

TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

AÇIK KAYNAK İSTİHBARATI: TWITTER ANALİZİ

Salih Zeki ATALI

DİSİPLİNLERARASI ADLİ BİLİMLER ANABİLİM DALI
YÜKSEK LİSANS TEZİ

DANIŞMAN
Dr. Öğr. Üyesi Bülent TUĞRUL

ANKARA
2020

ETİK BEYAN

Ankara Üniversitesi
Adli Bilimler Enstitüsü Müdürlüğü'ne,

Yüksek Lisans tezi olarak hazırlayıp sunduğum “Açık Kaynak İstihbaratı: Twitter Analizi” başlıklı tez; bilimsel ahlak ve değerlere uygun olarak tarafımdan yazılmıştır. Tezimin fikir/hipotezi tümüyle tez danışmanım ve bana aittir. Tezde yer alan deneysel çalışma/araştırma tarafımdan yapılmış olup, tüm cümleler, yorumlar bana aittir.

Yukarıda belirtilen hususların doğruluğunu beyan ederim.

Öğrencinin Adı Soyadı: Salih Zeki ATALI

Tarih:

İmza:

İÇİNDEKİLER

Etik Beyan	ii
İçindekiler	iii
Önsöz	iv
Simgeler ve Kısaltmalar	v
Şekiller	vi
Çizelgeler	vii
1. GİRİŞ	1
1.1 Açık Kaynak İstihbaratı	2
1.1.1 Açık Kaynak İstihbaratında Bilgi Toplama Yöntemleri	3
1.1.2. Açık Kaynak İstihbaratının Faydaları ve Zorlukları	4
1.1.3. Açık Kaynak İstihbaratının Önemi	5
1.2. Twitter	8
1.2.1 Twitter'ın Önemi	9
1.3. Veri Seti, Büyük Veri ve Veri Madenciliği	9
1.4. Semantik ve Elasticsearch	12
2. GEREÇ VE YÖNTEM	14
2.1. Keyhole ve Zemberek	14
2.2. Mentionmapp	16
2.3. Elasticsearch	17
2.4. RapidMiner ve Tinfoleak	18
3. BULGULAR	22
4. TARTIŞMA	42
5. SONUÇ VE ÖNERİLER	44
ÖZET	46
SUMMARY	47
KAYNAKLAR	48
EKLER	51
EK 1 : Anket Formu	51
ÖZGEÇMİŞ	54

ÖNSÖZ

Açık Kaynak İstihbaratı, insanların veya sistemlerin sunduğu verileri elde ederek açık kaynaklardan gelen verilerin bilgiye dönüştürülmesini sağlamaktadır. İnsanların sosyal medyalar ve internet kullanımı konusunda alışkanlıkları arttıkça, açığa çıkan veri boyutu da o ölçüde artış göstermektedir.

İnsanların duygularını, düşüncelerini, bilgilerini ve daha birçok veriyi belirli platformlar üzerinde toplayarak bunların analiz edilmesi, analiz sonrasında da işe yarayacak bilgiye dönüştürülme işleminin tümü açık kaynak istihbaratı olarak adlandırılmaktadır.

Twitter kullanımı son 10 yılda özellikle ciddi artış göstermekte, insanlar duygu ve düşüncelerini rahatlıkla paylaşmaktadır. Herhangi bir siyasi konuyla, toplumsal olaylarla alakalı duygu düşüncelerini, kendileriyle alakalı verileri burada paylaşmaktadır. Bu platformda tüm veriler doğru ve geçerli olmamakla birlikte, toplandıktan sonra insanlar veya yazılımlar aracılığıyla temizlenip duygu analizi gibi yöntemler kullanılarak faydalı bilgilere dönüştürülmektedir.

Bu çalışmanın yapılması konusunda bana desteklerini esirgemeyen değerli danışman hocam Dr. Öğr. Üyesi Bülent Tuğrul'a, her daim bana desteklerini esirgemeyen arkadaşlarım Bilgi Güvenliği Uzmanı İsmail ŞEN'e, Açık Kaynak Araştırmacısı İsmail AKSU'ya ve motivasyon konusunda desteklerini hiç esirgemeyen arkadaşım Abdullah Murat Özkan'a teşekkür ederim.

SİMGELER VE KISALTMALAR

ABD	Amerika Birleşik Devletleri
BM	Birleşmiş Milletler
OSINT	Açık Kaynak İstihbaratı (Open Source Intelligence)
RSS	Zengin Site Özeti (Rich Site Summary)



ŞEKİLLER

Şekil 2.1. Keyhole'de Twitter'dan Veri Alımı	14
Şekil 2.2. Zemberek Normalizasyonu	15
Şekil 2.3. Mentionmapp ile Sahte Kullanıcı Ayrımı	16
Şekil 2.4. Elasticsearch ile Benzer Tweetlerin Belirlenmesi	17
Şekil 2.5. Elasticsearch ile Kullanıcı Adı Ayrıştırması	18
Şekil 2.6. Tinfoleak ile Kullanıcı Analizi	19
Şekil 2.7. RapidMiner ile Duygu Analizi	21



ÇİZELGELER

Çizelge 3.1. Cinsiyet	22
Çizelge 3.2. Yaşanılan Coğrafi Bölge	23
Çizelge 3.3. Yaş	23
Çizelge 3.4. Eğitim Durumu	24
Çizelge 3.5. Aile/arkadaşlar ile iletişim ve/veya tanışma için kullanma durumu	24
Çizelge 3.6. Fotoğraf/video paylaşımı ve/veya eğlenmek için kullanma durumu	25
Çizelge 3.7. Gündem takibi ve/veya fikir belirtmek için kullanma durumu	25
Çizelge 3.8. Ticaret için kullanma durumu	26
Çizelge 3.9. Sosyal medyada geçirilen zaman (1 gün içinde)	26
Çizelge 3.10. Paylaşımlarda konum belirtme durumu	27
Çizelge 3.11. Verilerin 3. kişiler satılma/paylaşılma olasılığını bilme durumu	27
Çizelge 3.12. "10yearschallenge" aktivitesini görme/katılma durumu	28
Çizelge 3.13. Demografik bilgi paylaşma durumu	28
Çizelge 3.14. Özel Hayat bilgisi paylaşma durumu	29
Çizelge 3.15. Ekonomik bilgi paylaşma durumu	29
Çizelge 3.16. Hiç bilgi paylaşmama durumu	29
Çizelge 3.17. Sosyal medya aracılığıyla tanışma durumu	30
Çizelge 3.18. Dolandırılma durumu	30
Çizelge 3.19. Gerçek isim/fotoğraf kullanma durumu	31
Çizelge 3.20. Popüler konulara # ile katılma durumu	31
Çizelge 3.21. Paylaşılan verilerin kendine karşı kullanımından endişe duyma durumu	31
Çizelge 3.22. Sosyal ağlar güvenli/güvensiz seçimi	32
Çizelge 3.23. Paylaşımları Kimler Görebiliyor - Dolandırılma Durumu	32
Çizelge 3.24. Popüler Konulara Katılım - Dolandırılma Durumu	33
Çizelge 3.25. 10yearschallenge - Dolandırılma Durumu	33
Çizelge 3.26. Gerçek İsim/Fotoğraf Kullanma - Dolandırılma Durumu	34

Çizelge 3.27. Sosyal Medyada Tanışma - Dolandırılma Durumu	34
Çizelge 3.28. Kişisel Veri Paylaşma - Dolandırılma Durumu	35
Çizelge 3.29. Eğitim Düzeyi - Dolandırılma Durumu	35
Çizelge 3.30. Sosyal Medya Kullanma Süresi - Dolandırılma Durumu	36
Çizelge 3.31. Cinsiyet - Dolandırılma Durumu	36
Çizelge 3.32. Yaş - Dolandırılma Durumu	37
Çizelge 3.33. Sosyal Medya Kullanma Nedeni - Dolandırılma Durumu	38
Çizelge 3.34. Coğrafi Bölge - Dolandırılma Durumu	39
Çizelge 3.35. Huawei - Subjektiflik Analizi	40
Çizelge 3.36. iPhone - Subjektiflik Analizi	40
Çizelge 3.37. Samsung - Subjektiflik Analizi	41

1. GİRİŞ

Açık kaynak istihbaratı, sosyal medyanın tüm dünyada aktif olarak kullanılmaya başlanmasıyla birlikte çok daha kolay bir hâl almıştır. İnsanların düşüncelerine, duygularına, dini bilgilerine, çalıştığı iş yerine ulaşmak sosyal medya yokken açık kaynaklar üzerinden erişimi zor olan bilgilerdi. Sosyal medyanın kullanımı ile birlikte, kullanıcılar veri paylaşmasa bile açık kaynaklarda onlara ait izleri bulmak mümkün hâle gelmiştir.

Twitter ile birlikte, insanlar burayı duygularını, düşüncelerini ve hayatlarından kesitleri paylaştıkları platform olarak kullanmaya başladılar. Paylaşılan veriler tek başına bir işe yaramasa bile toplanıp veri seti hâline getirilerek analiz edildiğinde önemli bilgiye dönüşmektedir.

Bu çalışmada açık kaynakta paylaşılan verilerin, kullanıcılar hakkında olumsuz sonuçlar doğurabileceği, doğacak sonuçlardan nasıl etkilenebilecekleri gibi durumlar incelenecek, bu analiz sonuçları yorumlanacaktır. Bu kullanıcıların attığı tweetlerin doğruluğu, geçerliliği test edilip doğru bir veri seti elde edildikten sonra veri setleri analiz edilecek, analiz sonrasında elde edilen veriler ve sonuçlar yorumlanacaktır.

1.1. Açık Kaynak İstihbaratı

Açık kaynak istihbaratı (OSINT) herkesin ulaşabildiği kaynaklardan bilgi bulma, seçme, elde etmeyi ve elde edilen bilgiyi analiz etmeyi içeren istihbarat toplama disiplini. Açık kaynak bilgisi ve istihbaratının belirgin dört kategorisi vardır. (Steele, 2007, s. 131) Bunlar açık kaynak verisi, açık kaynak bilgisi, açık kaynak istihbaratı ve doğrulanmış açık kaynak istihbaratıdır.

- Açık Kaynak Verisi; Bir patent, telif hakkı ve başka kontrol mekanizması kısıtlaması olmadan herkesin ulaşabileceği veridir. Fotoğraf, ses kaydı, uydu görüntüsü veya şahsi mektup açık kaynak verisi olabilir.
- Açık Kaynak Bilgisi; Genellikle geniş bir kitleye yayılmış genel bilgidir. Gazeteler, kitaplar, yayınlar ve günlük haberler açık kaynak bilgi dünyasının bir parçasıdır. (Benavides, 2010, s. 7)
- Açık Kaynak İstihbaratı; Belirli bir istihbarat ihtiyacının ele alınması amacıyla zamanında uygun kitleden toplanan, faydalanılan ve yayılan halka açık bilgi üzerinden yapılır. (Benes, 2013, s. 23)
- Doğrulanmış Açık Kaynak İstihbaratı; Gizli istihbarat kaynaklarına erişebilen bir istihbarat uzmanı tarafından yüksek oranda doğru olduğu belirlenen veya medyanın yayınladığı canlı bir video gibi doğruluğu hakkında endişe doğurmayan güvenilir bir açık kaynaktan gelen bilgidir.

OSINT, hedefle ilgili her şeye ulaşmayı sağlamasa bile yapan kişiye perspektif kazandırır. OSINT'te kullanılan kaynaklar bilgi akışına göre altı kategoriye ayrılabilir:

- Medya; Gazeteler, dergiler, radyo, televizyon.
- İnternet; Çevrim içi yayınlar, internet günlükleri, tartışma grupları, sosyal medya.
- Kamu verisi; Halka açık bildirimler, bütçeler, komite oturumları, telefon rehberleri, basın toplantıları, internet siteleri, konuşmalar.

- Mesleki ve akademik yayımlar; Dergilerden elde edilen bilgiler, konferanslar, sempozyumlar, akademik bildiriler, tezler.
- Ticari veri; Ticari görüntüler, mali ve endüstriyel değerlendirmeler, veri tabanları.
- Bilimsel çalışmalar; Teknik raporlar, ön baskılar, patentler, ön makaleler, iş dokümanları, yayınlanmamış çalışmalar ve bültenler.

1.1.1. Açık Kaynak İstihbaratında Bilgi Toplama Yöntemleri

Bilgi toplamanın ana amacı herkesin ulaşabildiği kaynaklar aracılığıyla hedefle ilgili bilgi edinmektir. OSINT'te bilgi toplama yöntemi, ilgilenilen veri türüne ve toplama sürecinde kullanılacak senaryoya bağlı olarak değişkenlik gösterir. OSINT kaynakları üç ana metot kullanılarak toplanabilir: (Hassan, 2018, s. 14)

- Pasif Toplama; OSINT'te en yaygın olarak kullanılan bilgi toplama yöntemidir. Hedefle direkt olarak temasa geçmeden yapılır. Örnek olarak gazete/dergi okumak, televizyon izlemek, sosyal medya aracılığıyla bilgi toplamak verilebilir.
- Yarı Pasif Toplama; Hedef sunuculara genel bilgi toplamak için sınırlı trafik gönderilir. Gönderilen trafik keşif eylemlerinin dikkat çekmemesi için tipik internet trafiğine benzer. Bu tip bilgi toplamada hedefin çevrim içi kaynaklarının derin araştırması yapılamaz, sadece hedef tarafında herhangi bir uyarı oluşturmadan inceleme yapılabilir.
- Aktif Toplama; Hedefle direkt olarak temasa geçilerek yapılır. Bilgi toplayan kişi/tüzel kişi açık portlara erişim, zafiyet taraması gibi hedef bilgi sistemi altyapısı hakkında teknik veriler toplamak için ileri seviye teknikler kullanacağından dolayı hedef, keşif sürecinin farkına varabilir. Hedefe sosyal mühendislik saldırıları yapmak da aktif bilgi toplama yöntemidir.

OSINT'te, herkese açık kaynaklardaki bilgiyi gizlice toplayabilmek için pasif toplama tercih edilir.

1.1.2. Açık Kaynak İstihbaratının Faydaları ve Zorlukları

Günümüzde OSINT farklı istihbarat alanlarında hayati bir role sahiptir. İstihbarat toplamak için herkesin ulaştığı bilgiyi kullanmak, casusluk yapan uydulara veya insan kaynaklarını kullanmaya kıyasla risk içermez. OSINT'in diğer istihbarat edinme yöntemlerine göre maliyeti daha azdır (Gibson, 2004, s. 20). OSINT'in kaynaklarına her zaman ulaşılabilir. Bilgiler düzgünce toplanıp analiz edilirse, istihbarat bağlamında farklı yerlerde kullanılabilir. İstihbarat adına bilimsel çalışmalar haricinde kullanılacak bilgiler için telif hakkı tarzı kısıtlamalarla karşılaşılmaz. Resmî daireler, çeşitli kuruluşların/kişilerin sosyal medya hesaplarını, tatilde ne yaptıklarını ve hayat tarzlarını gözlemleyerek vergi kaçaklıklarını saptayabilirler. OSINT, sahte ürün/hizmet satımının da önüne geçer. OSINT, hükûmetlerin insanların tavırlarını anlayıp buna göre önlem almasını sağlar. Bu, OSINT'in en önemli faydalarındandır.

Tüm istihbarat toplama yöntemlerinin bazı kısıtlamaları vardır. Her ne kadar OSINT ile elde edilen veriyi analiz edecek otomatik araçlar olsa da toplanan büyük miktardaki veri, zaman ve analiz açısından zorluk çıkarır. OSINT'te istihbarat bağlamında kullanılacak bilginin doğruluğunun araştırılması gerekir. Çoğu hükûmet açık kaynak bilgisi/verisi toplama sürecini yanlış yönlendirmek için doğru olmayan bilgi/veri paylaşır. (Hassan, 2018, s. 15-17)

1.1.3. Açık Kaynak İstihbaratının Önemi

İnternet siteleri, internet günlükleri, forumlar, dergiler ve sosyal medya; hükûmete, güvenlik operasyonlarına, hukuki yaptırımlara ve kişinin dünyayı daha iyi anlamasına yardımcı olup tehditleri ve riskleri belirlemeleri konusunda analistlere destek olan değerli bilgi kaynaklarıdır. Bu kaynaklardaki bilgi gizli değildir ve çoğu kuruluşun ihtiyacı olan bilginin %80'ini karşılar.

OSINT, hazırlıklı olma ve korunma açısından insancıl eylemleri kökten değiştirir. OSINT, kriz anlarında bilginin paylaşılabilir olduğu sistemlerin kurulmasını mümkün kılar. Bu, olası tehditleri ve olanakları belirleyerek krize hazırlıklı olma parametrelerini sağlamak açısından önemlidir. Hazırlıklı olma konusunda bilgiye kolay erişim başlı başına kayda değer bir değişikliktir.

Yardım sürecinin can alıcı noktası, etkilenen kitle ihtiyaçlarının ve insani kriz faktörünün yol açtığı zararın boyutunu belirleme durumudur. Bu bağlamda, sadece ihtiyaçları değil aynı zamanda değişen ihtiyaç modellerini ve bunların yardım gereksinimleri üzerindeki etkilerini belirleyebilen artan sayıda platform vardır.

BM Mülteciler Yüksek Komiserliği 2009'dan beri mültecilerle diyalog içinde olmasını sağlayan sosyal medya ağını kullanmaktadır. Hollanda hükûmeti bir felaket anında insanlara bilgi ulaştırmak için irtibat noktası olarak özel bir internet sitesi kullanmaktadır -<http://www.crisis.nl>-.

2010 yılında gerçekleşen Haiti depreminin ardından, çeşitli sektörlerin ve sivil toplumun bilgi akışını baştan sona geliştirmesine yardımcı olan yazılım ve servisleri sağlayan Ushahidi platformu, kitlelerden bilgileri toplayıp yardımcı olunması için paylaşır. (Ushahidi) Benzer şekilde SensePlace2, değişen ihtiyaçları ve yardım süreçlerini izlemeye yardımcı olan bilgiyi seçmek için bir içeriğin topluca izlenmesine

olanak sađlayan ierik besleme yntemi olan RSS, haberler ve internet gnlkleri dhil olmak zere birok metin kaynađını bir araya getirir. (Chan, 2014, s. 11) Aık kaynak bilgisi, yardımların ulaşıp ulaşımadıđını ve ulaştıysa zamanını belirlemek iin kullanılabilir. Bu tarz kitle kaynaklı bilgi, alışıl gelmiş takip etme yntemlerine nazaran yapılan alıřmaların durumunu daha hızlı řekilde grmeyi sađlayabilir.

Kurumlar kaynaklarını arttırmak, varlıklarını srdrmek ve yardım giriřimleri iin nemli lde enerji tahsis ederler. Burada OSINT ve bunu sađlayan teknolojilerin, yardım sađlayanlar ve alanlar arasındaki iliřkiyi deđiřtirme konusunda kayda deđer bir potansiyeli vardır. OSINT sayesinde kitlesel fon sađlayanlar, yardım etmek iin fon sađlamadan sorumlu bir sivil toplum kuruluřuyla grřmeye gerek duymaksızın bir kriz blgesinden gelen kitleye dayalı bařvurudan haberdar olabilirler. Bu dođrultuda bir giriřim, bařta aileler olmak zere yardıma muhta olanlara mobil ađlar zerinden kaynak sađlayarak dođrudan yardım roln stlenen diaspora topluluklarla bařlamıřtır. 2010 yılındaki deprem sonrası kolera salgınına maruz kalan Haiti, alternatif yardım fonu kaynaklarının ve OSINT'in nemini gstermiřtir. Yardım edecek kiřiler/kurumlar yardımlarını ABD'de yařayan Haiti diasporası aracılıđıyla gnderdiler.

İnternetin nemi insani ihtiyalar iin kaynak seferberliđi bađlamında byk lde kabul grmektedir. 1998'de Kuzey Dakota'nın Grand Forks ilesindeki sel felaketi ve 1999'daki Floyd Kasırgası'ndan 2001 yılında ABD'deki terr saldırısı ve 2003 yılındaki Isabel Kasırgası'nın sebep olduđu yıkıma kadar afetlerin yol atıđı gereksinimleri karřılamak iin fonları arttırmak adına internet kullanımında srekli olarak bir artıř olmuřtur. İnternet kullanımının artmasıyla insanların ulařabilecekleri aık kaynak bilgisi de artmıřtır. Kaynak seferberliđinin bu řekli Katrina ve Rita Kasırgaları'nın ardından devam etti. Pew Arařtırma Merkezi'ne gre 13 milyon Amerikan vatandařı internet zerinden bađıř yapmıřtır. (Pew Research Center)

İnsani bağlamda sorumluluk sadece yardımlar sonrası oluşan etkilerle ilgili değildir. Bu bağlamdaki sorumluluk, krizlere hazırlıklı olmak veya krizleri önlemek için sorumluluk/rol alabilmiş olanları ve almış olması gerekenleri mesul tutmakla da ilgilidir. Tehditler konusunda çok iyi bilgi sahibi olup onları savuşturmayı amaçlayanlar sık sık tehditlere açık olanlardır ve genelde uzmanlar onlarla yakın temas kurmazlar. Genel olarak profesyonel aktörler, krizden etkilenen ve krize maruz kalan kişilere daha az yakından ilişki kurma eğilimindedirler. Uzmanlar kriz zamanlarında ne yapılması gerektiğine karar vermekle kalmayıp kriz sonrasında nelere ihtiyaç duyulacağına karar verirken, bu kitleye pasif izleyiciler gibi davranılır. Nispeten kısa bir zaman dilimi içinde projeleri gerçekleştirmekle yükümlü kuruluşların, 'müşterilerin' değişen ihtiyaçlarını, profillerini ve kendilerinin mal ile hizmetleri gerçekleştirmedeki yeterliliklerini gözlemleyip üzerinde düşünmek için çok az zamanları vardır. (White, 2008, s. 4)

OSINT'in önemi günümüzde giderek daha fazla hissedilmeye başlamıştır. Çevrim içi forumlar ve sosyal medya dâhil açık kaynak teknolojilerinden faydalanan kuruluşlar, yardıma muhtaç olanları; insani müdahaleleri bir şekilde sokma, planlama, izleme ve değerlendirme sürecinde bulunma, taleplerini ve isteklerini dile getirme yönünde teşvik ederek kalkınma ve insani yardımlar konusunda destekliyor. Bu bağlamda sorumlulukların artması artık bir seçenek değil zorunluluk olmuştur. Bir felaketten etkilenen kitleler seslerini medya aracılığıyla duyurabildiklerinden dolayı, ajanslar sorumluluklarının artık bir seçenek olmaktan çıktığının farkındalar. Haitililer 2010 yılında yardımların yeterli olmadığına dair seslerini Twitter üzerinden duyurdular. Bir grup Haitili BM'ye kolera salgınını körüklediği yönünde dava açmıştır. Yardımcı olan felaketedeler ve diaspora topluluklar direkt olarak yardım tepkilerini etkilerler. Endonezya'daki Merapi Yanardağı'nın patlaması sonrasında olanları dünyaya duyurmak için Facebook ve Twitter kullanılmıştır. Temel bulgular, BM İnsani İşler Koordinasyon Ofisi'ne sosyal medya aracılığıyla gönderilmiştir. Somali diasporası, Danimarka Mülteci Konseyi tarafından interaktif harita üzerinden internette yayımlanan gıda yardımı fon ihtiyaçlarını doğrudan karşılamıştır. Haiti diasporasındaki yüzlerce kişi; yollar, kafeler, tren istasyonları vb. konularda bilgi

sağlayan harita kullanan kişiler tarafından oluşturulan OpenStreetMap adındaki bir çevrim içi aracı kullanarak aşırı zarar gören bölgelerin haritasını çıkarmak konusunda birlikte hareket etmişlerdir. (OpenStreetMap) OSINT, insani krizlerin sonrasının yanı sıra başlangıcındaki sorumlulukların da dinamiklerini değiştirir.

1.2. Twitter

Twitter, dünyada olup bitenlerin aktarıldığı ve insanların görüşlerini yansıttıkları yerdir. (Twitter) Twitter'daki karakter sınırlaması; kullanıcıları yazılarını yenilemeleri yönünde teşvik eder. Gün geçtikçe artan mobil cihaz kullanımıyla birlikte çoğu kişinin Twitter ile meşgul olmasından dolayı Twitter ağına gönderilen bir hayli tweet vardır. Kullanıcıların çok miktardaki veri yüzünden bunalmaları, bu popülerliğin olumsuz tarafı olmuştur. Twitter bu sorunla mücadele etmek adına kullanıcıların tweetlerine ekleyebildikleri etiket (hashtag) kültürünü getirmiştir. Etiket, önünde # işaretinin olduğu boşluk içermeyen ve tweetin herhangi bir yerinde bulunabilen, belli bir gelişme ya da konuyla ilgili yazı veya sözcük öbeğidir. (Otsuka, 2014, s. 330) Etiket, insanların katkıda buldukları fikirleri konu olarak bir başlık altında toplamakla kalmayıp akımların başlamasına yol açabilir. (Van den Berg, 2014, s. 5) Kullanıcılar ve şirketler reklam, felaket durumlarında bilgilendirme ve yardım alma, farkındalık yaratmak gibi amaçlarla gündem olmalarını istedikleri konularda Twitter'daki etiketten faydalanırlar.

Twitter, kullanıcı mesajlarının hızlı ve kısa süreli tabiatı nedeniyle kullanıcıların dikkatlerinin yanı sıra ilgilerinin ve düşüncelerinin gerçek zamanlı yansımaları sağlar. Sonuç olarak Twitter, milyonlarca kullanıcısının yoğun ilgisini keşfetmek için kendisinden çıkarılabilecek 'trendlerin' yansıttıklarıyla zengin bir kaynak işlevi görür. (Naaman, 2011, s. 905) İnternet üzerinde etiketler aracılığıyla hangi konuların ilişkili olacağının bir ifadesi olarak oluşan trendlerin ortaya çıkışını ve kullanımını dikkate

almak gerekir. Bir veya daha fazla terim ve zaman periyodu içeren trend konusu, belli bir zaman periyodunda gönderilen, içerisinde terimler barındıran mesajların boyutu beklenen hareket seviyesini aştığında oluşur. Trend konuları etiketler aracılığıyla hızlı bir şekilde milyonlarca kullanıcıya ulaşabilir.

1.2.1. Twitter'ın Önemi

Etiketlerin kullanımı, mesajları yeniden tweetlemek ve kullanıcıları takip etmek daha büyük bilgi ağı oluşmasını sağlar. Şirketler çeşitli reklam kampanyalarında ticari mesajın can alıcı kısmını belirtmek, insanlar olağanüstü durumlarda birbirini bilgilendirmek, sosyal mesajları duyurmak için Twitter'dan faydalanırlar. Güney Afrika'daki bir cep telefonu şirketi olan Cell C, 2013'te #CellCBelieve etiketiyle çarpıcı bir reklam kampanyası başlatmıştır. (Cell C) Engelli atlet Oscar Pistorius, model olan sevgilisi Reeva Steenkamp'ı öldürme suçundan yargılanırken, #OscarPistorius etiketiyle tüm dünya gelişmeleri takip edebilmiştir. Arap Baharı'ndaki sivil protestolar sırasında #Bahrian en çok kullanılan etiketler arasına girmiştir. 2007 Ekim ayında olan orman yangınları sırasında #sandiegofire etiketinde ilgili bilgiler paylaşılmıştır. (Van den Berg, 2014, s. 3-4)

1.3. Veri Seti, Büyük Veri ve Veri Madenciliği

Veri seti, veri tabanı tablosu içeriklerine veya istatistiksel veri matrisine karşılık gelen bir veri yığıdır. Değişkenlerin veya niteliklerin sayısı, türü ve standart sapma gibi bunlara uygulanabilen çeşitli istatistiksel ölçüler, bir veri setinin yapısını ve özelliklerini tanımlar. Veri setleri ayrıca birtakım yazılım testleri amacıyla algoritmalar tarafından oluşturulabilir.

Büyük veri, bilgi işlem uygulama yazılımlarının ele alamayacakları kadar büyük veya karmaşık veri setlerini ifade etmek için kullanılan bir terimdir. Veri toplamak, depolamak, incelemek, araştırmak, paylaşmak, iletmek, görselleştirmek, sorgulamak, güncellemek, bilgi gizliliği ve veri kaynakları, büyük verinin zorluklarındandır.

Büyük veri herhangi bir şekilde oluşturulabilir. Dijital ağlarda neredeyse kaçınılmaz olan dijital izler, düşük maliyetli alternatif veri kaynaklarını bulmak için fazlasıyla fırsatlar yaratmıştır. Veri kaynağı (arama terimleri) dijital bir yan üründür ama gerçek zamanlı veriye sahip hastalık önleme ve kontrol yetki makamlarının resmî istatistik formlarının yerine geçecek potansiyele sahiptir. Bu, aynı zamanda çoğu büyük veri kaynağının alışlagelmiş veri kaynaklarının çoğundan farklı olan bir spesifik araştırma sorusunun sonucu olarak oluşturulmadığını gösterir. Böyle bir büyük veriye genelde olaydan sonra ihtiyaç duyulur, öncesinde yorumlanması gerekmez.

Büyük veri rastgele örnekleme yerini almıştır. Büyük veri gerçek dünyada yaşananların dijital izleri olarak genellikle var olanı yakalar. Dünya çapında %95'in üzerinde kullanımıyla evrensel veri kaynağı hâline gelen cep telefonlarının kayıtları, sosyoekonomik, demografik ve diğer davranışsal işlerde sonuç çıkarmak için kullanılabilir. Tahmin doğruluğu değişkenlerin kombinasyonlarına bağlıdır ama çağrı süresi veya sıklığı gibi veri kayıtlarını kullanmak genellikle %80-85 civarında doğru tahmin yürütülmesini sağlayacaktır. Büyük veriye genellikle gerçek zamanlı olarak erişilebilir. Büyük verinin yaygın gerçek zamanlı kaynaklarından biri, çevrim içi sosyal medyada sürekli olarak bulunan konuşmacı olmuştur. Bu kaynak özellikle gelişmekte olan ülkelerde önem arz eder. Twitter'ın mikro günlüklerinin dil içeriği; Malezya, Filipinler, Venezuela ve Endonezya gibi ülkeler dâhil kültürel kimlikleri yakınlaştırmak, turizm hareketliliği ve uluslararası göç için kullanılmıştır.

Büyük veri, farklı kaynakları bir araya getirir. Dijital davranışların ardından kalan genelde karmaşık ve eksik dijital izler farklı kaynaklardaki veri fazlalığı ile telafi

edilebilir. Büyük veri farklı duygusal durumların yanı sıra fikirleri ve spesifik konuları da değerlendirir. Büyük veri, büyük veri analizi olarak görülebilir. Büyük veri kavramı, gitgide veri miktarının ve niteliğinin ötesine geçip akıllı kararlar alınabilmesi için analize odaklanmaktadır. Bu süreç “Haberleşme Çağı” ve “Haberleşme Toplulukları”ndan “Bilgi Toplulukları”na evrilirken atılacak doğal adım olarak görülebilir. Büyük veri paradigması, dijital haberleşmeyi akıllı kararlar aldırın bilgiye dönüştürmeye odaklanır. Google grip trendinde, grip salgınlarını alışlagelmiş modellerden daha iyi öngörebilecek 45 arama terimini belirlemek için 450 milyon farklı matematiksel model uygulamıştır.

Büyük veri analizi, veri miktarı analitik modelin seçimini etkilediğinden dolayı alışlagelmiş istatistiksel analizden farklıdır. Algoritmaların verilerden bir şeyler öğrenmesini sağlayan makine öğrenimi ve veri madenciliği büyük miktarda veriye uygulanabilir. Araştırmaya dayalı veri madenciliği ve makine öğrenimi yöntemleri teoriyle yönlendirilmez ve sonuçların yorumlanmasını sağlamaz. Şablonları ve korelasyonları belirlerler. (Hilbert, 2016, s. 137-140)

Veri madenciliği; makine öğrenimi, istatistikler ve veri tabanı sistemlerinin kesişimindeki yöntemleri içeren büyük veri setindeki şablonları keşfetme sürecidir. Veri madenciliği yaparken büyük veri tabanlarından öngörülmeven bilgi alınır. Veri madenciliği için kullanılan araçlar alınan öngörülmeven veriden şablonları belirler, şablonlardaki ilişkileri ve kuralları yorumlarlar. Bu şablonlar ve kurallar sonrasında karar verme sürecine rehberlik ederler ve verilecek kararların etkileri konusunda öngörüde bulunurlar. (Benoit, 2002, s. 265) Veri madenciliği ve veri analizi arasındaki fark; veri analizi bir pazarlama kampanyasının etkinliğini analiz etmek gibi anı özetlerken, veri madenciliği ilerisini öngörüp verilerin içindeki şablonları keşfetmek için spesifik makine öğrenimi ve istatistiksel modelleri kullanmaya odaklanır.

Veri madenciliğinde, veriyi ve yaptığı işi bilmek önemlidir. Eldeki veri uygun ortamda analiz edilir. Verinin içeriği ve muhtemel formu doğrulanıp veri gerektiğinde araçlarla bütünleştirilir. Veride belli hatalar aranıp bulunan hatalar, hatalı kayıtlar veya önemsiz değerler ortadan kaldırılır. Pasif tekniklerle ve detaylandırmayla yeni veri özellikleri elde ederek, analiz yapılacak uygun modeli seçerek ve test edilecek hipotezler geliştirilerek basit veri incelemesi yapılır. Şablonları ve yeni bilgiyi ortaya çıkarmak veya geliştirilmiş hipotezleri keşfetmek için ana keşif prosedürleri uygulanır. Test setlerindeki tahminî modelleri denemek ve segmentasyonları analiz etmek de dâhil olmak üzere keşfedilmiş bilgi değerlendirilir. (Benoit, 2002, s. 270-271)

Veri madenciliğinde işler; özetleme, sınıflandırma, kümeleme, ilişkilendirme ve trend analizi olarak beş gruba ayrılabilir. Özetleme, işle ilgili veri setinin genel değerlendirmesidir. Sınıflandırma, nesnenin özelliklerine bağlı olarak sınıfını belirleyen modelin veya işlevin kaynağıdır. İlişkilendirme, nesnelerin birlikteliğinin veya bağlantısının keşfidir. Kümeleme, sınıfı bilinmeyen nesne setinin sınıf tanımlamasıdır. Trend analizi, nesnelerin değişim geçmişlerindeki şablonlarını ve değişen trendlerini bulur. (Luo, 2008, s. 4)

1.4. Semantik ve Elasticsearch

Semantik; dilde, programlama dillerinde, biçimsel mantıkta ve gösterge biliminde dil bilimsel ve anlamsal çalışmadır. Kelimeler, söz öbekleri, işaretler ve semboller gibi ifadeler ve manaları arasındaki ilişkiyle alakalıdır. Sesler, yüz ifadeleri, vücut dili ve vücudun bulunduğu ortamdaki konumu semantik içeriğe sahiptir. Yazı dilinde paragraf yapısı ve noktalama işaretleri semantik içerik taşır. Semantik; sözcük bilimi, edim bilimi ve köken bilimi gibi pek çok araştırma alanıyla kesişir.

(Focus On, 2018, Semantics Bölümü)

Elasticsearch, Java ile yazılmış dağıtık ve ölçeklenebilir platform bağımsız tam metin arama motorudur. Elasticsearch, bu özelliklerin yanı sıra esneklik sağlamasıyla gerçek zamanlı büyük veri analizinde çok kullanışlı olur. (Shah, 2018, s. 2) Elasticsearch büyük veri miktarını ele alıp işleyebilir ve anahtar kelimeleri gerçek zamanlıya yakın şekilde belirleyebilir. Bu özelliği Elasticsearch'ü, bütün dünyaya son dakika haberlerini ve mesajlarını ulaştırma konusunda hızlı bir aracı olan Twitter gibi sosyal medya uygulamaları için uygun kılar. Kümeleme (Cluster) dostu bir sistem olduğundan dolayı çok düğümlü kümelemelerin rahatça çalışmasına olanak sağlar. Twitter bu araçtan küresel boyuttaki trendlerini çabucak belirlemek için faydalanabilir. Twitter'daki mesajların metin yapısı göz önünde tutulursa, Elasticsearch gibi bir arama aracı, gelişmiş arama işlevselliği ve yorumlara, sorulara hızlı yanıt vermeye olanak sağlaması açısından idealdir. Ayrıca Elasticsearch'ün görsel veri arayüzü Kibana, yaygın olarak kullanılan önemli bir görsel açık kaynak aracıdır. Kibana basit kullanıcı arayüzüyle, biçimlendirilebilen popüler konu veya tweetler gibi etkileşimli görsel sunabilir. (Zhu, 2018, s. 361) Elasticsearch belli başlı özellikleriyle ön plana çıkar:

- Dağıtık; Elasticsearch sunucusu, tek düğümlü başlatılabilir ve gereksinimlere bağlı olarak yatay bir şekilde ölçeklendirilebilir. Daha fazla kapasiteye ihtiyaç duyulursa, daha fazla düğüm eklenebilir.
- Yüksek erişilebilirlik; Elasticsearch, Elasticsearch kümelemesi oluşturabilir. Kümeleme hata esnekliğine sahiptir. Herhangi bir hata tespit edilirse, otomatik olarak başarısız düğümleri kaldırıp verinin güvenli ve erişilebilir olduğundan emin olmak için kendini yeniden düzenler.
- Tam metin araması; Arama kapasitesine dayanan tam sorgulama sağlar.
- Doküman odaklı; Elasticsearch veriyi veya dokümanları depolar. Tüm dokümanlar çok hızlı sonuç sağlaması yüzünden varsayılan olarak dizinlenir.
- Şema bağımsız; Dokümanlar kolay şekilde depolanabilir. Elasticsearch otomatik olarak veri yapısı ve tipini belirler ve bu doğrultuda veriyi dizinler. Kullanıcı kendi şemasını tanımlayıp gerektiğinde değiştirebilir. (Zheng, 2014, s. 471)

2. GEREÇ VE YÖNTEM

2.1. Keyhole ve Zemberek

Keyhole, milyonlarca insanın sosyal medya ve internet üzerindeki etkileşimini anlamaya yardımcı olan bir sosyal medya analiz aracıdır. (Keyhole)

Where do you want to track #huawei?
We will collect all of its mentions from the following sources in real-time.

- Youtube Upgrade
- Twitter
- Instagram
- News
- Blogs
- Forums
- Facebook
Choose which public Facebook pages to track your search term in. You can add more pages or change the suggested ones.

Create Tracker

Real-time Tracker: #huawei JUL 02 - JUL 02

Search Posts Search Geography Q SENTIMENT EXCLUDE RETWEETS RESET EMBED CSV

USER	POST	REGION	COUNTRY	ENGAGEMENT	SENTIMENT	DATE
9samera @9samera	@Abdullah_iApp رحلت الهمزة رابتر طاران Huawei 5G CPE Pro H112-370 شكرا @Abdullah_iApp شكرا @STC_KSA #STC5G #Huawei...	-	-	0	😊	Jul 2, 2019 2:20 PM
Ceotechl @Ceotechl	Retweeted from @Ceotechl Tronsmart Element Groove Bluetooth Speaker CeoTech https://t.co/rCjGClA395 Link Amazon: https://t.co/diem7SKdoy @Tronsmart...	-	-	0	😊	Jul 2, 2019 2:20 PM
marypcbuk @marypcbuk	The problem that emerged in the UK Huawei review is that their development process is such that there is no chain of custody from source code to shipping produ...	-	-	0	😞	Jul 2, 2019 2:20 PM
Bigapple2la @Bigapple2la	@robert_spalding That was a ridiculous performance on #CNBC. @andrewsorkin caught you lying about #Huawei, didn't he	-	Hong Kong	0	😡	Jul 2, 2019 2:20 PM

Şekil 2.1. Keyhole'de Twitter'dan Veri Alımı

Twitter üzerinden takip edilecek etiket belirlenip dil, tarih, istenirse konum gibi özellikler seçilerek amaçlanan veri elde edilir. Keyhole üzerinden duygu analizi de yapılabilir.

Zemberek bir doğal dil işleme kütüphanesidir. İçinde bulundurduğu modüller sayesinde morfolojik analiz, sözcükler arasındaki belirsizliği giderme, sözcük türetme, cümle sonu tespiti, basit yazım denetleme, metin sınıflandırma, normalizasyon gibi özelliklere sahiptir.

```
Python 3.7.3 (v3.7.3:ef4e0ed12, Mar 25 2019, 22:22:05) [MSC v.1916 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\IA\Desktop\Zemberek\Zemberek-Python-Examples-master\examples\normalization\noisyy_text_normalization.py

Warning (from warnings module):
  File "C:\Users\IA\AppData\Local\Programs\Python\Python37\lib\site-packages\jpyype\core.py", line 210
    """
UserWarning:
-----
Deprecated: convertStrings was not specified when starting the JVM. The default
behavior in JPyype will be False starting in JPyype 0.8. The recommended setting
for new code is convertStrings=False. The legacy value of True was assumed for
this session. If you are a user of an application that reported this warning,
please file a ticket with the developer.
-----

Noisy Sentence 0: Yrn okua gidicem
Normalized Sentence 0: yarın okula gideceğim

Noisy Sentence 1: Tmm, yarın havuza giricem ve aksama kadar yaticam :)
Normalized Sentence 1: tamam , yarın havuza gireceğim ve akşama kadar yatacağım :)

Noisy Sentence 2: ah aynen ya annemde fark etti siz evinizden çıkmayın diyo
Normalized Sentence 2: ah aynen ya annemde fark etti siz evinizden çıkmayın diyor

Noisy Sentence 3: gercek mi bu? Yuh! Artık unutulması bile beklenmiyo
Normalized Sentence 3: gerçek mi bu ? yuh ! artık unutulması bile beklenmiyor

Noisy Sentence 4: Hayır hayat telağm olmasa alacam buraları gökdelen dikicem.
Normalized Sentence 4: hayır hayat telaşı olmasa alacağım buraları gökdelen dikeceğim .

Noisy Sentence 5: yok hocam kesinlikle öyle birşey yok
Normalized Sentence 5: yok hocam kesinlikle öyle bir şey yok

Noisy Sentence 6: herşeyi soyle hayatında olmaması gerek bence böyle insanların falan baskı yapıyorsa
Normalized Sentence 6: herşeyi söyle hayatında olmaması gerek bence böyle insanların falan baskı yapıyorsa

>>>
```

Şekil 2.2. Zemberek Normalizasyonu

Zemberek kullanarak yazım kurallarına uymayan kelimeler normalizasyon işlemiyle dile uygun hâle getirilebilir.

2.3. Elasticsearch

Elasticsearch ile büyük veri, istenen veri aralığına indirgenebilir.

ID	TwtAtil...	TwtAtanHesapAdi	TwtAtanKullaniciID	TwtAtanKullaniciAdi	Twt	TwtID	TwtKategorisi	Benzerlik	
1	1105	2019-0...	şeytanın gör dediği	127191429	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141066700307250000	samsung	172
2	1107	2019-0...	Doruk Yılmaz	1462326894	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141064275106160000	samsung	172
3	1112	2019-0...	mobayl	2789995601	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141068072175050000	samsung	172
4	1117	2019-0...	Faruk Tin	1311295172	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141058939213280000	samsung	172
5	1118	2019-0...	MKA.T.C. Suat Erdim	363725964	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141058401616550000	samsung	172
6	1132	2019-0...	çarılı	4825240101	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141048365494550000	samsung	172
7	1110	2019-0...	Dursun KUŞ	260044939	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141062387409340000	samsung	172
8	1136	2019-0...	Ege Akdeniz	562362372	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141047204389530000	samsung	172
9	1137	2019-0...	??Osman Ak	361086102	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141047105533960000	samsung	172
10	1142	2019-0...	kasım ozcan	844829607509409000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141041882249740000	samsung	172
11	1144	2019-0...	PROF. DR. FENERGE...	1463842212	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141040560096100000	samsung	172
12	1148	2019-0...	Esat Aday	2332451103	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141037568391340000	samsung	172
13	1151	2019-0...	ihsan	824903719078670000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141035036440970000	samsung	172
14	1154	2019-0...	Türkçü Erbek/Andımız v...	3650634075	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141032422399750000	samsung	172
15	1158	2019-0...	Mustafa seyri	2337090102	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141031815047710000	samsung	172
16	1159	2019-0...	C'HakoBabaC*	197727482	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141031764544100000	samsung	172
17	1161	2019-0...	MAHŞER	337692564	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141030062956970000	samsung	172
18	1169	2019-0...	Meral	1068877084754350000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141023130707730000	samsung	172
19	1170	2019-0...	ceylanıyağlıkara	968792183947124000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141022517953510000	samsung	172
20	1173	2019-0...	Ali aksoy 1919	1127986476120240000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141019688228580000	samsung	172
21	1174	2019-0...	Köle	4310569576	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	114101877389320000	samsung	172
22	1175	2019-0...	CEM	216485192	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141017923307390000	samsung	172
23	1179	2019-0...	? Güneşim ?	1627574041	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141015733146720000	samsung	172
24	1182	2019-0...	Murat İhsan	2759338667	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141011218083260000	samsung	172
25	1183	2019-0...	sedat gündüz	96000618	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141011075963530000	samsung	172
26	1190	2019-0...	Yavuz Çınar	1129915032408660000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	114100842330490000	samsung	172
27	1192	2019-0...	Sadi Arduç	2184728961	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141007118931240000	samsung	172
28	1194	2019-0...	yusuflb	74979431	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141005602023450000	samsung	172
29	1197	2019-0...	İrfan Sönmez	784305758	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1141003732664160000	samsung	172
30	1201	2019-0...	umit onal	1379465132	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140999907152660000	samsung	172
31	1205	2019-0...	(E.Y.T) Ülkay KÜÇÜKE...	824893548	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140997934365970000	samsung	172
32	1207	2019-0...	Firat	3356604347	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140997347419280000	samsung	172
33	1209	2019-0...	Tunay Öztürk	1054331461657130000	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140996187480290000	samsung	172
34	1219	2019-0...	Nadir AKAY	978972115	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140991333622000000	samsung	172
35	1220	2019-0...	Sessizçiglik	3600117628	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140991298989640000	samsung	172
36	1222	2019-0...	Kenan Yazıcı	909749672	NULL	rt @hazaluğur9 : 69 yıl önce kurtarmak için ask...	1140990939304550000	samsung	172

Şekil 2.4. Elasticsearch ile Benzer Tweetlerin Belirlenmesi

Elasticsearch, bir kullanıcıya ait olan tweetin başka kullanıcılar tarafından yayınlanması veya aynı içeriğe sahip tweetin atılması sonucunda ortaya çıkan benzer tweetleri belirleyebilir.

TwitAtilanZaman	TwitAtanHesapAdi	TwitAtanKullaniciD	TwitAtanKullaniciAdi	Twit	TwitID	TwitKategorisi	Benzerlik
2019-06-26 11:46:00.0000000	ozyavasyasn	225040461	@ozyavasyasn	atmosferin kalnlğının 1000 kilometreden fazla olduđu...	1143802810749910000	huawei	1
2019-06-26 11:46:00.0000000	Halis Türk	1895284297	@Halis_Turk	@haku_alkan huawei güçlenmesi telefon kullanıcıları a...	1143802601399640000	huawei	1
2019-06-26 11:40:00.0000000	ShiftDelete Net	28332200	@shiftdeletenet	huawei p20 lite ve mate 20 için emri 9.1 yayınlandı https...	1143801220454370000	huawei	1
2019-06-26 11:38:00.0000000	Hakki Alkan	28147896	@hakki_alkan	seni öldürmeyen şey , daha da güçlendirir reis ?? ? htt...	1143800662330880000	huawei	1
2019-06-26 11:36:00.0000000	Abdullah Ayan	75607527	@abdullahayan	huawei batı blokuna giren truva atı mı ? (11.6.2019) h...	1143800228044240000	huawei	1
2019-06-20 07:47:00.0000000	Türkçe	25992620	@turkce	huaweinin kurucusu 30 milyar dolarlık kaybı önemsemiy...	1141568124065430000	huawei	1
2019-06-20 07:09:00.0000000	Alone HERO	3429287247	@iamsuperherotr	@tinercimayki lan mal iphone leş gibi bir kere özgürlüğü...	1141558699858750000	huawei	1
2019-06-20 06:46:00.0000000	Türkçe	25992620	@turkce	çok iddialı honor 20 ailesi için avrupa çıkması ! : huaw...	1141552900037930000	huawei	1
2019-06-20 06:01:00.0000000	Teknoloji	637354321	@Teknoloji	huawei : zarara uğramak bizi yıldırmaz ! https://t.co/reqv...	1141541446501090000	huawei	1
2019-06-20 05:24:00.0000000	Ayrıntılı Haber	3147473994	@AyrıntılıHaberV	huaweinin kurucusu 30 milyar dolarlık düşüşü önemse...	1141532266583280000	huawei	1
2019-06-20 05:06:00.0000000	Göksel KAPLAN	3824066861	@gokskapln	huawei telefonu ilk çöp olan ben miyim . dokunmatik ke...	1141527765725790000	huawei	1
2019-06-20 02:59:00.0000000	Bülent Kocabyık	141223306	@twitvole	rt @gazetemag : huawei kurucusu ren zhengfei , en kö...	1141495715925040000	huawei	1
2019-06-20 02:58:00.0000000	GazeteMAG	333716756	@GazeteMAG	huawei kurucusu ren zhengfei , en kötü senaryoyu açıkl...	1141495426547440000	huawei	1
2019-06-20 02:22:00.0000000	Berat ÇETİN	4754608761	@beratcetin_mp4	herkesin telefon zil sesi neden huawei reklam müziğı ol...	1141486452179630000	huawei	1
2019-06-20 01:48:00.0000000	#ANDIMIZ #VATAN S...	1474707349	@oruncutvatan	rt @arbeitervon : #cnntrk @idli20231 coming ? huaw...	1141477978951600000	huawei	4
2019-06-20 01:22:00.0000000	work Net haber	329000121	@otylmaz	huaweinin kurucusu zhengfei , 30 milyar dolarlık düşüş...	1141471384759130000	huawei	1
2019-06-20 00:39:00.0000000	on5yirmi5	106050885	@on5yirmi5	huawei abd yaptırımlarına karşı ilginç bir uygulamaya ba...	1141460601031520000	huawei	1
2019-06-20 00:21:00.0000000	emre akkaya	250721978	@emredenizakkaya	rt @capamagekonomi : #huawei kurucu ve ceosu #ren...	1141455982901500000	huawei	2
2019-06-20 00:20:00.0000000	Melikovski?????????	1542725791	@MelikeKarakelli	rt @relaksol35 : 360 yıldı bir güneş sisteminde görüne...	1141455744228820000	huawei	1
2019-06-19 23:47:00.0000000	Av. Mehmet Bostancı	461973626	@AvMBoSS	@milimetrikgo amerikanın huawei'ye uyguladığı android...	1141447326910880000	huawei	1
2019-06-19 23:45:00.0000000	ş	1381805400	@cortexphan	@lovesyoumalk Huaweinin Google'le ilgili problemleri olm...	1141446828178700000	huawei	1
2019-06-19 23:32:00.0000000	barann	595702706	@brn_1914	merhaba ! bu ilanı şimdi letgoda yayınladım. ilgilenir mi...	1141443746501930000	huawei	1
2019-06-19 23:32:00.0000000	Adam gibi Adam	3086667724	@cesar_pavese	@durdenwashere yahu boş konuşuyorsun 5g nin nere...	1141443562783030000	huawei	1
2019-06-19 23:29:00.0000000	4 haber 4Haber.Com	2769340759	@4habercom	Huaweinin sahibinden abdye ayar ! bizim için çok küçük...	1141442847280890000	huawei	1
2019-06-19 23:23:00.0000000	Turkish Market	529464675	@kamerknc	huawei ceosu : trump'la konuşmak isterim .	1141441360542780000	huawei	1
2019-06-19 23:19:00.0000000	???	1319734920	@Whisper_ist	1 senede %100 artar mı bir makina ya ! şaşırms bunlar ...	1141440436718000000	huawei	1

Şekil 2.5. Elasticsearch ile Kullanıcı Adı Ayrıştırması

Elasticsearch'te kullanıcı kimlikleri yardımıyla Twitter üzerinden sorgulama yapıp kullanıcı adı elde edilmesi için gelen HTML cevabı kaydedilir ve gelen HTML cevap üzerinden HtmlAgilityPack aracılığıyla kullanıcı adı olan HTML kod içerisindeki kullanıcı adları ayrıştırılır.

2.4. RapidMiner ve Tinfoleak

Tinfoleak, Twitter kullanıcısıyla ilgili basit bilgileri araştırır. Twitter kullanıcıları tarafından kullanılan araçları, işletim sistemlerini, uygulamaları, sosyal ağları tespit eder. Twitter kullanıcısının tweet atarken kullandığı etiketleri ve kullandıkları tarihi saptar. (Tinfoleak)

Tinfoleak üzerinden yapılan açık kaynak analiziyle, bu çalışmada kullanılan tweet atan kullanıcının gerçek kullanıcı olup olmadığını tüm tweetler üzerinden analiz ederek bulma işlemi gerçekleştirilmektedir.

Şekil 2.6'da gösterilen rapor örneğinde ise, kullanıcının Twitter'a giriş tarihi, konumu, bulunduğu bölgenin saati, kullandığı dil, bio kısmında yazan veriler gibi bilgiler bulunmaktadır. Bunların dışında, istatistik alanında ise, tüm tweet sayısı, tüm takipçi sayısı, takip ettiklerinin sayısı, bunların oranı, tweet listelerine alınma durumu gibi istatistikler bulunmaktadır. Başlıklar, etiketler ve bahsedilme kısmında ise, genel olarak bahsettiği kelimeler, kullandığı etiketler ve bahsettiği/bahsedilen kişiler gibi bilgiler bulunmaktadır.

Kullanıcının analizi için gerekli olan kısım ise, tweet analizi olarak görünmektedir. Bu alanda kullanıcının attığı tweetlerin yüzde olarak yanıtlanma oranı, bahsedilme oranı, etiket kullanma oranı, yeniden tweetlenme oranı, bağlantı bulundurma oranı, medya bulundurma oranı, genel olarak kullandığı Twitter uygulaması (Twitter Web, Android için Twitter uygulaması, IOS için Twitter uygulaması, Twitter için masaüstü uygulaması) gibi alanları göstermektedir. Burada gösterilen oranlardan, atılan tweetlerin yanıtlanma oranı ve bahsedilme oranı %20'nin altında ise bu kullanıcı tweetleri güvenilir olarak görülmüştür. Tüm tweetlerinde bağlantı ve medya bulunan hesaplar, muhtemelen bir sistemle bağlantılı (Örn. Haber Sitesi) ve otomatik olarak tweet göndermektedir. Bu tweetlerin de son kullanıcı yorumu olarak değerlendirilmesi mümkün değildir. Bunların yanı sıra atılan tweetlerin popüler olarak görünmesi için kullanıcılar sahte kullanıcılardan yeniden tweetlenme desteği veya beğeni desteği almaktadır. Bütün tweetleri yeniden tweetlenmiş olan kullanıcı hesapları, başkaları tarafından kullanılan yüzlerce robot (bot) hesaplardan birisi olma ihtimalini güçlendirmektedir. Dolayısıyla tüm tweetleri yeniden tweetlenmiş olan hesapların yeniden tweetlediği konular ciddiye alınmamıştır.

Son olarak, analiz içerisinde kullanıcının attığı tüm tweetlerin saatler arasında bölünmüş göstergeleri bulunmaktadır. Bu zaman çizelgesi üzerinden, kullanıcının belli zaman aralıklarında attığı tweetlere yapılacak analizde, günün sadece belli bir saatinde (Örn: Gece 01.00'da) çalışan bir uygulamanın otomatik yaptığı beğeniler ciddiye alınmamıştır.

RapidMiner; veri hazırlama, makine öğrenimi, metin madenciliği ve tahminî analizler için bütünleşik bir ortam sağlayan veri bilimi yazılım platformudur. (RapidMiner)

Row No.	polarity_con...	subjectivity_...	polarity	subjectivity	Created-At	From-User	From-User-Id	Text	Id
1	0.922	0.756	neutral	subjective	6/26/19 10:20...	Abdullah Çiftçi	2787936004	microsoft : ' h...	1143781007...
2	0.932	1	neutral	subjective	6/26/19 10:03...	Sputnik Türkiye	424893128	microsoft ve i...	1143776808...
3	0.913	1.000	neutral	objective	6/20/19 8:05 ...	HAVELSAN	4772520083	5g ve uygula...	1141572884...
4	0.614	0.853	neutral	subjective	6/26/19 1:57 ...	1 Çift Kumru	1387070582	huawei sağol...	1143835754...
5	0.734	1.000	neutral	objective	6/26/19 1:54 ...	orkide	844141364	rt @abdullah...	1143834979...
6	0.734	1.000	neutral	objective	6/26/19 1:54 ...	Qguzz Devam...	2986615031	rt @abdullah...	1143834796...
7	0.937	0.695	neutral	objective	6/26/19 1:51 ...	Yusuf Erik	3341820987	huawei telefo...	1143834111...
8	0.734	1.000	neutral	objective	6/26/19 1:48 ...	kintar	218847108	rt @abdullah...	1143833429...
9	0.734	1.000	neutral	objective	6/26/19 1:44 ...	Hasan TOKG...	3222379804	rt @abdullah...	1143832450...
10	0.759	1.000	neutral	objective	6/26/19 1:42 ...	lil lila	8263930168...	@urienyum h...	1143831921...

Şekil 2.7. RapidMiner ile Duygu Analizi

RapidMiner'daki duygu analizi modülüyle atılan tweetin subjektif olup olmadığı belirlenir.

3. BULGULAR

Kişilerin sosyal medya kullanım alışkanlıklarını ölçerek sosyal medya kullanım farkındalığını, sosyal medyadaki güvenlik bilincini ortaya koyup dijital anket yapılmıştır. 4102 kişi üzerinde yapılan dijital ankette, kullanıcıların sosyal medya alışkanlıkları ve farkındalık bilinci ölçümlenmiştir. Bu ölçümlerde sosyal medyada karşılaşılan sorunlar, bilgi içeren paylaşımlar ve bu paylaşımların sebep olduğu sorunlar analiz edilmiştir. Kişilerin sosyal medya kullanımında, paylaşılan verilerin hangi amaçla üçüncü kişilerin eline geçtiği, bu verilerin üçüncü kişiler tarafından hangi maksatla kullanılabilceği ve ne tarz sorunlar yaşanabileceğine ilişkin farkındalık ölçümü yapılmıştır. Bu ölçümler yaş, konum, eğitim durumu, kullanım sıklığı, sosyal medya kullanım amacına göre tasnif edilmiştir.

Çizelge 3.1. Cinsiyet

	Sayı	Yüzde
Kadın	650	15,8
Erkek	3452	84,2
Total	4102	100,0

Dijital ankete katılan kullanıcıların 650'si kadın, 3452'si erkek olarak ölçümlenmiştir. Yüzde 84,2 ile erkeklerin daha fazla katılım gösterdiği gözlemlenmiştir.

Çizelge 3.2. Yaşanılan Coğrafi Bölge

	Sayı	Yüzde
Marmara Bölgesi	1433	34,9
Ege Bölgesi	495	12,1
Akdeniz Bölgesi	319	7,8
İç Anadolu Bölgesi	1261	30,7
Karadeniz Bölgesi	261	6,4
Doğu Anadolu Bölgesi	118	2,9
Güneydoğu Anadolu Bölgesi	160	3,9
Yurt dışı	55	1,3
Total	4102	100,0

Dijital ankete katılan kullanıcıların 1433'ü Marmara Bölgesi, 495'i Ege Bölgesi, 319'u Akdeniz Bölgesi, 1261'i İç Anadolu Bölgesi, 261'i Karadeniz Bölgesi, 118'i Doğu Anadolu Bölgesi, 160'ı Güneydoğu Anadolu Bölgesi, 55'i de yurt dışından katılım göstermiştir. Katılımcıların büyük çoğunluğunun yüzde 34,9 ile Marmara Bölgesi ve yüzde 30,7 ile İç Anadolu Bölgesi'nden katılım gösterdiği gözlemlenmiştir.

Çizelge 3.3. Yaş

	Sayı	Yüzde
18 yaş altı	480	11,7
18 - 21 yaş arası	908	22,1
22 - 25 yaş arası	923	22,5
26 - 29 yaş arası	723	17,6
30 yaş ve üzeri	1068	26,0
Total	4102	100,0

Dijital ankete katılan kullanıcıların 480'i 18 yaş altı, 908'i 18 – 21 yaş arası, 923'ü 22 – 25 yaş arası, 723'ü 26 – 29 yaş arası, 1068'inin ise 30 yaş ve üzeri olarak katılım gösterdiği ölçümlenmiştir. Katılımcıların yüzde 44,6'sının 18 – 25 yaş arası genç kullanıcılar olduğu gözlemlenmiştir.

Çizelge 3.4. Eğitim Durumu

	Sayı	Yüzde
İlkokul – Lise	1257	30,6
Lisans	2364	57,6
Lisansüstü	481	11,7
Total	4102	100,0

Dijital ankete katılan kullanıcıların 1257'sinin ilkokul – lise, 2364'ünün lisans, 481'inin lisansüstü eğitim düzeyine sahip olduğu ölçümlenmiştir. Katılımcıların yüzde 57,6'sının eğitim düzeyinin lisans olduğu gözlemlenmiştir.

Çizelge 3.5. Aile/arkadaşlar ile iletişim ve/veya tanışma için kullanma durumu

	Sayı	Yüzde
Hayır	1955	47,7
Evet	2147	52,3
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyayı aile/arkadaşlarla iletişim veya tanışmak için kullanım durumu; 1955 hayır, 2147 evet olarak ölçümlenmiştir. Kullanıcıların sosyal medyayı kullanma amacının aile/arkadaşlarla iletişim veya tanışma oranının neredeyse yarı yarıya olduğu gözlemlenmiştir.

Çizelge 3.6. Fotoğraf/video paylaşımı ve/veya eğlenmek için kullanma durumu

	Sayı	Yüzde
Hayır	976	23,8
Evet	3126	76,2
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyayı fotoğraf/video paylaşımı veya eğlenmek için kullanma durumu; 976 hayır, 3126 evet olarak ölçümlenmiştir. Kullanıcıların yüzde 76,2'lik büyük çoğunluğunun sosyal medyayı kullanım amacının fotoğraf/video paylaşımı veya eğlenmek için olduğu gözlemlenmiştir.

Çizelge 3.7. Gündem takibi ve/veya fikir belirtmek için kullanma durumu

	Sayı	Yüzde
Hayır	1931	47,1
Evet	2171	52,9
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyayı gündem takibi veya fikir belirtmek için kullanım durumu; 1931 hayır, 2171 evet olarak ölçümlenmiştir. Kullanıcıların sosyal medyayı kullanma amacının gündem takibi veya fikir belirtmek için kullanma oranının neredeyse yarı yarıya olduğu gözlemlenmiştir.

Çizelge 3.8. Ticaret için kullanma durumu

	Sayı	Yüzde
Hayır	3012	73,4
Evet	1090	26,6
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyayı ticaret için kullanım durumu; 3012 hayır, 1090 evet olarak ölçümlenmiştir. Kullanıcıların büyük çoğunluğunun sosyal medyayı ticaret amaçlı kullanmadığı gözlemlenmiştir.

Çizelge 3.9. Sosyal medyada geçirilen zaman (1 gün içinde)

	Sayı	Yüzde
1 saat	1132	27,6
2 saat	1082	26,4
3 saat	657	16,0
4 saat ve üzeri	1231	30,0
Total	4102	100,0

Dijital ankete katılan kullanıcıların bir gün içerisinde 1 saat kullanımının 1132 kişi, 2 saat kullanımının 1082 kişi, 3 saat kullanımının 657 kişi, 4 saat ve üzeri kullanımının 1231 kişi olduğu ölçümlenmiştir. Kullanıcıların yüzde 30'luk oranının bir gün içerisinde 4 saat ve üzeri sosyal medyada zaman geçirdiği gözlemlenmiştir.

Çizelge 3.10. Paylaşımlarda konum belirtme durumu

	Sayı	Yüzde
Paylaşıyor	868	21,2
Paylaşmıyor	3234	78,8
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medya paylaşımlarında konum belirtme durumu; paylaşan 868 kişi, paylaşmayan 3234 kişi olduğu ölçümlenmiştir. Kullanıcıların yüzde 78,8'lik oranının sosyal medya paylaşımlarında konum belirtmediği gözlemlenmiştir.

Çizelge 3.11. Verilerin üçüncü kişilere satılma/paylaşılma olasılığını bilme durumu

	Sayı	Yüzde
Biliyor	3423	83,4
Bilmiyor	679	16,6
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada paylaştığı verilerin üçüncü kişilere satılma/paylaşılma olasılığını bilme durumu; bilen 3423 kişi, bilmeyen 679 kişi olduğu ölçümlenmiştir. Kullanıcıların yüzde 83,4'lük oranının sosyal medyada paylaştığı verilerin üçüncü kişilerin ellerine geçtiğini bildiği hâlde paylaştığı gözlemlenmiştir.

Çizelge 3.12. "10yearschallenge" aktivitesini görme/katılma durumu

	Sayı	Yüzde
Gördüm ve kullandım	119	2,9
Gördüm ve kullanmadım	3122	76,1
Görmedim	861	21,0
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada trend hâline gelen "10yearschallenge" aktivitesine görme/katılma durumu; görüp kullanan 119 kişi, görüp kullanmayan 3122 kişi ve görmeyen 861 kişi olduğu ölçümlenmiştir. Kullanıcıların yüzde 76,1'nin aktiviteyi gördüğü fakat kullanmadığı gözlemlenmiştir.

Çizelge 3.13. Demografik bilgi paylaşma durumu

	Sayı	Yüzde
Hayır	1089	26,5
Evet	3013	73,5
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada demografik bilgi paylaşım durumu; 1089 hayır, 3013 evet olduğu ölçümlenmiştir. Kullanıcıların yüzde 73,5'inin sosyal medyada demografik bilgi paylaştığı gözlemlenmiştir.

Çizelge 3.14. Özel Hayat bilgisi paylaşma durumu

	Sayı	Yüzde
Hayır	2873	70,0
Evet	1229	30,0
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada özel hayatına ilişkin bilgi paylaşma durumu; 2873 hayır, 1229 evet olduğu ölçümlenmiştir. Kullanıcıların yüzde 70'lik oranının sosyal medyada özel hayatına ilişkin bilgi paylaşmadığı gözlemlenmiştir.

Çizelge 3.15. Ekonomik bilgi paylaşma durumu

	Sayı	Yüzde
Hayır	1587	38,7
Evet	2515	61,3
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada ekonomik bilgilerini paylaşma durumu; 1587 hayır, 2515 evet olduğu ölçümlenmiştir. Kullanıcıların yüzde 61,3'lük oranının sosyal medyada ekonomik bilgi paylaşımı yaptığı gözlemlenmiştir.

Çizelge 3.16. Hiç bilgi paylaşmama durumu

	Sayı	Yüzde
Hayır	3447	84,0
Evet	655	16,0
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada bilgi paylaşıp paylaşmama durumu; 3447 hayır, 655 evet olduğu ölçümlenmiştir. Kullanıcıların yüzde 84'lük oranının sosyal medyada hiç bilgi paylaşmadığı gözlemlenmiştir.

Çizelge 3.17. Sosyal medya aracılığıyla tanışma durumu

	Sayı	Yüzde
Evet, tanıştım	2521	61,5
Hayır, tanışmadım	1581	38,5
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medya aracılığıyla tanışma durumu; 2521 evet, 1581 hayır olduğu ölçümlenmiştir. Kullanıcıların yüzde 61,5'lik oranının sosyal medya aracılığıyla farklı insanlarla tanıştığı gözlemlenmiştir.

Çizelge 3.18. Dolandırılma durumu

	Sayı	Yüzde
Evet, dolandırıldım	323	7,9
Hayır, dolandırılmadım	3779	92,1
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medya üzerinden dolandırılma durumu; 323 evet, 3779 hayır olduğu ölçümlenmiştir. Kullanıcıların yüzde 92,1'lik oranının sosyal medya üzerinde dolandırılmadığı gözlemlenmiştir.

Çizelge 3.19. Gerçek isim/fotoğraf kullanma durumu

	Sayı	Yüzde
Kullanıyor	3569	87,0
Kullanmıyor	533	13,0
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada gerçek kimliğini kullanma durumu; 3569 evet, 533 hayır olduğu ölçümlenmiştir. Kullanıcıların yüzde 87'lik oranının sosyal medyada gerçek kimliğini kullandığı gözlemlenmiştir.

Çizelge 3.20. Popüler konulara # ile katılma durumu

	Sayı	Yüzde
Katılıyor	612	14,9
Katılmıyor	3490	85,1
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada popüler konulara etiket kullanarak katılma durumu; 612 evet, 3490 hayır olduğu ölçümlenmiştir. Kullanıcıların yüzde 85,1'lik oranının sosyal medyada popüler konulara etiket ile katılmadığı gözlemlenmiştir.

Çizelge 3.21. Paylaşılan verilerin kendine karşı kullanımından endişe duyma durumu

	Sayı	Yüzde
Endişe duyuyor	1420	34,6
Endişe duymuyor	2682	65,4
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyada paylaşılan verilerin kendine karşı kullanımından endişe duyma durumu; 1420 evet, 2682 hayır olduğu ölçümlenmiştir. Kullanıcıların yüzde 65,4'lük oranının sosyal medyada paylaşılan verilerin kendine karşı kullanımından endişe duymadığı gözlemlenmiştir.

Çizelge 3.22. Sosyal ağlar güvenli/güvensiz seçimi

	Sayı	Yüzde
Güvenli	305	7,4
Güvensiz	3797	92,6
Total	4102	100,0

Dijital ankete katılan kullanıcıların sosyal medyanın güvenliğine ilişkin durumu; 305 kişinin güvenli olarak düşündüğü, 3797 kişinin ise güvensiz olarak düşündüğü ölçümlenmiştir. Kullanıcıların yüzde 92,6'lık oranının sosyal medyayı güvensiz bulduğu gözlemlenmiştir.

Çizelge 3.23. Paylaşımları Kimler Görebiliyor - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Paylaşımlarınızı kimler görebiliyor	Paylaşım yapmıyorum	4	8,33%	44	91,67%
	Arkadaşlarım	237	7,13%	3087	92,87%
	Herkes	82	11,23%	648	88,77%

Sosyal medyada dolandırılan kişilerin çoğunluğunun paylaşımlarının herkes tarafından görülebildiği gözlemlenmiştir. Dolandırılmayan kişilerin çoğunluğunun ise paylaşımlarının sadece arkadaşları tarafından görüldüğü gözlemlenmiştir.

Çizelge 3.24. Popüler Konulara Katılım - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Popüler konulara # ile katılma durumu	Katılıyor	46	7,52%	566	92,48%
	Katılmıyor	277	7,94%	3213	92,06%

Sosyal medyada popüler konulara etiket ile katılan kişilerin dolandırılma durumu ölçümlenmiştir. Kullanıcıların popüler konulara etiket ile katılımıyla dolandırılma durumu arasında gözle görülebilir bir fark olmadığı gözlemlenmiştir.

Çizelge 3.25. 10yearschallenge - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
"10yearschallenge" aktivitesini görme/katılma durumu	Gördüm ve kullandım	12	10,08%	107	89,92%
	Gördüm ve kullanmadım	254	8,14%	2868	91,86%
	Görmedim	57	6,62%	804	93,38%

Sosyal medyada 10yearschallenge aktivitesini görme/kullanma durumuyla dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. 10yearschallenge etiketini görmüş ve kullanmış kişilerin dolandırılma durumunun daha fazla olduğu, görmeyen ve hiç kullanmayan kişilerin daha az dolandırıldığı gözlemlenmiştir.

Çizelge 3.26. Gerçek İsim/Fotoğraf Kullanma - Dolandırılma Durumu

			Dolandırılma durumu			
			Evet, dolandırıldım		Hayır, dolandırılmadım	
			Sayı	Yüzde	Sayı	Yüzde
Gerçek isim/fotoğraf kullanma durumu	Kullanıyor	282	7,90%	3287	92,10%	
	Kullanmıyor	41	7,69%	492	92,31%	

Sosyal medyada gerçek kimliğini kullanan kişilerin dolandırılma durumu ölçümlenmiştir. Gerçek kimliğini kullanan kişilerin dolandırılma durumunun daha fazla olduğu, gerçek kimliğini kullanmayan kişilerin ise dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.27. Sosyal Medyada Tanışma - Dolandırılma Durumu

			Dolandırılma durumu			
			Evet, dolandırıldım		Hayır, dolandırılmadım	
			Sayı	Yüzde	Sayı	Yüzde
Sosyal medya aracılığıyla tanışma durumu	Evet, tanıştım	237	9,40%	2284	90,60%	
	Hayır, tanışmadım	86	5,44%	1495	94,56%	

Sosyal medya aracılığıyla insanlarla tanışma durumuyla dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Sosyal medya aracılığıyla tanışan insanların dolandırılma durumunun daha fazla olduğu, tanışmayan insanların ise dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.28. Kişisel Veri Paylaşma - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Kişisel veri paylaşma durumu	Paylaşıyor	119	9,31%	1159	90,69%
	Paylaşmıyor	204	7,22%	2620	92,78%

Sosyal medyada kişisel verilerin paylaşımı ile dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Sosyal medyada kişiler verilerinin paylaşımı konusunda hassasiyet sahibi olmayan insanların dolandırılma durumunun daha fazla olduğu, hassasiyet sahibi olan insanların ise dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.29. Eğitim Düzeyi - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Eğitim Durumu	İlkokul - Lise	128	10,18%	1129	89,82%
	Lisans	165	6,98%	2199	93,02%
	Lisansüstü	30	6,24%	451	93,76%

Sosyal medyayı kullanan insanların eğitim düzeyi ile dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Eğitim düzeyi ilköğretim – lise olan kullanıcıların dolandırılma durumunun daha fazla olduğu, eğitim düzeyi lisans ve üzeri olan kullanıcıların dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.30. Sosyal Medya Kullanma Süresi - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Sosyal medyada geçirilen zaman (1 gün içinde)	1 saat	60	5,30%	1072	94,70%
	2 saat	83	7,67%	999	92,33%
	3 saat	41	6,24%	616	93,76%
	4 saat ve üzeri	139	11,29%	1092	88,71%

Sosyal medyada geçirilen zaman durumuyla dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Sosyal medyada 4 saat ve üzeri zaman geçiren insanların dolandırılma durumunun daha fazla olduğu, 1 saat zaman geçirenlerin ise dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.31. Cinsiyet - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Cinsiyet	Kadın	37	5,69%	613	94,31%
	Erkek	286	8,29%	3166	91,71%

Sosyal medyayı kullanan insanların cinsiyet durumuyla dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Erkeklerin sosyal medya üzerinde dolandırılma durumunun daha fazla olduğu, kadınların sosyal medya üzerinde dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.32. Yaş - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Yaş	18 yaş altı	56	11,67%	424	88,33%
	18 - 21 yaş arası	94	10,35%	814	89,65%
	22 - 25 yaş arası	72	7,80%	851	92,20%
	26 - 29 yaş arası	39	5,39%	684	94,61%
	30 yaş ve üzeri	62	5,81%	1006	94,19%

Sosyal medyayı kullanan kişilerin yaş durumu ile dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. 21 yaş ve altı kullanıcıların dolandırılma durumunun daha fazla olduğu, 26 yaş ve üzeri kullanıcıların dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.33. Sosyal Medya Kullanma Nedeni - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Aile/arkadaşlar ile iletişim ve/veya tanışma için kullanma durumu	Hayır	142	7,26%	1813	92,74%
	Evet	181	8,43%	1966	91,57%
Fotoğraf/video paylaşımı ve/veya eğlenmek için kullanma durumu	Hayır	55	5,64%	921	94,36%
	Evet	268	8,57%	2858	91,43%
Gündem takibi ve/veya fikir belirtmek için kullanma durumu	Hayır	158	8,18%	1773	91,82%
	Evet	165	7,60%	2006	92,40%
Ticaret için kullanma durumu	Hayır	226	7,50%	2786	92,50%
	Evet	97	8,90%	993	91,10%

Sosyal medyayı kullanım amacıyla dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Sosyal medyayı aile/arkadaşlar ile iletişim ve/veya tanışma için kullanma durumu evet olan kişilerin, fotoğraf/video paylaşımı ve/veya eğlenmek için kullanma durumu evet olan kişilerin ve ticaret için kullanma durumu evet olan kişilerin dolandırılma durumunun daha fazla olduğu, fotoğraf/video paylaşımı ve/veya eğlenmek için kullanma durumu hayır olan kişilerin ise dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Çizelge 3.34. Coğrafi Bölge - Dolandırılma Durumu

		Dolandırılma durumu			
		Evet, dolandırıldım		Hayır, dolandırılmadım	
		Sayı	Yüzde	Sayı	Yüzde
Yaşanılan Coğrafi Bölge	Marmara Bölgesi	116	8,09%	1317	91,91%
	Ege Bölgesi	50	10,10%	445	89,90%
	Akdeniz Bölgesi	31	9,72%	288	90,28%
	İç Anadolu Bölgesi	84	6,66%	1177	93,34%
	Karadeniz Bölgesi	20	7,66%	241	92,34%
	Doğu Anadolu Bölgesi	5	4,24%	113	95,76%
	Güneydoğu Anadolu Bölgesi	9	5,63%	151	94,38%
	Yurt dışı	8	14,55%	47	85,45%

Sosyal medyayı kullanan kişilerin yaşadıkları bölge ile dolandırılma durumu arasındaki bağlantı ölçümlenmiştir. Ege Bölgesi ve yurt dışında yaşayan insanların dolandırılma durumunun daha fazla olduğu, Doğu Anadolu Bölgesi ve Güneydoğu Anadolu Bölgesi'nde yaşayan insanların dolandırılma durumunun daha az olduğu gözlemlenmiştir.

Tinfoleak ve Mentionmapp aracılığıyla; Twitter'daki 'Samsung', 'iPhone' ve 'Huawei' etiketlerine atılan 7521 tweetin içinden hesabın aktifliği, takipçi oranları gibi olgular göz önünde bulundurularak sahte kullanıcı tweetlerinin çıkarılması sonrasında elde edilen verilere duygu analizi yapılmıştır. Toplamda 2957 verinin içeriğinin subjektif olup olmadığı ölçümlenmiştir.

Çizelge 3.35. Huawei - Subjektiflik Analizi

Subjektiflik	Sayı	Yüzde
Objektif	536	54,416%
Subjektif	447	45,381%
Belirsiz	2	0,203%

Twitter'daki 'Huawei' etiketine atılan 985 tweetin subjektiflik analizi ölçümlenmiştir. Objektif tweetlerin sayısının, subjektif tweetlerin sayısından fazla olduğu gözlemlenmiştir.

Çizelge 3.36. iPhone - Subjektiflik Analizi

Subjektiflik	Sayı	Yüzde
Objektif	564	57,143%
Subjektif	423	42,857%

Twitter'daki 'iPhone' etiketine atılan 987 tweetin subjektiflik analizi ölçümlenmiştir. Objektif tweetlerin sayısının, subjektif tweetlerin sayısından fazla olduğu gözlemlenmiştir.

Çizelge 3.37. Samsung - Subjektiflik Analizi

Subjektiflik	Sayı	Yüzde
Objektif	692	70,254%
Subjektif	292	29,645%
Belirsiz	1	0,102%

Twitter'daki 'Samsung' etiketine atılan 984 tweetin subjektiflik analizi ölçümlenmiştir. Objektif tweetlerin sayısının, subjektif tweetlerin sayısından fazla olduğu gözlemlenmiştir.

4. TARTIŞMA

Bu bölümde, araştırma sürecinde elde edilen bulgular yorumlanmıştır. İnternet üzerinden farkındalık ölçmek amacıyla yapılan dijital ankete 4102 kişi katılmıştır.

Sosyal medyada paylaşımların herkes tarafından görülebilmesi dolandırılmak başta olmak üzere birçok sorunu beraberinde getirmektedir. Sosyal medyada paylaşımlarını sadece arkadaşlarına yapan kişiler ile herkese açık yapan kişilerin dolandırılma sayıları karşılaştırıldığında, paylaşımlarını herkesin görebildiği kişilerin daha çok dolandırılma olayı yaşadığı görülmüştür. Bunların yanı sıra, bazı popüler etiketleri kullanan insanlara, kötü niyetli kişilerin ulaşmasının daha kolay olduğu da görülmüştür. Örneğin; #10yearschallenge etiketini kullanan kişilerin kullanmayanlara oranla daha çok dolandırıcılık olayı yaşadığı görülmüştür. Dolandırılan kişilerin ortak özelliklerine bakıldığında ise, çoğunluğunun sosyal ağlarda kendi isimlerini ve fotoğraflarını kullandıkları belirlenmiştir. Kendi isimlerini ve fotoğraflarını kullanmaları, sanal ortamdaki kimliklerinin gerçekte de keşfedilmesinin daha kolay olduğunu göstermektedir. Özellikle sosyal medyada gerçek kimliklerini kullanan kişilerin, sonrasında sosyal medya üzerinden tanıştığı kişiler ile gerçek hayatta görüşmesi de, dolandırılma oranını ciddi anlamda artırmaktadır. Sosyal medya üzerinden tanışarak, gerçek hayatta görüşen kişilerin sosyal medyada oluşturduğu kimliklerin gerçek kimliklerine dönüşmesi dolandırılma olayını da tetiklemektedir. Gerçek hayatta görüşmeler bile sosyal medyada gerçek kimliklerini ve kişisel bilgilerini paylaşan kişilerin dolandırılma oranının çok daha fazla olduğu tespit edilmiştir.

Sosyal medyada insanların geçirdikleri zaman ile dolandırılma durumu arasında da doğrusal bir ilişki bulunmaktadır. Sosyal medyada günde sadece 1 saat zaman geçiren kişilerin, 4 saat ve üzeri zaman geçiren kişilere oranla daha az dolandırıcılık olayına maruz kaldığı tespit edilmiştir.

İnsanlar sosyal medyayı farklı amaçlarla kullanmaktadır. Bazıları aile ve arkadaşlarıyla görüşmek için bazıları ise sadece gündemi takip etmek için kullanmaktadır. Yapılan araştırmada, sosyal medyayı ticaret amaçlı kullanan kişilerin, diğer kişilere göre daha fazla dolandırıcılık olayına maruz kaldığı tespit edilmiştir. Sosyal medyayı her yaşta ve her eğitim düzeyinden insan kullanmaktadır. Eğitim düzeyi ve dolandırılma durumları karşılaştırıldığında ise, ilkokul ve lise mezunlarının lisans ve lisansüstü eğitim görmüş insanlara oranla daha fazla dolandırıcılık olayına maruz kaldığı tespit edilmiştir. Yaş ve dolandırılma durumları karşılaştırıldığında ise, 22 yaş altı kişilerin, 23 ve üzeri yaşa sahip olan kişilere göre çok daha fazla dolandırıcılık olayına maruz kaldığı görülmüştür.

Coğrafi bölgeler ile sosyal medya kullanımı üzerine yapılan araştırmada, yurt dışından katılan kullanıcılar ile Ege bölgesinden katılanların daha fazla dolandırıcılık olayına maruz kaldığı tespit edilmiştir. Kötü niyetli kişilerin açık kaynak kullanarak verilerini elde ettiği kişileri dolandırmasının daha kolay olduğu, bunun önüne geçmek için ise sosyal medya kullanım farkındalığının artırılmasına yönelik çalışmalar yapılması gerektiği gözlemlenmiştir.

Sosyal medya üzerinden açık kaynak istihbaratı analizi yapılarak, üç farklı konu hakkında çeşitli veriler elde edilmiştir. Bu veri setini oluşturan kullanıcıların attığı tüm tweetler analiz edilerek kullanıcının sahte veya gerçek kullanıcı sorgulamaları yapılmıştır. Bazı kullanıcıların, para ile satılan takipçiler olduğu, sadece belirli tweetleri yeniden tweetleyerek para ödeyen kullanıcıları popüler göstermek için kullandıkları tespit edilmiştir. Bu kullanıcıların tweetleri listeden temizlenerek, analizi yapılacak veri setinin doğru, bütün ve geçerli olması sağlanmıştır. Veri seti olarak dünyada en çok kullanılan telefon markaları üzerinden iPhone, Samsung ve Huawei etiketine atılan tweetler seçilmiştir. Veri seti, objektif ve subjektif olarak değerlendirilmiştir. Bu bulgu ile insanların Samsung markası konusunda daha objektif tweetler attığı görülmüştür.

5. SONUÇ VE ÖNERİLER

İnsanların sosyal medya kullanımını konusunda ciddi bir eğitime ihtiyaç duyduğu gözlemlenmiştir. Özellikle 22 yaş altı kullanıcıların sosyal medya kullanırken nelere dikkat etmesi gerektiğini bilmeden, daha gerçek hayatı ve gerçek hayattaki kötülükleri öğrenmeden sosyal medyada birçok insanla tanışıyor olmasının üzerine bir de sosyal medya farkındalık eğitimi olmadığı da eklenince, karşılaşacakları kötü durumlar hayatlarında ciddi sorunlara yol açmaktadır. Bu yaş grubunun Türkiye’de ilk ve orta öğretim ile lise öğrencileri olduğu düşünülürse, lise ve öncesi eğitim kurumlarında sosyal medya farkındalık eğitimleri verilmesi gerektiği sonucuna varılmıştır.

Türkiye’de sosyal medya kullanıcılarının;

- Hangi verileri paylaşmaları gerektiği,
- Hangi verilerin özel nitelikli veri olduğu ve paylaşmaması gerektiği,
- Paylaşımlarını kimlerin görebileceği,
- Popüler konulara katılımın hangi sınırlar içinde kalması gerektiği,
- Görüşmelerin gerçek hayata geçmesi aşamasında nelere dikkat etmesi gerektiği,
- Kullanma sürelerinin sınırlandırılması gerektiği,

konularında eğitime ihtiyaçları olduğu gözlemlenmiştir. Sosyal medya farkındalık eğitimi her yaş için ve her grup için önem taşımaktadır. Paylaşılan verilerin hiçbir yerden, hiçbir zaman silinemeyeceği düşünüldüğünde geri dönüşü olmayan verilerin paylaşılmaması ve kullanılmaması sağlanarak insanların zarar görmesi engellenmelidir.

Açık kaynak istihbaratı sonucu elde edilen veri setleri analiz edilip yorumlandığında ise, insanların markalar hakkında yaptıkları yorumların markaların satışlarıyla doğrusal bir orana sahip olduğu görülmüştür. 2018 yılı 2. çeyrek resmî satış sayıları göz önüne alındığında en çok satış yapan marka olan Samsung hakkında en fazla objektif yorumun yapıldığı gözlemlenmiştir. Bu yıl içerisinde en çok satış gerçekleştiren üç telefona yapılan objektif yorumların, subjektif yorumlardan oran bakımından daha fazla olduğu görülmüştür.



ÖZET

Açık Kaynak İstihbaratı: Twitter Analizi

Günümüzde sosyal medyanın önem kazanmasıyla birlikte elde edilebilecek açık kaynak bilgisi arttı. Gerek istihbarat kuruluşları gerekse insanlar açık kaynak bilgisinden faydalanarak önemli verilere ulaşmaya başladılar. Açık kaynak istihbaratının gerekli verileri elde ettiği ve sosyal medya olarak yaklaşık 267 milyon kullanıcının tercih ettiği Twitter önemli büyük veri kaynaklarından birisidir.

Hesapların sahte olup olmadığını belirleyen, verileri semantik açıdan değerlendiren, Twitter'daki etiketler ve trend konular üzerinden veri madenciliği yapan araçlar ve Elasticsearch ile bir veri seti oluşturulup analizler yapılarak açık kaynak istihbaratı adına önemli sonuçlara ulaşılabilmektedir.

Anahtar Sözcükler: Açık kaynak istihbaratı, Büyük veri, Elasticsearch, Etiket, Semantik, Trend konusu, Twitter, Veri madenciliği, Veri seti

SUMMARY

Open Source Intelligence: Twitter Analysis

Nowadays the open source information which will be able to be gotten has increased with social media gaining importance. Both intelligence services and humans started to reach the important data by exploiting open source information. Twitter which open source intelligence gets the essential data and almost 267 million users prefer as social media is one of the important big data sources.

Significant results can be achieved in the name of open source intelligence by being created a data set with tools which determine whether accounts are fake or not, evaluate data in the semantic way and perform data mining via hashtags and trend topics in Twitter and Elasticsearch and being made analyses.

Keywords: Big data, Data mining, Data set, Elasticsearch, Hashtag, Open source intelligence, Semantics, Trend topic, Twitter

KAYNAKLAR

Benavides M E B (2010). *Open Source Intelligence (OSINT) Link Directory Targeting Tomorrow's Terrorist Today (T4) through OSINT*. Erişim Adresi: [https://www.slideshare.net/Prohest/benavides-online-osint-quick-reference-handbook-new-table-of-contents?from_action=save]. Erişim Tarihi: [7/9/2019].

Benes L (2013). OSINT, new technologies, education: Expanding opportunities and threats. A new paradigm. *Journal of Strategic Security*, **6**(3), 22-37.

Benoit G (2002). University of Kentucky. *Annual Review of Information Science and Technology*, 265-310.

Cell C. Erişim Adresi: [https://stuff.co.za/cell-c-launches-believe-campaign-and-competition/]. Erişim Tarihi: [8/9/2019].

Chan J C (2014). The Role Of Social Media In Crisis Preparedness, Response And Recovery. Erişim Adresi: [https://www.oecd.org/governance/risk/The%20role%20of%20Social%20media%20in%20crisis%20preparedness,%20response%20and%20recovery.pdf]. Erişim Tarihi: [7/9/2019].

Focus On (2018). 100 Most Popular Social Philosophy, Semantics Bölümü.

Gibson S (2004). Open source intelligence: An intelligence lifeline. *The RUSI Journal*, **149**(1), 16-22.

Hassan N A, Hijazi R (2018). *Open Source Intelligence Methods and Tools*. Apress. Erişim Adresi: [http://cyber.sci-hub.se/MTAuMTAwNy85NzgtMS00ODQyLTMtMTMg==/10.1007%40978-1-4842-3213-2.pdf]. Erişim Tarihi: [7/9/2019].

Hilbert M (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, **34**(1), 135-174.

Keyhole. Erişim Adresi: [<https://keyhole.co/>]. Erişim Tarihi: [8/9/2019].

Luo Q. (2008, Ocak). Advancing knowledge discovery and data mining. *First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008)*, 3-5.

Mentionmapp. Erişim Adresi: [<https://mentionmapp.com/>]. Erişim Tarihi: [8/9/2019].

Naaman M, Becker H, Gravano L (2011). Hip and trendy: Characterizing emerging trends on Twitter. *Journal of the American Society for Information Science and Technology*, **62**(5), 902-918.

OpenStreetMap. Erişim Adresi: [<https://www.openstreetmap.org/about>]. Erişim Tarihi: [7/9/2019].

Otsuka E, Wallace S A, Chiu D (2014). Design and evaluation of a twitter hashtag recommendation system. *Proceedings of the 18th International Database Engineering & Applications Symposium*, 330-333.

Pew Research Center. 13 million Americans made donations online after Hurricanes Katrina and Rita. Erişim Adresi: [<https://www.pewinternet.org/2005/11/24/13-million-americans-made-donations-online-after-hurricanes-katrina-and-rita/>]. Erişim Tarihi: [7/9/2019].

RapidMiner. Erişim Adresi: [<https://rapidminer.com/>]. Erişim Tarihi: [8/9/2019].

Shah N, Willick D, Mago V (2018). A framework for social media data analytics using Elasticsearch and Kibana. *Wireless Networks*, 1-9.

Steele R D (2007). Open source intelligence. *Handbook of intelligence studies*, 42(5), 129-147.

Tinfoleak. Erişim Adresi: [<https://tinfoleak.com/>]. Erişim Tarihi: [8/9/2019].

Twitter. Erişim Adresi: [https://about.twitter.com/en_us.html]. Erişim Tarihi: [7/9/2019].

Ushahidi. Erişim Adresi: [<https://www.usahidi.com/about>]. Erişim Tarihi: [7/9/2019].

Van den Berg J A (2014). The story of the hashtag (#): A practical theological tracing of the hashtag (#) symbol on Twitter. *HTS Theological Studies*, **70**(1), 1-6.

White S. (2008). Turning Ideas into Action: Innovation within the Humanitarian Sector. *A think piece for the HFP Stakeholders Forum*. Erişim Adresi: [<http://www.humanitarianfutures.org/wp-content/uploads/2013/06/Briefing-Paper-Turning-Ideas-into-Action.pdf>]. Erişim Tarihi: [7/9/2019].

Zheng Y, Deng F, Zhu Q, Deng, Y (2014, Kasım). Cloud storage and search for mass spatio-temporal data through Proxmox VE and Elasticsearch cluster. *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, 470-474. IEEE.

Zhu J, Tirumala S S, Babu G A (2018, Nisan). A Technical Evaluation of Neo4j and Elasticsearch for Mining Twitter Data. *International Conference on Advances in Computing and Data Sciences*, 359-369. Springer, Singapore.

EKLER

EK 1 : Anket Formu



Bu anket "Açık Kaynak Verilerinin Kullanımı ve Analizi" isimli tez çalışmam için hazırlanmıştır.

Amacımız, insanların sosyal ağlarda zaman geçirirken neler ile karşılaştıklarına dair farkındalığı ölçmektir.

Sosyal ağlar (Twitter, Facebook, Instagram vb.) insanların yaşam tarzını büyük ölçüde değiştirmiştir.

Çalışmamız, bu değişikliklerin gerçek hayata nasıl yansıdığı, gerçek hayatta ne sorunlar çıkardığını ölçmek, bu sorunları veya endişeleri yaşayan insanların ortak özelliklerini, farklı durumlarını analiz etmek üzere oluşturulmuş sorulardan ibarettir.

Ankete katılım sağladığınızda verdiğiniz bilgilerin tamamı gizli tutulacak ve başka herhangi bir uygulama-araştırmada kullanılmayacaktır.

Anketi doldurduğunuzda katılımı "gönüllü olarak" onaylamış olacaksınız.

Katılımınız için teşekkür ederiz.

Çalışma ile ilgili sorularınız için "sfhzktl@gmail.com" adresine e-posta gönderebilir veya anket sonundaki formu doldurabilirsiniz.

Araştırmacı: Salih Zeki Atalı
Danışman: Dr. Öğr. Üye. Bülent Tuğrul
Adli Bilimler Enstitüsü
Ankara Üniversitesi
2019

1. Cinsiyetiniz
 Kadın Erkek
2. Hangi bölgede yaşıyorsunuz?
 Marmara İç Anadolu Güneydoğu Anadolu
 Ege Karadeniz Diğer (belirtiniz)
 Akdeniz Doğu Anadolu
3. Yaşınız
Yaş
4. Eğitim durumunuz
 İlkokul Üniversite
 Lise Yüksek lisans - doktora
5. Sosyal medyayı neden kullanıyorsunuz?
 Arkadaşlarla/Aileyle iletişim hâlinde kalmak için Satın alınacak ürünleri araştırmak/bulmak için
 Pek çok arkadaş/aile üyesi sosyal medyada takıldığı için Diğer insanlarla genel ağ oluşturmak için
 Zaman geçirmek için Yeni insanlarla tanışmak için
 Fotoğraf veya video paylaşmak için Gündemi takip etmek için
 Fikirlerini paylaşmak için İş veya ticari amaçlı
 Komik ve eğlenceli içerikler bulmak için Doğa/gezi
6. Günde zamanınızın ne kadarını sosyal medyada geçiriyorsunuz (ortalama)?
Saat
7. Paylaşım yaparken konumunuzu gösteriyor musunuz?
 Evet Hayır
8. Paylaşılan verilerin istihbarat örgütleri gibi çeşitli kuruluşlara satılabileceğini veya çeşitli kuruluşlarca kullanılabileceğini biliyor musunuz?
 Evet Hayır
9. 10 yıl önceki ve sonraki görünüşlerle ilgili olan "10yearchallenge" etiketini (hashtag) gördünüz mü ve kullandınız mı?
 Gördüm ve kullandım Gördüm ve kullanmadım Görmedim
10. Kişisel veri (sahibi olduğu kişiyi tanımlayan, özel ve genel bilgiler içeren fotoğraf, video vb. veri) paylaşıyor musunuz?
 Evet Hayır
11. Paylaştığınız verileri kimler görebiliyor?
 Paylaşım yapmıyorum
 Arkadaşlarım
 Herkes
12. Paylaşımınızdan sizinle ilgili hangi veriler elde edilebilir?
 Cinsiyet Siyasi Görüş Hobiler/Fobiler
 Eğitim Durumu Ekonomik Durum Özel Hayat
 Meslek Hayatı İnanç Hiçbir şey
13. Sosyal medyada tanıştığınız biriyle yüz yüze görüşmeden sohbet ettiniz mi?
 Evet Hayır

14. Sosyal ağlarda dolandırıldınız mı?

Evet

Hayır

15. Sosyal medyada gerçek isminizi ve/veya fotoğrafınızı kullanıyor musunuz?

Evet

Hayır

16. Popüler konular hakkında etiket (hashtag) kullanarak yorum yapıyor musunuz?

Evet

Hayır

17. Paylaştığınız bilgileri/verileri birisinin okuyup size karşı kullanmasından endişe duyuyor musunuz?

Evet

Hayır

18. Sizce sosyal ağlar güvenli mi?

Evet

Hayır

19. Dilek / öneri / geri bildirim yazmak için aşağıdaki formu doldurabilirsiniz.



ÖZGEÇMİŞ

I- Bireysel Bilgiler

Adı: Salih Zeki

Soyadı: ATALI

Doğum yeri ve tarihi: 09.03.1991/KÜTAHYA

Uyruk: Türkiye

Medeni durum: Bekâr

Askerlik durumu: Muaf

İletişim adresi ve telefon: salihzekiatali@gmail.com/90 (530) 308 26 47

II- Eğitim

Üniversite (Yüksek Lisans): Ankara Üniversitesi Adli Bilişim (2017 - 2020)

Üniversite (Lisans): Kocaeli Üniversitesi Elektronik ve Haberleşme Mühendisliği
(2009 – 2013)

Lise: Süleyman Demirel Anadolu Lisesi (2005 – 2009)

Yabancı Dil: İngilizce (Profesyonel Çalışma Yetkinliği) (Yurt dışı)

III- Mesleki Deneyim

BT UZMANI - SARGEM (08/13 – 02/14)

Kullanıcıların internet, yazıcı, bilgisayar problemleriyle ilgilenildi.

NETWORK MÜHENDİSİ – ANKARA ÜNİVERSİTESİ (02/14 – 07/14)

Ankara Üniversitesi Gölbaşı Kampüsünde bulunan kullanıcıların ağ problemleriyle ilgilenildi.

BT UZMANI - SARGEM (07/14 – 10/16)

Kullanıcıların internet, yazıcı, bilgisayar problemleriyle ilgilenildi.

AĞ GÜVENLİK UZMANI – KOZA ALTIN İŞLETMELERİ A.Ş. (10/16 – 02/17)

Şirket içerisinde güvenli haberleşme yapılabilmesi için Cisco e-Token kullanılarak CUCM 7.1.5 vasıtasıyla telefon görüşmeleri şifrelendi. Loglar SIEM aracılığıyla incelendi. Kullanıcıların ağ problemleriyle ilgilenildi.

AĞ GÜVENLİK UZMANI – SİNCAN BELEDİYESİ (08/17 -)

Kullanıcıların ağ problemleriyle ilgileniliyor. Zabbix programıyla ağ izleniyor. Loglar SIEM aracılığıyla inceleniyor. Sunucuların günlük yedekleri alınıyor. Belediyedeki tüm birimler VLAN'lara ayrıldı. ISO 27001 uyumluluk süreci koordinatörlüğü yapılıyor.

Stajyer Mühendis : Ankara Büyükşehir Belediyesi Bilgi İşlem bölümünde staj yapıldı. (1 ay)

Stajyer Mühendis : TAI Elektronik ve Haberleşme Bölümünde staj yapıldı. (1 ay)

IV- Bilimsel Etkinlikler

Projeler:

1) İşaret İşleme

FFT, DFT, DTFT, FIR-IIR filtre tasarımı, windowing (Hamming, Hanning, Kaiser, Blackman) uygulamaları yapıldı. (MATLAB)

2) C/C++ Uygulamaları

3) Görüntü İşleme

Lokal kontrasta dayalı füzyon yöntemi sayesinde orijinal görüntüyle iyileştirilmiş görüntü birleştirildi. (MATLAB)

4) Lisans Bitirme Tezi

Matlab ile Fiber Optik Dalga Kılavuzlarında Elektromanyetik Dalgaların İncelenmesi (Kocaeli Üniversitesi, 2013)

Dielektrik dilim dalga kılavuzlarında TE karakteristik denklemlerine göre yayılım modunu etkileyen fiziksel parametreler incelendi. Eikonal denklem ile ışığın dalga kılavuzlarında yayılımını incelendi. Değişen indisli dalga kılavuzları incelendi.

5) Yüksek Lisans Semineri

Açık Kaynak İstihbaratı

V- Dięer Bilgiler

Kurslar:

CCNA (28.04.2014 – 19.06.2014) (70 Saat)

CCNP (27.09.2014 – 21.02.2015) (105 Saat)

CCNA Security (02.03.2015 - 06.03.2015) (35 Saat)

CEH (02.10.2017 – 06.10.2017) (35 Saat)

