**Duygu Analizi ve Metin Madenciliği Yöntemleri ile Twitter Verilerinin İncelenmesi**

**Semanur KONUR, Beyza Nil UYSAL, Engin BAŞ**

**Özet:** Çevrimiçi iletişimin gelişmesiyle birlikte insanların fikirlerini sosyal medya üzerinde paylaşması, sosyal medya verilerini önemli bir hale getirmiştir. Sosyal medya platformu olan Twitter’dan alınan düzensiz haldeki verilerden; kullanıma uygun ve düzenli veriler metin madenciliği yöntemleri ile sağlanmıştır. Bu çalışmada doğu coğrafyasında ve batı coğrafyasında gerçekleşen yakın dönem savaşları ve bu coğrafyalardan göç eden mülteciler ile ilgili twitter kullanıcılarının konuştuğu veriler toplanıp düzenlemiş ve duygu analizleri yapılmıştır, konular hakkında kullanıcıların hangi konulara yoğunluk verdiği konu modellemesi ile saptanmış ve iki farklı savaş diyalogları hakkında değerlendirmeler yapılmıştır.

**Anahtar Kelimeler — Twitter, duygu analizi, konu modellemesi, savaş,mülteci**

**Abstract:** With the development of online communication, people's sharing their ideas on social media has made social media data important. From the irregular data obtained from the social media platform twitter; Regular data suitable for use are provided by text mining methods. In this study, the data about the recent wars that took place in the eastern and western countries and the refugees who migrated from these geographies were collected and organized and sentiment analyzes were made.

**Keywords — Twitter, sentiment analysis, topic modeling, war,** **refugee**

# GİRİŞ

Günümüz dünyasında sosyal medya platformları insanların iletişim kurdukları, duygularını belirttikleri, reklam paylaşımlarını yaptıkları ve daha birçok etkileşimin sağlandığı potansiyeli oldukça yüksek olan, veri akışının sürekli olduğu ve insanların büyük kitlelere ulaşabildiği, dünyayı farklı gözlerden görebilme şansı tanıyan web siteleri ve uygulamalarıdır. Veri tabanlı teknikler ve analizler, dijital vatandaş hâline gelen bireyin tüm yaşamına dokunmakta [1] ve yaşam hızının arttığı dünyada toplumun yapıtaşını oluşturan iletişim, yoğun bir şekilde sosyal medyada akmaktadır. Bireyin buna adapte olmaya çalışmasının yanında “dijital özgürlük” kapsamında konu hakkında bilgisi fark etmeksizin yorum yapmaktadır. Sosyal bilimlerde veri toplamak ve verileri değerlendirmek geleneksel yöntemleri hızlı bir şekilde gerisinde bırakmaktadır. Twitter platformu üzerinden kullanıcıların paylaşmış oldukları metin verilerini anlamak ve analiz edilmesi için bu veriler noktalama işaretlerinden, tweet diyaloglarında geçen linklerden, özel karakter ve tekrar eden kelimelerden filtrelenmiştir. Böylelikle duygu analizi ve metin madenciliği için ön işleme adımları gerçekleştirilmiştir. Verilerin polarite puanlarını (pozitif/negatif/nötr) ve bu puanların normalleştirilip toplanması sonucu olan compound değerleri hesaplanmıştır. Bileşik puan, -1 (en negatif) ve +1 (en pozitif) arasında normalleştirilmiş derecelendirmelerinin toplamını hesaplayan bir metriktir. Bileşik puanın istatiksel sonucuna göre VADER tahmini oluşturulmuştur. Oluşturulan tahmin görselleştirilmiş ve modelin doğruluk puanı sklearn.metrics modülü ile hesaplanmıştır. Veri seti içerisinde her kelimenin sıklığını ve önemini göstermek amacıyla veri görselleştirme tekniklerinden biri olan Word Cloud kullanılmıştır. Konu modellemesi için öncelikle içerisinde İngilizce dilinde yaygın olarak kullanılan sözcüklerin bulunduğu bir “Stop Words” dosyası [2] oluşturduğumuz veri setinde ayıklanmıştır. Toplanan veriler içerisinde kelimenin önemini değerlendirmek için matematiksel olarak terim sıklığı ve ters doküman sıklığı (TF-IDF) metrikleri kullanılmıştır. Denetimsiz öğrenmenin bir şekli olan konu modelleme tekniklerinden biri olan Latent Dirichlet Allocation (LDA) uygulanmıştır. Bunlar sonucunda veri setinde kullanılan kelimeler modellenmiştir. Sonuçlar doğrultusunda insanların doğu ve batı coğrafyasında olan savaşlar ve bu coğrafyalardan göç eden mülteciler hakkında konuştukları veriler duygu analizi ve konu modellemesi ile değerlendirilmiştir.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil : Vader Modeli Hesaplama Örneği

1. **MATERYAL VE METOT**
   1. **Duygu Analizi**

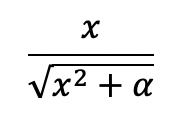
İnternet kullanımın yaygınlaşmasıyla insanlar, internet üzerinden iletişim kurabilecekleri Facebook, Instagram, Twitter gibi sosyal platformlar oluşturdu. İnternet ile paralel olarak kullanımı hızla yaygınlaşan sosyal platformlar, insanların, duygularını saniyeler içerisinde büyük kitlelere iletmesinde büyük bir katkı sağlamıştır. Bu platformlar, kullanıcılarının paylaştıkları veriler sayesinde sektör için büyük bir veri havuzu durumuna geldiler. Paylaşılan veriler sadece bilgi değil, veriyi oluşturan kullanıcıya ait duyguları da içermektedir. Verilerin analiz edilebilmesi, bilgisayarların bu verileri anlayabilmesi için “doğal dil işleme” (NLP) yöntemleri ortaya çıkmıştır [3]. Bu çalışmada metinde saklanan duyguları, doğal dil işleme yöntemlerinden biri olan “duygu analizi” (sentiment analysis) kullanarak incelemelerde bulunduk.

Literatürde duygu analizi, fikir madenciliği, duygu madenciliği gibi isimlerle de telaffuz edilmektedir. Duygu analizinin görevi temel olarak 3 başlık altında incelenebilir: özelliği belirleme, yönelimi belirleme ve yönelimin gücünü belirleme [4]. Metin analizi ile birlikte kullanıcılara ait veriler işlenir ve içerisindeki duygu çıkarılır. Ardından verilerin duygu analizi yapılarak duygu tahmin edilir. Çeşitli duygu analizi algoritmaları vasıtasıyla metin içerisindeki duygunun olumlu veya olumsuz düşünce olduğu veya herhangi bir duygu belirtmediği kanaatine varılır. Nihai sonuca ulaşabilmek adına metindeki verilerin, önceden hazırlanmış, ifade edilen duygunun gücüne göre puanlandırılmış “duygu analizi sözlüğü” ile karşılaştırılması gerekmektedir. Duygu analizi sözlüğü, analiz sonuçlarını değerlendirebilmek adına önceden hazırlanmış, kelimelere ve kelime guruplarına karşılık gelen “duygu puanlarının” bulunduğu sözlüktür [5]. Metin içerisindeki kelimelerin veya kelime guruplarının, sözlük içerisinde karşılık geldiği puanlar toplanarak, metin içerisindeki duygunun nihai sonucunu elde edilir. Elde edilen sonuçlar, algoritmalarda genellikle -1 ile +1 arasında puanlandırılır. Bu çalışmada, duygu analizi için, metnin polaritesine (kutbuna) ve metnin duygu yoğunluğuna duyarlı olan “Vader Sentiment Analysis” kullanıldı.

* + 1. **Vader**

Vader (Valence Aware Dictionary for Sentiment Reasoning), duygunun hem polaritesine (pozitif – negatif) hemde duygu yoğunluğuna duyarlı olan, duygu analizi için kullanılan bir modeldir [6]. 2014 de tanıtılan Vader; özellikle sosyal medya gönderilerinde ifade edilen duyguları ayrıştırmak ve tanımlamak için ayarlanmış insan merkezli bir yaklaşım kullanır. Vader duygu analizi, sözcüksel özellikleri (lexical feature), Sentiment Score (duygu puanı) adı verilen duygu yoğunluklarıyla eşleyen bir sözlüğe dayanır. Metindeki Sentiment Score, metindeki her kelimenin duygu yoğunluğu puanının toplanmasıyla elde edilir.

Lexical feature (sözcüksel özellik) ile kastedilen, mesajlaşma dilinde kullandığımız; emojileri, argo kelimeleri, kısaltmaları, büyük-küçük harf kullanımı dahil bütün kelimeler Vader duygu analizinde yoğunluk değerleri ile eşleştirilir. Metnin içerisindeki kelimeler, +4 ila -4 arasında puan alır. +4 en olumlu, -4 en olumsuz, 0 ise nötr duyguyu temsil eder. Puanlama işlemi yapılan kelimelerin puanları, cümle bazlı toplanarak elde edilen sonucun normalizasyon işlemi yapılır ve cümle, +1 ila -1 arasında bir duyarlılık puanı döndürür. Burda +1 olumlu, -1 olumsuz, 0 ise nötr duyguları temsil eder. Normalizasyon işlemi formül şekil 2’de gösterilmiştir.

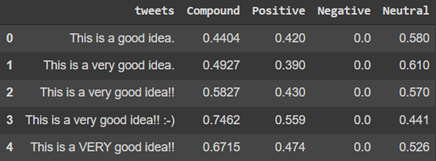


Şekil 2:Normalizasyon İşlemi Formülü

X=Cümleyi oluşturan kelimelerin duygu puanlarının toplamı

α=Normalizasyon parametresi

Duygu sözlüklerinin yanı sıra, doğası gereği tarafsız olan ancak duygunun kutupluluğunu değiştirebilen ("değil" ve "ama" gibi) veya tüm cümlenin yoğunluğunu değiştirebilen ("çok" ve "aşırı" gibi) yapılar vardır. Vader' da geliştiriciler, noktalama işaretleri, büyük harf kullanımı, zarflar ve karşıt bağlaçlar, kısaltmalar ve benzeri durumları ele alan birkaç buluşsal kural ekledi. Bu kuralların cümleye olan etkisi Tablo 1’de gösterilmiştir.



Tablo : Dilbilgisi kurallarının duygu yoğunluğu üzerindeki etki

Tablo 1’de görüldüğü üzere “This is a good idea.” Twteet’i üzerinde dil bilgisi kurallarını test edebilmek, tweet’in pozitiflik puanını arttırmak için aynı tweet’in 4 farklı türünü, üzerinde ufak değişiklikler yaparak teste tabi tuttuk. Test sonucu duygu yoğunluğunu arttıran “very” kelimesinin, “!” işaretinin kullanımının, gülümsemeyi temsil eden “:-)” emojisinin ve büyük harf kullanımının, total puanını (Compound) pozitife yaklaştırdığı sonucuna varılmıştır.

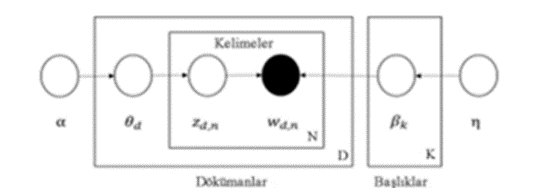
* 1. **Konu Modelleme**

Günden güne daha çok büyüyen veri ile birlikte bu verinin işlenmesi ve analiz edilmesi ihtiyacı da artmaktadır. Bu ihtiyacı karşılamak için en çok kullanılan yöntemlerden birisi metin madenciliğidir. Metin Madenciliği (Text Mining), yapısal olmayan ve düzensiz yapıdaki elektronik metin yığınlarından; önceden bilinmeyen, potansiyel olarak kullanışlı, yapısal ve düzenli veri elde etme sürecidir. Elde edilen bilgiyle, analiz edilen metin kaynaklarında açık olarak görülmeyen ilişkiler, hipotezler ve eğilimler tespit edilir [7]. Metin Madenciliğinin metinlerin sınıflandırılması, gruplandırılması, kategorilere ayrılması, duygu analizi ve konu modelleme gibi alanları bulunur. Metin Madenciliği, doküman koleksiyonlarının önişlemden geçirilmesi, ara sonuçların saklanması, ara sonuçların analiz edilmesi için çeşitli tekniklerin kullanılması ve nihai sonuçların görselleştirilmesi gibi aşamalardan oluşmaktadır [8].

Konu Modelleme metinler içerisindeki konuları saptamak için kullanılan bir çeşit istatiksel modeldir. Konu modelleme gruplandırma yapması yönüyle makine öğrenmesi tekniklerinden olan gözetimsiz (unsupervised) öğrenmeye dahil edilir. Bu projede yapılan araştırmalar sonucu en doğru sonucun alınabileceği düşünülerek LDA konu modelleme yöntemi kullanılmıştır. LDA uygulamak için elimizdeki veriler bir takım ön işlemlerden geçirildi.

İlk olarak veriler üzerinde stop words işlemi uygulandı. Bu yöntem veri setinin içinden ait olduğu dile dair en yaygın kullanılan etkisiz kelimeleri (bağlaçlar, zamirler, çoğul ekleri, soru ekleri vb.) filtreleme işlemidir. Daha sonra herhangi bir kelimenin veri içindeki önemini belirlemek için kullanılan TF-IDF algoritması kullanıldı. TF (Terim Sıklığı) değeri kelimenin doküman içinde geçme sayısının toplam kelime sayısına bölünmesiyle elde edilir. IDF (Ters Doküman Sıklığı) değeri ise toplam doküman sayısının, kelimenin bulunduğu doküman sayısına oranının logaritması alınarak elde edilir. Algoritmanın sonucunda ise elde edilen bu iki ayrı değer çarpılarak kelimenin ağırlık değeri hesaplanır.

Veriler üzerinde daha çok bilgiye sahip olmak için konu modelleme yöntemlerinden birisi olan LDA projeye dahil edildi. Bu, dokümanın ilişkili olduğu gizli anlamsal yapıları olasılıksal olarak ortaya çıkaran bir yöntemdir. Latent Dirichlet Allocation (Gizli Dirichlet Ayırımı) Algoritmasının temel amacı doküman içindeki kelimelere karşın o dokümanın konusunu bulmaktır. Bu işlem sırasında sadece dokümanlar gözlenebilir durumdadır ve Şekil 3’te siyaha boyanmıştır. Kelimelerin konularla eşlenmesi ve konulardaki dağılımları, konuların dokümandaki dağılımları ise gizlidir ve Şekil 1’de beyaz ile gösterilmiştir. LDA kelime torbası yaklaşımına dayalı çalışan denetimsiz bir yöntemdir. Kelime torbası yaklaşımında kelimelerin nerelerde bulunduğu göz ardı edilirken birlikte kullanılan kelimeler kullanılmaktadır.



Şekil :LDA Düzlemsel Gösterim [9]

Şekil 3’ de K kutusu konu başlık sayısını, D kutusu doküman sayısını ve N kutusu da dokümanda bulunan benzersiz toplam kelime sayısını temsil etmektedir. W\_(d,n) dokümandaki kelimeleri, z\_(d,n) ise her kelimenin bağlı olduğu konu başlığını gösterir. Bu gösterimdeki n kelimeyi (∀𝑛=1, …, N), d(∀d=1, …, N) ise dokümanı temsil etmektedir. Her dokümanın konu başlıklarına dağılımı θ\_d ile bu konu başlıklarının içeriği de β\_k ile temsil edilmiştir. β\_k, θ\_d, z\_(d,n) önceden bilinmeyen, W\_(d,n) bilinen değişkeninin işlenmesi sonucu elde edilen değerleri temsil etmektedir. 𝜂 ve α parametreleri ise β\_k ve θ\_d, parametrelerinin önceki değerlerini temsil etmektedir [10].

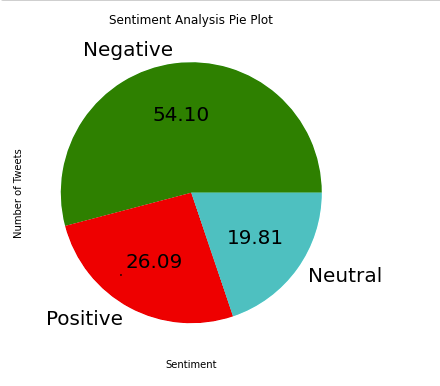
Gizli konuları bulma uygulamalarında hiper-parametre olarak α ve β değerleri önceden verilir. Bu parametrelerin alacağı değerler her bir konunun dağılımını ve yoğunluklarını belirler. α doküman-alt başlık yoğunluğunu belirlerken, β ise kelime-alt başlık yoğunluğunu belirler. Başka bir deyişle, yüksek α değeri fazla sayıda başlık bulabilirken, düşük α değeri daha az alt başlık bulur. Benzer şekilde, yüksek β değeri, bir başlığın fazla sayıda kelimelerden oluştuğunu gösterirken, düşük β değeri bir başlığın daha az kelimelerden oluştuğunu gösterir. Bu parametreler simetrik veya asimetrik dağılımla giriş değerleri verilebilir. Yapılan çalışmalara göre asimetrik α ve simetrik β parametreler daha iyi sonuç vermektedir [11].

1. **DENEYSEL SONUÇLAR**

Bu çalışma, dünya tarihi boyunca insanların farklı amaçlar doğrultusunda sonuçları ağır olmasına rağmen hiç durmadan gerçekleştirdikleri ve yakın dönemde iki farklı bölgede yaşanmış olan savaş konusunu sosyal medya platformu twitter üzerinde söylenen konuşmaların içerisinde geçen belirli kelimelerin duygu analizi gerçekleştirilmiştir. Twitter’da; Batı savaşı olarak nitelendirdiğimiz “Ukraine-Russia”,”Putin - Zelenski”, ”Russian-Ukrainian” kelimeleri bir grup, Doğu savaşı olarak nitelendirdiğimiz “Israel-Palestine”, ”Afghanistan-taliban”, ”Syrian” kelimeleri ise farklı bir grup olarak ele alınmış ve bu iki grup için ele alınmış ülke kelimelerinin mülteci kombinasyonu kelimelerini içeren tweetler her grup için kazınmıştır. Doğu ve batı savaşları olarak ayırdığımız iki veri seti için de 20.000, doğu ve batı coğrafyasından göç eden mülteciler için de 7000 adet tweet çekilmiş, veriler düzenlenmiş ve konuşmalar içerisinde geçen duygular pozitif, negatif ve nötr olarak kutuplaştırılmıştır. Veri seti konu modellemesi teknikleri ile modellenmiş ve sonuçları dokümante edilmiştir.

* 1. **Batı Savaşı Verileri ile Elde Edilen Duygu Analizi**

Batı savaşı tweetlerinin yüzdelik dağılımları Şekil 4’te verilmiştir. Analiz sonucunda konuşmaların duygu yoğunluğu %54,10 oranında negatif, %26,09 oranında pozitif ve %19,81 oranında ise nötr olmuştur.



Şekil : Batı savaşı tweetlerinin yüzdelik dağılımları

Batı savaşı tweetleri Word Cloud sonuçları şekil 5’te verilmiştir.

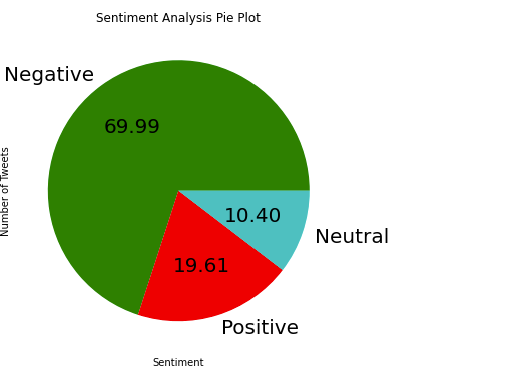
metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil : Batı savaşı tweetleri Word Cloud sonuçları

* 1. **Doğu** **Savaşı Verileri ile Elde Edilen Duygu Analizi**

Doğu savaşı tweetlerinin yüzdelik dağılımları Şekil 6’da verilmiştir. Analiz sonucunda konuşmaların duygu yoğunluğu %69,99 oranında negatif, %19,61 oranında pozitif ve %10,40 oranında ise nötr olmuştur.



Şekil : Doğu savaşı tweetlerinin yüzdelik dağılımları

Batı savaşı tweetleri Word Cloud sonuçları şekil 7’ de verilmiştir.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil : Batı savaşı tweetleri Word Cloud sonuçları

* 1. **Her İki Savaş Verisinin Karşılaştırmalı Analizleri**

Batı coğrafyasında gerçekleşen savaş terimleri ile taranmış olan veriler ve doğu coğrafyasında gerçekleşen savaş terimleri ile taranmış olan verilerin duygu analizinin yüzdelikleri Tablo 2’de kıyaslanmıştır.

tablo içeren bir resim

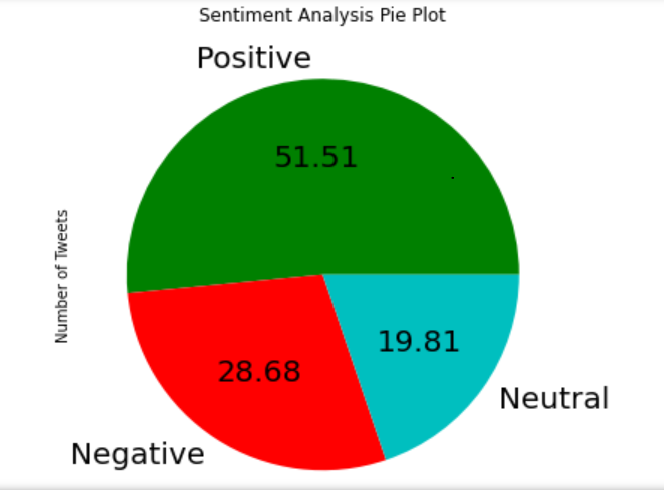
Açıklama otomatik olarak oluşturuldu

Tablo : Karşılaştırma

Tablo 2 incelendiğinde batı coğrafyasında gerçekleşen savaşlar hakkında yapılan konuşmaların duygu oranları %26,09 olumlu, %19,81 nötr ve %54,10 oranında olumsuz düşünceler olduğu analiz edilmiştir. Ancak Doğu coğrafyasında gerçekleşen savaşlar hakkında veriler incelendiğinde olumlu duygular %19,62, nötr düşünceler %10,40 oranına gerilerken olumsuz düşünceler %69,99 oranına yükselmiştir.

* 1. **Batı Kökenli Mülteci Verileri ile Elde Edilen Duygu Analizi**

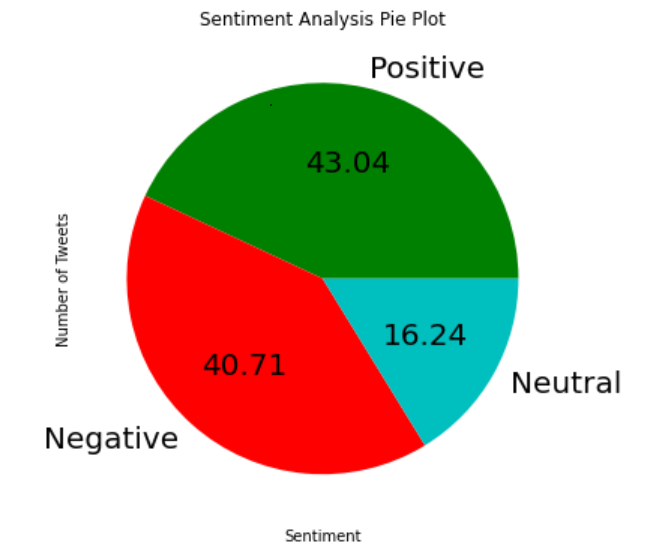
Batı kökenli mülteciler hakkında bulunan tweetlerin yüzdelik dağılımları Şekil 8’ de verilmiştir. Analiz sonucunda konuşmaların duygu yoğunluğu %28,68 oranında negatif, %51,51 oranında pozitif ve %19,81 oranında ise nötr olmuştur.



Şekil : Batı kökenli mülteciler hakkında bulunan tweetlerin yüzdelik dağılımları

* 1. **Doğu Kökenli Mülteci Verileri ile Elde Edilen Duygu Analizi**

Doğu kökenli mülteciler hakkında bulunan tweetlerin yüzdelik dağılımları Şekil 9’ da verilmiştir. Analiz sonucunda konuşmaların duygu yoğunluğu %40,71 oranında negatif, %43,04 oranında pozitif ve %16,24 oranında ise nötr olmuştur.



Şekil : Doğu kökenli mülteciler hakkında bulunan tweetlerin yüzdelik dağılımları

* 1. **Her İki Mülteci Verisinin Karşılaştırmalı Analizleri**

Batı kökenli mülteci terimleri ile taranmış olan veriler ve doğu kökenli mülteci terimleri ile taranmış olan verilerin duygu analizinin yüzdelikleri Tablo 3’te kıyaslanmıştır.

tablo içeren bir resim

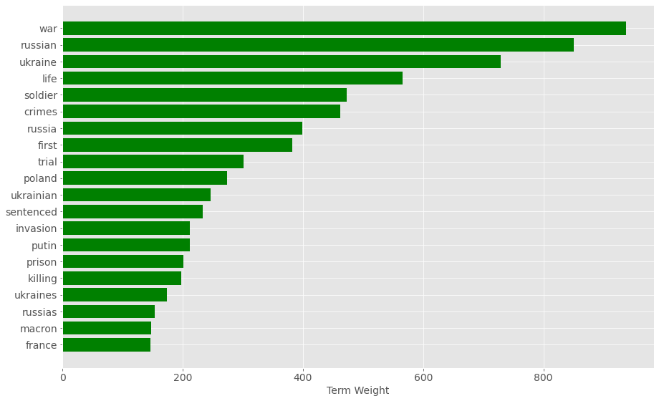
Açıklama otomatik olarak oluşturuldu

Tablo :Karşılaştırma

Tablo 3 incelendiğinde batı kökenli mülteciler hakkında yapılan konuşmaların duygu oranları %51,51 olumlu, %19,81 nötr ve %28,68 oranında olumsuz düşünceler olduğu analiz edilmiştir. Ancak Doğu coğrafyasında gerçekleşen savaşlar hakkında veriler incelendiğinde olumlu duygular %43,04, nötr düşünceler %16,24 oranına gerilerken olumsuz düşünceler %40,71 oranına yükselmiştir.

* 1. **Batı Savaşı Verileri Elde Edilen Konu Modellemesi**

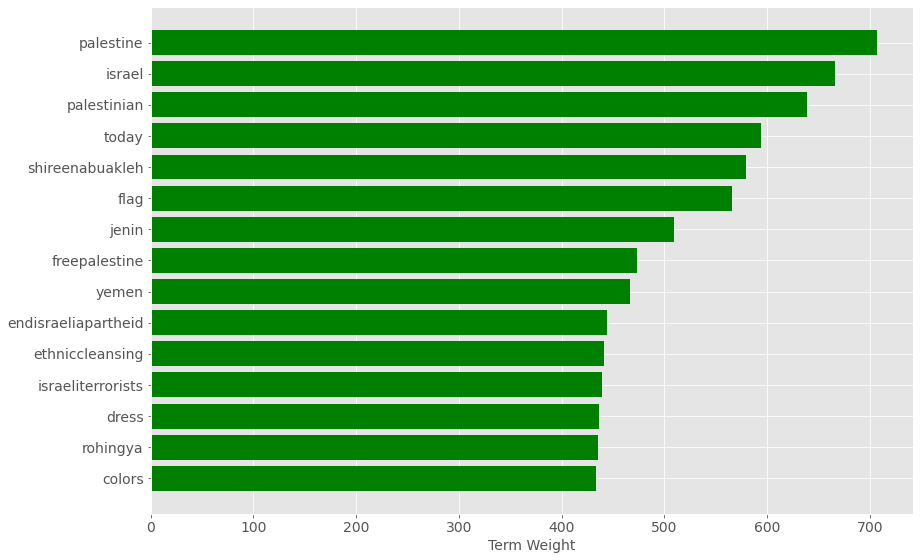
Batı savaşı tweetlerinden oluşan veri seti içerisinden en alakalı ve ağırlığı yüksek olan kelimeler şekil 10’da hesaplanmıştır.



Şekil : Batı Savaşı Verileri Elde Edilen Konu Modellemesi

1. **Doğu savaşı verileri elde edilen konu modellemesi**

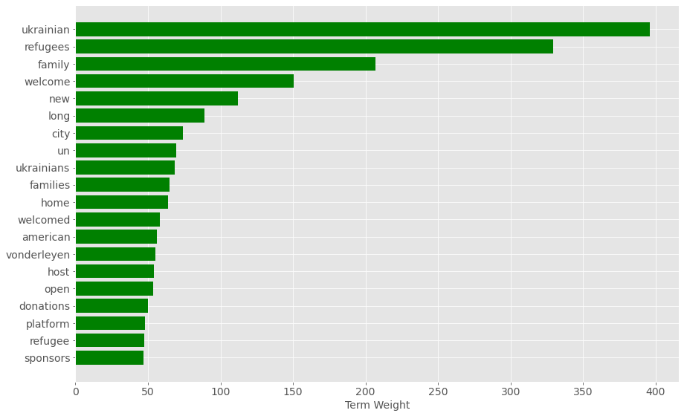
Doğu savaşı tweetlerinden oluşan veri seti içerisinden en alakalı ve ağırlığı yüksek olan kelimeler şekil 11’de hesaplanmıştır.



Şekil : Doğu savaşı verileri elde edilen konu modellemesi

1. **Batı kökenli mülteci verileri ile elde dilen konu modellemesi**

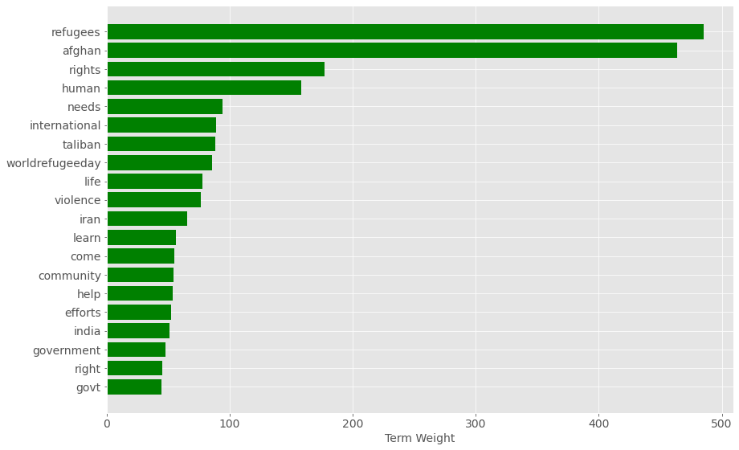
Batı kökenli mülteci tweetlerinden oluşan veri seti içerisinden en alakalı ve ağırlığı yüksek olan kelimeler şekil 12’de hesaplanmıştır.



Şekil : Batı kökenli mülteci verileri ile elde edilen konu modellemesi

**3.10 Doğu kökenli mülteci verileri ile elde edilen konu modellemesi**

Doğu kökenli mülteci tweetlerinden oluşan veri seti içerisinden en alakalı ve ağırlığı yüksek olan kelimeler şekil 13’te hesaplanmıştır.



Şekil : Doğu kökenli mülteci verileri ile elde edilen konu modellemesi

1. **SOSYAL MEDYADA BATI VE DOĞU TOPLUMLARININ YANSIMASI**

Bu çalışmada, toplumun batı ve doğu coğrafyalarında gerçekleşen savaş durumlarında sosyal medyanın savaş durumu ve savaş içerisindeki toplumlar hakkında görüşleri ve bu görüşlerin duygu durumu hakkında araştırma yapılmıştır.

Savaş hakkında bulunan duygu analizi sonuçlarında görüldüğü üzere sosyal medya platformu konuşmalarında batı savaşı hakkında diyalogların duygu analizi sonuçları %54,10 olumsuz iken doğu coğrafyasındaki savaş diyalogları %69,99 oranında olumsuz sonuçlanmıştır. Doğu ile ilgili sosyal medyada sunulan haberler, konuşmalar, yorumlar ve eleştiriler genellikle olumsuz olmakla birlikte “öteki” konumundaki doğu bireyi savaş durumunda iken bile konu modellemesi sonucunda görüldüğü üzere “etnik temizleme” yorumu yüksek oranda konuşulmuştur.

Yaratılıştan gelen korunma güdüsü gereği insan, kendisine yarar sağlayacağını düşündüğü şeylerden hoşnut olup, daha az yararlı olacağı hatta zararı gelebileceğini düşündüğü şeylerden uzaklaşır, onlarla arasına mesafe koyar. [12] Kendine ait olduğunu düşündüğü alanı ve beraber yaşamaya alıştığı toplumu benimseyen ve “öteki” konumunda kalan bireyleri Gelişmiş ülkeler ve geri kalmış ülkeler olarak sınıflandırır. Geri kalmış ülkelerden gelen sığınmacıları potansiyel bir tehlike olarak görür ve toplumsal yaşamda bireyleri ötekileştirir. Duygu analizi sonucunda görüldüğü üzere mültecinin kökeni batı iken olumsuz duyguların oranı %28,68, mültecinin kökeni doğu iken olumsuz duyguların oranı %40,71 oranına yükseldiği nötr ve olumlu duyguların ise azaldığı görülmüştür.

Mülteci bireyler hakkında oluşturulan veri setinin konu modellemesi ile en sık kullanılan kelimeler incelendiğinde batı veri seti içerisinde “aile” ve “hoş geldiniz” kavramları dikkat çekmiştir. Aynı konu üzerinden oluşturulan doğu sığınmacılarını ele aldığımız veri setinde bu gibi kelimeler görülmemiştir.

Sosyal medya üzerinden söylenen ırkçı ve yanlış söylemler istemeden tarihsel güç dengesizliklerini yeniden üretebilir [13].

Rönesans döneminde yaşayan felsefeci Erasmus (2009: 116-117) Thomas More’a yazdığı “Deliliğe Övgü” kitabında “İngilizler, başka şeyler bir yana, özellikle güzellik, müzik ve kaliteli yemek gibi konularda eşsiz olduklarını düşünürler. İskoçyalılar soyluluklarıyla, kraliyet unvanlarına sahip olmanın ayrıcalığıyla ve bir o kadar da diyalektik tartışmalarıyla övünürler..” söyleminde bulunmuştur. Her milletin kendi milletini diğer tüm toplum milletlerinden üstün görerek övündüğünü belirtmiştir. Geri kalmış toplum olarak sınıflandırılan toplumlar bir de sığınmacı konumuna geldiğinde önyargı ve nefret söylemlerinin yayılması ile birlikte güçsüz olarak bakılan toplum daha güçsüz ve saldırıya açık hale gelmektedir. Sosyal medya üzerinden yapılan yorumlar çok hızlı yayılmakta olduğundan yansıttığı duyguların da bulaşıcı olduğu kaçınılmazdır.

1. **SONUÇ**

Bu çalışma, yakın dönemde batı coğrafyası ve doğu coğrafyasında gerçekleşen savaşlar ve bu coğraflardan göç eden mülteciler hakkında sosyal medya platformu olan twitter’dan kullanıcıların tweetleri ile oluşturulmuş iki farklı veri seti üzerinden duygu analizi ve konu modellemesi yapılmıştır. Veriler eğitildikten ve sonuçlar alındıktan sonra duygu analizi modeli olan VADER modelinde sklearn metrics ile batı tweetlerinin analizinde doğruluk puanı şekil 5’te, doğu tweetlerinin analizinde doğruluk puanı şekil 6’da gösterilmiştir.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil : Batı tweetlerinin analizinde doğruluk puanı

metin, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil : Doğu tweetlerinin analizinde doğruluk puanı

**KAYNAKÇA**

[1] Türk, M. S. (2021). Büyük Veri ve Değişim. TRT Akademi, 6 (11), 5-9. Retrieved from https://dergipark.org.tr/tr/pub/trta/issue/60117/870962

[2] [What are Stop Words? - Opinosis Analytics (opinosis-analytics.com)](https://www.opinosis-analytics.com/knowledge-base/stop-words-explained/#.YrXBokbP3EY)

[3] Seker, S. E. 2016. Duygu Analizi (Sentimental Analysis). YBS Ansiklopedi, cilt 3, sayı3 (09/16) [2] Lou T., Xu G. (2013). Sentiment Analysis. DOI:10.1007/978-1-4614-7202-5\_4 In book: Trust-based Collective View Prediction

[4] [2] Lou T., Xu G. (2013). Sentiment Analysis. DOI:10.1007/978-1-4614-7202-5\_4 In book: Trust-based Collective View Predictio

[5] Albayrak M., Topal K., Altıntaş V. (2017). SOSYAL MEDYA ÜZERİNDE VERİ ANALİZİ: TWITTER. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi Y.2017, C.22, Kayfor15 Özel Sayısı

[6] C. J. H. E. Gilbert, “Vader: A parsimonious rule-based model for sentiment analysis of social media text,” in Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16)

[7] http://www.metinmadenciligi.com

[8] H. K. Küçükkartal, "Twitter'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması", Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi, c. 1, sayı. 2, ss. 10-13, Haz. 2020

[9] D.M. Blei, A.Y. Ng, M.I. Jordan, “Latent Dirichlet Allocation”, Journal of Machine Learning Research, 3: 993-1022, 2003.

[10] V. Altıntaş , M. Albayrak ve K. Topal , "Kanser hastalığı ile ilgili paylaşımlar için Dirichlet ayrımı ile gizli konu modelleme", Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, c. 36, sayı. 4, ss. 2183-2196, Eyl. 2021, doi:10.17341/gazimmfd.734730

[11] Wallach, H. M., Mimno, D., ve McCallum, A., Rethinking LDA: Why priors matter, Advances in Neural Information Processing Systems 22-Proceedings of the 2009 Conference, Vancouver- Kanada 1973– 1981, 7-10 Aralık, 2009.

[12] Journal of Humanities and Tourism Research 2020, 10 (2): 275-288

[13] Matamoros-Fernández A, Farkas J. Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*. 2021;22(2):205-224. doi:[10.1177/1527476420982230](https://doi.org/10.1177/1527476420982230)