



Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

6 Şubat 2023 Türkiye Depremi Hakkındaki Twitter Paylaşımlarının
Metin Madenciliği Yöntemleri ile İncelenmesi

Esmagül DENİZ

DANIŞMAN

Doç. Dr. Buket DOĞAN

İSTANBUL, 2024



Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

6 Şubat 2023 Türkiye Depremi Hakkındaki Twitter
Paylaşımlarının Metin Madenciliği Yöntemleri ile İncelenmesi

Esmagül DENİZ

170420036

DANIŞMAN

Doç. Dr. Buket DOĞAN

İSTANBUL, 2024

MARMARA ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği Öğrencisi Esmagül Deniz'in "6 Şubat 2023 Türkiye Depremi Hakkındaki Twitter Paylaşımlarının Metin Madenciliği Yöntemleri ile İncelenmesi" başlıklı bitirme projesi çalışması, 05/07/2024 tarihinde sunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

Jüri Üyeleri

Doç. Dr. Buket DOĞAN (Danışman)

Marmara Üniversitesi

Prof. Dr. Ali BULDU (Üye)

Marmara Üniversitesi

Dr. Öğr. Üyesi Gözde KARATAŞ BAYDOĞMUŞ (Üye)

Marmara Üniversitesi

İÇİNDEKİLER

ÖZET	6
ABSTRACT.....	7
SEMBOLLER LİSTESİ.....	8
KISALTMALAR LİSTESİ.....	9
ŞEKİL LİSTESİ.....	10
TABLO LİSTESİ.....	11
1. GİRİŞ.....	12
1.1. İlgili Çalışmalar.....	15
1.2. Amaç.....	20
2. MATERYAL VE YÖNTEM.....	21
2.1. Projede Kullanılan Araçlar.....	22
2.2. Veri Seti.....	23
2.3. Metin Madenciliği ile Ön İşleme.....	24
2.4. Veri Görselleştirme.....	26
2.5. Projede Yapılan Analizler için Kullanılan Yöntemler.....	26
2.6. Kısıtlar.....	31

3.	BULGULAR	31
3.1.	Hashtag Analizi.....	31
3.2.	Günlük Tweet Yoğunluk Analizi.....	33
3.3.	Dil Analizi.....	33
3.4.	11 İl için Hasar Tweet Yoğunluğu Analizi.....	34
3.5.	Mention/Kullanıcı Adı Analizi.....	35
3.6.	Tweet İçeriklerinde En Sık Kullanılan Kelimeler ve Kelime Grupları.....	37
3.7.	Tweet İçeriklerinde URL Kullanım Yoğunluğu Analizi.....	38
4.	SONUÇLAR	39
5.	KAYNAKÇA.....	43
6.	ÖZGEÇMİŞ.....	45

ÖZET

6 Şubat 2023 tarihinde Türkiye’de büyük bir deprem felaketi yaşanmıştır. Depremi merkez üssü olan bölgeler başta olmak üzere, tüm ülkeyi etkileyen bu olay sonrasında sosyal medya paylaşımlarının ana konusu deprem olmuştur. Depremi yarattığı yıkım ve kayıplar, insanlar üzerinde derin izler bırakmış ve sosyal medya araçları bu duyguların ve düşüncelerin ifade edildiği birer mecra haline gelmiştir. İnsanlar deprem sonrası yaşadıkları deneyimleri, hissettikleri duyguları ve gözlemlerini sosyal medya üzerinden paylaşarak geniş bir kitleye ulaşımlardır. Bu durum sosyal medya paylaşımlarının analiz edilmesiyle, toplumsal olayların anlaşılması ve değerlendirilmesi açısından önemli bir veri kaynağı olarak kullanılabileceğini göstermektedir.

Twitter, her konuya ilişkin tepkilerin, bilgi paylaşımının ve farkındalığın yoğun bir şekilde ifade edildiği bir platformdur. İnsanlar, Twitter üzerinden anlık olarak bilgi paylaşmakta, olaylara ilişkin tepkilerini dile getirmekte ve çeşitli konularda farkındalık yaratmaktadır. Deprem gibi ani ve yıkıcı olaylar sonrasında Twitter üzerinden yapılan paylaşımlar, halkın olaylara nasıl tepki verdiğini, hangi konularda endişe duyduğunu ve ne tür bilgi alışverişinde bulunduğunu anlamak için değerli bir kaynak sunar. Bu sebeplerden ötürü, bu çalışmada veri kaynağı olarak Twitter seçilmiştir.

Bu çalışma, Python programlama dili ve çeşitli kütüphaneleri kullanarak gerçekleştirilmiştir. Çalışmada, Twitter’da 6-21 Şubat 2023 tarihleri arasında atılan tweetlerden oluşan bir veri seti kullanılmıştır. Bu veri seti, deprem sonrasında halkın tepkilerini, paylaşımlarını ve bilgi alışverişini içermektedir. Veri seti üzerinde yapılan analizlerle depremle ilgili kamuoyu görüşleri ve tepkileri anlamaya yönelik çıkarımlar elde edilmiştir.

Çalışmanın sonuçları, veri görselleştirme araçları kullanılarak sunulmuştur. Bu çalışmanın bulguları, sosyal medya verilerinin toplumsal olayları anlama ve değerlendirme sürecinde ne kadar önemli bir kaynak olduğunu göstermektedir. Özellikle deprem gibi afet durumlarında sosyal medya verilerinin hızlı ve etkili bir şekilde analiz edilerek afet yönetimi ve kriz iletişimi süreçlerine katkı sağlayabileceği ortaya konmuştur.

Github Linki: https://github.com/denizsmagul/Bitirme_Projesi

Anahtar Kelimeler: Metin Madenciliği, Twitter, Deprem, Python, Veri Görselleştirme, Analiz

ABSTRACT

On February 6, 2023, a major earthquake disaster occurred in Turkey. The earthquake has been the main topic of social media shares after this event, which affected the entire country, especially the regions that are the epicenter of the earthquake. The destruction and losses caused by the earthquake have left deep traces on people, and social media tools have become a medium in which these feelings and thoughts are expressed. People have reached a wide audience by sharing their experiences, feelings and observations they have experienced after the earthquake via social media. This shows that by analyzing social media posts, they can be used as an important data source for understanding and evaluating social events.

Twitter is a platform where reactions, information sharing and awareness about every topic are expressed intensively. People share information instantly via Twitter, express their reactions to events and raise awareness about various issues. Sharing via Twitter after sudden and devastating events such as an earthquake provides a valuable resource for understanding how the public reacts to events, what issues they are concerned about, and what types of information they exchange. For these reasons, Twitter was chosen as the data source in this study.

This study was carried out using the Python programming language and various libraries. February 6-21, 2023 A data set consisting of tweets sent on Twitter between the dates of Dec. 2023 was used in the study. This data set includes public reactions, sharing and information exchange after the earthquake. With the analyses carried out on the data set, inferences have been obtained to understand public opinions and reactions about the earthquake.

The results of the study are presented using data visualization tools. The findings of this study show how important social media data is a resource in the process of understanding and evaluating social events. It has been shown that social media data can contribute to disaster management and crisis communication processes by analyzing them quickly and effectively, especially in disaster situations such as earthquakes.

Keywords: Text Mining, Twitter, Earthquake, Python, Data Visualization, Analysis

SEMBOLLER LİSTESİ

: Hashtag

@ : Twitter kullanıcı adı önek

KISALTMALAR LİSTESİ

NLTK: Natural Language Toolkit

URL: Uniform Resource Loader

NLP: Natural Language Processing

CSV: Comma Separated Values

TF-IDF: Term Frequency - Inverse Document Frequency

SVM: Support Vector Machine

ŞEKİL LİSTESİ

Şekil 2.1. Çalışmada analiz sonuçlarına ulaşmak için uygulanan adımlar	22
Şekil 2.2. Hashtag' lerin küçük harfe çevrilmesi.....	27
Şekil 2.3. En fazla kullanılan 10 hashtag ve kullanım sayılarının bulunması.....	27
Şekil 2.4. Date sütununda format dönüştürme işlemi.....	28
Şekil 2.5. Language sütununda format dönüştürme işlemi.....	28
Şekil 2.6. İl isimlerini içeren regex tanımlanması.....	29
Şekil 2.7. Mention içeren regex tanımlanması.....	29
Şekil 2.8. N-gram boyutu 1 olan n-gramların oluşturulması.....	30
Şekil 2.9. Url içeren tweetlerin sayılması.....	30
Şekil 3.1. En fazla kullanılan 10 hashtag için frekans grafiği.....	32
Şekil 3.2. Hashtag frekans WordCloud grafiği.....	32
Şekil 3.3. 06 – 21 Şubat tarihleri arasında günlük tweet yoğunluğu grafiği.....	33
Şekil 3.4. Tweet içeriklerinin dil dağılımı grafiği.....	34
Şekil 3.5. Belirlenen 11 ilin tweet içeriklerinde geçme yoğunluğu grafiği.....	34
Şekil 3.6. Belirlenen 11 il için WordCloud grafiği.....	35
Şekil 3.7. En fazla kullanılan 10 mention için frekans grafiği.....	36
Şekil 3.8. Mention frekans WordCloud grafiği.....	36
Şekil 3.9. Tweet içeriklerinde url geçme yoğunluğu grafiği.....	38

TABLO LİSTESİ

Tablo 2.1. Ön işleme yapılan tweet örnekleri.....	25
Tablo 3.1. Tweet içeriklerinde en sık kullanılan kelime ve kelime grupları listesi.....	37

1. GİRİŞ

Sosyal medya, günümüzde haber alma ve iletişimde önemli bir rol oynamaktadır. Özellikle Twitter gibi platformlar anlık olarak olayların takip edilmesini ve insanların duygularını, düşüncelerini paylaşmasını sağlar. Bu nedenle sosyal medya verilerinin analizi, toplumsal olayların anlaşılması ve değerlendirilmesi açısından büyük önem taşır. Sosyal medyanın sağladığı hızlı bilgi akışı, kullanıcıların olaylara dair anında geri bildirimde bulunabilmelerini sağlar. Böylece hem bireylerin hem de toplulukların duygusal ve düşünsel tepkileri gerçek zamanlı olarak izlenebilir ve analiz edilebilir. Bu analizler sadece bireysel kullanıcıların değil, aynı zamanda kamu kuruluşları, sivil toplum örgütleri ve medya kuruluşlarının da ilgisini çeker. Özellikle kriz anlarında sosyal medya, olayların yayılım hızını ve toplumsal tepkileri ölçmek için önemli bir araç haline gelir. Bu nedenle sosyal medya verilerinin analiz edilmesi, toplumsal olayların daha iyi anlaşılması ve değerlendirilmesi açısından büyük bir öneme sahiptir.

Araştırmacıların oldukça ilgisini çeken bu büyük veri, finans, duygu analizi, başarı tahmininde, bir konu ya da kavram ile ilgili neler konuşulduğunun saptanması amacıyla, kriz iletişimde veya toplumu etkileyen güncel olaylar hakkında anında bilgi almak, gündemdeki gelişmeler ile ilgili toplumun tepkilerini ölçmek amacıyla kullanılabilir (Ayan, Can & Gürsoy, 2019; Çobaner, 2019; Etaman, 2014). Bu tür analizler, yalnızca mevcut durumun anlaşılmasına yardımcı olmakla kalmaz, aynı zamanda gelecekteki olaylar ve tepkiler hakkında öngörülerde bulunmayı da sağlar. Örneğin bir kriz anında sosyal medya üzerinden yapılan paylaşımlar, krizin yayılımını ve halkın bu krize verdiği tepkileri ölçmek için kullanılabilir. Bu veriler kriz yönetimi stratejilerinin geliştirilmesinde önemli bir rol oynar ve toplumsal direncin artırılmasına katkıda bulunur.

Araştırmalar sosyal ağın, büyük ölçekli afetler sırasında durumu hızlı bir şekilde değerlendirmek için değerli olabileceğini ifade etmiştir (Mendoza, Poblete, & Valderrama, 2019). Özellikle deprem, sel ve diğer doğal afetler gibi büyük kriz durumlarında sosyal medya platformları, afetin etkilerini anında raporlayabilen ve yardım çağrılarını hızla iletebilen bir mecra olarak ön plana çıkar. Bu sayede afet yönetimi ve müdahale süreçleri daha etkin bir şekilde yürütülebilir. Ayrıca afet sonrası toparlanma sürecinde de sosyal medya, toplumsal desteğin organize edilmesinde ve yardımların koordinasyonunda önemli bir rol oynar. Kamuoyu görüşlerinin anlaşılması için bir kaynak olarak kullanılabilir (Doan, Ho Vo, & Collier, 2011). Sosyal medya verileri, toplumsal eğilimlerin, düşüncelerin ve

duyguların analizi için zengin bir kaynak sunar. Bu veriler politikacılar, araştırmacılar ve kamuoyu araştırmacıları tarafından halkın genel tutumlarını ve görüşlerini anlamak için kullanılabilir. Böylece daha bilinçli, etkili politika ve stratejiler geliştirilebilir.

Zheng vd. (2022) depremde yerinde ve hızlı müdahalelerin, afet ile ilgili bilgilerin zamanında elde edilmesine ve analiz edilmesine bağlı olduğunu belirtmiştir (Zheng, Shi, Zhou, Lu, & Lin, 2022). Bu tür bilgiler, afetin etkilerini azaltmak ve kurtarma operasyonlarını daha etkili hale getirmek için hayati öneme sahiptir. Ayrıca sosyal medya verileri, afet sonrası rehabilitasyon süreçlerinde de kullanılabilir. Bu veriler, afet bölgelerindeki ihtiyaçların belirlenmesi ve yardımların doğru bir şekilde yönlendirilmesi için kritik bilgiler sağlar. Sonuç olarak sosyal medya verilerinin analizi, sadece mevcut olayların anlaşılmasına yardımcı olmakla kalmaz, aynı zamanda gelecekteki krizlerin yönetimi ve toplumsal direncin artırılması için de önemli bir araçtır.

Türkiye, deprem riski yüksek bir coğrafyada bulunması sebebiyle sürekli olarak deprem tehlikesi altındadır. Ülkemizin bulunduğu bölge, aktif fay hatları ile çevrili olup tarih boyunca birçok yıkıcı depreme sahne olmuştur. Bu durum, Türkiye'deki şehirlerin ve yerleşim bölgelerinin sürekli bir tehdit altında olduğu anlamına gelir. Depremler, yıkıcı etkileriyle yaşamı ve çevreyi derinden etkileyebilir. Özellikle büyük ölçekli depremler, can kayıplarına, yaralanmalara ve maddi hasara yol açarak toplumu derinden sarsar. Bu nedenle depremlerin etkilerini anlamak, önceden tahmin edebilmek ve müdahale stratejileri geliştirmek oldukça önemlidir. Bu bağlamda, 6 Şubat 2023 tarihinde meydana gelen deprem Türkiye için önemli bir olaydır. Bu deprem hem büyüklüğü hem de yol açtığı yıkım açısından dikkat çekici olup toplumda büyük yankı uyandırmıştır. Depremin ardından sosyal medya platformlarında da yoğun bir bilgi akışı ve paylaşım yaşanmıştır. Twitter gibi sosyal medya araçları halkın duygularını, düşüncelerini ve deneyimlerini paylaştığı önemli bir alan haline gelmiştir.

Bu tez çalışmasında 6 Şubat 2023 Türkiye depremi hakkındaki Twitter verilerinin metin madenciliği ile incelenmesini amaçlanmıştır. Metin madenciliği, doğal dil işleme ve veri madenciliği tekniklerinin bir arada kullanılmasıyla gerçekleştirilir. Doğal Dil İşleme, bilgisayarların insan dilini anlamasını ve işlemesini sağlayan bir bilim dalıdır. Bu bilim dalı, metinlerdeki dil yapılarını ve anlamları analiz ederek bilgisayarların metin verilerini anlamlı bir şekilde işlemesine olanak tanır. Veri madenciliği ise büyük ölçekli veriler arasından faydalı bilgiye ulaşma, bilgiyi madenleme işidir (İnternet:Vikipedi). Metin madenciliği, doğal dil işleme ve veri madenciliği kavramlarının birleşimi olarak büyük miktarda metin verisini analiz

ederek bu verilerden anlamlı ve faydalı bilgiler çıkarılmasını sağlar. Bu işlemi yaparken çeşitli algoritmalar ve teknikler kullanır. Bu tez kapsamında, Twitter verileri üzerinde yapılacak metin madenciliği çalışmaları ile deprem sonrası toplumsal tepkiler ve bilgi akışının incelenmesi hedeflenmektedir. Bu analizler, deprem gibi büyük olaylar sonrasında toplumsal dinamiklerin ve tepkilerin daha iyi anlaşılmasına katkıda bulunacaktır.

Çalışmada Python dili ve kütüphaneleri kullanılmıştır. Python, veri analizi ve veri işleme konusunda güçlü ve esnek bir programlama dili olup oldukça kullanışlı kütüphanelerinin bulunması sayesinde veri bilimi projelerinde sıkça tercih edilmektedir. Çalışma için Twitter üzerinde 6-21 Şubat 2023 tarihleri arasında atılan tweetlerden oluşan bir hazır veri seti kullanılmıştır (Swaptr, 2023). Bu veri seti, depremin hemen ardından ve takip eden günlerde halkın tepkilerini, düşüncelerini ve paylaşımlarını içermektedir. Bu tür veriler, toplumsal olaylar ve kriz anlarında halkın nasıl tepki verdiğini anlamak için önemli bir kaynak sunar.

Bulunan veri seti üzerinde çeşitli analizler yapılarak sonuçlar elde edilmiştir. Bu sonuçlar, veri görselleştirme araçları kullanılarak sunulmuştur. Veri görselleştirme, karmaşık veri kümelerini daha anlaşılır ve yorumlanabilir hale getirmek amacıyla grafikler ve görseller kullanan bir tekniktir. Analizler arasında hashtag sayısı, günlük olarak atılan tweet sayısı, en çok kullanılan kelime ve kelime gruplarının belirlenmesi, atılan tweetlerin hangi dillerde olduğu, en fazla adı geçen illerin tespit edilmesi konuları yer almıştır. Bu analizler, deprem sonrasında sosyal medyada hangi konuların öne çıktığını, tweet paylaşımı yapan kişilerin hangi dillerde iletişim kurduğunu ve hangi bölgelerin en fazla konuşulduğunu ortaya koymuştur. Hashtag analizleri, belirli konuların ve etiketlerin ne kadar popüler olduğunu ve hangi konuların daha fazla ilgi gördüğünü anlamamıza yardımcı olmuştur. Günlük tweet sayısı analizleri ise belirli günlerde tweet aktivitesinin nasıl değiştiğini ve hangi günlerde daha yoğun paylaşımlar yapıldığını göstermiştir. Ayrıca en çok kullanılan kelime ve kelime gruplarının belirlenmesi, tweet paylaşımı yapan kişilerin en çok hangi kelimelerle duygularını ifade ettiğini ve hangi konuların daha fazla konuşulduğunu ortaya koymuştur. Tweetlerin hangi dillerde atıldığının tespit edilmesi, deprem sonrasında hangi dillerin daha fazla kullanıldığını ve uluslararası ilginin boyutlarını anlamamıza olanak sağlamıştır. En fazla adı geçen illerin tespit edilmesi ise, depremin hangi bölgelerde daha fazla hissedildiğini ve bu bölgelerle ilgili paylaşımların yoğunluğunu göstermiştir. Bu kapsamlı analizler, deprem sonrası sosyal medya dinamiklerini anlamak ve değerlendirmek için önemli bir kaynak oluşturmıştır.

1.1. İlgili Çalışmalar

Mendoza, Poblete, ve Valderrama yaptıkları çalışmada, depremlerin neden olduğu hasarların Twitter üzerinden erken tahmin edilmesine yönelik bir yöntem geliştirilmiştir (Mendoza, Poblete, & Valderrama, 2019). Çalışmada, Modifiye Mercalli Şiddet Ölçeği (Mercalli ölçeği) kullanılarak depremlerin şiddetinin ve hasar düzeyinin belirlenmesinde sosyal medya verilerinin nasıl kullanılabileceği incelenmiştir. Mercalli ölçeği, depremlerin algılanan şiddetini ve hasar düzeyini niteliksel olarak ifade eder ve acil durum müdahalesinin planlanmasında kritik bir rol oynar. Geleneksel yöntemlerle, bu raporların hazırlanması saatler veya günler sürebilir ve insan gözlemcilerin doğruluğuna bağlıdır. Bu çalışmada ise, depreme verilen tepkilere dayalı olarak erken tahminler yapabilen bir model önerilmiştir. Twitter'dan elde edilen verilere dayanarak, belirli bölgeler için Mercalli ölçeği raporlarının hızla üretilmesi hedeflenmiştir.

Coronel, García, ve Herrera-Tapia yaptıkları çalışmada, sosyal medya verilerini kullanarak doğal afetlerin yönetimini incelemişlerdir (Coronel, García, & Herrera-Tapia, 2019). Çalışma, 2017 Meksika depremi sırasında Twitter'dan elde edilen verileri analiz ederek, acil durum yönetiminde sosyal medyanın potansiyel kullanımını araştırmıştır. Twitter analizlerinin, afet sonrası yardımların organize edilmesi ve bilgi paylaşımının hızlandırılmasında kritik bir rol oynayabileceğini göstermiştir. Özellikle afetin ilk 72 saati içerisindeki tweetlerin zaman ve içerik analizine odaklanılmıştır.

Doan, Ho Vo ve Collier'in 2011 yılında gerçekleştirdiği çalışmada, Tohoku Depremi sırasında Twitter'dan toplanan 1,5 milyon tweet analiz edilmiştir (Doan Ho Vo, & Collier, 2011). Araştırma, Tokyo metropol bölgesinde farkındalık ve kaygı düzeylerini izlemeyi amaçlamıştır. Sonuçlar, tweet verilerinin deprem olaylarıyla yakından ilişkili olduğunu ve İngilizce ile Japonca tweetler arasında güçlü bir korelasyon olduğunu göstermiştir. Yerel dildeki tweetlerin erken uyarı sistemlerinde önemli bir rol oynadığı, ayrıca Japonya'daki insanların kaygılarının depremden sonra ne kadar hızlı normal seviyelere döndüğünü ortaya koymuştur. Araştırma, sosyal medya verilerinin doğal afetler sırasında halkın ruh halini izlemek ve erken uyarı sağlamak için yararlı bir kaynak olduğunu vurgulamaktadır. Elde edilen veriler, doğal afetlerin yönetiminde daha hızlı ve etkin müdahaleler için önemli bilgiler sağlamaktadır.

Fatma Kirman yaptığı çalışmada, 6 Şubat 2023'te Kahramanmaraş merkezli depremlerin ardından insanların sosyal medyada gösterdikleri travmatik stres tepkilerini incelemiştir (Kirman,2023). Bu çalışma, özellikle Twitter platformu üzerinden yapılan paylaşımlar analiz

edilerek gerçekleştirilmiştir. Deprem sonrasında insanların verdikleri tepkiler, yorumlarında dini referanslar ve depremle ilgili algıları, psikolojik açıdan değerlendirilmiştir. Araştırma, depremin ilk 20 günü içinde yapılan 160 paylaşımı incelemiş ve içerik analizi yöntemiyle değerlendirmiştir. Sonuçlara göre, depremin travmatik stres tepkileri arasında en sık rastlanan tepkilerden biri "sebeup arama" olarak belirlenmiştir. Bu tepkiler, dini ve seküler perspektifler olarak ikiye ayrılmıştır. Diğer yaygın tepkiler arasında acı ve korku duyma, suçluluk hissetme, anlamsızlık ve belirsizlik yaşama, öfke ve gerginlik hissetme, içe kapanma ve ümitsizlik gibi duygular yer almıştır. Çalışma, insan davranışının karmaşıklığını ve sosyal medyada paylaşılan tepkilerin çok boyutlu yapısını vurgulamıştır.

Zheng, Z. ve arkadaşlarının yaptıkları çalışmada, deprem etkilerinin analizinde metin madenciliği ve sosyal medya analitiğinin kullanımı incelenmiştir (Zheng, Shi, Zhou, Lu, & Lin, 2022). Bu çalışmada, sosyal medya platformlarından özellikle Twitter'dan elde edilen veriler kullanılarak, depremlerin etkilerini anlamak için metin madenciliği teknikleri uygulanmıştır. Araştırma, sosyal medyanın deprem hasarlarının hızla değerlendirilmesinde önemli bir kaynak olduğunu ortaya koymuştur. Sosyal medya gönderilerinin içeriğini ve desenlerini analiz ederek, depremden etkilenen ana bölgeler, rapor edilen hasar türleri ve etkilenen nüfusun acil ihtiyaçları belirlenmiştir. Elde edilen sonuçlar, sosyal medya analitiğinin geleneksel veri kaynaklarını tamamlayarak, afet müdahale çabaları sırasında daha hızlı ve daha ayrıntılı bilgi sağladığını göstermiştir.

Hatice Kübra Üçükkartal yaptığı çalışmada, sosyal medya etkileşimlerinin artan veri üretimini nasıl yönettiğini ve analiz ettiğini incelemektedir (Üçükkartal,2020). Bu çalışmanın amacı, Twitter platformunda insanların hangi konuları tartıştığını belirlemek, insanların görüşlerini göstermek, metin madenciliği konusunda altyapı oluşturmak ve veri görselleştirme yöntemleriyle verileri daha anlamlı hale getirmektir. Çalışmada, Twitter'dan toplanan verilere metin madenciliği yöntemleri uygulanmıştır. Metin madenciliği, veri bilimi ve veri görselleştirme araçlarıyla birleştirilerek veriler anlaşılır hale getirilmiştir. Özellikle duygu analizi kullanılarak insanların gönderilerinin pozitif, negatif veya nötr olup olmadığı analiz edilmiştir. Bu sayede sosyal medya verilerinin analiz edilmesi ve yorumlanması sağlanmıştır. Bu çalışma, sosyal medya verilerinin metin madenciliği yöntemleri ile analiz edilmesinin pratik uygulamalarını ve sonuçlarını göstermektedir. Çalışmanın bulguları, sosyal medya verilerinin analiz edilmesi ve anlamlandırılması konusundaki çeşitli yaklaşımları ve yöntemleri ortaya koymaktadır.

Fehime Elem YILDIRIM yaptığı çalışmada, 6 Şubat 2023 tarihinde Türkiye'de gerçekleşen depremin ardından Twitter'ın kullanımını incelemiştir (Yıldırım,2023). Deprem sonrası Twitter'da paylaşılan içerikler analiz edilerek, bilgi paylaşımı, yardım çağrıları ve kamuoyunun bilgilendirilmesi konusundaki işlevselliği değerlendirilmiştir. Çalışma, sosyal medyanın afet yönetimindeki rolünü ve etkisini vurgulamaktadır. Sonuçlar, Twitter'ın deprem sonrasında hızlı ve etkili bir bilgi paylaşım aracı olduğunu göstermektedir. Acil durumlarda vatandaşlar ve yetkililer Twitter'ı aktif olarak kullanmış, doğru ve yanlış bilgilerin hızla yayıldığı gözlemlenmiştir. Bu, sosyal medyanın kriz yönetiminde nasıl daha etkin kullanılabileceğine dair önemli bulgular sunmaktadır. Twitter, kriz anlarında hızlı bilgi yayılımı ve koordinasyon sağlama konusunda güçlü bir araç olmasına rağmen, yanlış bilgilerin yayılma riski de bulunmaktadır.

Hakan Aşan yaptığı çalışmada, deprem sonrasında sosyal medya hesaplarının güvenilirliği ve etkinliğini incelemiştir (Aşan,2024). Bu çalışma, afet yönetiminde sosyal medyanın nasıl kullanıldığını ve bu kullanımın güvenilirliğini analiz etmeyi amaçlamaktadır. Deprem sonrası sosyal medya platformlarında paylaşılan bilgilerin doğruluğu ve yayılım hızını ele alarak kriz anlarında sosyal medyanın rolünü değerlendirmiştir. Aşan' ın bulguları, sosyal medyanın afet yönetiminde kritik bir araç olduğunu ancak aynı zamanda yanlış bilgilerin hızla yayılabileceği riskini de taşımaktadır. Aşan' ın çalışması, deprem sonrası sosyal medyada yayılan içeriklerin doğruluk derecesini incelemiş ve bu içeriklerin afet yönetiminde nasıl kullanılabileceğine dair önemli sonuçlar ortaya koymuştur. Özellikle, doğru bilgi yayılımının kriz yönetiminde hayati öneme sahip olduğu, ancak yanlış bilgilerin de hızla yayılarak kaos yaratabileceği tespit edilmiştir. Bu çalışma, sosyal medya hesaplarının güvenilirlik analizinin afet yönetimi stratejilerinde dikkate alınması gerektiğini vurgulamaktadır. Sonuç olarak, sosyal medyanın afet sonrası bilgi paylaşımında etkin bir rol oynadığı, ancak bu platformların doğru ve güvenilir bilgi sunma kapasitesinin artırılması gerektiği belirtilmiştir.

Cemile Uzun'un yaptığı çalışmada, 6 Şubat 2023'te gerçekleşen Kahramanmaraş depremi sonrasında Twitter'da paylaşılan içeriklerin duygu analizi yapılmıştır (Uzun,2023). Uzun, çalışmasında deprem sonrası Twitter'da kullanıcıların paylaştığı tweetlerin duygusal tonunu analiz ederek halkın duygusal tepkilerini ve bu tepkilerin zaman içindeki değişimini incelemiştir. Bu analiz, deprem sonrası toplumun ruh halini ve sosyal medyanın, halkın duygusal tepkilerini nasıl yansıttığını anlamak açısından önemli bulgular sunmaktadır. Uzun'un çalışması, deprem sonrası Twitter'da paylaşılan tweetlerin olumlu, olumsuz ve nötr

duygusal içeriklerini analiz etmiş ve bu içeriklerin zaman içindeki dağılımını incelemiştir. Sonuçlar, deprem sonrasında halkın büyük çoğunluğunun olumsuz duygularla tweet attığını ve bu duygusal tepkilerin zamanla değişim gösterdiğini ortaya koymuştur. Çalışma, sosyal medyanın, kriz anlarında halkın duygusal tepkilerini ifade etme ve paylaşma aracı olarak önemli bir rol oynadığını vurgulamaktadır. Aynı zamanda, bu tür duygu analizlerinin kriz yönetimi ve halkla ilişkiler stratejilerinin geliştirilmesinde faydalı olabileceği belirtilmiştir.

Nezahat Koçyiğit, Mustafa Kocaoğlu ve Esra Yıldız yaptıkları çalışmada, 6 Şubat 2023 tarihinde meydana gelen depremler sonrasında Twitter'da yapılan paylaşımlarını incelemiştir (Koçyiğit, N., Kocaoğlu, M., & Yıldız, E., 2024). Yazarlar, sosyal medyanın afet yönetiminde ve bilgi yayılımında nasıl kullanıldığını ve dijital okuryazarlığın bu süreçlerdeki rolünü ele almışlardır. Çalışma, deprem sonrasında Twitter kullanıcılarının bilgi paylaşma, yardım çağrısında bulunma ve kamuoyunu bilgilendirme konusundaki etkileşimlerini analiz ederek dijital okuryazarlığın kriz anlarında ne kadar önemli olduğunu vurgulamaktadır. Bulgular, sosyal medyanın kriz yönetiminde etkin bir araç olduğunu, ancak dijital okuryazarlığın eksikliğinin yanlış bilgi yayılımına neden olabileceğini göstermektedir. Koçyiğit, Kocaoğlu ve Yıldız'ın çalışması, Twitter'da deprem temalı paylaşımların içerik analizini yaparak dijital okuryazarlığın afet yönetiminde nasıl bir rol oynadığını ayrıntılı bir şekilde incelemiştir. Deprem sonrası paylaşılan tweetlerin doğruluk ve güvenilirlik derecesi değerlendirilmiş, dijital okuryazarlık seviyesinin bu paylaşımlar üzerindeki etkisi analiz edilmiştir. Çalışma, sosyal medya platformlarının afet sonrası doğru bilgi yayılımında önemli bir rol oynadığını ancak dijital okuryazarlık eksikliklerinin bilgi kirliliğine yol açabileceğini ortaya koymuştur. Sonuç olarak, dijital okuryazarlığın artırılması ve sosyal medyanın daha etkin kullanılması, kriz yönetimi ve halkın doğru bilgilendirilmesi açısından hayati öneme sahiptir.

Yeşim Argın yaptığı çalışmada, 2023 Kahramanmaraş depremi sonrasında Twitter'ın kullanımını incelemiştir (Argın, 2023). Argın, çalışmasında deprem sonrası Twitter'da paylaşılan içeriklerin afet yönetiminde ve bilgi yayılımında nasıl bir rol oynadığını analiz etmiştir. Bulgular, Twitter'ın afet yönetiminde etkin bir araç olduğunu, ancak yanlış bilgilerin hızla yayılabileceği riskini de taşıdığını göstermektedir. Argın'ın çalışması, deprem sonrası Twitter'da paylaşılan tweetlerin içerik analizini yaparak, bu içeriklerin afet yönetimindeki etkilerini değerlendirmiştir. Özellikle, doğru bilgi yayılımının kriz anlarında hayati öneme sahip olduğu, ancak sosyal medyada yayılan yanlış bilgilerin halk arasında panik ve karmaşaya

yol açabileceği belirtilmiştir. Çalışma, sosyal medyanın kriz yönetiminde etkin kullanılması gerektiğini ve doğru bilgi sunma kapasitesinin artırılması gerektiğini vurgulamaktadır. Sonuç olarak, Twitter'ın afet sonrası bilgi paylaşımında önemli bir rol oynadığı, ancak yanlış bilgilerin yayılmasını önlemek için dijital okuryazarlığın artırılması gerektiği belirtilmiştir.

Esin Serin ve Simge Ünlü yaptıkları çalışmada, 6 Şubat 2023 Kahramanmaraş depremi sonrasında Twitter'da yayılan bilgi dezenformasyonunu ele almışlardır (Serin, Ünlü,2024). Serin ve Ünlü, çalışmalarında sosyal medyada yanlış bilgi yayılımını ve bu dezenformasyonun toplum üzerindeki etkilerini analiz etmişlerdir. Çalışma, deprem sonrası Twitter'da hızla yayılan yanlış bilgilerin kriz yönetimi ve halkın bilgilendirilmesi süreçlerini nasıl olumsuz etkilediğini vurgulamaktadır. Bulgular, sosyal medyada yayılan yanlış bilgilerin, toplumda panik ve karmaşa yaratma potansiyeline dikkat çekmektedir. Serin ve Ünlü' nün çalışması, deprem sonrası Twitter'da yayılan yanlış bilgi ve dezenformasyonun kriz yönetiminde karşılaşılan zorlukları ve bu durumun toplumsal etkilerini ayrıntılı bir şekilde incelemiştir. Özellikle sosyal medyada yayılan yanlış bilgilerin, halkın güvenilir bilgiye ulaşmasını zorlaştırdığı ve bu durumun afet yönetimi süreçlerini olumsuz etkilediği belirtilmiştir. Çalışma, sosyal medya platformlarının kriz anlarında doğru ve güvenilir bilgi sunma kapasitelerinin artırılması gerektiğini vurgulamaktadır. Sonuç olarak, sosyal medyada dezenformasyonun önlenmesi, etkili kriz yönetimi ve toplumun doğru bilgilendirilmesi açısından hayati öneme sahiptir.

Havva Nur Tarakçı yaptığı çalışmada, 6 Şubat 2023 Kahramanmaraş depremi sonrasında Twitter'da paylaşılan içeriklerinin analizini yapmıştır (Tarakçı,2023). Tarakçı, bu çalışmada deprem sonrası Twitter kullanıcılarının tweetlerinin içeriklerini ve duygusal tonlarını inceleyerek toplumun duygusal tepkilerini ve bu tepkilerin zaman içindeki değişimini analiz etmiştir. Bu çalışma, deprem sonrası toplumun ruh halini ve sosyal medyanın halkın duygusal ifadelerini nasıl yansıttığını anlamak açısından önemli bilgiler sunmaktadır. Tarakçı'nın çalışması, deprem sonrası Twitter'da paylaşılan tweetlerin olumlu, olumsuz ve nötr duygusal içeriklerini analiz etmiş ve bu içeriklerin zaman içinde nasıl değiştiğini incelemiştir. Bulgular, deprem sonrası halkın çoğunluğunun olumsuz duygularla tweet attığını ve bu duygusal tepkilerin zamanla değiştiğini göstermektedir. Çalışma, sosyal medyanın kriz anlarında halkın duygusal tepkilerini ifade etme ve paylaşma aracı olarak önemli bir rol oynadığını vurgulamaktadır. Ayrıca, bu tür duygu analizlerinin kriz yönetimi ve halkla ilişkiler stratejilerinin geliştirilmesinde yararlı olabileceği belirtilmiştir.

Bu çalışmada kullanılan veri setini kullanan diğer çalışmalar aşağıda verilmiştir:

Patil yaptığı çalışmada, Twitter üzerinde belirli konularla ilgili tweetleri makine öğrenimi teknikleri kullanarak sınıflandırmayı hedeflemektedir (Patil,2023). Proje kapsamında, tweet verileri öncelikle temizlenmiş ve hazırlanmıştır. Metinler küçük harfe dönüştürülmüş, gereksiz karakterler kaldırılmış ve stopword'ler çıkarılmıştır. Ardından, TF-IDF gibi yöntemler kullanılarak metinler sayısal özelliklere dönüştürülmüştür. Eğitilen Naive Bayes, SVM ve Random Forest gibi makine öğrenimi modelleri, tweetleri belirli konulara sınıflandırmak için kullanılmıştır. Modellerin performansı doğruluk, kesinlik, geri çağırma ve F1 skoru gibi metriklerle değerlendirilmiştir. Bu çalışma, sosyal medya verilerinin analiz edilmesinin önemini vurgulamakta ve kriz yönetimi ile halkla ilişkiler stratejilerinin geliştirilmesindeki potansiyel katkılarını ortaya koymaktadır. Yapılan tez çalışması, Patil'in yaptığı çalışmaya veri ön işleme aşamasında benzerlik göstermektedir. Patil, deprem konulu tweet paylaşımlarını makine öğrenimi teknikleri kullanarak sınıflandırmayı hedeflerken bu tez çalışmasında 2023 Türkiye depremi üzerine keşifsel veri analizi gerçekleştirilmiştir.

Abdulahdi yaptığı çalışmada, 2023 yılında Türkiye ve Suriye'deki depremler üzerine yapılan keşifsel veri analizi üzerine çalışmıştır (Abdulahdi,2023). Projenin temel unsurları veri toplama, veri görselleştirme, istatistiksel analiz ve coğrafi analiz olarak özetlenebilir. Veri toplama aşamasında depremlerle ilgili veriler çeşitli kaynaklardan elde edilmiş, bu verilerde depremlerin zamanlaması, konumu, derinliği ve şiddeti gibi önemli parametreler yer almaktadır. Veri görselleştirme yöntemleri kullanılarak depremlerin mekansal ve zamansal dağılımı haritalar ve grafiklerle analiz edilmiş, istatistiksel analizlerle depremlerin sıklığı ve şiddeti gibi veriler değerlendirilmiştir. Coğrafi analizler ise depremlerin büyük şehirlere ve fay hatlarına olan yakınlığının incelenmesiyle gerçekleştirilmiştir. Yapılan tez çalışması Abdulmahdi'nin yaptığı çalışmaya, yapılan analiz çeşitleri açısından benzerlik göstermektedir. Abdulmahdi, deprem konulu tweet paylaşımları üzerinde hashtag, dil, günlük tweet analizi gibi analizler gerçekleştirmiştir. Kalan analizler farklılık göstermektedir.

1.2. Amaç

Bu çalışma, ülkemizde 6 Şubat 2023 tarihli Türkiye depremiyle ilgili Twitter paylaşımları hakkında bilgi sahibi olmayı amaçlar. Türkiye, tarih boyunca birçok yıkıcı depreme maruz kalmış bir ülkedir ve bu tür doğal afetlerin etkileri toplumsal hafızada derin izler bırakmıştır. Sosyal medya araçları bu tür olaylar sonrasında halkın duygularını, düşüncelerini ve

deneyimlerini paylaştığı önemli bir alan haline gelmiştir. Özellikle Twitter, anlık bilgi akışı ve geniş kullanıcı tabanı ile bu tür analizler için değerli bir veri kaynağı sunar.

Literatür taraması sonuçlarına göre, Twitter içerikli birçok deprem konulu analiz çalışmaları olduğu görülse de toplumda gerçek afet bilincinin oluşturulması ve ayrıca yetkililerin daha etkili müdahale sistemleri oluşturmaya yönelik adımlar atabilmeleri için daha fazla çalışma yapılmasına ihtiyaç vardır. Deprem sonrasında sosyal medya verilerinin analizi, halkın afetlere karşı bilinç düzeyini artırmak ve yetkililerin müdahale stratejilerini optimize etmek için kullanılabilir.

Çalışma kapsamında literatürdeki benzer çalışmaların incelenmesi ve ardından ilgili kodların programa yazılarak analiz sonuçları alınması hedeflenmiştir. Bu süreçte, literatürde yer alan metodolojiler ve analiz teknikleri dikkate alınarak mevcut çalışmalardan faydalanılmıştır. Çalışma sonucu, insanların görüşlerinin sonuçlarını göstererek verileri daha anlamlı hale getirmeyi amaçlar.

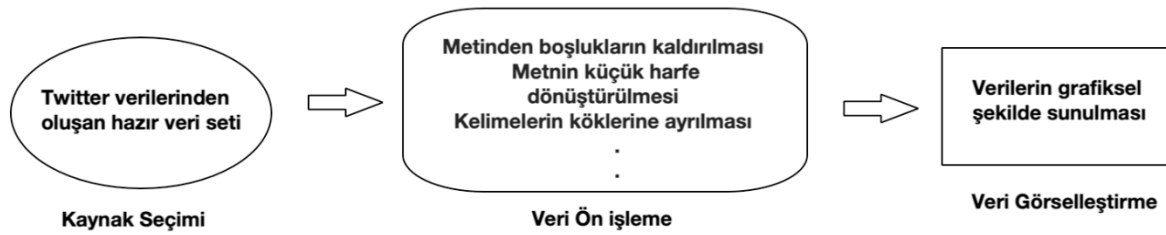
2. MATERYAL VE YÖNTEM

Twitter'da paylaşılan içerikler yapısal olmayan veri türlerindendir (Üçükkartal,2020). Bu içerikler, kullanıcıların çeşitli duygu ve düşüncelerini ifade ettiği serbest metin formatında paylaşımlar içerir. Yapısal olmayan bu veriler, doğrudan analiz edilmeye uygun değildir ve karmaşık bir yapıya sahiptir. Bu nedenle, bu verilerin etkili bir şekilde incelenebilmesi için belirli aşamalardan geçerek işlenebilir bir forma dönüştürülmesi gerekmektedir.

Metin tabanlı yapısal olmayan verilerin metin madenciliği yöntemleri ile incelenmesi, bu dönüşüm sürecinin temelini oluşturur. Verilerin uygun bir formata dönüştürülmesi, analiz edilebilir ve anlamlı hale getirilmesi için kritik bir adımdır. Bu süreçte, metin madenciliği teknikleri kullanılarak verilerin düzenlenmesi, temizlenmesi ve yapılandırılması sağlanır. Böylece, Twitter'daki geniş ve çeşitli veri yelpazesi daha kapsamlı şekilde analiz edilebilir ve önemli bilgiler elde edilebilir hale gelir. Bu yöntemler, sosyal medya analizleri, duygu analizleri ve toplumsal olayların incelenmesi gibi çeşitli alanlarda önemli katkılar sunar.

Tweet ölçümleri (tweet sayısı, kelime sayısı, hashtag sayısı, vb.) yapılarak verilerin basit ama temel bir sonucunu sunmak amaçlanır. İnsanların görüşlerinin sonuçlarını göstermeyi hedefleyen bu çalışmada Şekil 2.1' de bulunan adımlar sırasıyla gerçekleştirilmiştir. Çalışmada metin kaynağı olarak Twitter verilerinden oluşan hazır veri seti kullanılmıştır. Bu veri seti üzerinde analizler yapılabilmesi için metinler çeşitli veri ön işleme aşamalarından geçirilmiştir.

Veri ön işleme aşaması metinlerin küçük harfe dönüştürülmesi, kelimelerin köklerine ayrılması, metinlerden noktalama işaretlerinin kaldırılması gibi işlemleri içermektedir. Ön işleme adımlarının ardından verilerin grafiksel olarak sunulması amacıyla veri görselleştirme aşamasına geçilmiştir. Bu aşamada, verilerin daha anlaşılır ve yorumlanabilir olması için çeşitli grafik ve görselleştirme teknikleri kullanılmıştır.



Şekil 2.1. Çalışmada analiz sonuçlarına ulaşmak için uygulanan adımlar

2.1. Projede Kullanılan Araçlar

Bu projede kullanılacak olan araçlar şu şekildedir;

- Kaggle
- Python
- Visual Studio Code
- Google Colab

Kaggle, veri setini indirmek için kullanılacak olan arayüzdür.

Python, veri analizi ve görselleştirme işlemlerinin yapıldığı açık kaynak kodlu programlama dilidir.

Visual Studio Code, Python programlama dili ile yazılacak olan kodların çalıştırıldığı ortamdır.

Google Colab, kelime kök analizi için kullanılacak olan Python programlama dili ile yazılacak kodların çalıştırıldığı ortamdır.

Kullanılan kütüphaneler;

- Pandas
- Matplotlib
- WordCloud
- NLTK (Natural Language Toolkit)
- Re

- Collections
- Seaborn
- Zemberek

Pandas, Python programlama dilinde yazılmış bir veri analizi ve işleme kütüphanesidir. Veriler üzerinde hızlı ve kolay işlemler yapmayı mümkün kılar.

Matplotlib, Python'da grafik çizmek için kullanılan bir temel kütüphanedir. Birçok farklı grafik türünü çizmek için çeşitli fonksiyonlar sunar.

WordCloud, metin verilerindeki kelimelerin sıklığını görselleştirmek için kullanılan bir kütüphanedir. Kelimeleri, sıklıklarına göre farklı boyut ve renklerde gösteren bir bulut oluşturur.

NLTK (Natural Language Toolkit), doğal dil işleme için kullanılan bir Python kütüphanesidir. Metin analizi, kelime köklerini bulma ve cümle parçalama gibi işlemler gerçekleştirmek için kullanılır.

Re, Python'un düzenli ifadeler modülüdür. Metinlerde belirli desenleri aramak, değiştirmek veya eşleştirmek için kullanılır. Bu modül ile bir dizenin belirtilen arama modeline uyup uymadığını kontrol edebiliriz.

Collections, Python'un veri yapılarını genişleten bir modüldür. Nesnelerden oluşan bir topluluğu bir arada tutan yapıların oluşturulmasını sağlar.

Seaborn, Matplotlib'e dayalı bir veri görselleştirme kütüphanesidir. Daha çekici ve bilgilendirici grafikler oluşturmayı kolaylaştıran yüksek seviyeli bir arayüz sunar, özellikle istatistiksel grafikler için çok kullanışlıdır.

Zemberek, Türkçe dilinin yapısı dikkate alınarak geliştirilmiş NLP (Doğal Dil İşleme) kütüphanesidir. Zemberek yazım denetimi, hatalı kelimeler için öneri, heceleme, hatalı kodlama temizleme işlevlerinin yanı sıra, belirteçlere ayırma, cümle bölümlenme ve sözcük ayırma (tokenizations, sentence segmentation, lemmatizations) gibi işlevlere sahiptir (Tokcaer, 2021). Özellikle Türkçe'nin zengin eklemeli yapısını dikkate alarak doğru sonuçlar elde etmeye yönelik tasarlanmıştır. Metin ön işleme aşamasında kelimelerin köklerine ayrılması işleminde kullanılmıştır.

2.2. Veri Seti

Bu çalışmada Kaggle web sitesinde bulunan hazır veri setinden yararlanılmıştır (Swaptr,2023). Bu veri setinde 06-21 Şubat tarihleri arasında Twitter üzerinden atılmış olan toplam 478.053 adet tweet bulunmaktadır. Veri setinde bulunan bilgiler ve veri setinin içeriği şu şekildedir;

- Content sütunu, atılan tweetleri tutar.
- Date sütunu, atılan tweetler için atıldığı tarih ve saat bilgisini tutar.
- Hashtags sütunu, atılan tweetler içerisinde kullanılan hashtag bilgisini tutar.
- Like count sütunu, atılan tweet için beğenilme sayısını tutar.
- Rt count sütunu, atılan tweet için tekrar paylaşılma sayısını tutar.
- Followers count sütunu, atılan tweetin hesabının takipçi sayısını tutar.
- Language sütunu, atılan tweetin hangi dilde olduğu bilgisini tutar.
- Source sütunu, atılan tweet için hangi kaynaktan olduğu bilgisini tutar.

Kullanılan veri seti csv dosyası şeklinde ilgili web sitesinden indirilmiştir. Python programlama dilinin Pandas kütüphanesi kullanılarak veri setinin kodda okunması sağlanmıştır. Ardından bu csv dosyasında bulunan verilere metin madenciliği yöntemleri ile ön işleme yapılır.

2.3. Metin Madenciliği ile Ön İşleme

Yapılandırılmamış olan Twitter verilerine verinin temizlenmesi, düzenlenmesi ve analize uygun hale getirilmesi amacıyla metin madenciliği ile ön işleme aşamasında bir dizi işlemler uygulanır. Yapılan işlemler sonucu analiz sürecinin verimi artar ve daha doğru sonuçlar elde edilmektedir. Bu işlemler Türkçe veriler üzerinde yapılmıştır. Veriler üzerinde yapılan ön işlemler aşağıdaki gibidir;

- Metinden noktalama işaretleri kaldırıldı.
- Metin küçük harfe dönüştürüldü.
- Metinden karakterler, semboller, rakamlar kaldırıldı.
- Metinden hashtag, url ve kullanıcı isimleri kaldırıldı.
- Metinden özel karakterler ve boşluklar kaldırıldı.
- Metin başındaki, sonundaki ve metin içerisindeki fazla boşluklar kaldırıldı.
- Veri tekrarını önlemek amacıyla benzersiz veriler bulunarak kalanlar silindi.
- Türkçe içerikler için tek başlarına bir anlam ifade etmeyen stopwords olarak adlandırılan kelimeler metinden çıkartıldı.
- Kelimeler köklerine ayrıldı.

Metin ön işleme sonucu veri setinde toplam 87.609 adet Türkçe tweet kalmıştır. Metin ön işleme, Türkçe tweetler için en sık kullanılan kelime ve kelime gruplarının analizi için yapılmıştır. Kalan analizler veri setinde bir işlem gerçekleştirilmeden ya da bazı işlemler gerçekleştirilerek yapılmıştır. Örneğin hashtag analizi için metin ön işleme adımlarının tamamı değil sadece metnin küçük harflere dönüştürülmesi adımı uygulanmıştır. Çünkü veri setinin ‘Hashtags’ sütununda, her tweet için hangi hashtag kullanıldığı bilgisi bulunmaktadır. Bu nedenle yapılan analiz, bazı durumlarda tweetlerin temizlenmesi işleminden bağımsız gerçekleştirilmiş diyebiliriz. Bu duruma yol açan sebep, sadece Türkçe veriler üzerinde çalışılmamasıdır. Sadece Türkçe veriler üzerinde çalışılmadığı için tüm metin ön işleme adımlarını aynı anda farklı dillere de uyarlayamayız. Çünkü her dilin morfolojik yapısı birbirinden farklıdır. Bu duruma çözüm olarak, yapılacak olan her analiz için ayrı kod dosyası açılmış ve yapılan analize bağlı oluşabilecek durumlara göre kodlar yazılmıştır.

Projede kodların çalıştırıldığı ortam Visual Studio Code’ dur. Zemberek kütüphanesi, Visual Studio Code ortamında çalıştırılırken bazı sorunlar yaşanmıştır. Bu nedenle yalnızca kelimelerin köklerine ayrılması aşamasında yazılan kodlar, Google Colab ortamında çalıştırılmıştır. Kelime köklerine ayrılan veri seti üzerinde analiz, yine Visual Studio Code ortamında yapılmıştır.

Tablo 2.1’de, örnek tweetler üzerinde yapılan metin ön işleme sonuçları gösterilmektedir. Bu tablonun ilk sütununda, kullanıcılar tarafından gönderilen tweetlerin ilk halleri yer almaktadır. İkinci sütunda, noktalama işaretleri, rakamlar ve stopword’ lerden arındırılmış halleri, üçüncü sütunda ise kelime köklerine ayrılmış halleri gösterilmektedir.

Tablo 2.1. Ön işleme yapılan tweet örnekleri

Tweet	Temizlenmiş Tweet	Kelime Köklerine Ayrılmış Tweet
Allah ülkemizi bu büyük imtihandan selamete çıkarsın inşallah 🤲 #deprem #AFAD #earthquake	allah ülkemizi büyük imtihandan selamete çıkarsın inşallah	allah ülke büyük imtihan selamet çık inşallah
acil durumlarda yardımcı olabilecek uygulamalar, lütfen telefonunuzda bulunsun. #Turkey #deprem #earthquake #HelpTurkey #PrayForTurkey https://t.co/9QkAaNFwEP	acil durumlarda yardımcı olabilecek uygulamalar lütfen telefonunuzda bulunsun	acil durum yardımcı ol uygula lütfen telefon bulun

2.4. Veri Görselleştirme

Veri görselleştirme, verilerin insanlar tarafından daha kolay anlaşılır ve yorumlanabilir olması için grafik gibi görsel bir bağlama çevrilmesidir. Veri görselleştirme, bilgilerin grafiksel şekilde sunulmasıdır.

Çalışmada veri görselleştirme yöntemleri olarak WordCloud ve Matplotlib kullanılmıştır. WordCloud, Türkçe karşılığı kelime bulutu şeklindedir. Kelime bulutları genellikle kelime sıklıklarına göre analiz sonrasında, farklı renk ve desenlerle oluşturulan metin görselleridir (İnternet: Hüseyin, D. 2014. Kelime Bulutu). Metin içerisindeki kelimelerin sıklığını görsel olarak ifade ederek önemli kelimelerin ve temaların belirlenmesini sağlar. Kelimeler, sıklıklarına göre farklı boyutlarda ve renklerde gösterilir. Böylece veriler daha çekici ve anlaşılır bir şekilde sunulmuş olur. Depremden etkilenen 11 ilin tweetler içerisinde geçme yoğunluğunun hesaplanması sonuçları, WordCloud kütüphanesi kullanılarak görselleştirilmiştir.

Matplotlib ise grafikler ve çizimler oluşturmak için güçlü ve esnek bir araçtır. Çizgi grafik, bar grafik, histogram ve daha birçok grafik türünü destekleyerek verilerin etkili bir şekilde sunulmasına yardımcı olur. Günlük tweet yoğunluk analizi, hashtag analizi, atılan tweetler için dil tespiti analizi gibi hesaplamaların sonuçları ise Matplotlib kütüphanesi kullanılarak görselleştirilmiştir.

2.5. Projede Yapılan Analizler için Kullanılan Yöntemler

- # Hashtag Analizi: Hashtag, kullanıcıların aynı konu üzerine attıkları tweetleri kolayca bulabilmesini sağlamak için kullanılan bir etiket kullanımıdır. Hashtag analiz sonucuna varmak için kullanılan yöntem, veri setindeki 'hashtags' kolonunda yapılan işlemlere dayanmaktadır. Bu kolonda bulunan veriler, her tweet için tweetler içerisinde kullanılan hashtaglerden oluşmaktadır. Toplanan bu hashtag verileri, veri setine 'hashtags' adındaki kolon ile entegre edilmektedir. Bu entegrasyon işlemi, veri setinin analiz edilebilirliğini artırmak için önemlidir. Ayrıca, hashtag'lerin doğru bir şekilde sayılması ve analiz edilmesi için verilerin küçük harfe çevrilmesi gerekmektedir. Şekil 2.2' de bulunan kod parçası ile bu işlem gerçekleştirilmektedir. Bu kod hashtag'leri küçük harfe dönüştürmek için 'lowercase_hashtags' fonksiyonunu kullanır ve bu dönüşümü 'hashtags' kolonuna uygular.

```
#Hashtag'leri küçük harfe çevir  
tweets['hashtags'] = tweets['hashtags'].apply(lowercase_hashtags)
```

Şekil 2.2. Hashtag'lerin küçük harfe çevrilmesi

Hashtags kolonunda bulunan veriler küçük harfe çevrilerek aynı hashtag'in farklı yazım biçimlerinin ayrı ayrı olarak sayılması engellenmektedir. Bu sayede yapılan işlemlerde daha doğru sonuçlar elde edilebilir. Küçük harfe çevirme işlemi, veri setindeki tutarlılığı artırarak aynı hashtag'in farklı varyasyonlarının tek bir kategoride toplanmasını sağlar.

Hashtag analiz yöntemi, hangi hashtag'lerin daha çok kullanıldığını belirlemek için uygulanır ve elde edilen sonuçlar, sosyal medya kullanıcılarının belirli konulara olan ilgisini ölçmeye yardımcı olur. Şekil 2.3' de verilen kod parçası ile bu işlemi yapabiliriz. Bu kod parçası, tüm hashtag'leri içeren bir listeyi (all_hashtags) kullanarak en sık kullanılan ilk 10 hashtag'i ve bunların kaç kez kullanıldığını içeren bir Series oluşturur. Sonuç olarak, top_hashtags değişkeni, en sık kullanılan 10 hashtag ve bunların kullanım sayısını içeren bir Series olur. Bir Series, pandas'ta tek boyutlu, dizinlenmiş bir veri yapısıdır ve liste gibi davranır.

```
#En çok kullanılan hashtag'leri ve kullanım sayılarını bul  
top_hashtags = pd.Series(all_hashtags).value_counts().head(10)
```

Şekil 2.3. En fazla kullanılan 10 hashtag ve kullanım sayılarının bulunması

- **Günlük Tweet Yoğunluk Analizi:** Günlük tweet yoğunluk analizi sonucuna varmak için kullanılan yöntem, veri setindeki 'date' kolonunda yapılan işlemlere dayanmaktadır. Bu kolonda bulunan veriler, her tweet için tweet mesajına ait tarih ve saat bilgisinden oluşmaktadır. Analizin yapılması için 'date' kolonu tarih formatına dönüştürülmüştür. Şekil 2.4' de bulunan kod parçası bu işlemi yapmaktadır. Bu işlem tweetlerin tarih bazında doğru bir şekilde analiz edilmesini sağlamaktadır.

```
# 'date' sütununu tarih formatına dönüştür  
df['date'] = pd.to_datetime(df['date'])
```

Şekil 2.4. Date sütununda format dönüştürme işlemi

06 Şubat 2023 ve 21 Şubat 2023'e kadar olan tarih aralığında, her gün için atılan toplam tweet sayıları hesaplanmıştır. Sonuçlar 'y' ekseninde logaritmik ölçek kullanılarak çizgi grafiğinde görselleştirilmiştir.

- Dil Analizi: Dil analizi sonucuna varmak için kullanılan yöntem, veri setindeki 'language' kolonunda yapılan işlemlere dayanmaktadır. Bu kolonda bulunan veriler, her tweet için hangi dilde olduğu bilgisini verir. Analizin yapılması için 'language' kolonunda boş gelen değerler doldurulmuş ve metinler listeye dönüştürülmüştür. Şekil 2.5' de bulunan kod parçası bu işlemi yapmaktadır.

```
# 'language' kolonundaki NaN değerleri boş bir string ile doldur  
df['language'] = df['language'].fillna('')  
  
# 'language' kolonundaki metinleri bir liste olarak al  
documents = df['language'].tolist()
```

Şekil 2.5. Language sütununda format dönüştürme işlemi

Bu işlemler eksik veri nedeniyle oluşabilecek hataları önlemek ve dil frekanslarını hesaplamak için önemlidir. Toplam frekanslar toplanmış ve dil isimleri sıralı bir şekilde elde edilmiştir. Sonuçlar 'y' ekseninde logaritmik ölçek kullanılarak çizgi grafiğinde görselleştirilmiştir.

- 11 İl için Hasar Tweet Yoğunluğu Analizi: İller için hasar tweet yoğunluğu analizi sonucuna varmak için kullanılan yöntem, veri setindeki 'content' kolonunda yapılan işlemlere dayanmaktadır. Bu kolonda bulunan veriler, atılan tweetlerden oluşmaktadır. İl isimlerinin tweetler içerisindeki dağılımını bulmak için, il isimlerini içeren bir düzenli ifade (regex) pattern'i tanımlanmıştır. Düzenli ifadeler bir metin içerisinde kalıplar kullanarak esnek arama yapmanızı sağlar (İnternet: Uludaş, N.2019). Bu

pattern, projede tweet mesajlarında belirtilen illeri tespit etmek için kullanılmaktadır. Şekil 2.6’ da bulunan kod parçası bu işlemi yapmaktadır.

```
# İl isimlerini içeren bir regex pattern tanımlaması
pattern = r'\b(Adana|Gaziantep|Malatya|Adıyaman|Kahramanmaraş|Diyarbakır|Hatay|Elazığ|Osmaniye|Kilis|Şanlıurfa)\b'
cities = re.findall(pattern, tweet)
```

Şekil 2.6. İl isimlerini içeren regex tanımlanması

Tweetlerin her biri üzerinde bu regex pattern'i ile arama yapılmış ve bulunan il isimleri bir sayaç kullanılarak sayılmıştır. Böylece her il için tweet yoğunluğu toplanarak hesaplanmıştır.

- @ Mention/Kullanıcı Adı Analizi: @ işareti, tweetlere kullanıcı eklemek için kullanılan bir işarettir. Kullanıcı adı önüne @ işareti konarak tweet ilgili kişinin sistemine düşürülür. Kullanımı @kullanıcıadı şeklindedir. Kullanıcılar sizden bahsetmek, size mesaj göndermek veya profilinize bağlantı göndermek için @ işaretiyle birlikte kullanıcı adı kullanır. Mention analizi için kullanılan yöntem, veri setindeki ‘content’ kolonunda yapılan işlemlere dayanmaktadır. Mention analizi, tweetlerdeki @mention kullanımlarının incelenerek hangi kullanıcıların daha fazla etkileşim aldığını ve hangi kullanıcılarla daha sık etkileşime geçildiğini belirlemeye yönelik bir yöntemdir. Bu analizde, veri setindeki 'content' kolonundaki tweetler taranarak mention' lar bulunur ve bu mention' lar sayılır. Şekil 2.7’ de bulunan kod parçası, tweet içeriklerindeki mentionları bulmak için bir regex tanımlamasıdır.

```
# Mention'ları bulmak için regex deseni
mention_pattern = re.compile(r'@(\w+)')
```

Şekil 2.7. Mention içeren regex tanımlanması

En çok kullanılan mention' lar belirlendikten sonra bu veriler çubuk grafik şeklinde görselleştirilir.

- En Sık Kullanılan Kelime ve Kelime Grupları Analizi: Kelime ve kelime grupları analizi sonucuna varmak için kullanılan yöntem, veri setindeki ‘content’ kolonunda yapılan işlemlere dayanmaktadır. Metin madenciliği ile ön işleme adımıdaki tüm

maddeler gerçekleştirildikten sonra en sık kullanılan kelimeler ve ikili yapıda olan kelime grupları bulunmuştur. Veri setinde bulunan tweet mesajlarında en çok kullanılan kelimeler ve kelime gruplarını belirlemek amacıyla Counter aracı kullanılarak n-gram değerleri 1 ve 2 olarak ayarlanmış ve analizler yapılmıştır. Şekil 2.8’ de bulunan kod parçası, 1-gram ‘ları oluşturmaktadır.

```
# n-gram = 1 oluşturulur
n = 1
ngrams_list1 = ngrams(tokens, n)
```

Şekil 2.8. N-gram boyutu 1 olan n-gramların oluşturulması

Boş gelen alanlar silinerek yanlış hesaplamaların önüne geçilmiştir. İlk 10 adet en sık kullanılan kelime ve kelime grupları bulunarak konsola yazdırılmıştır. Ardından analiz sonuçlarının görsel olarak daha anlaşılır olması için bir tabloya aktarılmıştır.

- URL Kullanım Yoğunluğu Analizi: Url kullanım yoğunluğu analizi sonucuna varmak için kullanılan yöntem, veri setindeki ‘content’ kolonunda yapılan işlemlere dayanmaktadır. Analiz için tarih bilgilerinde bir format dönüştürme işlemi yapılmıştır. Ardından 06-21 Şubat 2023 tarihleri arasındaki tweetler filtrelenmiştir. Filtrelenen tweetler içerisinde URL bulunup bulunmadığı belirlenmiştir. Şekil 2.9’da bulunan kod parçası bu işlemi yapmaktadır. Kod parçasında url için bir regex tanımlaması yapılmış ve metin içinde bu ifade ile eşleşen tüm url’ ler bulunmuştur.

```
# URL içeren tweetleri saymak için gerekli fonksiyon
def count_urls(text):
    url_pattern = re.compile(r'http[s]?://\S+')
    return len(url_pattern.findall(text)) > 0
```

Şekil 2.9. Url içeren tweetlerin sayılması

URL içeren ve içermeyen tweetlerin sayıları hesaplanarak toplam tweet sayısı ile karşılaştırılmış ve günlük olarak URL içeren tweet sayıları elde edilmiştir. Ayrıca en yoğun kullanılan adresler de tespit edilmiştir.

2.6. Kısıtlar

Çalışma kapsamındaki kısıtlar;

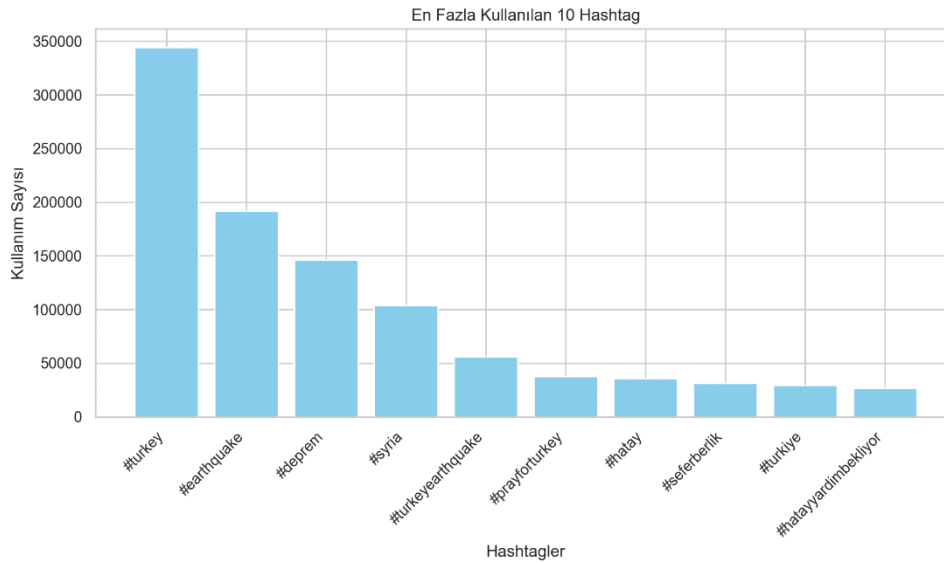
- Veri seti yalnızca Twitter verilerini içermektedir.
- Kayıtlar 06-21 Şubat 2023 tarih aralığındaki tweetlerden oluşmaktadır.
- Çalışmada sadece metin içerikli tweetler değerlendirmeye alınmıştır.
- Kelime ve kelime grupları analizi için sadece Türkçe veriler dikkate alınmıştır. Kalan analizler için ise hem Türkçe hem de yabancı dildeki veriler kullanılmıştır.

3.BULGULAR

Bu çalışmada Python programlama dili kullanılarak veri seti üzerinde Pandas, Zemberek gibi kütüphaneler ile silme, dönüştürme, köklerine ayırma vb. işlemler ve analizler yapılmıştır. Elde edilen analiz sonuçları, matplotlib ve wordcloud kütüphaneleri kullanılarak görselleştirilmiştir. Yapılan analizler ve sonuçları aşağıda sırasıyla sunulmuştur.

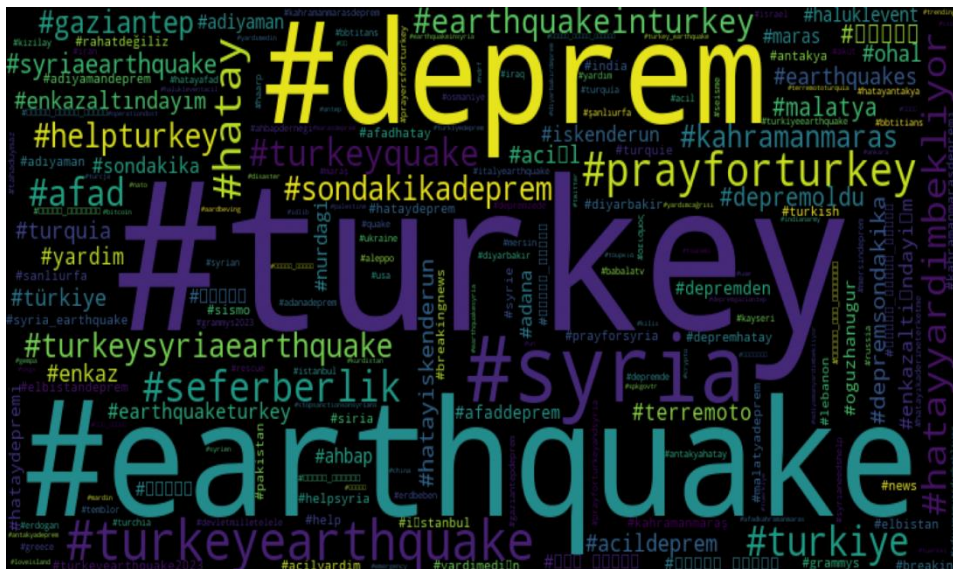
3.1. #Hashtag Analizi

Hashtag analizi, sosyal medya kullanıcılarının görüşleri, eğilimleri, ilgi alanları ve genelde veri seti içeriği hakkında bizlere ipuçları verir. Şekil 3.1’ de en fazla kullanılan 10 hashtag için frekans bar grafiği sunulmuştur. Yatay ekseninde kullanım sayıları, dikey ekseninde ise hashtag isimleri yer almaktadır. Grafik incelendiğinde, en fazla 344.381 adet ile #turkey hashtag’ının kullanıldığı görülmüştür. Veri setinin içeriği dil yönünden çeşitli olduğu için hashtag analizinde hem Türkçe hem de yabancı dillerden sonuç gelmiştir. Ayrıca toplam 478.052 tweet içeren veri setinde 18 adet tweet mesajında hashtag kullanılmadığı bulunmuştur.



Şekil 3.1. En fazla kullanılan 10 hashtag için frekans grafiği

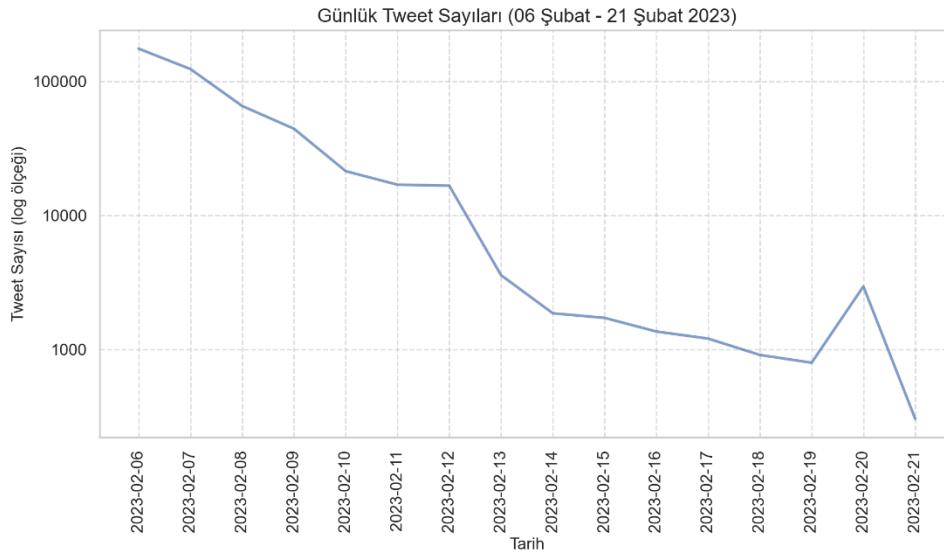
Hashtag frekans Wordcloud grafiğinin daha anlaşılır olması amacıyla en az 2000 kez kullanılan hashtagler ekrana basılmıştır. Şekil 3.2’ de tweet içeriklerinde kullanılan hashtagler için WordCloud frekans grafiği sunulmuştur. Bu grafik incelendiğinde en büyük boyutlu yani en fazla kullanılan hashtag, #turkey olduğu görülmektedir.



Şekil 3.2. Hashtag frekans WordCloud grafiği

3.2. Günlük Tweet Yoğunluk Analizi

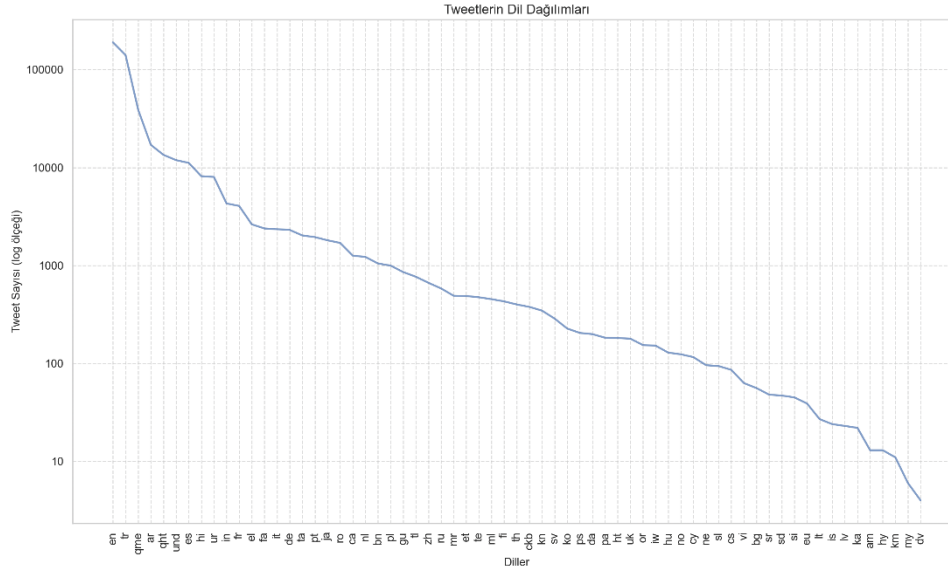
Bu çalışmada 06 – 21 Şubat aralığında atılan tweet yoğunluğuyla ilgili bir analiz gerçekleştirilmiştir. Şekil 3.3’ de 06-21 Şubat tarihleri arasında günlük olarak atılan tweet yoğunluğu çizgi grafiği sunulmuştur. Bu grafikte sonuçlar görselleştirilirken logaritmik ölçekten yararlanılmıştır. Yatay eksen logaritmik ölçekte tweet sayıları, dikey eksen ise 06-21 Şubat aralığında bulunan tarihler yer almaktadır. Grafik incelendiğinde 06 Şubat 2023 gününün en yoğun tweet atılan gün olduğu bulunmuştur.



Şekil 3.3. 06 – 21 Şubat tarihleri arasında günlük tweet yoğunluğu grafiği

3.3. Dil Analizi

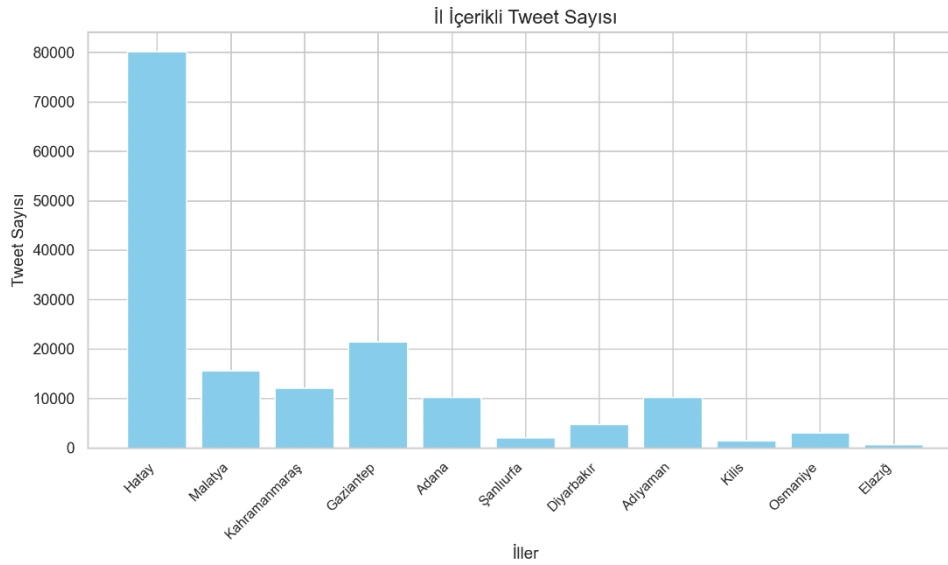
Bu çalışmada 06 – 21 Şubat aralığında atılan tweetlerin hangi dile sahip olduğuyla ilgili bir analiz gerçekleştirilmiştir. Şekil 3.4’ de tweet içeriklerinin dil dağılımı çizgi grafiği sunulmuştur. Bu grafikte sonuçlar görselleştirilirken logaritmik ölçekten yararlanılmıştır. Yatay eksen logaritmik ölçekte tweet sayıları, dikey eksen ise dil isimleri yer almaktadır. Grafik incelendiğinde veri setindeki tweetlerin birçok farklı dilden geldiği tespit edilmiştir. Veri setindeki tweet içeriklerinin en fazla İngilizce dilinde olduğu görülmüştür. İngilizce dilinden sonra tweet içeriklerinde en fazla kullanılan ikinci dil Türkçe olmuştur. Ayrıca toplam 65 farklı dilde tweet mesajı atıldığı bulunmuştur.



Şekil 3.4. Tweet içeriklerinin dil dağılımı grafiği

3.4. 11 İl için Hasar Tweet Yoğunluğu Analizi

Depremden etkilenen 11 ilin adlarının tweet içeriklerinde geçme yoğunluğuyla ilgili bir analiz gerçekleştirilmiştir. Şekil 3.5’ de belirlenen 11 ilin tweet içeriklerinde geçme yoğunluğu bar grafiği sunulmuştur. Yatay ekseninde tweet sayıları, dikey ekseninde ise il isimleri yer almaktadır. Grafik incelendiğinde Hatay ilinin diğer illere göre tweet içeriklerinde daha fazla geçtiği bulunmuştur. Hatay ilinden sonra tweet içeriklerinde en fazla adı geçen ikinci il Gaziantep olmuştur.



Şekil 3.5. Belirlenen 11 ilin tweet içeriklerinde geçme yoğunluğu grafiği

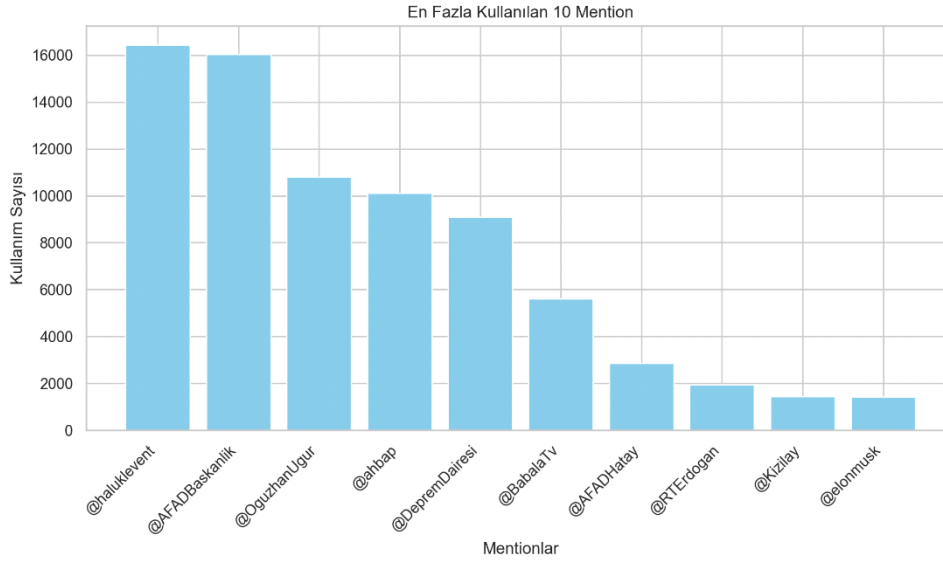
Şekil 3.6’ da ise belirlenen 11 il için WordCloud grafiği sunulmuştur. Bu grafik incelendiğinde en büyük boyutlu yani tweet içeriklerinde en fazla geçen ilin, ‘Hatay’ olduğu görülmektedir.



Şekil 3.6. Belirlenen 11 il için WordCloud grafiği

3.5. Mention/Kullanıcı Adı Analizi

Veri setindeki tweet analizlerinde @mention ve bunların toplam kullanılma sayıları, kullanıcıların kiminle etkileşimde bulunduğunu gösterir. Bu çalışmada, tweetlerin hangi kullanıcılara yönlendirildiğini ve hangi kullanıcıların daha fazla etkileşim aldığını belirlememizi sağlayan bir analiz gerçekleştirilmiştir. Şekil 3.7’ de tweet içeriklerinde en fazla kullanılan 10 mention için frekans bar grafiği sunulmuştur. Yatay ekseninde kullanım sayıları, dikey ekseninde ise kullanıcı adları yer almaktadır. Grafik incelendiğinde en fazla @haluklevent kullanıcısının 16430 adet tweette paylaşıldığı görülmüştür. Ayrıca toplam 478.052 tweet içeren veri setinde 392.493 adet tweet mesajında mention kullanılmadığı bulunmuştur.



Şekil 3.7. En fazla kullanılan 10 mention için frekans grafiği

Mention frekans Wordcloud grafiğinin daha anlaşılır olması amacıyla en az 500 kez kullanılan kullanıcı adları ekrana basılmıştır. Şekil 3.8’ de tweet içeriklerinde kullanılan mentionlar için WordCloud frekans grafiği sunulmuştur. Bu grafik incelendiğinde en büyük boyutlu yani tweet içeriklerinde en fazla kullanılan mention, @haluklevent ya da @AFADBaskanlik olduğunu söyleyebiliriz. Wordcloud grafiği bu sonuç için kesinlik vermemektedir. Bar grafiğinden de anlaşılacağı üzere kullanım sayıları birbirlerine oldukça yakındır.



Şekil 3.8. Mention frekans WordCloud grafiği

3.6. Tweet İçeriklerinde En Sık Kullanılan Kelimeler ve Kelime Grupları

İnsanların ilgi odağı olan ve tartışılan konuları belirlemek için paylaşılan tweet mesajlarındaki kelimelerin frekans analizi yapılır. Ancak kelime frekans analizi, # etiketlerinin analizinden farklı olarak kelimenin öncelikle kök haline dönüştürülmesi gerekir (Güneş & Arıkan, 2023). Bu aşamada zemberek kütüphanesi kullanılmıştır.

Tweet içeriklerinde en sık kullanılan kelime ve kelime gruplarını bulmak amacıyla N-Gram analizi yapılmıştır. N-gram bize hangi kelimelerin ayrı ayrı, hangilerinin birlikte kullanıldığını anlamamızda yardımcı olur (Addo, Akpatsa, & Dela, 2023). Bu analizde n-gram değerleri 1 ve 2 verilmiştir. Böylece kelimeler ve ikili kelime grupları incelenmiştir.

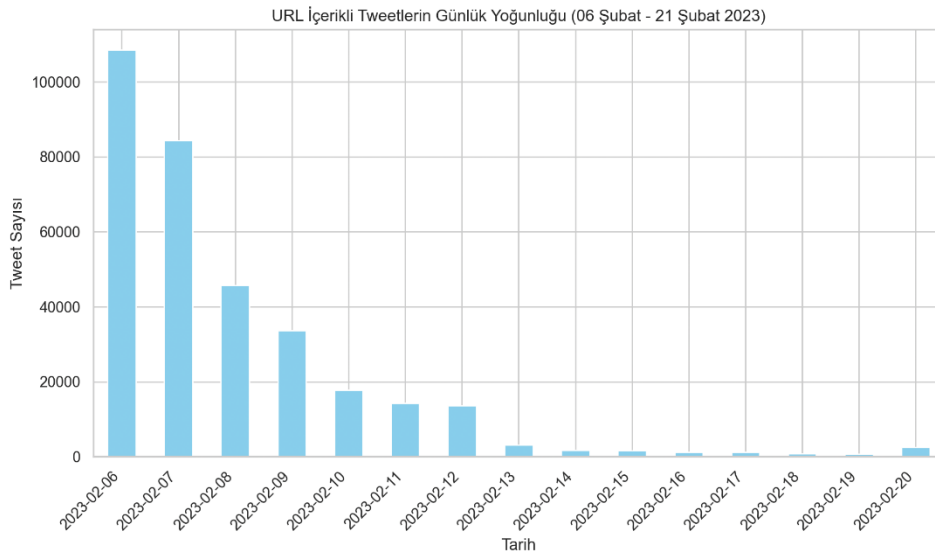
N-gram=1 değeri için frekansı en yüksek terimler “ol”, “yardım”, “lütfen” bulunmuştur. N-gram=2 değeri için frekansı en yüksek terimler “yardım et”, “enkaz alt”, “lütfen yar” bulunmuştur. Tablo 3.1’ de tweet içeriklerinde en sık kullanılan kelime ve kelime grupları listesi sunulmuştur. N-gram = 1 tek başına kullanılan kelimeleri, N-gram = 2 ise ikili kelime gruplarını temsil etmektedir.

Tablo 3.1. Tweet içeriklerinde en sık kullanılan kelime ve kelime grupları listesi

Sıra	N-Gram = 1		N-Gram = 2	
	Terim	Frekans	Terim	Frekans
1	ol	30300	yardım et	7190
2	yardım	24841	enkaz alt	6644
3	lütfen	21576	lütfen yar	3603
4	et	16987	yardımcı ol	2983
5	deprem	14524	deprem bölge	2972
6	var	14299	göçük alt	2002
7	enkaz	12079	haber al	1823
8	arkadaş	10554	ihtiyaç var	1645
9	alt	9951	lütfen yay	1550
10	hatay	9476	geçmiş ol	1536

3.7. Tweet İçeriklerinde URL Kullanım Yoğunluğu Analizi

06 – 21 Şubat tarihleri arasında paylaşılan tweet içeriklerinde URL kullanım yoğunluğu ile ilgili bir analiz gerçekleştirilmiştir. Şekil 3.9’ da günlük atılan tweet içeriklerinde url kullanım yoğunluğu bar grafiği sunulmuştur. Yatay ekseninde tweet sayıları, dikey ekseninde ise 06-21 Şubat aralığında bulunan tarihler yer almaktadır. Grafik incelendiğinde URL bulunduran tweetlerin en yoğun 06 Şubatta, ikinci en yoğun ise 07 Şubatta atıldığı bulunmuştur. URL adreslerinin kullanımının ilerleyen günlerde giderek azaldığını söyleyebiliriz.



Şekil 3.9. Tweet içeriklerinde url geçme yoğunluğu grafiği

En çok kullanılan URL adreslerinin analizinde, çoğunlukla sismik izleme ve acil durum bilgilendirmesi amaçlarıyla kullanıldığı gözlemlenmiştir. Örneğin, bu adresler arasında sismolojik cihaz ve yazılım geliştiren bir web sitesi, küresel afet ve acil durumlar hakkında bilgi sunan platformlar, gerçek zamanlı uyarılar sağlayan bir uygulama önerisi ve Telegram sayfalarına yönlendirme yapan bağlantılar bulunmaktadır. Bazı bağlantıların çalışmadığı ve erişilebilir olmadığı da bulunmuştur.

Telegram, çoklu platform desteği sunan, güvenli anlık mesajlaşma servsidir (İnternet:Vikipedi).

4. SONUÇLAR

Bu tez çalışmasında, 6 Şubat 2023 tarihinde Türkiye'de meydana gelen büyük deprem sonrası Twitter paylaşımları metin madenciliği yöntemleriyle analiz edilmiştir. Yapılan analizler toplumun bilgi paylaşımını, gündemini belirlemek için önemli veriler sağlamıştır. Bu bölümde yapılan analizlerin sonuçları ve metin ön işleme adımlarının etkileri detaylı bir şekilde ele alınmıştır.

Yapılan analizler sonucunda elde edilen bulgular, halkın sosyal medyada nasıl tepkiler verdiğini göstermektedir. Çalışma sonucu elde edilen analizler aşağıda tartışılmıştır.

Hashtag kullanımı, kullanıcıların deprem sonrası bilgi ve duygu paylaşımını organize etmek ve geniş kitlelere ulaşmak için sıkça tercih ettikleri bir yöntemdir. Hashtagler, belirli konuların takip edilmesini kolaylaştırır ve aynı konuya ilgi duyan kullanıcıları bir araya getirir. Deprem sonrası Twitter'da en çok kullanılan hashtag #turkey olmuştur. Bu hashtag, toplam 344.381 kez kullanılmıştır.

Toplam 478.052 tweet içeren veri setinde 18 tweet mesajında hashtag kullanılmadığı tespit edilmiştir. Bu, hashtag kullanımının yaygın olmasına rağmen bazı kullanıcıların tweetlerinde hashtag kullanmamış olabileceğini göstermektedir.

Hashtag analiz sonuçları, halkın deprem sonrası bilgi paylaşımı ve farkındalık oluşturma çabalarını yansıtmaktadır. Özellikle belirli hashtaglerin yoğun kullanımı, bu hashtaglerin toplumsal hafızada ve iletişimde önemli bir yer tuttuğunu göstermektedir.

Günlük tweet yoğunluk analizleri, halkın deprem sonrasında sosyal medyada nasıl bir tepki verdiğini ve hangi günlerde daha aktif olduğunu ortaya koymaktadır. 6 Şubat 2023 günü, en yoğun tweet atılan gün olarak tespit edilmiştir. Bu durum, depremin hemen ardından halkın yoğun bir şekilde Twitter'ı kullandığını ve anında bilgi paylaşımı yaptığını göstermektedir. Bu yoğunluk, depremin yarattığı ani şokun ve bilgi ihtiyacının bir göstergesidir. Bu bilgiler, kriz anlarında halkın bilgiye ulaşma ve paylaşma davranışlarını anlamak için değerlidir.

Analizler, deprem sonrası atılan tweetlerin farklı dillerde olduğunu göstermektedir. En fazla kullanılan dil İngilizce olup, bunu Türkçe takip etmektedir. İngilizce' nin yaygın kullanımı, depremin uluslararası boyutta ilgi gördüğünü ve farklı dillerde paylaşımlar yapıldığını göstermektedir. Toplamda 65 farklı dilde tweet atıldığı belirlenmiştir. Bu, depremin geniş bir coğrafyada ilgi uyandırdığını ve dünya genelinde takip edildiğini göstermektedir. Dil

çeşitliliği, deprem sonrası paylaşımların uluslararası boyutunu ve küresel topluluğun desteğini yansıtmaktadır.

Depremden etkilenen 11 ilin adlarının tweet içeriklerinde geçme yoğunluğunun analiz edilmesi sonucunda, en fazla adı geçen ilin Hatay olduğu belirlenmiştir. Hatay'ı Gaziantep takip etmektedir. Bu durum, bu illerin depremden en fazla etkilenen bölgeler olduğunu ve halkın bu bölgeler hakkında yoğun bir şekilde paylaşımlarda bulunduğunu göstermektedir. Bu analiz, depremden en çok etkilenen bölgeler hakkında bilgi sunmakta ve halkın bu iller hakkında yoğun bir şekilde paylaşımlarda bulunduğunu göstermektedir. Hasar yoğunluğu analizleri, afet yönetimi ve yardım çalışmalarının odaklanması gereken bölgeleri belirlemek için kritiktir.

En fazla mention yapılan kullanıcının @haluklevent olduğu tespit edilmiştir. Bu kullanıcı, toplam 16.430 tweette mention edilmiştir. Bu, Haluk Levent'in deprem sonrası yardım çağrıları ve koordinasyon çabalarında önemli bir rol oynadığını göstermektedir. Mention analizi, halkın hangi kullanıcılarla etkileşimde bulunduğunu ve bu kullanıcıların toplumsal etkilerini anlamamıza yardımcı olmaktadır. Özellikle kriz anlarında belirli kullanıcıların ön plana çıkması, bu kişilerin güvenilir bilgi kaynağı veya yardım organizatörü olarak kabul edildiğini göstermektedir.

Tweet içeriklerinde en sık kullanılan kelimeler "ol", "yardım", "lütfen" olarak belirlenmiştir. Bu kelimeler, halkın acil yardım taleplerini ve çağrılarını yansıtmaktadır. İkili kelime grupları analizinde ise "yardım et", "enkaz alt", "lütfen yar" terimleri öne çıkmıştır. Bu analiz, halkın deprem sonrası en çok hangi konularda konuştuğunu ve hangi kelimeleri kullandığını ortaya koymaktadır. Kelime ve kelime grubu analizleri, halkın önceliklerini ve acil ihtiyaçlarını anlamak için değerlidir.

En yoğun URL kullanımı 06 Şubat tarihinde gözlemlenmiş olup, özellikle sismik izleme ve acil durum bilgilendirmesi içerikli tweetlerde bu trend belirgin şekilde görülmüştür. URL içeren tweetlerin ikinci en yoğun olduğu tarih ise 07 Şubat olarak belirlenmiştir.

En çok paylaşılan URL adreslerinin incelenmesi sonucunda, genellikle sismolojik cihaz ve yazılım geliştiren siteler, küresel afet ve acil durumlar hakkında bilgi sunan platformlar, gerçek zamanlı uyarılar sağlayan uygulamalar ve Telegram sayfalarına yönlendirme yapan bağlantılar sıklıkla yer almıştır. Ancak bazı bağlantıların erişilebilir olmadığı da tespit edilmiştir. Bu analiz, belirli dönemlerde Twitter kullanıcılarının özellikle sismik aktiviteler ve acil durumlar

konusunda bilgi edinmek için URL içeriklerine yoğun bir şekilde başvurduğunu ve bu bilgileri aktararak toplumsal farkındalık oluşturma çabası gösterdiğini göstermektedir.

Sosyal medya platformlarından biri olan Twitter, kriz anlarında halkın duygusal ve düşünsel tepkilerini anında paylaşabildiği bir alan sunar. Bu tez çalışmasının faydalarından bahsedecek olursak deprem sonrasında halkın duygusal tepkilerini ve önceliklerini belirleyerek, afet yönetimi ve kriz iletişimi süreçlerine katkı sağlamaktadır. Sosyal medya verilerinin hızlı ve etkili bir şekilde analiz edilmesi, doğru ve zamanında müdahalelerin yapılmasına olanak tanır. Veri görselleştirme teknikleri, elde edilen bilgilerin geniş kitlelere ulaştırılmasını kolaylaştırır. Çalışma ayrıca, sosyal medya platformlarının toplumsal bilinç oluşturma ve kriz anlarında etkin iletişim sağlama konusundaki potansiyelini ortaya koyar. Literatürdeki benzer çalışmaların analizleri, bilimsel bilgiye katkı sağlar ve metin madenciliği yöntemlerinin sosyal medya verilerinin analizindeki önemini vurgular.

Bu çalışma, deprem sonrası sosyal medya platformlarından biri olan Twitter için yapılan paylaşımların analiziyle toplumsal tepkilerin, bilgi paylaşımının ve afet yönetimi süreçlerinin daha iyi anlaşılmasına katkı sağlayarak önemli bir veri kaynağı sunmaktadır.

Gelecekte yapılacak çalışmalar için şu önerilerde bulunulabilir:

Daha Geniş Veri Setleri Kullanımı: Bu çalışmada kullanılan veri seti belirli bir zaman aralığını kapsamakta olup, daha geniş zaman dilimlerini kapsayan veri setleri ile daha kapsamlı analizler yapılabilir. Böylece, depremin uzun vadeli etkileri ve halkın tepkilerinin zaman içerisindeki değişimi daha iyi anlaşılabilir.

Farklı Metin Madenciliği Tekniklerinin Kullanımı: Bu çalışmada belirli metin madenciliği teknikleri kullanılmıştır. Gelecekteki çalışmalar, farklı teknikler ve algoritmalar kullanarak daha detaylı ve çeşitli analizler yapabilir. Özellikle makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak daha ileri seviye analizler gerçekleştirilebilir.

Sosyal Medya ve Diğer Veri Kaynaklarının Entegrasyonu: Sosyal medya verilerinin yanı sıra, diğer veri kaynaklarının (örneğin, haberler, resmi raporlar) entegrasyonu ile daha kapsamlı analizler yapılabilir. Bu, depremin etkilerinin daha geniş bir perspektiften değerlendirilmesine olanak tanır.

Dil Analizlerinde Gelişmiş Yöntemler: Çalışmada kullanılan dil analizi yöntemleri, farklı dillerdeki tweetlerin analiz edilmesine odaklanmıştır. Gelecekte, daha gelişmiş doğal dil işleme yöntemleri kullanılarak dil analizleri derinleştirilebilir ve dilsel farklılıkların etkileri daha ayrıntılı olarak incelenebilir.

Bu öneriler doğrultusunda yapılacak çalışmalar, sosyal medya verilerinin afet yönetimi, kriz iletişimi ve toplumsal olayların analizi konularında daha etkili ve kapsamlı bir şekilde kullanılmasına katkı sağlayacaktır. Bu çalışmalar hem akademik araştırmalar hem de pratik uygulamalar için önemli bilgiler sunarak, toplumsal dayanıklılığın artırılmasına ve afetlere karşı daha hazırlıklı olunmasına yardımcı olabilir.

Sonuç olarak, bu tez çalışması, 6 Şubat 2023 Türkiye depremi sonrası Twitter paylaşımlarının metin madenciliği yöntemleriyle analiz edilmesiyle önemli bulgular sunmuştur. Yapılan analizler ve elde edilen sonuçlar, sosyal medya verilerinin afet yönetimi ve kriz iletişimi süreçlerinde ne kadar değerli bir kaynak olduğunu göstermektedir. Gelecekte yapılacak çalışmalar, bu bulguları genişletmek ve daha kapsamlı analizler yapmak için önemli fırsatlar sunmaktadır.

5. KAYNAKÇA

Ayan, B., Can, M., & Gürsoy, U. T. (2019). Sosyal Medya Madenciliği İle Firmaların Twitter Verilerinin İncelenmesi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi UIİİD-IJEAS*, 2019,(23):237-256 ISSN 1307-9832.

Çobaner, A. A. (2019). Doğal Afetlerde Twitter Kullanımı: Mersin’de 2016 Yılı Sel Felaketi Örneği. *MER- AK Akademik Araştırmalar Sosyal Bilimler Kitap-1*, 41-67. Mersin: Mer-Ak Mersin Akademi Yayınları.

Eteman, G. S. (2014). NodeXL ile Sosyal Ağ Analizi: #akademikzam Örneği. 15. *Uluslararası Ekonometri, Yöneylem Araştırmaları ve İstatistik Sempozyumu Bildiriler Kitabı*, 464-482.

Mendoza, M., Poblete, B., & Valderrama, I. (2019). Nowcasting earthquake damages with Twitter. <https://doi.org/10.1140/epjds/s13688-019-0181-0>. *EPJ Data Sci.* 8(3), 1-23.

Doan, S., Ho Vo, B., & Collier, N. (2011). An analysis of Twitter messages in the 2011 Tohoku Earthquake https://doi.org/10.1007/978-3-642-29262-0_8, *International Conference on Electronic Healthcare*, 91, 58- 66.

Zheng, Z., Shi, H. Z., Zhou, Y. C., Lu, X. Z., & Lin, J. R. (2022). Earthquake Impact Analysis Based on Text Mining and Social Media Analytics. arXiv:2212.06765, <https://doi.org/10.48550/arXiv.2212.06765>.

Coronel, P., García, J., & Herrera-Tapia, J. (2019). Management of natural disasters based on Twitter analytics. 2017 Mexico Earthquake. *Explorations in Technology Education Research*, 3-12.

Doan, S., Ho Vo, B., & Collier, N. (2011). An analysis of Twitter messages in the 2011 Tohoku Earthquake https://doi.org/10.1007/978-3-642-29262-0_8, *International Conference on Electronic Healthcare*, 91, 58- 66.

Kirman, F. (2023). Deprem Sonrası Travmatik Stres Tepkileri ve Sosyal Medyada Yansımaları, *Antakiyat*, 6(1), 14-29.

Üçükkartal, H. (2020). Twitter'daki verilere metin madenciliği yöntemlerinin uygulanması. *ESTUDAM Bilişim Dergisi*, 1(2), 10-13.

Yıldırım, F. E. (2023). 6 Şubat 2023 Depreminde Türkiye’de Twitter’ın Kullanımı <https://doi.org/10.56133/intermedia.1354699>, 10(19), 276 – 292.

Aşan, H. (2023). 6 Şubat 2023 Kahramanmaraş Depremi Sonrası Afet Yönetimi Açısından Sosyal Medya Hesaplarının Güvenilirlik Analizi ve Değerlendirilmesi. *Akademik Yaklaşımlar Dergisi*, 15(1).

Uzun, C. (2023). 6 Şubat Depremi ve Twitter Duygu Analizi. *Türk Dili ve Edebiyatı Dergisi*, 63(2), 503-517.

Argın, Y. (2023). Doğal Afetlerde Sosyal Medya Kullanımı: 2023 Kahramanmaraş Depremi Özelinde Twitter Örneği. *Sanat Tasarım ve Mimarlık Araştırmaları Dergisi*, 3(1), 140-165.

Tarakçı, H. (2023). Afet İletişimi ve Twitter: Kahramanmaraş Depremi Özelinde Bir İçerik ve Duygu Analizi Araştırması. *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi*, 11(2).

Serin, E., Ünlü, S. (2024). Sosyal Medya Çağının Salgını Bilgi Dezenformasyonu: Twitter Üzerinden 6 Şubat Kahramanmaraş Depremi Hakkında Bir Değerlendirme. Volume. 4 – Issue. 1 p. 1-16.

Koçyiğit, N., Kocaoğlu, M. ve Yıldız, E. (2024). “Sosyal Medya Kullanımı ve Dijital Okuryazarlık: 6 Şubat Depremleri Temalı Twitter Paylaşımları Örneği”, *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, sayı 62, Denizli, ss. 227-238.

Tokcaer, S. (2021). Türkçe Metinlerde Duygu Analizi Sentiment Analysis in Turkish Texts. *Journal of Yasar University*, 16(63), 1514-1534.

WEB KAYNAKLARI

URL 1: https://tr.wikipedia.org/wiki/Veri_madenciliği

URL 2: <https://www.kaggle.com/datasets/swaptr/turkey-earthquake-tweets>

URL 3: <https://huseyindemirtas.net/kelime-bulutlu>

URL 4: [https://tr.wikipedia.org/wiki/Telegram_\(yaz%C4%B1%C4%B1m\)](https://tr.wikipedia.org/wiki/Telegram_(yaz%C4%B1%C4%B1m))

URL 5: <https://medium.com/fedeveloper/regex-hakk%C4%B1nda-2c4e80501802>

URL 6: <https://www.kaggle.com/code/sachinpatil1280/tweet-classification-machine-learning>

URL 7: <https://www.kaggle.com/code/yassinabdulelah/eda-turkey-and-syria-earthquake>

6. ÖZGEÇMİŞ

ESMAGÜL DENİZ

Phone Number: (506) 1496599

E-mail: denizesmagul@gmail.com

Address: Başakşehir / İstanbul

EDUCATION

Computer Engineering

Marmara University / 2020- Expected July 2024

Elazığ Anatolian High School / 2014-2018

EXPERIENCES

Software Test Engineer / Biletbank / February 2024 - present

Software Test Engineer Intern / Biletbank / July 2023 – January 2024

Sap Abap Developer Intern / Prodea Information Technologies / August 2022-October 2022