

Detection of Islamophobic Tweets on Twitter Using Sentiment Analysis

 Buğra AYAN^{1,*} Birol KUYUMCU² Bünyamin CİYLAN³
¹Turkish Presidency Bestepe / Ankara

²SefaMerve R&D Center , 34854, Maltepe / İstanbul

³Gazi University Faculty of Technology, Department of Computer Engineering, 06500, Yenimahalle/ANKARA

Graphical/Tabular Abstract

Article Info:

 Received:
 8/05/2019
 Revision
 13/06/2019
 Accepted:
 17/06/2019

Highlights

- Machine Learning on Twitter
- Islamophobia on Twitter

Keywords

 Sentiment analysis,
 Islamophobia,
 Twitter, Machine Learning

In this study, Tweets on Twitter are sought to be found out whether they are Islamophobic or not by using emotion analysis. Estimations are made using precision, recall and F1 measures via models trained with linear ridge model. In the end, accurate results, in the range of 96.3 per cent to 96.5 per cent on average, are obtained for positive tweets.

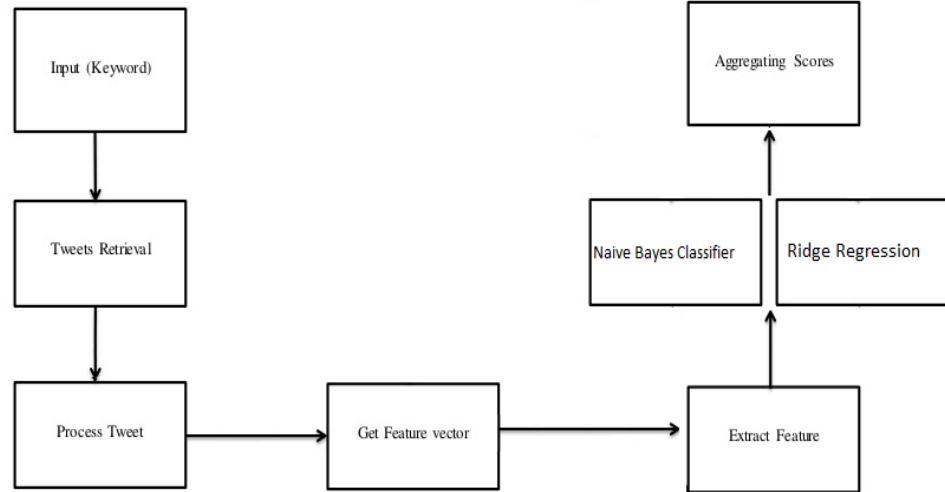


Figure A. Block schema of the system

Purpose: Although there are many studies on machine learning in social networks, a study based on Islamophobia, especially a study that brings a machine learning approach to the problems about this subject, has not been done scientific research. In this thesis, it is aimed to carry out a study on the detection of hate speech on Twitter on Islamophobia with different machine learning algorithms.

Theory and Methods: Twitter has been marked firstly in our study. Then, preprocessing studies have been performed and trained with two different algorithms.

Results: The system has been trained using 80% of the data set consisting of 162,000 tweets and tests have been performed using 20%. In this way, the accuracy rate of the system has been determined. After the application of word-based TF-IDF weighting on the data set, Ridge Regression and Naive Bayes model has been applied. For these models, the rates obtained in Tweet classification study were 96.9% in Ridge Regression in F1 criterion while this value has been decreased to 95.4% in the Naive Bayes. In addition, Ridge Period was shorter than Naive Bayes Classifier as training time.

Conclusion: Suggestions that can be put forward as a result of this study are that; researches should be done with similar national language for different languages, too and tweets should be researched and got more detailed results based on location and date based. In addition, by providing Tweets specific to the location, the "Islamophobia Score" can be calculated for different locations and countries, and the Global Islamophobia Index which includes the countries can be calculated.



Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti

Buğra AYAN^{1,*} Birol KUYUMCU² Bünyamin CİYLAN³

¹Türkiye Cumhuriyeti Cumhurbaşkanlığı Beştepe / Ankara

²SefaMerve Ar-Ge , 34854, Maltepe / İstanbul

³Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, 06500, Yenimahalle/ANKARA

Öz

İnternet kullanımının artmasıyla daha çok erişilebilir hale gelen sosyal ağlar insanların çeşitli konular hakkındaki duygu ve görüşlerini paylaştığı mecralar haline gelmiştir. Dijital ortamda var olan verinin yüzde 90'ı geçen 2 yılda oluşturulmuştur. Fakat bu veriler üzerinde geleneksel yollarla analizler yapmak günler hatta aylar sürebilmektedir. Bu sebeple makine öğrenmesi yöntemleriyle sosyal ağlar üzerinden duygu analizleri yapılması tercih edilmektedir. Bu çalışmada Twitter üzerindeki tweetlerin İslamofobik olup olmadığı duygu analizi ile öğrenilmeye çalışılmıştır. Çalışmada Lineer ridge regresyonu ve Naive Bayes Sınıflandırıcı ile eğitilen modeller üzerinden precision, Recall, F1 ölçütlerinde hesaplamalarda bulunulmuştur. Sonuç olarak pozitif Tweetler için Ridge modelinde Naive Bayes sınıflandırıcıya göre daha iyi sonuçlar alınmış ve Ridge Regresyonunda %96,3 , Naive Bayes Sınıflandırıcıda %95.3 oranında doğru sonuca ulaşılmıştır.

Makale Bilgisi

Başvuru: 8/05/2019

Düzeltilme: 13/06/2019

Kabul: 17/06/2019

Anahtar Kelimeler

Duygu Analizi, İslamofobi,
Twitter, Makine Öğrenmesi

Keywords

Sentiment analysis,
Islamophobia, Twitter,
Machine Learning

Detection of Islamophobic Tweets on Twitter Using Sentiment Analysis

Abstract

Social networks, becoming more accessible as the Internet usage increases, have turned into platforms where people share their feelings and opinions on various subjects. The 90 per cent of the data we have today has been generated over the past two years. Yet, carrying out traditional analyses through this data might take days, if not months. Therefore, emotional analyses are preferred to be made via social networks by means of machine learning. Tweets on Twitter are sought to be found out whether they are Islamophobic or not by using emotion analysis. Estimations are made using precision, recall and F1 measures via models trained with linear ridge model. In the end, accurate results, in the range of 96.3 per cent to 96.5 per cent on average, are obtained for positive tweets.

1. GİRİŞ (INTRODUCTION)

İnternetin web 2.0 evresiyle sosyal ağlar popüler hale gelmiş ve insanlar vakitlerinin büyük bölümünü bu sosyal ağlarda geçirmeye başlamıştır [1-2]. Teknolojinin birçok alanında olduğu gibi sosyal ağ teknolojileri de çeşitli risk ve fırsatı aynı anda firmaların ve bireysel kullanıcıların karşısına çıkartmıştır. Güncel rakamlara göre internette her saniyede 8.000'in üzerinde Tweet paylaşılmaktadır [3]. Bu tweetlerden elde edilen bilgilerle gerek ekonomik gerek toplumsal analizler yapılabilir. Bu analizler ile bireylerin ileriye yönelik tercihleri üzerine de çıkarımlarda bulunulabilmektedir [4]. Twitter'ın kendine has bir dili olması, hatalı yazılan kelimeler gibi sebeplerle sahip olduğu verinin insan algısıyla anlaşılması güçtür. Bu sebeple doğal dil işleme ve veri madenciliği ile ilgili çeşitli çalışmalar yapılmakta ve sektörel ürünler yine bu yöntemlerle endüstriye kazandırılmaktadır.

1.1.Literatür Özeti

Makine öğrenmesi çalışmaları; altyapı çalışmalarıyla internet hızlarının artması ve mobil kullanım ile internet kullanım sürelerinin uzamasıyla beraber cazip hale gelmiştir. Bu çalışmalar kapsamında makine öğrenmesi yöntemleriyle akademik başarının tahmin edilmesinden [5], müşteri kaybı analizine kadar [6]

geniş bir yelpazede akademik çalışmalar yapılmaktadır. Bu çalışmaların bir kısmı da sahip olduğu veri hacminin yüksek olmasından ötürü Twitter üzerinde yoğunlaşmıştır.

Andranik Tumasjan ve arkadaşları tarafından ABD seçim sürecinde atılan Tweetler gerçek zamanlı olarak incelenerek sınıflandırma algoritmalarıyla duygu analizi yapıldığı çalışmada 100.000'in üzerindeki Tweet'in incelemesi sonucu %1.65 hata ile siyasal yönelimler tahmin edilmiştir. [7]

Xing Fang ve Justin Zhan tarafından yapılan çalışmada ürünlere yapılan yorumlar üzerinde duygu analizi yapılmıştır. Naive Bayes, Rassal Orman ve Destek Vektör Makinesi algoritmalarındaki sonuçların karşılaştırıldığı çalışmada. True Positive için en yüksek başarıma %98 ile Rassal Orman ulaşırken F1-Score'da ise Naive Bayes yüzde 82 doğruluğa ulaşmıştır.[8]

Adam Sadilek, Henry Kautz ve Vincent Silenzio tarafından yapılan çalışmada Twitter üzerinden tahmin edilen ruh halleri ile borsa hareketleri tahmin edilmeye çalışılmıştır. Toplam 15,9 milyon Tweetin kullanıldığı verisetinde destek vektör makinesi algoritması kullanılmış ve precision göstergesine göre %98 recall göstergesine göre ise %97 doğruluğa erişilmiştir. [9]

Rohan Kshirsagar ve arkadaşları tarafından yapılan çalışmada Twitter üzerindeki nefret konuşmaları makine öğrenmesi yaklaşımıyla tahmin edilmeye çalışılmıştır. Altmışbine yakın Tweet'in kullanıldığı çalışmada cinsiyetçi söylem için %76, ırkçı söylem için %78 ve diğer nefret söylemlerinin dahil olduğu ana küme için %86 oranıyla F1 skoru bulunmuştur. [10]

Hajime Watanabe ve arkadaşları tarafından yapılan araştırmada 2010 Tweetten oluşan verisetinden bir çalışma yapılmış ve %78 doğruluk oranıyla bir Tweet'in nefret dolu mu, saldırgan mı yoksa normal tonda yazılmış bir Tweet mi olduğu tahmin edilebilmiştir. [11]

Bu çalışmaların yanında Twitter üzerindeki nefret içeren konuşmaların makine öğrenmesi yöntemleriyle tespit edilmesi üzerine de çeşitli araştırmalar yapılmaktadır. Fakat bu çalışmalarda İslamofobi gibi özel bir nefret teması seçilmemiş, genel olarak nefret üzerine çalışma yapılmıştır.

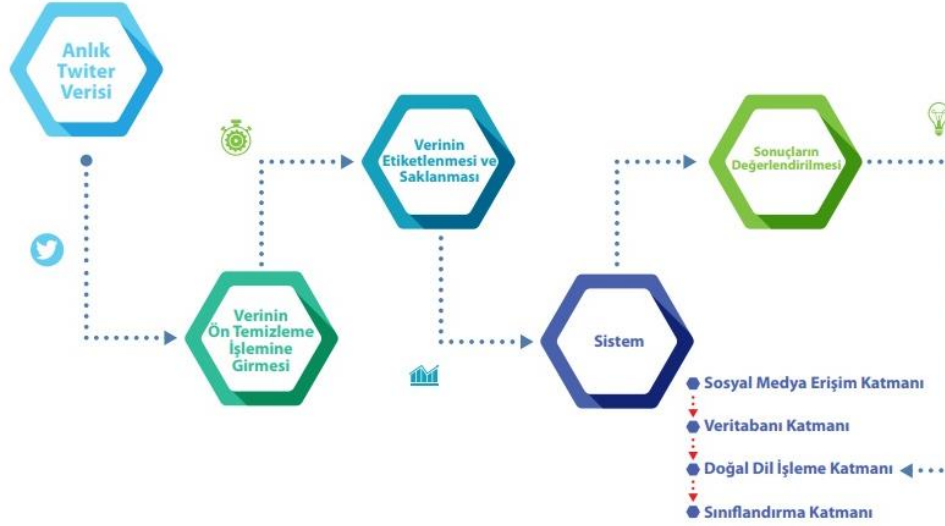
1.2.İslamofobi Seçimi

İslam Korkusu anlamına gelen İslamofobi son yıllarda Müslümanlara ve İslam dinine karşı yapılan saldırıların artmasıyla gün geçtikçe daha da büyük bir sorun haline gelmektedir.[12] İslamofobik vakalar Belçika [13], İsviçre [14], Avusturya [15], Almanya [16] gibi birçok ülkede incelenerek bu konuda çeşitli raporlar yayınlanmıştır. Fakat bu raporlar ağırlıklı olarak basına yansımış büyük şiddet olaylarına yer verilirken, internet üzerindeki daha adi denilebilecek suçlar kayıt altına alınamamaktadır. Bunun sebebi internet verisinin çok fazla olması ve bu alanda çalışan kişilerin geleneksel yöntemlerle bu verileri taramasının zor olmasıdır. Bu çalışmada ise, büyük veriyi işlemenin zorluklarından dolayı dağıtık hesaplama altyapısı kullanılmıştır. Bu yolla elde edilecek sonuçlarla İslamofobi'nin yöntemiyle incelenmesi farklı bir bakış açısı sunabilmektedir. Çalışmanın neticelerinin ilgili kurumlar ile paylaşılmasıyla bu nefret yöneliminin daha önceden tespit edilmesi ve önleyici adımlar atılması sağlanabilecektir.

1.3.Twitter Seçimi

Sosyal ağlar, geliştiriciler için API isminde ara yüzler sunmaktadır. Bu arayüz üzerinden verilen geliştiriciye özel verilen Access Token ve Access Token Secret değerleri ile belirli limitler dahilinde sosyal ağdaki verinin izin verilen kısmı kullanılabilir. [17] Bu çalışmada gerek sunduğu API'nin verilere erişim noktasındaki kolaylığı, gerek üretilen içeriğin yazı odaklı ve kısa olması nedeniyle Twitter platformu tercih edilmiştir.

Çalışma ile ilgili akış diyagramı Şekil 1 de verilmiştir.



Şekil 1. Sistemin Akış Diyagramı

2. TWITTER VERİLERİNİN TOPLANMASI (DATA COLLECTION FROM TWITTER)

Twitter, kullanıcılarının “twit” olarak adlandırdığı en fazla 280 karakterden oluşan mesajlar göndermesine olanak tanıyan, zamanla görsel ve video içerikler paylaşmayı da destekleyen ABD merkezli sosyal ağ ve mikroblog sitesidir. Mobil uygulamasına haber kategorisinde yer vermeyi tercih eden platform önemli bir haber kaynağı da olma yolunda ilerlemektedir. Twitter dünyanın en çok ziyaret edilen internet siteleri arasında yer almaktadır [18]. Platform ilk kuruluşunda sadece yazı bazlı paylaşımlara izin verirken, ilerleyen yıllarda resim, video, canlı yayın gibi farklı içerik türlerini de desteklemeye başlamıştır. Fakat ana odak noktası metin olduğu için ağırlıklı kullanım metin içerikli Tweetler’de gerçekleşmektedir.

Twitter kullanıcıları gerek teknik sınırlamalar gerekse iletişimi daha efektif hale getirmek amacıyla kendine has çeşitli konuşma biçimleri kullanmaktadır. Bunlardan ilki “hashtag” olarak adlandırılan ve diyez(#) karakteriyle belli bir konuya atıfta bulunulmasıdır. Bir başkası @ işaretiyle başka bir kullanıcının mesajdan haberdar edilmesidir. Yine bunların yanında Twitter bazı özel günler için özel simgeler barındıran konseptler geliştirmektedir. Örneğin dünya kupasında belirli bir başlığa gönderilen mesajların yanında futbol topu simgesi çıkmıştır [19].

Bu çalışma kapsamında geliştirilen sosyal medya arama katmanı yardımıyla Twitter API’ sine erişilmiş ve belirlenen anahtar sözcükler için 01.08-2018-05.09-2018 tarihleri arasında sorgulama yapılarak İngilizce dilinde Tweetler toplanmıştır.

2.1. Verilerin Saklanması ve Etiketlenmesi

Çalışmada elde edilen verilerin saklanabilmesi için MYSQL veritabanı kullanılmıştır. Bu sayede yüksek erişilebilirlik sağlanmıştır. Okuma ve yazma performansları güçlü olmuştur. İngilizce dilinde anahtar kelimelere göre çekilen 162.000 farklı tweetin tamamı 5 gönüllü kişi tarafından işaretlenmiş ve salt çoğunluğun görüşüne göre pozitif veya negatif olarak değerlendirilmiştir. Çekilen Tweetler’de harici internet sitesi bağlantıları silinmiş, bu silinme işleminin ardından da benzer olan Tweetler elenmiş, küçük harfe dönüştürülmüş ve işaretleme işlemine geçilmiştir. Bu işaretlemede Tweetlerin İslamofobik içerik barındırıp barındırmadığı belirlenmiştir. İşaretleme sonucu veri setindeki Tweetlerde olumlu ve olumsuz yüzdeleri %51-%49 oranında olmuştur. Buradaki pozitif (olumlu) tanımlaması İslamofobi içeren, negatif (olumsuz) tanımlaması ise İslamofobi içermeyen Tweetler için kullanılmıştır.

3.GELİŞTİRİLEN SİSTEM (SYSTEM)

Çalışmada dört uygulama katmanı kullanılmıştır. Bunlar sosyal medya erişim katmanı, veritabanı yönetim katmanı, doğal dil işleme modülü ve sınıflandırma modülüdür.

3.1.Sosyal Medya Erişim Katmanı

Sosyal medya platformlarında geliştiriciler için API ismi verilen özel araçlar kullanılmaktadır. Bu yolla sosyal medya platformları geliştiricilerin hangi verileri kullandığını görebilmekte ve veri aktarımına çeşitli sınırlamalar getirerek kullanıcılara verilen hizmetin aksamasının önüne geçebilmektedir. Bu çalışma esnasında da yukarıda değinildiği gibi Twitter API'si kullanılmıştır. Oluşturulan uygulamada Twitter içeriklerini okuma izni alınmıştır. Bu izin içeriği herkese açık hesapları kapsamaktadır. Kilitli profillerin içeriklerine ulaşılamamaktadır. API izinleri alındıktan sonra, PHP tabanlı olarak geliştirilen yazılım yardımıyla belirlenen anahtar sözcüklere göre Twitter verileri toplanmıştır. Bu toplama işleminde PHP Cron Job kullanılarak zamanlayıcı uygulama ile periyodik olarak yazılım arama yapmakta ve gelen sonuçlar veritabanına yazılmaktadır. Kullanım sırasında Twitter'ın verdiği limitler doğrultusunda işlem yapılabilir.

3.2.Veritabanı Katmanı

Veritabanı yönetim katmanı MYSQL üzerine iki farklı tablo üzerine veriler eklenmiştir. Mysql seçilirken PHP ile sorunsuz çalışması göz önünde bulundurulmuştur[20]. İlk olarak anahtar kelime yapısına göre İslamofobik olma ihtimali yüksek Tweetler bir tabloya, İslamofobik olmama ihtimali yüksek Tweetler ise başka bir tabloya eklenmiştir. Buradaki tablo seçimi esnasında “muzrad” gibi Müslümanlara hakaret etmek için kullanılan kelimeleri içeren Tweetler İslamofobik olma ihtimali yüksek tabloya eklenirken, “mashallah” benzeri Müslüman jargonundaki sözcükler İslamofobik olma ihtimali düşük tabloya eklenmiştir. Bu şekilde işaretleme sırasında 5 gönüllü kişinin işlemleri daha pratik bir şekilde yapılabilmesi sağlanmıştır.

3.3 Doğal Dil İşleme Modülü

Doğal dil işleme modülünde Python dili kullanılmıştır. Python modülünün Mysql ile bağlantısı web servis kullanarak sağlanmıştır. İlk olarak metin ve etiket olarak veriler birleştirilmiştir. Ardından metni sayıya dönüştürmek için 'kelime' bazlı ve hesaplamak için Scikit-learn kütüphanesinin içerisinde bulunan TfIdfVectorizer metodu kullanılmıştır. Scikit-learn yapay zeka alanında en yaygın olarak kullanılan ve bir çok regresyon ile sınıflandırma yöntemini içeren bir kütüphanedir [21]. Bu aşamada eğitim kümesinde geçen kelimeler TfIdfVectorizer metodu ile terim frekansı (tf) , ters belge sıklığı (idf) özellik matrisine dönüştürülmüştür

3.4 Sınıflandırma Modülü

Metod seçimi ardından %80-%20 oranıyla veri eğitim ve test olarak ikiye ayrılmıştır ve gözetimli öğrenme kullanılarak eğitilmiştir. %80-%20 seçiminde Scikit-learn tarafından önerilen bölme oranları esas alınmıştır[22]. Bu işlemin ardından eğitim seti 'Negatif' ve 'Pozitif' durumları için küçük varyanslı tahminler için öncelikle uygun olan iki ayrı lineer Ridge Modeli eğitilmiştir. Eğitilen bu modeller 'Negatif.pkl' ve 'Pozitif.pkl' isimleri ile kaydedilmiştir. Ridge Regresyonu veya bir diğer ismiyle Tikhonov normalizasyonu regresyon katsayılarının tahmininde, bağımsız değişkenlerin birbirleri üzerindeki etkilerini minimum yapmayı hedefleyen ve kararlı katsayı tahminleri elde edebilmeyi sağlayan bir yöntemdir [23]. Ridge regresyonunun ardından aynı eğitim ve test seti Naive Bayes Sınıflandırıcıyla eğitilmiştir. Naive Bayes sınıflandırıcı Twitter üzerindeki duygu analizlerinde çeşitli araştırmalarda kullanılan olasılıkçı bir yaklaşımdır [24].

Sınıflandırma modülünde eğitilen modeller ile test verisi tahmin edilmeye çalışılmıştır. Tahmin sonuçları ile gerçek sonuçların karşılaştırma matrisi hesaplanmış ve çeşitli ölçütlerle karşılaştırmalar yapılmıştır. Bu ölçütler precision, recall, f1-score türleridir. Precision ölçütü Pozitif olarak tahmin edilen bir durumdaki başarıyı gösterir. Recall pozitif durumların ne kadar başarılı tahmin edildiğini gösterirken, F1 ölçütü ise Recall ve Precision ölçütlerinin harmonik ortalamasıdır [25].

4.SONUÇ VE ÖNERİLER (RESULT AND SUGGESTIONS)

162.000 tweetten oluşan veri setinin %80'i kullanılarak sistem eğitilmiş ve %20'si kullanılarak testler yapılmıştır. Bu yolla sistemin doğruluk oranı tespit edilmiştir.

Veri kümesi üzerinde kelime bazlı TF-IDF ağırlıklandırma işleminin uygulanmasının ardından, ayrı ayrı Ridge Regresyonu ve Naive Bayes Sınıflandırıcı modeli uygulanmıştır. Bu modeller için Tweet sınıflandırma çalışmasında elde edilen oranlar F1 ölçütünde Ridge Regresyonu'nda %96.9 olurken Naive Bayes Sınıflandırıcıda bu değer %95.4'e düşmüştür. Ayrıca eğitim süresi olarak Ridge Regresyonu Naive Bayes Sınıflandırıcıya göre daha kısa sürmüştür.

F1 Score	Pozitif Tweet	Negatif Tweet	Ortalama
Ridge	%97.5	%96.3	%96.9
Naive Bayes	%94.5	%95.9	%95.4
Precision	Pozitif Tweet	Negatif Tweet	Ortalama
Ridge	%96.9	%97.2	%97.0
Naive Bayes	%89.9	%99.5	%95.7
Recall	Pozitif Tweet	Negatif Tweet	Ortalama
Ridge	%98.1	%95.5	%97.0
Naive Bayes	%99.5	%92.7	%95.3

Tablo 1. Tahmin Sonuçlar

Ölçütlerine Göre

Bu sonuçlara göre her üç ölçüt sonuçlarının ortalamalarının birbirine yakın olduğu görülmekte, fakat negatif tweetlerde Precision ölçütü, pozitif tweetlerde ise Recall ölçütünün daha iyi sonuçlar verdiği görülmüştür. Ayrıca ölçütlerin her birinde Ridge Regresyonu Naive Bayes Sınıflandırıcıya göre daha yüksek doğruluk oranına ulaşmıştır.

Çalışma sonucunda elde edilecek veriler ile sosyal medya üzerinden yapılabilecek sosyal mühendislik çalışmalarına karşı erken uyarı sistemi kurulabilir, bölgesel veriler incelenerek karşı enformasyon çalışmaları yapılabileceği öngörülmüştür.

Bu çalışma sonucunda ortaya konulabilecek öneriler, farklı diller için de benzer doğal dil işleme çalışmalarıyla araştırmalar yapılması, lokasyon ve tarih bazlı olarak tweetlerin incelenerek daha detaylı sonuçlar alınabilmesidir.

Öneri olarak derin öğrenme kullanan bir metod denenebilir ve veri setinin genişletilmesi üzerine çalışmalar yapılabilir. Ayrıca görüntü işleme kütüphaneleri kullanılarak görsel ve video içeren Tweetler üzerine çalışmalar yapılabilir. Fakat bu kısımda veri boyutunun artması nedeniyle daha yüksek işlemciye gücüne ihtiyaç duyulacaktır.

Bunların yanında konuma özel olarak Tweetlerin temin edilmesiyle farklı konular, ülkeler için "İslamofobi Skoru" hesaplanabilir, bu skorlar üzerinden ülkelerin yer aldığı Küresel İslamofobi indeksi çıkarılabilir.

TEŞEKKÜR (ACKNOWLEDGMENT)

Çalışmaya katkıda bulunan Salih Çağlar, Berk Çelebi, Şebnem Ayan'a teşekkür ederim.

KAYNAKLAR (REFERENCES)

- [1] Buzzi, M. C., Buzzi, M., & Leporini, B. Web 2.0: Twitter and the blind. In Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity (pp. 151-156), ACM. (2011, September).
- [2] Pak, A., & Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326), (2010, May).
- [3] Murthy, D. (2018). Twitter. Cambridge, UK: Polity Press, 24,25.
- [4] Bermingham, A., & Smeaton, A. On using Twitter to monitor political sentiment and predict election results. In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011) (pp. 2-10), (2011).
- [5] GÖK, M . "MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE AKADEMİK BAŞARININ TAHMİN EDİLMESİ". Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji 5 (2017): 139-148
- [6] KAYNAR, O , TUNA, M , GÖRMEZ, Y , DEVECİ, M . "Makine öğrenmesi yöntemleriyle müşteri kaybı analizi". Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi 18 (2017): 1-14
- [7] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. Predicting elections with twitter: What 140 characters reveal about political sentiment. Icwsm, 10(1), (178-185.), (2010).
- [8] Fang, X., & Zhan, J. Sentiment analysis using product review data. Journal of Big Data, 2(1), 5, (2015).
- [9] Sadilek, A., Kautz, H. A., & Silenzio, V. Modeling Spread of Disease from Social Interactions. In ICWSM (pp. 322-329), (2012, June).
- [10] Kshirsagar, R., Cukovac, T., McKeown, K., & McGregor, S. Predictive Embeddings for Hate Speech Detection on Twitter. arXiv preprint arXiv:1809.10644, (2018).

- [11] Watanabe, H., Bouazizi, M., & Ohtsuki, T. Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. *IEEE Access*, 6, 13825-13835, (2018).
- [12] Er, T., & Ataman, K. İslamofobi ve Avrupa’da birlikte yaşama tecrübesi üzerine. *Uludağ Üniversitesi İlahiyat Fakültesi Dergisi*, 17(2), (2008).
- [13] AYDIN, M. Z. Belçika’da İslamofobi ve Müslümanlara Yönelik Ayrımcılık. *Cumhuriyet Üniversitesi İlahiyat Fakültesi Dergisi*, 12(1), 7-32, (2008).
- [14] Çebi, M. S. İsviçre’deki Minare Karşısı Referandum Afişlerinde İslamofobi’nin Söylemsel İnşası. *Bilig/Türk Dünyası Sosyal Bilimler Dergisi*, (73), 99-140, (2015).
- [15] HAFEZ, F. Avusturya Özgürlük Partisi FPÖ’nün Sağcı Popülizminin Bir Aracı Olarak İslamofobi. Şu kitapta: Haz. Kadir CANATAN ve Özcan HIDİR. *Batı Dünyasında İslamofobi ve Anti-İslamizm*. Ankara: Eski Yeni Yayınları, 355-372, (2007).
- [16] Yavuzcan, İ. H. Almanya’da İslamofobi. *Batı Dünyasında İslamofobi ve Anti-İslamizm içinde*, ed. Canatan, Kadir-Hıdır, Özcan, Eskiyei Yayınları, Ankara, (2007).
- [17] Kwak, H., Lee, C., Park, H., & Moon, S. What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). AcM., (2010, April).
- [18] Resce, G., & Maynard, D. What matters most to people around the world? Retrieving Better Life Index priorities on Twitter. *Technological Forecasting and Social Change*, 137, 61-75. (2018).
- [19] Fan, M., Billings, A., Zhu, X., & Yu, P. (2019). Twitter-Based BIRGing: Big Data Analysis of English National Team Fans During the 2018 FIFA World Cup. *Communication & Sport*, 2167479519834348.
- [20] Kofler, M. What Is MySQL?. In *MySQL* (pp. 3-19), (2001). Apress, Berkeley, CA.
- [21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830, (2011).
- [22] Saseendran, A. T., Setia, L., Chhabria, V., Chakraborty, D., & Roy, A. B. (2019). Impact of Data Pruning on Machine Learning Algorithm Performance. *arXiv preprint arXiv:1901.10539*.
- [23] Saleh, A. M. E., Arashi, M., & Kibria, B. G. (2019). *Theory of Ridge Regression Estimation with Applications* (Vol. 285). John Wiley & Sons.
- [24] Go, A., Huang, L., & Bhayani, R. Twitter sentiment analysis. *Entropy*, 17, 252, (2009).
- [25] Powers, D. M. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, (2011).