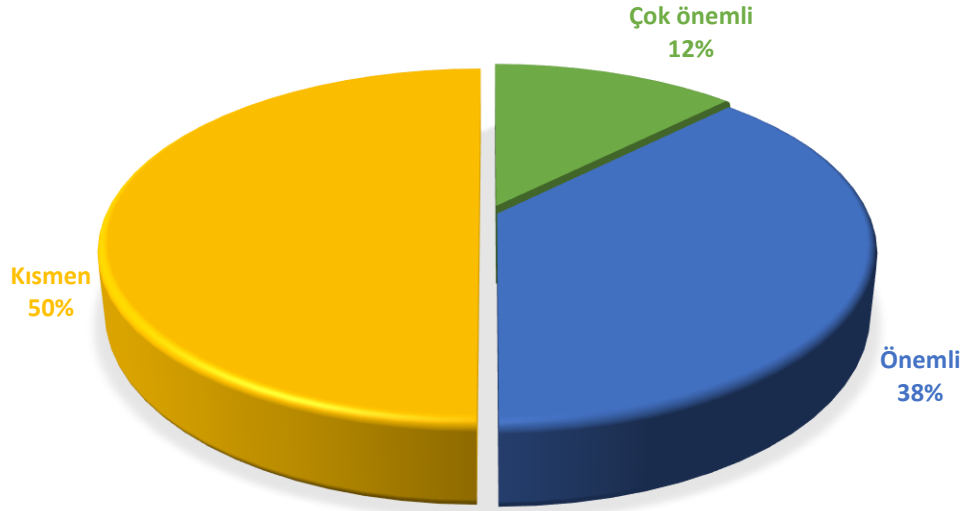
Büyük Verinin kurumlar için “gelecekteki büyük şey” olup olmadığı sorulduğunda, katılımcıların çeřitli cevapları vardı. Katılımcıların çođu Büyük Veri için “gelecek sınırın” ne olabileceğini kesin olarak öngöremedi.

Bir yönetici řuna deđindi, tüm řirketler aynı řekilde – özellikle bir grup yerine bizim müşterilerimize ürünlerimizi uygun hale getirmek yoluyla – Büyük Veriden faydalanmayı umar. Bundan dolayı, [Büyük Veri ile] pazarın geleceğinde ne olacağını tam olarak öngörmek çok zordur.”

Başka bir katılımcı ise “Büyük veri sürekli gelişen sistemler tasarlanmasına yol açar. Tasarlanarak değişen sistem üzerinden de veriler birikmeye devam eder. Bu da yeni yöntem ve sistemlere sebep olacağından sürekli yaşayan, yaşadıkça gelişen, geliştikçe değişerek büyüyen bir sistem ve bu sistemin kararlarını oluşturan etkenlerin içinde bulunulan değişken duruma göre bağlı fonksiyonların izlendiği sürecin çıktısı sadece değişim, gelişim ve yeni kararlar ile bir sonraki dönemdeki faaliyetlere karar vermeye yarayan yeni verilerin Büyük Veriye eklenmesi olacaktır. Kanımca tek sınır kişisel hak, hürriyet ve kanunla belirlenmiş kurallar olabilir” şeklinde bir yaklaşımda bulundu.

Bu görüşmecinin veri ile birlikte yönetsel pozisyonundan ve bu ifadelerinden dolayı Büyük Verinin, yönetim anlayışına etki etmesi beklentisi zihinlerde bulunmaktadır.

Kurumlarda elde edilmek istenen sonuçlara etkisi ve neden/nasıl sorusunu ise çoğu katılımcı ihtiyatlı bir şekilde yanıtladı. Bir katılımcı, “Evet, faydalı olabilir, Neden sorusunun cevabını vererek başlıyorum şöyle ki: Şahısların uzun yıllar süren tecrübelerinden oluşan kararların herhangi bir psikolojik duruma bağımsız olarak mevcut verilerin yorumlanması sonrası daha az risk ve insan faktörünün payı minimumda olacak şekilde alınması sağlanabilecektir. Evrensel Büyük verinin bir parçası olan sizin tuttuğunuz büyük veri elbette tüm soruların cevaplarını vermeyecektir. Bu konuda mevcut durumunuzu yorumlamaya yönelik eksik büyük verinin diğer Büyük Veri kullanıcılarından temini ve/veya Büyük Veri paylaşımı durumunda çok daha tutarlı, entegre ve yönetilebilir faaliyetler tasarlanarak tüm girdileri ve çıktıları belli olan bir matematiksel model yerine belirli koşullara bağlı olarak değişen, sizin yerinize düşünen, karar vermede sizlere seçeneklerle yardımcı olan bir Stokastik Model tasarımı yapılabilir. Nasıl sorusuna vereceğim yanıt ise basit bir şekilde her alınan karar veya riskin, yapılacak kar odaklı çalışmaların öznel sebeplerinin bulunduğu matematiksel dayanağı olacağı şeklindedir” dedi.



Şekil 5.5: Sizce Büyük Veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir?

(Kaynak: Analiz çıktısı, 2015)

Katılımcılardan elden edilen genel düşünce şuydu, Büyük Veri girişimlerimizin kritik bir parçası olmaya devam edecek, fakat kendimizde fazla abartmadan ve gerçek bir fikir çerçevesini korumalıyız. Sadece pazardaki bütün diğer önemli gelişmeler gibi davranmalıyız. Şirketler için son derece önemlidir, çünkü “Şirket için yapılandırılmış veri kullanımı artık ihtiyaç duydukları avantajı sağlamada yeterli değildir. Daha hızlı bir şekilde internet ortamındaki yapılandırılmamış veriyi toplama ve işlemek için, Büyük Veriyi anlamak ve veri yönetimindeki en son gelişmelerle ilgili güncel kalmak kurumlar için son derece önemlidir.”

Yine bir katılımcı Büyük Veri fırsat-şans mı yoksa zorluk-engel mi sorusuna, Büyük Veri kurumlar için bir fırsat-şans olarak karşımıza çıkar diyerek şu şekilde yorum getirmiştir; “Doğru yorum ve değerlendirme sonucu iyi tasarlanmış ve işletilen kararlar doğru yönetildiği takdirde başarıya ulaşacaktır. Bu sürecin başarısı süreç sahibinin genel kabiliyeti ve iradesi ile doğru orantılıdır. Büyük Veri ilgili süreç sahibine yukarıda saymış olduğum; yorum, değerlendirme, tasarım, işletim, kararın niteliği ve bağlı unsurları, yönetim fonksiyonları gibi daha birçok başarı odaklı çalışmalar süresince gerekli ya da gerekebilecek nicel veriyi nitel fonksiyonlar halinde önüne serecektir. Burada en önemli

unsur yorumlanacak büyük verinin yorumlanması için gerekli bilginin doğru kaynaklardan doğru şekilde elde edilmesi olacaktır. Bu konudaki başarısızlık; gelişimdeki zorluk-engel ve sonrasında zarar olarak karşımıza çıkacaktır” şeklinde yanıt verdi.

5.3. SONUÇ

Literatür araştırmasını ortaya koyduğumuz Genel Kısımlar bölümünde 2.2. Büyük Verinin Gelişimi ve Veri Yönetimi maddesi “Büyük Veri mevcut ve geleceğin araştırma sınırlarından birisi oldu. Gartner “Top 10 Strategic Technology Trends For 2013” ve “Top 10 Critical Tech Trends For The Next Five Years” listeleri yayınladı ve Büyük Veri her ikisine de girdi.” ifadesi ile başlamıştır. Bu Büyük Veri konusunun uluslararası alanda dikkatle gözlemlendiği ve önem ifade ettiğini göstermektedir. Genel Kısımlarda ve özellikle 2.2 ve 2.3 maddelerinde elde ettiğimiz literatür verilerinin, yaptığımız görüşmeler ve analizleri neticesinde elde ettiğimiz sonuçlar ile uyumlu olduğu görülmektedir ki, aynı önem ve dikkat yapılan görüşmelerde de ortaya konulmuş ve gözlemlenmiştir.

Büyük Veri için teknik altyapısal gereksinimlerin önemi, yetkin personel konusunun hassas bir önemi olduğu görülmüştür. Yönetimlerin konuyu benimseyiş biçimleri, Büyük Verinin bazı fırsatlar getirmesinin yanında zorlukları da beraberinde getirmesi gibi konularda literatür verileri ile analiz verilerimiz ciddi anlamda örtüşmektedir.

Sonuçlar gösterdi ki, veri miktarı çok hızlı bir şekilde artmasına rağmen Büyük Veri daha yavaş bir şekilde gelişmektedir, fakat onun varlığı çeşitli farklı sektörlerde ve çeşitli şekillerde operasyonları etkilenen kurumların departmanlarında hissedilmektedir. Kurumların yöneticileri Büyük Verinin potansiyel kullanımlarının farkındalar ve bir bütün olarak “Büyük Veri” kavramını kullanma niyetindedirler, sadece sektördeki hedefleri için değil, rakiplerine karşı bir avantaj kazanmak için bir girişim içindedirler.

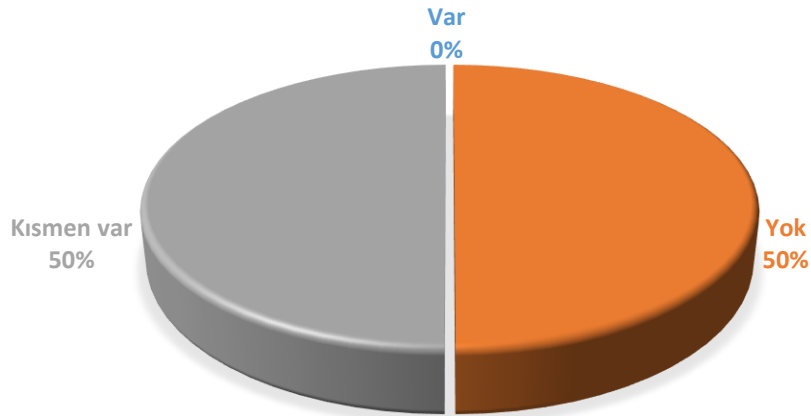
Büyük Verinin kurumlar üzerindeki en büyük etkisi, ağırlıklı olarak, müşteri analizi ve ürün geliştirme konularında hissedilmektedir. Artan veri miktarı müşterilerin artık değişen çeşitli gruplarının azalması demektir. Bunun yerine, bir müşterinin bilgisi, örneğin önceki satın almaları ve sosyal medya beğenileri kişiselleştirilmiş bir profil oluşturmak için korelasyon olabilir. Elbette, kurumlar her bir müşteri için kişiselleştirilmiş kampanyalar ve teklifler oluşturamayabilir. Ancak, daha fazla bireysel

düzye reklamını belirlemenin yanı sıra, yeni ürün geliřtirmede ve kişiselleřtirilmiř kampanyalar oluřturmada kullanabilirler. Bu durum Büyük Veri analizinin yatırımlar üzerindeki olumlu etkisi ve fırsat olarak deęerlendirilmektedir.

Bunların yansıra katılımcıların řu noktayı da gözden kaçırmadıkları görülmüřtür. Büyük Verinin gerçek görünüřü, yani onun potansiyelleri kadar zorluklarının da olduęunu fark etmiřlerdir. Onlar Büyük Veriyi ne sektörlerinin “kurtarıcısı” olarak görmekte, ne de basit bir řekilde ilgilenilmesi gereken bir sorun ya da bir engel olarak görmektedirler. Bu da bize göstermektedir ki, basındaki abartıların çoęundan ayrı olarak, yöneticiler aslında iřleriyle ilgili olarak günlük bazda Büyük Veri ile ilgilenmeye bařlamıřtır, onlar bu konuya gerçekçi ve pratik bir řekilde yaklařmaktadır.

Kurumlar řirketin iç yeniden yapılanmasını önemli ölçüde deęiřtirmemektedir ve bu yüzden Büyük Verinin konularıyla bařlıca olarak genelde IT departmanları ilgilenmekte ve pazarlama gibi departmanlara bilgi desteęi saęlamaktadır.

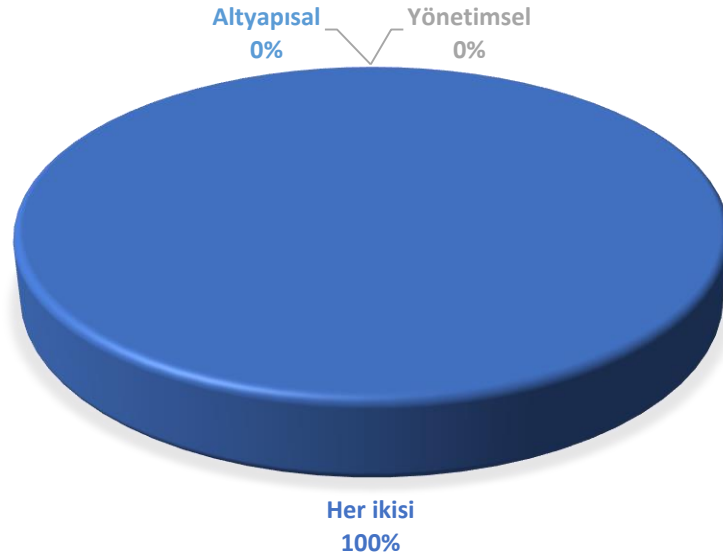
Önemli ölçüde teknik bir kaç konuya da katılımcılar tarafından deęinilmiřtir; veri depolama maliyetinin kontrolü, veri sızıntısı ve veri güvenlięi konuları endiře duyulan taraflardır. Ayrıca hukuki ve gizlilik konusu da Büyük Veri konusunun önde gelen kaygılarındanadır. Bunlar Büyük Verinin potansiyel zorlukları olarak ortaya konulmuřtur.



Şekil 5.6: Türkiye’de Büyük Veri için altyapısal yeterlilik var mıdır?

(Kaynak: Analiz çıktıısı, 2015)

Büyük Verinin yatırımlar üzerindeki etkisi sorumuzda “Kısmen” vardır şeklinde cevap veren “%50”lik grup içerisindeki bir katılımcı, “Kısmen etkili olması konunun önemsizliği değil, aksine bu kadar önemli bir konuda yapılan az çalışma yüzündendir. Önümüzdeki dönemde bahsetmiş olduğum konuların gerçekleştirilmesine yönelik yoğun bir Büyük Veri çalışmasının olmayacağını düşünüyorum. Yapılan çalışmaların da belirli bir konuya özel ve parça halinde olacağını, kendimce bütünsel bir yorum için yeterli olmayacağı kanaatindeyim. Yatırım için bir düşünce söz konusu olduğunda uzun bir süre daha Büyük Veri sonuçlarından ziyade alfa numerik veri ve değerlere ulaşmanın daha kolay olacağı, Büyük Veriden beklediğim fonksiyonların yine kişiler ve insan deneyimleri temasında gerçekleştirileceğini düşünüyorum. Kendi adıma günümüz şartlarında tutmuş olduğumuz verilerin yorumlanması konusunda kaybedilen zamanın deneyimler ile kısa sürede söylenmiş bir karara kıyasla çok daha az efektif olduğunu düşünüyorum. Günümüz koşulları birçok durumda Büyük Veri kullanımının deneyim kullanmadan daha pahalı olduğunu bana düşündürmektedir. Yapılacak çalışmaların artması ve topluma iletilen bir veri akışının söz konusu olması durumunda kullanılacak düşük maliyetli çözümler gelinceye değin Büyük Veri tarafının söz konusu olası yatırımlarımda ikinci planda kalacaktır. Teknik Yetersizlik.” şeklinde ifade etmiştir.

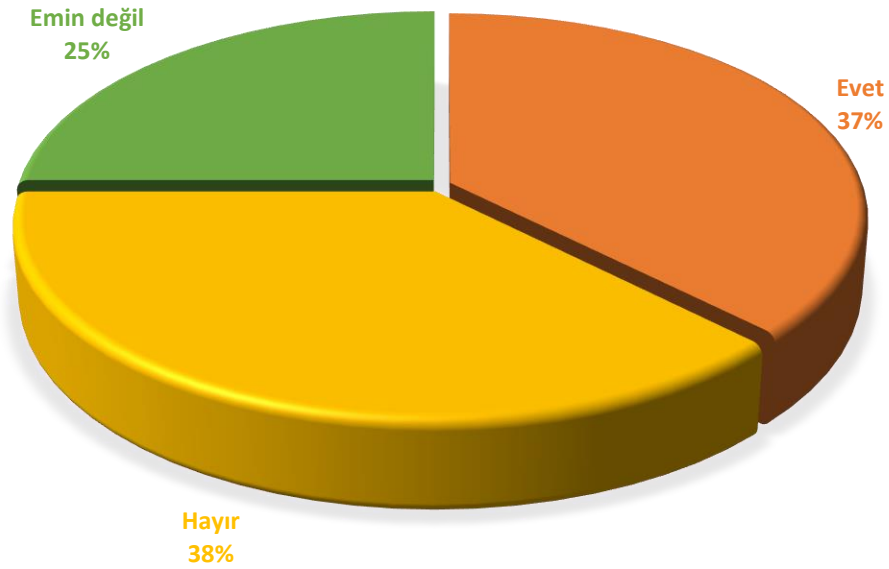


Şekil 5.7: Büyük Veri Yönetimine olan ihtiyaç, size göre altyapısal mı yoksa yönetimsel bir konu mudur?

(Kaynak: Analiz çıktısı, 2015)

Katılımcılar ile yaptığımız görüşmelerden ve aldığımız cevaplardan kapsamlı olarak şu neticeyi çıkartmaktayız:

Katılımcılar Büyük Veri nedir, neler yapabilir, zorlukları ve fırsatları nelerdir farkındadırlar. Bununla birlikte genel kısımlarda da bahsedilen verinin işlenmesi zorluğu, teknik kısıtlar, deneyimli personel eksikliği vb. gibi sebeplerle de Büyük Veri konusuna gayet temkinli ve gerçekçi yaklaşmaktalar.



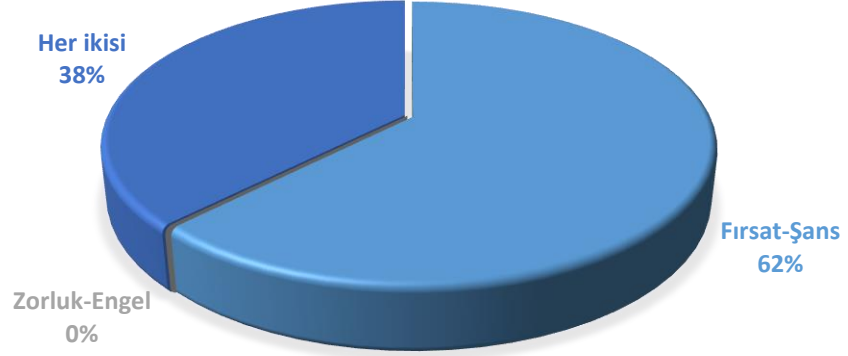
Şekil 5.8: Büyük Verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz?

(Kaynak: Analiz çıktısı, 2015)

Nihai olarak, görüşme analiz tablosunda da (Tablo 4.2) gördüğümüz sonuçlar ve karşılıklı edinimler üzerine şunları söyleyebiliriz:

Kişi ve kurumların Büyük Veri kullanımı eğiliminde olduğu, fakat altyapısal ve yönetimsel yetersizlikler nedeniyle tam olarak faydalanamadığı, kökten bir değişiklik

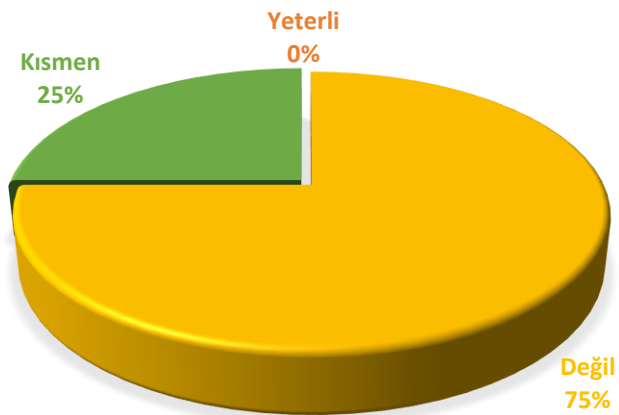
yerine “etkili ilave bir kaynak” olarak düşünüldüğü ve sınırları şu anda öngörülemeyen ve daha çok fırsat-şans olarak görülen bir paradigma olduğu sonucuna ulaşılmıştır.



Şekil 5.9: Sizce Büyük Veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır?

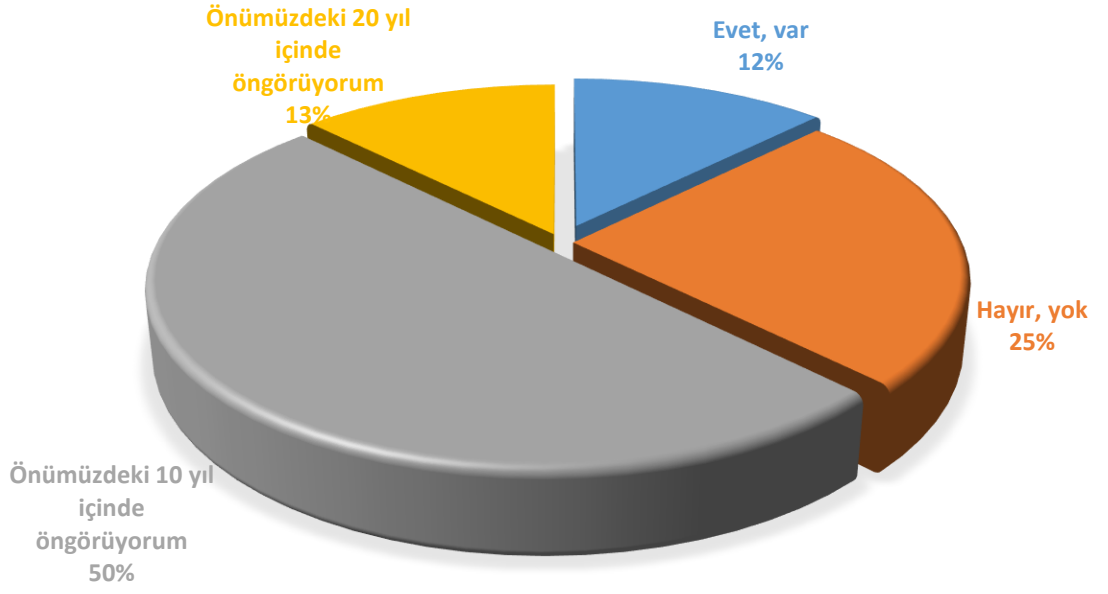
(Kaynak: Analiz çıktısı, 2015)

Bunlara ilaveten Türkiye açısından kısmi altyapı yeterliliği, faydalanma derecesinin henüz başlangıç seviyesinde olduğu ve kurumlar tarafından ağırlıklı olarak muhtemelen önümüzdeki 10 yıl içerisinde daha efektif kullanılacağı düşünülmektedir.



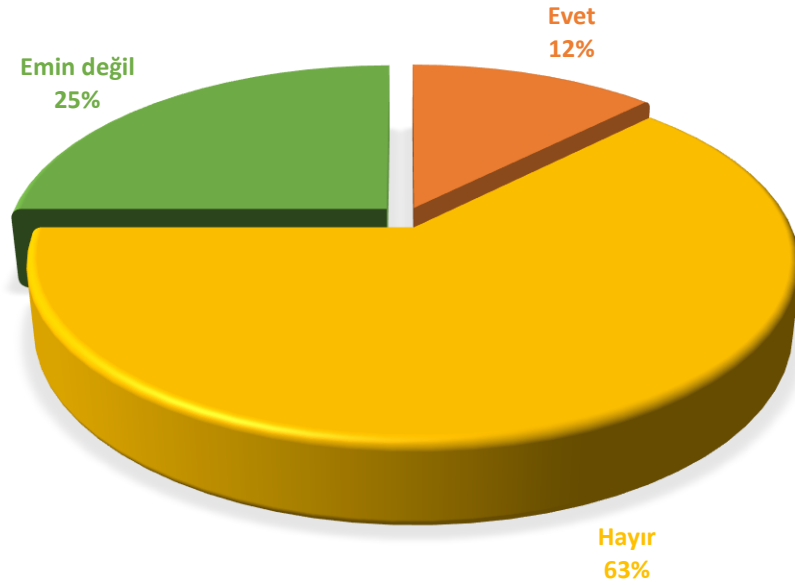
Şekil 5.10: Türkiye’de Büyük Veriden faydalanma seviyeleri ile ilgili neler söylemek istersiniz?

(Kaynak: Analiz çıktısı, 2015)



Şekil 5.11: Sizce Türkiye’de Google, Facebook, Amazon vb. gibi Büyük Veriden ciddi anlamda faydalanan benzeri firmalar var mıdır?

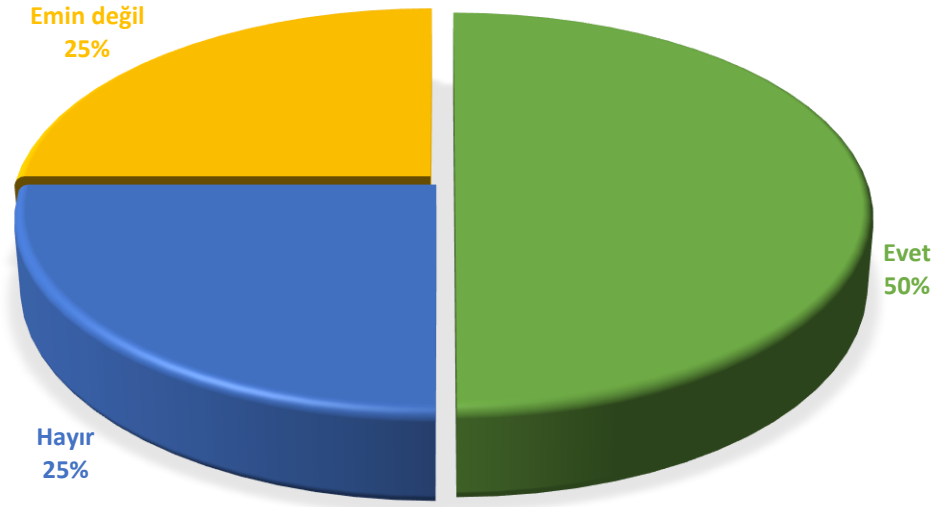
(Kaynak: Analiz çıktısı, 2015)



Şekil 5.12: Büyük Verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz?

(Kaynak: Analiz çıktısı, 2015)

Büyük Veri “her derde deva bir kurtarıcı” değildir, fakat o rekabet dünyasında, elde edilmek istenen sonuçlara ulaşmada kullanılması gereken yeni bir olgudur.

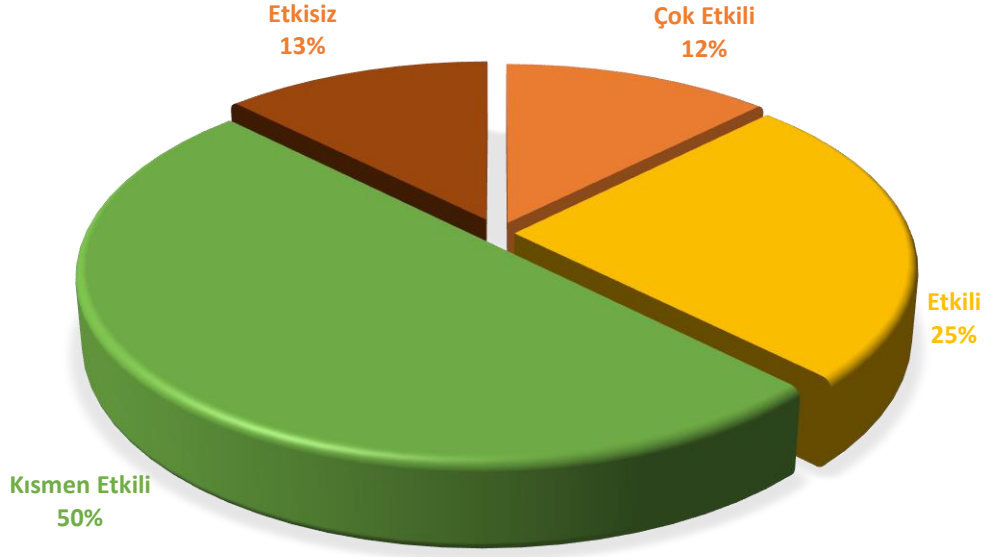


Şekil 5.13: Büyük Verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz?

(Kaynak: Analiz çıktısı, 2015)

Katılımcıların yatırımlara olan etkisine bakışı ise Büyük Veri literatür araştırmasındaki uluslararası çaptaki veri uzmanlarının düşüncelerinden çok uzak düşmeyecek şekilde gerçekçi çıkmıştır.

Katılımcılar %12,5’arlık dilimler ile iki uça çok etkili ve etkisiz olarak değerlendirmiş, %25’lik bir dilimde etkili ve %50’lik kısım düşünce ise kısmen etkili olarak tespit edilmiştir.



Şekil 5.14: Büyük Veri yatırımlarınız üzerinde ne kadar etkili olacaktır?

(Kaynak: Analiz çıktısı, 2015)

Bu değerlendirmenin Türkiye’de henüz Büyük Veri konusunun çok yeni olmasından ve yeterince fayda sağlanamadığından dolayı olduğu ifade edilebilir. Büyük Veri analizi hizmeti yapan şirketler ve de bu analizleri uygulayan kurumlar arttıkça muhtemelen bu oranlar ciddi anlamda artacaktır.

Belki de, Büyük Veri analizlerinin en gözle görülür uygulamaları iş dünyasındadır.

Tahmini olarak bir perakendeci analizin gücünü kullanarak operasyon marjını %60’a kadar arttırabilir. Yeni fırsatlar kullanılarak (örneğin yer-farkındalıklı ve yer-tabanlı hizmetler) yeni gelirler için önemli oranda başı çekmektedir.

Kapsamlı bir analitik çerçeve müşteri işlemleri, envanter yönetimi, mağaza tabanlı video verileri, reklam ve müşteri ilişkileri, müşteri özellikleri ve fikirleri, satış yönetimi altyapısı, finansal veri ve diğerlerini içeren çok modlu geniş miktarda veriler toplayan

işletmelerin tedarik zinciri yönetimi, müşteri yönetimi, satış sonrası destek, reklam vb. entegrasyonunu gerektirebilir.

Büyük perakendeciler için tüm verilerin toplanması kolay bir şekilde exabaytlar seviyesine yaklaşmaktadır. Envanter izlemek için RFID'nin kapsamlı yayılması, tedarikçi veri tabanlarına bağlanma ile, müşteri özellikleri ve profilleri ile entegrasyon (mağaza sadakat programlarıyla) ve finansal sistemlere tam entegrasyonu ile gelişmiş verimlilik potansiyeli muazzamdır ve şimdi sadece farkındalık başlamaktadır.

Ayrıca Büyük Verinin uluslararası alanda ne denli ilgiyle takip edildiği ve önem verildiğini ifade etmesi açısından şu dikkat çekici bilgiye değinmekte de fayda vardır:

https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf adresinden tamamına ulaşılabilen, Büyük Veri ile ilgili olarak ABD Başkanının talimatıyla Beyaz Saray'da 1 Mayıs 2014 tarihli "Big Data: Seizing Opportunities, Preserving Values" – Büyük Veri: Fırsatları Yakalama, Değerleri Koruma – adlı, Başkan Danışmanı John Podesta, Ticari Sekreter Penny Pritzker, Enerji Sekreteri Ernest J. Moniz, Bilim Ofisi ve Teknoloji Politikası Direktör John Holdren ve Ulusal Ekonomi Konseyi Direktör Jeffrey Zients imzasıyla detaylı bir rapor hazırlanmıştır.

Raporun giriş kısmında şu ifadelere yer verilmiştir.

"Sevgili Başkan,

Sosyal, ekonomik ve teknolojik devrimin tam ortasındayız. Nasıl iletişim kuracağımız, nasıl sosyalleşeceğimiz, nasıl zaman harcayacağımız ve işimizi nasıl yöneteceğimiz internet üzerine taşındı. İnternet telefonlarımıza, evlerimizde ve şehirlerimizde cihazlara ve endüstriyel ekonominin gücü olan fabrikalara taşındı. Veri ve keşif patlamasının sonucu dünyamızı değiştiriyor.

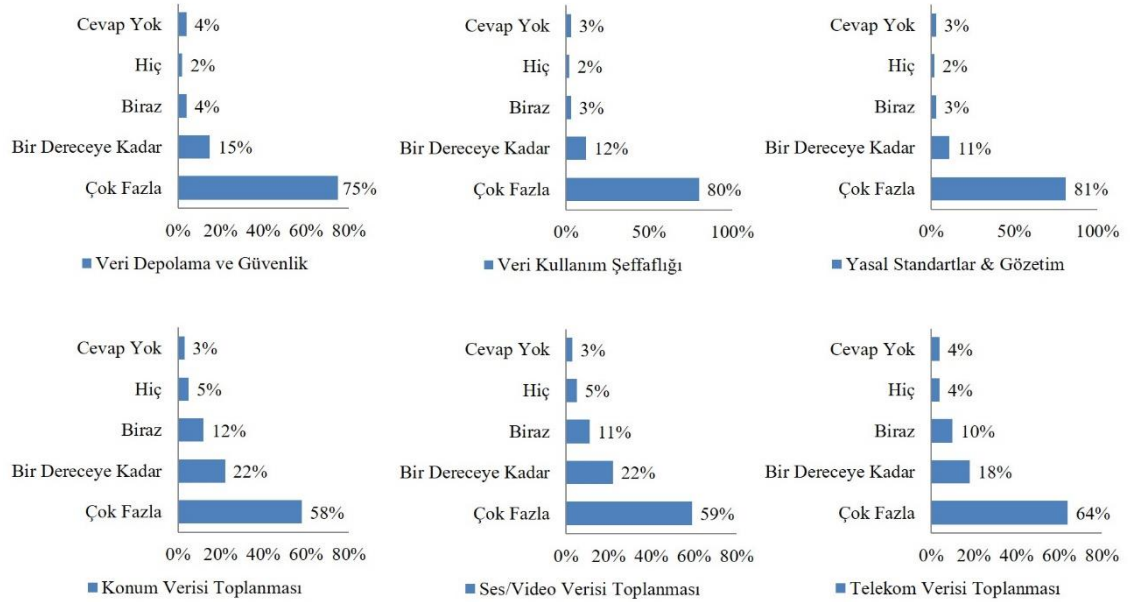
Ocak ayında, Büyük Verinin yaşamımızı, devlet ilişkilerini, vatandaşlarımızı, işletmeleri ve tüketicileri nasıl dönüştüreceği ile ilgili bir çalışma istediniz. Bu inceleme kamu ve özel sektörün Büyük Verinin risklerini minimize ederken, faydalarını nasıl maksimize edebileceği üzerine odaklanmıştır.

Aynı zamanda bu çalışma, ekonomimizi büyütmek, sağlık ve eğitimi geliştirmek, ulusal güvenliğimizi sağlamak ve daha fazla enerji verimliliği oluşturmak için Büyük Veri fırsatlarını açıklamaktadır.”

Raporda Büyük Veri nedir, veriden farkları nelerdir, zorlukları ve fırsatları, riskleri, endişeler, hangi alanlarda nasıl kullanılabilir, kamuda ve özel sektörde veri yönetimi vb. sorulara cevaplar verilmiştir.

Yine bu raporda gösterilen birkaç ankete de burada değinmek önemli görünmektedir. Beyaz Saray tarafından 4 hafta boyunca 24.092 kişi den toplanan yanıtlarla elde edilen veriler, tam bir istatistiksel veri olarak düşünülmesi de Beyaz Saray tarafından rapora yansıtılmış ve önem verilmiştir.

Veri uygulamaları ile ilgili sorulan sorulara verilen cevapların oranları aşağıdaki grafiklerde görünmekte olup, katılımcıların ortalama %61 oranında “çok fazla endişeli” oldukları sonucu elde edilmiştir. Katılımcıların %80 i veri kullanımının şeffaflığı konusunda “çok fazla endişeli” görüşündeler.



Şekil 5.15: Büyük veri ile ilgili ABD Beyaz Saray anketi.

(Kaynak: White House Report, Big Data: Seizing Opportunities, Preserving Values, Executive Office of the President, 2014)

Aslında bu rapor Büyük Verinin uluslararası çapta ne denli önemli bir yere sahip olduğunu göstermesi açısından oldukça önemlidir. Bu yüzden Büyük Veri üzerine giriş mahiyetinde olacak bu tez çalışmasının ülkemiz açısından da bir kilometre taşı olması sevindiricidir.

Şunu da ifade etmek gerekir ki, bilişim dünyasının bu derece geliştiği bir çağda ki Bilgi Çağı olarak ifade de edilmektedir. Bilişimin ve aygıtların “son kullanıcı” olan insan hayatının içerisine girmiş olmasından dolayı Büyük Veri kavramı çok önemli bir hal almıştır. Kısaca ‘Bilgi, 21. yüzyılın petrolü olacaktır!’ (Peter Sondergaard, Gartner, 2011) denilmektedir.

İnsanlar artık ifade etmek istediklerini, karşılıklı diyalog şeklinde müşteri temsilcilerine, kurum danışmanlarına, dilekçe ile ya da başka bir tekil şekilde sadece karşısındaki muhatabının bileceği şekilde değil, herkesin rahatça okuyabileceği, görebileceği, üzerine yorumlar yazabileceği, öneri verebileceği bir şekilde ortak platformlar üzerinden paylaşmaktadırlar.

İnsanların paylaşımları bilerek/isteyerek ya da bilmeden/istem dışı her an olmaktadır. Mobil telefonlar aracılığı ile konum bilgisi, arama motorlarındaki anahtar kelimelerin saklanması, web logları vb. gibi bilmeden/istem dışı bilginin yanında Facebook, Twitter, Google+, forumlar vb. gibi bilerek/isteyerek yapılan paylaşımlar aracılığıyla bilgi türemektedir.

Bunlar gibi literatür kısmında bahsedilen onlarca kaynaktan oluşan devasa veri “Büyük Veri Yönetimi” ile ele alınmalı ve kontrol edilmelidir. Çünkü kontrol edilmeyen ve işletmenin/kurumun haricinde büyümeye devam eden veri o işletmeye/kuruma hatta yöneticilerine negatif etkiler oluşturulabilir. Özellikle rekabetçi piyasada doğru şekilde analiz edilmeyen müşteri düşünce ve talepleri işletmelerin kısa ve uzun vadede negatif etkilenmesine, iyi bir şekilde analiz edilen veriler ise, hem kısa hem uzun vadede literatür kısmında da bahsedildiği örneklerde görüldüğü gibi pozitif kazanımlar sağlayacaktır.

Türkiye’deki en belirgin örneklerden birisi olan Hürriyet Gazetesinin yaptığı Büyük Veri çalışması bize bunu göstermesi açısından somut bir örnek ortaya koymaktadır: Hürriyet Gazetecilik Yönetim Kurulu Başkanı Vuslat Doğan Sabancı, büyük verinin yayıncılığı hızlı bir şekilde değiştirdiğini ifade etti. Hürriyet gazetesinin 20 yıl önce internetle

tanıştığını söyleyen Vuslat Doğan Sabancı, ilk önce yayın yönetmenliğinin yanı sıra bir de trafik editörlüğü oluşturduklarını anlattı. Vuslat Doğan Sabancı, “Trafikimizi çok daha farklı ölçebilmeye başladık. Yayınlarımızdaki etkiyi tirajlarımızda veya ertesi günlerde görmek yerine gerçek zamanlı olarak tespit etmeye başladık. Birdenbire erişimimiz o kadar çok arttı ki biz bile inanmadık. Her zaman doğru bir şekilde yatırım yaptık ve buna istikrarlı bir şekilde devam ettik. Ayda 1 milyon erişimden 15 milyon erişime ulaştık. Aylık sayfa gösterimimiz 1 milyara ulaştı. Müthiş bir trafik. Hâlâ da bu yolda gitmeye devam ediyoruz” dedi. Vuslat Doğan Sabancı, Büyük Verinin de yaklaşık 1.5 yıl önce Hürriyet Dünyası’nın hayatına girdiğini belirtti. Vuslat Doğan Sabancı, “Promosyonlar gazetelerin önemli bir harcamasıdır. Sadece 2-3 aylık bir çalışma ile aynı satışlarımızı tutmuş olmamıza rağmen yüzde 40 maliyetten tasarruf ettik. Bu hayal bile edemeyeceğimiz bir şeydi” diye konuştu. Büyük veriyi okur tatminini arttırmak ve reklam verenlere daha iyi hedeflenmiş reklam envanteri sunabilmek için kullandıklarını söyleyen Vuslat Doğan Sabancı sözlerine şu şekilde devam etti: “En doğru yerde, en doğru zamanda, en doğru cihazla hizmet vermeye çalışıyoruz. Hürriyet Dünyası’nın okunduktan sonra çıkılan haberlerini hesaplıyoruz. İçerik öneri motoruyla siteden çıkma oranını yüzde 10 düşürdük. Ayrıca çoğunlukla sadece erkeklerin tercih ettiği otomobil modeli veya markasının reklamını sadece erkeklere göstererek yüzde 20 daha doğru reklam hedeflemesi yaptık.” (<http://www.hurriyet.com.tr/buyuk-veri-cagi-29115095>, 2015).

Son olarak şunu da söyleyebiliriz ki, kurumlar daha iyi hizmet verebilmek, işletmeler daha iyi ürünler ve hizmetler sunmak, eksik yönlerini tespit edebilmek, pazar paylarını ve kar marjlarını arttırmak vb. işletme faaliyetlerinde üstünlük sağlayabilmek için Büyük Veriden yararlanmak zorundadırlar.

Bütün bu değerlendirmeler sonrası bir yerde yatırım söz konusu ise; “Büyük Veri olmadan alınacak stratejik kararlar her zaman eksik kalacaktır” diyebiliriz. Büyük Veriyi “yeni çağın yeni danışmanı” olarak tanımlamak hiç de abartı olmayacaktır.

KAYNAKLAR

- Akgül, M. K., 2014, *Büyük (Mega) Veriler ve Süper İşlemler Çağı-1*, <https://anahtar.sanayi.gov.tr/tr/news/buyuk-mega-veriler-ve-super-islemler-cagi-1/526>, [Ziyaret Tarihi: 22 Şubat 2015].
- Akgül, M. K., 2014, *Büyük (Mega) Veriler ve Süper İşlemler Çağı-2*, <https://anahtar.sanayi.gov.tr/tr/news/buyuk-mega-veriler-ve-super-islemler-cagi-2/548>, [Ziyaret Tarihi: 25 Nisan 2015].
- Baran, A., 2013, '*Big Data*': *Büyük Veri Bileşenleri ve Pazarlama Değişimleri*, <http://www.pazarlamasyon.com/2013/04/big-data-buyuk-veri-bilesenleri-ve-pazarlama-degisimleri/>, [Ziyaret Tarihi: 28 Şubat 2015].
- Barriball K.L. ve White, A., 1994, Collecting Data Using a Semi Structured. Interview: A Discussion Paper. *Journal of Advanced Nursing*, 19, 328-335.
- Boyacıgiller, N. A. ve Adler, N. J., 1991, The parochial dinosaur: Organizational science in a global context. *Academy of management Review*, 162, 262-290.
- Boyd, D. ve Crawford, K., 2011, *Six Provocations for Big Data A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, <http://ssrn.com/abstract=1926431> or <http://dx.doi.org/10.2139/ssrn.1926431>, [Ziyaret Tarihi: 9 Eylül 2014].
- Bryman, A., 2012, *Social research methods*, Oxford University press, Oxford, ISBN: 978-0-19-958805-3, 157-590.
- Buhalis, D. ve Deimezi, O., 2004, E-tourism developments in Greece: Information communication technologies adoption for the strategic management of the Greek tourism industry, *Tourism and Hospitality Research*, 52, 103-130.
- Buytendijk, F., Horiuchi, H., Heudecker, N., Adrian, M., McIntyre, A. ve Ingelbrecht, N., 2014, *Cool Vendors in Big Data*, <http://www.clearstorydata.com/gartner-cool-vendors-report/>, [Ziyaret Tarihi: 17 Eylül 2014].
- Callery, C. A., 2012, "*Data-Driven Decision Making In Community Colleges: An Integrative Model For Institutional Effectiveness*" *Dissertations*, <http://digitalcommons.nl.edu/diss/59>, [Ziyaret Tarihi: 22 Kasım 2014].
- Chen, C. L. ve Zhang, C., 2014, *Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Information Sciences*, <http://www.sciencedirect.com/science/article/pii/S0020025514000346>, [Ziyaret Tarihi: 6 Mayıs 2015].
- Chen, M., Mao, S. ve Liu, Y., 2014, *Big Data: A Survey*, <http://link.springer.com/article/10.1007%2Fs11036-013-0489-0>, [Ziyaret Tarihi: 4 Mayıs 2015].

- Cisco, 2009, *Cisco Basın Bültenleri*, <http://www.cisco.com/web/TR/news/press/archive/2009/070209.html>, [Ziyaret Tarihi: 16 Eylül 2015].
- Cisco, 2011, *Cisco Basın Bültenleri*, <http://www.cisco.com/web/TR/news/press/archive/2011/020611.html>, [Ziyaret Tarihi: 16 Eylül 2015].
- Collis, J. ve Hussey, R., 2009, *Business research: A practical guide for undergraduate and postgraduate students*, Palgrave Macmillan, New York, ISBN: 978-0-230-30183-2.
- Computing Research Association, 2011, “Challenges and Opportunities with Big Data.” <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>, [Ziyaret Tarihi: 12 Ağustos 2014].
- Conway, D. ve Diego K., 2013, *Innovation Patterns and Big Data*, Big Data and Business Analytics, In: Liebowitz, J., 8, CRC Press, Boca Raton, ISBN: 978-1-4665-6579-1, 131 – 147.
- Davenport, T. ve Dyché J., 2013, *Big Data in Big Companies*, https://www.sas.com/content/dam/SAS/en_us/doc/research1/big-data-big-companies-executive-summary-106462.pdf, [Ziyaret Tarihi: 1 Mayıs 2015].
- Desouza, K. C. ve Smith, K. L., 2014, *Big Data for Social Innovation*, http://ssir.org/articles/entry/big_data_for_social_innovation, [Ziyaret Tarihi: 6 Mayıs 2015].
- Dobre, C. ve Xhafa, F., 2013, *Intelligent services for Big Data science*, <http://www.sciencedirect.com/science/article/pii/S0167739X13001593>, [Ziyaret Tarihi: 6 Mayıs 2015].
- Fielding, N. G. ve Raymond, M. L., 2002, New patterns in the adoption and use of qualitative software, *Field Methods*, 14, 197-216.
- Flick, U., 2009, *An introduction to qualitative research*, Sage Publications Ltd., Hamburg, ISBN: 978-1-84787-1.
- George, G., Haas, M. R. ve Pentland, A., 2014, *Big Data and Management*, <http://dx.doi.org/10.5465/amj.2014.4002>, [Ziyaret Tarihi: 4 Mayıs 2015].
- Göksu, C., 2014, “*Big Data*” nedir? Geleneksel “Veri Yönetimi”ne etkisi ne olur?, <http://datawarehouse.gen.tr/big-data-nedir-geleneksel-veri-yonetimine-etkisi-ne-olur/>, [Ziyaret Tarihi: 28 Şubat 2015].
- Gülle, M. T., 2013, *Büyük Veri ya da İçgörü*, <http://www.tk.org.tr/index.php/tk/article/view/955>, [Ziyaret Tarihi: 22 Şubat 2015].
- Hansen, M. M., Miron-Shatz, T., Lau, A. Y. S. ve Paton, C., 2014, *Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives*, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287084/>, [Ziyaret Tarihi: 30 Eylül 2015].

- Hurwitz, J., Alan N., Fern H., ve Marcia K., 2013, *Big Data for Dummies*, John Wiley & Sons, ISBN: 978-1-118-50422-2.
- Hürriyet Gazetesi, 2015, *Büyük veri çağı*, <http://www.hurriyet.com.tr/buyuk-veri-cagi-29115095>, [Ziyaret Tarihi: 23 Ekim 2015].
- Jao, J., 2015, *Why Big Data Is A Must in Ecommerce*, <http://www.bigdatalandscape.com/news/why-big-data-is-a-must-in-ecommerce>, [Ziyaret Tarihi: 23 Ekim 2015].
- Kambatla, K., Kollias, G., Kumar, V. ve Grama, A., 2014, Trends in Big Data Analytics, *Journal of Parallel and Distributed Computing*, 74 (7), 2543-2686.
- Kuş, E. S., 2006, *Sosyal Bilimlerde Bilgisayar Destekli Nitel Veri Analizi: Örnek Program NVivo ile Gösterimler*, Anı Yayıncılık, Ankara.
- Kuş-Khalilov, M. C. ve Gündebahar, M., 2014, *Bankacılıkta Büyük Veri Uygulamaları: Bir İnceleme, Kuveyt Türk Katılım Bankası*, <http://ab.org.tr/ab14/bildiri/191.pdf>, [Ziyaret Tarihi: 25 Nisan 2015].
- LaborComm 2014 Bildiriler Kitabı, 2014, *LaborComm 2014 – 5. Uluslararası İşçi ve İletişim Konferansı*, http://laborcomm.org/wp-content/uploads/2014/08/LaborComm2014-bildiriler_kitabi.pdf, [Ziyaret Tarihi: 1 Mart 2015].
- Laney, D. ve Sussin, J., 2014, *Customer Analytics and the Art of the Possible With Big Data*, <https://www.gartner.com/doc/2703617/customer-analytics-art-possible-big>, [Ziyaret Tarihi: 25 Nisan 2015].
- Lester, S., 1999, *An introduction to phenomenological research*, Stan Lester Developments.
- Lewins, A. ve Silver, C., 2007, *Using software in qualitative research: a step-by-step guide*, Sage Publications Ltd., Hamburg, ISBN: 978-0761949220.
- Lindlof, T. R. ve Taylor, B. C., 2002, *Qualitative communication research methods*, Sage Publications Ltd., Hamburg, ISBN: 978-1412974738.
- Luo, S., Wang, Z. ve Wang, Z., 2013, *Big-Data Analytics: Challenges, Key Technologies and Prospects*, http://www.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, [Ziyaret Tarihi: 1 Mayıs 2015].
- Mayer-Schonberger, V. ve Cukier, K., 2013, *Big Data: A Revolution that will transform how we live, work and think*, John Murray An Achette UK Company, ISBN: 978-1-84854-790-2.
- McKinsey Global Institute Report, 2011, *Big data: The next frontier for innovation, competition, and productivity*, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, [Ziyaret Tarihi: 2 Mayıs 2015].

- Microsoft Blog, 2012, *Büyük Veri Nedir?*, <http://blog.microsoft.com.tr/buyuk-veri-nedir.html>, [Ziyaret Tarihi: 22 Şubat 2015].
- Oracle Corporation White Paper, 2012, *Oracle: Big Data for the Enterprise*, <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>, [Ziyaret Tarihi: 5 Mart 2015].
- Orb, A., Eisenhauer, L. ve Wynaden, D., 2001, Ethics in qualitative research, *Journal of nursing scholarship*, 331, 93-96.
- Orlikowski, W. J. ve Baroudi, J. J., 1991, Studying information technology in organizations: Research approaches and assumptions, *Information systems research*, 21, 1-28.
- Özbilgin, İ. G., 2015, *Kamuda Büyük Veri ve Uygulamaları*, <http://ab.org.tr/ab15/bildiri/483.pdf>, [Ziyaret Tarihi: 29 Eylül 2015].
- Özdoğan, O., 2014, *Big Data Kavramı Nedir? – Oracle Big Data Appliance*, <http://www.bilgicozumleri.com/big-data-kavrami-nedir-oracle-big-data-appliance>, [Ziyaret Tarihi: 1 Mart 2015].
- Parkhe, A., 1993, “Messy” research, methodological predispositions, and theory development in international joint ventures, *Academy of Management review*, 182, 227-268.
- Paul K., Radhika K., ve Udo S., 2013, *Finding Big Value in Big Data: Unlocking the Power of High-Performance Analytics*, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.
- Richards, N. M. ve King, J. H., 2014, *Big Data Ethics*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174, [Ziyaret Tarihi: 5 Mayıs 2015].
- Rouse, M., 2014, *Data mart (datamart) definition*, <http://searchsqlserver.techtarget.com/definition/data-mart>, [Ziyaret Tarihi: 25 Ekim 2015].
- Rouse, M., 2015, *Data silo definition*, <http://searchcloudapplications.techtarget.com/definition/data-silo>, [Ziyaret Tarihi: 25 Ekim 2015].
- Rubin, H.J. ve Rubin, I.S., 2005, *Qualitative interviewing: The art of hearing data*, Sage Publications Ltd., USA, ISBN: 978-1-4129-7837-8.
- Sart, G., 2013, Fikri Mülkiyet Hakları Hakkında Üniversitedeki Araştırmacıların Algısı, 8. Ulusal Eğitim Yönetimi Kongresi (Okul Geliştirme Sürecinde Öğretmen ve Yönetimin Niteligi: Politika ve Uygulamalar, İstanbul, 7-9 Kasım 2013, 101-103.
- Siegel, E., 2013, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*, Wiley, ISBN: 978-1-118-35685-2.

- Stan Lester Developments, 2012, *An introduction to phenomenological research*, http://www.academia.edu/1936094/An_introduction_to_phenomenological_research, [Ziyaret Tarihi: 17 Eylül 2014].
- Streubert, H. J. C., 1999, *Qualitative Research in Nursing Advancing the Humanistic Imperative*, Wolters Kluwer, China, ISBN: 978-0-7817-9600-2.
- Tambe, P., 2014, *Big Data Investment, Skills, and Firm Value*, <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2014.1899>, [Ziyaret Tarihi: 6 Mayıs 2015].
- Taylan, E., 2012, *Big Data konusunda bilinmesi gereken 6 araç*, <http://webrazzi.com/2012/12/26/big-data-konusunda-bilinmesi-gereken-6-arac/>, [Ziyaret Tarihi: 25 Nisan 2015].
- Tewksbury, R., 2009, Qualitative versus Quantitative Methods: Understanding Why Qualitative Methods are Superior for Criminology and Criminal Justice, *Journal of Theoretical and Philosophical Criminology*, 1 (1), 38 – 50.
- Thomas, I., 2013, *Putting Big Data at the Heart of the Decision Making Process*, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.
- Thomson Reuters, 2015, *Filling 9 billion bowls: The answer is data*, <http://blog.thomsonreuters.com/index.php/filling-9-billion-bowls-answer-data/>, [Ziyaret Tarihi: 20 Eylül 2015].
- Thomson Reuters, 2015, *How Will We Fill 9 Billion Bowls By 2050?*, <http://reports.thomsonreuters.com/9billionbowls/>, [Ziyaret Tarihi: 20 Eylül 2015].
- Topal, H. C., 2011, *2012'nin modası Big Data*, <http://kodcu.com/2011/12/2012nin-modasi-big-data/>, [Ziyaret Tarihi: 20 Kasım 2014].
- Trajman, O., 2013, *The Intrinsic Value of Data*, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.
- Tucker, D. S. ve Wellford, H. B., 2014, *Big Mistakes Regarding Big Data*, *The Antitrust Source*, http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/dec14_tucker_12_16f.authcheckdam.pdf, [Ziyaret Tarihi: 6 Mayıs 2015].
- Ularu, E., Puican, F., Apostu, A. ve Velicanu, M., 2012, *Perspectives on Big Data and Big Data Analytics*, http://dbjournal.ro/archive/10/10_1.pdf, [Ziyaret Tarihi: 2 Mayıs 2015].
- Vaitheeswaran, G. ve Arockiam, L., 2014, *Big Data for Education in Students' Perspective*, <http://research.ijcaonline.org/icaccthp2014/number4/icaccthp6041.pdf>, [Ziyaret Tarihi: 28 Eylül 2015].
- Vikipedi, 2015, *PlayStation Network erişiminin kesilmesi*, https://tr.wikipedia.org/wiki/PlayStation_Network_erişiminin_kesilmesi, [Ziyaret Tarihi: 28 Eylül 2015].

White House Report, 2014, *Big Data: Seizing Opportunities, Preserving Values*, Executive Office of the President, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf, [Ziyaret Tarihi: 4 Mayıs 2015].

World Economic Forum, 2012, *Big Data, Big Impact: New Possibilities for International Development*, http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf, [Ziyaret Tarihi: 25 Nisan 2015].

Zhao, D., 2013, *Frontiers of Big Data Business Analytics: Patterns and Cases in Online Marketing*, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.

EKLER

EK 1. “Stratejik büyük verinin yatırımlar üzerindeki etkileri” tezi araştırması görüşme rehberi.

Kurum Adı :

Kurumun Sektörü :

Görüşülen Kişi :

Pozisyonu :

1. Büyük Veri sektörünüzde kullanılıyor mu?

- Evet
Hayır
Yeni yeni

2. Nasıl? Açıklar mısınız?

3. Büyük Veri kurumunuzda kullanılıyor mu?

- Evet
Hayır
Yeni yeni

4. Nasıl? Açıklar mısınız?

5. Büyük Verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz? Nasıl? Açıklar mısınız?

- Evet
Hayır
Emin değil

6. Büyük Verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz? Nasıl? Açıklar mısınız?

- Evet
- Hayır
- Emin deęil

7. Sizce Büyük Veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir? Açıklar mısınız?

- Çok önemli
- Önemli
- Kısmen önemli
- Önemsiz

8. Büyük Verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz? Nasıl daha faydalı olabilir? Neden daha faydalı olabilir?

- Evet
- Hayır
- Emin deęil

9. Büyük Veri Yönetimine olan ihtiyaç, size göre altyapısal mı yoksa yönetsel bir konu mudur? Açıklar mısınız?

- Altyapısal
- Yönetsel
- Her ikisi

10. Türkiye’de Büyük Veri için altyapısal yeterlilik var mıdır?

- Var
- Yok
- Kısmen var

11. Büyük Veri ve Veri Yönetiminin sınırının olacağını tahmin edebiliyor musunuz? Açıklar mısınız?

- Evet
- Hayır

12. İş dünyasında Büyük Veri Yönetimi yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir? Açıklar mısınız?

- Evet, Yeni bir paradigma veya olgu
- Hayır, Verinin genişleyen devamı

13. Sizce Büyük Veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır? Açıklar mısınız?

- Fırsat-Şans
- Zorluk-Engel
- Her ikisi

14. Türkiye’de Büyük Veri’ den faydalanma seviyeleri ile ilgili neler söylemek istersiniz?

- Yeterli
- Yeterli değil
- Kısmen yeterli

15. Sizce Türkiye’de Google, Facebook, Amazon vb. gibi Büyük Veri’ den ciddi anlamda faydalanan benzeri firmalar var mıdır?

- Evet, var
- Hayır, yok
- Önümüzdeki 10 yıl içinde öngörüyorum
- Önümüzdeki 20 yıl içinde öngörüyorum

16. Nihai olarak, yaptığınız açıklamalar ve değerlendirmeler doğrultusunda, Büyük Veri yatırımlarınız üzerinde ne kadar etkili olacaktır? Bu etkilerin nedenlerini açıklayabilir misiniz? Eğer yanıtınız olumlu ise, Büyük Verinin yatırımlar üzerinde ne tür etkileri olacaktır? Açıklar mısınız?

- Çok etkili
- Etkili
- Kısmen etkili
- Etkisiz

ÖZGEÇMİŞ

Kişisel Bilgiler



| | |
|--------------------|---|
| Adı Soyadı | Ümit DÜLGER |
| Uyruğu | T.C. |
| Doğum tarihi, Yeri | 1978, Kadıköy |
| Telefon | 0 533 162 10 80 |
| E-mail | dulgerumit@gmail.com |
| Web adres | https://tr.linkedin.com/in/dulgerumit |

Eğitim

| Derece | Kurum/Anabilim Dalı/Programı | Yılı |
|---------------|--|------|
| Yüksek Lisans | İ.Ü. Fen Bilimleri Enstitüsü/ Mühendislik Bilimleri Anabilim Dalı / Mühendislik Programı | 2015 |
| Lisans | İ.Ü. Mühendislik Fakültesi / Elektrik-Elektronik Mühendisliği | 2002 |
| Lise | Pendik Lisesi, İstanbul | 1995 |