# VERİ MADENCİLİĞİ TEKNİKLERİ İLE AÇIK KAYNAK İSTİHBARATI: SOSYAL MEDYA İLE BBC, CNN, AL JAZEERE VE REUTERS HABER AJANSLARINA AİT GÜNDEMİN MUKAYESESİ[1]

**Emre Cihan ATEŞ**
Dr.
Jandarma Genel Komutanlığı
emre_cihan_ates@hotmail.com

**Erkan BOSTANCI**
Doç.Dr.
Ankara Üniversitesi
Bilgisayar Müh.Böl.
ebostanci@ankara.edu.tr

**Mehmet Serdar GÜZEL**
Doç.Dr.
Ankara Üniversitesi
Bilgisayar Müh.Böl.
mguzel@ankara.edu.tr

**Özet**

Sosyal medya, günümüzde insanların çeşitli alanlarda kendilerine özgü duygu ve düşünceleri sanal bir platform üzerinde özgürce paylaştıkları yegâne platformdur. Bu sebeple özellikle son 10 yılda toplumda hemen hemen herkesin fiziksel hayattaki somut kişilik özelliklerine ilave olarak, sanal platformlarda oluşturduğu sanal kişilikler mevcuttur. Söz konusu sanal platformlarda birçok konuda anlık olarak görüşler paylaşılmakta, söz konusu görüşler "trend" olarak popülerlik kazanmasıyla birlikte tüm toplum için yönlendirici veya birleştirici olabilmektedir. Sanal kişilikler üzerinden, bireyler takipçilerine çeşitli konularda mesajlar yayınlamakta ve yayınlanan mesajlar, kimi takipçilerce beğenilmekte veya tekrar paylaşılmaktadır. Sosyal medya şirketleri ise olumlu tepki alan paylaşımları bazı algoritmalar yardımıyla benzer yönelimde olan kullanıcılara önermesi ile bireylerin attığı mesajlar, toplumsal kitle tepkisine dönüşebilmektedir. Yapılan paylaşımlar, elbette çoğu konuda toplumda olumlu gelişmelerin yol açmasına sebep olmakla birlikte, kime zaman suçu yönlendirici veya ihtiva edici olabilmekte, bu durum da açık kaynak temelli istihbarat elde etmenin önemini vurgulamaktadır.

Sosyal medya kullanımının çok yoğun olduğu günümüzde açık kaynak temelli istihbarat elde etmek ancak veri madenciliği teknikleri ile yapılacak analizler ile mümkün olabilmektedir. Bu çalışmada sosyal medyadan elde edilen veriler ile haber ajanslarının paylaşımlarının mukayese edilerek, ne kadar benzerlik gösterdiğinin belirlenmesi amaçlanmıştır. Yapılan bu araştırmadaki veriler, kullanıcıların herkese açık daha fazla paylaşım yaptığı ve verileri elde etmede kaynak

[1] Bu çalışma "Önleyici Kolluk Faaliyetleri Kapsamında Sosyal Medyada Veri Madenciliği: Suçları Önlemek İçin Twitter Analytic'i Kullanma" isimli doktora tez çalışmasından faydalanılarak hazırlanmıştır.

kod kullanılmasına imkân tanıyan "Twitter" isimli sosyal ağdan 04 Eylül - 04 Aralık 2019 tarihleri arasını ihtiva edecek şekilde, İngilizce dil süzgecinde "#" simgesi içeren 4.030.945 adet tweet Twitter API aracılığıyla Python ile çekilmiştir. Sosyal ağlarda "#" etiketiyle konuşulan konular ile aynı tarihlerde "BBC News (World)", "Reuters", "Al Jazeere News", "BBC Breaking News", "CNN International" ve "CNN"'e ait haber paylaşımlarının tamamının birleşimi üzerinden TF-IDF kullanılarak vektörel benzerlik analizi yapılmıştır. Araştırmada yalnızca Twitter API üzerinden metin madenciliği çalışması ile elde edilen birincil veriler kullanılmıştır. Metin madenciliği işleminde, öncelikle ham veriler kullanılabilir bir formata dönüştürülmüş ve müteakibinde veriler üzerinde temizlik yaparak, elde edilen veriler ön işleme tabi tutulmuştur. Çalışmanın yukarıda belirtilen tarih aralığında olması ile Twitter özelinde yapılması çalışmanın kısıdını oluşturmaktadır.

Çalışma sonucunda Twitter içindeki paylaşımların %18 ile %51 oranları arasında haber ajanslarına ait paylaşımlarla benzerlik gösterdiği ve en fazla benzerliğin %51 ile BBC News (World) ile olduğu belirlenmiştir. Haber ajanslarından elde edilen bilgi, insanların sosyal medyada yazdıklarının ne kadarını karşıladığı tespit edilmiş olup, yapılan tespitte sosyal medyanın istihbarat için yadsınamaz öneme haiz olduğu belirlenmiştir. Zaman zaman sosyal medya da yalan haberlerin bile dolaştığı bilinmekle birlikte, paylaşımın kendisi yalan bile olsa bunun açık kaynak istihbaratı açısından kolluk birimleri tarafından tespiti önem arz etmektedir. Yapılan bu örneklem çalışmadan yola çıkarak, sosyal medyadaki hareketliliğin güvenlik güçleri ve araştırmacılar tarafından incelenerek, bilgi çıkarımı yapılmasının suç sorununun kökeninin belirlenmesinde ve yapılacak suçu önleyici çalışmalarda fayda sağlayacağı değerlendirilmektedir.

**Anahtar Kelimeler:** Veri Madenciliği, Açık Kaynak İstihbaratı, Twitter

## 1. GİRİŞ

Sosyal medya, çeşitli konularda bilgi edinmek ve bilginin yayılımı için özellikle son yıllarda önemli bir iletişim platformu haline gelmiş ve televizyon, gazete, radyo vb. gibi birçok geleneksel iletişim araçlarına göre popülerliği oldukça yükselmiştir. Sosyal medyayı geleneksel iletişim araçlarından daha popüler yapan en önemli konu, kişi ya da grupların fikirlerini özgürce paylaşabilmeleri ve başkalarıyla çevrimiçi bağlantı kurabilmeleridir.

Sosyal medya platformu üzerinden kullanıcılar, kendilerini takipçilerine tanıtacak bir dizi mesajlar yayınlamaktadır. Söz konusu mesajlar, takipçiler tarafından beğenilmekte ve tekrar paylaşılabilmektedir. Beğeni yapıldığı taktirde, çeşitli algoritmalar yardımıyla benzer yönelim

gösteren kullanıcılara önerilmekte olup, tekrar paylaşıldığı taktirde ise başka kullanıcılar tarafından da görülebilmekte ve sosyal ağ içinde hızla yayılmaktadır. Sosyal medyanın ana amacı olan iletişim kurabilme yeteneğinin temeli, paylaşımların diğer insanlar üzerinde ilgi görmesidir. Çünkü ilgi gören yani trend olarak nitelenen mesajlar, kullanıcılar tarafından okunmakta, beğenilmekte, yorum yapılmakta veya tekrar paylaşılmaktadır.

Sosyal medyadaki paylaşımlardan yola çıkarak sosyal ağ içindeki hareketliliğin tespiti ve veri madenciliği teknikleri ile analizi açık kaynak istihbaratı anlamında toplumsal hareketlilik ile ilgili fikir verebilmektedir. Bu çalışmada sosyal medyadan # etiketiyle çekilen paylaşımlar ile BBC News (World), Reuters, Al Jazeere News, BBC Breaking News, CNN International ve CNN'e ait paylaşımların metin ön işleme teknikleri de kullanılarak, vektörel düzeyde birbirine ne kadar benzerlik gösterdiğinin ortaya konarak, sosyal medyanın geleneksel ajans paylaşımlarından ne kadar ayrıştığının ve açık kaynak temelindeki önemi ortaya konulmak istenmiştir.

Çalışma kapsamında giriş kısmına müteakip, sırasıyla veri madenciliği, doğal dil işleme, sosyal medyada bilginin yayılımı, açık kaynak istihbaratı kavramlarının tanımlandığı ve bazı literatürdeki çalışmaların incelendiği kavramsal çerçeve hazırlanmıştır. Metodoloji kısmında verilerin elde edilmesi, ön işleme, benzerlik analizi ile ilgili yöntem bilgisi verilmiştir. Bulgular kısmında elde edilen sonuçlar paylaşılmış olup, tartışma ve sonuç bölümünde ise elde edilebilecek bilgilerin kullanılabilirlik durumu ve önemi ortaya konulmaya çalışılmıştır.

## 2. KAVRAMSAL ÇERÇEVE

Kavramsal çerçeve başlığı altında veri madenciliği, doğal dil işleme, sosyal medya ve açık kaynak istihbaratı kavramları tanımlanacaktır.

### 2.1. Veri Madenciliği

Veri madenciliği, analiz yoluyla geçmiş verileri açıklayarak, ondan dersler çıkarma ve geleceğe ilişkin çıkarımlarda bulunma çabası olup, literatürde veri madenciliğinin tanımına ilişkin karşılaşılan bazı önemli tanımları aşağıda verilmiştir;

- Önceden bilinmeyen verilerle ilgili potansiyel olarak yararlı bilgileri gizli olmayan bir şekilde ortaya çıkarmak (Srinivas ve ark., 2010),
- İstatistikleri; kavramlar, araçlar, algoritmalar, makine öğrenimi ve çok büyük veri setlerinin analizi ile birleştirerek anlama ve eyleme dönüştürülebilir bilginin elde edilmesini sağlamak (Williams, 2009).

- Büyük miktarda veriden bir bilgi alma veya madencilik çıkarma süreci olarak tanımlanabilen tüm bilgi keşif sürecinde bir adımdır ve özellikle belirli bir alandaki sorunları çözmek için gerekli olan bir bilgi keşfi biçimidir (Beniwal ve Arora, 2012).

Bu tanımlardan yola çıkarak, veri madenciliğinin diğer alanlarla ilişkisini de dikkate alarak şu tanımı yapmak mümkündür: Veri madenciliği, geçmişte büyük veriden yararlanılarak, istatistik, makine öğrenimi, yapay zekâ gibi yöntemler kullanılarak bilgiyi elde etmeyi amaçlayan bir süreçtir ve genellikle bu bilgiyi gelecekle ilgili karar destek faaliyetlerinde kullanmaktır (Campbell ve Ying, 2011).

### 2.2. Doğal Dil İşleme

Dünya üzerinde yer alan verilerin % 90'ı yapılandırılmamış şekilde olup, söz konusu veriler; resim, video ve metin gibi birçok farklı biçimde olabilmektedir (Kulkarni ve Shivananda, 2019). Söz konusu verilerin yapılandırılmamış konumda olması, veri gürültüsü (noisy data) sorununu da beraberinde getirmektedir. Bu konuda veri dünyasında sıkça söylenilen "Garbage in, garbage out (GIGO)" özdeyişinde olduğu üzere, düşük kalitede yer alan verilerle yapılan algoritmik öğrenmeler de düşük düzeyde olacaktır. Örneğin "selllam dstum, umarim iyisindir." cümle yapısını aynı dili konuşan bir insanın dahi anlaması kimi zaman mümkün olmamaktadır. Veri madenciliğinin başarılı şekilde yapılabilmesi için en önemli adımlardan biri veri üzerinde yapılan temizlik ile verinin içeriği ve yapısı gibi temel etmenlerinin ortaya konulmasıdır, farklı bir ifadeyle mümkün olduğunca gürültüden arındırılmasıdır. Makine öğrenimi veya derin öğrenme algoritmaları ile metin üzerinde işlem gerçekleştirmek için öncelikle metin verilerinin işlenmesi ve yorumlamasının yapılması gerekir. İster denetimli veya yarı denetimli, ister denetimsiz olsun, tüm makine öğrenimi algoritmaları sayısal olan giriş özellikleri ile çalışır. Metin verilerini sayısallaştırmadan önce ilk olarak metin verilerinin temizlenmesi, normalleştirilmesi ve önişleme tabi tutulması gerekmektedir.

### 2.3. Sosyal Medya

Günümüzde sosyal ağların hızlı gelişimi ile birlikte, internet kullanımının kapsamı oldukça genişlemiştir. Her geçen gün insanların duygularını ifade etme ve diğer insanlarla bir şeyler paylaşma gibi amaçlarla sosyal medyada sanal bir kişilik oluşturduğu dönemdeyiz (Mulyadi ve Fitriana, 2018). Genelde sosyal medya sitelerine üyelikle başlayan sanal kişilik oluşumu sonrasında, metin, ses, resim ve video gibi veri türlerinin transferi yapılabilmektedir. Veri transferinin yanı sıra sosyal medya, insanların birbirlerine mesaj, durum veya yorum gibi mesajları da atabildiği bir iletişim ortamıdır. Sosyal medyanın kullanımı iletişim başta olmak

üzere pek çok disiplin için devrim niteliği taşımaktadır. Sosyal medya, insanlar arasındaki fiziksel iletişimi sanal boyuta taşıyan dijital ağ ortamıdır.

### 2.4. Sosyal Medyada Açık Kaynak İstihbaratı

Kullanımı için özel bir kısıtlama olmaksızın, kişilerin herkese açık olarak yaptığı paylaşımların diğer bütün insanlar gibi kolluk kuvvetleri tarafından da kullanılması sosyal medyada açık kaynak istihbaratını oluşturmaktadır (Bartlett ve ark., 2013).

Wearesocial (2020) üzerinden ülkemiz özelinde elde edilen veriler incelendiğinde;

- Sosyal medya hesaplarını nüfusun %64'ünün kullandığı (Tüm yaşlar 0- ∞),
- Günde yaklaşık ortalama 7 saat 29 dakika internetin kullanıldığı (4 saati telefon, 3 saat 29 dakikası bilgisayar, tablet, laptop vb.) ve söz konusu sürenin ortalama 2 saat 51 dakikasında sosyal ağların kullanıldığı,
- Bazısı aktif şekilde kullanmasa da ülkemizde kişi başına düşen sosyal ağ hesabı sayısının 9.1 olduğu,
- Sosyal ağların kullananların %44'ünün iş amacı için bireysel kullanım olduğu,
- 37.000.000 Facebook kullanıcısı olduğu ve yapılan paylaşımların kişi başı ayda ortalama 3 yorum aldığı,
- 38.000.000 Instagram kullanıcısı olduğu,
- 11.800.000 Twitter kullanıcısı olduğu,
- 8.400.000 Linkedin kullanıcısı olduğu,
- 7.700.000 Snapchat kullanıcısı olduğu belirlenmiştir.

Tüm yaş grubu ortalamasında %64 kullanım oranı, kişi başı 9.1 hesap ve günlük ortalama 2 saat 51 dakikanın sosyal medyada geçirildiği ülkemiz, açık kaynak olarak devasa bir veri tabanı özelliği taşımaktadır. Kolluk kuvvetleri genel olarak iki şekilde bilgi temininde sosyal ağları kullanabilmektedir; bunlardan ilki gizlilik ihtiva etmeyen suç verilerinin yayılımına fırsat sağlayarak, sosyal medyanın popülaritesi ve kolay erişimi nedeniyle potansiyel suç veya suça ait bulguları toplamayı amaçlar (Brandl, 2019). Fakat sosyal medyada bilgilerin paylaşımı sonrası, veri üzerinde kontrolün kaybolması ve üzerine yapılması muhtemel sahte haberler risk unsurudur. İkincisi ise, kolluk kuvvetlerinin aradığı bir konuda elde ettiği bilgiyi ilgili siteden çekmesidir (Brandl, 2019). Bu yaklaşım herkese açık olarak yapılan paylaşımları ihtiva eder. Kullanıcın gizlilik ayarlarına bağlı olarak; ilgi alanı, mesleği, adresi, iletişim bilgileri, arkadaşları, mesajları, fotoğrafı, ses kaydı, yorumu, konumu ve videoları gibi bilgilere erişim sağlanabilmektedir. Organize grupların (çete, mafya, örgüt vb.) üyeleriyle yaptığı iletişimdeki örüntüler, kullanılan şifreli kelimeler, kelime ağı, anomali hareketler ve olası kullanılan özel

kelime ve emojiler bulunabilmektedir (Cox ve ark., 2017). Açık kaynak istihbaratı suç örgülerini çözmede yardımcı pozisyonda olup, sosyal medya vasıtalarıyla bilgi almak, soruşturmaya ilişkin ipuçları verebilmektedir.

## 3. METODOLOJİ

Bu çalışmada, Twitter örnekleminde hashtag ifadesiyle birlikte yer alan ifadelerin metin madenciliği temelinde doğal dil işleme yöntemleri ile analiz edilerek, açık kaynak istihbaratı açısından değerlendirilmesi metot olarak amaçlanmıştır. Çalışma kapsamında çeşitli haber ajanslarından elde edilen haber paylaşımları, sosyal medyada hashtag altında konuşulan konularla benzerlik analizi yapılmıştır.

Yapılan bu araştırmadaki veriler, kullanıcıların herkese açık daha fazla paylaşım yaptığı ve verileri elde etmede kaynak kod kullanılmasına imkân tanıyan "Twitter" isimli sosyal ağdan 04 Eylül - 04 Aralık 2019 tarihleri arasını ihtiva edecek şekilde, İngilizce dil süzgecinde "#" simgesi içeren 4.030.945 adet tweet Twitter API aracılığıyla Python ile çekilmiştir. Sosyal ağlarda "#" etiketiyle konuşulan konular ile aynı tarihlerde "BBC News (World)", "Reuters", "Al Jazeere News", "BBC Breaking News", "CNN International" ve "CNN"'e ait haber paylaşımlarının tamamının birleşimi üzerinden TF-IDF kullanılarak vektörel benzerlik analizi yapılmıştır. Araştırmada yalnızca Twitter API üzerinden metin madenciliği çalışması ile elde edilen birincil veriler kullanılmıştır. Metin madenciliği işleminde, öncelikle ham veriler kullanılabilir bir formata dönüştürülmüş ve müteakibinde veriler üzerinde temizlik yaparak, elde edilen veriler ön işleme tabi tutulmuştur.

Veri ön işlemede, veri seti içerisinde yer alan metinlerin ortak bir forma dönüştürülmesi amaçlanmıştır. Bu sebepler tweetler küçük harfe dönüştürülerek, aşağıdaki gereksiz içerikler metinden kaldırılmıştır. Metinler içindeki boşluk, URL adres, noktalama, numara/özel karakter, ve durağan kelimeler temizlendikten sonra stemming ve lemmatization işlemi TextBlob kütüphanesi aracılığıyla uygulanmıştır.

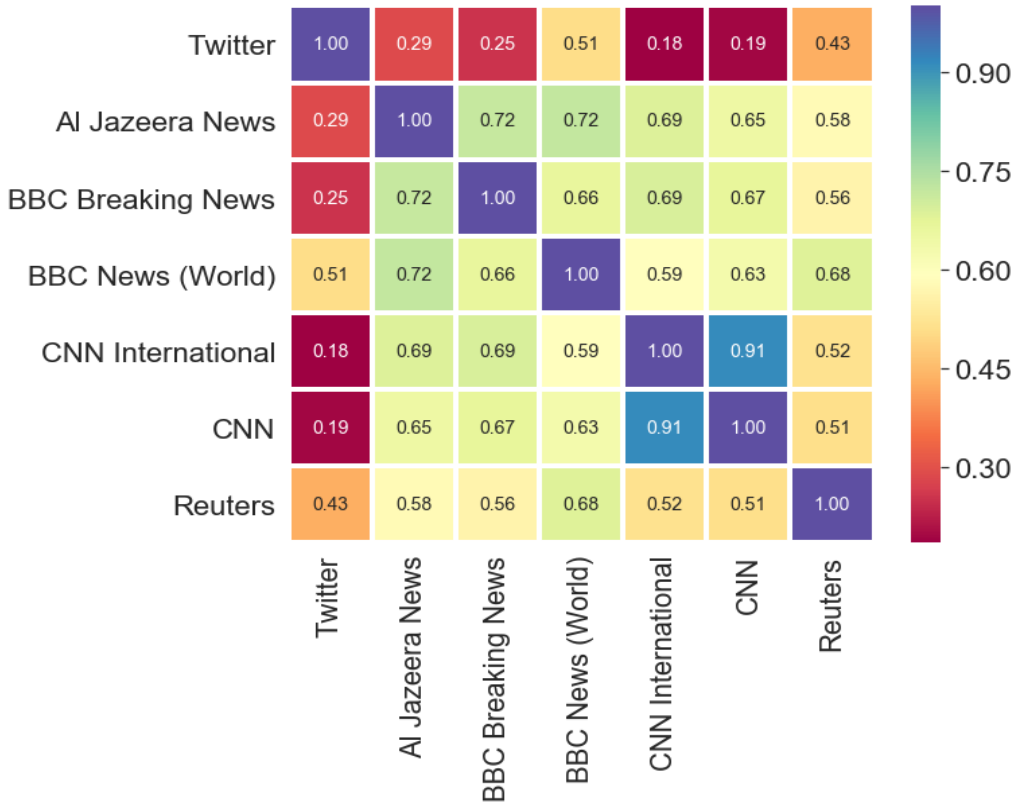
Benzerlik ölçümü sözcüksel düzeyde, korpus temelli yaklaşımla yapılmıştır. Metin benzerliğini ortaya koymak için hamming mesafesi, kosinüs mesafesi, öklid mesafesi, manhattan mesafesi, Jaccard benzerliği, BM25 benzerliği, TF-IDF, Levenshtein mesafesi ve n-gram gibi yöntemler kullanılmakta olup, en çok kullanılanı ise kosinüs ve TF-IDF yöntemleridir (Sarkar, 2019). Bu çalışmada TF-IDF doküman benzerliğinde terim frekansı baz alınmıştır.

Çalışma kapsamında Python 3.8 kullanılmış olup, tüm çalışmalar 12 GB RAM ile destekli Windows PC kullanılarak gerçekleştirilmiştir. Kullanılan algoritmalar, Pycharm kullanılarak,

Python'da programlanmıştır. Parametre değerlerinin belirlenmesinde her bir algoritma için hiperparametre optimizasyonu yapılmıştır.

## 4. BULGULAR

Twitter örnekleminde sosyal medyada # etiketi ile yapılan paylaşımlar ile aynı tarihler arasında "Al Jazeere News", "BBC Breaking News", "BBC News (World)", "CNN International", "CNN" ve "Reuters" tarafından yapılan paylaşımlar kendi içinde birer korpusta toplanmış ve TF-IDF kullanılarak benzerlik durumlarının ortaya konması amaçlanmıştır.




**Şekil 1.** Benzerlik Matrisi

Şekil 1 incelendiğinde;

– Twitter üzerinden konuşulan konulara en benzer haber paylaşımlarının %51 BBC News (World) ve müteakibinde ise %43 ile "Reuters" olduğu,

– En yüksek benzerliğin olduğu %51'lik oran üzerinden konuşulsa dahi sosyal ağlarda konuşulan konuların neredeyse yarısının haber metinlerinde yer almadığı ve bu sebeple de sosyal medyadan haber temini için kullanılabileceği,

– Haber paylaşımları arasında en yüksek benzerliğin %91 ile "CNN International" ve "CNN" arasında olduğu,

– Haber paylaşımları arasında en düşük benzerliğin %51 ile "Reuters" ve "CNN" arasında olduğu anlaşılmaktadır.

## 5. TARTIŞMA VE SONUÇ

Son yılarda sosyal medya, çeşitli konularda bilgi edinmek ve yaymak için önemli bir iletişim platformu haline gelmiştir. Sosyal medyayı geleneksel iletişim araçlarından daha popüler yapan en önemli konu, kişi ya da grupların fikirlerini özgürce paylaşabilmeleri ve başkalarıyla çevrimiçi bağlantı kurabilmeleri olmuştur (Ahmed ve ark., 2019).

Sosyal medya platformu üzerinden kullanıcılar, kendilerini takipçilerine tanıtacak bir dizi mesajlar yayınlamaktadır. Söz konusu mesajlar, takipçiler tarafından beğenilmekte ve tekrar paylaşılabilmektedir. Beğeni yapıldığı taktirde, çeşitli algoritmalar yardımıyla benzer yönelim gösteren kullanıcılara önerilmekte olup, tekrar paylaşıldığı taktirde ise başka kullanıcılar tarafından da görülebilmekte ve sosyal ağ içinde hızla yayılmaktadır. Sosyal ağlar içindeki iletişim kişisel ağlar içinde olabileceği gibi, fiziksel mesafeler ve asimetrik programlar nedeniyle çevrimdışı etkileşimlerin imkânsız olacağı ve daha önceden hiçbir bağı olmayan bireyler de olabilmektedir (Barberá, 2015).

Fiziksel ortamda bir araya gelme olma ihtimali olmayan insanların bile etkileşim kurduğu sosyal medyada örneklem olarak;

- Comunello ve Anzera (2012) tarafından yapılan çalışmada arap baharı olayları,
- Karataş ve Saka (2017) tarafından yapılan çalışmada ise gezi olayları,
- Bernard ve ark. (2018) tarafından yapılan çalışmada bulaşıcı hastalıklarının yayılması,
- Jones ve Mattiacci (2019) tarafından yapılan çalışmada sosyal medyadaki isyan diplomasisi,
- Elmahjub (2013) tarafından yapılan çalışmada ise Libya'da yaşanan iç savaşta sosyal medya açık kaynak istihbaratı açısından incelenmiştir.

Sosyal medyanın gündelik vatandaşlar tarafından sağlanan yorum ve gözlemlerin varlığı nedeniyle "zayıf sinyalleri" tanımlamak için bir kaynak olarak kullanıldığı ve sosyal medyadaki gönderilerin haber kaynakları, gazeteciler ve diğer ilgi gruplarından toplanan verilerle karşılaştırıldığında özellikle daha nadir olduğu Andrews ve ark. (2018) tarafından yapılan çalışmada ortaya konulmuştur.

Bu çalışma kapsamında da sosyal medyadaki gönderilerin haber kaynakları ile karşılaştırıldığında daha nadir olduğu tespitine istinaden "Al Jazeere News", "BBC Breaking News", "BBC News (World)", "CNN International", "CNN" ve "Reuters" haber ajanslarının paylaşımları ile sosyal medyada # etiketiyle konuşulan diğer konular mukayese edilmiştir. Elde edilen sonuçlarda en fazla benzerliğin %51 ile BBC News (World)'de olduğu, diğer sonuçların

oldukça düşük olduğu gözlemlenmiştir. Bu durum da istihbarı açıdan önemi tartışmalı olsa da açık kaynak istihbaratı yapılarak, nadir bilgilere erişebileceğinin göstergesi konumundadır. Haber ajanslarından elde edilen bilgi, insanların sosyal medyada yazdıklarının ne kadarını karşıladığı belirlenmiş olup, yapılan tespitte sosyal medyanın istihbarat için yadsınamaz öneme haiz olduğu belirlenmiştir. Araştırma sonucu elde edilen veriler, sosyal ağda suçla mücadeleye dair örneklem seviyede bilgi vermekte olup, yapılan tespitlerin suç ve suçla mücadele açısından yürütülen önleyici ve adli hizmetlerdeki hareket tarzları için fikir verdiği değerlendirilmektedir.

**Kaynakça**

Ahmed, Y. A., Ahmad, M. N., Ahmad, N., & Zakaria, N. H. (2019). Social media for knowledge-sharing: A systematic literature review. Telematics and informatics, 37, 72-112.

Andrews, S., Brewster, B., & Day, T. (2018). Organised crime and social media: a system for detecting, corroborating and visualising weak signals of organised crime online. Security Informatics, 7(1), 1-21.

Barberá, P. (2015). How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. Job Market Paper, New York University, 46.

Bartlett, J., Miller, C., Crump, J., & Middleton, L. (2013). Policing in an information age.

Beniwal, S., & Arora, J. (2012). Classification and feature selection techniques in data mining. *International journal of engineering research & technology (ijert)*, *1*(6), 1-6.

Bernard, R., Bowsher, G., Milner, C., Boyle, P., Patel, P., & Sullivan, R. (2018). Intelligence and global health: assessing the role of open source and social media intelligence analysis in infectious disease outbreaks. Journal of Public Health, 26(5), 509-514.

Brandl, S. G. (2019). Criminal investigation. SAGE Publications.

Campbell, C., & Ying, Y. (2011). Learning with support vector machines. *Synthesis lectures on artificial intelligence and machine learning*, *5*(1), 1-95.

Comunello, F., & Anzera, G. (2012). Will the revolution be tweeted? A conceptual framework for understanding the social media and the Arab Spring. *Islam and Christian–Muslim Relations*, *23*(4), 453-470.

Cox, S. M., Massey, D., Koski, C. M., & Fitch, B. D. (2017). *Introduction to policing*. Sage Publications.

Elmahjub, E. M. J. (2013). Facebook versus Ghaddafi: social networking as a tool for democratic change in Libya. *Space, Place and Culture*, 1-18.

Jones, B. T., & Mattiacci, E. (2019). A manifesto, in 140 characters or fewer: Social media as a tool of rebel diplomacy. British Journal of Political Science, 49(2), 739-761.

Karatas, D., & Saka, E. (2017). Online political trolling in the context of post-Gezi social media in Turkey. International Journal of Digital Television, 8(3), 383-401.

Kulkarni, A., & Shivananda, A. (2019). Natural language processing recipes. Apress.

Mulyadi, U., & Fitriana, L. (2018). Hashtag (#) as message identity in virtual community. Jurnal The Messenger, 10(1), 44-53.

Sarkar, D. (2019). Text analytics with Python: a practitioner's guide to natural language processing. Apress.

Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE)*, *2*(02), 250-255.

Wearesocial (2020). Digital 2020: Global digital yearbook, July Global Statshot Report. Erişim Adresi: [https://wearesocial.com/digital-2020] Erişim Tarihi:20/10/2021.

Williams, G. (2009). Rattle: a data mining GUI for R.