

T.C.
AYDIN ADNAN MENDERES ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
UYGULAMALI EKONOMETRİ PROGRAMI
2020-YL 010

ULUSLARARASI ÇATIŞMALARIN SOSYAL AĞ ANALİZİ
İLE İNCELENMESİ


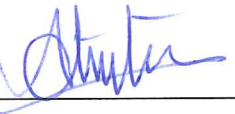
HAZIRLAYAN
Zühal MOLLAOĞULLARI

TEZ DANIŞMANI
Dr. Öğr. Üyesi Ahmet ÜNLÜ

AYDIN-2020

T.C.
AYDIN ADNAN MENDERES ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE
AYDIN

Ekonometri Ana Bilim Dalı Uygulamalı Ekonometri Programı öğrencisi **Zühal Mollaoğulları** tarafından hazırlanan **Uluslararası Çatışmaların Sosyal Ağ Analizi İle İncelenmesi** başlıklı tez, 20/12/2019 tarihinde yapılan savunma sonucunda aşağıda isimleri bulunan jüri üyelerince kabul edilmiştir.

| Ünvanı, Adı ve Soyadı | Kurumu | İmzası |
|----------------------------|--|--|
| Prof. Dr. Cem DİŞBUDAK | Muğla Sıtkı Koçman Üniversitesi İİBF İktisat Bölümü |  |
| Dr. Öğr. Üyesi Ahmet ÜNLÜ | Aydın Adnan Menderes Üniversitesi Nazilli İİBF Ekonometri Bölümü |  |
| Dr. Öğr. Üyesi Engin ÇAKIR | Aydın Adnan Menderes Üniversitesi Nazilli İİBF İşletme Bölümü | |

Jüri üyeleri tarafından kabul edilen bu (Tezin Türü) tezi, Enstitü Yönetim Kurulununsayılı kararıyla(Tarih) tarihinde onaylanmıştır.

Prof. Dr. Ahmet Can BAKKALCI
Enstitü Müdürü V.

T.C.
AYDIN ADNAN MENDERES ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE
AYDIN

Bu tezde sunulan tüm bilgi ve sonuçların, bilimsel yöntemlerle yürütülen gerçek deney ve gözlemler çerçevesinde tarafımdan elde edildiğini, çalışmada bana ait olmayan tüm veri, düşünce, sonuç ve bilgilere bilimsel etik kuralların gereği olarak eksiksiz şekilde uygun atıf yaptığımı ve kaynak göstererek belirttiğimi beyan ederim.

20/01/2020

Zühal MOLLAOĞULLARI

ÖZET

ULUSLARARASI ÇATIŞMALARIN SOSYAL AĞ ANALİZİ İLE İNCELENMESİ

Zühal MOLLAOĞULLARI

Yüksek Lisans Tezi, Uygulamalı Ekonometri Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Ahmet ÜNLÜ

2020, XIV + 72 sayfa

Sosyal Ağ Analizi (SAA), günümüzün popüler çalışma alanlarından biridir. Ağ metodolojilerini meydana getiren ilişkileri temel alan SAA ile olaylara çok farklı bakış açıları geliştirilebilir. Bu çalışmanın temel amacı, sosyal ağ analizini yöntemiyle dünyadaki çatışmaların incelenmesidir. Araştırma kapsamında kullanılan veriler Global Database of Events Language and Tone (GDELT) projesi kapsamındaki ülkelerin çatışma sayılarına ilişkin verilerdir. Araştırma kapsamında betimsel istatistikler, merkezîyet ölçüleri ve kümeleme analizi kullanılmıştır. Yapılan analizler sonucunda bugün için dünyadaki çatışmaların merkezindeki en büyük aktörün ABD olduğu bulunmuştur. ABD'den sonra sırasıyla başlıca aktörlerin Rusya, Irak, İsrail, İran, Filistin olduğu görülmektedir. En az çatışma içerisinde bulunan ülkelerin ise sırasıyla Çin, Britanya, Peru, Doğu Timor, Endonezya ve Eritrea (Eritre) olduğu görülmektedir. Aynı zamanda sosyal ağ analizi ile oluşturulan sosyal ağın yedi (7) farklı kümeden meydana geldiği belirlenmiştir. Çatışmaların genel itibarıyla büyük bir bölümünün Ortadoğu'da gerçekleştiği ve Ortadoğu'nun her anlamda dünya siyasetinde çok önemli bir bölge olduğu sonucuna ulaşılmıştır.

ANAHTAR SÖZCÜKLER: Sosyal Ağ Analizi, Uluslararası Çatışmalar, UCINET, Büyük Veri.

ABSTRACT

INVESTIGATION OF INTERNATIONAL CONFLICTS WITH SOCIAL NETWORK ANALYSIS

Zühal MOLLAOĞULLARI

Master Thesis, at Applied Econometrics

Supervisor: Asst. Prof. Ahmet ÜNLÜ

2020, XIV + 72 pages

The main purpose of this study is to investigate the conflicts in the world by using social network analysis method. The data used in the research are data on the number of conflicts in the countries covered by the GDELT project. As a result of the analysis, it was found that the US is the biggest actor at the center of the conflicts in the world. Russia, Iraq, Israel, Iran and Palestine are the main actors after the US. The least conflicting countries are China, Britain, Peru, East Timor, Indonesia and Eritrea, respectively. At the same time, it is determined that the social network formed by social network analysis consists of 7 different clusters. It is concluded that the majority of the conflicts took place in the Middle East and that the Middle East was a very important region in the world politics in every sense.

KEYWORDS: Social Network Analysis, International Conflicts, UCINET, Big Data.

ÖNSÖZ

Günümüzde, yaptığımız hemen hemen her eylem dijital ortamda bir iz bırakmaktadır. Çevrimiçi olduğumuzda, GPS donanımlı akıllı telefonlarımızı taşıdığımızda, arkadaşlarımızla sosyal medya veya sohbet uygulamaları aracılığıyla iletişim kurduğumuzda ve alışveriş yaptığımızda veri üretilmektedir. Bu ve benzeri birçok günlük faaliyetle dijital ayak izi bıraktığımızı söyleyebiliriz ki bu durum her şey anlamına gelmektedir. Bunun üzerine, makine tarafından üretilen veri miktarı da hızla artmaktadır. “Akıllı” ev cihazlarımız birbirleriyle veya ev sunucularıyla iletişim kurduğunda veriler oluşturulur ve paylaşılır. Dünyanın dört bir yanındaki üretim ortamlarında bulunan endüstriyel makine, giderek veri toplayan ve ileten sensörlerle donatılmıştır. Büyük Veri terimi, tüm bu bahsedilen verilerin toplanmasını ve bunları iş dahil olmak üzere çok çeşitli yaşama dair kolaylıklara yönelik kullanma yeteneğimizi ifade etmektedir. Bu aşamada eldeki büyük veriden anlamlı sonuçlar çıkarmak büyük önem taşımaktadır.

Bu araştırma dört bölümden oluşmaktadır. Birinci bölüm giriş (problem durumu, amaç, önem, varsayımlar, sınırlılıklar, tanımlar, ilgili yayınlar); ikinci bölüm yöntem (araştırma modeli, evren ve örneklem, veri toplama araçları, verilerin toplanması ve analizi); üçüncü bölüm bulgular ve dördüncü bölüm sonuç, tartışma ve önerilerden oluşmaktadır.

Bu çalışmanın gerçekleşmesinde değerli birçok insanın emeği bulunmaktadır. Bu çalışmanın gerçekleşmesi için çalışmamın her aşamasında görüş ve önerileriyle bana destek olan, engin bilgi ve tecrübesiyle bana yol gösteren, yardımlarını hiçbir zaman esirgemeyen değerli danışman hocam Dr. Öğr. Üyesi Ahmet Ünlü’ye en derin saygı ve teşekkürlerimi sunuyorum.

Yüksek lisansa başladığım ilk günden bu güne kadar sıkıntılı zamanlarımda desteğini hiçbir zaman esirgemeyen, tez sürecinde benimle bilgi ve deneyimlerini paylaşan sevgili eşim Törehan MOLLAOĞULLARI’na sonsuz teşekkürlerimi sunarım.

Zühal MOLLAOĞULLARI

İÇİNDEKİLER

| | |
|---|------|
| KABUL VE ONAY SAYFASI..... | iii |
| BİLİMSEL ETİK BİLDİRİM SAYFASI..... | iv |
| ÖZET | v |
| ABSTRACT | vi |
| ÖNSÖZ..... | vii |
| SİMGELER DİZİNİ..... | xi |
| ŞEKİLLER DİZİNİ | xii |
| TABLolar DİZİNİ..... | xiii |
| KISALTMALAR DİZİNİ | xiv |
| GİRİŞ..... | 1 |
| 1. BÖLÜM | 6 |
| 1. KURAMSAL VE KAVRAMSAL ÇERÇEVE..... | 6 |
| 1.1. Veri Bilimi Nedir? | 6 |
| 1.2. Veri Bilimcisi Kimdir, Ne Yapar? | 7 |
| 1.3. Veri Analisti..... | 8 |
| 1.4. Büyük Veri Yaşam Döngüsü | 10 |
| 1.5. Büyük Veri Nedir?..... | 13 |
| 1.6. Büyük Veri'nin 3V'si | 14 |
| 1.6.1. Veri Büyüklüğü (Volume)..... | 14 |
| 1.6.2. Hız (Velocity) | 14 |
| 1.6.3. Çeşitlilik (Variety) | 14 |
| 1.6.4. Doğrulama (Validation)..... | 15 |
| 1.6.5. Değer (Value) | 15 |
| 1.7. Büyük Veri Analizi Nedir? | 16 |
| 1.8. Veri Türleri | 17 |
| 1.8.1. Yapılandırılmış Veri | 17 |

| | |
|---|----|
| 1.8.2. Yapılandırılmamış Veri | 17 |
| 1.8.3. Yarı Yapılandırılmış Veri | 18 |
| 1.9. Veri Kaynağı Nedir? | 19 |
| 1.9.1. Veri Kaynakları Türleri | 19 |
| 1.9.1.1. Makine veri kaynakları..... | 19 |
| 1.9.1.2. Dosya veri kaynakları..... | 20 |
| 1.10. Büyük Veri Uygulamaları..... | 20 |
| 1.11. Büyük Veri Teknolojileri..... | 22 |
| 1.12. Hadoop Nedir? | 23 |
| 1.13. ApacheSpark..... | 25 |
| 1.14. R Programlama Nedir? | 25 |
| 1.15. Bulut Bilişim..... | 28 |
| 1.16. SQL Veritabanı | 32 |
| 1.17. NoSQLVeritabanı | 32 |
| 1.18. BigQuery..... | 33 |
| 1.19. Büyük Veri Analizinde Kullanılan Teknikler..... | 36 |
| 1.19.1. A/B Testi..... | 36 |
| 1.19.2. Doğal Dil İşleme..... | 37 |
| 1.19.3. Veri Madenciliği..... | 37 |
| 1.19.4. Kümeleme Analizi | 38 |
| 1.19.5. Sınıflandırma | 39 |
| 1.19.6. Kalabalığın Gücü | 39 |
| 1.19.7. Makine Öğrenmesi..... | 39 |
| 1.19.8. Sinir Ağları | 39 |
| 1.19.9. Sosyal Ağ Analizi | 39 |
| 1.19.10. Örüntü Tanıma..... | 39 |
| 1.19.11. Zaman Serileri | 40 |

| | |
|--|-----------|
| 1.19.12. Duygu Analizi..... | 40 |
| 1.19.13. Veri Görselleştirme..... | 41 |
| 1.19.14. Denetimli Öğrenme | 41 |
| 1.19.15. Denetimsiz Öğrenme | 42 |
| 2. BÖLÜM | 43 |
| 2. YÖNTEM | 43 |
| 2.1. Veri Seti (GDELT Projesi) | 43 |
| 2.2. Literatür Taraması..... | 44 |
| 2.3. Sosyal Ağ Analizi (SAA) | 46 |
| 2.4. Merkezilik Ölçütleri..... | 49 |
| 3. BÖLÜM | 51 |
| 3. BULGULAR | 51 |
| 3.1. Ülkeler Arasında Yaşanan Çatışmalara İlişkin Veri Setinin Elde Edilmesi | 51 |
| 3.2. Ülkeler Arasında Yaşanan Çatışmaların Sosyal Ağ Analizi ile İncelenmesi | 53 |
| 3.3. Merkezilik Ölçütleri Bakımından Sosyal Ağ Özellikler..... | 53 |
| 4. SONUÇ, TARTIŞMA VE ÖNERİLER | 61 |
| 4.1. Öneriler | 64 |
| 5. KAYNAKLAR | 65 |
| ÖZGEÇMİŞ | 72 |

SİMGELER DİZİNİ

α : Ton yayılımı yakınsama hız katsayısı

β : Ton yayılımı komşu düğümlerin yakınsama hız katsayısı

μ : Ortalama

σ Sigma : Standart Sapma



ŞEKİLLER DİZİNİ

| | |
|--|----|
| Şekil 1.1. Veri Biliminin Aşamaları | 9 |
| Şekil 1.2. Veri Bilim Yaşam Döngüsü | 11 |
| Şekil 1.3. İstatistiksel Analiz Akışı | 11 |
| Şekil 1.4. Model Planlama Araçları..... | 12 |
| Şekil 1.5. Model Oluşturma Araçları | 13 |
| Şekil 1.6. Yapılandırılmamış Veri Örneği..... | 18 |
| Şekil 2.1. Düğümleri ve Kenarları Gösteren Basit Bir Sosyal Ağ | 48 |
| Şekil 3.1. Veri Çekimine İlişkin Komut Dizinleri..... | 52 |
| Şekil 3.2. Ülkelerin Çatışma Sayılarına İlişkin Histogram Grafiği..... | 52 |
| Şekil 3.3. Ülkelerarası Çatışma Çizgi Grafiği | 53 |
| Şekil 3.4. Ülkelere İlişkin Sosyal Ağ | 54 |
| Şekil 3.5. Amerika Dışındaki Ülkelere İlişkin Sosyal Ağ..... | 55 |
| Şekil 3.6. Ego Ağına İlişkin Görsel..... | 56 |
| Şekil 3.7. Girvan-Newman Algoritmasına Göre Küme Sayıları..... | 59 |
| Şekil 3.8. Sosyal Ağda Yer Alan Ülkelerin Girvan-Newman Algoritmasına Göre Kümelene Durumu..... | 59 |

TABLolar DİZİNİ

| | |
|---|----|
| Tablo 2.1. Komşuluk Matrisi..... | 48 |
| Tablo 3.1. Ülkeler ve Ülkelerin Türkçe Karşılıkları | 51 |
| Tablo 3.2. Ülkelerin ve Bağlantılarının Oluşturduğu Ağın Merkezilik Ölçütleri | 57 |



KISALTMALAR DİZİNİ

| | |
|----------|---|
| ABD | : Amerika Birleşik Devletleri |
| AWS | : Amazon Web Services |
| BI | : Business Intelligence – İş Zekası |
| Big Data | : Büyük Veri |
| BT | : Bilişim Teknolojileri |
| EMR | : ElasticMapReduce |
| ETLT | : (Eliminate, Transform, Load, Transform: Ayıkla, Dönüştür, Yükle ve Aktar) |
| GDELT | : Global Database of Events, Language (Location) and Tone |
| HDFS | : Hadoop Dağıtılmış Dosya Sisteminde |
| HDFS | : Hadoop Distributed File System – Hadoop Dağıtık Dosya Sistemi |
| IoT | : Nesnelerin İnterneti |
| ML | : Makine öğrenimi |
| NLP | : Doğal Dil İşleme |
| NoSQL | : Not Only SQL – Yalnızca SQL değil – İlişkisel Veritabanı Yönetim Sistemi |
| PB | : PetaBayt |
| SAA | : Sosyal Ağ Analizi |
| SQL | : Structured Query Language – Yapılandırılmış Sorgu Dili |
| UCINET | : Ağ analizinin gerçekleştirilebildiği kapalı sistem istatistik programı. |
| ZB | : Zettabayt |

GİRİŞ

Araştırmanın Amacı ve Önemi

Bu çalışmanın temel amacı, sosyal ağ analizi yöntemiyle dünyadaki çatışmaların incelenmesidir. Bu çalışma, dünyadaki çatışmaların sosyal ağ analizi ile incelenmesi vasıtasıyla ülkelerin çatışma anlamında dünya siyasetinde nasıl bir konumda oldukları, hangi ülkelerin uluslararası çatışmalar söz konusu olduğunda başat aktör konumunda olduğunun belirlenmesi anlamında önem taşımaktadır. Çalışma aynı zamanda, ülkelerin dünya siyasetinde yer aldıkları konumda bulunma sebeplerine ilişkin bir tür delil toplama süreci olarak görülebilir ve bu noktada bu özellikler bu çalışmanın neden önemli olduğunu ortaya koymaktadır. Araştırma kapsamında GDELT projesi kapsamındaki ülkelerin çatışma sayılarına ilişkin veriler kullanılmıştır. Veri setinde birbirleri ile çatışma içerisinde olan ülkeler, çatışma tarihleri ve çatışma sayıları yer almaktadır. Çatışma sayılarının çok ayrıntılı olmasa da tarih değişkeni göz önünde bulundurularak incelenmesi ve çalışma ile oluşturulan sosyal ağ içerisindeki kümelerin belirlenmesi araştırmanın alt amaçlarından bir tanesi olarak düşünülebilir. Çalışmanın ana amacına bağlı olarak alt amaçlar aşağıda listelenmiştir.

- Uluslararası çatışmaların tarih aralıkları göz önünde bulundurulduğunda nasıl bir dağılım gösterdiğini belirlemektir.
- Uluslararası çatışma sayıları ve ülkeler kullanılarak elde edilen sosyal ağda aktör konumunda olan ülkelerin belirlenmesi
- Uluslararası çatışma sayıları ve ülkeler kullanılarak elde edilen sosyal ağda izole konumda olan ülkelerin belirlenmesi
- Uluslararası çatışma sayıları ve ülkeler kullanılarak elde edilen sosyal ağın kaç alt gruptan (kümeden) meydana geldiğinin belirlenmesi

Araştırma Problemi

Günümüzde ülkeler hakkında farklı ortamlarda ve farklı amaçlarla çok fazla veri toplanmaktadır. Uluslararası ilişkiler disiplini, sosyal ağlarla çok daha net bir şekilde sebep-sonuç anlamında anamca genişletilebilecek bir yapıya sahiptir. Ülkelerin herhangi bir sebepten ötürü birbirleriyle çatışma içerisinde olması ülkelere ilişkin oluşturulacak olan sosyal ağın yapısını rahatlıkla etkilemektedir. Adı ister çatışma ister ilişki olsun ülkeler arası etkileşimlerde aktör ülkeler ve oluşumda herhangi bir paya sahip olmayan ülkeler mevcuttur.

Sosyal ağlar incelenerek ülkelerin birbirleriyle olan olumlu ya da olumsuz ilişkiler rahatlıkla ortaya konabilmekte, değişken sayısı arttıkça ülkeler arası ilişkilerin şiddeti ve yönüne ilişkin net yorumlar yapılabilmektedir. Bu araştırmada, GDELT Project üzerinden çekilen verilerle gerçekleştirilen sosyal ağ analizi sonucunda çatışmalarda rol oynayan aktör ülkeler, rol oynamayan izole ülkeler, çatışmaların daha çok hangi bölgede gerçekleştiği ve oluşturulan ağdaki ideal küme sayısı nedir sorularına cevap aranmaktadır.

Araştırmanın Alt Problemi

Dünyadaki belli başlı çatışmalar nasıl bir dağılım göstermektedir?

Alt problemler. Araştırmanın ana problem cümlesine bağlı olarak oluşturulan alt problemler aşağıda listelenmiştir.

1. Uluslararası çatışmalar tarih aralıkları göz önünde bulundurulduğunda nasıl bir dağılım göstermektedir?
2. Uluslararası çatışma sayıları ve ülkeler kullanılarak elde edilen sosyal ağda aktör konumunda olan ülkeler hangileridir?
3. Uluslararası çatışma sayıları ve ülkeler kullanılarak elde edilen sosyal ağda izole konumda olan ülkeler hangileridir?
4. Uluslararası çatışma sayıları ve ülkeler kullanılarak elde edilen sosyal ağ kaç alt gruptan (kümeden) meydana gelmektedir?

Varsayımlar

Bu araştırmanın varsayımları aşağıda maddeler halinde sıralanmıştır.

1. GDELT projesi kapsamında yer alan veriler güvenilirdir. Bu verilerin analiz edilmesinde kullanılacak UCINET in gerekli analiz yeteneğine sahip olduğu varsayılmıştır. Dolayısıyla elde edilen sonuçların güvenilirlik sorunu olmayacaktır.

Sınırlılıklar

Bu araştırmanın sınırlılıkları aşağıda maddeler halinde sıralanmıştır.

1. Çalışma kapsamında kullanılan veri setinde sınırlı sayıda ülke yer almaktadır.

2. Çalışma kapsamında sadece tarih, ülke ve çatışma sayı değişkenleri kullanılmıştır.

Tanımlar

Sosyal ağ analizi:Sosyal ağ analizi, toplumsal yapıları ağ ve çizge (graf) teorileri üzerinden incelemeye yarayan analiz türüdür.

Uluslararası çatışmalar:Uyumsuz amaçların bulunduğu ya da bulunduğu inanan en az iki ülke arasındaki ilişki.

UCINET: Sosyal ağ analizinin gerçekleştirilebildiği kapalı sistem istatistik programı.

Büyük veri: Büyük veri, artan hacimlerde ve her zamankinden daha yüksek hıza ulaşan daha fazla çeşitlilik içeren verilerdir.

Her ne kadar büyük veri kavramı yeni olsa da, büyük veri setlerinin kökenleri, ilk veri merkezleri ve ilişkisel veri tabanının geliştirilmesiyle, veri dünyasının yeni başladığı 1960'lı ve 70'li yıllara dayanmaktadır. 2005 yılı civarında, insanlar Facebook, YouTube ve diğer çevrimiçi hizmetler yoluyla ne kadar veri ürettiğini fark etmeye başladılar. Aynı yıl Hadoop (özellikle büyük veri setlerini depolamak ve analiz etmek için oluşturulan açık kaynaklı bir çerçeve) geliştirildi. NoSQL ayrıca bu sırada popülerlik kazanmaya başladı. Hadoop (ve daha son zamanlarda Spark) gibi açık kaynaklı çerçevelerin geliştirilmesi, büyük verilerin büyümesi için çok önemliydi çünkü Hadoop sayesinde büyük veriler ile çalışılmasını daha kolay ve depolamasını daha ucuz yapılmasını sağlamaktadır. O zamandan beri, büyük verilerin hacmi arttı. Kullanıcılar hala çok büyük miktarlarda veri oluşturuyor. Ancak bunu yapan insanlar değildir. Nesnelerin İnterneti'nin (IoT) ortaya çıkmasıyla, müşteriye kullanım şekilleri ve ürün performansı hakkında veri toplayarak internete daha fazla nesne ve cihaz bağlanmıştır (Oracle, 2019).

Her şey, dijital çağın başlangıcından beri ürettiğimiz veri miktarındaki patlama ile başlar. Bu büyük ölçüde bilgisayarların, İnternet'in ve içinde yaşadığımız dünyadan veri toplayabilen teknolojinin yükselmesinden kaynaklanmaktadır. Kendi içindeki veriler yeni bir icat değildir. Bilgisayarlar ve veritabanlarından önce bile geriye dönersek, kâğıt işlem kayıtlarımız, müşteri kayıtlarımız ve arşiv dosyalarımız vardı. Aslında hepsi birer veriydi. Bilgisayarlar ve özellikle elektronik tablolar ve veritabanları, verileri kolayca ve kolayca erişilebilir bir şekilde depolamak ve düzenlemek için bir yol sağladı. Aniden, bir fare tıklamasıyla bilgiler büyük bir hızla artmaya başladı.

Yine de, erken elektronik tablolardan ve veritabanlarından beri çok yol kat edildi. Bugün, iki günde bir, zamanın başlangıcından 2000 yılına kadar yaptığımız kadar veri yaratıyoruz. Ve yarattığımız veri miktarı hızla artmaya devam ediyor; 2020 yılına kadar, mevcut dijital bilgi miktarı bugün yaklaşık 5 zettabayttan 50 zettabayt'a yükselmiş olacak.

Günümüzde, yaptığımız hemen hemen her eylem dijital bir iz bırakıyor. Çevrimiçi olduğumuzda, GPS donanımlı akıllı telefonlarımızı taşıdığımızda, arkadaşlarımızla sosyal medya veya sohbet uygulamaları aracılığıyla iletişim kurduğumuzda ve alışveriş yaptığımızda veri üretiyoruz. Dijital ayak izlerini, yaptığımız her şeyle dijital ayak izi bıraktığımızı söyleyebiliriz ki bu hemen hemen her şeydir. Bunun üzerine, makine tarafından üretilen veri miktarı da hızla artmaktadır. "Akıllı" ev cihazlarımız birbirleriyle veya ev sunucularıyla iletişim kurduğunda veriler oluşturulur ve paylaşılır. Dünyanın dört bir yanındaki fabrikalarda ve fabrikalarda bulunan endüstriyel makine, giderek veri toplayan ve ileten sensörlerle donatılmıştır.

"Büyük Veri" terimi, tüm bu verilerin toplanmasını ve bunları iş dahil olmak üzere çok çeşitli alanlarda avantajımıza kullanma yeteneğimizi ifade eder (Marr, 2019).

İşletme için Büyük Veri

Büyük Veri veritabanları ile işletmeler, para tasarrufu yapabilir, gelirlerini artırabilir ve diğer birçok iş hedefine ulaşabilir (Davenport and Harris 2007).

- Yeni uygulamalar oluşturur: Büyük veriler bir şirketin ürünleri, kaynakları veya müşterileri üzerinde milyarlarca gerçek zamanlı veri noktası toplamasına ve ardından müşteri deneyimini veya kaynak kullanımını en iyi duruma getirmek için bu verileri anında yeniden paketlemesine izin verebilir. Örneğin, büyük bir ABD şehri, coğrafi verileri 30'dan fazla farklı departmandan gerçek zamanlı olarak toplayıp analiz ederek suçları azaltmak ve belediye hizmetlerini iyileştirmek için MongoDB kullanıyor.
- Etkinliği artırır ve mevcut uygulamaların maliyetini düşürür: Büyük veri teknolojileri, yüksek düzeyde özelleştirilmiş, pahalı eski sistemleri, mal donanımı üzerinde çalışan standart bir çözümle değiştirebilir. Ve birçok büyük veri teknolojisinin kullanımı açık ve ücretsiz olduğundan, tescilli teknolojilerden çok daha ucuza uygulanabilir. Örneğin, referans veri yönetimi uygulamasını MongoDB'ye geçirerek, bir Kademe 1 bankası, daha önce koştugu tescilli ilişkisel veritabanına ilişkin lisans ve donanım

maliyetlerini önemli ölçüde düşürürken, aynı zamanda başvurusunu düzenleme gereklilikleri ile daha iyi bir uyumluluğa getirmiştir.

- Yeni rekabet kaynakları gerçekleştirir: Büyük veriler, işletmelerin rakiplerinden daha hızlı değişiklik yapmalarına olanak tanıyarak daha çevik davranmalarına yardımcı olabilir. Örneğin, MongoDB en geniş İnsan Kaynakları Yönetimi (HCM) çözüm sağlayıcılarından birinin çok çeşitli farklı kaynaklardan gelen verileri entegre eden mobil uygulamaları hızla oluşturmaya izin verdi (Davenport and Harris 2007)..
- Müşteri sadakatini artırır: Kuruluş içinde paylaşılan veri miktarını (ve güncelleme hızını) artırmak, işletmelerin ve diğer kuruluşların müşteri talebine daha hızlı ve doğru şekilde yanıt vermelerini sağlar.



1. BÖLÜM

1. KURAMSAL VE KAVRAMSAL ÇERÇEVE

1.1. Veri Bilimi Nedir?

Veri Bilimi terimi günlük hayatta giderek daha fazla kullanılmaktadır. Ancak Veri Bilimi tam olarak ne anlama geliyor? Veri Bilimcisi olmak için hangi becerilere ihtiyaç vardır? İş Zekası ve Veri Bilimi arasındaki fark nedir? Veri Biliminde kararlar ve öngörüler nasıl yapılır? Bunlar cevaplanması gereken sorulardan sadece bir kaçıdır. Veri Bilimi, ham verilerden gizli kalıpları keşfetme amacıyla çeşitli araçların, algoritmaların ve makine öğrenme ilkelerinin bir karışımıdır (Gökçen, 2011).

Veri bilimi, bilginin nereden geldiği, neyi temsil ettiği ve İşletme-Bilişim Teknolojisine ilişkin stratejilerin oluşturulmasında nasıl değerli bir kaynağa dönüştürülebileceğinin incelenmesidir. Modelleri tanımlamak için büyük miktarda yapılandırılmış ve yapılandırılmamış veri madenciliği yapmak, bir kuruluşun maliyetlerinin azaltılmasına, verimliliğinin artırılmasına, yeni pazar fırsatlarının tanınmasına ve kuruluşun rekabet avantajını artırmasına yardımcı olabilmektedir (McGuire, Manyika ve Chui, 2012).

Veri biliminin kendine özgü yararları, şirketin hedefine ve sektörüne bağlı olarak değişmektedir. Örneğin, satış ve pazarlama departmanları, dönüşüm oranlarını iyileştirmek veya bire bir pazarlama kampanyaları oluşturmak için müşteri verilerini değerlendirebilmektedir. Yine bankacılık kurumları sahtekarlık tespitini geliştirmek için veri madenciliğinden yararlanmaktadır. Netflix gibi akış hizmetleri, kullanıcılarının ilgilendiklerini belirlemek için kullandığı verileri ve hangi TV programlarını veya filmlerini üreteceğini belirlemek için bu verileri kullanır. Veri tabanlı algoritmalar Netflix'te, kullanıcının görüntüleme geçmişini temel alan kişiselleştirilmiş öneriler oluşturmak için de kullanılır. DHL, FedEx ve UPS gibi sevkiyat şirketleri en iyi teslimat yollarını ve zamanlarını ve ayrıca gönderileri için en iyi ulaşım seçeneklerini bulmak için veri bilimini kullanır (Canbay, Vural ve Sağıroğlu, 2017).

Veri bilimi işletme içinde sürekli gelişmekte olan bir alandır, çünkü büyük boyutlarda ham (yapılandırılmamış) verilerin tanımlanması ve analizi şirketler için çok karmaşık, pahalı ve zaman alıcı olabilir. Bu nedenle işletmenin yararına kullanılacak verilerin hazırlanması ve işlenmesi, bütçe yönetimi açısından önemli bir maliyet kalemi ortaya çıkarabilir.

Genellikle veribilimi süreci bazı aşamaları içermektedir. Bu aşamalar aşağıdaki gibidir(Canbay, Vural ve Sağırođlu, 2017):

- Veri toplama ve depolama
- Keşif ve hedef belirleme
- Verilere erişim, alma ve bütünleştirme
- Verilerin işlenmesi ve temizlenmesi
- İlk veri araştırması ve keşifsel veri analizi
- Bir veya daha fazla potansiyel model ve algoritma seçmek
- Veri bilimi yöntem ve tekniklerini uygulama
- Sonuçların ölçülmesi ve iyileştirilmesi (dođrulama ve ayarlama)
- Kesin sonuçlar sunma ve iletme
- Yeni bir sorunu çözmek için işlemi tekrarlamak

1.2. Veri Bilimcisi Kimdir, Ne Yapar?

Veri bilimcileri özellikle de geçtiğimiz 10 yılda yerleri doldurulamayan personel haline geldi. Bu profesyonel kişiler soruları yanıtlamak ve kuruluşlarında stratejiyi yönlendirmek için kullanılan büyük miktarda bilgiyi organize etmek ve sentezlemek için karmaşık nicel algoritmalar oluşturabilen üst düzey teknik becerilere sahip, çok yönlü, veri odaklı bireylerdir (McGuire, Manyika ve Chui, 2012).

Veri bilimcileri, teknik olmayan meslektaşlarına yüksek teknik sonuçlar açıklamalarını sağlayan, sektöre özel bilgi ve iletişim becerilerine sahip, meraklı ve sonuç odaklı olmaları gerekmektedir. Veri bilimciler, istatistikler ve lineer cebir üzerine güçlü bir nicel altyapıya sahip oldukları gibi, algoritmaları oluşturmak ve analiz etmek için veri depolama, madencilik ve modellemeye odaklanan programlama bilgisine de sahiptirler. Ayrıca, veri bilimciler, R, Python, ApacheHadoop, MapReduce (Eşleİndirge), ApacheSpark, NoSQL veritabanları, Bulut Bilişim, ApachePig, Tableau ve GitHub gibi araçları kullanabilen kimselerdir.

Veri Bilimcisi, hangi soruların cevaplanması gerektiğini ve ilgili verileri nerede bulabileceklerini incelemektedir. İş zekası ve analitik becerilerinin yanı sıra verileri ayıklama,

temizleme ve sunma yeteneklerine sahiptirler. İşletmeler, büyük miktarda yapılandırılmamış veriyi depolamak, yönetmek ve analiz etmek için veri bilimcilerini kullanır. Sonuçlar daha sonra sentezlenir ve organizasyonda stratejik karar vermeyi yönlendirmek için kilit paydaşlara iletilir (McGuire, Manyika ve Chui, 2012).

Veri bilimcileri, tipik modern işletme tarafından üretilen veri miktarı arttıkça, ham verileri değerli işletme bilgilerine dönüştürmelerine yardımcı olmak için kuruluşlar tarafından işe alınan veri bilimcilerin saygınlığı da giderek artmaktadır. Veri çıkarma, daha fazla işlem ve inceleme için yapılandırılmamış veya kötü yapılandırılmış veri kaynaklarından belirli verileri alma işlemidir. Veri bilimcileri, analitik, makine öğrenmesi, veri madenciliği ve istatistiksel becerilerin yanı sıra algoritma ve kodlama tecrübesine sahip olmalıdır. Büyük miktarda veriyi yönetme ve yorumlamanın yanı sıra, birçok veri bilim insanına dijital bilginin işletme değerini göstermeye yardımcı olan veri görselleştirme modelleri oluşturma görevi de verilmektedir. Bununla birlikte, etkili olabilmek için veri bilimcileri, eğitim ve veri analitiği deneyimine ek olarak duygusal zekâyâ sahip olmalıdır. Belki bir veri bilimcisinin sahip olması gereken en önemli beceri, veri görüşlerini başkalarına sunma ve verinin önemini kolayca anlaşılabilir bir şekilde açıklama yeteneğidir (Gökçen, 2011).

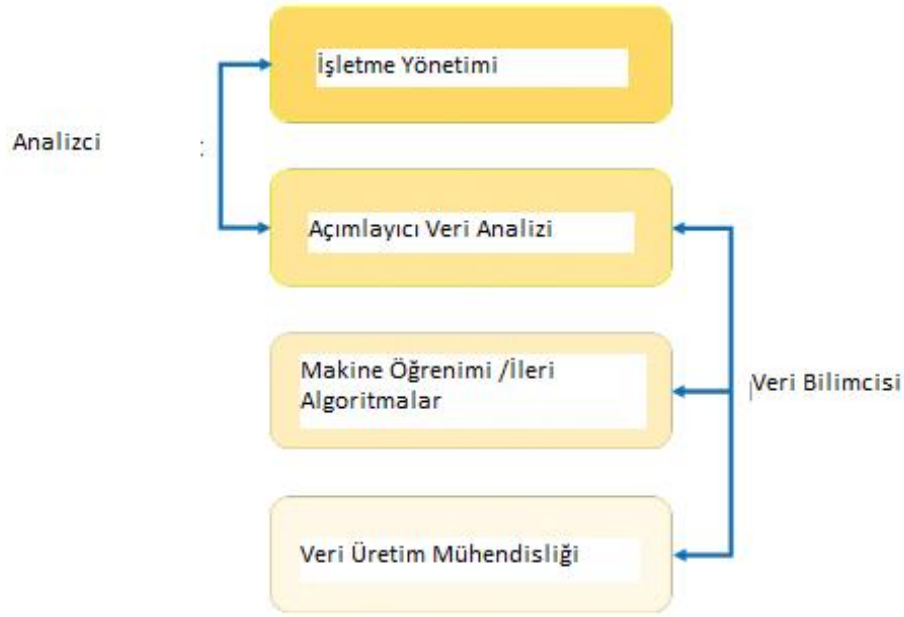
Veri bilimcileri, çalıştıkları dijital bilgileri, akıllı telefonlar, nesnelerin interneti (Internet of Things-IoT) cihazları, sosyal medya, anketler, satın almalar ve internet aramaları ve davranışları da dahil olmak üzere büyüyen kanallar ve kaynaklar listesinden almaktadır. Veri bilimciler bu büyük veri kümelerini sıralayarak, veri madenciliği olarak bilinen bir süreç olan veri analizi yoluyla sorunları çözecek kalıpları belirleyebilmektedirler (McGuire, Manyika ve Chui, 2012).

1.3. Veri Analisti

Veri analistleri, veri bilimcileri ve iş analistleri arasındaki boşluğu kapatmaktadır. Veri analistleri, teknik analizleri niteliksel eylem maddelerine dönüştürmekten ve bulgularını farklı paydaşlara etkin bir şekilde iletmekten sorumludur.

Aşağıdaki şekilde görüleceği üzere, Veri Analisti genellikle verilerin geçmişini işleyerek neler olduğunu açıklar. Öte yandan, veri bilimi, keşif analizini yalnızca bu konudaki görüşleri keşfetmek için yapmakla kalmaz, aynı zamanda gelecekteki belirli bir olayın oluşumunu tanımlamak için çeşitli gelişmiş makine öğrenme algoritmaları kullanır. Bir veri

bilimcisi, verilere birçok açıdan, bazen de daha önce bilinmeyen açılardan bakmaktadır (Gökçen, 2011).



Şekil 1.1. Veri Biliminin Aşamaları

Şekil 1.1 incelendiğinde veri bilimi öncelikle, öngörüsül, nedensel analitik, kurallı analitik (öngörücü artı karar bilimi) ve makine öğrenimini kullanan kararlar ve tahminler yapmak için kullanılır.

Öngörüsül Nedensel Analiz: Gelecekte belirli bir olayın olanaklarını tahmin edebilecek bir model isteniliyorsa, öngörüsül nedensel analizin uygulanması gerekmektedir. Krediyile para sağlıyorsanız, müşterilerin gelecekteki kredi ödemelerini zamanında yapmama olasılıkları sizin için endişe vericidir. Burada, gelecekteki ödemelerin zamanında yapılp yapılmayacağını tahmin etmek için müşterinin ödeme geçmişinde tahmine dayalı analitik gerçekleştirebilen bir model oluşturulabilmektedir (Aksu, Eser ve Güzeller, 2019).

Reçeteli Analitik: Kendi kararlarını alma zekasına ve dinamik parametrelerle değiştirebilme yeteneğine sahip bir model istendiğinde kesinlikle reçeteli analitik uygulaması kullanılmalıdır. Reçeteli analitik stratejisi ile verilerin akışı incelenerek sürecin nereye gittiği öngörüsül, anormallik durumunda operatöre ya da doğrudan otomatik olarak sisteme bir çözüm önerilmektedir (Aksu, Eser ve Güzeller, 2019).

Tahmin Yapmak için Makine Öğrenmesi: Bir finans şirketinin işlem verileriniz varsa ve gelecekteki eğilimi belirlemek için bir model oluşturmanız gerekiyorsa, makine

öğrenimi algoritmaları en iyi yöntemdir. Bu, denetimli öğrenme paradigması kapsamındadır. Denetimli olarak adlandırılır, çünkü makinelerinizi nasıl çalıştırabileceğinize dayanan verilere zaten sahipsiniz. Örneğin, bir sahtekarlık algılama modeli, geçmiş bir sahte alım kaydı kullanılarak eğitilebilir (Ersöz ve Ersöz, 2015).

Kalıp Keşfi için Makine Öğrenmesi: Tahmin yapabilmek için temelde parametreler yoksa, anlamlı tahminler yapabilmek için veri kümesindeki gizli kalıpları bulmak gerekmektedir. Bir telefon şirketinde bir bölgeye kuleler koyarak bir ağ kurmanız gerektiğini varsayalım. Ardından, tüm kullanıcıların en iyi sinyal gücünü almasını sağlayacak kule konumlarını bulmak için kümeleme tekniğini kullanmak mümkündür (Ersöz ve Ersöz, 2015).

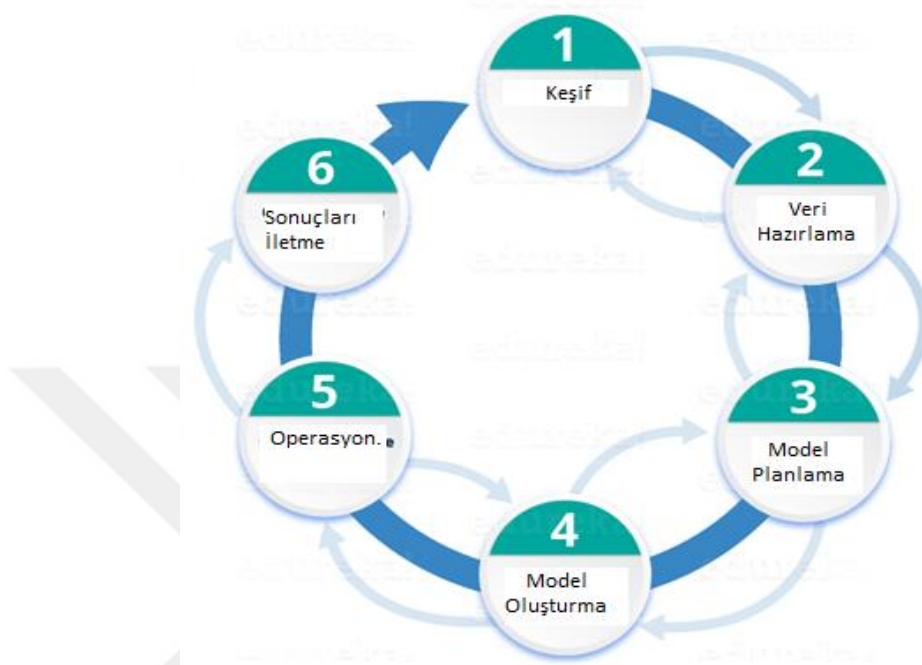
Şimdi de yukarıda açıklanan yaklaşımların oranının Veri Analizi ve Veri Bilimi için nasıl farklılaştığını görelim. Aşağıdaki şekilde görüldüğü gibi, Veri Analizi açıklayıcı analitik ve bir ölçüde tahmin içermektedir. Öte yandan, Veri Bilimi, Tahmin Edici Nedensel Analitik ve Makine Öğrenimi hakkında daha fazla bilgi vermektedir (Ersöz ve Ersöz, 2015).

1.4. Büyük Veri Yaşam Döngüsü

Büyük Veri Yaşam Döngüsü (Big data life cycle) kavramı henüz yeni ve oturmamış bir kavramdır. Birçok kişinin kullandığı bu terim, çok farklı anlamlara gelebilmektedir. Yeni modelde, iş dünyasıyla veri ambarları ve iş zekâsı süreçlerinin bir yaşam döngüsünde bağlanmaktadır. Öncelikle bir iş süreci var ve iş dünyasının belli ihtiyaçları ve sorguları bu modelde ele alınmış. İkinci adımda veri ambarı (Data Warehouse) süreci işin içine girmektedir. Burada görev yapan veri bilim insanından bahsedilebilir. Veri bilim insanı, veri ambarının çıkartmış olduğu farklı veri kaynaklarından topladığı, ön işlendiği ve analiz ettiği sonuçları alır ve bunların üzerinde kendine göre analiz/analitik modelleri geliştirir. Bu aşamada, veri madenciliği algoritmalarından bahsedilebilir. Belirli tekrarlar, belirli şablonlar yakalanmaya çalışılır. Sonraki adımda bir iş zekâsı süreci başlar. İş zekâsı sürecinde yeni fikirlerin ortaya çıkması çok önemlidir. Yani veri bilimcisi açısından baktığımızda, iş zekâsında çıkanlar sadece rapordan ibarettir ama iş dünyası açısından baktığımızda o raporlar aslında iş dünyasına şekil veren iş dünyasını bazen komple değiştiren yeni bir sürece sürükleyen yeni fikirler getiren bir yapıya sahiptir. Ardından bu çıktıların iş dünyasındaki etkilerinin tekrar ölçülmesi, verinin anlaşılması, kararları etkilemesi, stratejik kararları yönetim sürecini etkilemesi, işlevsel anlamda çalışanların verimliliğini arttırması, kalitenin arttırılması gibi çok farklı yansımaları olur ve yeni ihtiyaçlar ortaya çıkar. Bu yeni ihtiyaçlar

yeni veri işleme süreçlerini başlatır. Böylelikle, yeni büyük veri yaşam döngüleri başlar ve büyük veri yaşam döngüsü bu şekilde dönmeye devam eder (Şeker, 2015).

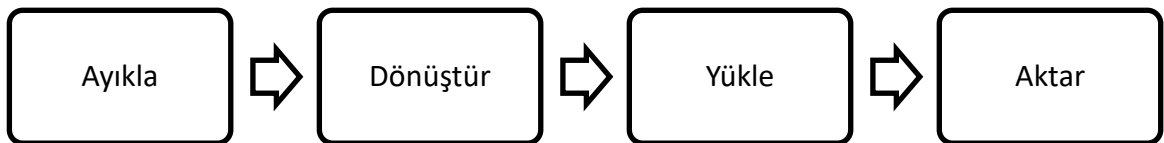
Veri Bilim Yaşam Döngüsünün ana aşamaları Şekil 1.2’de verilmiştir:



Şekil 1.2. Veri Bilim Yaşam Döngüsü (Sharma, 2019)

1. Aşama (Keşif): Projeye başlamadan önce, çeşitli özellikleri, gereksinimleri, öncelikleri ve gerekli bütçeyi anlamak önemlidir. Doğru soruları sorma yeteneğine sahip olmak gerekmektedir. Burada, projeyi desteklemek için insan, teknoloji, zaman ve veri bakımından gerekli kaynakların olup olmadığını değerlendirilmelidir. Bu aşamada, iş problemini çerçevelemek ve test etmek için ilk hipotezleri formüle etmek gerekmektedir.

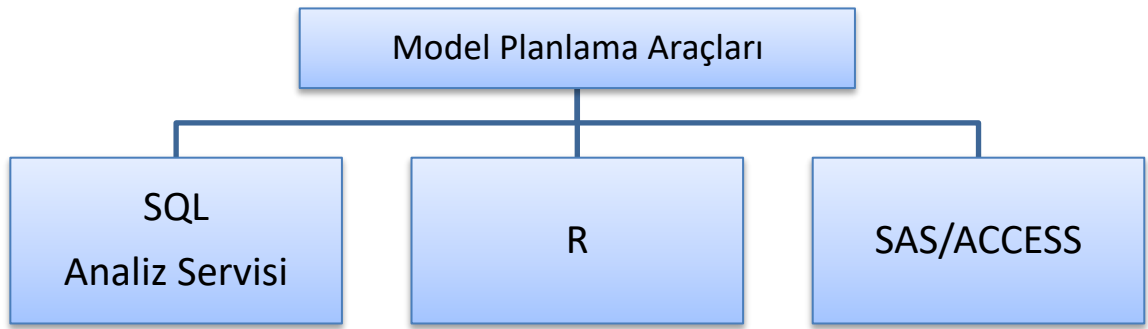
2. Aşama (Veri Hazırlama): Bu aşamada, projenin tamamı boyunca analitik gerçekleştirebilecek analitik sanal alana ihtiyaç vardır. Modellemeden önce verileri araştırmak, ön işlemeye koymak ve koşullandırmak gerekmektedir. Ayrıca, verileri sanal alana almak için ETLT (Eliminate, Transform, Load, Transform: Ayıkla, Dönüştür, Yükle ve Aktar) gerçekleştirilmektedir. Şekil 1.3’te İstatistiksel Analiz akışı verilmiştir.



Şekil 1.3. İstatistiksel Analiz Akışı

Veri temizleme, dönüştürme ve görselleştirme için R kullanılabilir. Bu, aykırı noktaları tespit etmeye ve değişkenler arasında bir ilişki kurmaya yardımcı olacaktır. Verileri temizledikten ve hazırladıktan sonra, üzerinde keşifsel analizler yapmak için hazır hale gelir.

3. Aşama (Model Planlama):Burada değişkenler arasındaki ilişkileri çizecek yöntem ve teknikleri belirlenmektedir. Bu ilişkiler, bir sonraki aşamada uygulanacak olan algoritmaların temelini oluşturacaktır. Açıklayıcı Veri Analitiği, çeşitli istatistiksel formüller ve görselleştirme araçları kullanarak uygulanmaktadır. Şekil 1.4'te model planlama araçları gösterilmektedir.



Şekil 1.4. Model Planlama Araçları

R programı, tam bir modelleme yeteneklerine sahiptir ve yorumlayıcı modeller oluşturmak için iyi bir ortam sağlamaktadır.

SQL Analiz Hizmetleri, ortak veri madenciliği işlevlerini ve temel öngörücü modelleri kullanarak veritabanı içi analitik işlemleri gerçekleştirmektedir.

SAS/ACCESS, Hadoop'tan verilere erişmek için kullanılabilir ve tekrarlanabilir ve yeniden kullanılabilir model akış şemaları oluşturmak için kullanılır.

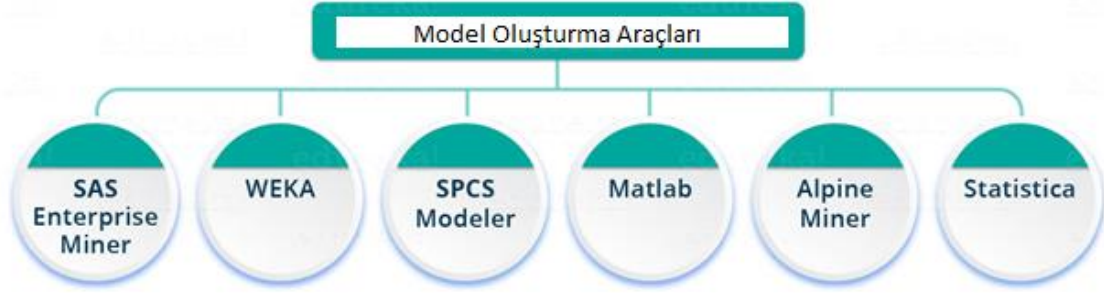
Bununla birlikte, piyasada birçok araç mevcut olmakla birlikte, R programı en yaygın kullanılan araçtır.

Artık verilerin doğası hakkında fikir sahibi olunduktan sonra kullanılacak algoritmalara karar verilir. Bir sonraki aşamada algoritmayı uygulayacak ve bir model oluşturulur.

4. Aşama (Model Oluşturma):Bu aşamada, eğitim ve test amaçlı veri kümeleri geliştirilir. Mevcut araçların modelleri çalıştırmak için yeterli olup olmayacağını veya daha

sağlam bir ortama (hızlı ve paralel işleme gibi) ihtiyaç duyulacağı düşünülür. Modeli oluşturmak için sınıflandırma, ilişkilendirme ve kümeleme gibi çeşitli öğrenme tekniklerini analiz kullanılır.

Model oluşturma işlemi Şekil 1.5'teki yazılım araçlarıyla yapılabilir.



Şekil 1.5. Model Oluşturma Araçları (Sharma, 2019)

5. Aşama (Operasyonelleştirme): Bu aşamada, nihai raporlar, brifingler, kodlar ve teknik belgeler sunulur. Ek olarak, bazen gerçek zamanlı bir üretim ortamında pilot bir proje de uygulanmaktadır. Bu, tam dağıtımdan önce performansın ve diğer ilgili kısıtlamaların küçük ölçekte net bir şekilde görüntülenmesini sağlayacaktır.

6. Aşama (Sonuçları İletme): İlk aşamada planlanan hedefe ulaşıp ulaşılmadığını değerlendirmek önemlidir. Böylece, son aşamada tüm kilit bulguları belirler, paydaşlarla iletişim kurar ve proje sonuçlarının Faz 1'de geliştirilen kriterlere dayanarak başarılı mı yoksa başarısız mı olduğu belirlenir.

1.5. Büyük Veri Nedir?

Teknolojideki gelişmeler nedeniyle, Büyük Veri'nin tanımı yıllar içinde değişmiştir. Ancak, aynı kalan tek şey, veri miktarının son derece hızlı bir oranda sürekli olarak artmasıdır. Herhangi bir biçimde, iç görü kazanmak ve değer üretmek için kullanılan tüm veriler "Büyük Veri" olarak kabul edilir (Mauro, Greco ve Grimaldi, 2015).

Büyük veri, artan hacimlerde ve daha da yüksek hızda gelen daha fazla çeşitliliği içeren verilerdir. Özellikle, Büyük Veri, hacim, hız ve çeşitlilik açısından dikkat çekici olan veri oluşturma, saklama, geri alma ve analiz ile ilgilidir. Durum "3 V" olarak bilinmektedir.

1.6. Büyük Veri'nin 3V'si

1.6.1. Veri Büyüklüğü (Volume)

Veri miktarı önemlidir. Büyük verilerle, yüksek hacimli, düşük yoğunluklu ve yapılandırılmamış verileri işlemek gerekir. Bu, Twitter veri yayınları, bir web sayfasındaki tıklama akışları veya mobil uygulama veya sensör etkin donanımlar gibi bilinmeyen değerdeki veriler olabilmektedir. Bazı kuruluşlar için, bu onlarca terabayt veri olabilmektedir. Diğerleri için, yüzlerce PB (PetaBayt) olabilmektedir (Watson, 2014).

1.6.2. Hız (Velocity)

Verilerin alındığı ve (belki de) üzerinde etkili olduğu hızdır. Normalde, en yüksek veri hızı, diske yazılmak yerine doğrudan belleğe aktarılır. Bazı İnternet özellikli akıllı ürünler gerçek zamanlı olarak veya gerçek zamanlıya yakın çalışır ve gerçek zamanlı değerlendirme ve eylem gerektirmektedir.

1.6.3. Çeşitlilik (Variety)

Mevcut olan birçok veri türünü ifade etmektedir. Geleneksel veri tipleri yapılandırılmış ve ilişkisel bir veri tabanına tam olarak oturmuştur. Büyük verilerin artmasıyla birlikte, veriler yeni yapılandırılmamış veri türlerinde gelir. Metin, ses ve video gibi yapılandırılmamış ve yarı yapılandırılmış veri türleri, anlam çıkarmak ve meta verileri desteklemek için ek ön işlemler gerektirir (Mauro, Greco ve Grimaldi, 2015).

Son zamanlarda, büyük veri uygulayıcıları ve düşünce liderleri ek V'ler önermişlerdir. Bunlar;

1.6.4. Doğrulama (Validation)

Bu, toplanan verilerin kalitesini ifade eder. Kaynak veri doğru değilse, analizler değersiz olacaktır. Dünya otomatik karar verme yolunda ilerlerken, bilgisayarların insanlar yerine seçimler yaptığı yerlerde, kuruluşların verilerin kalitesine güvenebilmesi zorunlu hale gelir.

1.6.5. Değer (Value)

Büyük verinin dünya çapında çok fazla ilgi görmesinin temel sebebi değer sağlama potansiyelidir. “Değer” kavramından kastedilen veriye ilişkin değerdir. Sınırsız miktarda veri herhangi bir değere dönüştürülmediği sürece hiçbir anlam ifade etmemektedir. Veriler ve öngörüler arasında açık bir şekilde bir bağlantı olsa da bu durum her zaman büyük veride bir karşılık bulamamaktadır. Büyük veri kullanılarak elde edilen değer her durumda olmasa da bazı durumlarda öznel ve zaman bağlıdır. Bir başka deyişle, bir veriye ilişkin değer zamanla değişmektedir. Veriler kendi başlarına çoğu zaman sınırlı değerdedir. Ancak, veriler o ana kadar kullanılmayan veya erişilemeyen anlayışlara yol açtığına önemli farklılıklar yaratacak bir değer farklılaştırıcısı haline gelmektedir (Watson, 2014).

Basitçe söylemek gerekirse, büyük veriler, özellikle yeni veri kaynaklarından daha büyük, daha karmaşık veri kümeleridir. Bu veri kümeleri o kadar büyük hacimlidir ki, geleneksel veri işleme yazılımı onları yönetemez. Ancak, bu büyük miktardaki veriler, daha önce çözemediğiniz iş sorunlarını çözmek için kullanılabilir.

Büyük veri, sayısız kaynaktan gelir: Akıllı telefonlar ve sosyal medya yayınları; trafik sinyalleri ve fayda sayaçları gibi sensörler; elektronik sağlık kayıtları gibi.

Büyük veri, belirli bir zaman diliminde verileri yönetmek ve analiz etmek için geleneksel veri yönetim sisteminin işleme kapasitesinin ötesinde büyük miktarda veri anlamına gelmektedir. Büyük Veri birçok kaynaktan gelmektedir, bazıları dijital medya, çevrimiçi işlem kayıtları, cep telefonu sinyalleri vb.

Örneğin, 10 TB (TeraByte) görüntü dosyanız varsa, belirli zaman dilimlerinde görüntülerin yeniden boyutlandırılması ve geliştirilmesi gibi bazı işlemlerin yapılması gerekir. Bu sistemi yerine getirmek için geleneksel sistemleri kullanırsanız, bu görevi belirli bir zaman diliminde gerçekleştiremezsiniz, çünkü geleneksel sistemin hesaplama kaynakları

bu görevi yerine getirmek için yeterli değildir. Bu nedenle, bu 10 TB veri Büyük Veri olarak adlandırılır.

Büyük Veri'nin ana zorluğu, verileri belirli bir zaman aralığında depolamak ve işlemektir. Geleneksel yöntemleri kullanarak bu görevi gerçekleştirmek iyi bir uygulama değildir. Büyük verileri analiz etmek ve saklamak için bu geleneksel yaklaşımdan kurtulmak için Hadoop teknolojisi geliştirilmiştir (Mauro, Greco ve Grimaldi, 2015).

Veri gölleri, birçok kaynaktan elde edilen büyük verileri ham, ayrıntılı bir biçimde tutan merkezi bir depolama deposudur. Yapılandırılmış, yarı yapılandırılmış veya yapılandırılmamış verileri depolayabilir, bu da verilerin gelecekteki kullanım için daha esnek bir biçimde tutulabileceği anlamına gelmektedir. Veri depolarken, bir veri gölü daha hızlı erişim için tanımlayıcılarla ve meta veri etiketleriyle ilişkilendirilir. Veri bilimciler, veri göllerini kullanarak verilere daha hızlı ve daha doğru bir şekilde erişebilir, hazırlayabilir ve analiz edebilir. Analitik uzmanları için, çeşitli geleneksel olmayan formatlarda bulunan bu geniş veri havuzu, duyarlılık analizi veya sahtekarlık tespiti gibi çeşitli kullanım durumlarında verilere erişim için benzersiz bir fırsat sağlamaktadır.

1.7. Büyük Veri Analizi Nedir?

Büyük Veri Analizi, yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış veriler dahil olmak üzere her türlü verilere modern analitik yazılım araçlarını uygulama pratiğidir. Büyük Veri Analizi'nin amacı, öngörülerini ve usulsüzlükleri keşfetmek ve iş performansı ve müşteri davranışlarını anlayışını arttırmaktır. Analitik odaklı bu bilgiler, işletme sonuçlarını şekillendirmek, rekabet avantajını artırmak, finansal kararları iyileştirmek ve daha özlü tahminler geliştirmek için kullanılabilir.

Büyük Veri, geleneksel teknolojiler, beceriler ve altyapının verimli bir şekilde ele alınması için çok çeşitli, hızlı değişen veya büyük veri içeren teknolojiler ve girişimler anlamına gelmektedir. Farklı şekilde söylendiği gibi, verilerin hacmi, hızı veya çeşitliliği çok fazladır (McGuire, Manyika ve Chui, 2012).

Ancak günümüzde yeni teknolojiler Büyük Veri'den değer elde etmeyi mümkün kılmaktadır. Örneğin, perakendeciler, kampanyaları, fiyatlandırmayı ve stoklamayı iyileştiren davranış eğilimlerini belirlemek için kullanıcının web tıklamalarını izleyebilmektedir. Kamu hizmetleri, kesintileri tahmin etmek ve daha verimli enerji tüketimini teşvik etmek için hane

halkı enerji kullanım seviyelerini yakalayabilmektedir. Hükümetler ve hatta Google bile hastalık salgınlarının ortaya çıkışını sosyal medya sinyalleriyle tespit edip izleyebilmektedir. Petrol ve gaz şirketleri, daha verimli ve daha güvenli delme kararları almak için sondaj ekipmanlarındaki sensörlerin çıktısını alabilmektedir.

1.8. Veri Türleri

Büyük Veri yapısını oluşturan veriler üç farklı biçimde bulunabilir:

1. Yapılandırılmış
2. Yapılandırılmamış
3. Yarı Yapılandırılmış

1.8.1. Yapılandırılmış Veri

Yapılandırılmış veri tipi belirli kurallar ve sistemler doğrultusunda depolandıkları için kolay erişilebilir, düzenlenebilir, kategorize edilebilir vb. yapıdadır. Bu veri tipi, tablolar üzerinde satır ve sütunlar halinde düzenlenmiş verilerdir. Bu tür veriler kağıt üzerinde olabileceği gibi, bilgisayarlarda istatistik paket programlarının satır ve sütunlarından oluşan matris şeklindeki yapılara da kayıt edilmiş olabilir (Oğuzlar,2011).

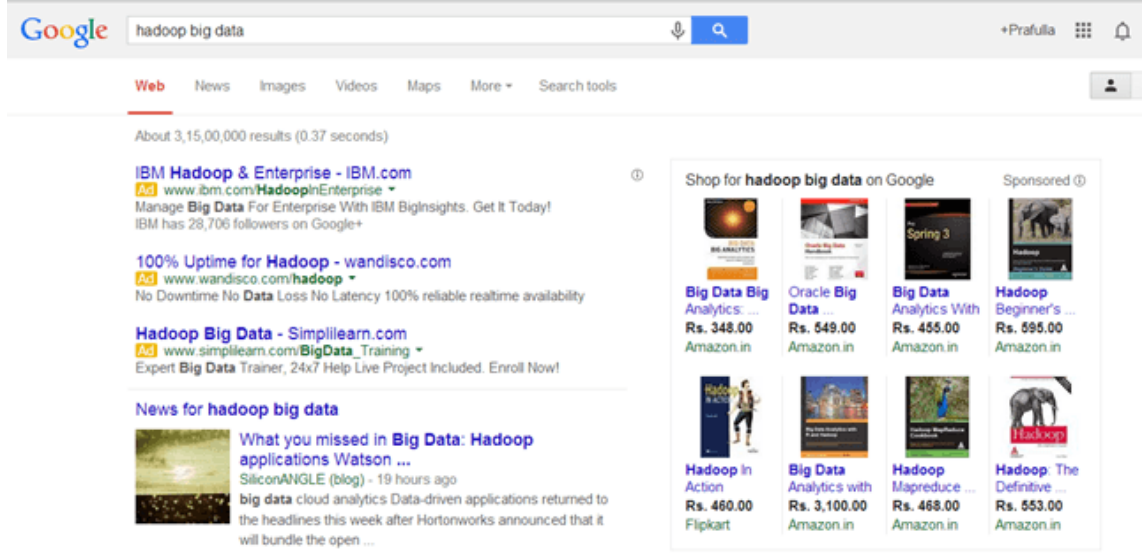
1.8.2. Yapılandırılmamış Veri

Yapılandırılmamış veriler ise mektup, doküman, kitap gibi kağıt üzerinde bulunan veya e-mail, web sayfaları gibi elektronik ortam metinlerinden; fotoğraf gibi durağan ya da film gibi hareketli görüntülerden ve/veya seslerden oluşur (Oğuzlar, 2011).

Yukarıdaki açıklamalara göre yapılandırılmış veri ile yapılandırılmamış veri arasındaki temel farkın, yapılandırılmış verinin üzerinde her türlü işlem ve sorgulamanın yapılabileceği ilişkilerinin kolaylıkla kurulabileceği bir veritabanı yönetim sistemi üzerinde bulundurulması olduğu görülebilir (Doğan ve Arslantekin, 2016).

Yapılandırılmamış Veri Örnekleri

Şekil 1.6'da 'Google Arama' tarafından döndürülen çıktı yer almaktadır.



Şekil 1.6. Yapılandırılmamış Veri Örneği

1.8.3. Yarı Yapılandırılmış Veri

Tanımlı bir format haricinde sunulan ve depolanan kayıt türleridir. Genellikle, kitaplar, makaleler, belgeler, e-postalar gibi serbest formatlardaki metinlerden ve resim, ses ve video gibi medya dosyalarından oluşmaktadır. Bu türdeki verinin katı bir şekilde sunulmasının zor olması, veri işleme süreçlerinde NoSQL (Not only SQL) gibi yeni mekanizmaların ortaya çıkmasına neden olmuştur (AKTAN,2018).

Yarı yapılandırılmış veri örnekleri

Bir XML dosyasında saklanan kişisel veriler:

```
<rec><name>PrashantRao</name><sex> Erkek </sex><age> 35 </age></rec>  
<rec><name>Seema R. </name><sex> Kadın </sex><age> 41 </age></rec>  
<rec><name>SatishMane</name><sex> Erkek </sex><age> 29 </age></rec>  
<rec><name>SubratoRoy</name><sex> Erkek </sex><age> 26 </age></rec>  
<rec><name>Jeremiah J. </name><sex> Erkek </sex><age> 35 </age></rec>  
<rec><name>PrashantRao</name><sex> Erkek </sex><age> 35 </age></rec>
```

```
<rec><name>Seema R. </name><sex> Kadın </sex><age> 41 </age></rec>
```

```
<rec><name>SatishMane</name><sex> Erkek </sex><age> 29 </age></rec>
```

```
<rec><name>SubratoRoy</name><sex> Erkek </sex><age> 26 </age></rec>
```

```
<rec><name>Jeremiah J. </name><sex> Erkek </sex><age> 35 </age></rec>
```

1.9. Veri Kaynağı Nedir?

Bir veri kaynağı, bilgisayar bilimleri ve bilgisayar uygulamaları bağlamında, kullanılan verilerin nereden geldiği konumdur. Bir veritabanı yönetim sisteminde, birincil veri kaynağı bir diskte veya uzak bir sunucuda bulunabilen veritabanıdır. Bir bilgisayar programının veri kaynağı bir dosya, bir veri sayfası, bir elektronik tablo, bir XML dosyası ve hatta programın içinde kodlanmış veriler olabilmektedir (Ersöz ve Ersöz, 2015)

Bir veri kaynağının amacı, verilere erişmek için gereken tüm teknik bilgileri (sürücü adı, ağ adresi, ağ yazılımı vb.) tek bir yerde toplamak ve kullanıcıdan gizlemektir. Kullanıcı, Bordro, Envanter ve Personel içeren bir listeye bakabilmeli, listeden Bordro'yu seçebilmeli ve uygulamanın, bordro verilerinin nerede olduğunu veya uygulamanın nasıl elde edildiğini bilmeden, bordro verilerine bağlanmasını sağlamalıdır (Ersöz ve Ersöz, 2015).

1.9.1. Veri Kaynakları Türleri

İki tür veri kaynağı vardır: makine veri kaynakları ve dosya veri kaynakları. Her ikisi de veri kaynağı hakkında benzer bilgiler içermesine rağmen, bu bilgilerin depolanma şekli bakımından farklılık gösterir. Bu farklılıklar nedeniyle, biraz farklı şekillerde kullanılırlar.

1.9.1.1. Makine veri kaynakları

Makine veri kaynakları, kullanıcı tanımlı bir adla sistemde saklanır. Veri kaynağı adıyla ilişkili olarak, Sürücü Yöneticisi ve sürücünün veri kaynağına bağlanması gereken tüm bilgiler bulunur. Bir Xbase veri kaynağı için, Xbase sürücüsünün adı, Xbase dosyalarını içeren dizinin tam yolu ve sürücüye, tek kullanıcı mod veya salt okunur gibi bu dosyaları nasıl kullanacağını söyleyen bazı seçenekler olabilmektedir. Oracle veri kaynağı için bu, Oracle sürücüsünün adı, Oracle DBMS'nin bulunduğu sunucu, kullanılacak SQL * Net sürücüsünü tanımlayan SQL * Net bağlantı dizesi ve sunucudaki veritabanının sistem kimliği olabilir.

1.9.1.2. Dosya veri kaynakları

Dosya veri kaynakları bir dosyada saklanır ve bağlantı bilgilerinin tek bir kullanıcı tarafından tekrar tekrar kullanılmasına veya birkaç kullanıcı arasında paylaşılmasına izin verir. Bir dosya veri kaynağı kullanıldığında, Sürücü Yöneticisi bir dsn dosyasındaki bilgileri kullanarak veri kaynağına bağlantı kurar. Bu dosya diğer dosyalar gibi değiştirilebilir. Bir dosya veri kaynağında, bir makine veri kaynağında olduğu gibi bir veri kaynağı adı yoktur ve herhangi bir kullanıcı veya makineye kayıtlı değildir.

Bir dosya veri kaynağı, bağlantı işlemini kolaylaştırır, çünkü dsn dosyası SQL Driver Connect işlevine yapılan bir çağrı için yapılması gereken bağlantı dizesini içermektedir. Dsn dosyasının bir diğer avantajı, herhangi bir makineye kopyalanabilmesidir, böylece aynı veri kaynakları, uygun sürücünün kurulu olduğu sürece birçok makine tarafından kullanılabilir. Bir dosya veri kaynağı, uygulamalar tarafından da paylaşılabilir. Paylaşılabilir bir dosya veri kaynağı bir ağa yerleştirilebilir ve aynı anda birden fazla uygulama tarafından kullanılabilir.

Bir .dsn dosyası da paylaşılabilir. Paylaşılabilen bir .dsn dosyası, tek bir makinede bulunur ve bir makine veri kaynağına işaret eder. Uygulanamayan dosya veri kaynakları, esas olarak, makine veri kaynaklarının dosya veri kaynaklarına kolayca dönüştürülmesini sağlamak için mevcuttur, böylece bir uygulama yalnızca dosya veri kaynaklarıyla çalışmak üzere tasarlanabilir. Sürücü Yöneticisi bilgileri paylaşılabilen bir dosya veri kaynağına gönderdiğinde, .dsn dosyasının işaret ettiği makine veri kaynağına gerektiği gibi bağlanır.

1.10. Büyük Veri Uygulamaları

Akıllı Şebeke: Ulusal elektronik güç tüketimini gerçek zamanlı olarak yönetmek ve Akıllı şebeke operasyonlarını izlemek çok önemlidir. Bu akıllı sayaçlar, Ulusal kontrol merkezleri ve diğer altyapılar arasındaki çoklu bağlantılarla sağlanır. Büyük Veri analitiği, risk altındaki trafoların belirlenmesine ve bağlı cihazların anormal davranışlarının tespit edilmesine yardımcı olur. Grid Utilities bu sayede en iyi tedavi veya işlemi seçebilir. Üretilen Büyük Veri'nin gerçek zamanlı analizi, olay senaryolarının modellenmesine izin verir. Bu, düzeltici maliyetleri azaltmak için stratejik önleyici planlar oluşturmayı sağlar. Ek olarak, Enerji Tahmini analitiği, güç talep yükünü daha iyi yönetmeye, kaynakları planlamaya ve dolayısıyla ispatları maksimize etmeye yardım eder (Stimmel, 2014).

E-sağlık: bağlı sağlık platformları zaten sağlık hizmetlerini kişiselleştirmek için kullanılıyor (örneğin, CISCO çözümü)(Nambiar ve ark., 2013). Büyük Veriler farklı heterojen kaynaklardan üretilir (Örn. Laboratuvar ve klinik veriler, uzak sensörlerden yüklenen hasta semptomları, hastaneler işlemleri, farmasötik veriler). Tıbbi veri setlerinin gelişmiş analizi birçok faydalı uygulamaya sahiptir. Sağlık hizmetlerini kişiselleştirmeyi sağlar (örneğin, doktorlar reçeteyi ayarlamak için çevrimiçi hasta semptomlarını izleyebilir); Halk sağlığı planlarını nüfus belirtileri, hastalık evrimi ve diğer parametrelere göre uyarlamak. Ayrıca hastane operasyonlarını optimize etmek ve sağlık harcamalarını azaltmak için de faydalıdır.

Nesnelerin İnterneti(IoT): Büyük veri uygulamalarının ana pazarlarından birini temsil eder. Çok çeşitli nesnelere nedeniyle, IoT uygulamaları sürekli olarak gelişmektedir. Günümüzde lojistik işletmeleri destekleyen çeşitli Büyük Veri uygulamaları bulunmaktadır. Aslında, araçların konumlarını sensörler, kablosuz adaptörler ve GPS ile izlemek mümkündür. Böylece, bu tür veri odaklı uygulamalar, şirketlerin yalnızca çalışanları denetlemelerini ve yönetmelerini değil aynı zamanda teslimat yollarını optimize etmelerini sağlar. Bu geçmiş sürüş deneyimi de dahil olmak üzere çeşitli bilgileri kullanarak ve birleştirerek Smart şehirde IoT verilerinin uygulanmasına dayalı sıcak bir araştırma alanıdır (Chen ve diğerleri, 2014b)

Kamu hizmetleri: Su tedarik kuruluşları gibi tesisler, karmaşık su temin şebekelerindeki su akışını izlemek için boru hatlarına sensörler yerleştirilir. Basında, Bangalore (Hindistan'da bir şehir) Su Temini ve Kanalizasyon Kurulunun, şehrin farklı bölgelerine eşit miktarda su teminini sağlamak için sızıntıları, yasa dışı bağlantıları ve uzaktan kontrol vanalarını tespit etmek için gerçek zamanlı bir izleme sistemi uyguladığı bildiriliyor. Vana operatörlerine duyulan ihtiyacı azaltmada ve sızan su borularını zamanında tespit etmede ve sabitlemede yardımcı olur.

Ulaştırma ve lojistik:[Birçok toplu taşıma şirketi, otobüsleri takip etmek için RFID ve GPS kullanıyor ve orada hizmetleri geliştirmek için ilginç verileri araştırıyor. Örneğin, farklı rotalarda otobüs kullanan yolcu sayısı hakkında toplanan veriler otobüs rotalarını ve seyahat sıklığını optimize etmek için kullanılır. Çeşitli gerçek zamanlı sistemler sadece yolculara önerilerde bulunmakla kalmayıp, aynı zamanda onu istenen varış noktasına götürecek olan bir sonraki otobüsü ne zaman bekleyecekleri konusunda değerli bilgiler sunmak için de uygulanmaktadır. Büyük Veri Madenciliği, kamu ya da özel ağlara olan talebi tahmin ederek

seyahat işini iyileştirmeye de yardımcı olur. Örneğin, dünyanın en büyük demiryolu ağlarından birine sahip olan Hindistan'da, her gün verilen toplam ayrılmış koltuk sayısı 250.000 civarındadır ve 60gün öncesinden rezervasyon yapılabilir. Bu tür verilerden tahminler yapmak karmaşık bir konudur çünkü hafta sonları, festivaller, gece treni, başlangıç veya ara istasyon gibi birçok faktöre bağlıdır. Makine öğrenmesi algoritmalarını kullanarak, geçmiş ve yeni büyük veri toplamaya ileri düzeyde analitik uygulamak ve uygulamak mümkündür. Aslında, gelişmiş analizler birçok konuda yüksek sonuç elde edilmesini sağlayabilir

Siyasi hizmetler ve devlet izleme: Hindistan ve ABD gibi birçok hükümet, siyasi eğilimleri izlemek ve nüfus duygularını analiz etmek için veri madenciliğini kullanmaktadır. Birçok veri kaynağını birleştiren birçok uygulama vardır: sosyal ağ iletişimi, kişisel görüşmeler ve seçmen kompozisyonları. Bu tür sistemler, ulusal sorunlara ek olarak yerel sorunları tespit etmeyi de sağlar. Ayrıca, hükümetler, değerli kaynakların ve yardımcı programların kullanımını optimize etmek için Büyük Veri sistemlerini kullanabilir. Örneğin, büyük ağlardaki su akışını izlemek için su tedarik zinciri boru hatlarına sensörler yerleştirilebilir. Bu nedenle birçok ülkenin, şehrin farklı bölgelerine eşit miktarda su temini sağlamak için kaçakları, yasadışı bağlantıları ve uzaktan kontrol vanalarını tespit etmek için gerçek zamanlı izleme sistemine güvenmek mümkündür (Rajaraman, 2016).

1.11. Büyük Veri Teknolojileri

Büyük veri hızıyla başa çıkma ihtiyacı, temel bilgi işlem altyapısına benzersiz talepler getirir. Büyük hacimleri ve veri türlerini hızlıca işlemek için gereken bilgi işlem gücü tek bir sunucu veya sunucu kümesini zorlayabilir. Kuruluşlar, gerekli hızı elde etmek için büyük veri görevlerine yeterli işlem kapasitesi uygulamalıdır. Bu, potansiyel olarak, işlemeyi dağıtabilecek ve kümelenmiş bir mimaride işbirliği içinde çalışabilecek yüzlerce veya binlerce sunucuyu talep edebilir(Simon, 2014).

Bu hızın düşük maliyetli bir şekilde elde edilmesi de bir zorluktur. Bir çok işletme lideri, özellikle 7/24 çalışmayanlar olmak üzere büyük veri iş yüklerini desteklemek için kapsamlı bir sunucu ve depolama altyapısına yatırım yapmaktan çekinmektedir. Sonuç olarak, genel bulut bilişim artık büyük veri sistemlerini barındırmak için birincil araç olmuştur. Açık bir bulut sağlayıcısı, büyük bir veri analizi projesini tamamlayacak kadar uzun süre boyunca veri depolayabilir ve gereken sayıda sunucuyu ölçeklendirebilir. İşletme yalnızca gerçekten

kullanılan depolama ve hesaplama zamanını öder ve bulut örnekleri tekrar ihtiyaç duyulana kadar kapatılabilir (Weng ve Lin, 2014; Yadav ve Kumar, 2015).

Hizmet seviyelerini daha da artırmak için, genel bulut sağlayıcıları, yüksek oranda dağıtılmış Apache Hadoop hesaplama örnekleri, ApacheSpark işleme motoru ve ilgili büyük veri teknolojilerini içeren yönetilen servisler aracılığıyla büyük veri yetenekleri sunar. Amazon Web Services'den (AWS) Amazon ElasticMapReduce (EMR), genel bir bulutta çalışan büyük bir veri hizmetine bir örnektir; diğerleri Microsoft'un Azure HDInsight™ ve Google CloudDataproc'tur. Bulut ortamlarında, büyük veriler Hadoop Dağıtılmış Dosya Sisteminde (HDFS) veya Amazon Basit Depolama Hizmeti (S3 gibi düşük maliyetli bulut nesnesi deposunda saklanabilir)); NoSQL veri tabanları, kendileri için uygun olan uygulamalar için buluttaki başka bir seçenektir (Weng ve Lin, 2014).

Şirket içi büyük veri sistemlerini dağıtmak isteyen kuruluşlar için, Hadoop ve Spark'ın yanı sıra yaygın olarak kullanılan Apache açık kaynak teknolojileri, Hadoop'un yerleşik kaynak yöneticisi ve iş zamanlayıcısı, Başka Bir Kaynak Müzakerecisi (YARN), MapReduce Programlama çerçevesi, bir uygulama-uygulama mesajlaşma ve veri aktarım platformu olan Kafka, HBase veri tabanı ve Drill, Hive gibi SQL-on-Hadoop sorgu motorları, Impala ve Presto'dur. Kullanıcılar, teknolojilerin kendilerinin açık kaynaklı versiyonlarını kurabilirler veya bulutta da desteklenen Cloudera, Hortonworks ve MapR Technologies tarafından sunulan ticari büyük veri platformlarına başvurabilirler (Canbay, Vural ve Sağıroğlu, 2017).

1.12. Hadoop Nedir?

ApacheHadoop, büyük miktarda veriyi bir kümede etkili bir şekilde saklayabilen java tabanlı bir özgür yazılım çerçevesidir. Bu çerçeve bir küme üzerinde paralel olarak çalışır ve tüm düğümlerdeki verileri işlememize izin verme yeteneğine sahiptir. Hadoop Dağıtılmış Dosya Sistemi (HDFS), büyük veri toplayan ve bir kümedeki birçok düğüme dağıtan Hadoop'un depolama sistemidir. Bu aynı zamanda kümedeki verileri çoğaltır ve böylece yüksek kullanılabilirlik sağlar(Ramachandran, 2013).

Hadoop, kümelenmiş sistemlerde çalışan büyük veri uygulamaları için veri işlemeyi ve depolamayı yöneten açık kaynaklı bir dağıtım çerçevesidir. Tahmini analitik, veri madenciliği ve makine öğrenimi uygulamaları da dahil olmak üzere, öncelikle gelişmiş analitik girişimleri desteklemek için kullanılan, büyüyen büyük veri teknolojileri ekosisteminin merkezinde yer almaktadır. Hadoop, çeşitli yapılandırılmış ve

yapılandırılmamış veri biçimlerini kullanabilir ve kullanıcılara ilişkisel veritabanlarının ve veri ambarlarının sağladığından daha fazla veri toplama, işleme ve analiz etme esnekliği sunar.

Hadoop bilgisayar mühendisleri DougCutting ve Mike Cafarella tarafından, başlangıçta Nutch açık kaynaklı arama motorunda ve web tarayıcısında işlemeyi desteklemek için yaratıldı. Google 2003 ve 2004 yıllarında Google Dosya Sistemini (GFS) ve MapReduce Programlama çerçevesini ayrıntılandıran teknik makaleler yayınladıktan sonra, Cutting ve Cafarella, önceki teknoloji planlarını değiştirdi ve Java tabanlı bir MapReduce uygulamasını ve Google'da modellenmiş bir dosya sistemi geliştirdi.

Hadoop uygulamaları ve kullanım örnekleri

Hadoop öncelikli olarak analitik kullanımlarına yöneliktir ve farklı türdeki verileri işleme ve saklama yeteneği onu büyük veri analitiği uygulamaları için özellikle uygun hale getirir. Büyük veri ortamları tipik olarak yalnızca büyük miktarda veriyi değil, aynı zamanda yapılandırılmış işlem verilerinden, internet tıklama kayıtları, web sunucusu ve mobil uygulama günlükleri, sosyal medya yayınları, müşteri e-postaları ve algılayıcı gibi yapılandırılmış ve yapılandırılmamış bilgi formlarına kadar çeşitli türleri içerir. Hadoop tabanlı büyük veri sistemleri için yaygın olarak kullanılan bir durum müşteri analitiğidir. Örnekler arasında müşteri kaybını tahmin etme, çevrimiçi reklamları web kullanıcılarına daha iyi hedeflemek için tıklama akışını analiz etme ve bir şirketin sosyal ağlardaki yorumlarına dayanarak müşteri duyarlılığını izleme çabaları yer alıyor. Sigortacılar poliçe fiyatlandırmasını analiz etmek ve güvenli sürücü indirim programlarını yönetmek gibi uygulamalar için Hadoop'u kullanıyor. Sağlık kuruluşları, Hadoop'un yardımıyla tedavileri ve hasta sonuçlarını iyileştirmenin yollarını ararlar (Grover, Malaska, Seidman and Shapira, 2015).

YARN, Hadoop kümelerinin, ve ApacheFlink gibi işleme motorlarıyla paralel olarak çalışan akış işleme ve gerçek zamanlı analitik uygulamalarını içerecek şekilde uygulayabileceği uygulamaları büyük ölçüde genişletti. Örneğin, bazı üreticiler, öngörücü bakım uygulamalarında, meydana gelmeden önce ekipman arızalarını tespit etmeye çalışmak için Hadoop'a akan gerçek zamanlı verileri kullanıyor. Sahtekarlık tespiti, web sitesi kişiselleştirme ve müşteri deneyimi puanlaması diğer gerçek zamanlı kullanım durumlarıdır.

Hadoop bu kadar geniş bir veri yelpazesini işleyebildiği ve saklayabildiği için, kurumların gelen bilgi akışları için geniş kapsamlı rezervuarlar olarak veri gölleri kurmalarını sağlar. Bir Hadoop veri deposunda, ham veriler çoğu zaman olduğu gibi depolanır, böylece veri bilimciler ve diğer analistler gerektiğinde tüm veri kümelerine erişebilir; veriler daha sonra diğer uygulamaları desteklemek için gerektiğinde analitik veya BT ekipleri tarafından filtrelenir ve hazırlanır.

Veri gölleri genellikle temizlenmiş işlem verisi kümelerini tutan geleneksel veri ambarlarından farklı amaçlara hizmet eder. Ancak, bazı durumlarda, şirketler Hadoop veri göllerini günümüz veri ambarları olarak görüyorlar. Her iki şekilde de, büyük veri analitiklerinin işletme karar alma konusundaki artan rolü, veri yönetimi ve veri güvenliği süreçlerini veri göl dağıtımlarında öncelikli hale getirmiştir.

1.13. ApacheSpark

Apache, büyük veri süreçleri için toplu ve akış terimlerini birleştirmekten adını alan bir proje modelidir. Her iki durumda da kullanabileceğimiz tek bir model.

ApacheSpark en yaygın kullanılan Apache projeleri ve inanılmaz hızlı büyük veri işleme (küme hesaplama) için popüler bir seçimdir. Gerçek zamanlı veri akışı, SQL, makine öğrenmesi, grafik işleme ve daha birçok özellik için yerleşik yeteneklere sahiptir.

1.14. R Programlama Nedir?

R programlama dili farklı amaçlar için geliştirilmiş pek çok uygulamayı içerisinde barındırmaktadır. Bu kapsamda, istatistik, meta analiz, veri madenciliği, yapay zeka vb. uygulamalara yönelik olarak geliştirilen ve kullanılan pek çok yazılımı (SPSS, SAS, CMA, WEKA, MATLAB vb.) kendi içerisinde barındırdığını belirtmek doğru olacaktır. R web tabanlı büyük veri ve veri yönetimi süreçlerinde de sıklıkla kullanılmaktadır. Bu kadar geniş bir alanda kullanımı bulunan R programlama dili; üniversiteler, bankalar, politik kampanyalar, teknolojik girişimler, fabrikalar, uluslararası kalkınma ve yardım kuruluşları, hastaneler gibi çok kuruluş tarafından değişik alanlarda sıklıkla kullanılmaktadır. R, yazılım dillerini bilen veya yazılım dillerine aşina olan programcılar ve araştırmacılar tarafından analizlerin açılır menüler vasıtasıyla gerçekleştirilmemesi sebebiyle daha sık kullanılmaktadır (Aksu, Eser ve Güzeller, 2019).

R istatistiksel yazılım geliştirme ortamı, son yıllarda popülerite anlamında diğer istatistiksel yazılımlara göre çok büyük bir yol almıştır. Bu popülerite göz önünde bulundurularak, R'ın yeni ve gelişmekte olan bir dil olduğunu söyleyebiliriz. Şaşırtıcı bir şekilde, R, 1992'de Robert Gentleman ve Ross Ihaka tarafından başlatılan bir araştırma projesi kapsamında geliştirilmeye başlanmıştır ve 1995'ten bu yana açık kaynak kodlu bir yazılım olarak piyasada yer almaktadır. R programlama dilinin S programlama diline rakip olarak çıktığı söylenebilir. Yazılımın popüleritesinin aniden ve çok hızlı bir şekilde artış göstermesi, veri biliminin bir kariyer ve çalışma alanı olarak ortaya çıkması ile doğru orantılıdır. Ancak, veri biliminin temelleri uzun yıllara dayanmaktadır. İstatistik, doğrusal cebir, yöneylem araştırması, yapay zeka ve makine öğrenimi, modern bir veri bilimcisinin kullandığı araçlara katkıda bulunmaktadır. R istatistiksel yazılım geliştirme ortamı, bu araçlar ile gerçekleştirilenlerin neredeyse hepsini tek bir ara yüz ile gerçekleştirebilme yeteneğini sahiptir. R istatistiksel yazılım geliştirme ortamının tarihsel olarak nasıl bir gelişim gösterdiği aşağıda açıklanmıştır (Aksu, Eser ve Güzeller, 2019; Satman, 2017):

1992: R, Yeni Zelanda'da bir araştırma projesi kapsamında Robert Gentleman ve Ross Ihaka tarafından geliştirilmeye başlandı.

1993: R'ın iki kodlu ilk sürümü Statlib'de yayınlandı.

1995: R, GPL2 lisansı altında açık kaynak kodlu bir yazılım olarak piyasaya sürüldü.

1997: R geliştiricileri "The R Core Group" adı altında bir araya geldiler.

1997: CRAN (The Comprehensive R Archive Network; Kapsamlı R Arşiv Ağı) oluşturuldu.

1999: R web sitesi, www.r-project.org kuruldu

1999: The R Core Group Viyana'da İstatistik konulu konferansta bir araya geldi.

2000: R'ın 1.0.0 sürümü yayınlandı.

2000: 1998'de ACM Yazılım Sistemleri Ödülünü S yazılımı ile alan John Chambers The R Core Group'a katıldı.

2001: R News (daha sonra The R Journal; R Dergisi oldu) kuruldu.

- 2003:** R Vakfı (The R Foundation) kuruldu.
- 2004:** Viyana’da ilk UseR! Konferansı gerçekleştirildi.
- 2004:**R’ın 2.0.0 sürümü yayınlandı.
- 2009:** R dergisinin ilk sayısı yayınlandı.
- 2013:**R’ın 3.0.0 sürümü yayınlandı.
- 2015:** R Vakfının katılımıyla ‘‘R Birliđi’’ kuruldu.
- 2016:** Yeni logo kullanılmaya başlandı.
- 2017:**R 3.3.3, R 3.4.0, R 3.4.1, R 3.4.2 ve R 3.4.3 yayınlandı.
- 2018:**R 3.4.4, R 3.5.0, R 3.5.1 ve R 3.5.2 yayınlandı.
- 2019:**Günümüzde R 3.5.2 hali hazırda en güncel R sürümüdür.
- 1992:** R, Yeni Zelanda’da bir araştırma projesi kapsamında Robert Gentleman ve Ross Ihaka tarafından geliştirilmeye başlandı.
- 1993:**R’ın iki kodlu ilk sürümü Statlib’de yayınlandı.
- 1995:** R, GPL2 lisansı altında açk kaynak kodlu bir yazılım olarak piyasaya sürüldü.
- 1997:** R geliştiricileri ‘‘The R CoreGroup’’ adı altında bir araya geldiler.
- 1997:** CRAN (The ComprehensiveR Archive Network; Kapsamlı R Arşiv Ađı) oluşturuldu.
- 1999:** R web sitesi, www.r-project.org [www.r-project.org] kuruldu
- 1999:** The R CoreGroup Viyana’da İstatistik konulu konferansta (InauguralDirections in Statistical Computing Conference) bir araya geldi.
- 2000:** R’ın 1.0.0 sürümü yayınlandı (Şubat).
- 2000:** 1998’de ACM Yazılım Sistemleri Ödülünü S yazılımı ile alan John Chambers The R CoreGroup’a katıldı.

2001: R News (daha sonra The R Journal; R Dergisi oldu) kuruldu.

2003: R Vakfı (The R Foundation) kuruldu.

2004: Viyana’da ilk UseR! Konferansı gerçekleştirildi.

2004:R’ın 2.0.0 sürümü yayınlandı.

2009: R dergisinin ilk sayısı yayınlandı.

2013:R’ın 3.0.0 sürümü yayınlandı.

2015: R Vakfının katılımıyla ‘‘R Birliđi’’ kuruldu.

2016: Yeni logo kullanılmaya başlandı.

2017:R 3.3.3, R 3.4.0, R 3.4.1, R 3.4.2 ve R 3.4.3 yayınlandı.

2018:R 3.4.4, R 3.5.0, R 3.5.1 ve R 3.5.2 yayınlandı.

2019:Günümüzde R 3.5.2 hali hazırda en güncel R sürümüdür.

1.15. Bulut Bilişim

Basitçe söylemek gerekirse, bulut bilişim, daha hızlı yenilik, esnek kaynaklar ve ölçek ekonomileri sunmak için, İnternet üzerinden (“bulut”) bilgi işlem hizmetlerinin sunucular, depolama, veritabanları, ağ iletişimi, yazılım, analitik, istihbarat ve daha fazlasının sunulmasıdır. Genellikle yalnızca kullanılan bulut hizmetleri için ödeme yapılır, işletme maliyetleri düşürülür, altyapı daha verimli bir şekilde çalıştırılır ve iş ihtiyaçları değıştikçe ölçeklenir (Microsoft Azure, 2019).

Bulut bilişimin arkasındaki temel bir kavram, hizmetin konumunun ve üzerinde çalıştığı donanım veya işletim sistemi gibi birçok detayın, büyük ölçüde kullanıcıyla ilgisiz olmasıdır. Bu akılda tutulması, bulutun metaforunun, genel telefon şebekesinin (ve daha sonra internetin) çoğu zaman bunun önemsiz olduğunu belirtmek için bir bulut olarak temsil edildiđi eski telekom ağ şemalarından ödünç alınmıştır. Bu, elbette aşırı basitleştirmedir (Ranger, 2018).

Bulut bilişim sayesinde, kendi verilerinizi saklamakla gelen bu farklı sorunları ortadan kaldırır, çünkü donanım ve yazılımı yönetmezsiniz bu Salesforce gibi deneyimli bir satıcının sorumluluğudur. Paylaşılan altyapı, bir yardımcı program gibi çalıştığını gösterir: Yalnızca ihtiyacınız olan şey için ödeme yaparsınız, yükseltmeler otomatiktir ve yukarı veya aşağı ölçeklendirmek kolaydır. Bulut tabanlı uygulamalar günler veya haftalar içinde kullanıma hazır olabilir ve çalışıyor olabilir ve daha düşük maliyetlidir (Salesforce, 2019).

Bulut bilgi işlem popülaritesi arttıkça, binlerce şirket bulut dışı ürünlerini ve hizmetlerini “bulut bilgi işlem” olarak yeniden değerlendiriyor. Bulut tekliflerini değerlendirirken her zaman daha derine inmek, donanım ve yazılım satın almak ve yönetmek zorunda kalınmaktadır.

Bulut bilişimin birçok avantajı vardır. Hizmeti sağlamak genellikle daha hızlıdır ve çoğu durumda bu hizmete anında erişilebilmektedir. Uzak kullanıcılar, bulut kaynaklarına fiziksel coğrafyayla sınırlı kalmak yerine, bağlantı kurdukları her yerden erişebilirler.

Genellikle, üç kategoriye ayrılır: hizmetlere veya altyapıya kimlerin erişimi olduğunu gösteren özel, genel ve karma. Genel bulut hizmetleri, hizmetleri satın almak veya kiralamak isteyen herkes için kullanılabilir. Özel bulut hizmetleri, işletmeler tarafından yalnızca çalışanları ve ortakları tarafından kullanılmak üzere kurulur. Hibrit bulut hizmetleri ikisini birleştirmektedir.

Konsept zaman içinde, genellikle Web veya mobil uygulamalar kullanılarak bir ağ bağlantısı üzerinden hızlı bir şekilde sağlanabilecek hemen hemen her hizmeti içerecek şekilde büyümüştür. Örneğin, bir müşteri bulutta Web barındırma siparişini (Amazon veya Rackspace) sipariş edebilir veya bulutta talep edilen filmler ve müzik (Apple iTunes, Amazon ve Netflix), depolama (Dropbox veya Google Drive) gibi dijital medya servislerini kullanabilir.), e-posta veya hatta konut ve ulaşım hizmetleri için sözleşme yapabilir (AirBnB veya Uber).

Bulut Bilişim (Cloud Computing) aşağıdaki temel özellikleri gösterir:

- Bulut bilişim, teknolojik altyapı kaynaklarını yeniden sağlama, ekleme veya genişletme konusunda kullanıcıların esnekliğini arttırabileceğinden, kuruluşlar için dinamiklik arttırılabilir ve kullanıcı dostu uygulamalar geliştirilebilir.
- Maliyet azaltma, bulut sağlayıcıları tarafından talep edilmektedir. Bir kamu bulutu dağıtım modeli, sermaye harcamalarını(örneğin, sunucu satın alma) operasyonel harcamaya dönüştürür. Altyapı genellikle üçüncü taraflarca sağlandığı ve bir kerelik ya da sık olmayan yoğun bilişim görevleri için satın alınması gerekmediğinden, bu iddia edilen giriş engellerini azaltır. Faydalı hesaplama bazında fiyatlandırma, kullanıma dayalı faturalama seçenekleriyle “iyi ayarlanmıştır”. Ayrıca, bulut bilişim kullanan projelerin uygulanması için daha az şirket içi BT becerileri gerekir. e-FISCAL projesinin son teknoloji ürünü depo maliyet yönlerini daha ayrıntılı olarak inceleyen çeşitli makaleler içerir; çoğu, maliyet tasarrufunun desteklenen faaliyetlerin türüne ve kurum içi mevcut altyapı türüne bağlı olduğu sonucuna varır.
- Cihaz ve konum bağımsızlığı, kullanıcıların konumlarından veya hangi cihazı kullandıklarından (Örneğin, PC, cep telefonu) bağımsız olarak bir web tarayıcısı kullanarak sistemlere erişmelerini sağlar. Altyapı saha dışında (genellikle bir üçüncü tarafça sağlanır) ve Internet üzerinden erişildiğinden, kullanıcılar herhangi bir yerden buna bağlanabilir.
- Bulut bilgi işlem uygulamalarının bakımı daha kolaydır, çünkü her kullanıcının bilgisayarına yüklenmesi gerekmez ve farklı yerlerden (örneğin, seyahat ederken farklı çalışma yerleri vb.) erişilebilir.
- Çoklu kiralama, kaynakların ve maliyetlerin büyük bir kullanıcı havuzunda paylaşılmasını şu şekilde sağlamaktadır:
 - Altyapının düşük maliyetli yerlerde (gayrimenkul, elektrik vb.) merkezileştirilmesi
 - Azami yük kapasitesi artar (kullanıcıların mümkün olan en yüksek yük seviyelerini karşılaması için kaynaklar ve ekipman için mühendisliğe ve ödemeye ihtiyaç duymazlar)
 - Genellikle sadece% 10 - 20 olan sistemler için kullanım ve verimlilik iyileştirmeleri yapılmıştır.

- Performans, servis sağlayıcıdan BT uzmanları tarafından izlenir ve tutarlı ve gevşek bir şekilde birleştirilmiş mimariler sistem ara yüzü olarak web servisleri kullanılarak oluşturulur.
- Verimlilik birden fazla kullanıcı aynı anda aynı veri çalışabilirsiniz zaman yerine kaydedilip gönderilecek şekilde beklemek yerine, artırılabilir. Alanlar eşleştirildiğinde bilgilerin tekrar girilmesi gerekmediğinden veya kullanıcıların bilgisayarlarına uygulama yazılımı yükseltmeleri yüklemeleri gerekmediğinden zaman kazanılabilir.
- Güvenilirlik, birden fazla yedekli sitenin kullanımıyla artar, bu da iyi tasarlanmış bulut bilişimi iş sürekliliği ve felaket kurtarma için uygun hale getirir.
- Güvenlik, verilerin merkezileştirilmesi, artan güvenlik odaklı kaynaklar vb. nedeniyle gelişebilir, ancak bazı hassas veriler üzerindeki kontrolün kaybı ve depolanan çekirdekler için güvenlik eksikliği konusunda endişeler devam edebilir. Güvenlik çoğu zaman diğer geleneksel sistemlerden daha iyidir, çünkü kısmen servis sağlayıcılar, birçok müşterinin başa çıkamayacağı veya çözecek teknik becerilerinden yoksun olan güvenlik sorunlarını çözmek için kaynakları tahsis edebilmektedir. Bununla birlikte, veriler daha geniş bir alana veya daha fazla sayıda cihaza ve ilgisiz kullanıcılar tarafından paylaşılan çok kiracılı sistemlere dağıtıldığında, güvenlik karmaşıklığı büyük ölçüde artmaktadır. Ayrıca, güvenlik denetim günlüklerine kullanıcı erişimi zor veya imkânsız olabilir. Özel bulut kuruluşları, kullanıcıların altyapı üzerindeki kontrolünü elinde tutma ve bilgi güvenliği kontrolünü kaybetmeme konusundaki istekleri kısmen motive edilmektedir.
- Bulut altyapımız yalnızca yaptığınız kadar güvenli. Bulut güvenliğini sağlama sorumluluğu sadece güvenlik ekiplerine değil aynı zamanda uygun güvenlik kontrollerinin kullanılmasını sağlamakla görevlendirilen DevOps ve operasyon ekiplerine de aittir. İşletmeler, kişisel tanımlayıcı bilgileri, finansal bilgileri veya sağlık bilgilerini yöneten veya içeren herhangi bir uygulama dahil olmak üzere buluta daha fazla düzenlenmiş iş yükü getirmek için isteklidir.
- Bulut bilgi işlem risklerinden kaçınmak için, bir bulut tarafından yönetilen hizmet sağlayıcısı, her seviyede yerleşik veri katmanından işletim sistemine kadar yerleşik güvenlik katmanları içermelidir; sektör lideri fiziksel güvenlik ve yüksek güvenliktir.

kişiler tarafından gerçekleştirilen düzenli güvenlik açığı taramalarıyla tam yapılandırılmış bir çözüm sunar uzmanları.

1.16. SQL Veritabanı

SQL (Structured Query Language) ilişkisel veritabanlarını yönetmek ve bunlardaki veriler üzerinde çeşitli işlemleri gerçekleştirmek için kullanılan standart bir programlama dilidir. Başlangıçta 1970'lerde oluşturulan SQL, yalnızca veritabanı yöneticileri tarafından değil, aynı zamanda analitik sorgular oluşturmak ve çalıştırmak isteyen veri entegrasyon komut dosyaları ve veri analistleri yazan geliştiriciler tarafından da düzenli olarak kullanılmaktadır (Rouse, 2016).

SQL kullanımı, veritabanı tablosunu ve dizin yapılarını değiştirmeyi; veri satırlarının eklenmesi, güncellenmesi, silinmesi, işlem işleme ve analitik uygulamaları için bir veri tabanından bilgi alt kümelerini almaktır. Sorgular ve diğer SQL işlemleri ifadeler olarak yazılan komutlar biçimini almaktadır. Yaygın olarak kullanılan SQL ifadeleri arasında seçim, ekleme, güncelleme, silme, oluşturma, değiştirme ve kesme sayılabilir (Rouse, 2016).

SQL, 1970'lerin sonunda ve 1980'lerin başında ortaya çıktıktan sonra ilişkisel veritabanları için fiili standart programlama dili haline geldi. SQL veritabanları olarak da bilinen ilişkisel sistemler, satır ve sütunlarda veri içeren bir dizi tablo içermektedir. Tablodaki her sütun bir veri kategorisine karşılık gelir. Örneğin, müşteri adı veya adresi her satır kesişen sütun için bir veri değerini içerir (Rouse, 2016).

1.17. NoSQL Veritabanı

NoSQL, anahtar-değer, belge, sütunlu ve grafik formatları da dahil olmak üzere çok çeşitli veri modellerini barındırabilen bir veritabanı tasarımı yaklaşımıdır. “Yalnızca SQL’i değil” anlamına gelen NoSQL, verilerin tablolara yerleştirildiği ve veri şeması veritabanı oluşturulmadan önce dikkatlice tasarlandığı geleneksel ilişkisel veritabanlarına bir alternatiftir. NoSQL veritabanları özellikle büyük dağıtılmış veri kümeleriyle çalışmak için yararlıdır (Rouse, 2017).

NoSQL terimi, ilişkisel veritabanı yönetim sistemini hazırlayan bazı veritabanlarına uygulanabilir, ancak daha çok bulut ve web uygulamalarında büyük ölçekli veritabanı kümelemesi amacıyla 2000’li yılların başında oluşturulan veritabanlarına atıfta bulunur. Bu

uygulamalarda, performans ve ölçeklenebilirlik gereklilikleri, RDBMS'nin işlemsel kurumsal uygulamalara sağladığı hızlı ve katı veri tutarlılığına duyulan gereksinime karşı geldi (Rouse, 2017).

Özellikle, NoSQL sistemlerinin yerleşik bir ilişkisel şema izlemesi gerekmemektedir. Google ve Amazon gibi büyük ölçekli web organizasyonları, dar operasyonel hedeflere odaklanmak için NoSQL veritabanlarını kullanmakta ve yüksek dereceli veri tutarlılığının gerekli olduğu durumlarda ek olarak ilişkisel veritabanlarını kullanmaktadırlar (Rouse, 2017).

Web ve bulut uygulamaları için ilk NoSQL veritabanları, veri yönetiminin çok özel özelliklerine odaklanma eğilimindeydi. Çok büyük miktarlarda veri işleme ve bu verileri bilgisayar kümelerine hızlı bir şekilde dağıtma yeteneği, web ve bulut tasarımında istenen özelliklerdi. Bulut ve web sistemlerini uygulayan geliştiriciler, sürekli güncellenen uygulamalarda hızlı değişikliklerin daha iyi yapılmasını sağlamak için esnek veri şeması (hiç şema içermemesi veya hiç şema oluşturmamayı) kullanmaktadır (Rouse, 2017).

1.18. BigQuery

Büyük veri, biçimlendirilmiş ve biçimlendirilmemiş verilerin büyük yığınının işlenmesi içindir. Bu kadar büyük veri analizinin amacı pazarlama ve iş analizidir. Piyasada sayısız veri depolama desteği bulunmaktadır. BigQuery, bulut platformu üzerinden büyük veri desteği en son ve en iyi çözümdür. Piyasadaki üç Cloud rakibi arasında Google, Cloud platformuna yatırım yapmak için yatırım yapıyor. Google BigQuery, Google'ın şu anda büyük veri analizi alanını işgal eden tamamen yönetilen, sunucusuz veri ambarı çözümüdür (Verma, 2018).

Google BigQuery, tüm ölçeklerde Büyük veri analitiklerindeki veri analistlerine yardımcı olan işletmeler için oldukça ölçeklenebilir ve hızlı bir veri deposudur. Ayrıca, Google BigQuery, veri analistlerinin daha üretken olmalarını sağlayan düşük maliyetli bir depodur. Yönetilmesi gereken bir altyapı yoktur ve her zaman veri analizinden elde edilen anlamlı görüşlere odaklanılmaktadır (Verma, 2018).

Google BigQuery'nin düşük maliyetli bir veri ambarı çözümü olmasından kaynaklanmaktadır. Buna ek olarak, BigQuery işletmelerin depoyu birkaç saniye içinde kurmalarına izin verir ve biri verileri hemen sorgulayabilir. BigQuery, hızlı veri entegrasyonu için JDBC ve ODBC sürücülerini sağlamak için SQL kullanır. Ayrıca, Google BigQuery

kullanarak, büyük erişim ve kapasite ile sorunsuz ölçeklendirme yapmakta mümkündür. Ayrıca, verileri analiz eden bir kullanıcı olarak performans optimizasyonu ile birlikte paralel yürütmede yapılabilmektedir (Verma, 2018).

Google'ın BigQuery Veritabanı, GDELT gibi veri kümeleri için özel olarak tasarlanmış ve tüm veri kümesinde gerçek zamana yakın anlık sorgulama sağlamaktadır. Bu, GDELT'e nasıl eriştiğiniz, hangi tür operatörleri kullandığınız veya sorgunuzun karmaşıklığına bakılmaksızın, sonuçların neredeyse gerçek zamanlı olarak görüleceği anlamına gelmektedir(The GDELT Project, 2019).

BigQuery'nin aşağıdaki özellikleri, kullanıcıların GDELT veri kümeleriyle verimli bir şekilde etkileşim kurmalarını sağlamaktadır (Leetaru ve Hoffa, 2015):

Ölçeklenebilirlik ve Esneklik: GDELT 'nin veri setleri, birden fazla formatta yüzlerce trilyon veri noktasını toplu olarak kodlar. Protestolar ve çatışma olayları gibi bazı veri akışları, İlişkisel Veritabanı Yönetim Sistemleri (RDBMS) kullanılmak üzere tasarlanmış ve on yıllarca kullanım için optimize edilmiş, yüksek düzeyde yapılandırılmış şemalara sahiptir.

Anlatıların ve duyguların katalogları gibi diğer akışlar, son derece küçük ölçeklerde kullanılmak üzere tasarlanan temelde yeni meta veri uygulamalarını temsil eder, bu ölçeklerde kodlamanın çok az önceliği vardır. Karmaşıklığa ek olarak, değerlendirilecek boyutların sayısı sürekli olarak artmakta ve sürekli genişleyebilen bir akışkan şemasını gerektirmektedir. Değerlendirilen her boyut diğer bilgilere göreceli yakınlığı veya yoğunluğu gibi sayısal bilgileri kodlamalıdır. Bu, BigQuery'nin sağladığı karmaşık yuvalamayı, sayısal değerleri ve sabit genişlemeyi destekleyen esnek bir veri formatını gerektirmektedir (Leetaru ve Hoffa, 2015).

Yeni Sütunların Sürekli Eklenmesi: GDELT'nin veri kümelerinden biri, izlenen haberlerin her birinde milyonlarca tema ve binlerce duygu varlığını, içeriğini ve yoğunluğunu belirlemektedir. Temaların ve duyguların listesi, zaman içinde sürekli artmakta ve her biri puanla ilişkili sayısal değerleri depolaması gereken satır başına milyonlarca boyutun ne olduğunu etkili bir şekilde sorgulama ve analiz etme yeteneği gerektirmektedir. GDELT, verileri yuvalanmış ayrılmış biçimlerde depolamak ve seçilen değerleri sorgu zamanında çıkarmak için BigQuery'nin gelişmiş düzenli ifade desteğini kullanmaktadır (Leetaru ve Hoffa, 2015).

Kombine Gerçek Zamanlı ve Tarihi Veriler: GDELT'in gerçek zamanlı güncellemelerinden çıkan desenler, önemlerini, uygunluklarını ve altında yatan itici güçleri belirlemek için tarihi arşivler bağlamında düşünülebilmektedir. BigQuery'nin yüksek hızlı akış ekleme API'si, gerçek zamanlı analizler için güçlü bir temel sunmaktadır. BigQuery, en güncel iş verilerini analiz etmek için anında hazır hale getirerek, gerçek zamanlı süreçleri analiz etmeyi sağlar. Kırılma olaylarının analizini etkinleştirmek için gerçek zamanlı güncellemelerin hemen kullanılabilir olması gerektiğinden, BigQuery'nin akış eklemeleri yoluyla desteklediği hem gerçek zamanlı hem de geçmiş veri depolarında birleşik sorgulamaya izin veren bir ortam gereklidir. BigQuery'de GDELT'e sahip olmanın en çığır açan yanı, yalnızca karmaşık sorgulama ve verilerin çıkarılması için değil aynı zamanda ilk kez gerçek dünya analizlerinin tamamen veritabanında yapılmasına izin vermesidir. Son 40 yılda, dünyadaki en önemli çatışma etkileşimini aylara göre hesapladığımızı veya bir dizi ülke arasındaki farklı ilişki sınıfları arasında çapraz sekanslı bir korelasyon yapılabilir. Bu tür sorgular tamamen BigQuery'nin içinde çalıştırılabilir ve birkaç saniye içinde geri gönderilebilir(Google Cloud Platform Blog, 2014).

*Birçok Sütunda Özel Endeks İçermeyen Arama :*GDELT'in veri setlerinden biri, 310 milyondan fazla satır ve 59 sütundan oluşan 40 yıllık küresel olay arşividir. Sorgular genellikle birçok sütunu birleştirir ve her bir sorgu farklı bir sütun kombinasyonuna erişmektedir. Tek bir sütun veya sütun kümesi yeterli azaltma gücü sağlamaz; bu da geleneksel bir dizine alınmış RDBMS modelini kullanmayı imkansız hale getirmekte ve bunun yerine BigQuery tarafından kullanılan benzer bir dizinsiz sorgu işleme modeli gerekmektedir (Leetaru ve Hoffa, 2015).

Kamu Erişimi: Tüm GDELT veri akışları açık veri olarak serbestçe kullanılabilir. Bu, veri barındırma ve yönetmeyle ilgili kaynakları sorgulama ile ilgili olanlardan ayıran bir platformda çalışması gerektiği anlamına gelmektedir. BigQuery, veri kümelerinin genel olarak erişilebilir olmasını sağlamaktadır (Leetaru ve Hoffa, 2015). Birden çok konumda ücretsiz veri ve işlem çoğaltma olanağı, verilerinizin olağanüstü arıza durumlarında bile sorgulamaya hazır olmasını sağlamaktadır. BigQuery, şeffaf ve otomatik bir şekilde, ek ücret ve kurulum olmadan yüksek kullanılabilirlik olanağının yanı sıra dayanıklı ve çoğalmış depolama alanı sunmaktadır (Leetaru ve Hoffa, 2015).

Gelişmiş hesaplama: GDELT sorguları genellikle, belirli bir belgedeki konumlarla temaların eşleştirilmesi gibi potansiyel olarak haritalama için son bir coğrafi histogramın

çıkması için terabayt büyüklüğündeki verilerin işlenmesini gerektiren karmaşık bir mantık içermektedir. Bu nedenle, GDELT, tamamen veri tabanı platformunda eşit bir şekilde sofistike kontrol akışları ve algoritmalar gerçekleştirebilmeli ve bu sayede bilgiyi verilere getirebilmelidir. BigQuery'nin kullanıcı tanımlı fonksiyonları bunu mümkün kılmaktadır (Leetaru ve Hoffa, 2015).

Tüm Verinin Veritabanındaki Analizi: Yukarıda belirtilen daha geleneksel analizlere ek olarak, bazı analiz türleri, veritabanının tamamını kendilerine karşı etkin bir şekilde analiz etme kabiliyetine ihtiyaç duymaktadır. Örneğin, haber medyasında ifade edildiği şekilde tarihin döngülerini ve modellerini incelemek, şeffaf bir hesaplama ve veri hareketi ölçeklendirmesi gerektiren tüm veritabanını hareketli bir pencerede çapraz ilişkilendirme yeteneğini gerektirmektedir. Böyle bir analiz için gerekli olan çok sayıda işlemci, BigQuery gibi bulutta barındırılan bir ortama ihtiyaç duyulmaktadır (Leetaru ve Hoffa, 2015).

1.19. Büyük Veri Analizinde Kullanılan Teknikler

Bu kısımda, büyük veri analizinde en çok kullanılan tekniklerden bahsedilmiştir.

1.19.1. A/B Testi

A / B testi (ayrıca bölünmüş test veya kova testi olarak da bilinir), hangisinin daha iyi performans gösterdiğini belirlemek için bir web sayfasının veya uygulamanın iki sürümünü birbiriyle karşılaştırmanın bir yöntemidir. AB testi esasen, bir sayfanın iki veya daha fazla varyantının kullanıcılara rastgele gösterildiği ve belirli bir dönüşüm hedefi için hangi varyasyonun daha iyi performans gösterdiğini belirlemek için istatistiksel analiz kullanılan bir deneydir.

Bir A / B testi çalıştırmak için, tek bir değişkende değişiklik yaparak bir içerik parçasının iki farklı sürümünün oluşturulması gerekir. Ardından, bu iki sürüm benzer boyuttaki iki izleyiciye gösterilerek hangisinin belirli bir zaman diliminde daha iyi performans gösterdiği analiz edilir.

1.19.2. Doğal Dil İşleme

Özellikle büyük miktarlarda doğal dil verilerini işlemek ve analiz etmek için bilgisayarları nasıl programlayacağınızla ilgili olarak bilgisayar ve insan (doğal) dilleri arasındaki etkileşimlerle ilgili bir bilgisayar bilimi, bilgi mühendisliği ve yapay zeka alanıdır. Doğal Dil İşleme, bilgisayarların insanın doğal dilini anlamalarına yardımcı olmak için kullanılan teknolojidir. Doğal Dil İşleme genellikle NLP olarak bilinmektedir. Doğal dili kullanan bilgisayarlar ve insanlar arasındaki etkileşimle ilgilenen yapay zekanın bir dalıdır. NLP'nin nihai amacı insan dillerini değerli bir şekilde okumak, deşifre etmek, anlamak ve anlamlandırmaktır. NLP tekniklerinin çoğu insan dilinden anlam çıkarmak için makine öğrenmesine dayanır. NLP, yapılandırılmamış dil verilerinin bilgisayarların anlayabileceği bir formata dönüştürülmesi için doğal dil kurallarını tanımlamak ve çıkarmak için algoritmalar uygulama gerektirmektedir.

1.19.3. Veri Madenciliği

Gizli bağlantıları keşfetmek ve gelecekteki eğilimleri tahmin etmek için verileri araştırmak çok uzun bir geçmişe sahiptir. Bazen “veri tabanlarında bilgi keşfi” olarak adlandırılan “veri madenciliği” terimi 1990'lara kadar kullanılmamıştır. Ancak temeli iç içe geçmiş üç bilimsel disiplinden oluşur: “istatistik” (veri ilişkilerinin sayısal çalışması), “yapay zeka” (yazılım ve/veya makineler tarafından görüntülenen insan benzeri zeka) ve “makine öğrenmesi“(verilerden tahmin yapmayı öğrenebilecek algoritmalar).

Veri madenciliği, şirketler tarafından ham verileri yararlı bilgilere dönüştürmek için kullanılan bir işlemdir. İşletmeler, büyük veri gruplarındaki kalıpları aramak için yazılım kullanarak, daha etkili pazarlama stratejileri geliştirmek, satışları artırmak ve maliyetleri düşürmek için müşterileri hakkında daha fazla bilgi edinebilirler. Veri madenciliği, etkili veri toplama, depolama ve bilgisayar işlemeye bağlıdır.

Haberleşme: Büyük veri, iletişimciler için verimli bir temeldir ve şirketlere toplum için bilgi üretiminde eşlik etme konusunda eşsiz bir fırsattır. Dahili olarak, bir şirketin çalışan etkileşim araçları ve kanalları aracılığıyla topladığı veriler, şirketin iç iletişimini ve mesajlarını ayarlamasına yardımcı olabilir.

Sigorta: Sigorta bilgilerinin daha önemli uygulamalarından bazıları, yatırım portföylerinden ziyade müşteri hizmetleri kayıtları ile daha yakından ilgilidir. Sigortada

müşteri hizmet kayıtları temelli olan büyük verilerin kullanımını son yıllarda hızla artmaya devam etmektedir. Bu yüzden birçok şirket iş operasyonlarını canlandırmak için daha yaratıcı yollar aramaktadır. Bu da, sigorta şirketlerinin bilgileri takip etmelerine yardımcı olmak için teknoloji çözümleri geliştiren bütün bir sektörün yaratılmasına öncülük etmektedir.

Eğitim: Büyük veri sistemleri, eğitimcilerin öğrencileri doğru değerlendirmelerine yardımcı olurken, ilerlemelerini ve ilerleme olasılıklarını sürekli izler. Bu, gelecekteki kolej ve üniversiteye kabul şanslarını artırmak için öğrencilerin lisede yeterince performans göstermelerini sağlamak için önemli bir kaynaktır.

İmalat: Proses üretimi, tarihçilerde on yıllarca depolanan büyük miktarda veri içermektedir. Bu veriler, üreticiler için potansiyel olarak büyük değer taşımaktadır. Veri analitiği, tesis personelinin makine verilerini hızlı bir şekilde yakalamasına, temizlemesine ve analiz etmesine ve performanslarını geliştirmelerine yardımcı olabilecek bilgileri ortaya çıkarmasına yardımcı olmaktadır.

Bankacılık: Bankacılıktaki büyük veri çözümleri, şirketlerin departmanlar arasında şube (bireysel çalışan) performans ölçümlerini gerçek zamanlı olarak toplamasını, anlamlandırmasını ve paylaşmasını sağlar. Bu, günlük işlemlerde daha iyi görünürlük ve herhangi bir sorunu proaktif olarak çözme kabiliyeti artışı anlamına gelir.

Perakende: Gittikçe daha rekabetçi bir pazar alanında, perakendecilerin müşterileri hakkında her şeyi bilmeleri gerekir: kim olduklarını, ne aldıklarını, ne zaman aldıklarını, neden aldıklarını... Ayrıca, müşteri satın alma davranışları hakkında dolaşan veri miktarı sayesinde, bu bilgileri toplamak, düzenlemek, temizlemek ve analiz etmek için gerekli teknolojiye sahip olmaları koşuluyla, tüm bu soruları ve daha fazlasını cevaplayabilirler. Bir iş zekası girişiminin bir parçası olarak veri madenciliği ile perakendeciler gerçek zamanlı olarak gerçek sorulara gerçek cevaplar verebilir.

1.19.4. Kümeleme Analizi

Kümeleme analizi, veri içindeki yapıları tanımlamaya çalışan keşifsel bir analizdir. Kümeleme analizi ayrıca bölümlenme analizi veya sınıflandırma analizi olarak da adlandırılır. Daha spesifik olarak kümeleme analizi ile grüplamanın daha önce bilinmediği durumlarda homojen vaka grüpları belirlenmeye çalışılır. Açıklayıcı bir analiz türü olduğu için kümeleme analiziyle bağımlı ve bağımsız değişkenler arasında bir ayırım gerçekleştirilmemektedir.

1.19.5. Sınıflandırma

Sınıflandırma analizi, analizimizin doğruluğunu artırmak için kategorilere/sınıflara öge atama işleminin denetlenmesi işlemidir.

1.19.6. Kalabalığın Gücü

Kalabalığın gücüne ilişkin platformlar, çevrimiçi kitle arasında kimin akıllıca bir ses olduğunu belirlemek için makine öğrenmesi ve gelişmiş algoritmalar kullanır ve sonra da bu girişi iç görünlere dönüştürmektedir.

1.19.7. Makine Öğrenmesi

Makine öğrenimi (ML), bilgisayar sistemlerinin açık talimatlar kullanmadan belirli bir görevi yerine getirmek için kullandığı algoritmaların ve istatistiksel modellerin bilimsel bir çalışmasıdır, kalıplara ve çıkarımlara dayanır. Yapay zekanın bir alt kümesi olarak görülür.

1.19.8. Sinir Ağları

Sinir ağları, insan beyninin çalışma prensibine benzer bir şekilde modellenen, kalıpları tanımak için tasarlanmış bir algoritmalar kümesidir. Duyusal verileri bir tür makine algısı, ham girdiyi etiketleme veya kümeleme yoluyla yorumlarlar.

1.19.9. Sosyal Ağ Analizi

Sosyal Ağ Analizi (SAA), sosyal yapıları ağlar ve grafik teorisi kullanarak araştırma sürecidir. Ağa bağlı yapıları, düğümler (tek tek aktörler, insanlar veya ağ içindeki şeyler) ve bunları bağlayan bağlar, kenarlar veya bağlantılar (ilişkiler veya etkileşimler) açısından karakterize eder. Araştırmanın ilerleyen bölümlerinde Sosyal Ağ Analizinden daha ayrıntılı bir şekilde bahsedilmiştir.

1.19.10. Örüntü Tanıma

Örüntü tanıma, verilerdeki örüntülerin ve düzenlerin otomatik tanınmasıdır. Örüntü tanıma, yapay zeka ve makine öğrenmesi ile yakından ilgilidir, veri madenciliği ve veri tabanlarındaki bilgi keşfi (VBB) gibi uygulamalar ile birlikte ve bu terimlerle sıklıkla birbirinin yerine kullanılır.

1.19.11. Zaman Serileri

Zaman Serileri, bir süre boyunca eşit zaman aralıklarında ölçülen iyi tanımlanmış bir veri noktaları dizisidir. Özel olarak veya düzensiz bir şekilde toplanan veriler bir zaman serisi oluşturmaz. Zaman serisi analizi, zaman serisi verilerini analiz etmek ve verilerle ilgili anlamlı istatistik ve özellikleri çıkarmak için istatistiksel yöntemlerin kullanılmasıdır. Zaman Serileri Analizi, zaman serileri veri noktalarında belirli bir eğilime yol açan temel güçlerin neler olduğunu anlamamıza yardımcı olur ve ona uygun modelleri uygulayarak veri noktalarını tahmin etmemize ve izlememize yardımcı olur (Researchoptimus, 2019).

1.19.12. Duygu Analizi

Duygu Analizi, bir yazının olumlu, olumsuz ya da tarafsız olduğunu belirleme sürecidir. Metin analizi için bir duyarlılık analiz sistemi bir cümle veya cümle içindeki varlıklara, konulara, temalara ve kategorilere ağırlıklı duyarlılık puanları atamak için doğal dil işleme (NLP) ve makine öğrenme tekniklerini birleştirir (Lexalytics, 2019). Duygu analizi, bir şirketin çevrimiçi konuşmaları izlerken marka, ürün veya hizmetlerinin sosyal duyarlılığını anlamalarına yardımcı olur. Bununla birlikte, sosyal medya akışlarının analizi genellikle sadece temel duyarlılık analizleri ve sayıya dayalı ölçümler ile sınırlıdır (Gupta, 2018).

Derin öğrenmedeki son gelişmelerle, algoritmaların metni analiz etme kabiliyeti önemli ölçüde gelişmiştir. Gelişmiş yapay zeka tekniklerinin yaratıcı kullanımı, derinlemesine araştırma yapmak için etkili bir araç olabilmektedir. Bir müşteri için gelen müşteri görüşmesini, satırlara dayalı olarak sınıflandırmanın önemlidir (Gupta, 2018).

Duygu analizi, büyük işletmelerdeki veri analistlerinin kamuoyunu ölçmesine, farklı pazar araştırmaları yapmasına, marka ve ürün itibarını izlemesine ve müşteri deneyimlerini anlamasına yardımcı olur. Bununla birlikte, veri analizi şirketleri, kendi müşterilerine yararlı bilgiler sunmak için genellikle üçüncü taraf duyarlılık analizi API'lerini kendi müşteri deneyimi yönetimi, sosyal medya izleme veya işgücü analizi platformu ile birleştirirler (Lexalytics, 2019).

1.19.13. Veri Görselleştirme

Veri görselleştirme, verileri ham rakamlardan grafik, harita ve karmaşık gösterge tabloları gibi grafiksel bir gösterime dönüştürme pratiğidir. Neyin önemli olduğunu (anlamını), nasıl geliştiğini (tarihini) ve gerçek hayatta tam olarak nasıl çalışabileceğini (örnekler) göstermektedir (Hackernoon, 2018).

Veri görselleştirme, uygulamalı istatistik ve makine öğrenmesinde önemli bir tekniktir. İstatistikler aslında nicel açıklamalara ve verilerin tahminlerine odaklanır. Veri görselleştirme, nitel bir anlayış kazanmak için önemli bir araç takımı sunar. Görselleştirme, bir veri kümesini keşfetme ve tanıma konusunda yardımcı olabilir ve kalıpların, bozuk verilerin, aykırı değerlerin ve çok daha fazlasının tanımlanmasına yardımcı olabilmektedir.

Veri görselleştirme yöntemleri, bilgilerin grafiksel gösterimlerinin oluşturulmasına karşılık gelir. Görselleştirme, veri analizinin önemli bir bölümünü oynar ve karmaşık sayısal veya gerçek rakam kümelerini kullanarak büyük verileri gerçek zamanlı bir yapıda yorumlamaya yardımcı olur.

Araştırmacılar görselleştirmenin en etkin duyumuz olduğu konusunda hemfikirdir: algıladığımız, öğrendiğimiz veya işlediğimiz bilgilerin %80-85'ine görselleştirme aracılığıyla öğrenilir. Görselleştirme, verileri anlamaya ve yorumlamaya çalıştığımızda ya da göreceli önemlerini belirlemek için yüzlerce veya binlerce değişken arasındaki ilişkileri ararken daha da fazla önem kazanır. Önemli ilişkileri ayırt etmenin en etkili yollarından biri, gelişmiş analizler ve anlaşılması kolay görselleştirmelerdir. Veri görselleştirme pratik olarak her bilgi alanına uygulanır. Çeşitli disiplinlerdeki bilim adamları, karmaşık olayları modellemek ve hava durumu modelleri, tıbbi durumlar veya matematiksel ilişkiler gibi doğrudan gözlenemeyen olayları görselleştirmek için bilgisayar tekniklerini kullanır (Medium, 2019).

1.19.14. Denetimli Öğrenme

Denetimli öğrenmede, algılanan *algoritmalar*, algoritma eğitimi sırasında tahminlerin doğruluğu hakkında geri bildirim sağlamanın yanı sıra, hem girdi hem de istenen çıktıyı sağlamak için makine öğrenim becerileri ile bir veri bilimcisi veya veri analisti gerektirir. Veri bilimcileri, modelin hangi değişkenleri veya özellikleri analiz edeceğini ve tahminleri geliştirmek için kullanacağını belirler. Eğitim tamamlandığında, algoritma öğrenilenleri yeni verilere uygulayacaktır.

1.19.15. Denetimsiz Öğrenme

Denetimsiz öğrenmede ise istenen sonuç verileriyle eğitilmesi gerekmez. Bunun yerine, verileri gözden geçirmek ve sonuçlara varmak için derin öğrenme denen yinelemeli bir yaklaşım kullanılır. Denetimsiz öğrenme algoritmaları (aynı zamanda sinir ağı olarak da adlandırılır) görüntü tanıma, konuşma-metin ve doğal dil üretimi de dahil olmak üzere denetimli öğrenme sistemlerinden daha karmaşık işlem görevleri için kullanılır. Bu sinir ağı, milyonlarca eğitim verisi örneğini tarayarak ve birçok değişken arasında sıklıkla ince korelasyonları otomatik olarak tespit ederek çalışır.



2. BÖLÜM

2. YÖNTEM

Bu bölümde öncelikle, dünya haber medyasını takip eden GDELT veri seti hakkında temel bilgiler verilmiştir. Daha sonra sosyal ağ analizinin kullanıldığı çatışma olayları ile ilgili literatür taranması yapılmış ve sosyal ağ analizinin teorik alt yapısını oluşturan temel kavramlar ve verilmiştir. Son olarak GDELT veri setinden elde edilen çatışma verileri UCINET ve Netdraw programı kullanılarak analiz edilmiş ve elde edilen bulgular sunularak yorumlanmıştır.

2.1. Veri Seti (GDELT Projesi)

Bu çalışmada kullanılan veriler dünyanın en büyük açık veritabanı olan GDELT'ten elde edilmiştir. GDELT Projesi kendisini “şimdiye kadar oluşturulan insan toplumunun en büyük, en kapsamlı ve en yüksek çözünürlüklü açık veritabanı” olarak tanımlamaktadır (Pratt, 2015). GDELT, dünyanın her köşesinden 1979'a kadar uzanan binlerce haber kaynağına dayanan, insan toplumunun nicel veri tabanıdır.

Georgetown Üniversitesi araştırmacısı olan Kalev Leetaru tarafından oluşturulan GDELT Projesi, var olan insan toplumundaki en büyük açık erişimli veritabanıdır. Olay Veri Tabanı, 1979'daki fiziksel etkinlikleri içerir ve Global Bilgi Grafiği, dünyadaki herkesi, organizasyonu, şirketi ve konumu birbirine bağlayan büyük bir ağıdır. Google Ideas tarafından desteklenen GDELT Projesi, dünyanın her köşesinden 100'den fazla dilde dünyanın yayın, baskı ve web haberlerini izler ve global toplumu yönlendiren insanları, yerleri, kuruluşları, sayıları, temaları, kaynakları ve olayları tanımlar. GDELT her gün her saniye, tüm dünyada bilgi işlem için ücretsiz bir açık platform oluşturmaktadır.

Google Ideas tarafından desteklenen GDELT Projesi, gerçek zamanlı bilgileri ve meta verileri tüm dünyayla paylaşır. Bu kodlanmış meta veriler (ancak makalelerin metni değil) daha sonra her 15 dakikada bir güncellenen ve bilgilerin çok dilli bir açıklamalı indeksini sağlayan açık bir veri akışı olarak yayınlanır. Yayın, baskı ve çevrimiçi haber kaynaklarını içerir. Proje trilyonlarca veri noktasına sahip bir veri tabanını paylaşıyor. Veriler indirilebilir csv uzantılı dosyalar olarak mevcut olmasına rağmen, az sayıda kullanıcı terabayt boyutunda veriyi indirmek ve etkili bir şekilde sorgulamak ve analiz etmek için depolama kapasitesine ve işlem gücüne sahiptir. Google'ın BigQuery platformu, bu devasa bilgi kaynağı ile

etkileşime girmenin bir yolunu sunar. GDELT, Büyük Veri'nin net bir örneğidir (Sagi ve Labeaga, 2016).

GDELT 2.0 Olay Veritabanı, protestolar ve askeri saldırılardan barışçıl itirazlara ve diplomatik değişimlere kadar 300'den fazla kategoride dünya çapındaki olayların küresel bir kataloğudur. Her olay kaydı, olayın birçok farklı özelliğini yakalayan 58 alanı ayrıntılandırır. GDELT 2.0 Şubat 2015'ten günümüze kadar çalışmakta, 15 dakikada bir güncellenmekte ve 19 Şubat 2016'dan itibaren 326 milyon 103 milyon farklı olaydan oluşmaktadır. Bu veri seti, izlenen tüm içeriğin 65 ana dilde makine çevirisi kapsamını kullanmaktadır. Ayrıca, her bir olayın her bir sözünü kaydeden ayrı bir "Mentions" tablosu sağlayarak GDELT 1.0'a genişler ve bu sözlerin her birinin dengelenmesi, bağlamı ve güvenini sunar (Google Cloud Platform, 2018).

Bu genel veri kümesi Google BigQuery'de barındırılmaktadır ve BigQuery'nin 1TB/ay ücretsiz seviye işlemesine dahil edilmiştir. Bu, her kullanıcının her ay 1 TB ücretsiz BigQuery işlemi alması anlamına gelir; bu, bu halka açık veri kümesinde sorguları çalıştırmak için kullanılabilir (Google Cloud Platform, 2018).

2.2. Literatür Taraması

Son yıllarda, fiziksel ve sosyal bilimler arasında ağ araştırmalarına olan ilgide büyük artış yaşanmıştır. Sosyal bilimciler için, ağlar teorisi, bir altın madeni olmuştur ve bu da psikolojiden ekonomiye kadar çok çeşitli disiplinlerde sosyal olayların açıklamalarını vermektedir. Sosyal bilimlerdeki en güçlü fikirlerden biri, bireylerin sosyal ilişkiler ve etkileşimlerin karmaşık ağlara gömülü olduğu düşüncesidir. Sosyal ağ teorisi, Platon zamanından bu yana sosyal felsefeyi meşgul eden bir soruya, yani sosyal düzen sorununa cevap verir: özerk bireylerin kalıcı, işleyen toplumlar oluşturmak için nasıl bir araya gelebileceğini göstermektedir. Ağ teorisi ayrıca bireysel yaratıcılıktan kurumsal karlılığa kadar sayısız sosyal fenomen için açıklamalar da sunmaktadır (Borgatti ve ark., 2009).

Savaşın nedenleri tipik olarak, savaşçı devletlerin kendi siyasi, askeri, ekonomik veya sosyal özellikleri gibi bağıntıları incelenerek anlaşılabilir (Vasquez 2000). Bununla birlikte, uluslararası ilişkilerin hem tanım olarak hem de teorik temellerle, karmaşık bir uluslararası sistemdeki sayısız devlet arasındaki etkileşimlerin sonucu olduğu düşünüldüğünde, bu garip bir bakış açısidir (Jervis, 1997). Birimlerin birbirine bağlı bir sistemdeki davranışını anlamak, bireysel devletlerin özelliklerinin ötesine geçen bir yaklaşım olan farklı bir yaklaşımı

gerektirmektedir. Bu durum, savaşın devletler ağı içinde bir fenomen olarak ortaya çıkmasını inceler.

Çatışmayı anlamak için bir ağ modelinin uygulanmasının en iyi örneği, Wayne Zachary (1977)'nin çalışmasıdır. Zachary, bir ağın kırk üyesinin gelişmekte olan bir anlaşmazlığın hangi tarafının sadece biriyle olan önceki ilişkileri hakkında biraz bilgi sahibi olacağı konusunda tam olarak ne olacağını tahmin edebileceğini bulmuştur.

Ağlardaki birkaç farklı merkezilik biçimi oldukça farklı sonuçların olabileceğini göstermiştir. Bu farklı merkezilik türleri, doğrudan iktidar ve özerklik ile ilgili oldukları için, çatışma ve barış sorunlarıyla büyük ölçüde ilgilidir. Tek fark, yakınlık merkeziyle iyilik merkezi arasında vardır. Herhangi bir belirli ağ veya sistem sadece belirli derecelerde yakınlık veya netlik sergileme olarak nitelendirilemez, aynı zamanda analist bir ağdaki her bir merkez için geçici gözlemci tarafından görünmeyen bir bireysel merkezilik indeksi ve bir merkezilik merkezini belirleyebilir (Johnson & Jordan, 2007).

Birkaç merkezilik indeksi birbiriyle ilişkili olsa da, ayırt edici özellikleri ağı anlamak veya değiştirmek için çok önemli olabilmektedir. Bireysel düğümler için, yüksek yakınlık merkeziliği özerklik, başkaları tarafından kontrol edilmekten bağımsızlık anlamına gelmektedir. Diğer taraftan, Arasındalık Merkeziliği (Betweenness Centrality) güç, başkalarının kontrolü için potansiyel anlamına gelir. Bu tür sonuçlarla, her türlü organizasyona uygulanabilecek bu resmi ağ özelliklerinin yönetim, yönetim ve çatışmaların çözümü için çok önemlidir. Herhangi bir sistemdeki kişiler, ofisler veya kuruluşlar arasındaki bağımlılık, özerklik ve iktidar endekslerini, küçük bir grubunkinden uluslar üstü veya küresel düzeyde birini belirleyebilmek için ek analitik çalışmalara ihtiyaç duyulur.

Her ne kadar çatışmanın yayılması karmaşık bir olgu olsa da, rassal değildir. Çatışma, ortak şikayetleri olan sınır ötesi gruplar, silahlara erişim veya paylaşılan kaynaklar gibi komşu ülkeler arasındaki spesifik bağlantı türleri nedeniyle yayılmaktadır. Bu nedenle, çatışmaların başlama sürecini anlamak için, ara bağlantıların anlaşılmasında en uygun olan nicel yöntemlerden ağ analizi tavsiye edilir (yani, grafik teorisinin uygulanması). Sadece son zamanlarda Maoz ve ark. (2007) tarafından yapılan çalışmalar, uluslararası politika ve devletlerarası savaşları incelemek için ağ analizi araçlarını sistematik olarak kullanmaya başlamıştır (Maoz ve ark., 2007). Ağ perspektifinden bir ilk tahmin, yapısal merkezilikte olduğu gibi, bir devletin içeriğinin içeriğinden daha önemli olduğudur. Bununla birlikte, bir

ağ yaklaşımı bunu dışlamaz ve aslında alternatifini test etmek için yeni bir yol sunar: olağandışı durum özellikleri, ideolojiler veya politikalar ağ üzerindeki etkileri bir bütün olarak modellenmektedir (Johnson & Jordan, 2007).

Çok sayıda siyaset bilimci, tarihçi ve politika yapıcı, bölgesel faktörlerin sivil çatışmanın yayılmasındaki önemli etkilerine dikkat çekmiştir (Cederman, 2002). Bununla birlikte, iç savaşın niceliksel çalışması, tek tek devletlerin hangi özelliklerinin onları iç savaşa daha fazla ya da daha az eğilimli hale getirdiğine odaklanmaya yönlendirmiştir. Gleditsch'in (2007), uluslararası faktörlerin komşu ülkelerden etkilenmenin iç savaş olasılığını arttırdığını gösteren ilk nicel araştırmayı yayınlamıştır. Gleditsch'in belirttiği gibi, "iç savaş riski sadece bir ülkenin iç veya yerel özellikleri tarafından belirlenmiyor, bir ülkenin diğer devletlerle olan bağlarına bağlı olarak temelde farklılaşıyor" (Gleditsch 2007: 293). "Ulus ötesi faktörlerin, iç savaşın başlaması riskini önemli derecede etkilediğini ve bu tür gelecekteki araştırmalar için bir basamak taşı olduğunu açıkça ortaya koyuyor" (Gleditsch 2007: 306). Çelik (2019) GDELT veri setini kullanarak yaptığı çalışmada dünyadaki çatışmaların en büyük iki aktörün sırasıyla Amerika ve Rusya olduğunu bulmuştur.

2.3. Sosyal Ağ Analizi (SAA)

Bu çalışmada elde edilen bilgileri değerlendirmek için sosyal ağ analizi yöntemi kullanılmıştır. Sosyal ağ terimi, bireyler, aileler, haneler, köyler, topluluklar, bölgeler vb. arasında atfedilen veya erişilen bir sosyal ilişkiyi ifade etmektedir (Laumann, 1976). Sosyal ağ analizi, sosyal bütünlük, aracılık ve değişim ile sosyal grupların içindeki sosyal sıralamanın analizinde uygun görülmektedir. En basit ifadeyle, Sosyal Ağ Analizi (SAA), düğümün herhangi bir fiziksel dünya varlığı olabileceği veya Facebook gibi sanal dünyada bile bulunabileceği bir ağ şeklinde modellenen düğümlerin analizidir. Örneğin, Facebook arkadaşlarınızın (bir düğüm gibi davranan her bir arkadaşın) bir grafik veya düğüm ağı etkileşimi analiz etmek için yaratılabilir ve böylece çok faydalı bilgiler edinmeye yardımcı olur. Sosyal Ağ Analizi, bu bilgilerin Sosyal Ağlarda nasıl dağıldığını anlamamızı sağlar.

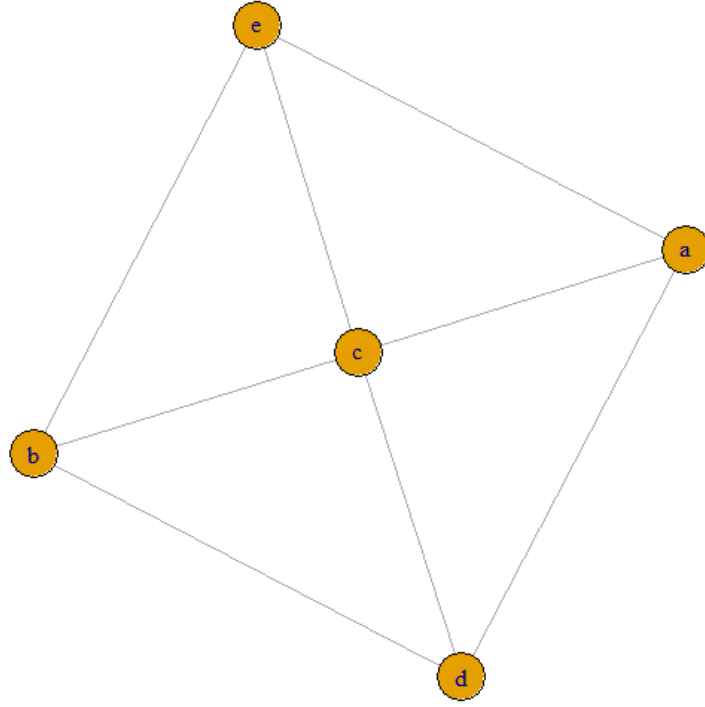
Ağ düşüncesi, sosyal güç ile ilgili bir dizi önemli anlayışa katkıda bulunmuştur. Belki de en önemlisi, ağ yaklaşımı gücün doğal olarak ilişki olarak olduğunu vurgulamasıdır. Bireyin soyutta iktidarı yoktur, iktidarları vardır, çünkü başkalarına hükmedebilirler, ego'nun gücü, değişime bağlıdır. Çünkü iktidar, ilişki kalıplarının bir sonucudur, sosyal yapılarda güç miktarı değişebilir. Bir sistem çok gevşek bağlanmışsa (düşük yoğunluk) çok fazla güç

uygulanamaz; yüksek yoğunluklu sistemlerde daha fazla güç için potansiyel vardır. Güç hem sistemik (makro) hem de ilişkisel (mikro) bir özelliktir. Bir sistemdeki güç miktarı ve aktörler arasındaki dağılımı birbiriyle ilişkilidir, ancak aynı şey değildir. İki sistem aynı miktarda güce sahip olabilir, fakat eşit olarak dağıtılabılır ve bir diğerinde eşit olarak dağıtılamaz. Sosyal ağlarda güç, ya bir mikro özellik olarak (yani, aktörler arasındaki ilişkileri tarif eder) ya da bir makro özellik olarak (yani, tüm evreni tanımlayan) görülebilir; Diğer kilit sosyolojik kavramlarda olduğu gibi, makro ve mikro sosyal ağ düşüncesinde yakından bağlantılıdır (Hanneman & Riddle, 2005).

“Bir resim bin kelime konuşur” en çok kullanılan ifadelerden biridir. Ancak bir grafik bundan çok daha fazla konuşur. Verilerin grafikler halinde görsel olarak gösterilmesi, eyleme geçirilebilir iç görüler edinmemize ve bunlara dayanarak daha iyi veri odaklı kararlar almamıza yardımcı olmaktadır. Ancak hangi grafiğin ne olduğunu ve neden kullanıldığını gerçekten anlamak için, Graf Teorisi (Çizge Kuramı) olarak bilinen bir kavramı anlamamız gerekir. Bu kavramı anlamak bizi daha iyi programcılar ve daha iyi veri bilimciler yapmaktadır. Graf Teorisi kavramları, Sosyal Ağları, Sahtekarlık kalıplarını, Güç tüketim kalıplarını, Viralite ve Sosyal Medyadaki Etkileri incelemek ve modellemek için kullanılır. Sosyal Ağ Analizi (SAA) muhtemelen Veri Bilimi için Grafik Teorisi'nin en iyi bilinen uygulamasıdır. Grafikler, ilişkiler ve etkileşimler gibi soyut kavramlarla baş etmenin daha iyi bir yolunu sağlar. Ayrıca, Sosyal Ağ Analizi bu kavramlar hakkında sezgisel olarak düşünmenin görsel bir yolunu sunmaktadır. Grafikler ayrıca sosyal bağlamda ilişkileri analiz etmek için doğal bir temel oluşturmaktadır. Grafikler, analitik iş akışlarını modellemek için de kullanılır. Bazı Sinir Ağı Yapıları, çeşitli işlemleri farklı katmanlarda modellemek için grafikleri kullanmaktadır (Analytics Vidhya, 2018).

Ağ analizi tüm sosyal durumları düğümler ve bunların bağlantıları, kişiler ve ilişkiler, gruplar ve ilişkiler açısından kavramsallaştırmakla başlar. Bu açıdan bakıldığında, tüm sistemler ağlardır, ancak ağlar farklı özelliklere sahiptir ve bu varyasyon çok önemlidir. Ağ analizi bilgi ve kaynakların etkin akışını sağlamak için pratik yöntemler geliştirmek için kurumsal ve bütünlük kuralları hakkında bilgi verebilir.

En temelde bir ağ, bir tür ilişkilerle veya dernekler ('kenarlar' olarak bilinir) ile birbirine bağlanan 'düğümler' (insanlar, ürünler veya web sayfaları gibi şeyler) topluluğu olarak tanımlanabilir. Aşağıdaki Şekil 2.1 de basit bir ağ yapısı verilmiştir.



Şekil 2.1. Düğümleri ve Kenarları Gösteren Basit Bir Sosyal Ağ

Ağ analizinin temel anlayışı, bir düğümün ağdaki konumunun önemini yansıtmasıdır. Örneğin, bir bağlantı ağının ortasındaki bir web sitesi ve bir insan topluluğunun ortasındaki bir kişi, kenarlardakilerden daha fazla öneme veya etkiye sahiptir. Gerçek dünyadaki bir olguyu bir ağ olarak modellendiği takdirde yeni fikirler ortaya çıkmasını sağlar. Şekil 2.1'deki bulunan daireler ağdaki düğümleri, oklar ise düğümler arasındaki bağlantıları belirtmektedir. Ülkeler arasında yaşanan çatışmaların Sosyal Ağ Analizi ile incelenmesi durumunda ağdaki düğümler ülkeleri çizgiler ise çatışma ilişkilerini ifade edecektir.

Bir sosyal ağ genellikle aktör ya da oyuncuların bir düğüm olarak gösterildiği bir yapıdır. Sosyal ağ analizine başlamadan önce ağda kullanılan verinin bir matrise dönüştürülerek analiz edilmesi gerekmektedir. Bunun için en çok komşuluk matrisi kullanılmaktadır. Örneğin, Tablo 2.1'de sosyal ağa ilişkin oluşturulan komşuluk matrisi verilmiştir.

Tablo 2.1. Komşuluk Matrisi

| | a | b | c | d | e |
|----------|----------|----------|----------|----------|----------|
| a | 0 | 0 | 1 | 1 | 1 |
| b | 0 | 0 | 1 | 1 | 1 |
| c | 1 | 1 | 0 | 1 | 1 |
| d | 1 | 1 | 1 | 0 | 0 |
| e | 1 | 1 | 1 | 0 | 0 |

Bir sosyal ağ analizinde verilerin analizini yapmak için öncelikle verilerin Tablo 2.1'deki gibi bir matris formatında yazılması gerekir.

Çok kenarlı bir graf yapısı için komşuluk matrisi aşağıdaki (1) denklemindeki gibi ifade edilmektedir (Çelik, 2019).

$$M_{i,j} = \begin{cases} a, & (v_i, v_j) \in E \\ 0, & (v_i, v_j) \notin E \end{cases} \quad (1)$$

Denklem (1)'deki a , v_i ve v_j düğümleri arasında kalan kenar sayısını ifade etmektedir.

Şekil 2.1'de görüldüğü gibi bir sosyal ağda bulunan her birim bir düğümle (noktayla), düğümler arasında bulunan bağlantılar çizgiler ile gösterilmektedir.

Şekil 2.1'deki sosyal ağ yapısı graf teorisinde $G = (V, E)$ şeklinde tanımlanmaktadır. V , bir dizi köşe kümesini ve E de kenar kümesini belirtmektedir. Bir e_{ij} kenarı v_i ve v_j köşesi ile birleştirilir. Şekil 1'de bulunan sosyal ağdaki düğüm kümesi $V = \{a, b, c, d, e\}$, bağlantı kümesi ise $E = \{(a, d), (b, e), (c, a), (c, b), (c, d), (c, e), (d, b), (e, a)\}$ şeklindedir. Şekil 2.1'deki bulunan daireler ağdaki düğümleri, oklar ise düğümler arasındaki bağlantıları belirtmektedir.

2.4. Merkezilik Ölçütleri

Merkezilik (Centrality): Ağları analiz etmek için en yaygın kullanılan ve önemli kavramsal araçlardan biridir. Merkeziyet, bir ağdaki en önemli düğümleri bulmayı amaçlar. Bir ağda farklı “önemli” kavramlar olabilir ve bu nedenle birçok merkezilik ölçütü vardır. Merkezilik ölçütlerinin birçok türü vardır. En sık kullanılan merkezilik ölçütlerinden bazıları aşağıdaki gibidir (Analytics Vidhya, 2018):

Derece Merkeziliği (Degree Centrality): İlk ve kavramsal olarak en basit merkezilik tanımıdır. Derece merkeziliği, bir düğüme bağlı kenarların sayısıdır. Yönlendirilmiş bir grafik durumunda, 2 derece merkezi ölçü ölçüsü alınabilir. Bunlar, Giriş ve Çıkış merkezleridir (Analytics Vidhya, 2018).

Arasındalık Merkeziliği (Betweenness Centrality): Bir düğümün, diğer iki düğüm arasındaki en kısa yolda mevcut olma sayısıdır.

Yakınlık Merkeziliği (Closeness Centrality): Bir düğümünden, diğer tüm düğümlere olan en kısa yolun ortalama uzunluğudur. Dahası, ağ üyelerinin “köprü” aracılığıyla bilgiye erişme yeteneğini yansıtır. Bu şekilde yakınlık, her birey ve ağda bulunan diğer herkes arasındaki en kısa mesafenin (bazen jeodezik mesafe olarak adlandırılan) toplamının tersi olarak kabul edilir. N düğüm sayısı olan bir ağ için, yakınlık matematiksel olarak şu şekilde temsil edilir (Ghali ve ark., 2012):

$$C_c(n_j) = \frac{n - 1}{\sum_{k=i, j=k}^n d(n_i, n_j)} \quad (2)$$

Denklem (2)'deki $C_c(n_k, j$ ve $d(n_i, n_j)$ düğümünün standartlaştırılmış yakınlık merkezini tanımladığından, j ve k arasındaki jeodezik mesafeyi belirtir.

Ağ Yoğunluğu (Network Density): Ağ Yoğunluğu, bir grafiğin kaç kenarı olduğuna dair bir ölçüdür. Gerçek tanımı, Grafik türüne ve sorunun sorulduğu içeriğe bağlı olarak değişmektedir. Tam yönlendirilmemiş bir ağ için, yoğunluk 1, boş bir ağ için 0'dır. Ağ yoğunluğu bazı durumlarda 1'den büyük olabilmektedir (Analytics Vidhya, 2018).

3. BÖLÜM

3. BULGULAR

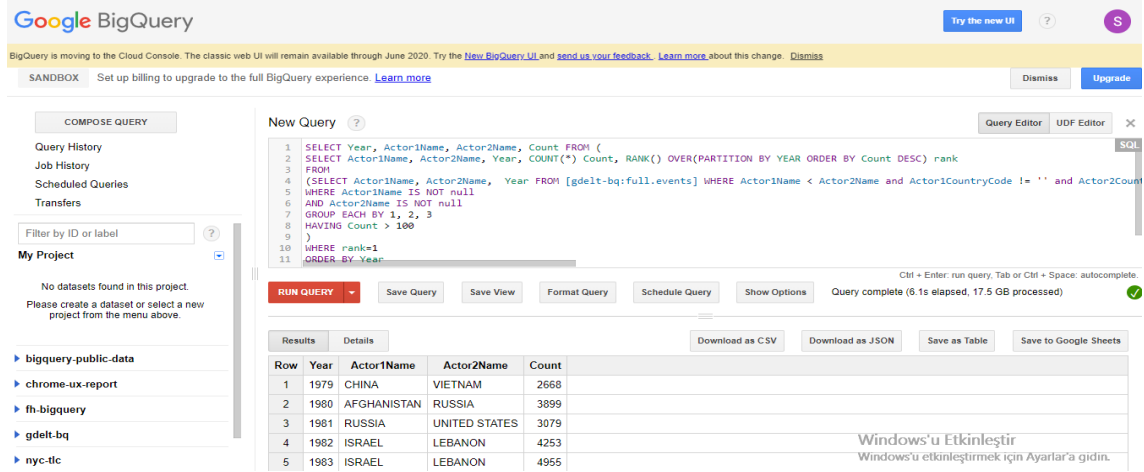
3.1. Ülkeler Arasında Yaşanan Çatışmalara İlişkin Veri Setinin Elde Edilmesi

Araştırmada kullanılan veriler Google CloudBigQuery aracılığıyla www.gdeltproject.org sitesinden çekilmiştir. Çekilen veriler üzerinde veri madenciliğinin veri yönetimi basamağı kapsamında temizleme ve arındırma süreçleri başvurulmuştur. Nihai veri seti 38 ülkeden meydana gelmektedir. Araştırma kapsamında “*Veri setini meydana getiren 38 ülkenin zaman aralığı ve çatışma sayısı göz önünde bulundurularak oluşturulan sosyal ağdaki konumları hakkında ne söylenebilir*” sorusuna cevap aranmıştır. Araştırmada kullanılan veri setini meydana getiren ülkeler ve ülkelerin Türkçe ad karşılıkları Tablo 3.1’de yer almaktadır.

Tablo 3.1. Ülkeler ve Ülkelerin Türkçe Karşılıkları

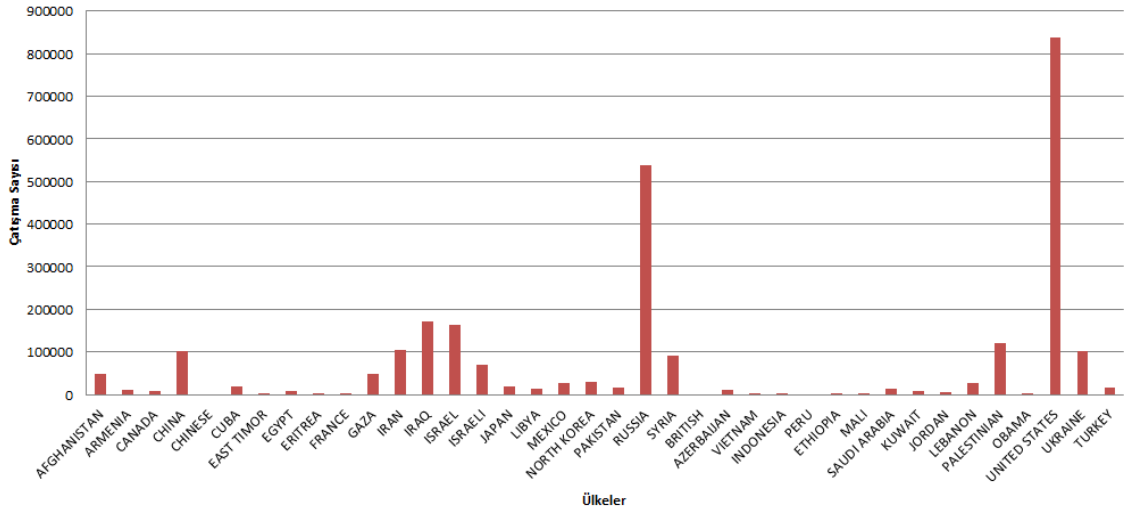
| | | | |
|-------------|------------|---------------|-----------------|
| Afghanistan | Afganistan | Pakistan | Pakistan |
| Armenia | Ermenistan | Russia | Rusya |
| Canada | Kanada | Syria | Suriye |
| China | Çin | British | İngilizler |
| Chinese | Çinliler | Azerbaijan | Azerbaycan |
| Cuba | Küba | Vietnam | Vietnam |
| East Timor | Doğu Timor | Indonesia | Endonezya |
| Egypt | Mısır | Peru | Peru |
| Eritrea | Eritrea | Ethiopia | Etiyopya |
| France | Fransa | Mali | Mali |
| Gaza | Gazze | Saudi Arabia | Suudi Arabistan |
| Iran | İran | Kuwait | Kuveyt |
| Iraq | Irak | Jordan | Ürdün |
| Israel | İsrail | Lebanon | Lübnan |
| Israeli | İsraililer | Palestinian | Filistin |
| Japan | Japonya | Obama | Obama Dönemi |
| Libya | Libya | United States | Amerika |
| Mexico | Meksika | Ukraine | Ukrayna |
| North Korea | Güney Kore | Turkey | Türkiye |

Google BigQuery kullanılarak çekilen verilere ilişkin komut dizileri Şekil 3.1 kapsamında yer almaktadır.



Şekil 3.1. Veri Çekimine İlişkin Komut Dizini

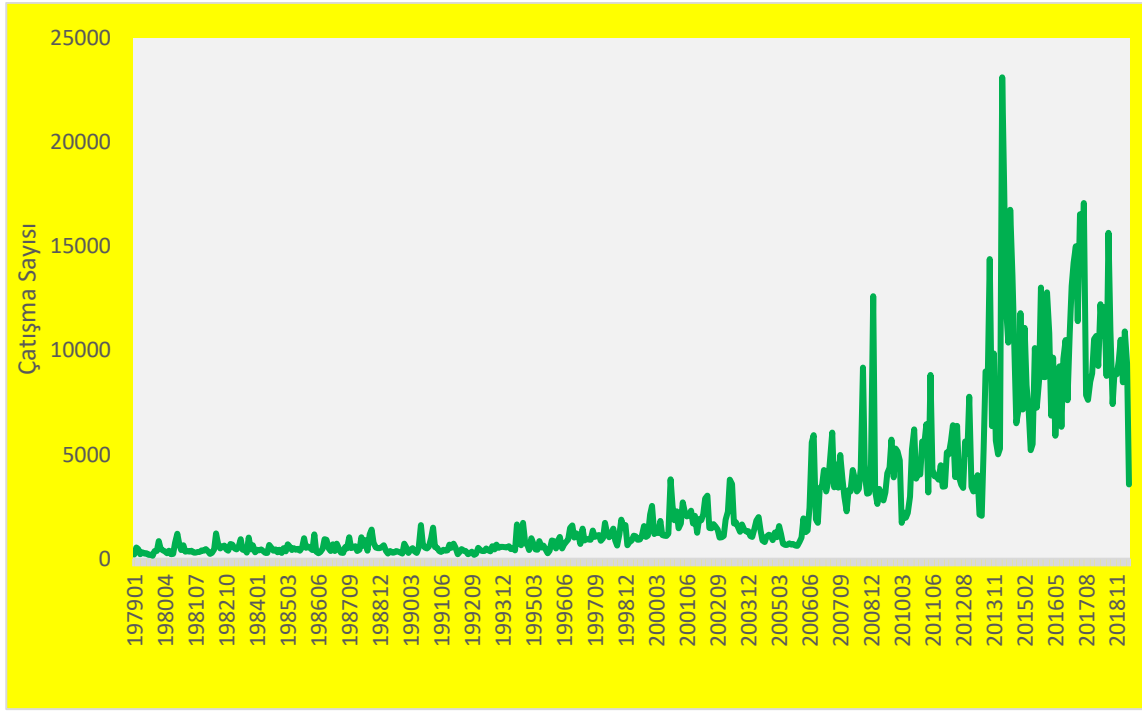
Bu işlemin ardından Şekil 3.2’de Excel programı kullanılarak araştırma kapsamında kullanılan veri setinde yer alan ülkelerin çatışma sayılarına ilişkin bilgi vermektedir. Şekil 3.2 incelendiğinde en fazla çatışma gerçekleştiren 6 aktör ülkenin sırasıyla Amerika (837.919), Rusya (536.938), Irak (171.637), İsrail (162.153), İran (103.477), Filistin (121.621) olduğu göze çarpmaktadır. Söz konusu ülkelerin üçte ikisinin ortadoğu da bulunması dikkat çekicidir. En az çatışma içerisinde bulunan ülkelerin ise sırasıyla Çin, Britanya, Peru, Doğu Timor, Endonezya ve Eritrea olduğu görülmektedir. Bu ülkelerden Çin’in dünyanın en yüksek nüfusuna sahip olmasına rağmen çatışmanın bulunmaması da ilginçtir. Bunun temel nedeni olarak ülkede uygulanmakta olan (baskı veya) aşırı denetleyici rejimin iktidarda hüküm sürmesinin olduğu düşünülebilir.



Şekil 3.2. Ülkelerin Çatışma Sayılarına İlişkin Histogram Grafiği

3.2. Ülkeler Arasında Yaşanan Çatışmaların Sosyal Ağ Analizi ile İncelenmesi

Arındırılmış GDELT veri setinde çatışma olaylarında ismi birlikte geçen ülkeleri analiz etmek için UCINET programından yararlanılmıştır. Şekil 3.3, veri setinde yer alan ülkelerin 1979'dan 2018 yılına kadarki çatışma trendleri hakkında kümülatif bir bilgi sunmaktadır.

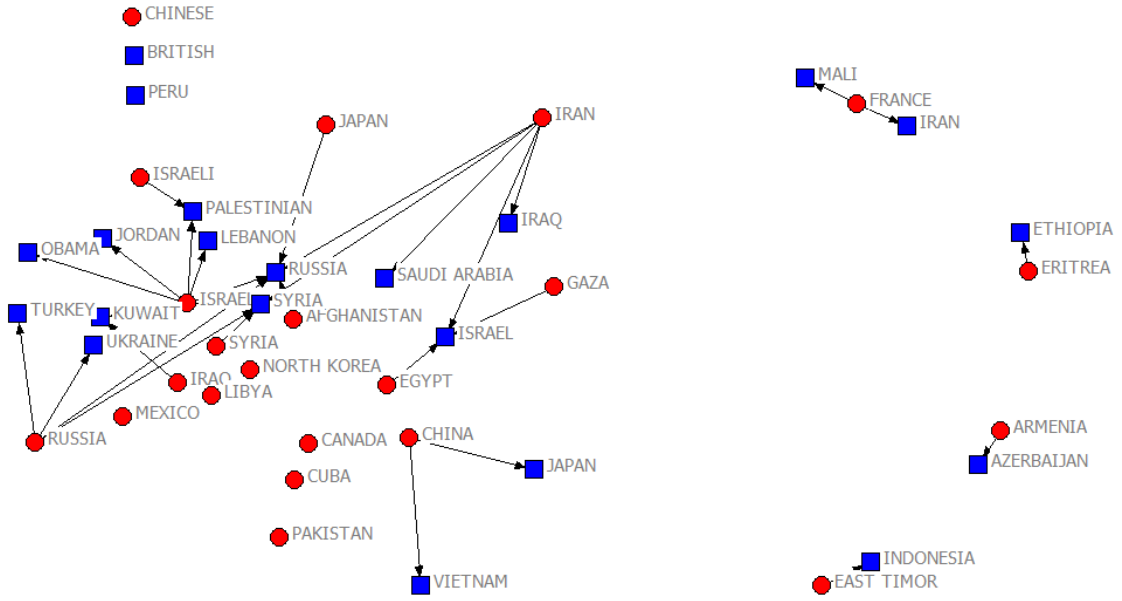


Şekil 3.3. Ülkelerarası Çatışma Çizgi Grafiği

Şekildeki grafik incelendiğinde grafiğin artan bir trende sahip olduğu görülmektedir. Bu da dünyanın büyük bir çatışma ortamına doğru sürüklendiğini göstermektedir. Yani dünya düzenden düzensizliğe veya bir kaos ortamına doğru sürüklenmektedir. Şekil incelendiğinde son yıllarda yaşanan çatışmaların daha yoğun olduğu görülmektedir. Bu çatışma trendinin oluşmasında genellikle Amerika, Rusya, Irak, İsrail, İran ve Filistin gibi aktör rolde olan ülkelerin payının büyük olduğu düşünülmektedir.

3.3. Merkezilik Ölçütleri Bakımından Sosyal Ağ Özellikler

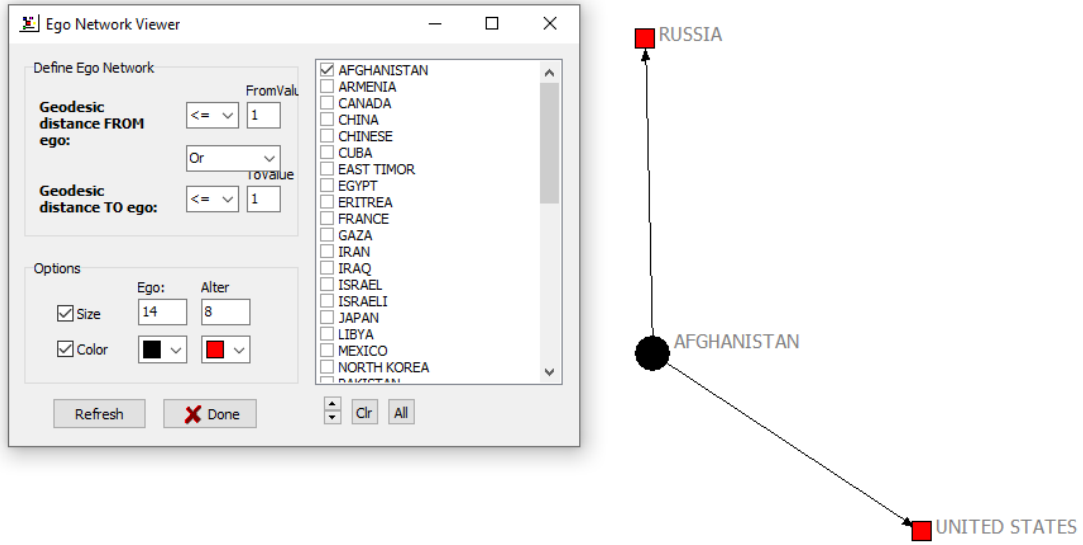
UCINET ve Netdraw programı kullanılarak analize dahil edilecek olan ülkeler satır ve sütun meydana getirecek şekilde matris oluşturulmuş ve UCINET programına bu matris okutularak analizler gerçekleştirilmiştir. Hangi ülke(ler)in hangi ülke(ler)le çatışma içerisinde bulunduğuna ilişkin görsel Şekil 3.4'te yer almaktadır.



Şekil 3.5. Amerika Dışındaki Ükelere İlişkin Sosyal Ağ

Şekil 3.5 incelendiğinde, Şekil 3.4'e göre bir ülkenin bir diğer ülke ile en az bir kere çatışmaya girdiğini simgeleyen ok sayısında gözle görülür bir azalma olduğu gözle çarpılmaktadır.

Amerika ve Rusya'nın çatışmalara ilişkin oluşturulan sosyal ağdaki rolüne daha iyi gözlemlenmek için ağdaki ego(lar)ı temel alarak jeodezik uzaklığın 1 ya da 1'den küçük olduğu durumlara ilişkin görsel elde edilmiştir. Bu görsel Şekil 3.6'da yer almaktadır. Şekil 3.6 incelendiğinde, Afganistan'ın hem Amerika hem de Rusya ile çatışma içerisinde olduğu görülmektedir. Afganistan'ın Amerika ve Rusya ile olan çatışma sayısının diğer ağ düğümlerinden (ülkeleri) fazla olması Afganistan'a da bir ego olma özelliği vermektedir. Kısacası, Afganistan da oluşturulan sosyal ağ içerisindeki en önemli ülkelerden bir tanesidir.



Şekil 3.6. Ego Ağına İlişkin Görsel

Yukarıdaki şekillerle açıklanan durumların sayısal bazı sosyal ağ ölçüleri ile doğrulanması gerekmektedir. Tablo 3.2, oluşturulan sosyal ağa ilişkin ölçüleri içermektedir. Tablo 3.2 çalışmada kullanılan veri setini oluşturan ülkelere ilişkin sosyal ağ ölçülerini içermektedir.

Dünyadaki çatışmalar arasındaki ilişkinin keşfine yönelik olarak sosyal ağ analizinde her bir ülkenin (düğümün) konumunu belirlemek, ülkeler arasındaki ilişkiyi açıklamak ve ağın tümüne ilişkin çıkarımlar yapmak için bazı ölçütlerin hesaplanması gerekmektedir. Yapılan analiz sonucunda elde edilen ülkelerin ve bağlantılarının oluşturduğu ağın merkezilik ölçütlerine ilişkin sayısal veriler Tablo 3.2’de verilmiştir.

Tablo 3.2. Ülkelerin ve Bağlantılarının Oluşturduğu Ağın Merkezilik Ölçütleri

| Aktörler | Derece Merkeziliği | Yakınlık Merkeziliği | Arasındalık Merkeziliği |
|-------------|--------------------|----------------------|-------------------------|
| USA | 19 | 0.0018 | 818.66 |
| Russia | 10 | 0.0018 | 383.33 |
| Israel | 7 | 0.0018 | 140.33 |
| Iran | 7 | 0.0018 | 231.33 |
| Syria | 6 | 0.0018 | 75.50 |
| China | 4 | 0.0017 | 86.83 |
| Iraq | 3 | 0.0018 | 68.00 |
| Afghanistan | 2 | 0.0018 | 0.00 |
| Libya | 2 | 0.0017 | 68.00 |
| Turkey | 2 | 0.0017 | 0.00 |
| Japan | 2 | 0.0017 | 13.33 |
| Egypt | 2 | 0.0017 | 68.00 |
| British | 0 | 0 | 0.00 |
| Lebanon | 2 | 0.0017 | 0.00 |
| Armenia | 1 | 0.0004 | 0.00 |
| Azerbaijan | 1 | 0.0004 | 0.00 |
| Chinese | 0 | 0 | 0.00 |
| Cuba | 1 | 0.0017 | 0.00 |
| East Timor | 1 | 0.0004 | 0.00 |
| Eritrea | 1 | 0.0016 | 35.00 |
| Ethiopia | 0 | 0.0015 | 0.00 |
| France | 2 | 0.0017 | 68.00 |
| Gaza | 1 | 0.0017 | 1.33 |
| Indonesia | 1 | 0.0004 | 0.00 |
| Italy | 2 | 0.0017 | 33.00 |
| Jordan | 1 | 0.0017 | 0.00 |
| Kuwait | 1 | 0.0016 | 0.00 |
| Mexico | 1 | 0.0017 | 0.00 |
| North Korea | 1 | 0.0017 | 0.00 |
| Pakistan | 1 | 0.0017 | 0.00 |
| Peru | 0 | 0 | 0.00 |
| Sudi Arabia | 1 | 0.0017 | 0.00 |
| Ukraine | 1 | 0.0017 | 0.00 |
| Canada | 1 | 0.0017 | 0.00 |
| Vietnam | 1 | 0.0016 | 0.00 |

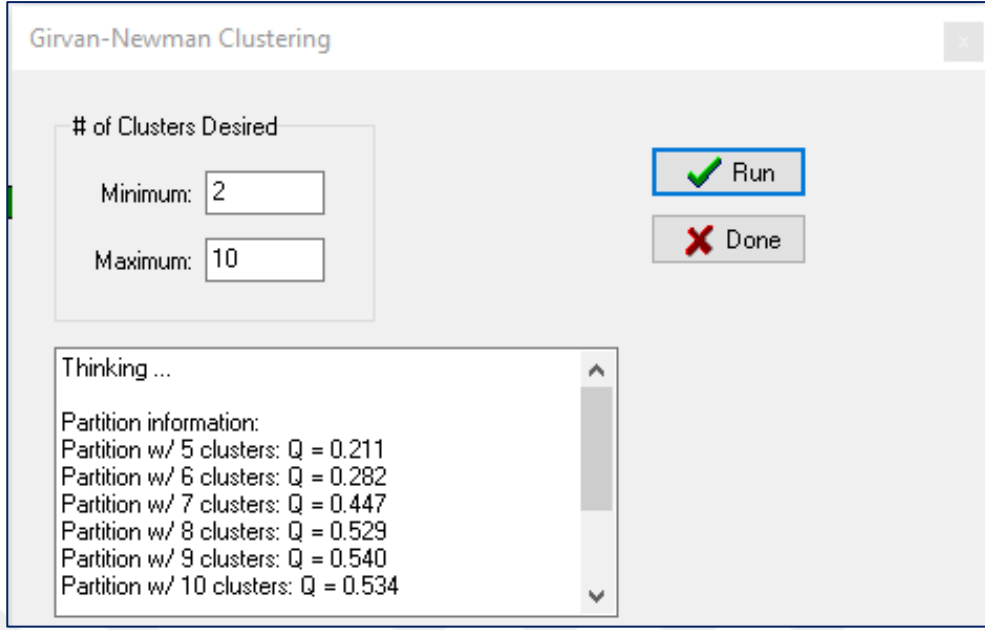
Tablo 3.2'deki analiz sonuçları dikkate alındığında, merkezilik derecesi bakımından en yüksek ülkenin ABD olduğu görülmektedir. Analizler sonucunda, ABD'nin merkezi derecesinin 19 bulunmuştur. ABD'yi ikinci sırada Rusya ve üçüncü sırada İsrail izlemekte olup merkezilik dereceleri sırasıyla 10 ve 7 olarak bulunmuştur. Yine merkezilik yakınlık ve merkezilik arasındalık ölçütlerine bakıldığında en büyük ülkenin ABD olduğu ve onu ikinci sırada Rusya ve üçüncü sırada İsrail'in takip ettiği bulunmuştur. Bu verilere dayanarak ABD'nin diğer ülkelerle en çok bağlantılı (korelasyonlu) olduğu ve dünyadaki çatışmaların oluşmasında en büyük aktör olduğu söylenebilir. Peru, Etiyopya ve Britanya'nın ise görsellerdeki izole yani belirlenen sene aralığında çatışma içerisinde bulunmama durumu Tablo 3.2'deki sonuçlar ile uyusmaktadır. Yine Tablo 3.2'deki sonuçlardan; Rusya, İsrail,

İran ve Suriye gibi ülkelerin de diğer ülkelerle yüksek korelasyonlu oldukları sonucuna varılmıştır. Tüm bu sonuçlar dünyada yaşanan çatışmaların büyük kısmının Ortadoğu ülkelerinde yaşandığını göstermektedir.

Tablo 3.2’de elde edilen sonuçlarda az sayıda ülkenin çok bağlantıya, çok sayıda ülkenin ise az bir bağlantıya sahip olduğu bulunmuştur. Bu sonuçlar 1999 yılında Barbarasi ve Albert’in geliştirdiği “ölçekten bağımsız” ağ modeline işaret etmektedir. Barbarasi ve Albert, 1999 yılında geliştirdikleri “tercihli eklenti” (preferential attachment) modelleri ile ölçekten bağımsız ağların derece dağılımlarının kuvvet yasasına uygun olduğunu bulmuşlardır (Albert ve Barbarasi, 1999).

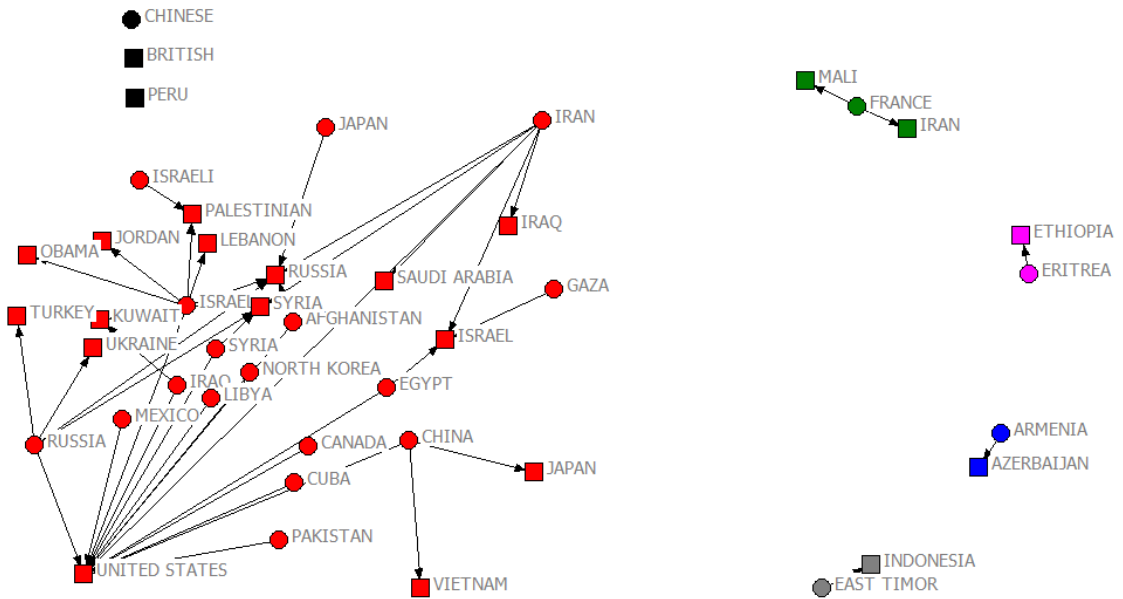
Ülkelerarasında yaşanan çatışmalara ilişkin kritik bağlantıların ortaya çıkartılması anlamında çok büyük olmayan sosyal ağlarda kullanılması önerilen Girvan-Newman algoritması kullanılarak kümeleme analizi gerçekleştirilmiştir. Gerçekleştirilen kümeleme analizinin sayısal çıktısına ilişkin görsel Şekil 3.7’de yer almaktadır.

Girvan-Newman algoritması, bir sosyal ağ içerisindeki alt grupları tanımlamak için tasarlanmış iteratif (yinelemeli) bir süreçtir. Algoritma, ağdaki tüm köşelere ilişkin ağırlık merkezini hesaplar ve ardından en yüksek değere sahip kenar veya kenarları siler. Süreç sonunda zayıf bileşenlerin sayısı artmaktadır. Bu bileşenler birleştirici alt gruplar olarak adlandırılır ve orijinal verilerin bir bölümünü (partition) oluşturur. Bileşenlerin sayısı arttığında, yeni bir bölüm elde edildiğinde, bu bölümler iç içe geçmekte ve bileşenlerin sayısı azami bir kullanıcı tarafından belirlenen bir sayıdan daha azken işlem devam etmektedir. Şekil 3.7 incelendiğinde, Girvan-Newman algoritması kullanılarak gerçekleştirilen kümeleme sürecinin 5 küme ile başladığı ve 10 küme ile sonlandığı görülmektedir. UCINET programı, araştırma kapsamında oluşturulan sosyal ağa ilişkin ideal küme sayısını 6 (altı) olarak belirlemiştir.



Şekil 3.7. Girvan-Newman Algoritmasına Göre Küme Sayıları

Bununla birlikte Şekil 3.8, UCINET programı tarafından Girvan-Newman algoritmasına göre ülkelerin kümelenme durumuna ilişkindir.



Şekil 3.8. Sosyal Ağda Yer Alan Ülkelerin Girvan-Newman Algoritmasına Göre Kümelenme Durumu

Şekil 3.8 incelendiğinde toplam ağın 7 gruptan yani kümeden oluştuğu ve gruplardan en göze batan ülkenin Amerika (çatışma sayısını temsil eden bağlantı sayısının fazla olması) olduğu görülmektedir. Elde edilen bu sonuç, ABD'nin diğer ülkelerle yüksek düzeyde ilişkili (korelasyonlu) olduğunu göstermektedir. Şekil 3.8 tekrardan incelendiğinde, Çinlilerin, Britanya ve Peru'nun belirlenen dönem aralığında herhangi bir çatışma içerisinde

bulunmayarak yani izole konumda olarak ayrı birer küme oluşturduđu; Mali, Fransa ve İran'ın Fransa'nın EGO görevini yani baş aktör görevini üstlenerek ayrı birer küme oluşturduđu; Etiyopya ve Eritrea çatışmasının ayrı birer küme oluşturduđu; Ermenistan ve Azerbaycan çatışmasının ayrı birer küme oluşturduđu; Endonezya ve Dođu Timor çatışmasının da ayrı birer küme oluşturduđu görölmektedir.



4. SONUÇ, TARTIŞMA VE ÖNERİLER

Dünya çapında grupsal ya da internet tabanlı olarak çok fazla çeşitlilik gösteren sistemler, çok farklı ve karmaşık bir topolojiye sahip birbirine bağlı ağlar tarafından tanımlanmaktadır. Yapılan son araştırmalar, bu ağların, basit ama genel yasalarla yönetilen, kendi kendini organize eden süreçlerin birbirine rassal olarak beklenenden çok daha fazla benzer olmasını sağlayan mimari yapılar sonucunda ortaya çıktığını göstermektedir. Bu ağlar temel olarak ilişkilerin yapısı hakkında bilgi vermektedir. Resmi ve gayri resmi ilişkileri analiz yoluyla inceleyerek bilgi, ilişki, çatışma ve savaş durumları gibi bilgilere ulaşılabilir. Bu noktada sosyal ağ analizi devreye girmektedir. Sosyal Ağ Analizi kapsamında karşılıklı bağımlılık önlemleri sağlayan merkezlilik ve yoğunluk gibi belirli istatistiklerin ölçümleri geliştirilmiştir. Böylece, bireysel merkezli bir bakış açısı yerine, tüm ağın incelenmesi (tüm bireylerin etkisi) olasılığı ortaya çıkmaktadır.

Yeryüzünde karmaşık yapıya sahip olan bir sistemi anlamak ilk olarak sistem içerisinde yer alan bileşenlerin birbiriyle olan etkileşimi konusunda bilgi sahibi olmak gerekmektedir. Başka bir deyişle, bağlantı şemasının bir haritasına ihtiyaç duyulmaktadır. Bir ağ, genellikle düğümler veya köşe noktaları olarak adlandırılan sistem bileşenlerinin ve aralarında bağlantılar veya kenarlar olarak adlandırılan doğrudan etkileşimlerin bir kataloğudur. Bu ağ gösterimi, doğada, görünümde veya kapsamda büyük ölçüde farklılık gösterebilecek sistemleri incelemek için ortak bir dil sunmaktadır (Barbara, 2016). Birçok büyük ağın ortak özelliği, köşe bağlantılarının ölçekten bağımsız bir kuvvet yasası dağılımını göstermesidir. Bu özellik, ağların yeni köşe eklenmesiyle sürekli büyüdüğü ve ağdaki yeni köşe noktalarının tercihli olarak bağlantı sayısı fazla olan bağlı düğümlere bağlandığını belirtmektedir. Bu iki özelliğe dayanan bir model, büyük ölçekli ağların geliştirilmesinin, kendi kendini organize eden fenomenler (olaylar) tarafından yönetildiğini belirten ölçekten bağımsız dağılımı (kuvvet yasası dağılımı) yeniden üretmektedir.

Bu çalışmanın temel amacı, sosyal ağ analizini yöntemiyle dünyadaki çatışmaların incelenmesidir. Yapılan analizler sonucunda bugün için dünyadaki çatışmaların merkezindeki en büyük aktörün ABD olduğu bulunmuştur. ABD'den sonra sırasıyla başlıca aktörlerin Rusya, Irak, İsrail, İran, Filistin olduğu göze çarpmaktadır. En az çatışma içerisinde bulunan ülkelerin ise sırasıyla Çin, Britanya, Peru, Doğu Timor, Endonezya ve Eritre olduğu görülmektedir.

Elde edilen bu bulgular iktidar teorisi göz önünde bulundurulduğunda, ABD ve Rusya'nın nispeten eşit bir güç dengesine sahip olduklarından bu iki devletin birbiriyle savaşa (veya askeri çatışmaya) girme olasılığı oldukça düşüktür. En azından bu iki ülkenin birbirlerinin topraklarını işgal ederek savaşmayacakları düşünülebilir. Dolayısıyla bu iki ülke başka ülkelerin toprakları üzerinde yenedünya düzeninde en büyük söz sahibi olan ülke konumunda olma amacıyla birbirlerine güç gösterisi yapabilmektedir. Kısacası, ABD ve Rusya diğer devletlerden daha güçlü olduklarından, dünyadaki konumlarını güçlendirmek için başka topraklar üzerinde savaşa (veya askeri çatışmaya) girme olasılıkları oldukça yüksek olan iki ülkedir. Analiz sonuçlarına göre bu iki ülkeye savaş alanı olan ülkelerden bir tanesinin Afganistan olduğu çok açık bir şekilde görülmektedir. Diğer taraftan iktidar dengesi tezinin güç eşitliği koşulunda zaferi sorunlu hale getirmesi ve dolayısıyla ortaya çıkan belirsizliği arttırmasının saldırganlığı caydırdığı savı göz önünde bulundurularak ABD ve Rusya'nın birbiriyle çatışmaya girmesinin oldukça zor olduğu söylenebilir.

Diğer taraftan iktidar dengesizliği (yani, güç üstünlüğü) açısından değerlendirildiğinde ABD ve Rusya'nın diğer devletlere karşı başarılı olma şansları yüksek olduğundan diğer devletlere karşı daha saldırgan bir tutum sergilemektedirler. Bu durumda zayıf ülkeler ABD ve Rusya ile çatışmaya girmeyi göze almaz ve bu da daha güçlü tarafların hedeflerine ulaşmalarında büyük kolaylık sağlar. Aynı zamanda Amerika birleşik devletleri ve Rusya'nın oluşturulan sosyal ağda yüksek bir merkeziliğe sahip oldukları göz önünde bulundurulduğunda bu iki ülkenin çatışmalarda daha fazla yer aldığı ve bu iki ülkenin çatışmalardaki en etkin iki ülke olduğu söylenebilir.

Son yıllarda Orta Doğu bölgesindeki ülkelerin yaşadığı iç karışıklıklar ve birbirleriyle olan çıkar çatışmaları göz önünde bulundurulduğunda Irak, İsrail, İran ve Filistin'in Amerika ve Rusya'dan sonra çatışma konusunda başlıca aktör oldukları göze çarpmaktadır. Araştırma kapsamında yer alan görseller dikkatlice incelendiğinde özellikle de İsrail'in Kuveyt, Ürdün, Lübnan, Filistin, Rusya ve Suriye ile çatışma içerisinde olduğu dikkat çekmektedir. Ortadoğu devletlerinin demokratikleşmesinin öngörüldüğü proje olan *Büyük Orta Doğu Projesi* kapsamında Türkiye, İran, Irak, Suriye, Suudi Arabistan, İsrail ve Filistin merkezi Orta Doğu hattını oluşturmaktadır. Sosyal ağ analizi kapsamında elde edilen İsrail'in Kuveyt, Ürdün, Lübnan, Filistin, Rusya ve Suriye ile ciddi çatışma içerisinde olduğu ve bu çatışmalardaki baş aktörün İsrail olduğu sonucu aslında bir nevi bu projenin düşündürücü bir proje olduğu şeklinde yorumlanabilir. Bu proje kapsamında dünya siyaseti literatüründe kendisine yer

bulmuş olan kavramlar olan “İlimli İslamiyet”, “İsrail’in güvenliğinin sağlanması” ve “Silah ticareti için yeni Pazar alanlarının ele geçirilmesi” kavramları göz önünde bulundurulduğunda İsrail’in bu ülkelerle olan çatışmaları ve bölgedeki karışıklıklarda baş aktörlerden bir tanesi olması manidar bir şekilde yorumlanabilir.

Ortadoğu coğrafyası, tarihin her döneminde savaşların ve çatışmaların gerçekleştiği yeryüzünün en talihsiz bölgelerindedir. Özellikle de Ortadoğu’da çok zengin enerji kaynaklarının yer alması bölgenin önemini maksimum düzeye çıkarmaktadır. Başat Ortadoğu ülkeleri arasındaki çatışmalar araştırma kapsamında oluşturulan sosyal ağ analiziyle de geçerlik kazanmıştır. Amerika Birleşik Devletleri ve Rusya’nın sosyal ağ analizi sonucuna göre en önemli aktörler olduğu göz önünde bulundurulduğunda ABD ve Rusya gibi küresel aktörlerin, zaten çatışmaya meyilli bir ortamda hızlandırıcı bir etkileri olduğundan bahsedilebilir.

Tüm bu sonuçlar elde edilen ağ yapısının ölçekten bağımsız olduğu ve ağın derece dağılımının kuvvet yasasına uygun olduğunu göstermektedir. Son olarak araştırma kapsamında elde edilen bulgular göz önünde bulundurulduğunda toplam ağın 6 kümeden (gruptan oluştuğu) ve gruplardan en büyüğünün merkezinde sırasıyla başat aktörler olarak Amerika ve Rusya’nın yer aldığı bulunmuştur. Çinlilerin, Britanya ve Peru’nun belirlenen dönem aralığında herhangi bir çatışma içerisinde bulunmayarak yani izole konumda olarak ayrı birer küme oluşturduğu; Mali, Fransa ve İran’ın Fransa’nın EGO görevini yani baş aktör görevini üstlenerek ayrı birer küme oluşturduğu; Etiyopya ve Eritre çatışmasının ayrı birer küme oluşturduğu; Ermenistan ve Azerbaycan çatışmasının ayrı birer küme oluşturduğu; Endonezya ve Doğu Timor çatışmasının da ayrı birer küme oluşturduğu da araştırmanın diğer sonuçları olarak ortaya çıkmaktadır. Araştırma sonucunda elde edilen sosyal ağda Fransa’nın Mali ile ayrı bir küme altında çatışma halinde görüntülenmesi Batı Afrika’da Fransa’nın gerçekleştirdiği sömürgecilik çalışmalarıyla ilgili olan Wooten (2002), Bannelier, Karine ve Christakis (2013) ve Guitta (2010)’nın çalışma sonuçlarıyla paralellik göstermektedir. Aynı zamanda çalışma sonucunda Etiyopya ve Eritre arasındaki çatışmanın sosyal ağ üzerinde ayrı bir küme olarak görüntülenmesi Etiyopya ve Eritre arasındaki sınır krizini konu eden Lyons (2006), Bezabih (2014)’in araştırma sonuçları ile paralellik göstermektedir. Çalışma sonucunda Ermenistan ve Azerbaycan’ın arasındaki çatışmanın sosyal ağ üzerinde ayrı bir küme olarak görüntülenmesi ise Gül ve Ekici (2002), Huntington (2005), Çimen ve Gögebakan (2009)’ın çalışma sonuçlarıyla paralellik göstermektedir.

4.1. Öneriler

Sembolik sermayenin bölgeselleşmesi ve sosyal ağ analizi temelli bu araştırma sonucuyla da paralellik gösteren Ortadoğu'nun şiddetli çatışmalarını ele almak için daha etkili olabilecek analiz türlerine başvurulabilir ve bu tür araştırmalar sonucunda küresel bazda refaha ulaşılabilecek bazı politikalar önerilebilir. Bu politikalar aracılığıyla bölgedeki ihtilafların sebepleri daha somut ve sayısal bir hale getirilebilir. Benzeri araştırmalar gerçekleştirilerek karmaşık olmayan kimlik politikaları kavramı da gözden geçirilip bütünleştirilebilir.

Bu ve benzeri araştırmalar yapılarak özellikle de Ortadoğu ülkelerinin birbirleri arasındaki hesaplaşmaların ana sebepleri ve bu hesaplaşmalarda Amerika ve Rusya'nın rolü daha iyi anlaşılabilir. Benzeri araştırmalar gerçekleştirilerek çatışma ile ilgili mallarda ve personelde sınır ötesi trafikle ilgili önerilen analiz seviyeleri, daha net bir şekilde bölgenin çeşitli çatışmalarının nasıl birbirine bağlandığını kavramsallaştırma ve çatışmalar üzerinde Amerika ve Rusya'nın rolü ve asıl gayeleri daha net ortaya koyulabilir.

Her ne kadar ilgili çatışmaların sebepleri genel olarak farklı zenginlik kaynaklarına ulaşma, sınır ötesi hareket, popüler efsaneler ve komplolar ve politik aktörlerin bunları sembolik sermayeye çevirme kabiliyetleri olsa da arka planda bu sebepleri tetikleyen unsurları inceleme ve belirleme amaçlı çok daha fazla ülkeden meydana gelen ve sadece çatışma sayılarını ve tarihleri değil aynı zamanda farklı değişkenleri de temel alan çok farklı araştırmalar gerçekleştirilebilir. Gerçekleştirilebilecek bu ve benzeri araştırmalar sonucunda elde edilen bulgular ışığında çatışma içerisinde olan ülkelere farklı ekonomik ve stratejik politikalar önerilebilir.

Araştırma sonuçlarına göre her ne kadar Türkiye oluşturulan sosyal ağda diğer Ortadoğu ülkeleri kadar başat aktör rolüne bürünmüş gibi görünmese de ülkelerin diğer akademik disiplinlere ilişkin bazı özelliklerini içeren değişkenler benzer konulu gelecek araştırmalara dahil edilerek Türkiye'nin hem küresel hem de genel anlamda Ortadoğu temelinde yaşanan bu çatışmalardaki asıl konumu daha açık bir şekilde belirlenebilir.

5. KAYNAKLAR

- Aksu, G., Eser, M. T., ve Güzeller, C. O. (2019). *R Programlama Dili ile Temel İstatistikler ve Raporlama*. Ankara: Pegem Akademi.
- Aktan, E. (2018). Büyük veri: Uygulama alanları, analitiği ve güvenlik boyutu. *Bilgi Yönetimi*, 1(1), 1-22.
- AnalyticsVidhya (2019). An Introduction to Graph Theory and Network Analysis (with Python codes), 10.08.2019 tarihinde <https://www.analyticsvidhya.com/blog/2018/04/introduction-to-graph-theory-network-analysis-python-codes/> adresinden alınmıştır.
- Azure (2017). What is Cloud Computing, 04.11.2019 tarihinde <https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/> adresinden alınmıştır.
- Bannelier, K. ve Christakis, T. (2013). Under Security Council's Watchful Eyes: Military Intervention by Invitation in the Malian Conflict. *Leiden Journal of International Law*, (26), ss. 855-874.
- Becominghuman (2018). A Simple Introduction to Natural Language Processing, 02.11.2019 tarihinde <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32> adresinden alınmıştır.
- Berkeley (2019). What is Data Science, 03.11.2019 tarihinde <https://datascience.berkeley.edu/about/what-is-data-science/> adresinden alınmıştır.
- Bezabih, W. F. (2014). Fundamental Consequences of the Ethio Eritrean War [1998-2000]. *Journal of Conflictology*, Vol: 5, No: 2, November, 2014, s. 39-47.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science*, 323(5916), 892-895
- Canbay, Y., Vural, Y., Sağiroğlu, Ş. (2017). "Büyük Veri Mahremiyetinde kullanılan Kavramlar ve Yöntemler", *Büyük Veri ve Açık Veri Analitiği: Yöntemler ve Uygulamalar*, Grafiker Yayınları.
- Cederman, L.-E. 2002. Endogenizing Geopolitical Boundaries with Agent-Based Modeling. *Proceedings of the National Academy of Sciences* 99 (Supplement 3):7296-7303.
- Cimen, A., Gögebakan, G. (2009). Tarihi değiştiren savaşlar. İstanbul, Timaş yayınları, 432 s.
- Cloudplatform (2017). Google BigQuery, 02.11.2019 tarihinde <https://cloudplatform.googleblog.com/2014/05/worlds-largest-event-dataset-now-publicly-available-in-google-bigquery.html> adresinden alınmıştır.
- Computerworld (2019). The GDELT Project, 02.11.2019 tarihinde <https://www.computerworld.com/article/2977530/the-gdelt-project>. Adresinden alınmıştır

- Craondata (2018). Top Big Data Technologies, 02.11.2019 tarihinde <https://www.crayondata.com/blog/top-big-data-technologies-used-store-analyse-data/> adresinden alınmıştır.
- Çelik, S. (2019). Dünyadaki Çatışmaların Sosyal Ağ Analizi Yöntemiyle İncelenmesi. *Öneri Dergisi*, 14(52), 236-254.
- Davenport, T. H. and Harris, J. G. (2007). The Dark Side of Customer Analytics, *Harvard Business Review*, R0705A, 1-9.
- De Mauro A, Greco M, Grimaldi M (2015) What is big data? A consensual definition and a review of key research topics. *In: AIP conference proceedings*, 1644, 97–104.
- Dellemc (2019). Big Data, 01.11.2019 tarihinde <https://www.dellemc.com/en-us/big-data/definitions.htm> adresinden alınmıştır.
- Edureka (2019). What is Data Science, 04.11.2019 tarihinde <https://www.edureka.co/blog/what-is-data-science/> adresinden alınmıştır.
- Ersin, N. (2003). Ortadoğu Savaşları'nın Perde Arkası, Gündem Yayınları, İstanbul.
- Ersöz, S. ve Ersöz, O. Ö. (2015). İşletmelerde Bilgi Sistemleri Uygulamalı Örneklerle. (1. Baskı). Ankara: Nobel Akademik Yayıncılık.
- GDELT Project (2019). Data, araştırmada kullanılan veri seti <https://www.gdelproject.org/data.html#googlebigquery> adresinden Google BigQuery vasıtasıyla alınmıştır.
- Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Socialnetworksanalysis: Tools, measuresandvisualization. *In Computationalsocialnetworks* (pp. 3-23). Springer, London p. 7.
- Gleditsch, K. S. (2007). TransnationalDimensions of CivilWar. *Journal of PeaceResearch* 44 (3):293-309.
- Google Cloud Platform (2018). GDELT 2.0 Event Database, 9.10.2019 tarihinde <https://console.cloud.google.com/marketplace/details/the-gdelt-project/gdelt-2-events> adresinden alınmıştır.
- Google Cloud Platform Blog ((2014). World's largest event dataset now publicly available in BigQuery, 02.11.2019 tarihinde <https://cloudplatform.googleblog.com/2014/05/worlds-largest-event-dataset-now-publicly-available-in-google-bigquery.html> adresinden ulaşılmıştır.
- Gökçen, H. (2011). Yönetim Bilgi / Bilişim Sistemleri: Analiz ve Tasarım. (2. Baskı). Ankara: Afşar Matbaacılık
- Greycampus (2019). Big Data Developer, 04.11.2019 tarihinde <https://www.greycampus.com/opencampus/big-data-developer> adresinden alınmıştır.

- Gupta, S. (2018). Sentiment Analysis: Concept, Analysis and Applications, 14.06. 2019 tarihinde <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> adresinden alınmıştır.
- Guru99 (2018). What is Big Data, 02.11.2019 tarihinde <https://www.guru99.com/what-is-big-data.html> adresinden alınmıştır.
- Gül, N., Ekici, G. (2002). Ortak Tehdit Algılamaları ve Stratejik İttifaklığa Doğru İlerleyen İran-Ermenistan İlişkileri. *Stratejik Analiz*, 2(22): 37- 43.
- Hacısalıhoğlu, Y. (2004), BOP; Avrupa, Rusya, Çin ve Hindistan'ın Yaşam Alanını Daraltıyor, ABD'nin Kalıcı Egemenlik Arayışı, Cumhuriyet Strateji.
- Hackernoon (2018). What is Data Visualization, 02.11.2019 tarihinde <https://hackernoon.com/what-is-data-visualization-definition-history-and-examples-e51ded6e444a> adresinden alınmıştır.
- Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods. University of California, Riverside. CA (*Online book*).
- Hubspot (2017). How to Do AB Testing, <https://blog.hubspot.com/marketing/how-to-do-a-b-testing>
- Huntington, S.P. (2005). Medeniyetler Çatışması ve Dünya Düzeninin Yeniden Kurulması. İstanbul, Okuyan Us yayınları, 536 s.
- IBM (2018). Cloud Computing, 01.11.2019 tarihinde <https://www.ibm.com/cloud/learn/cloud-computing> adresinden alınmıştır.
- Ihaka, R. and R. Gentleman (1996). R: A languagefordataanalysisandgraphics. *Journal of ComputationalandGraphicalStatistics*, 5, 299–314.
- Innoarchitech (2019). What is Data ScienceandWhatdoes Data Scientist Do?, 04.11.2019 tarihinde <https://www.innoarchitech.com/what-is-data-science-does-data-scientist-do/> adresinden alınmıştır.
- Investopedia (2017). Datamining, 02.11.2019 tarihinde <https://www.investopedia.com/terms/d/datamining.asp> adresinden alınmıştır.
- Johnson, D.,& Jordan, F. (2007). The Web of War: A Network Analysis of the Spread of CivilWars in Africa. In *Annual Meeting of thePoliticalScienceAssociation* (Vol. 28, No. 02.09, p. 2007) p.2-3.
- Kdnuggets (2018). GDELT Big Data News, 02.11.2019 tarihinde <https://www.kdnuggets.com/2014/09/gdelt-big-data-news-conflict.html> adresinden alınmıştır.
- Laumann, E., ve Pappi, F. (1976). Networks of Collective Action: A Perspective on Community Influence Systems. New York: Academic Press.

- Leetaru, K. ve Hoffa, F. (2015). Analyzing the world' news: Exploring the GDELT Project through Google BigQuery, 25.11.2019 tarihinde https://www.oreilly.com/ideas/analyzing-the-worlds_news_exploring_the_gdelt_project_through_google_bigquery adresinden alınmıştır.
- Lexalytics (2019). Sentiment Analysis Explained, 13.07.2019 tarihinde <https://www.lexalytics.com/technology/sentiment-analysis> adresinden alınmıştır.
- Lyons, T. (2006). Avoiding Conflict in the Horn of Africa U.S. Policy Toward Ethiopia and Eritrea. The Centre for Preventive Action, Vol: 21, s. 14-3
- M. Grover, T. Malaska, J. Seidman, and G. Shapira. (2015). Hadoop Application Architectures, USA: O'Reilly Media Inc.
- Madhivanan, Ramachandran (2013), Top big data technologies used to store and analyse data
- Mailchimp (2017). AB Tests, 02.11.2019 tarihinde <https://mailchimp.com/marketing-glossary/ab-tests/> adresinden alınmıştır.
- Maoz, Z., Lesley, G., Terris, R., D. Kuperman, I. (2007). What Is the Enemy of My Enemy? Causes and Consequences of Imbalanced International Relations, 1816–2001. Journal of Politics 69 (1):100–115.
- Marr, B. (2019). What Is Big Data? 12.11.2019, <https://www.bernardmarr.com/default.asp?contentID=766> adresinden alınmıştır.
- McGuire, T., Manyika, J., ve Chui, M. (2012). Why Big Data is the New Competitive Advantage. Ivey Business Journal Online.
- Medium (2015). Introduction to Natural Language Processing, 01.11.2019 tarihinde <https://medium.com/m/global-identity?redirectUrl=https%3A%2F%2Fbecominghuman.ai%2Fsimple-introduction-to-natural-language-processing-ea66a1747b32> adresinden alınmıştır.
- Medium (2019). Big Data Visualization Techniques, 01.11.2019 tarihinde <https://medium.com/sciforce/best-data-and-big-data-visualization-techniques-e07b897751dd> adresinden alınmıştır.
- Microsoft (2019). 02.11.2019 tarihinde <https://docs.microsoft.com/en-us/sql/odbc/reference/machine-data-sources?view=sql-server-2017> adresinden alınmıştır.
- Microsoft (2019). Data Sources, 02.11.2019 tarihinde <https://docs.microsoft.com/en-us/sql/odbc/reference/data-sources?view=sql-server-2017> adresinden alınmıştır.
- Microsoft (2019). File Data Sources, 02.11.2019 tarihinde <https://docs.microsoft.com/en-us/sql/odbc/reference/file-data-sources?view=sql-server-2017> adresinden alınmıştır.
- Microsoft (2019). Types of Data Sources, 02.11.2019 tarihinde <https://docs.microsoft.com/en-us/sql/odbc/reference/types-of-data-sources?view=sql-server-2017> adresinden alınmıştır.

- Microsoft Azure (2019). What is cloud computing? 20.08.2019 tarihinde <https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/> adresinden alınmıştır.
- MongoDB (2019). Big Data Explained, 04.11.2019 tarihinde <https://www.mongodb.com/big-data-explained> adresinden alınmıştır.
- Olivier, G. (2010). Al-Qaeda in the Islamic Maghreb: A Threat for the West. Defence Against Terrorism Review, Vol.3(1), s.67
- Optimizely (2016). AB Testing, 01.11.2019 tarihinde <https://www.optimizely.com/optimization-glossary/ab-testing/> adresinden alınmıştır.
- Oracle (2019). The Definition of Big Data, 15.10.2019 <https://www.oracle.com/big-data/guide/what-is-big-data.html> adresinden alınmıştır.
- Oreilly (2019). GDELT Project through Google BigQuery, 02.11.2019 tarihinde <https://www.oreilly.com/ideas/analyzing-the-worlds-news-exploring-the-gdelt-project-through-google-bigquery> adresinden alınmıştır.
- Oussous, A., Zahra, F., Lahcen, A., Belfkih, S. (2018). Big Data Technologies: A Survey, 01.11.2019 tarihinde <https://www.sciencedirect.com/science/article/pii/S1319157817300034> adresinden alınmıştır.
- Parlar, S. (2002). Ortadoğu Vaadedilmiş Topraklar, İstanbul: Yar Yayınları, 2. Baskı, Mayıs 2002, s. 31.
- Pratt, M. K. (2015). The GDELT Project, 20.10. 2019 tarihinde <https://www.computerworld.com/article/2977530/the-gdelt-project>. adresinden alınmıştır.
- Ranger, S. (2018). What is cloud computing, in simple terms?, 25.02.2019 tarihinde <https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-from-public-and-private-cloud-to-software-as-a/> adresinden alınmıştır.
- Researchoptimus (2018). What is Time Series Analysis, 02.11.2019 tarihinde <https://www.researchoptimus.com/article/what-is-time-series-analysis.php> adresinden alınmıştır.
- Researchoptimus (2019). Time Series Analysis and Forecast Service for Your Business, 23.02.2019 tarihinde <https://www.researchoptimus.com/article/what-is-time-series-analysis.php> adresinden alınmıştır.
- Rouse, M. (2016). SQL (Structured Query Language), 15.08.2019 tarihinde <https://searchsqlserver.techtarget.com/definition/SQL> adresinden alınmıştır.
- Rouse, M. (2017). NoSQL (Not Only Structured Query Language), 12.08.2019 tarihinde <https://searchdatamanagement.techtarget.com/definition/NoSQL-Not-Only-SQL> adresinden alınmıştır.

- Sagi, D. J. B., & Labeaga, J. M. (2016). Using GDELT Data to Evaluate the Confidence on the Spanish Government Energy Policy. *IJIMAI*, 3(6), 38-43.
- Salesforce (2019). What is cloud computing?, 14.10.2019 tarihinde <https://www.salesforce.com/products/platform/best-practices/cloud-computing/#> adresinden alınmıştır.
- SAS (2016). Data Mining, 01.11.2019 tarihinde https://www.sas.com/en_us/insights/analytics/data-mining.html adresinden alınmıştır.
- Satman, M. H. (2017). *R ile Programlama*. İstanbul: Türkmen Kitabevi
- Sdxcentral (2016). What is Cloud, 01.11.2019 tarihinde <https://www.sdxcentral.com/cloud/definitions/what-is-cloud/> adresinden alınmıştır.
- Searchdatamanagement (2018). Not Only SQL, 03.11.2019 tarihinde <https://searchdatamanagement.techtarget.com/definition/NoSQL-Not-Only-SQL> adresinden alınmıştır.
- Searchsqlserver (2019). SQL, 01.11.2019 tarihinde <https://searchsqlserver.techtarget.com/definition/SQL> adresinden alınmıştır.
- Sharma, P. (2019). What Is Data Science? A Beginner's Guide To Data Science, 15.07.2019 <https://www.edureka.co/blog/what-is-data-science/>
- Simon, P. (2014). *The visual organization: data visualization, big data, and the quest for better decisions*. Wiley-Blackwell.
- Stephen R. Wooten (2002). Mal, *Encyclopedia of Twentieth-Century African History*, Ed. Paul Tiyambe Zeleza, Routledge, London, s. 345
- Şahin, A. (2004). *Büyük Orta Doğu Projesi ve Türkiye*. Truva Yayınları, İstanbul.
- Şeker Sadi Evren, *Büyük Veri ve Büyük Veri Yaşam Döngüleri*, YBS Ansiklopedi , v. 2, is. 3, Eylül, 2015)
- Talend (2019). Future of Big Data, 03.11.2019 tarihinde <https://www.talend.com/resources/future-big-data/> adresinden alınmıştır.
- Techopedia (2016). Data Source, 02.11.2019 tarihinde <https://www.techopedia.com/definition/30323/data-source> adresinden alınmıştır.
- The GDELT Project (2019) Google BigQuery, 04.12.2019 tarihinde <https://www.gdeltproject.org/data.html#googlebigquery> adresinden alınmıştır.
- Towardsdatascience (2017). Sentiment Analysis Concept and Applications, 02.1.2019 tarihinde <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> adresinden alınmıştır.
- Trcharget (2019). Data Science, 04.11.2019 tarihinde <https://searchenterpriseai.techtarget.com/definition/data-science> adresinden alınmıştır.

- Vasquez, John A., ed. 2000. *What Do We Know About War?* Lanham, MD: Rowan&Littlefield.
- Verma, A. (2018). Google BigQuery and Its Role in Big Data Analytics 12.10.2019 tarihinde <https://www.whizlabs.com/blog/google-bigquery/> adresinden alınmıştır.
- Watson, H. J. (2014). Tutorial: Big data anal',lies: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34 (1). 1247–1268.
- Weng, W.-H.,& Lin, W.-T. (2014). Development Trends and Strategy Planning in Big Data Industry. *Contemporary Management Research*, 10(3).
- Whizlabs (2017). Google BigQuery, 02.11.2019 tarihinde <https://www.whizlabs.com/blog/google-bigquery/> adresinden alınmıştır.
- Wisconsin (2019). What is Big Data, 03.11.2019 tarihinde <https://datasciencedegree.wisconsin.edu/data-science/what-is-big-data/> adresinden alınmıştır.
- Yadav, R.,& Kumar, T. (2015). Usage of Big Data Analytics for Customer Relationship Management. *International Journal of Advanced Research in Computer Science*, 6(2).
- Yıldız, Y.G., (2004). *Oyun İçinde Oyun: Büyük Ortadoğu*. IQ Kültür Sanat Yayıncılık, 3. Baskı, Sayfa: 22.
- Zdnet (2018). What is Cloud Computing, 01.11.2019 tarihinde <https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-from-public-and-private-cloud-to-software-as-a/> adresinden alınmıştır.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Zühal Mollaoğulları

Doğum Yeri ve Tarihi : Bor \ 20.04.1987

Eğitim Durumu

Lisans Öğrenimi : Muğla Sıtkı Koçman Üniversitesi

Lisansüstü Öğrenimi : Aydın Adnan Menderes Üniversitesi

Bildiği Yabancı Diller : İngilizce

İletişim

e- posta Adresi : zuhal.mollaogullari@adu.edu.tr

Tarih :20.01.2020