



BÜYÜK VERİ, BÜYÜK VERİ ANALİZİ VE UYGULAMA ALANLARI

Ayşegül İrem ULADI¹, Emin Sertaç ARI²

¹ Sosyal Bilimler Enstitüsü, Osmaniye Korkut Ata Üniversitesi, Osmaniye Türkiye

² Yönetim Bilişim Sistemleri, İktisadi ve İdari Bilimler Fakültesi, Osmaniye Korkut Ata Üniversitesi, Osmaniye Türkiye

ÖZET

Akıllı cihazların ve özellikle internetin çok yaygınlaştığı günümüzde üretilen veriler geçmişe göre katbekat artmıştır. Bu veri yığınlarını analiz etmek ve yapılan analizlerden anlamlı sonuçlar elde edebilmek klasik veri işleme yöntemleri ile mümkün olamamaktadır. Bu durum büyük veri kavramını ve büyük veri analizini ortaya çıkarmıştır. Çok büyük miktardaki verilerin analizini kolaylaştırdığı için büyük veri analizi işletmelere ve organizasyonlara çok büyük faydalar sağlamaktadır. Büyük veriden faydalanan işletmeler diğerlerine göre hem zamansal açıdan hem de ekonomik açıdan avantajlı konuma gelmektedir. Bu durum büyük veri kavramının ve analizinin önemini arttırmaktadır. Bu sebeple bu çalışmada büyük veri kavramından ve büyük veri analizinden detaylı bir şekilde bahsedilerek büyük verinin uygulama alanları açıklanmıştır. Çalışmada önce, büyük veri bileşenleri açıklanmıştır. Sonra, büyük veri analizi faydaları ve dezavantajları ile birlikte açıklanarak, büyük veri işleme platformlarından bahsedilmiştir. Daha sonra veri toplama ve veri çeşitliliği işlenmiş ve son olarak büyük verinin kullanıldığı sektörler, uygulama alanları ve sonuç kısmına yer verilmiştir.

Anahtar Kelimeler: Büyük Veri, İşletme, Hadoop

BIG DATA, BIG DATA ANALYSIS AND APPLICATION AREAS

ABSTRACT

Today, because of the smart devices and the internet extensive uses, the data produced has increased many times compared to the past. It is not possible to analyze these data stacks and to obtain meaningful results from the analyzes made with classical data processing methods. This situation has revealed the concept of big data and big data analysis. Big data analysis provides great benefits to businesses and organizations as it facilitates the analysis of large amounts of data. Businesses that benefit from big data come to an advantageous position compared to others both in terms of time and economy. This situation increases the importance of big data concept and analysis. For this reason, in this study, the concept of big data and big data analysis are mentioned in detail and the application areas of big data are explained. First, the big data components are explained in the study. Then, big data analytics are explained with their benefits and disadvantages, and big data processing platforms are mentioned. Afterwards, data collection and data diversity are processed and finally, the sectors where big data is used, application areas and the conclusion part are given.

Keywords: Big Data, Management, Hadoop

GİRİŞ

Veri; tek başına anlamı olmayan, herhangi bir işlem yapılmadan, gözlem veya ölçümle elde edilen ve analiz edilmesi gereken ham bilgidir (Doğan ve Arslantekin, 2016). Geleneksel veri işleme tekniklerinin kullanılmasıyla analizi mümkün olmayan farklı hacimdeki veri topluluğuna ise büyük veri denmektedir (Aktan, 2018). Büyük veri kavramı 21. Yüzyılın başlarında ortaya çıkmıştır. Günümüzde veri üretimi gerçekleştiren birçok araç ve teknoloji bulunmaktadır. Teknolojinin çok hızlı

gelişmesiyle beraber yapılan internet alışverişleri, sosyal medya paylaşımları ve bilgisayarlar sayesinde çok büyük veriler üretilmeye başlamıştır. Üretilen veriler çok büyük bir veri yığını oluşturmaya başlamıştır. Büyük veri teknolojileri bu verileri toplayıp, analiz edip, sınıflandırarak daha anlaşılabilir nihai verilere dönüştürür (Aksoy vd., 2017). Günümüzde şirketlerin büyük çoğunluğu büyük veriden faydalanmaktadır. Hatta büyük veri sadece şirketler tarafından değil devletler tarafından da çok büyük organizasyonlarda kullanılmaktadır (Bozkurt, 2016).

Büyük veriyi geleneksel veri analiz teknikleriyle analiz etmek mümkün değildir. Rekabetin çok yoğunlaştığı günümüzde şirketlerin rekabet avantajı sağlamak için büyük veri analizi gerçekleştirecek teknolojiler kullanması gerekmektedir. Şirketlerin çok kısa sürede toplanan büyük miktardaki veriyi; verimliliği yükseltecek, maliyetleri azaltacak, müşterilerle iletişimi kolaylaştıracak şekilde analiz etmesi gerekmektedir (Ayvaz ve Salman, 2020).

Bu çalışmanın amacı, büyük veri ve büyük veri analizinden detaylı bir şekilde bahsedip, büyük verinin uygulama alanlarını açıklamaktır. Büyük veri teknolojilerinin günümüzde çok önemli bir noktaya geldiğini düşündüğümüzde bu çalışmanın büyük veriyi anlatmak açısından literatürde önemli bir yer alacağı düşünülmektedir.

BÜYÜK VERİ KAVRAMI

Akademisyenler ve araştırmacılar büyük veri ile tam olarak ne kastedildiği konusunda farklı görüşlere sahiptirler. En geniş şekilde büyük veri; geleneksel veri tabanı yöntemleri ve araçlarıyla verimli bir şekilde analiz edilemeyen veri hacmi anlamına gelmektedir (Kaisler vd., 2013). Büyük veri hem yönetilen veri türünü hem de veriyi depolayıp analiz etmek için kullanılan teknolojileri anlatır (Doğan ve Arslantekin, 2016). Geleneksel veriden farkı; büyük veri daha fazla analiz ve işlem gerektirmektedir (Demirel vd., 2019).

Endüstri 4.0'ın gelişimi ile birlikte gelen süreçte büyük veri analizi alanında karmaşık ve zor çözümlenmeleri kolaylaştıran programlar geliştirilmiş ve bu sayede büyük veri analizlerinin kullanımı yaygınlaşmıştır.

Büyük verinin oluşmasını sağlayan veri tipleri şu şekildedir:

- Yapılandırılmamış veriler: katı bir formatı olmayan veri türüdür. Kitaplar, makaleler, e-postalar gibi formlardan toplanan metin, resim, ses ve video gibi dosyalar örnek verilebilir (Aktan, 2018).
- Yarı yapılandırılmış veriler: kısmen yapısal bir veri türünü ifade etmekle birlikte katı bir model değildir (Aktan, 2018).
- Yapılandırılmış veriler: modellenmesi, saklanması, analizi, sorgulanması ve görselleştirilmesi kolay olan veri türüdür (Aktan, 2018).

Bunların yanı sıra büyük veriye katkıda bulunan başka veri tipleri de bulunmaktadır. Sosyal medya verileri, web günlükleri ve tıklama akışları, konum verileri, makine ve sensörlerin oluşturduğu veriler, bilimsel veriler bunlardan bazılarıdır (Russom, 2011).

Büyük veri ile küçük veri arasındaki farklar aşağıda verildiği gibidir (Altunışık, 2015).

Tablo 1. Büyük Veri ile Küçük Veri Arasındaki Farklar

KÜÇÜK VERİ	BÜYÜK VERİ
Veriler belirli bir ortamda, belirli bir amaç için toplanır.	Veri farklı platformlardan analiz ve işlemler sırasında toplanır.
Verileri genellikle bireyler veya takımlar toplar.	Veriler veri toplama araçlarıyla otomatik toplanır.
Veri toplama işleminin belirli standartları vardır.	Veri toplama standartları veri toplanan ortama göre değişmektedir.
Veri toplama işleminde etik ve sosyal konular önemlidir.	Veri toplamada etik ve sosyal konular zayıf ve daha önemsizdir.

Veri toplanırken, verilerin elde edileceği Büyük veriler toplanırken varsayımsal onay kişilerin onayı beklenir. alınır.

Büyük Veri Bileşenleri

Büyük veri teknolojileri, çok büyük hacimlerden ekonomik olarak değer elde etmek için tasarlanmış yeni nesil teknolojileri ve mimarileri tanımlar (Kaisler vd., 2013). Büyük verinin 5 temel bileşeni bulunmaktadır. 5V olarak ifade edilen bu bileşenler; veri hacmi (volume), veri hızı (velocity), veri çeşitliliği (variety), veri doğruluğu (veracity) ve veri değeri (value) şeklindedir.

Veri hacmi (Volume)

Veri hacmi bir işletmenin erişebildiği veri miktarını ifade eder (Kaisler, 2013). Büyük veri ifadesindeki büyük kelimesi hacmi ifade eder (Katal vd., 2013). Günümüzde büyük veri mevcut veri tabanlarının kaldıramayacağı seviyelere gelmiştir. Akıllı telefonlar, bilgisayarlar, GPS sensörleri, akıllı cihazlar her gün terabayt düzeyinde veriler üreterek büyük veriye çok büyük katkı sağlamaktadır (Aktan, 2018). İlerde bu büyük verilerle nasıl başa çıkılması gerektiği araştırmacıların üzerinde durması gereken bir konudur (Aksoy vd., 2017). Bankacılık işlemleri, bilimsel deneyler ve araştırmalar, GPS verileri ve sosyal medya verileri büyük veri hacminin genişlemesine katkıda bulunan veri çeşitlerinden bazılarıdır (Demirold vd., 2019).

Hacim, geleneksel bilişim teknolojileri altyapısı ve araçlarıyla toplanması, yönetilmesi ve analiz edilmesi zor olan büyük miktardaki yapılandırılmış ve yapılandırılmamış verileri ifade eder. Bu nedenle, bu büyük veri kümeleri verileri yakalamak, depolamak ve analiz etmek için yeni ve yenilikçi araçlar ve yaklaşımlar gerektirir (Alharthi vd., 2017).

Veri hızı (Velocity)

Büyük verilerde hız, çeşitli kaynaklardan gelen verilerin hızıyla ilgilenen bir kavramdır. Veri hızı, büyük veri setlerinin genellikle gerçek zamanlı veri akışlarından oluştuğu ve sürekli genişletildiği gerçeğini ifade eder (Alharthi vd., 2017). Veriler devamlı hareket durumundadır, bu yüzden hız bileşeni araştırmacılar için çok önemli bir konu haline gelmiştir (Aktan, 2018). Çok hızlı veri üretimi gerçekleştiği için veriler kısa sürede büyük boyutlara ulaşmaktadır. Bu şekilde büyüyen veriler için gereken işlem sayısı da artmaktadır (Aksoy vd., 2017).

Veri çeşitliliği (Variety)

Büyük veri çok farklı türlerde ve formatlarda üretilmektedir ve bu veriler arasında belirli bir kural bulunmamaktadır (Aktan, 2018). Veri çeşitliliği verinin zenginliğinin bir ölçüsüdür; metin, görüntü, video, ses gibi (Kaisler vd., 2013). Büyük verilerin analizinde en büyük zorluklardan birisi verilerin çeşitliliğidir. Üretilen veriler yalnızca geleneksel veriler değil, aynı zamanda web sayfaları, web günlük dosyaları, sosyal medya siteleri, e-posta, belgeler, sensörler gibi çeşitli kaynaklardan gelen yarı yapılandırılmış verileri de içerdiğinden verileri kategorize etmek kolay değildir (Katal vd., 2013).

Veri doğruluğu (Veracity)

Büyük verinin doğruluğunu ve güvenilirliğini ifade etmektedir. Kullanılan büyük veri hatalara ve yanlış yorumlara neden olabileceğinden, değerli bilginin kullanılması için veri doğruluğu çok önemlidir (Aktan, 2018). Çok hızlı büyüyen verilerin, akışta bir sorun çıkmaması için güvenli olup olmadığının denetlenmesi gerekmektedir (Aksoy vd., 2017). Geçerliliği olmayan verileri analiz etmek, veri gürültüsünü gidermek veri doğruluğu bileşenini ilgilendirmektedir (Demirold vd., 2019).

Veri değeri (Value)

Toplanan veriler analiz edilip, sınıflandırıldıktan sonra işletmelere değer katması gerekmektedir (Aksoy vd., 2017). Veriler şirketler ve organizasyonlar için ekonomik değer oluşturmaya başladığında büyük verinin değeri ortaya çıkmaktadır (Aktan, 2018). Veri değerini veriler analiz edildikten sonra işletmeye sağladığı fayda olarak da tanımlayabiliriz. Değer ve doğruluk bileşenleri birbirleriyle doğrudan ilişkilidir. Ne kadar doğru veriler elde edilirse, veriden çıkarılan bilgi o kadar fazla değer taşır (Demirold vd., 2019).

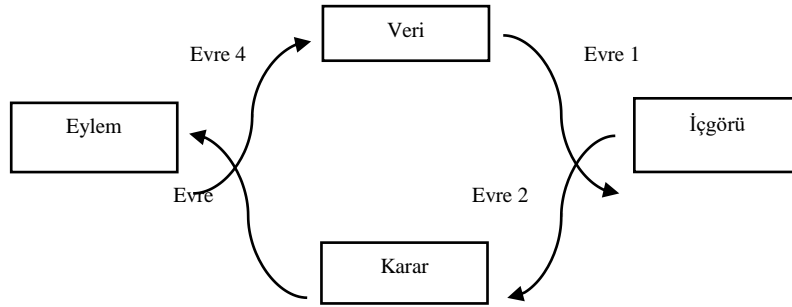
BÜYÜK VERİ ANALİZİ

Büyük veri analizi, çok geniş ve farklı türdeki verileri işlemek ve analiz etmek için kullanılan tekniklerdir. Büyük veri analizinde, geleneksel veri analiz teknikleriyle analiz edilemeyen yapısal, yarı yapısal ve yapısal olmayan verilerin analizi gerçekleştirilir (Aktan, 2018). İşletmelerin büyük veri analizini performansı artıracak, maliyetleri düşürecek, verimliliği yükseltecek şekilde gerçekleştirmesi gerekmektedir. Büyük veri analizi için kaynaklar tespit edilmeli ve işletmeler arasında büyük veri projeleri yapılmalıdır. Ayrıca büyük veri analizi için çalışanlara eğitimler verilmelidir (Özdemir ve Sağiroğlu, 2018). Büyük veri analizinin şirketlere çok büyük faydaları bulunmaktadır. Bunlardan bazıları şu şekildedir (Ning Dai vd., 2019):

- Fabrika operasyonlarının ve üretimin iyileştirilmesi,
- Makinelerin arıza sürelerinin azaltılması,
- Ürün kalitesinin iyileştirilmesi,
- Tedarik zincirinin verimliliğinin artırılması,
- Müşteri deneyiminin iyileştirilmesi.

Büyük veri analizinin amacı; organizasyonel karar verme ve yürütme süreçlerini geliştirmektir. Tabesh ve arkadaşlarına göre (2019) büyük veri analizi 4 evreden oluşmaktadır.

Şekil 1. Büyük Veri Analizi Evreleri (Tabesh vd., 2019).



Evre 1: Büyük, çeşitli ve genellikle yapılandırılmamış veriler iç ve dış kaynaklardan toplanır ve iç görüler oluşturmak için gelişmiş analitik araçları ve algoritmalar kullanılarak işlenir. Bu iç görüler daha sonra karar vericiler tarafından yorumlanır ve karar verme sürecinde kullanılır.

Evre 2: 1. Evrede üretilen iç görüler kararlara dönüştürülür.

Evre 3: Kararlar belirli operasyonel eylemlere dönüştürülür.

Evre 4: Kararların eyleme dönüştürülmesi, gelecekte karar verme çabaları için sürece geri döndürülen ek sonuçlar üretir.

Büyük Veri Analiz ve İşleme Platformları

Büyük veri işleme platformları yığın veri işleme ve akan veri işleme platformları olarak ikiye ayrılır. Yığın veri işleme teknolojilerinden en çok kullanılanı eşle-indirge paradigmasını kullanan Hadoop'tur. Akan veri işleme için ise; Storm, S4, Samza, Flink Streaming, Spark Streaming ve Apache Spark'tır (Kayabay vd., 2016). Bu bölümde veri analiz platformları incelenecektir.

Hadoop

Hadoop, büyük veri kümelerini analiz etmek için çalışan, dağıtık dosya sistemi olarak adlandırılan, Java tarafından geliştirilmiş açık kaynak kodlu bir yazılımdır (Özbilgin, 2015). Hadoop'ta veri analizi yapılırken ölçeklenebilirlik, süreklilik, hatadan kurtarma ve esneklik gibi parametreler kullanılmaktadır (Doğan, 2014). Büyük veri analizinde en çok kullanılan yazılım Hadoop'tur. Günümüzde Amazon, Facebook, Twitter, Yahoo, Adobe gibi büyük şirketler büyük veri analizinde Hadoop kullanmaktadır (Demirold vd., 2019).

Hadoop sayesinde kullanıcıların diskleri bir araya toplanarak büyük bir sanal disk meydana getirilir. Bu şekilde değişken boyutlu dosyalar bu diskte depolanabilir. Hadoop çok miktardaki ve çok

büyük boyuttaki veri ve dosyaları okuma kolaylığı sağlar (Özbilgin, 2015). En çok kullanılan yöntem Hadoop olsa da büyük veri analizi tek bir yöntemin kaldıramayacağı kadar karmaşıktır. HBase, HCatalog, PIG, Mahout, Cassandra ve In-memory gibi tekniklerden de faydalanılmaktadır (Demiroglu vd., 2019).

Hadoop ile elde edilen çıktılar kesinlik içermemektedir bu sebeple örneğin, banka hesaplarının ayrıntılarının sorgulanması gibi durumlarda Hadoop'un kullanılması pek güvenilir olmamaktadır. Ancak kesin ve net cevapların gerekmediği durumlarda ilişkisel veri tabanlarına göre çok hızlıdır (Doğan ve Arslantekin, 2016, s. 30).

Örneğin, Visa 73 milyar civarında işlemin yaklaşık 2 yıllık süre zarfında tutulmuş tüm kayıtları için gerekli işlem süresini Hadoop kullanarak 1 ay gibi bir süreden 13 dakikalık bir süreye indirmiştir (Schönberger ve Cukier, 2013, s. 53). Hadoop üç ana bileşenden oluşmaktadır. Bunlar;

- HDFS (Hadoop Dağılık Dosya Sistemi),
- MapReduce (Haritalama ve İndirgeme),
- Kaynak Yöneticisi

HDFS

HDFS, Hadoop'un ana depolama alanıdır. Eldeki çok büyük veriyi farklı kaynaklara aktarma işlemini yapar. Bu işlemi yaparken aynı zamanda farklı kaynaklardaki verileri işlem aşamasında sanki tek bir kaynaktan gibi değerlendirebilmektedir. Ayrıca verinin türüne ve amacına göre bir kopyasını farklı kaynaklarda tutabilmektedir. Bu sayede örneğin Facebook sunucularından biri arızalandığında farklı bir sunucudan veri akışını sağlayarak sistemin sorunsuz akışını sağlayabilmektedir.

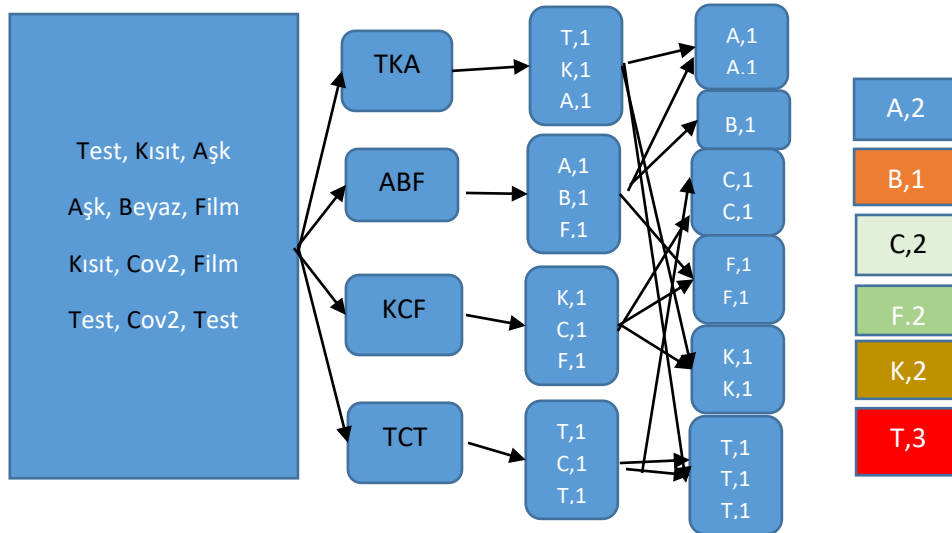
MapReduce

MapReduce aşamasında gerçekleşen işlemler aşağıda bir örnek ile gösterilecektir.

Örnek Durum:

Twitter üzerinden rasgele alınmış basit üçlü kelime grupları için kelime saydırma uygulaması MapReduce kullanılarak Şekil 2 'de olduğu gibi yapılmaktadır.

Şekil 2. Twitter MapReduce Örneği



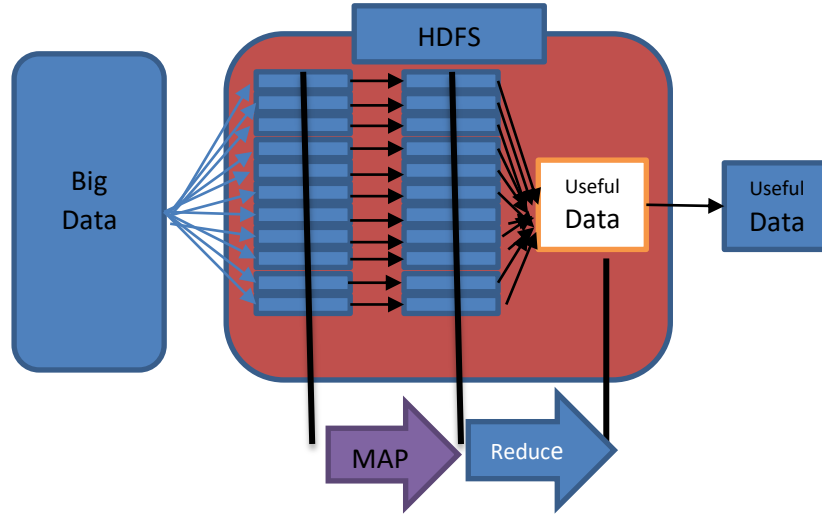
Burada, MapReduce bileşeninin temel çalışma prensibi basit bir twitter örneği ile anlatılmak istenmiştir. Burada görüldüğü üzere MapReduce aşamasında önce farklı kaynaklardan elde edilen veriler üzerinde çalışılmış ve elde edilen sonuçlar birleştirilerek ana sonuç elde edilmiştir.

Kaynak yöneticisi

HDFS ile MapReduce arasında geçişin koordinasyonun ve yönetimin sağlandığı sistemdir.

Aşağıda Şekil 3'te büyük veri ekosistemi içerisinde HDFS ve MapReduce bileşenlerinde yapılan işlemler gösterilmektedir.

Şekil 3. Büyük Veri Ekosistemi



NoSQL veri tabanları

NoSQL; günümüzde ilişkisel yazılımlara alternatif olarak ortaya çıkmış sistemlerdir. Bu tip yazılımlar tasarım ve kullanım basitliği ve erişilebilirliğin kolay olması gibi sebeplerle tercih edilmektedir (Özbilgin, 2015). Çağdaş veri yönetimi çözümleri, büyük verinin çeşitliliği ve doğruluğu için sınırlı entegrasyon yetenekleri sunarken, bulut bilişim ve NoSQL'deki son gelişmeler yeni çözümlerin kapısını açmaktadır. NoSQL veri tabanları, büyük verinin gereksinimlerini yüksek ölçeklenebilirlik, kullanılabilirlik ve hata toleransı ile karşılamaktadır (Yang vd., 2016).

NoSQL; farklı çeşitlikte ve büyüklükteki verileri bünyesinde barındırmakta ve bu verilerin arasında arama yapılmasına imkân sağlamaktadır. Bu veri tabanı veri dağılımına olanak tanıdığı için daha fazla işlem ve depolama alanı gerektirmektedir. Veri boyutunun artması ve geleneksel veri tabanlarının yetersiz kalmasıyla, büyük veri uygulamalarında çoğunlukla NoSQL kullanılmaktadır (Doğan ve Arslantekin, 2016).

Storm

Gerçek zamanlı akış işleme için en popüler platform, sınırsız veri akışı için açık kaynaklı, ölçeklenebilir, dağıtılmış ve hataya dayanıklı olan Storm'dur. Twitter tarafından geliştirilmiştir (Kayabay vd., 2016). Storm özellikle kullanımı çok kolay olan ve tüm verilerin işlenmesini sağlayan veri akışı için geliştirilmiştir (Jan vd., 2019).

S4

Yahoo tarafından piyasaya sürülmüştür. Büyük veri ortamında sınırsız veri akışı için genel amaçlı, hataya dayanıklı, dağıtılmış ve genişletilebilir bir platformdur. S4; Yahoo gibi çeşitli uygulamalarda kullanılmış ve iyi bir performans elde etmiştir (Jan vd., 2019). Bazı farklılıkları olsa da Storm ile benzerdirler.

Samza

Storm ve S4'ten daha sonra ortaya çıkmıştır. LinkedIn tarafından geliştirilmiştir. Analizlere ölçeklenebilirlik, hata toleransı ve işlem kolaylığı sunmaktadır. Akış katmanı, yürütme katmanı, işleme katmanı olmak üzere 3 katmandan oluşmaktadır (Kayabay vd., 2016).

Apache Spark

Büyük veri ile ilgili işlemler yapılmasını sağlayan hızlı bir uygulamadır. Akan verileri kolaylıkla işleyebilmekte ve Hadoop'a göre daha az kod yazarak daha kolay ve hızlı veri işlemesi yapmaktadır. Bu uygulama ile yapılan işlemlerde disk üzerinde Hadoop'dan 10 kat daha hızlı sonuç

alınabilmektedir. Hatta bellekte 100 kata kadar daha hızlı veri işlenmesine olanak tanıyan bir platformdur (Karabay ve Ulaş, s.7).

Büyük Veri Analizinin Faydaları

Dünyanın dört bir yanındaki kuruluşlar, büyük ve karmaşık veri setlerini analiz etme ve kullanma yeteneğinin 21. Yüzyılda en önemli rekabet avantajı olacağını düşünmektedir. Büyük veri, daha iyi müşteri deneyimi sağlama, iç verimliliği geliştirme ve nihayetinde tüm sektörlerdeki kuruluşların kârlılığını ve rekabet gücünü geliştirme potansiyeline sahiptir (Alharthi vd., 2017). Büyük veri; iş, kamu yönetimi, ulusal güvenlik, bilimsel araştırma, sağlık gibi hayatın hemen hemen her alanında kullanılmaktadır (Jan vd., 2019). Russom'a (2011) göre Büyük veri analizinin faydaları şu şekildedir:

- Daha iyi hedeflenmiş sosyal pazarlama,
- Daha doğru iş anlayışı,
- Satış ve pazar fırsatlarının tanınması,
- Müşteri davranışlarının tanımlanması,
- Dolandırıcılığın tespiti,
- Risklerin ölçülmesi,
- Daha iyi planlama ve tahmin,
- Tıklamalara göre tüketici davranışını anlama.

Son yıllarda teknolojinin hızlı gelişmesiyle hayatımıza giren büyük veri analizini, geleneksel veri analiz yöntemlerinden ayıran en önemli özellik her tipteki verilerin analizini yapabilmesidir. Büyük veri analizini en iyi şekilde yapıp, büyük verinin olanaklarından maksimum düzeyde faydalanabilmek için şirketlerin kaliteli insan kaynağına ihtiyacı vardır (Ayvaz ve Salman, 2020).

Büyük Veri Analizinin Zorlukları

Çok büyük boyutlarda verilerin analiz edilmesinden dolayı büyük veri analizinde birtakım zorluklar bulunmaktadır. Etkili bir veri analizi için bu zorlukların üstesinden gelinmesi gereklidir. Büyük veri analizinin önünde; büyük veri analizi için yetersiz personel, fazla maliyet, büyük veri analiz sistemini tasarlama zorluğu, ölçeklenebilirlik sorunları, uzman eksikliği, yüksek işletme giderleri gibi engeller bulunmaktadır (Russom, 2011). Bu bölümde büyük veri analizinin zorlukları detaylı bir şekilde incelenecektir.

Veri toplama ve veri çeşitliliği

Sensör ağları gibi bazı veri kaynakları şaşırtıcı miktarda veri üretebilir. Bu verilerin arasında ilgisiz veriler de bulunmaktadır. Bu verilerin filtrelenip, gerekli olanların ayırt edilmesi gereklidir (Labrinidis ve Jagadish, 2012). Büyük veri analizi için farklı kaynaklardan farklı özellikte çok fazla veri toplanmaktadır. Bu durum veri analizini zorlaştırmaktadır (Özdemir ve Sağiroğlu, 2018).

Altyapı hazırlığı

Büyük veri analizi için bilişim teknolojileri altyapısının geliştirilmesi, gerçek zamanlı olarak yüz milyonlarca kayıdın analizi için yazılım ve donanımına önemli yatırımlar yapılması gereklidir. Kullanılan mevcut teknolojilerin çoğu büyük veri analizinin gereksinimlerini karşılamak için tasarlanmamıştır. Bulut ve heterojen bilgi işlem mimarilerinin büyük veri için çözümler sunduğuna inanılmaktadır. Ancak bu teknolojiler büyük miktarda veri işlendiğinde teknik ve ekonomik açıdan genellikle başarısız olur ve bu sistemin montajının maliyeti yüksek olacaktır (Alharthi vd., 2017).

Teknik zorluklar

Büyük veri geleneksel analiz yöntemleriyle analiz edilemeyecek kadar büyük boyutlu ve hızlıdır. Büyük veri analizini gerçekleştirebilecek etkili ve yüksek performanslı tekniklere ihtiyaç vardır (Özdemir ve Sağiroğlu, 2018). Büyük veri projeleri için kullanılan veri tabanı yazılımları satan RainStor CEO'su John Bantleman'a göre geleneksel veri yönetimi araçları yeni veri oluşturma hızına ayak uydurmak için ölçeklenebilir olmadığından büyük veri analiz sistemlerinin yenilenmiş bilişim teknolojileri altyapılarına ihtiyacı vardır (Tabesh vd., 2019).

Veri görselleştirme

Veri görselleştirme büyük veri analizi gerçekleştirildikten sonra analizin daha iyi anlaşılmasını sağlayan bir araçtır. Çok büyük boyutta olduğu için büyük veride görselleştirme yapmak zordur (Özdemir ve Sağiroğlu, 2018). Veri Görselleştirme makalenin ilerleyen sayfalarında daha detaylı şekilde incelenecektir.

İnsan kaynağı ihtiyacı

Büyük veri analizini gerçekleştirilebilmek için kalifiye insan kaynağına ihtiyaç duyulmaktadır (Özdemir ve Sağiroğlu, 2018). Şu anda büyük veriye veya genel analitik becerilere sahip çalışanların eksikliği, büyük veri analizi yapmak isteyen kuruluşların karşılaştığı en büyük zorluklardan biridir. Mevcut çalışanlar arasında veri analitiği becerilerinin olmaması, bilgilerin yanlış kayda yerleştirilmesine, değerli bilgilerin kaybedilmesine ve işletmenin yakaladığı verilerden elde edeceği değer sınırlandırılmasına yol açabilir (Alharthi vd., 2017).

Yönetimsel zorluklar

İşletmelerde genel olarak en önemli kararlar üst düzey yöneticiler tarafından verilir. Bu kararlar genellikle veriye dayalı şekilde alınmaktadır. Eğer yönetici veriye güvenmezse ve ya veriler yöneticinin kendi deneyimleriyle uyumsuzsa sezgisel olarak karar verilebilir. İşletmelerde verilere yönelik karar alma kültürünün oluşması büyük veriyle çalışılmasını kolaylaştıracaktır (Altunışık, 2015).

BÜYÜK VERİNİN KULLANILDIĞI SEKTÖRLER VE UYGULAMA ALANLARI

Büyük veri teknolojileri günümüzde birçok sektörde kullanılmaktadır. Bu bölümde büyük verinin kullanıldığı sektörlerden bahsedilip örnekler verilecektir.

Taşımacılık, Lojistik, Perakendecilik

GPS sensörleri, RFID okuyucuları, akıllı telefonlar sayesinde hızlı bir şekilde veri toplanmaktadır. Bu veriler sayesinde işletmeler karşılıklı çıkan fırsatları daha çabuk fark edebilmektedirler ve süreçleri daha iyi optimize edebilmektedirler (Aktan, 2018). Perakendecilik sektöründe büyük veri analizi, çok sık kullanılmaktadır. Örneğin, cep telefonlarının GPS sinyalleri sayesinde marketler yakınlarında bulunan müşterilere, müşterilerin durumuna göre ürün teklifi gönderebilirler (Doğan, 2014).

Sağlık Hizmetleri

Sağlık sektöründe hastalar, hastalıklar, hastalıkların belirtileri, ilaçlar ve ilaçların yan etkileri konularında elde edilen veriler harmanlanarak büyük veri kümeleri elde edilir. Bu veri kümeleri kullanılarak detaylı hastalık analizleri yapılabilir ve hastalıklar hakkında yeni bilgiler elde edilebilir.

Örneğin, Covid-19 gibi salgın hastalıkların araştırılması aşamasında hem hastaların taşıdıkları belirtilerin verileri, hem ilaçların hastalar üzerinde etkisinin verileri hem de elektronik tıbbi görüntüleme verileri gerekmektedir. Bu durumda büyük veriden faydalanılmaktadır (Aktan, 2018).

Sağlık sektöründe büyük veriye hasta randevu planlama konusunda da başvurulabilmektedir. Sağlık kurumuna gelen hastaların ne zaman, hangi hastalık için geldiğini ve bu durumda tekrar ne zaman kontrole geleceğini hastalara bildirme işlemi büyük veri sayesinde gerçekleştirilebilir (Aksoy vd., 2017).

Devlet

Devlet kurumları; nüfus sayımı, seçim sonuçlarının analizi, bütçeleme gibi işlemlerde büyük veriyi kullanmaktadır (Aktan, 2018). Devletler aşağıdaki konularda büyük veri analizini kullanabilirler (Doğan, 2014):

- Tehditleri tahmin edip engelleme,
- Sosyal hizmetlerde sahtecilik tespiti,
- Vergide sahtecilik ve kaçakçılık tespiti,

- Suç tahmini ve engelleme.

Kamuda gerçekleştirilen büyük veri uygulamalarının kaynaklarına örnek verecek olursak;

- Vatandaşların işlemleriyle oluşan veriler,
- Kamu kuruluşlarından gelen veriler,
- Diğer ülkelerde bulunan vatandaşların/ kurumların oluşturduğu veriler

şeklinde sıralayabiliriz.

Büyük veri çalışmalarının sağladığı faydalar ve büyük veri örnekleri kamuda çalışan yöneticilere anlatılırsa kamu işlemlerinde çok büyük faydalar sağlayabilir (Özbilgin, 2015).

E-Ticaret

E-ticaret sektöründe büyük verinin kullanımı müşterilerin daha hızlı ve daha doğru anlaşılabilmesi açısından büyük önem taşımaktadır.

Örneğin, büyük veri kullanılarak siteyi gezen ziyaretçinin sitedeki hareketleri incelenerek ilgi alanlarını ve ihtiyaçlarını içeren profili maksimum hızda ve doğrulukta elde edilir ve bu ihtiyaçlara göre ürünler sergilenerek satışların artırılması sağlanabilir (Aksoy vd., 2017).

Sektörün temellerini atan ve günümüzde de sektörün en önemli şirketlerinden biri olan Amazon büyük veri kullanımına oldukça önem vermiş hatta bu alanda herkesin erişebileceği kendi platformu AWS'yi (Amazon Web Services) kurmuştur. AWS'nin etkin kullanılmaya başlandığı 2016 yılından sonra her yıl kâr marjını kat ve kat artırarak büyük başarı elde etmiştir.

Sigortacılık

Sigortacılık sektöründe büyük veri yardımıyla daha iyi fiyat ayarlaması yapılır ve müşteri ilişkileri geliştirilir böylelikle kârlılık artırılabilir. Sigortacılıktaki en önemli büyük veri unsurları; müşterilerin cinsiyeti, yaşı, yaşadıkları bölge ve sigortalı olup olmadıklarıdır (Demirel vd., 2019).

Bankacılık

Bankacılık sektöründe büyük veri analizi gerçekleştirecek şirketler kâr, kazanç ve risk durumlarını daha net görebilirler, böylelikle rakiplerine karşı çok büyük avantaj kazanabilirler. Bankacılık sektöründe çağrı analizi ile müşteri memnuniyeti ve dolandırıcılık tespiti gibi konularda büyük veri analizi yapılabilir (Doğan, 2014).

Büyük Veri Kullanım Örnekleri

Tablo 2. Büyük Veri Kullanım Örnekleri

Birleşik Arap Emirlikleri'ne ait Etihad havayolları, dünya genelinde 89'dan fazla noktaya uçmakta ve yılda 10 milyon yolcu taşımaktadır. 2003 yılından bu yana, filosuyla ilgili verileri oluşturmak için her uçakta yüzlerce sensör kullanmaya başlamıştır. Bu düzenleme bir dizi elektronik ve elektronik olmayan nesnelerin interneti kavramına benzer. Nesnelere tek bir ağa bağlanır ve gerçek zamanlı bilgi alışverişi yapar. Ağ büyük hacimlerde dijital veri üretir. Bu veriler Etihad tarafından analiz edilir ve tüm uçak filosunu kontrol etmek için kullanılır. Şirket filosunun durumunu ve konumunu gerçek zamanlı olarak izler ve sorunları önlemeye çalışır. Bu şirket için önemli ölçüde maliyet tasarrufu sağlar (Alharthi vd., 2017).

Hollanda İstatistik Kurumu, tüm yollara sensörler yerleştirerek, yolların kullanım oranlarını belirlemiştir. Tüm araçların tiplerini ve hızlarını tespit edip sisteme iletmektedir. Bu şekilde ulaşımda alınacak tedbirler belirlenebilir (Özbilgin, 2015).

Her yıl dünyanın dört bir yanındaki Walt Disney parklarını yaklaşık 100 milyon kişi ziyaret etmektedir. 2013 yılında şirket RFID sensörleriyle park girişleri için bir bileklik geliştirmiştir. RFID sensörleri, bağlı oldukları nesne hakkında dijital bilgileri otomatik ve kablosuz olarak, önemli bir mesafeden ve görüş hattı gerektirmeden iletebilmektedir. Magic Band adı verilen bu bileklikler, ziyaretçilere kuyruklarda atlama ve ön rezervasyon gibi avantajlar sunmaktadır. Magic Band bu

avantajları sunarken aynı zamanda ziyaretçinin hareketini de kaydeder ve ziyaretçilerin Disney Park ile etkileşimleri hakkında büyük hacimli, karmaşık ve değerli verilerin toplanmasını kolaylaştırır. Disney, satın alma geçmişi, bekleme süresi ve tercihler dahil olmak üzere müşteri davranışlarını analiz etmek için çeşitli büyük veri araçlarıyla birlikte bu veri setlerini kullanır. Bu bilgiler, Disney'in park ziyaretleri için daha iyi deneyimler sağlamasına ve pazarlama etkinliğini artırmasına yardımcı olur. Tüm bunlar şirketin daha fazla müşteri çekmesine ve müşteri başına gelirlerini artırmasına yardımcı olur (Alharthi vd., 2017).

Asya Sağlık Kurumu, hasta verilerinin Hadoop'ta analizini gerçekleştirmiş, radyoloji ve patoloji bölümlerinde daha hatasız ve hızlı teşhis konulmasını sağlamıştır (Özbilgin, 2015).

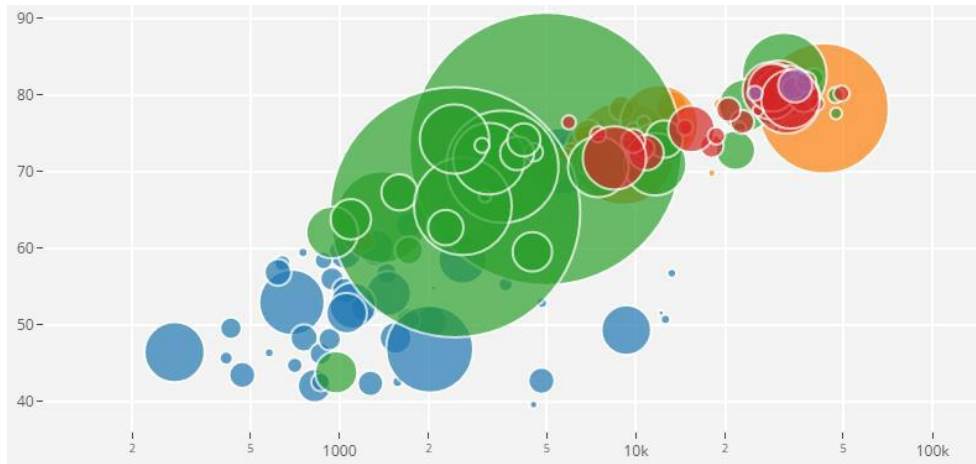
VERİ GÖRSELLEŞTİRME

Büyük veri analizlerinde veri miktarının yoğun olması çok önemlidir ancak asıl önemli olan büyük veriyi yönetebilmek ve bunları değerlere dönüştürebilmektir. Büyük veride hem analizler yapılmaktadır hem de ileriye yönelik tahminleme yapılmaktadır (Özcan, 2021). Bu durumda uygun grafiklerden faydalanmak, analiz edilen veriyi ilgi çekici görsellerle anlatmak etkili yollardan birisidir. Bu sebeple büyük veri ve görselleştirme sık sık birlikte anılmaktadır (Erkurt, 2021).

Görselleştirme, büyük verinin içindeki karmaşık ilişkileri sade bir şekilde özetleyen, anlaşılması kolay olan, grafiksel bir yöntemdir (Yıldırım ve Erkurt, 2020). Görselleştirme kavramı, her türlü veriyi grafiklerle özetlemeyi hedefler (Erkurt, 2021). Veri görselleştirmenin amacı; istatistiksel olarak karmaşık olan verileri, kolay anlaşılacak grafikler şeklinde sunmaktır. Yapılan araştırmalara göre insan beyni görsel uyaranlara daha iyi tepki vermektedir. İnsan beyni görsel bir veriyi, normal bir metinden 60.000 kat daha hızlı işlemektedir. Veri görselleştirme yöntemleriyle çok boyutlu veriler 2 veya 3 boyuta indirgenir ve daha kolay anlaşılır hale getirilir (Çelik ve Akdamar, 2018). Veri görselleştirme, analiz sonuçlarının gösterilmesi için en etkili yollardandır. Veri analizi gerçekleştirilirken her bir adım için görsel yorumlama yapmak mümkündür. Veri görselleştirmenin farklı yolları bulunmaktadır (Eken, 2020).

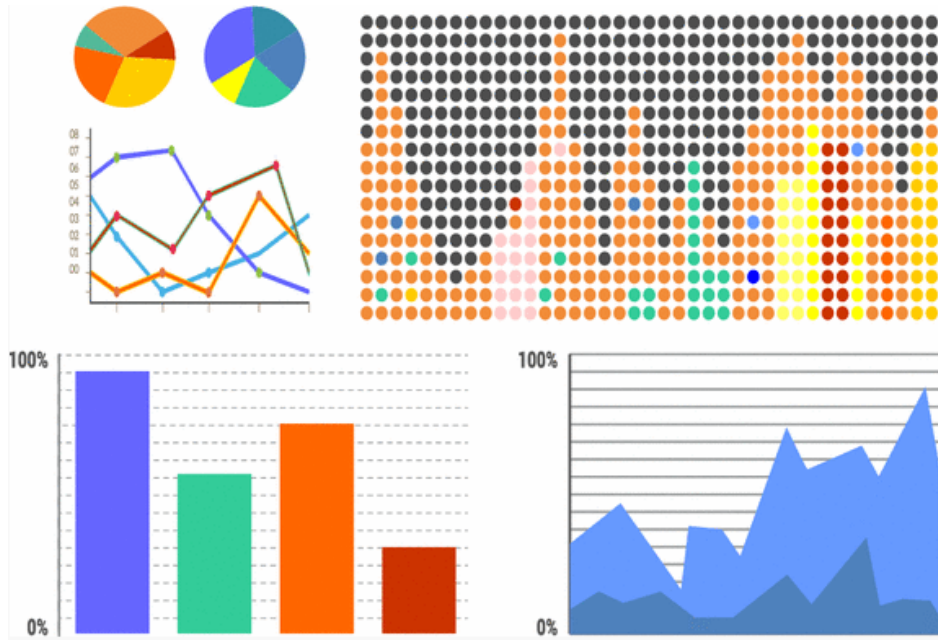
Hareket Grafikleri: Çok değişkenli ve büyük miktardaki verilerin analiz edilmesini sağlar (Çelik ve Akdamar, 2018).

Şekil 4. Hareket Grafikleri



Sözcük Bulut: Bir metinde geçen kelimeleri frekansına göre tabloştıran bir yöntemdir (Çelik ve Akdamar, 2018).

Şekil 7. Gösterge Tablosu



Büyük veri kavramı, kelime anlamından çok daha fazlasını ifade etmektedir. Büyük veriyi toplamak, depolamak, analizini yapmak, sonuçlardan tahminler yapmak gibi süreçlerin tümü büyük veri olarak ifade edilebilir. Büyük veri, 100 terebayttan 2 petabayta kadar büyüklüğü değişen, yapılandırılmamış verilerdir. Bu büyüklükteki veriler için uygun analiz programları gerekmektedir. Büyük verilerin analiz edilmesi sonucu; anlaşılır bilgilerin sunulması ve gerekli mesajların iletilmesinde en önemli yaklaşımlardan birisi görselleştirmedir (Erkurt, 2021).

SONUÇ

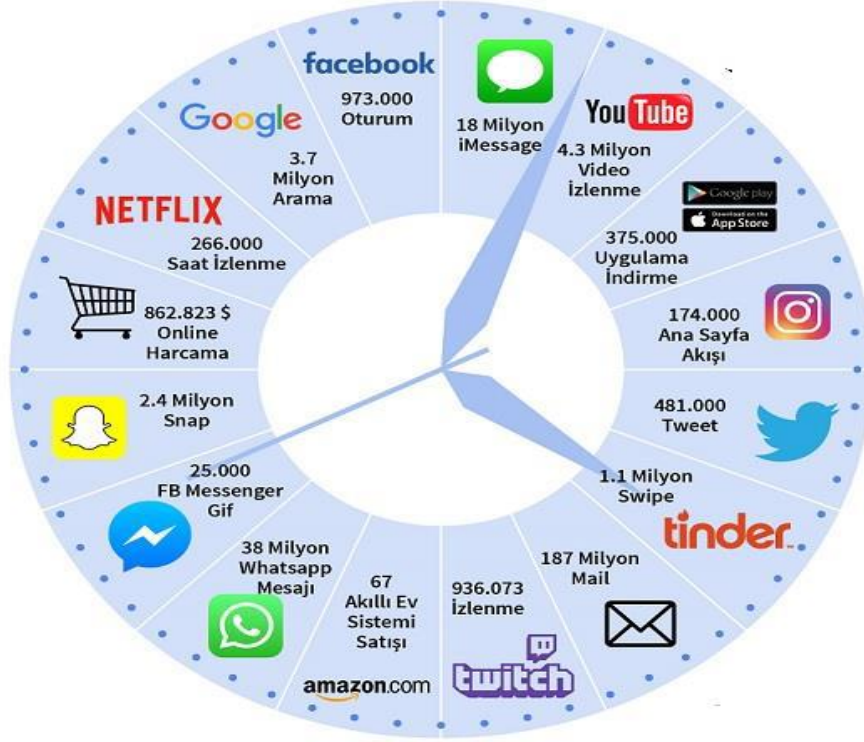
Büyük veri; sensörler, akıllı cihazlar, akıllı telefonlar, sosyal medya, internet ve dijital dünyadan gelen verileri ifade eder (Yang vd., 2016). Büyük veriye birden farklı yöntemle ulaşılabilir. Bunlar (Oktay, 2020):

- Büyük kuruluşların sunucularından doğrudan paylaşılan veriler,
- Uygulama ara yüzleri,
- Web sayfalarından toplanan verilerdir.

Büyük veri analizi ile verilerin daha hızlı ve düzgün şekilde analizi yapılabilmektedir. Standart veri analiz teknikleri ile analizi yapılamayan büyük miktardaki verilerin analizi büyük veri analiz teknikleri ile yapılabilmektedir. Büyük veri analizi ile büyük hacimli verileri ayrıştırmak, depolamak, modellemek ve sonuç çıkarmak kolaylaşmıştır (Altındış ve Kıran Morkoç, 2018). Büyük veri analizinin faydalarını fark eden birçok şirket büyük veri teknolojilerinden faydalanmaya başlamıştır. İletişimden ticarete, sağlıktan ulaştırmaya kadar birçok sektörde de büyük veriden faydalanılmaktadır. Ayrıca devletler rutin yapılan birçok işlemde büyük veriyi kullanmaktadırlar. Büyük veri analizi ile çok büyük yararlar sağlayan işletmeler hem rekabet avantajı elde etmekte hem etkili ve verimli müşteri ilişkileri yönetimi gerçekleştirmekte hem de kârlılıklarını artırabilmektedirler. Büyük verinin elde edilmesinde bazı güvensizlikler olabilmektedir. Verilerin günden güne artmasıyla ve teknolojinin sürekli gelişmesiyle beraber veri güvenliği çok önemli bir konu haline gelmiştir. Farklı mecralardan toplanan verilerin kötü niyetli insanların eline geçmesiyle istenmeyen sonuçlar oluşabilir. Bu sebeple büyük veri ile ilgilenen işletme ve kurumların çok dikkatli olması gerekmektedir (Aksu ve Güzeller, 2019).

Bu çalışmada büyük veri kavramına, büyük veri kavramının içinde büyük veri bileşenlerine yer verilmiştir. Bir sonraki kısımda büyük veri analizi detaylıca incelenmiş; büyük veri analiz ve işleme platformları, büyük veri analizinin faydaları ve büyük veri analizinin zorlukları anlatılmıştır. Üçüncü bölümde büyük verinin kullanıldığı sektörler ve uygulama alanları ile büyük veri kullanım örnekleri açıklanmıştır. Daha sonra veri görselleştirme ve sonuç bölümlerine yer verilmiştir. Bu çalışmanın büyük veri literatüründe açıklayıcı bir çalışma olması beklenmektedir.

Şekil 8. İnternette 1 Dakikada Neler oluyor?



KAYNAKÇA

- Aksoy B., Bayrakçı H. C., Bayrakçı E., Uğuz S. (2017). Büyük Verinin Kurumlarda Kullanımı. Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences, 22.
- Aksu G., Güzeller C. O. (2019). Büyük Veri: Sosyal Bilimler ile Eğitim Bilimlerinde Kullanımı ve Uygulama Alanları.
- Aktan E. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu.
- Alharthi A., Krotov V., Bowman M. (2017). Addressing Barriers to Big Data. Business Horizons, 60(3), 285-292.
- Altındış S., Kıran Morkoç İ. (2018). Sağlık Hizmetlerinde Büyük Veri. Academic Review of Economics & Administrative Sciences, 11(2).
- Altunışık R. (2015). Büyük Veri: Fırsatlar Kaynağı mı Yoksa Yeni Sorunlar Yumağı mı?. Yıldız Social Science Review, 1(1), 45-76.
- Ayvaz S., Salman Y. B. (2020). Türkiye’de Firmaların Büyük Veri Teknolojileri Bilinirliği ve Kullanımı Analizi. Avrupa Bilim ve Teknoloji Dergisi, (18), 728-737.
- Bozkurt A. (2016). Öğrenme Analitiği: E-öğrenme, Büyük Veri ve Bireyselleştirilmiş Öğrenme. Açıköğretim Uygulamaları ve Araştırmaları Dergisi, 2(4), 55-81.
- Çelik S., Akdamar E. (2018). Büyük veri ve veri görselleştirme. Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi, (65), 253-264.
- Dai H. N., Wang H., Xu G., Wan J., Imran M. (2019). Big Data Analytics for Manufacturing Internet of Things: Opportunities, Challenges and Enabling Technologies. Enterprise Information Systems, 1-25.

- Demirel D., Das R., Hanbay D. (2019, September). Büyük Veri Üzerine Perspektif Bir Bakış. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-9). IEEE.
- Doğan K., Arslantekin S. (2016). Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum. Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi, 56(1).
- Doğan M. (2014). Büyük Veri'nin kişiler ve kurumlar üzerindeki etkileri (Doctoral dissertation, İstanbul Bilgi Üniversitesi).
- Eken S. (2020). Büyük Verinin İnteraktif Görselleştirilmesi: Tableau Üzerine Öğrenci Deneyimleri. Avrupa Bilim ve Teknoloji Dergisi, (18), 262-271.
- Erkurt E., Yıldırım E. (2021). Bir büyük veri görselleştirme uygulaması olarak konut tercih infografikleri. Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 23(1), 36-52.
- Jan B., Farman H., Khan M., Imran M., Islam I. U., Ahmad A., Jeon G. (2019). Deep Learning in Big Data Analytics: A Comparative Study. Computers & Electrical Engineering, 75, 275-287.
- Kaisler S., Armour F., Espinosa J. A., Money, W. (2013, January). Big data: Issues and Challenges Moving Forward. In 2013 46th Hawaii International Conference on System Sciences (pp. 995-1004). IEEE.
- Karabay, B., & Ulaş, M. (2017). Büyük Veri İşlemede Yaygın Kullanılan Araçların Karşılaştırılması. In 8th International Advanced Technologies Symposium (pp. 1-18).
- Katal A., Wazid M., Goudar R. H. (2013, August). Big Data: Issues, Challenges, Tools and Good Practices. In 2013 Sixth International Conference on Contemporary Computing (IC3) (pp. 404-409). IEEE.
- Kayabay K., Gökalp M. O., Akyol M. A., Koçyiğit A., Eren P. E. (2016). Big Data for Future Enterprises: Current State and Trends. In 3rd International Management Information Systems Conference, İzmir (pp. 298-307).
- Labrinidis A., Jagadish H. V. (2012). Challenges and Opportunities With Big Data. Proceedings of the VLDB Endowment, 5(12), 2032-2033.
- Oktay H. T. (2020). Büyük Veri Çağında Sosyal Medya Verilerinin Sosyal Bilimler İçin Önemi; Twitter Örneği. Üçüncü Sektör Sosyal Ekonomik İnceleme , 55 (2), 1090-1110.
- Özbilgin İ. G. (2015). Kamuda Büyük Veri ve Uygulamaları. AB2015 Akademik Bilişim Konferansı, 31.
- Özcan A. (2021). Büyük veri: Fırsatlar ve tehditler. TRT Akademi, 6(11), 10-31.
- Özdemir İ., Sağiroğlu Ş. (2018). Denetimlerde Büyük Veri Kullanımı ve Üzerine Bir Değerlendirme. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 6(2), 470-480.
- Russom P. (2011). Big Data Analytics. TDWI Best Practices Report, Fourth Quarter, 19(4), 1-34.
- Schönberger, V. M., & Cukier, K. (2013). Büyük Veri-Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim. Çeviren: B. Erol, Paloma Yayınları, İstanbul, Türkiye.
- Tabesh P., Mousavidin E., Hasani S. (2019). Implementing Big Data Strategies: A Managerial Perspective. Business Horizons, 62(3), 347-358.
- Yang C., Huang Q., Li Z., Liu K., Hu, F. (2017). Big Data and Cloud Computing: Innovation Opportunities and Challenges. International Journal of Digital Earth, 10(1), 13-53.
- Yıldırım E., Erkurt E. (2020). Büyük Veri Görselleştirme: Emlak Sektörüne İlişkin Bir Uygulama. Ankara Hacı Bayram Veli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 38-57.