



Twitter’da Duygu Analizi

Nagehan İLHAN^{1,*}, Duygu SAĞALTICI²

¹Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 63100, Haliliye/ŞANLIURFA

²Harran Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 63100, Haliliye/ŞANLIURFA

Öz

Sosyal medya platformları, kullanıcılara tüm kategorilerde düşüncelerini ve duygularını paylaşacakları bir meca oluşturmuş ve günlük rutinin bir parçası haline gelmiştir. Yaygın sosyal medya kullanımı ile devasa boyutlarda veriler oluşmaya başlamıştır. Veri hacmindeki bu hızlı artışla, bu verilerin yönetilme ve içerisinden anlamlı bilgi çıkarılması ihtiyacı doğmuş ve akıllı hesaplama yöntemlerinin bu verileri analiz etmesi son derece kritik bir hale gelmiştir. Duygu analizi, anlamlı bilgi elde etmek için verilere uygulanan süreçler bütünüdür. Bu makalede, kullanıcıların durum güncellemelerini “tweet” şeklinde yayınladığı, çok sayıda kullanıcı tarafından kullanılan popüler sosyal medya sitesi olan Twitter verileri üzerinde duygu analizi yapılmıştır. Bu analizi gerçekleştirmek için Naïve Bayes ve Support Vector Machine gibi makine öğrenme yöntemleri kullanılarak tweetleri pozitif ve negatif sınıflara ayırmak için akıllı bir model oluşturulmuş ve karşılaştırılmalı sonuçlar verilmiştir

Makale Bilgisi

Başvuru: 24/07/2020
Düzelme: 10/08/2020
Kabul: 19/08/2020

Anahtar Kelimeler

Duygu Analizi
Veri Madenciliği
Sosyal Medya

Keywords

Sentiment Analysis
Data Mining
Social Media

Sentiment Analysis in Twitter

Abstract

Social media platforms have set users a space to share their opinions and feelings on all categories of topics and become a part of daily routine in human lives. The emergence of social media platforms has led to an exponential increase in the volume of data. With a rapid increase in the volume of data, computationally intelligent methods are proving to be extremely important to analyse the data in many fields. The sentiment analysis is the automated process applied to data to extract meaningful information. This paper focuses mainly sentiment analysis of Twitter which is one of the popular social media sites used by numerous users where users publish status update in the form of tweets. We build model for classifying “tweets” into positive and negative sentiment. To perform this analysis an intelligent model has been created by using machine learning methods such as Naïve Bayes and Support Vector Machine and the compared results has been given.

1. GİRİŞ (INTRODUCTION)

Günümüze kadar varan süreçte internetin gelişmesi bireylerin fikirlerini paylaşabilecekleri sosyal platformlara olanak sağlamıştır. Sosyal medya, bloglar, kişisel siteler gibi pek çok seçenek ve gittikçe yaygınlaşan kullanım alanları internette mevcuttur. İnsanların fikirlerini, bir konu hakkındaki görüşlerini paylaşmaları, bir toplumun hangi durumlar karşısında nasıl tepkiler verdiğini analiz edebilme olanağı sunmaktadır. Özellikle de Twitter gibi sosyal paylaşım ağlarının yaygın kullanılması buralardan elde edilecek verilerin değerini arttırmaktadır. Twitter, API aracılığıyla kolayca veri erişimine izin veren bir platformdur. Ancak Twitter yapı itibarıyla düzenlenmesi gereken bir veriye sahiptir. Bu nedenle verilerin bazı ön işlemlerden geçmesi ve kullanılacak yöntemlere yapısal olarak uyabilecek hale getirilmesi gerekmektedir [1].

*İletişim yazarı, e-mail: nagehanilhan@harran.edu.tr

Bireylerin fikirlerini içeren veriler çok yoğun ve büyük olabilir. Bu verilerden anlamlı sonuçlar çıkarabilmek adına doğal dil işleme yöntemleri kullanılarak duygu analizi yapılabilmektedir. Duygu analizi temelde bir doğal dil işleme yöntemidir. Bir cümlenin, kelimenin ya da metnin ifade ettiği anlamı bulmak için duygu analizi yöntemleri kullanılmaktadır. Temel olarak üç sonuca ulaşılabilir: olumlu, olumsuz, nötr [2].

Duygu analizi yapabilmek için verilerin temizlenmesi şarttır. Metin ya da cümle içindeki her bir kelimenin eklerine ve köklerine ayrılması, noktalama işaretleri ve simgelerin temizlenmesi, edat, bağlaç ve zarf gibi etiketlerinin belirlenmesi gerekmektedir. Farklı yöntemler kullanılmasına bağlı olarak her bir kelimenin metin içerisindeki terim ve ters doküman frekansları belirlenmelidir. Bu şekilde çeşitli sınıflandırma yöntemleri kullanılarak duygu analizi sonuçlarına ulaşılabilir olanağı oluşacaktır [3].

Duygu analizi günümüzde özellikle de ürün pazarlama ve ticari olarak birçok alanda kullanılmaktadır. Gerek toplum gerekse bireysel olarak elde edilen bilgiler ticari ve siyasi olarak da kullanılabilir nitelik taşımaktadır. Literatürde de duygu analizine dayalı örnekler yer almaktadır [4].

Bu çalışmada Twitter'den alınan 1,578,627 adet sınıflandırılmış tweet üzerinde duygu analizi yapmak amacıyla Destek Vektör Makineleri, Naïve Bayes gibi makine öğrenmesi yöntemleri kullanılmıştır. N-gram kullanılarak elde edilen sonuçlarda iyileştirme amaçlanmıştır.

2. LİTERATÜR TARAMASI (RELATED WORK)

Toplumsal düşünceyi veya duyguları analiz etmek, ürünlerinin piyasadaki tepkisini bulmaya çalışan, siyasi seçimleri tahmin eden ve borsa gibi sosyo-ekonomik etkenleri tahmin eden firmalar için yardımcı olabilir. Duygu analizi teknikleri bu amaç için yaygın olarak popülerdir. Uzun zamandır büyük kitlelere ait duyguları bulmak için çeşitli duygu sınıflandırma yaklaşımları ve yöntemleri kullanılmıştır. Duygu analizi yapılan çalışmalarda araştırmacılar çoğunlukla pozitif ve negatif olmak üzere iki sınıf belirlemişlerdir. [4].

Tansa Trisna Astono Putri ve arkadaşları [5] Endonezya'daki 3 büyük şehir, Jakarta, Bandung ve Medan ile ilgili görüşleri içeren tweetler üzerinde olumlu, olumsuz veya tarafsız tweet sınıflandırması yapmışlardır. Bu çalışmalarında duygu analizi için Destek Vektör Makinesi (SVM) yöntemini kullanmışlardır. Çalışmalarını önışleme, hedefe bağlı bir yaklaşımla sınıflandırma, SVM yöntemi ve son olarak fikir sınıflandırmasının belirlenmesi olmak üzere 4 aşamada gerçekleştirmişlerdir. Sonuç olarak hedefe bağlı yaklaşım Bandung, Jakarta ve Medan şehirlerinin duyarlılık analizi durumunda başarılı bir şekilde uygulanmış, daha sonra SVM yöntemi kullanılarak duyarlılık sınıflandırması yapılmıştır. Hedefe bağlı yaklaşımın Bandung, Cakarta ve Medan şehirleri hakkındaki olumlu, olumsuz ve tarafsız görüşlerin sayısını etkilediği görülmüştür.

Xia ve diğerleri [6] duygu analizinde Ensemble Learning yönteminin etkinliğini araştırmışlardır. Amaçları, daha güçlü bir sınıflandırıcı oluşturmak için çeşitli özellik kümelerini ve çeşitli sınıflandırma algoritmalarını verimli bir şekilde birleştirmektir. Duygu sınıflandırması için farklı düzenleme prosedürlerinin birleştirilmesiyle elde edilen bir toplama sistemi kullanmışlardır. Kelime torbası (BOW) gibi geleneksel metin sınıflandırma yaklaşımları, bazı kelimeleri atladığı için duyarlı sınıflamaya uygun değildir. Bu çalışmada, iki özellik türü (POS ve Word ilişkileri) ve üç sınıflandırıcı (NB, MaxEnt ve SVM) kullanılmıştır. Üç tür topluluk sınıflandırıcısı önerilmiş ve değerlendirilmiştir: ağırlıklı gruplandırma, sabit gruplama ve meta-sınıflandırıcı gruplama. Sonuçlar, Ensemble Learning yöntemlerinin ayrı ayrı sınıflandırıcıya kıyasla belirgin iyileşmelere yol açtığını göstermiştir. Ayrıca, sonuçlar farklı özellik kümelerine sahip çeşitli sınıflandırıcılardan oluşan topluluğun çok önemli gelişmeler sağladığını kanıtlamıştır.

Go ve L. Huang [7] tweet duygularını sınıflandırmak için bir yöntem geliştirmişlerdir. Duygu analizini, tweetlerin pozitif ve negatif olarak sınıflandırıldığı ikili bir sınıflandırma olarak ele almışlardır. İfadeleri olan tweetler içeren eğitim verileri, Read [8] tarafından önerilen denetim yaklaşımına göre toplanmıştır. Go ve arkadaşları [7] ifadeleri içeren tweetleri çekmek için Twitter API'sini kullanmışlardır. Yeniden yayınlanan tweetler ve tekrarlanan tweetleri çıkarmışlardır. Ayrıca, olumlu ve olumsuz duygular içeren tweetler filtrelenmiştir. Tweetleri sınıflandırmak için NB, MaxEnt ve SVM gibi çeşitli sınıflandırıcıları kullanılmıştır. Unigramlar, bigramlar, bigramlı unigramlar ve POSlu unigramlar gibi farklı özellikleri çıkarmışlardır. En iyi sonuçların, unigram ve bigram özellikleriyle birlikte MaxEnt sınıflandırıcısı

tarafından elde edildiğini görmüş; % 82.7'lik bir sınıflandırma doğruluğu ile NB'ye kıyasla yaklaşık % 83'lük bir doğruluğa ulaşmışlardır.

Malhar ve Ram [9], Twitter verilerini sınıflandırmak için denetimli (supervised) yöntemini önermişlerdir. Yaptıkları çalışmada, SVM'nin diğer sınıflandırıcılardan daha iyi performans gösterdiğini ve hibrit bir özellik seçimi kullanılarak % 88'lik bir doğruluk elde ettiğini göstermişlerdir. Nitelik azaltmak için Temel Bileşen Analizini (PCA) SVM sınıflandırıcısıyla birleştirmeyi denemişlerdir. Ayrıca unigram, bigram, hibrid (unigram ve bigram) özellik çıkarma yöntemleri kullanılmıştır. Malhar ve Ram, PCA'nın SVM ile entegre edilmesinin niteliklerin azaltılmasına yardımcı olabileceğini ve sonuçların %92'lik bir sınıflandırma doğruluğu elde ettiğini göstermişlerdir.

Sharma ve diğerleri [10], metin belgelerinin polaritelerine göre duyarlılık yönünü belirleyebilen, gözetimsiz bir belge tabanlı duyarlılık analiz sistemi (unsupervised document-based sentiment analysis system) önermişlerdir. Bu sistem belgeleri pozitif ve negatif olarak sınıflandırır [10] [11] ve duyarlılık kelimelerini belge koleksiyonlarından çıkarır ve kutuplarına göre sınıflandırır. Gözetimsiz sözlük tabanlı strateji (unsupervised dictionary-based strategy), bu sistemin bir parçası olarak kullanılır ve bu da ek olarak olumsuzlukla ilgilendirir. WordNet, fikir kelimelerini, eşdeğer kelimelerini ve zıt anlamlı sözcükleri tanımlamak için kullanılan bir sözlüktür [10]. Bu özel çalışmada, belgelerin polarite duyarlılığını tespit etmek için girdi olarak kullanılmak üzere film incelemeleri toplanmıştır. Sistem her birini pozitif, negatif ve tarafsız olarak sınıflandırmış ve toplam pozitif, negatif ve tarafsız belge sayısını sunarak özet çıktılar üretmiştir. Böylece sistem tarafından hazırlanan özet rapor, karar vericilere yardımcı olmuştur. Bu sistemle, herhangi bir belgenin duyarlılık polaritesine, belgelerde görünen fikir kelimelerinin çoğuna göre karar verilir.

Po-Wei Liang ve diğerleri [12] tarafından sosyal medya mesajlarının duygularını otomatik olarak araştıran ve tespit eden “opinion miner (görüş madencisi)” adı verilen bir sistem tasarlanmıştır. Açıklamalı tweetleri analizin yapılması için birleştirmişler ve bu çerçevede, tarafsız tweetler silinerek sadece duyguları içeren mesajlara ait polariteler belirlenmiştir (yani, pozitif veya negatif). Bunu başarmak için, Po-Wei Liang ve arkadaşları unigramlı Naïve Bayes sınıflandırıcısını kullanarak tweetleri “görüş” ve “görüş dışı” olarak sınıflandırmışlardır. Aynı şekilde, Karşılıklı Bilgi ve Ki-Kare çıkarma stratejisini kullanarak ilgisiz özellikleri elemişlerdir. Sonuçlar, gerçek mikroblog uygulamalarında duygu analizi için sistemin yeterliliğini doğrulamıştır.

3. ARKA PLAN BİLGİSİ (BACKGROUND INFORMATION)

3.1 Sentiment Analizi

Duygu analizi olarak bilinen sentiment analizi temelde bir metin işleme yöntemidir. Bir kelime, cümle ya da metnin ifade ettiği anlamı analiz etmek için kullanılır. Temelde olumlu, olumsuz ve nötr olmak üzere üç farklı sonuç elde edilebilir. Duygu analizinde genellikle pozitif ve negatif sınıflar kullanılır ancak bazı çalışmalarda düşüncenin nötr olabileceği düşünülerek üç sınıf üzerinden çalışmalar gerçekleştirilmiştir. Fikir madenciliği ise bir bilginin sosyal medya ortamlarında iletmeyi amaçladığı düşüncenin anlambilimsel olarak incelenmesidir. [2].

Sentiment analizi yapılırken, öncelikle bir eğitim veri seti oluşturulur. Eğitim veri setinde sınıfı belirlenmiş olan (pozitif / negatif /nötr) veriler kullanılır. Daha sonra bu veriler çeşitli veri madenciliği yöntemleri ile gerekli ön işlemlerden geçirilir ve temizlenerek sınıflandırılmaya uygun hale getirilir. Bu ön işlemlerden bazıları metindeki simge ve noktalama işaretlerinin temizlenmesi, metnin tek tek kelimelere ayrılması ve her bir kelimenin çeşitli yöntemlerle köklerinin bulunarak terim listelerinin oluşturulması, metin içerisinde yer alan edat, bağlaç ve zamirlerin oluşturduğu durak kelimelerin (stop words) kaldırılması, terim frekansları (TF) ve ters doküman frekansları (TDF) yardımıyla vektörel modelinin oluşturulmasıdır. Vector uzay modellerinde kullanılan yöntemler arasında n-gram, binary, ters döküman frekansı, terim frekansı metotları sayılabilir. Vektörel modeli oluşturulan bu veriler Karar Ağaçları, Destek Vektör Makinaları (SVM), Yapay Sinir Ağları (NN) gibi veri madenciliğinde sıklıkla kullanılan sınıflandırma algoritmaları kullanılarak eğitilmektedir. Sınıflandırma yöntemleriyle eğitilen modeller yeni gelen duygu ve görüş metinlerinin sınıflandırılmasında kullanılmaktadır [3].

3.2 WordNet

Princeton Üniversitesi Bilişsel Bilimler Laboratuvarı'nda hazırlanmış İngilizce kelimelerin kavramsal olarak ilişkilendirildiği bir sözlük veri tabanıdır. Wordnet'in oluşturulma amacı, anlamların bir araya getirilmesidir. Bunun için alt kavram-üst kavram ilişkisi kullanılmıştır. Ayrıca içerisinde eş anlamlı (synset) sözcükleri de barındırmaktadır ve sözcüklerin anlamlarına yer vermektedir [13].

Wordnet bir kaç kaynaktan ontoloji olarak alınmaktadır ancak oluşturulma amacı dikkate alındığında böyle bir amacı yoktur. Asıl amaç, İngilizce sözcüklerin bir takım ilişkilerle bir yapı haline getirilmesidir. Böylece alt kavram-üst kavramlar oluşturulmuştur. Bu sebeple Wordnet kavramsal bir ontoloji olarak görülebilmektedir.

İngilizce Wordnet'in, doğal dil işleme çalışmalarında kullanım alanı artınca, diğer diller için de Wordnet'ler geliştirilmeye başlanmıştır. Hâlihazırda kırkın üzerinde farklı dil için Wordnet çalışması yapılmıştır. Ayrıca bunun üzerine EuroWordnet ortaya çıkmıştır. EuroWordnet 8 farklı Avrupa dili için geliştirilmiştir. Her dil için ayrı birer Wordnet oluşturulmuştur [13].

WordNet isimleri, fiilleri, sıfatları ve zarfları sözcük kategorilerini içerir, ancak edatları, belirleyicileri ve diğer işlev sözcüklerini göz ardı eder. Bu kategorilerin her biri bilişsel eş anlamlılar (synsets) kümeleri halinde gruplandırılır. Synsets kavramsal-anlamsal ve sözcüksel ilişkiler aracılığıyla birbirine bağlanır. Ortaya çıkan anlamlı kelime ve kavramlardan oluşan ağ, tarayıcı ile gezilebilir. WordNet ayrıca ücretsiz ve herkese açık olarak indirilebilir. WordNet yapısı dolayısıyla, hesaplamalı dilbilim ve doğal dil işleme için kullanışlı bir sözlük veri tabanıdır [14].

3.3 SentiWordNet

SentiWordNet, İngilizce için oluşturulmuş olan bir görüş sözlüğüdür. SentiWordNet, olumlu, olumsuz ve nötr olmak üzere üç çeşit eş anlamlılık kümesi barındırır. Eş anlamlılık kümesi, anlamsal açıdan birbiriyle aynı olan terimlerin bir arada bulunduğu kümedir. SentiWordNet'teki her veri satırı Şekil 1'de gösterilen yapıya sahiptir [15].

Tablo 1: SentiWordNet Yapısı [16]

POS	Eş anlamlılık kümesindeki terimlerin türü. Bu terimler 4 çeşit türe sahip olabilir: Sıfat (a), Fiil (v), Zarf (r) ve İsim (n)
ID	Eş anlamlılık kümesinin benzersiz numarası
PosScore (Pozitiflik Skoru)	Eş anlamlılık kümesindeki terimlerin pozitiflik skoru. Bu skor 0 ile 1 aralığında bir değerdir ve nötr terimler için 0 değerine sahiptir. Bu skorun değeri ne kadar fazlaysa, eş anlamlılık kümesindeki terimlerin pozitiflik şiddeti o kadar fazladır
NegScore (Negatiflik Skoru)	Eş anlamlılık kümesindeki terimlerin negatiflik skoru. Bu skor 0 ile 1 aralığında bir değerdir ve nötr terimler için 0 değerine sahiptir. Bu skorun değeri ne kadar fazlaysa, eş anlamlılık kümesindeki terimlerin negatiflik şiddeti o kadar fazladır.
SynsetTerms (Eş Anlamlılık Kümesi)	Bu küme, anlamsal açıdan birbiriyle aynı olan terimleri içerir.
Gloss (Anlam)	Eş anlamlılık kümesindeki terimlerin anlamıdır.

3.4 N-Gram

N-gram dil modeli $n-1$ dereceden bir Markov Modeli sıralamasında bir sonraki elemanı tahmin eden istatistiksel bir dil modelidir. N-gram analizinde bir cümleyi oluşturan kelimeler belirlenmiş olan N parçaya bölünür. Bir dizi veriyi N-grama dönüştürerek, iki veya daha fazla veri akışını verimli bir şekilde

karşılaştırmak sıkça kullanılan bir doğal dil işleme yöntemidir. Temel olarak belirli bir pencerede bir arada oluşan kelimeler kümesidir ve N-gramları hesaplariken tipik olarak bir kelime ileri taşınır. Bir n-gram modeli, olasılığı bilinen kelime sırası baz alınarak sıradaki kelimenin ne olabileceğini tahmin etmeye çalışır. Bu yöntem kullanılarak n-1 kelimeye bakıldığında bir sonraki kelimenin kullanılma olasılığı hesaplanabilmektedir. N-gramlık metinler, metin madenciliği ve doğal dil işleme görevlerinde yaygın olarak kullanılmaktadır [1].

3.5 Sınıflandırıcılar

3.5.1 Naïve Bayes

Naïve Bayes sınıflandırıcısı olasılık tabanlı bir sınıflandırıcıdır. Sınıflandırıcı, özneliklerin birbirinden bağımsız olduğunu varsayar. Bu, bir özneliğinin var olmasının veya olmamasının, başka bir özneliğinin var olması, olmaması veya değerini etkilemeyeceği anlamına gelir. Naïve Bayes sınıflandırıcısı Bayes teoremine dayanmaktadır [5]. Bayes teoremi formülü aşağıda verilmiştir:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3.1)$$

Naïve Bayes sınıflandırma algoritması şu şekilde açıklanabilir:

1. D'nin veri seti kümesini temsil ettiği ve D'deki her elemanın sınıfının belli olduğu var sayılırsa, X ise n tane nitelikten oluşan bir vektör olarak kabul edilirse $X=(x_1, x_2, \dots, x_n)$ olarak temsil edilmektedir [5].
2. C ise sınıfları temsil ederken m tane sınıf olduğu varsayılınsın. Naïve Bayes sınıflandırma algoritması bir X vektörünün C'de yer alan herhangi bir sınıfa ait olup olmadığını belirlemek amacıyla, bütün sınıflar içinde en yüksek olasılığa sahip değeri bulmaya çalışır.
3. Sonuçta, sınıflandırıcı en büyük olasılık değerine sahip olan C_i sınıfını, X vektörünün sınıfı olarak seçer [5].

3.5.2 Destek Vektör Makineleri (Support Vector Machine- SVM)

Destek Vektör Makinesi algoritmasının amacı, N-boyutlu bir alanda (N - özellik sayısı) veri noktalarını ayrı ayrı sınıflandıran bir hiper düzlem bulmaktır. İki veri noktası sınıfını ayırmak için seçilebilecek birçok olası hiper alan vardır. Amacımız maksimum marj (snır), yani her iki sınıfın veri noktaları arasındaki maksimum mesafe olan bir düzlem bulmaktır. Marj mesafesini en üst düzeye çıkarmak, gelecekteki veri noktalarının daha güvenli bir şekilde sınıflandırılabilmesi için bir miktar takviye sağlar. Hiper alanlar veri noktalarını sınıflandırmaya yardımcı olan karar sınırlarıdır. Hiper düzlemin her iki tarafına düşen veri noktaları farklı sınıflarla ilişkilendirilebilir. Ayrıca, hiper düzlemin boyutu özelliklerin sayısına bağlıdır. Giriş özelliklerinin sayısı 2 ise, hiper düzlem sadece bir çizgidir. Giriş özelliklerinin sayısı 3 ise, hiper düzlem iki boyutlu bir düzlem haline gelir. Özellik sayısı 3'ü aştığında hayal etmek zorlaşır [5].

4.4 Sınıflandırma ve Deneyler

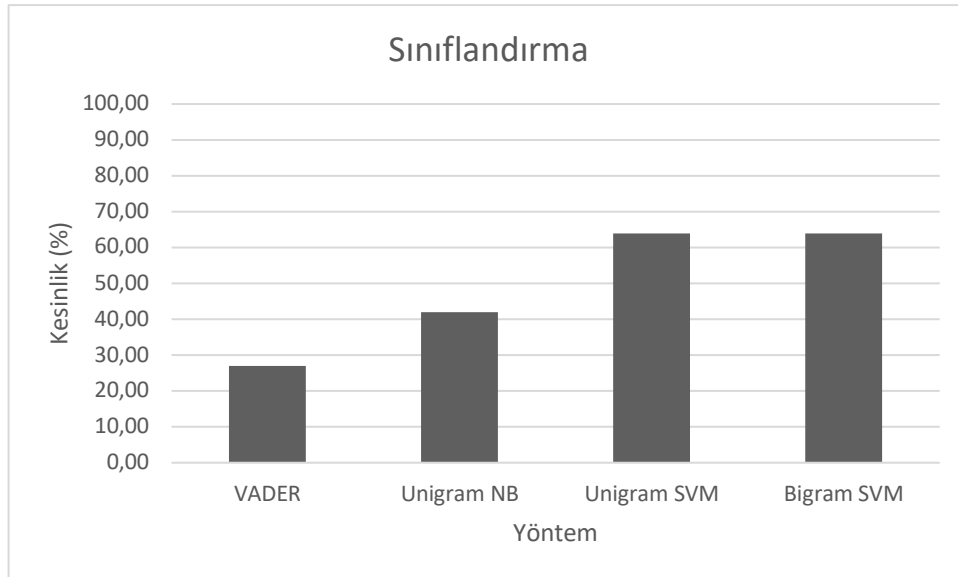
Twitter kullanılarak elde edilen verilerin duygu analizi ve sınıflandırılması için farklı yöntemler kullanılmıştır. Kısaca bahsedilecek olursa Python dili ile duygu analizinde sıklıkla kullanılan WordNet ve N-gram gibi yöntemler kullanılmıştır. Sınıflandırmada ise Naïve Bayes ve SVM sınıflandırıcıları ile başarı ölçülmüştür. Sonuçlar Şekil 6'da verilmiştir.

Unigram kullanılarak duygu analizi yapılmış ve Naïve Bayes sınıflandırma yöntemi ile sonuçlar elde edilmiştir. Python dilinde NLTK kütüphanesine ait SentimentAnalyzer tool kullanılmıştır. Burada cümleler kelimelere ayrılmış, frekans değerleri hesaplanmış ve Naïve Bayes sınıflandırması yapılarak sonuç elde edilmiştir.

Tüm değerler ilk aşamada %66 train ve %33 test olarak ayrılmıştır ve Naïve Bayes başarı oranı %42 olarak bulunmuştur.

NLTK ayrıca VADER (Valence Aware Dictionary ve Sentiment Reasoner) Sentiment Analyzer'ı da içerir. Özellikle günlük sosyal medya metinleri ile çalışmak için oluşturulan bir sözlük ve kural tabanlı duyarlılık analiz aracıdır. Bu hazır bir tool olmakla beraber deneme için kullanılmış ve %27 gibi başarısız bir kesinlik sonucu elde edilmiştir.

SVM(Support Vector Machine) sınıflandırma yöntemi unigram ve bigram yöntemleri ile kullanılmıştır. SVM yöntemi veri setleri için Python da sıklıkla tercih edilen Scikit-learn kütüphanesi ile kullanılır. Scikit-learn yapay öğrenme alanında en yaygın olarak kullanılan kütüphanelerden biridir. Doğrusal regresyon, lojistik regresyon, karar ağaçları, rastgele orman gibi birçok temel yöntemi içermektedir. Bu çalışmada word_tokenize, mark_negation, CountVectorizer, Pipeline, LinearSVC ve TransformerMixin kütüphaneleri kullanılarak unigram ve bigram yöntemi ile SVM sınıflandırma yapılmıştır. Her iki yöntemde de 0.64 başarı elde edilmiştir.



Şekil 6: Test verileri sınıflandırma sonuçları

Özetleyecek olursak, yöntemlerden en iyi sonucu %64 ile Unigram ve Bigram SVM vermektedir. Toplam veri setinde ise Naïve Bayes % 42 başarı oranı verirken VADER %27 başarı oranına sahiptir. Sonuçlara bakıldığında Unigram ve Bigram ile SVM birlikte kullanıldığında sonuçların iyileştiği görülmektedir. Bu çalışmada veriler iki polariteye ayrılmaktadır. SVM'in daha iyi sonuç vermesinin diğer nedenleri arasında ikili bir sınıflandırıcı olması ve büyük boyutlu metin verilerinde başarılı olması sayılabilir.

5.SONUÇ (CONCLUSION)

Bu çalışmada Twitter'da kullanıcılarına ait sınıflandırılmış tweet verileri üzerinde çeşitli makine öğrenme teknikleri kullanılarak bir duygu analizi çalışması yapılmıştır. N-gram metodu ile olumlu ve olumsuz

duyguların analizi yapılmış, makine öğrenmesinde sıkça kullanılan Naïve Bayes ve Destek Vektör Makineleri yöntemleri kullanılarak ilgili sınıflandırıcıların performans karşılaştırmaları yapılmıştır. Sonuçlar incelendiğinde, en yüksek değerın Destek Vektör Makineleri sınıflandırıcısına ait olduğu görülmüştür.

KAYNAKLAR (REFERENCES)

- [1] E. S. Akgül, C. Ertano and B. Diri, "Twitter Verileri ile Duygu Analizi," *Pamukkale University Journal of Engineering Sciences*, vol. 22, no. 2, 2016.
- [2] S. E. Şeker, "Duygu Analizi (Sentimental Analysis)," *YBS Ansiklopedi*, vol. 3, no. 3, pp. 21-36, 2016.
- [3] A. Yeşilyurt and S. E. Şeker, "Metin Madenciliği Yöntemleri ile Twitter Duygu Analizi," *YBS Ansiklopedi*, vol. 4, no. 2, 2017.
- [4] R. S. Gound, P. V. Tikone, S. S. Suryawanshi and D. Nagpal, "Twitter Data Sentiment Analysis and Visualization," *International Journal of Computer Applications*, vol. 180, no. 20, pp. 14-16, 2018.
- [5] T. T. Astono Putri, M. D. Mendoza and M. F. Alie, "Sentiment Analysis On Twitter Using The Target-Dependent Approach And The Support Vector Machine (SVM) Method," *Jurnal Mantik*, vol. 4, no. 1, pp. 20-26, 2020.
- [6] R. Xia, C. Zong and S. Li, "Ensemble of Feature Sets and Classification Algorithms for Sentiment Classification," *Information Sciences*, vol. 181, no. 6, pp. 1138-1152, 2011.
- [7] A. Go, R. Bhayani and L. Huang, "Twitter Sentiment Classification using Distant Supervision," CS224N Project Report, Stanford, 2009.
- [8] J. Read, "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification," *Proceedings of the ACL student research workshop*, pp. 43-48, 2005.
- [9] M. Anjaria and R. M. R. Guddeti, "Influence factor based opinion mining of Twitter data using supervised learning," in *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, India, 2014.
- [10] R. Sharma, S. Nigam and R. Jain, "Opinion Mining of Movie Reviews at Document Level," *International Journal on Information Theory (IJIT)*, vol. 3, no. 3, 2014.
- [11] R. Sharma, S. Nigam and R. Jain, "Polarity Detection at Sentence Level," *International Journal of Computer Applications*, vol. 86, no. 11, 2014.
- [12] P.-W. Liang and B.-R. Dai, "Opinion Mining on Social Media Data," in *2013 IEEE 14th International Conference on Mobile Data Management*, Milan, Italy, 2013.
- [13] O. Bilgin, Ö. Çetinoğlu and K. Oflazer, "Building a Wordnet for Turkish," *Romanian Journal of Information Science and Technology*, vol. 7, no. 1-2, pp. 163-172, 2004.
- [14] Y. Aktaş, E. Yılmaz İnce, A. Çakır and A. Kutlu, "Wordnet ve Bilgisayar Ağ Terimleri Sözlüğünün Oluşturulması," in *Akademik Bilişim*, 2016.
- [15] K. Denecke, "Using SentiWordNet for multilingual sentiment analysis," in *2008 IEEE 24th International Conference on Data Engineering Workshop*, Cancun, Mexico, 2008.
- [16] B. İ. Sevindi, *Türkçe Metinlerde Denetimli ve Sözlük Tabanlı Duygu Analizi Yaklaşımlarının Karşılaştırılması*, Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü, 2013.
- [17] R. Gandhi, "Support Vector Machine — Introduction to Machine Learning Algorithms," 2018. [Online]. Available: www.towardsdatascience.com.