

(Araştırma Makalesi)

## Twitter Veri Kümesi ile Depresyon için Duyarlılık Analizi Özellik Çıkarma

Kübra ÖZTÜRK<sup>1,\*</sup>

<sup>1</sup>Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Eskişehir,  
ORCID No: <http://orcid.org/0000-0001-9372-884X>

**Anahtar Kelimeler:**

Depresyon,  
Metin Madenciliği,  
Twitter Veri Seti,  
Duygu Analizi.

**Özet:** Depresyon son yıllarda özellikle de sosyal medyanın yaygınlaşması ile gittikçe artan ve bir çok insanın iş, ev, sosyal yaşamını etkileyen mental bir sağlık problemidir. Erken tespitinin önemini Dünya Sağlık Örgütü tarafından sunulan istatistikler göstermektedir. Depresyonun küresel olarak birincil endişe nedeni olduğunu ve tespit edilmediği takdirde vakaların çoğunda intihara yol açtığı gözlemlenmiştir. Günümüzde sosyal medya kullanımı ile insanların fikirlerini metin, ifadeler, fotoğraflar veya videolar aracılığıyla ifade etmeleri ve böylece duygularında ve ruh hallerindeki değişimleri yansıtmaları ve tespit edilmesi için harika bir araç olmuştur. Bu çalışmanın amacı, kullanıcıların Twitter'daki kişiliğini ve gönderilerini incelemek ve çevrimiçi kullanıcıların depresif belirtilerini gösterebilecek nitelikleri belirlemektir. Bunun için anksiyete ve depresyon hastalarının yazımlarında en çok kullandıkları kelimeler tespit edilmiş ve bu kelimeler üzerinden analiz aşaması tamamlanmıştır.

(Research Article )

## Sensitivity Analysis Feature Extraction for Depression with Twitter Dataset

**Keywords:**

Depression,  
Text-Mining,  
Twitter,  
Sentimental-Analysis.

**Abstract:** Depression is a mental health problem that has increased in recent years, especially with the spread of social media, and affects many people's work, home and social life. Statistics presented by the World Health Organization demonstrate the importance of early detection that depression is the primary cause of concern globally and, if left undetected, leads to suicide in the majority of cases. Today, with the use of social media, it has become a great tool for people to express their ideas through text, emoticons, photos or videos, thereby reflecting and detecting changes in their emotions and moods. The purpose of this study is to examine the personality and posts of users on Twitter and to identify the qualities that may indicate depressive symptoms of online users. For this, the words that anxiety and depression patients use most in their writing were determined and the analysis phase was completed based on these words.

### 1. GİRİŞ

Son zamanlarda sosyal medya aracılığıyla ruhsal hastalıkların tespitine yönelik araştırma alanı, sosyal medya platformlarının popüleritesinin artması ve bu alandaki güncel çalışmaların sadece İngilizce verileri kapsamı ile heyecan verici bir konu olmaya başlamıştır [1]. Sosyal medyanın gelişmesiyle birlikte, biri Twitter olmak üzere giderek daha fazla platform kullanılabilir [2]. Twitter, gerçek zamanlı olarak başkalarıyla bilgi

paylaşmak için kullanılan bir sosyal medyadır. Sosyal medyada hikayelerini, görüşlerini ve etkinliklerini aktaran çok sayıda Twitter kullanıcısı olduğundan, bu bir bilgi verisi kaynağı olabilir; Twitter'ı olan bazı kişilerin farklı bir yöntemi kullanarak düşüncelerini ifade etmesi nedeniyle sosyal çalışmalarda kullanılmak üzere bir kamu duyarlılığı verisi kaynağıdır [3].

Depresyon, üzgün, boş veya sinirli ruh halinin tezahürü, normalde eğlenceli aktivitelere ilgi veya zevk kaybı, kilo değişikliği ve enerji eksikliği ile karakterize edilen bir duygudurum bozukluğudur. Majör depresif bozukluğun

tanı ölçütlerini karşılaması için depresif belirtilerin en az iki hafta sürmesi gerekir. Araştırmalar, depresyonun genel olarak yazı stili ve buna karşılık gelen dil kullanımı üzerinde bir etkisi olduğunu göstermektedir [4]. Ayrıca, sosyal medyadaki kullanıcı kişiliği de bize kullanıcının zihinsel durumu hakkında birçok bilgi sağlayabilir. Araştırmalar, akıl hastalığı olan kişilerin kişilik, ilgi alanları, değerler ve yaratıcılık açısından benzer olduğunu göstermektedir. Ek olarak, bazı belirtiler ve deneyimler akıl hastalığına özgüdür - ama hepsi değil [5].

Bu çalışmada Twitter'da depresyon ve kaygı için belirlenen kelimeler taranarak analiz için bir veri seti oluşturulacaktır. Oluşturulacak bu veri seti üzerinde depresyon için kullanılan kelimelerin kullanım sıklıklarına yönelik duygu analizi ve karşılaştırmalı analiz yapılacaktır. Kelimeler pozitif, negatif ve nötr olarak sınıflandırılıp kullanıcıların bunları en çok hangi yıl ay attığı gösterilecektir. Ayrıca kullanıcıların kaç beğenileri (likes) olduğu, hangi tweetleri alıntıladıkları (retweet) kaç arkadaş oranlarına göre bakılacaktır. Tweetlerdeki emojiler çıkarılıp negatif, pozitif ve nötr olarak sınıflara ayrılıp, sayıları özellik çıkarmak için kullanılmak üzere virgülle ayrılan değerler (csv) olarak kaydedilecektir [10].

## 2. MATERYAL VE METOT

Bu çalışma kapsamında sosyal medya platformlarından Twitter üzerinden elde edilen veri setini kullanarak depresyona eğilimli kullanıcılar için sınıflandırma algoritmalarını kullanıldı. Bu inceleme için text-preprocessing, NLP ve Text-blob kullanıldı.

### 2.1. Veri Toplanması

Verilerin elde edilmesi için öncelikle Twitter API kullanılarak oluşturulan "Mental-Health-Twitter.csv" kullanıldı ve veri tabanına kaydedildi. Veri tabanındaki kolonlar şu şekildedir:

"post\_id,post\_created,post\_text,user\_id,followers,friends ,favourites,statuses,retweets,label.Post\_created" kolonundan tarih objesi yaratılarak year ve month kolonları eklenmiştir [9].

### 2.2. Text-Preprocessing

Çalışmamızın temeli, veri setinin ön işlenmesi ile başlar. Ön işlemenin arkasındaki fikir, yalnızca kelime fazlalığı artıran ve duygu sınıflandırmasına hiçbir katkısı olmayan sözcükleri ve noktalama işaretlerini kaldırarak yapılıdır. Bu adımı gerçekleştirerek, genel çalışma kolay ve verimli hale gelir. Bu nedenle, ön işleme birincil bir görevdir.

Küçük harfe Dönüştürme: Veri kümesindeki tüm kelimeler küçük harfe dönüştürülür. İşlevsellik eksikliği nedeniyle, veri ön işleme adımlarının bir parçası olarak aşağıdakiler kaldırılmıştır.

- Gereksiz noktalama işaretleri
- Sayısal veriler

İfadelerin Değiştirilmesi: Farklı duyguları belirten ifadeler, anahtar kelimelerle değiştirilir, yani, Pozitif ifadeler- EPOS, Negatif ifadeler ENEG, Nötr ifadeler-ENEUT.

### 2.3.Stop Words

Durdurulacak sözcükler, veri kümesine anlamlı içerik eklemeyen sözcüklerdir (yani zamirler, edatlar, bağlaçlar vb.). Bu nedenle, bunların kaldırılması, eğitim ve test kümesindeki öğelerin alanını önemli ölçüde azaltır. Çalışmada stop-words için Python nltk kütüphanesindeki "stopwords" kullanılmıştır [5].

Belirteçleştirme (Tokenization): Belirteçleştirme, bir grup olarak ele almak istediğimiz bir dizi karaktere bir belirteç verme işlemidir. Metni belirteç olarak ele almak, özellik olarak kullanılacak belirteç sayıları oluşturmaya olanak tanır [6].

Stemming: Stemming, anlam bakımından birbirine yakın kelimeleri gruplama işlemidir. Amaç, fazlalığın azaltılabilmesi için kelimenin temel biçimini elde etmek için eki kaldırmaktır. Bu şekilde, bu tür farklı kelimeler tek bir terim altında gruplandırılabilirse, NLP geliştirilecektir.

### 2.4.Duygu Analizi

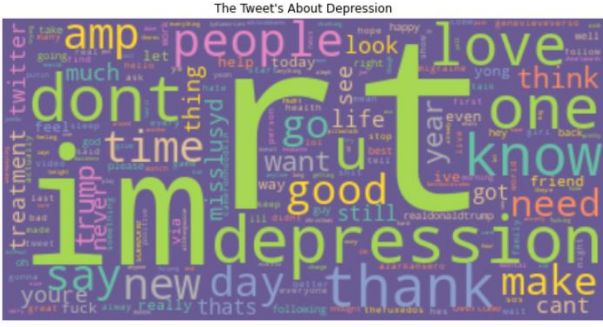
Metnin görüşü, olumsuz, olumlu veya nötr duygular içerip içermediğini belirlemek için analiz edilir. Duygu analizi veya fikir madenciliği, bireylerin veya sosyal grupların sahip olduğu fikirleri belirlemek için doğal dil işleme (NLP) ve makine öğrenimini kullanır [7].

## 3. BULGULAR VE TARTIŞMA

### 3.1 Work Cloud

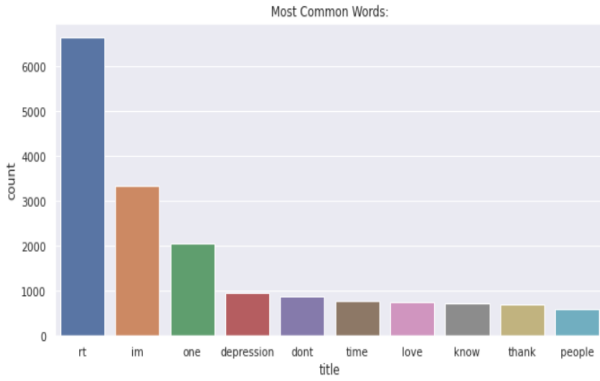
Bu araştırma kapsamında incelenen araştırma problemi tweetlerinin duygusal bütünlüğüne göre bir kişide depresyon teşhisi koymak mümkün mü sorusudur. Bu soruyu cevaplamak için farklı duygular değerlendirilecek ve bir her bir tweet için ağırlıklı puan hesaplanacak ve tek bir gün için ortalaması alınacak ve böylece nihai puan veya büyüklük oluşturulacaktır. Ayrıca depresyonu daha iyi anlamak için twitter tweetleri ile büyüklük arasındaki ilişki araştırılacaktır. Bu makalenin birincil amacı, son birkaç aydır tweet'lerine dayalı olarak twitter kullanıcılarının depresyona meyilli veya depresyonda olan kişileri tespit edebilmektir. Depresyonla alakalı atılan tweetlerden en çok kullanılanları kelime bulutu oluşturulup görselleştirilmiştir.

Oluşturduğumuz veri setlerini birleştirip text-blob kullanarak duygu analizi yaptığımızda depresyon için atılan tweetlerden pozitif ve negatif olarak dağılımı Şekil 7'de gösterilmiştir.

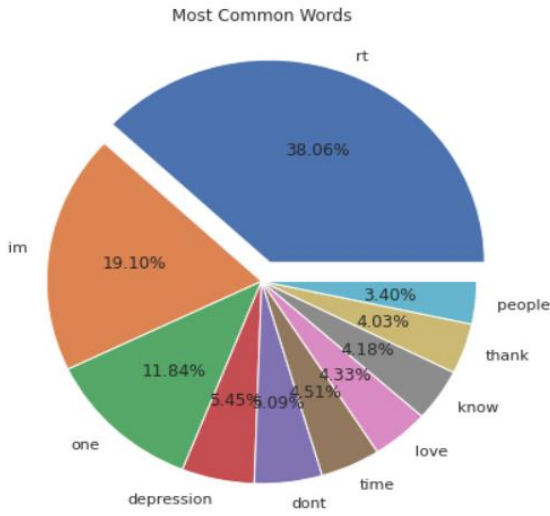


Şekil 1. Depresyon hakkındaki tweetler [2].

En çok kullanılan kelimeler ve sayıları;

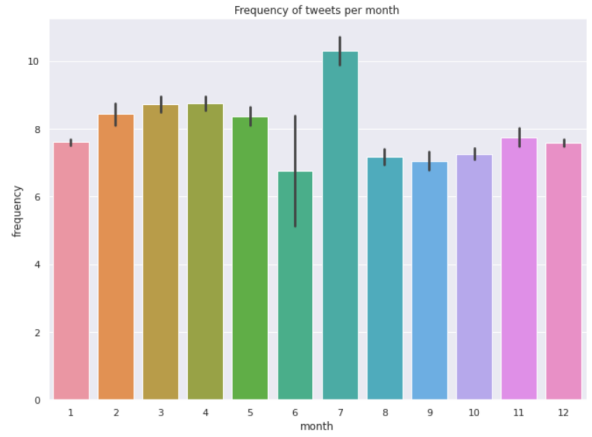


Şekil 2. En çok kullanılan kelimeler.



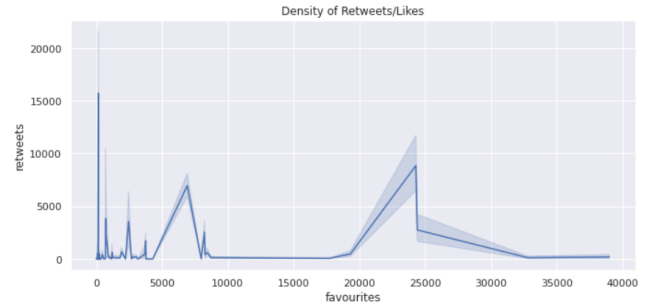
Şekil 3. En çok kullanılan kelimeler.

Atılan tweetlerin aylara göre sıklığı;

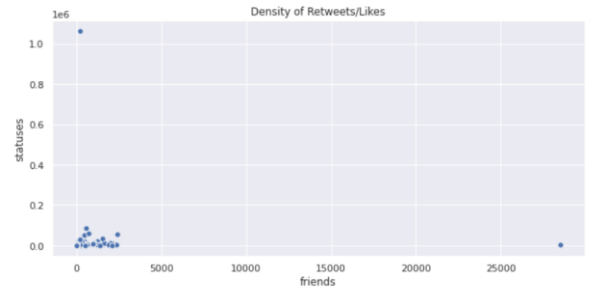


Şekil 4. En çok kullanılan kelimelerin aylara göre sıklığı.

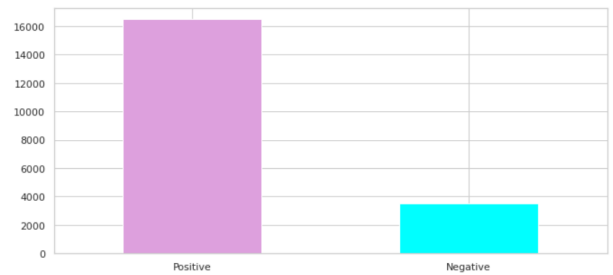
Retweet ve beğenme yoğunlukları;



Şekil 5. Retweet ve beğenme yoğunlukları.



Şekil 6. Retweet ve beğenme yoğunlukları.



Şekil 7. Tweetlerin pozitif ve negatif dağılımları.

#### 4. SONUÇ

Günümüzde insanlar günlük aktivitelerini sosyal medyaya yüklemeyi bunu diğer insanlarla paylaşmayı seviyorlar, biz de bu sayede depresyondan mustarip insanları belirleyebilir ve onların daha iyi ve müreffeh bir yaşam sürmelerine yardımcı olabiliriz çünkü akıl hastalığı toplum ve kendisi için zararlı olabilir [11][12].

Bu araştırmanın birincil amacı, twitter kullanıcılarının tweet'lerine dayalı olarak depresyon ile ilgili duygularını nasıl ifade ettiklerini, en sık hangi kelimeleri kullandıklarını, diğer kullanıcılarla etkileşimlerini incelemektir.

Gelecekteki çalışmalar açısından neden pozitif tweetlerin çok daha fazla çıktığı incelenebilir. Ayrıca uygulama depresyonda olabilecek kişilerin tespit edilerek onlara yardımcı olabilecek, psikolog gibi, kişilerin iletişim bilgilerinin önerilmesi olabilir.

#### Etik Hususlar

#### Etik kurallara uyum

Yazarlar etik görev ve sorumluluklara riayet edildiğini beyan ederler.

#### Finansman

Yazarlar kamu, ticari veya kar amacı gütmeyen sektörlerdeki fon kuruluşlarından özel bir hibe alınmadığını beyan ederler.

#### Çıkar çatışması

Yazarlar herhangi bir kişi veya kurumla çıkar çatışmasının bulunmadığını beyan ederler.

#### Teşekkür

Yazar, bu araştırmanın yaşam döngüsü boyunca sürekli desteği için akıl hocası ve rehber Dr. Eyüp Gülbandır'a teşekkür eder.

#### KAYNAKÇA

- [1] Wangler, T. P. 2008. RF Linear Accelerators. 2nd, completely revised and enlarged edition. WILEY-VCH Verlag GmbH&Co. KGaA, Weinheim, 450s.
- [2] Özyurt, B., Akçayol, A.A. 2018. Fikir Madenciligi ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. Selçuk Üniversitesi Bilim ve Teknoloji Dergisi, 6:4 (2018), 668-693.

- [3] Dani Castro, "Emoji Sentiment Perception between Readers and Writers, Jose Berengueres," IEEE International Conference on Big Data (BIGDATA), pp. 4321-4328, 2017.
- [4] Diveesh Singh and Alineen Wang, "Detecting Depression Through Tweets" Stanford University CA 9430, ;pp.1-9
- [5] Yair Neuman, Yohai Cohen, Dan Assaf, Gabbi Kedma, "Proactive screening for depression through metaphorical and automatic text analysis," Artificial Intelligence in Medicine, Vol. 56, No. 1, pp. 19-25, 2012.
- [6] De Choudhury, M., Counts, S., Horvitz, E.: Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference, WebSci 2013, New York, NY, USA, pp. 47–56. ACM (2013)
- [7] Çetin, F.S.2018. Türkçe Hedef Tabanlı Duygu Analizi İçin Alt Görevlerin İncelenmesi-Hedef Terim, Hedef Kategori Ve Duygu Sınıfı Belirleme. Bilişim Teknolojileri Dergisi, 11:1(2018), 43-55.
- [8] Arda, 2022 , Sentiment analysis of opinions about depression on twitter. <https://www.kaggle.com/code/ardawrld/twitter-sentiment-analysis-about-the-depression/notebook#Data-Visualization>(Erişim Tarihi:11.01.2023)
- [9] InFamousCoder, 2022, Depression: Twitter Dataset + Feature Extraction. <https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media> (Erişim Tarihi:11.01.2023)
- [10] InFamousCoder, 2022, Emoji-Sentiment-Features. <https://www.kaggle.com/code/infamouscoder/emoji-sentiment-features/notebook> (Erişim Tarihi:11.01.2023)
- [11] Hahn, T., Nierenberg, A. A., & Whitfield-Gabrieli, S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. Molecular psychiatry, 22(1), 37-43.
- [12] Giambattista Amati, Marco Bianchi, and Giuseppe Marcone. 2014. Sentiment estimation on twitter. In Proceedings of the 5th Italian Information Retrieval Workshop (IIR'14), Vol. 1127. CEUR Workshop Proceedings, 39—50
- [13] Association A. P. et al., Diagnostic and statistical manual of mental disorders (DSM-5®), American Psychiatric Pub, 2013.