

# Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi

*Araştırma Makalesi/Research Article*

 Önder ŞAHİNASLAN<sup>1\*</sup>,  Hüseyin DALYAN<sup>2</sup>,  Ender ŞAHİNASLAN<sup>3</sup>

<sup>1</sup>Bilişim Bölüm Başkanlığı Maltepe Üniversitesi, İstanbul, Türkiye

<sup>2</sup>Bilgisayar Mühendisliği Bölümü, Maltepe Üniversitesi, İstanbul, Türkiye

<sup>3</sup>Bilgisayar Mühendisliği Bölümü, Trakya Üniversitesi, Edirne, Türkiye

[ondersahinaslan@maltepe.edu.tr](mailto:ondersahinaslan@maltepe.edu.tr), [huseyindalyan@hotmail.com](mailto:huseyindalyan@hotmail.com), [dr.endsa@gmail.com](mailto:dr.endsa@gmail.com)

(Geliş/Received:10.10.2021; Kabul/Accepted:14.04.2022)

DOI: 10.17671/gazibtd.999960

**Özet**— Bu çalışma, Naive Bayes sınıflandırma algoritması kullanılarak sosyal medya verileri üzerinden dünya dillerinde yapılan yorum ve paylaşımların anlaşılmasına katkıda bulunmayı amaçlamaktadır. Duygu analizi veya fikir madenciliği, doğal dil işleme konularının bir alt bölümünde yer alır ve sosyal medya üzerinde yer alan verilerin anlamlandırılmasını sağlar. Yapay zekâ ve makine öğrenmesi gibi konularla birlikte günümüzde son derece popüler bir alandır. Küresel bir köye dönüşen dünyada, insanlar sosyal medya araçları üzerinden görüş, düşünce ve o konu hakkında memnuniyet durumları gibi pek çok veriyi paylaşmaktadır. Anlamlandırılmayı bekleyen bu veriler üzerinden insanların duygu ve düşünceleri ortaya çıkartabilmek mümkündür. Son dönemlerde Türkçe dilinde bir takım duygu analizi çalışmalarının yapıldığı görülmektedir. Ancak dünyadaki farklı dil ve lehçelerde yapılan yorumlar üzerinden çok dilli duygu analiz çalışması sınırlıdır. Bu çalışmada, Türkçe, İngilizce, Almanca, Fransızca, Arapça, Rusça ve Korece gibi farklı dünya dillerinden YouTube yorum verileri kullanılmıştır. PHP ile elde edilen bu ham veriler üzerinden anlamsız veriler temizlenerek belirli bir kalitede veri seti elde edilmiştir. Bu yorum verilerini İngilizce'ye çevirmek için Google çeviri aracı API'sı kullanılmıştır. Duygular, PHP Niiknow kütüphanesi üzerinden Naive Bayes algoritması kullanılarak pozitif, negatif ve nötr olarak sınıflandırılır. WEKA ile yapılan veri analizinde doğru sınıflandırma oranı %65,56 olarak bulunmuştur. Bu çalışma, 15.082 veri seti ve Google Çeviri aracı tarafından desteklenen 108 dil sayısı ile sınırlıdır.

**Anahtar Kelimeler**— duygu analizi, makine öğrenmesi, naive bayes, weka, youtube

## Multilingual Sentiment Analysis on YouTube Data Using Naive Bayes Classifier

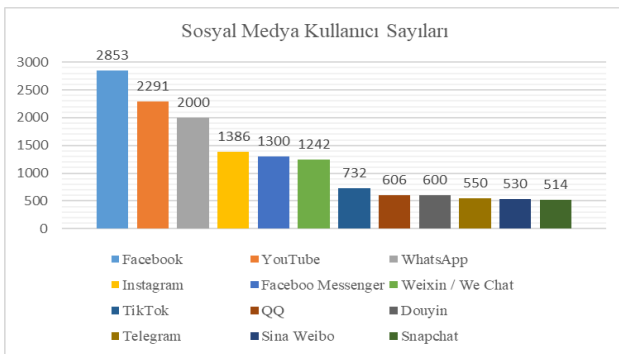
**Abstract**— This study aims to contribute to the understanding of comments and shares made in world languages over social media data using the Naive Bayes classification algorithm. Sentiment analysis or opinion mining is a subsection of natural language processing and it is ensured that the data on social media can be interpreted. It is an extremely popular field today, along with topics such as artificial intelligence and machine learning. In the world that has turned into a global village, people share a lot of data such as opinions, thoughts and satisfaction about that subject through social media tools. It is possible to reveal people's feelings and thoughts through these data waiting to be interpreted. It is seen that some sentiment analysis studies have been carried out in Turkish language recently. However, multilingual sentiment analysis work is limited on interpretations made in different languages and dialects in the world. In this study, YouTube comment data from different world languages such as Turkish, English, German, French, Arabic, Russian and Korean have been used. A certain quality data set has been obtained by cleaning the meaningless data over these raw data obtained with PHP. The Google translation tool API has been used to translate this comment data into English. Emotions are classified as positive, negative and neutral using the Naive Bayes algorithm via the PHP Niiknow library. In the data analysis performed with WEKA, the correct classification rate has found to be 65.56%. This study is limited to 15,082 data sets and the number of 108 languages supported by the Google Translate tool.

**Keywords**— sentiment analysis, machine learning, naive bayes, weka, youtube

## 1. GİRİŞ (INTRODUCTION)

Duygu, bireyin iç dünyasında yaşanan etki ve yankılara bağlı olarak davranışlarının şekillenmesinde rol alan, psikolojik temelli önemli bir kavramdır. Bireyin durum veya olaylar karşısında vereceği tepkiyi belirler. Bireyler çevreden algıladığı duygu ile farklı durumlar karşısında farklı davranış gösterirler. Tepkilerini, jest ve mimikleri aracılığı ile fiziksel olarak yansıtabildiği gibi yazılı olarak ta ortaya koyabilirler. Bu tepkiler olumlu, olumsuz ve tepkisiz şeklinde üç sınıfa ayrılır.

Günümüzde nesnelerin interneti (IoT), makineler arası iletişim (M2M), yapay zekâ, kuantum bilgisayarlar, blok zincir teknolojisi gibi birçok gelişmeler vardır [1]. Endüstri 4.0, bilgi teknolojilerinin gelişmesi ve internet/mobil uygulamaların yaygınlaşması ile birlikte dijital platformlar, insan hayatında giderek daha büyük yer almaktadır [2]. Bu hızlı gelişim ve teknolojik birliktelik birçok fırsat sunmaktadır [3]. Yeni teknoloji ve uygulamalar marka ve ürünlerin pazarlanmasında farklı yöntem ve metotların kullanılmasına aracı olmuştur. Bu değişim sosyal medya platformları üzerinde kendisini göstermektedir. Sosyal ağlarda artan kullanıcı yoğunluğu, etkin kullanım ve sıkı takip yeni fırsatlara kapı aralamaktadır. Ürün, marka, kurum veya kişiler hakkında düşünceler kolayca paylaşılmaktadır. İnsanlar sosyal ağlar aracılığıyla duygu, düşünce, ihtiyaç ve beklentilerini de dile getirmektedir [4]. Yapılan bir yorum milyonlarca kişiye anında ulaşabilmektedir. Sosyal ağlardan edinilen bir bilgi kısa zamanda dolaşıma uğrayabilmekte, bireylerin tercihlerinde değişikliklere neden olmaktadır. İhtiyaç duyulan bir konu hakkında öncelikli olarak arama motorları ve sosyal medyadan yapılan yorumlar incelenmektedir. Bazı alanlarda çeşitli uygulamalar kullanılarak yapılan bu yorumlar karşılaştırmalı olarak değerlendirilebilmektedir. Dijital platformlar üzerinde yer alan yorum verileri üzerinden fikir madenciliği sınıflandırması algoritmaları ile duygu analizi çalışmaları yapılmaktadır. Bu araştırmaların birçoğunun tek bir dil üzerinde gerçekleştirilmiştir. Örneğin sadece İngilizce veya sadece Türkçe yorumlar elde edilerek analiz edilmiştir. Bu çalışmada ise çok dilli duygu analizi üzerinde bir farklılık oluşturulmaya çalışılmıştır. Günümüzde Facebook, Twitter, Youube ve Instagram gibi sosyal medya uygulamaları yoğun olarak kullanılmaktadır.



Şekil 1. Sosyal medya kullanıcı sayıları (milyon)  
(Social media user numbers) (millions)

Statista [5], Temmuz 2021 verilerine göre dünya genelinde en popüler sosyal medya kullanıcı sayıları Şekil 1'de gösterilmektedir. Sunulan bu veriler incelendiğinde tüm dünya genelinde en çok tercih edilen ve kullanılan sosyal medya platformlarından Facebook, Youtube ön planda yer aldığı görülmektedir. Kullanıcı sayıları bakımından ilk beş uygulama detaylı incelendiğinde 2021 yılı ikinci çeyreğinde Facebook (2 Milyar 895 Milyon), Youtube (2 Milyar 291 Milyon), WhatsApp (2 Milyar), Instagram (1 Milyar 386 Milyon), Facebook Messenger (1 Milyar 300 Milyon) kullanıcıya sahiptir. Bu sayı ile Facebook dünya çapında en fazla kullanıcı bulunan sosyal medya platformu iken Youtube ise en büyük ikinci platformdur.

Youtube, çok farklı ve sayıda içeriğin barındıran bir videolu içerik platformudur. Yoğun kullanılan bu sosyal medya platformu üzerinde insanların duygu ve düşüncelerini barındıran yorumlar zamanla büyük veri yığınlarına dönüşmektedir. Depolanan bu veriler duygu analizi gibi sınıflandırma işlemlerinde kullanılabilir. Bu platformda, farklı ülkelerden içerik üreticileri farklı ülkelerde video içeriği üretmektedir. Dünya'nın herhangi bir yerinden internete erişebilen herkes bu platform üzerinden kendi e-posta adresini kullanarak bağlanabilmekte, video üretip yayımlayabilmekte ve yayımlanan videolar üzerinde yorumlarda bulunabilmektedir. Bu bağlamda Youtube çok dilli bir içeriğe sahiptir. Çeşitli ürün grupları veya kurumlar Youtube içerik üreticileriyle reklam anlaşmaları yapabilmekte ve onların takipçilerine ürün tanıtımı yaptırabilmektedir. İçerik üreticiler buradan gelir elde etmektedir. Bu anlamda, çok dilli evrensel bir yapı ve büyük bir veri birikimi söz konusudur [6].

Duygu analizi, insanların herhangi bir hizmet, ürün veya durumlara karşı olan görüşlerini değerlendirme ve analiz etme imkânı sağlar. Veri analiz ve değerlendirmesi sonucunda elde edilen bilgilerle daha doğru kararların daha kolay alınabilmesine imkân sunmaktadır. Toplanan verilerin anlamlandırılmasında birtakım yöntemler geliştirilmiştir. Bunlar temelde makine öğrenme ve sözlük tabanlı yöntemler olarak ikiye ayrılmaktadır. Şekil 2'de duygu analizi yaklaşımları görülmektedir.



Şekil 2. Duygu analizi yaklaşımları [7]  
(Sentiment analysis approaches)

Bu yöntem ve yaklaşımlar sadece duygu analizinde değil aynı zamanda doğal dil işleme, e-posta filtreleme ve spam

temizleme gibi diğer alanlarda kullanılmaktadır. Makine öğrenmesi, öğrenme biçimlerini benzer bir biçimde yeniden oluşturmak yani taklit etmek için, algoritma ve verilerin kullanılması ile kendini yenilemekte ve doğruluk oranlarını sürekli arttırmakta olan, bir bilgisayar bilimi alanıdır [8]. Duygu analiz çalışmasında kullanılan olasılık tabanlı Naive Bayes sınıflandırma yöntemi aynı zamanda makine öğrenmesi yaklaşımıdır. Duygu sınıflandırmasında insan ve makine zekâsının temelinde yer alır [9].

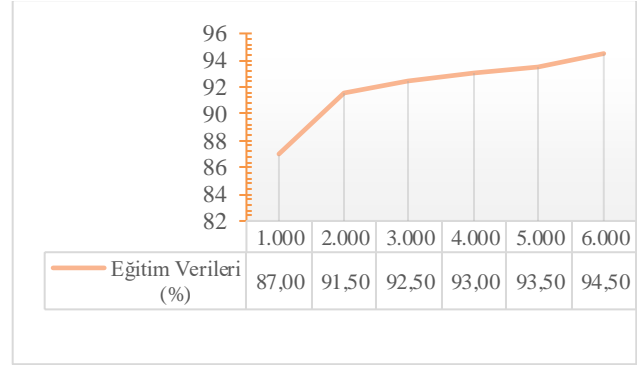
## 2. LİTERATÜR TARAMASI (LITERATURE REVIEW)

Sosyal medya, sahip olduğu büyük miktardaki veri hacmiyle sayesinde duygu analizi çalışmalarında araştırmacıların ilgi noktası olmuştur. Fikir madenciliği olarak bilinen duygu analizi için büyük veri hacmine veriye ihtiyaç duyulur. Youtube video yorumları analiz için yeterli veri hacmine sahiptir. Youtube yorumları çok dilli fikir madenciliğinde kullanılmaktadır [10]. İnsanlar bu video gösterim platformu üzerinden Google çeviri ve alt yazı desteği sayesinde daha önce hiç duymadığı farklı dillere ait videoları bile izleme ve hatta bunlar üzerinden kendi görüşlerini bildirebilmektedir. Bu biriken veriler, farklı toplum bireylerinin birer görüş ve duygusu anlamak için bir fırsattır. Günümüzde bilgi teknolojileri imkânları kullanılarak bu sosyal medya verilerin analizini yapabilmek mümkündür. Bu analiz sonucu toplumların duygularının daha iyi anlaşılmasına yardımcı olacaktır. Sosyal medya üzerinden yapılan çalışmalar İngilizce ağırlıklı tek bir dil üzerinden olması nedeniyle sınırlı kalmaktadır. Bu sınırlılık diğer dünya dil ve toplumların anlaşılmasında zorluklar yaşatmaktadır [11]. Yaşanan bir takım zorluklara rağmen gelişen teknolojik imkânlarla veriler üzerinden çok dilli analizlerin yapılabilmesi mümkündür. Aksi takdirde tek bir dil üzerinde yapılan duygu analiz çalışmaları küresel ölçekte duyguların anlaşılması önünde bir engel teşkil edecektir. Bu bağlamda, çok dilli duygu analizi diller arası duygu analiz süreçlerinde önemli bir safhayı temsil etmektedir [12]. Dünyanın farklı bölgelerindeki insanlar sosyal medyada kendi dillerinde paylaştığı bir yorumu ve deneyimi, diğer toplumların takip etmesi, karşılaştırması ve tahlil ederek katkıda bulunması büyük bir zenginlik katmaktadır.

Globalleşen dünyada çok dilli duygu analizi çalışmaları artan şekilde popülerlik kazanmaktadır. İnternet üzerinden her dile, kültüre, duygu ve düşünceye ait bilgi elde edilebilmektedir. Birçok duygu analizi yöntemi sözcüklere ve bunları içermekte olan duygulara dayanmaktadır. Bunların analizinde sözlük tabanlı duygu analizleri çözüm gösterilmektedir. İnternet üzerinde sürekli artmakta olan çok dilli verilerin yorumlanması ve duygu analizi aşamasında bu duygu analizi yöntemi artık bir gerekliliktir. Hogenboom ve ark. [13] göre, metin otomatik olarak hedef dile çevrilmekte ve bu dil üzerinde geliştirilen sözlük aracılığı ile duygu analizi gerçekleştirilmektedir.

Ma Ma ve arkadaşları tarafından yapılan çalışmada [14], örnek veri setleri aracılığıyla 'spam' olarak değerlendirilen veya normal e-postalar Naive Bayes algoritması kullanılmıştır. Bu yöntemle öğrenilen bilgiler kullanarak

gelecek e-posta iletilerinin daha doğruya yakın oranda sınıflandırılmasına, yüksek doğrulukta sonuçlar ürettiği gözlemlenmiştir. Bu çalışmada elde edilen sınıflandırma sonuçları Şekil 3'de gösterilmektedir.



Şekil 3. E-posta sınıflandırma sonuçları  
(Email classification results)

Çalışmada kullanılmakta olan verilerin miktarı artış gösterdikçe doğruluk oranının %96'ya kadar yaklaştığı görülmektedir. Yeterli veri seti olduğu durumlarda Naive Bayes algoritması sınıflandırma başarı düzeyini artırır. Bu aynı zamanda Naive Bayes sınıflandırıcısıyla öğretilen veri setleri kullanılarak çok dilli duygu analiz çalışması içinde kullanılabileceğinin bir göstergesidir. Khan ve arkadaşlarının yaptığı çalışmada [15]; Naive Bayes sınıflandırıcısı ile Youtube sosyal medya platformu içerisinde yer alan yorum verileri temel alınarak duygu analizi işlemi gerçekleştirilmiştir. Duygu analizi aşamasında, en az iki ve üzeri markanın (Iphone, Android) karşılaştırıldığı videoların altına yazılan yaklaşık 8.000 adet yorum verileri kullanılmıştır. Bu verilerde, kullanıcının kişisel görüşü olumlu, olumsuz veya nötr biçimde değerlendirilmiştir. Yapılan yorumlar üzerinden kullanıcı duygularını anlamlandırmak amacıyla Naive Bayes sınıflandırıcısı kullanılmış ve WEKA ile doğruluk analizi yapılmış ve çok etiketli sınıflandırma başarı düzeyi %48.42 - %61.18 arasında bulunmuştur.

Duygu analizinde birçok makine öğrenme ve veri madenciliği yöntemlerinden yararlanır. Naive Bayes sınıflandırma algoritması, makine öğrenmesi alanında yaygın bir kullanıma sahiptir. Kullanımı kolay ve anlaşılır olduğunda dolayı tercih edilmektedir. Ayrıca, basit verilerle hızlı ve yüksek doğrulukta sonuçlar üretebilmesi başka bir tercih sebebidir [16]. Çok dilli duygu analizi, artmakta olan veri miktarı ve evrenselleşme sonucunda büyük bir önem kazanmaktadır [17]. Belli bir ölçüğe sahip veri seti üzerinden yapılan çalışmalarda Naive Bayes algoritması başarılı sonuçlar üretmiştir. Veri setinin hacmi ile doğruluk oranı artışında pozitif bir ilişki gözlemlenmiştir. Yeterli veri setine sahip basit verilerde Naive Bayes algoritmasının kullanımı sonucunda (%96) başarı oranı elde edilmiştir [16]. Yılmaz ve Orman yaptıkları çalışmada [18], LSTM derin öğrenme tekniği kullanarak Twitter verileri üzerinden duygu analizi yapmışlardır. Farklı veri kümeleri üzerinde %85 - %97 arasında doğruluk başarıları elde edilmiştir. Bu başarı

oranının veri kümesindeki veri sayısı ile ilişkili olduğu sonucuna varmışlardır. Sel ve Hanbay çalışmalarında [19], ön eğitilmiş olan modellerden faydalanarak 5.292 adet Türkçe Twitter gönderileri üzerinden cinsiyet tespiti yapılmıştır. Çalışmada makine öğrenmesi yöntemleri, derin öğrenme modelleri ve ön eğitilmiş dil modelleri kullanılmıştır. Yaş, meslek ve cinsiyet gibi çeşitli parametreleri değerlendirmişlerdir. Çalışmada, SVM makine öğrenmesi tekniği, Bi-LSTM, CNN derin öğrenme yöntemleri ile Bert, Bert-128k, Distilbert, Electra Discriminator ve Electra Generator yöntemleri kullanılmıştır. Çalışmada sonucunda bu modellerde (%64,52 - %80,12) arasında doğruluk başarı oranı elde edilmiştir. En yüksek başarı doğruluk oranı Bert-128k modelinde bulunmuştur. Bert modeli ile ise  $n=1$  için %70  $n=2$  için %73 başarı elde edilmişlerdir.

Karahmetoğlu ve ark. [20], Twitter verilerini kullanarak covid-19 süreci hakkında 4.575 adet Türkçe veri üzerinden bir duygu analiz çalışması yapmışlardır. Çalışma sözlük tabanlı ve Bert makine tabanlı algoritma üzerinden yapılmıştır. Sözlük tabanlı duygu analizinde SentiTurkNet Türkçe kelime sözlüğü kullanılmış ve başarı oranı %99 elde edilmiştir. Bert tabanlı makine öğrenmesi modeli kullanıldığında doğruluk başarı oranı %98 olarak bulunmuştur. Samuel ve ark. [21], Twitter verileri üzerinden benzer bir çalışmada Naive Bayes yöntemi ve Lojistik Regresyon modeli kullanılmıştır. Naive Bayes yöntemi kullanıldıklarında 77 karakterden küçük Twitter verilerinde başarı oranı %91.43 gibi yüksek seviyede bulunurken 120 karakterden küçük olan veriler üzerinde test edildiğinde bu doğruluk başarı oranının %57.14'e düştüğü gözlemlenmiştir.

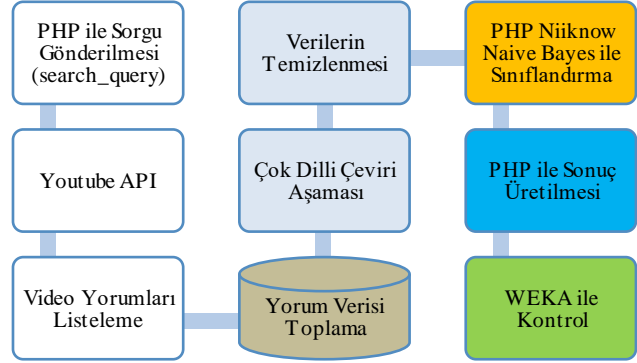
### 3. YÖNTEM (METHOD)

Bu çalışmada, çok dilli duygu analizi uygulamaları aktarılmaktadır. Veriler Youtube sosyal medya platformu üzerinden elde edilmiştir. Bu platform, dünya genelinde Facebook uygulamasından sonra en çok kullanılan sosyal medya platformu olup veri madenciliği, duygu analizi gibi alanlar için yeterli miktarda veriye sahiptir. Çok dilli duygu analiz çalışmasında farklı dünya dillerindeki verilere ihtiyaç duyulur. Bu sosyal medya uygulamasının farklı dillere ait yorumları barındırması ve bunları Youtube API üzerinden paylaşılması nedeniyle çalışmada Youtube uygulama verileri tercih edilmiştir. Farklı dillere ait yorumlar içeren bu veriler, Google Çeviri aracı kullanılarak İngilizce diline çevirisi yapılmıştır. Çeviri aracının desteklediği veriler ihmal edilmiştir. WEKA, veri madenciliği işlerini kolaylıkla yapabilmek için makine öğrenme algoritmalarını barındırmakta olan bir koleksiyon yazılımı olmaktadır. Veri sınıflandırma, regresyon, hazırlama, birliktelik ve kümeleme kuralları ile görselleştirme ve madencilik araçlarını içermektedir [22].

#### 3.1. Süreç Akışı (Process Flow)

Bu çalışmada uygulanan yönteme ilişkin ana süreç aşamaları ham veri kaynağını belirleme, veri hazırlama, veri toplama, işleme, analiz, sınıflandırma, sonuç ve

değerlendirme aşamalarından oluşmaktadır. Süreç aşamaları Şekil 4'de gösterilmektedir.



Şekil 4. Çok dilli duygu analizi süreç aşamaları (Multilingual sentiment analysis process stages)

Çok dilli duygu analizi ana süreç aşamasında; ham veri kaynağı olarak Youtube sosyal medya kullanılmıştır. Youtube API'si üzerine PHP ile gönderilen sorgu sonucunda elde edilen yorum verileri toplanarak veri hazırlama evresine geçilmiştir. Veri işleme evresinde, toplanan bu verilerin İngilizce diline çevirisi yapıldıktan sonra veri temizleme ve dönüştürme aşamasına geçilmiştir. Veri sınıflandırma aşamasında PHP yazılım dilinde Naive Bayes temelli bir sınıflandırmaya sahip olan Nüknov duygu kütüphanesi kullanılmıştır. PHP üzerinden üretilen sonuçlar WEKA aracı kullanılarak değerlendirilmiştir.

#### 3.2. Örneklem (Sample)

Bu çalışmada yer alan örneklem grubu, herhangi bir dilde bir cümlelerin veya anahtar kelimelerin aranması sonucunda ortaya çıkan videolara ait toplam 15.200 adet yorum verisinden oluşmaktadır. Bu ham verilerden çalışma kapsam ve amacına uymayan 118 adet veri temizlenerek veri analiz ve sınıflandırma çalışmaları 15.082 adet veri seti üzerinden yürütülmüştür. Bu veriler Türkçe, İngilizce, İtalyanca, Almanca, Fransızca, Rusça, Çince, Korece, Japonca gibi farklı dünya dillerinde yapılan yorumlardan oluşmaktadır. Dillere göre veri hacimleri karışık durumda olup belirli bir ağırlıklıdırma gözlemlenmiştir.

#### 3.3. Veri Toplama (Data Collection)

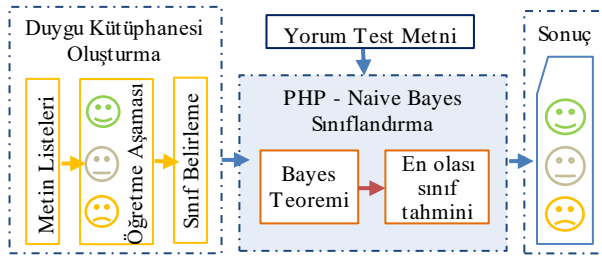
Veri toplama aşamasında Youtube sosyal medya uygulaması üzerinde yer alan ve herkese açık olan veriler üzerinden PHP ile Youtube API [23] kullanılarak elde edilmiştir. Youtube API sayesinde belirli bir video, oynatma listesi, kanal kaynağı, bir konu, kelime veya terime ait istenilen videolara ait paylaşılmış yorum verilerini elde etmek mümkündür. Toplanan bu veriler MySQL [24], Microsoft SQL(MSSQL) [25] gibi istenilen bir veri tabanına yüklenebileceği gibi CSV (Virgülle Ayrılan Değerler) metin dosyası veya Excel dosyasına taşınabilir. Yapılan çalışmada verilerin toplanmasında geliştirilen bir uygulama aracı kullanıldı ve veriler CSV metin dosyası üzerine kaydedildi.

### 3.4. Veri İşleme (Data processing )

Bu çalışma setinde yer alan verilerin tamamı Google çeviri API [26] aracılığıyla İngilizce diline tercümesi yapıldı. Halen, Google çeviri aracı kullanarak 109 farklı dil üzerinde çeviri işlemini yapabilmektedir. Bu özelliği sayesinde, farklı diller kullanan dünya nüfusun neredeyse tamamına yakınına ait verilerinin değerlendirilebilmesi mümkündür. Bu aşamada ayrıca çeviri sonucunda elde edilen veri seti üzerinde belirli veri kontrolleri gerçekleştirilmiştir. Yapılan kontrolde, sadece sayı, sembol (@, %, /, \* vb.), noktalama işaretleri, e-posta, internet adresi gibi çalışma amaç ve kapsamına girmeyen veriler tespit edilmiştir. Tespit edilen bu verilerin temizliği yapılarak çalışma kapsamından çıkartılmıştır.

### 3.5. Veri Sınıflandırma (Data Classification)

Verilerin sınıflandırma aşamasında test verisi olarak çalışmada toplanan yorum metinlerinin İngilizce dilindeki karşılıkları kullanılmıştır. Pozitif, negatif ve nötr duygu sınıflarında etiketlenmiş metinlerden oluşan duygu kütüphanesi verileri ile yorum test metni verileri üzerinden PHP Naive Bayes sınıflandırmasın tabii tutulmuştur. Sınıflandırmada, Bayes teoremi sonucunda elde edilen olasılık değerleri üzerinden en olası duygu sınıfının belirlenmesine çalışılmıştır. Sınıflandırma süreç aşamalarına ait mimari yapı Şekil 5’de gösterilmektedir.



Şekil 5. Duygu sınıflandırma mimarisini (Sentiment classification architecture)

Naive Bayes sınıflandırmasında PHP 'Niiknow\Bayes' kütüphane fonksiyonlarından yararlanılmıştır [27]. Sınıflandırma program akışı, kullanılan fonksiyonlar ve kullanım amaçları şu şekildedir:

- `use Niiknow\Bayes`; PHP programı içerisine belirtilen kütüphanesi eklenir.
- `$classifier = new \Niiknow\Bayes()`; yeni bir naive bayes sınıflandırıcı çalışmaya hazır hale getirilir.
- `$classifier->learn(metin, duygu_sınıfı)`; metodu ile metnin pozitif, negatif ve nötr duygu sınıflarından hangisine ait olduğu öğretilerek bir duygu kütüphanesi oluşturulur. Bu metinler test aşamasında daha yeni bir metnin sınıflandırmasında kullanılır.
- `$classifier->probabilities(metin)`; Belirlenen her duygu kategorisi için olasılıkları hesaplanır.
- `$classifier->categorize(metin)`; Test edilen yoruma ait metnin en olası duygu sınıfı belirlenerek kategorize edilir.
- `$classifier->toJson()`; sınıflandırma sonucunu JSON standart metin türünde üretimi gerçekleştirilir.

### 3.6. Naive Bayes Sınıflandırma (Naive Bayes Classifier)

Naive Bayes sınıflandırıcısı veri madenciliği, makine öğrenmesi, duygu analizi gibi alanlarda yaygın kullanıma sahip denetimli öğrenme sınıfında bir algoritmadır. Belirli bir özellik kümesinin belirli bir sınıfa ait olma olasılığını tahmin etmede kullanılır. Sınıflandırmada, Bayes karar teoremini esas alır [28], olasılık hesaplamalarını kullanarak en yüksek olasılığa sahip kararı seçmeyi amaçlar [29]. Her bir niteliğin sınıf içerisindeki diğer niteliklerden bağımsız olduğu kabul edilir. Bu sınıflandırıcı bir takım basit hesaplamalar yaparak olasılık tabanlı olarak bir olayın gerçekleşme ihtimalini hesaplar.

Bayes karar teoremi eşitliği (1)'de yer almaktadır [30] [31].

$$P(S_i|X) = \frac{P(X|S_i) P(S_i)}{P(X)} \quad (1)$$

X olarak seçilen bir örneğin i duygu sınıfına (pozitif, negatif, nötr) ait olma ihtimali  $P(S_i|X)$ , i duygu sınıfının önsel yani henüz veri testine başlamadan öncesinde duygu kütüphanesinde yer alan duygu sınıfına ait ilk olasılığı  $P(S_i)$ , seçilen bir örneğin X olma önsel olasılığı  $P(X)$ , i duygu sınıfına ait bir örneğin X olma olasılığı ise  $P(X|S_i)$  olarak ifade edilmektedir.

Yeni bir örnekle karşılaşılması durumunda bu örneğin üyelik olasılığını bulmada hesaplanan olasılık değerleri dikkate alınarak olasılığı en yüksek olan sınıf bulunur [32]. Bu amaçla ele alma bir örnek için eğitim seti kullanılarak her bir duygu sınıfı için  $P(S_i|X)$  olasılıkları ayrı ayrı hesaplanır. Bu hesaplamalar sonucunda olasılık değeri en büyük hesaplanan duygu sınıfı ele alınan örneğin sınıfı olarak kabul edilir. Bu eşitlikte  $P(X)$  tüm olası sınıflar için sabit olduğundan  $P(S_i|X) = P(X|S_i) P(S_i)$  olasılığı için en büyük değer aranması yeterlidir. Test kümesindeki bir verinin eğitim kümesinde bir karşılığı mevcut değilse o veri için tahmin yapılmamaktadır [33]. Naive Bayes yöntemi tahmine dayalı modelleme yapmak için basit bunun yanında güçlü bir algoritmadır [34]. Az sayıdaki veriyle bile güçlü tahmin kabiliyetine sahiptir. Bu tür kullanışlı ve avantajlı yanlarından dolayı pek çok alanda tercih edilen ve uygulanan bir sınıflandırıcıdır.

### 3.7. Veri Kontrol (Data Control)

Elde edilen sonuçların kontrol ve değerlendirilmesinde WEKA uygulaması tercih edilmiştir. Bu araç açık kaynak kodlu bir veri madenciliği programıdır [22].

## 4. BULGULAR (RESULTS)

Bu çalışmada, toplam 15.082 adet çok dilli veri üzerinden analiz ve sınıflandırma çalışması yürütülmüştür. Bu veriler olumlu (pozitif), olumsuz(negatif) ve tarafsız(nötr) olmak üzere üç farklı duygu sınıfında toplanmıştır. Verilerin duygu sınıflandırma durumlarına ilişkin sayı ve yüzde dağılımları Tablo 1’de yer almaktadır.

Tablo 1. Veri sınıflandırma tablosu  
(Data classification table)

Duygu	Adet	Oran (%)
Pozitif	5.167	34,25
Negatif	4.644	30,81
Nötr	5.271	34,94
Toplam	15.082	100,00

Elde edilen bu duygu sınıflandırma sonuçlarına göre 5.167 adet yorum verisi olumlu (%34,25), 4.644 adet veri olumsuz (%30,81) ve 5.271 adet yorum verisi ise tarafsız (%34,94) duyguya sahiptir. Çalışmada pozitif olarak sınıflandırılan yorumlar üzerinden seçilen ve Türkçe diline çevrilen bazı cümle örnekleri Şekil 6'da gösterilmektedir.

- "Çok başarılı "
- "Memnun kaldım "
- "Çok güzeldi sadece beğendiğimi söylemek istedim."
- "Sektörde resmi olarak açık bir lider, bundan gerçekten hoşlanıyorum, bravo."
- "Süper"
- "Tebrikler"



Şekil 6. Pozitif duygu örnekleri  
(Samples of positive emotions)

Bunlardan "başarılı", "beğendim", "süper" veya "hoşuma gitti" gibi olumlu anlam yüklü kelimeleri içeren sözcükler pozitif olarak değerlendirilirken "beğenmedim", "kötü" "yetersiz", "memnun değilim" veya "hoşlanmadım" tarzı olumsuz kelimeleri barındırmakta olanlar ise negatif olarak sınıflandırılmaktadır. Veri seti içerisinde yer alan ve negatif olarak değerlendirilen duygu sözcüklerine ait örneklerden birkaç tanesi Şekil 7'de gösterilmektedir.

- "Bu ürünü almanızı tavsiye etmiyorum, beğenmedim."
- "Tasarım oldukça eksik ve yetersiz."
- "Cihazı 3 gün önce aldım ama memnun değilim."
- "Fare, büyük ve ağır oyunlar için iyi çalışmıyor"
- "Kullanımı biraz rahatsız edici."
- "Hoşlanmadım"
- "Çok kötü bir ürün"



Şekil 7. Negatif duygu örnekleri  
(Samples of negative emotions)

Çalışmada herhangi bir duygu barındırmayan tarafsız sözcükler ise nötr olarak değerlendirilmiştir. Bu türde değerlendirilen yorum sözcüklerine ait örneklerden birkaç tanesi Şekil 8'de gösterilmektedir.

- "2 yıldır kullanıyorum"
- "Ne desem bilemiyorum"
- "Telefonun şarj süresi 21 ile 24 saat arasındadır."
- "Daha klasik ve ince tasarımlar daha çoksatar."
- "Kararsızım."
- "Ürün bugün elime geçti"
- "Kullanıyorum, şöyle, böyle"



Şekil 8. Nötr duygu örnekleri  
(Samples of neutral emotions)

Bu duygu örneklerinde ise bir tarafsızlık söz konusu olup herhangi bir beğenme veya beğenmeme durumu söz konusu değildir.

#### 4.1. WEKA Analiz Sonuçları ( WEKA Analysis Results )

Bu çalışmada Naive Bayes algoritması kullanılarak sınıflandırılan verilerin analiz ve kontrolünde WEKA yazılımı kullanılmıştır. Bu araç üzerinden gerçekleştirilen veri sınıflandırma analizine ait sonuçları Şekil 9'da yer almaktadır.

Correctly Classified Instances	9888	65.5616 %
Kappa statistic	0.4835	
Mean absolute error	0.2476	
Root mean squared error	0.4112	
Relative absolute error	55.8032 %	
Root relative squared error	87.2866 %	
Total Number of Instances	15082	

Şekil 9. Weka sınıflandırma analiz sonuçları  
(Weka classification analysis results)

Analiz işleminde kullanılan toplam örnek sayısının 15.082 olduğu görülmektedir. Bu veriler içerisinde doğru sınıflandırması yapılan örnek sayısı 9888 ve oranı %65,56 olarak bulunmuştur. İki farklı ölçüm arasındaki karşılaştırmalı uyuşmanın güvenilirliğini ölçen kappa istatistik katsayısı 0.4835, mutlak hata ortalaması 0.2446, kare hata ortalaması 0.4112. Göreceli mutlak hata oranı %55,82, kök ortalama kare hata oranı %87,29 olarak elde edilmiştir.

Analiz sonuçlarında yer alan kesinlik oranları tahmin edilmekte olan pozitif olayların oranını içermektedir. Kesinliği doğru kabul edilen pozitif sayısının tahmin edilen pozitif sayısına oranını aktarmaktadır. Hassasiyet sonuçları ise, doğru olarak tespit edilmekte olan pozitif olayların miktarını içermektedir. Doğru olarak işaretlenen pozitif tahmin miktarının toplam pozitif miktarına oranını aktarmaktadır. Yorum sözcük verilerinin sınıflandırılması sonucunda bu veriler olumlu, olumsuz veya nötr olarak belirlenmektedir. Verilere ait duygu sayı ve ağırlıklarına ilişkin WEKA uygulaması üzerinden elde edilen değerler Şekil 10'da gösterilmektedir.

Selected attribute			
Name: Duygu		Type: Nominal	
Missing: 0 (0%)		Distinct: 3	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	1	5167	5167.0
2	0	4644	4644.0
3	2	5271	5271.0

Şekil 10. Weka duygu sayı ve ağırlık dağılımları  
(Weka sentiment number and weight distributions)

'Label' sütun etiketinde yer alan rakamlar duygu durumlarını ifade etmektedir. Burada 1:pozitif, 0:negatif, 2 değeri ise nötr duygu durumunu göstermektedir. Bu duygu

sınıflarına göre ayrıntılı doğruluk oranları Şekil 11'de gösterilmektedir. Gerçek pozitif değeri (*TP Rate*), gerçek negatif değeri (*FP Rate*), hassasiyet değeri (*Precision*), gerçek olumlu oran (*Recall*), gerçek olumlu oranı ile hassasiyet değerinin ağırlıklı ortalaması (*F-measure*) ve MCC değerleri yer almaktadır.

=== Detailed Accuracy By Class ===						
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0,674	0,227	0,607	0,674	0,639	0,437
	0,751	0,140	0,704	0,751	0,727	0,600
	0,554	0,151	0,664	0,554	0,604	0,423
Weighted Avg.	0,656	0,174	0,657	0,656	0,654	0,482

Şekil 11. Duygu sınıfları doğruluk oranları  
(Sentiment classes accuracy rates)

Duygu sınıfı doğruluk oranlarına göre gerçek pozitif oran 0,67 ile 0,65 arasında değişiklik göstermektedir. Hassasiyet değerleri ise 0,60 ile 0,65 olarak görülmektedir. Hassasiyet ağırlıkları ortalaması 0,63 ile 0,65 arasında ki çizgidedir. Bu ölçümde kullanılmakta olan puan gerçek olarak varsayılan pozitiflerin yani hassasiyetin ve ayrıca kesinlik oranının ağırlıklı ortalaması olarak belirlenmektedir. MCC kısaltması Mathews'in korelasyon katsayısına karşılık gelmektedir. Korelasyon hesaplamada kullanılan bir matristir.

## 5. TARTIŞMA (DISCUSSION)

Duygu analizi, makine öğrenmesi alanında yer alan ve popülerliği gün geçtikçe artan bir konudur. Analiz yapılacak veriler sosyal medya platformlarından elde edilmektedir. Sosyal medya üzerinden gerçekleştirilen çalışmalar ise güncel olmakla birlikte artış göstermektedir. Sosyal medya uygulamalarından verilerin elde edilmesinde, ilgili uygulama geliştiricilere önemli bir pratiklik sağlamaktadır. Araştırmacılar için sosyal medya API'leri üzerinden verilerin elde edilebilir olması, makine öğrenmesi, yapay zekâ alanında yaşanan gelişmeler ve bilgi teknolojilerindeki diğer imkânlar yanında insanı/toplumunu alama duygusunu öğrenme arzusu bu konuya duyulan ilginin artışında birer etkidir. Sosyal medya uygulamaları üzerinde yer alan ve çok farklı dillerde yapılan yorumlara ilişkin veriler bu tür çalışmalar için yeterli veri kaynağıdır. Youtube platformu barındırmakta olduğu veri miktarı ve sağladığı veri desteği bakımından duygu analizi çalışmalarına uygundur.

Duygu analizi uygulamaları genellikle tek dilli olarak yapılmaktadır. Google, Yandex gibi dil çeviri araç desteği çok dilli duygu analizi çalışmalarının önünü açmıştır. Çeviri uygulamaları üzerinden gerçekleşen diller arası tercüme doğruluk oranı yapılmak istenen duygu analiz başarısına etkilemektedir. Diğer bir yandan çok dilli duygu analizinde dönüşümlü yapılacak ortak dilin seçimi önemlidir. Veri setine göre en yaygın olan bir dilin tercih edilmesi sonuçları bakımından çalışmanın başarısında önemli bir etkidir. Farklı dillerde yer alan kısaltma, deyim veya yerel kullanım farklılıkları gibi değişkenlerse çeviri araç başarısını düşürdüğü gibi yapılan çalışmanın başarı

düzeyini olumsuz etkilemektedir. Toplanan veriler üzerinden Naive Bayes algoritması kullanılarak duygu pozitif, negatif ve nötr olarak üç kategoride sınıflandırılmıştır. WEKA üzerinden yapılan analizde sınıflandırma başarı oranı %65,56 olarak elde edilmiştir. Kappa istatistik katsayısı, mutlak hata ortalaması ve kare hata ortalama verilerine göre veri setinin daha dengeli dağılımı için veri setin miktarı artırılabilir.

Yeterli veri setine sahip [14] basit verilerde [16] Naive Bayes algoritmasının kullanımı sonucunda (%96) başarı oranı elde edilmiştir. Twitter farklı veri kümeleri üzerinden LSTM derin öğrenme tekniği kullanarak %85-%97 arasında doğruluk başarı elde edilmiştir [18]. Türkçe Twitter gönderilerinden cinsiyet tespitine yönelik Türkçe Twitter çalışmada ön eğitimli 5.292 adet veri kullanılarak SVM, Bi-LSTM, CNN, Bert, Bert-128k, Distilbert, Electra Discriminator ve Electra Generator yöntemleri üzerinde yapılan çalışmada 5-10 arası tweet gönderenler çalışmaya dâhil edilmiştir. Çalışılan modellere göre (%64,52-%80,12) arasında doğruluk başarı oranları elde edilmiştir. Bu çalışmada Bert-128k modeliyle en yüksek başarı oranına ulaşıırken Bert modelinde bu oran (%73) olarak bulunmuştur. Benzer bir çalışmada [20] Türkçe ve bu dile uygun kelime sözlüğü kullanıldığında Bert yönteminde (%98) başarı elde edilmiştir. Ancak bu yüksek başarı oranları çoğunlukla tek dil(Türkçe) üzerinde yapılan çalışmalardan elde edilen sonuçlardır. Bu yüzden Türkçe gibi tek dil üzerine yapılan çalışmalardan bu başarı oranlarının yüksek çıkması normal sayılabilir. Başarı düzeyini etkileyen etmenlerin başında seçilen veri setinde yer alan verilerin o dilin doğru kullanımı ve gramer yapısı, veri temizlik ve ön işleme çalışmalarındaki yaklaşım, seçilen makine öğrenme yöntemi gibi unsurlar etkilidir. Çok dilli çalışmanın doğasından kaynaklı bu tür zorluklarına rağmen çalışmamızda farklı dillerden oluşan 15.082 adet veri seti üzerinden gerçekleştirilen %65,56'lık başarı oranı elde edildi. Naive Bayes algoritması kullanarak yapılan çalışmada [15] yaklaşık 8.000 adet YouTube yorum verilerinin kullanımı sonucunda WEKA doğruluk düzeyi (%48.42 - %61.18) arasında bulunmuştur. Bu çalışmada çoklu dil kullanılıp kullanılmadığı hakkında bir bilgi elde edilemedi. 200 adet Twitter veri üzerinden yapılan başka bir çalışmada [11] veri setinin %65-80'i İngilizce dilinde, geri kalanı Telugu, Hintçe, Bengalce, Arapça, Urduca ve tanınmayan farklı yerel dillerden oluştuğu belirtilmektedir. Bu çalışmada, destek vektör makineleri(SVM) yöntemiyle doğruluk oranının %95'e kadar ulaşabileceği belirtilmiştir. Ancak değerlendirilen veri seti ve kullanılan verilerin %80'e yakınının çok büyük oranda İngilizce dilinde olduğu dikkate alındığında çoklu bir dil analizi için yetersiz sayılabilir. Twitter verileri üzerinden kullanılarak yapılan diğer bir çalışmada [21], Naive Bayes yöntemi başarı düzeyi veri karakter sayısının 77'den küçük olması durumunda %91.43 gibi yüksek seviyede bulunmasına rağmen aynı çalışmada veri karakter sayısı 120'den küçük yorumlarda bu başarı oranı %57.14'e düşmüştür. Tüm bu sonuçlar dikkate alındığında örneklem sayısı ve çok dilli yapısıyla çalışmamızda elde edilen başarı düzeyi ve kullanılan veri hacmi yeterli seviyededir.

## 6. SONUÇ (CONCLUSION)

Her türlü duygu, görüş ve düşüncelerin yoğun olarak paylaşıldığı sosyal medya araçları üzerinden verilerin serbestçe elde edilmesi kıymetlidir. Bu tür çalışmalar birey ve toplumların duygu durumlarını anlamaya yardımcı olur. Bu çalışmaların başarısı duygu durumlarının doğru anlaşılabilmesinde önemli paya sahiptir. Bu alanda yapılacak benzer çalışmaların çeşitliliği ve zenginliği elde edilen doğruluk oranlarının daha yüksek tutarlılık seviyesine ulaşabilmesine yardımcı olur. Bu çalışma çeviri aracının desteklediği 108 farklı dilde paylaşılan metinler üzerinden duyguların analizine imkân vererek toplumların bir birlerini anlamasına, toplumsal olayların önlenmesine, daha anlaşılabilir ve paylaşılabilir bir dünyanın oluşturmaya katkı sunar niteliktedir. Çalışmada çok yüksek kullanıma sahip, içerik bakımından zengin yorumların paylaşıldığı ve duygu analizi yapmada oldukça elverişli olan YouTube sosyal medya platformu tercih edilmiştir. PHP uygulamasıyla YouTube API'si üzerinden toplanan 15.200 adet yorum verisi Google çeviri aracı kullanılarak ortak dil olarak kabul edilen İngilizce diline dönüştürüldü. Belirli ön kontroller gerçekleştirdi, duyguları anlamada hiçbir katkısı olmayan çeşitli semboller gibi anlamsız 118 adet veri temizlenerek çalışmadan çıkartıldı. Naive Bayes sınıflandırma algoritması kullanılarak 15.082 adet veri pozitif (5.167), negatif (4.644) ve nötr (5.271) duygu sınıfına ayrıldı. WEKA ile yapılan analizinde doğru sınıflandırma başarımları %65,56 bulundu. Bu başarımları kullanılan veri seti, yöntem ve araçların yetkinlikleriyle sınırlıdır. Benzer yöntemlerle yapılan çalışmalar, uzun ve farklı dillerde kullanılan verileri içermesi, bu verilerin çeviri aracı üzerinden İngilizce ortak dile çevrilerek analizin gerçekleştirildiği dikkate alındığında elde edilen sonuç başarılı olarak değerlendirilebilir. Farklı öznelik seçimi ve veri ön işleme teknikleriyle bu başarımları artırılabilir.

Bu çalışmada Youtube platformu, PHP Niiknow Kütüphanesi Naive Bayes algoritması, Google çeviri ve WEKA uygulama aracı kullanılmıştır. Çalışmanın başarısı elde edilen yorum verilerinin sayı ve içeriklerin doğruluğuna, kullanılan çeviri programı, yöntem ve uygulama araçların performansıyla sınırlıdır. Farklı dil ve kültürler için toplumların belli konular üzerindeki her türlü duygu ve düşüncelerini anlamada ve ona göre çeşitli öngörülerde bulunabilmede bu tür çalışmaların varlığı ve başarısı önemlidir. Bu yönüyle yapılan çalışma, çok dilli duygu analizinin desteklemesine katkı sağlamaktadır. Araştırmacılar benzer çalışmaları diğer sosyal medya uygulamaları üzerinden, Python, KNIME, TensorFlow benzeri makine öğrenme program araçlarıyla, Naive Bayes yerine destek vektör makineleri(SVM), en yakın komşu algoritmasını(KNN), derin öğrenme LSTM, CNN, BERT gibi diğer yöntemler üzerinden çalışabilirler. Duygu analizinde kullanılan verinin kalitesini artırılmaya yönelik yöntemleri geliştirilebilirler. Farklı dillerde toplanan veriler üzerinden yapılacak bu tür çok dilli bir analiz çalışmasında dil çeviri aracı olarak Yandex API kullanılabilir.

## KAYNAKLAR (REFERENCES)

- [1] O. Sahinaslan ve E. Sahinaslan, "Cross-object information security: A study on new generation encryption", *AIP Conference Proceedings*, 2086, 030034, 2019, doi:10.1063/1.5095119.
- [2] O. Sahinaslan, E. Sahinaslan ve E. Gunes, "Review of the contributions of contactless payment technologies in the COVID-19 pandemic process", *AIP Conference Proceedings*, 2334, 070002, 2021, doi:10.1063/5.0042225.
- [3] E. Sahinaslan, "On the internet of things: Security, threat and control", *AIP Conference Proceedings*, 2086, 030035, 2019.
- [4] N. Tuna, A. Sebatlı Sağlam ve F. Çavdur, "Covid-19 Salgını ile İlgili Paylaşımlar Üzerinde Veri Analizi", *Bilişim Teknolojileri Dergisi*, 15(1), 13-23, 2022, doi:10.17671/gazibt.d.928990.
- [5] İnternet: Most popular social networks worldwide as of July 2021, ranked by number of active users, <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>, 12.08.2021.
- [6] İ. Erkan, "Youtube Kullanma Amacının Youtube'da Yapılan Marka İşbirliklerine Yönelik Tutuma Etkisi", *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 128-145, 2020.
- [7] W. Medhat, A. Hassan ve H. Korashy, "Sentiment Analysis Algorithms And Applications: A Survey", *Ain Shams Engineering Journal*, 1093-1113, 2014.
- [8] İnternet: Makine Öğrenmesi Nedir, <https://www.ibm.com/tr-tr/cloud/learn/machine-learning>, IBM, 20.08.2021.
- [9] D. Jurafsky, J. H. Martin, "Naive Bayes and Sentiment Classification", *Speech and Language Processing*, 2021.
- [10] A. Severyn, A. Moschitti, O. Uryupina, B. Plank ve K. Filippova, "Multi-lingual Opinion Mining on YouTube", *Information Processing and Management*, 46-60, 2016.
- [11] K. Arun ve A. Srinagesh, "Multi-Lingual Twitter Sentiment Analysis Using Machine Learning", *International Journal of Electrical and Computer Engineering*, 5992-6000, 2020.
- [12] M. M. Agüero-Torales, J. I. Abreu Salas ve A. G. López-Herrera, "Deep Learning And Multilingual Sentiment Analysis On Social Media", *Applied Soft Computing*, 2021.
- [13] A. Hogenboom, B. Heerschop, F. Frasincaar, U. Kaymak ve F. De Jong, "Multi-Lingual Support For Lexicon-Based Sentiment Analysis Guided", *Decision Support Systems*, 43-53, 2014.
- [14] T. Ma Ma, K. Yamamori, A. Thida, "A Comparative Approach to Naive Bayes Classifier and Support Vector Machine for Email Spam Classification", *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 324-326, 2020.
- [15] K. Asad Ullah Rafiq, K. Madiha ve M. B. Khan, "Naive Multi-label Classification Of Youtube Comments Using", *Procedia Computer Science*, 2016.
- [16] C. R. Stephens, H. F. Huerta ve A. R. Linares, "Why The Naive Bayes Approximation is Not As Naive", *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2015.
- [17] I. Blekanov, M. Kukarkin, A. Maksimov, S. Bodrunova, "Sentiment Analysis for Ad Hoc Discussions Using Multilingual Knowledge-Based Approach", *Proceedings of the 3rd*



- International Conference on Applications in Information Technology**, 2018.
- [18] M. C. Yılmaz ve Z. Orman, "LSTM Derin Öğrenme Yaklaşımı ile Covid-19 Pandemi Sürecinde Twitter Verilerinden Duygu Analizi", *Acta Infologica*, <https://doi.org/10.26650/acin.947747>.
- [19] İ. Sel ve D. Hanbay, "Ön Eğitimli Dil Modelleri Kullanarak Türkçe Tweetlerden Cinsiyet Tespiti", *Fırat Üniversitesi Müh. Bil. Dergisi*, 33, 2, 675-684, 20211.
- [20] E. Karamehmetoğlu, S. Ersöz ve O. Karamehmetoğlu, "Sosyal Ağ Tabanlı Verilerden Faydalanarak Korona Virüs Konulu Duygu Analiz Çalışması", *Ergonomi*, 4(1), 47 - 54, 2021.
- [21] J. Samuel, G. G. M. Nawaz Ali, M. Rahman, E. Esawi ve Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification", *Information*, cilt 11(6), 314, 2020.
- [22] İnternet: WEKA Data Mining <https://www.cs.waikato.ac.nz/~ml/WEKA>, 08.09.2021.
- [23] İnternet: YouTube Data API Overview, <https://developers.google.com/youtube/v3/getting-started>, 24.06.2021.
- [24] İnternet, MySQL Database, <https://www.mysql.com/>, 24.09.2021.
- [25] İnternet: Microsoft Veri Platformu, <https://www.microsoft.com/tr-tr/sql-server/>, MsSQL, 24.09.2021
- [26] İnternet: Translation APIs & Reference, <https://cloud.google.com/translate/docs/apis>, 18.06.2021.
- [34] G. Harman , "Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini", *Avrupa Bilim ve Teknoloji Dergisi*, 32, 7-13, 2021, doi:10.31590/ejosat.1041186.
- [27] İnternet: Niiknow-Naive Bayes, <https://github.com/niiknow/bayes/blob/e754145d52900fdf4c1e620cc561132954721cf9/src/Bayes.php>, 19.06.2021.
- [28] S. Alqaraleh , "Efficient Turkish Text Classification Approach for Crisis Management Systems", *Gazi University Journal of Science*, 34(3), 718-731, 2021, doi:10.35378/gujs.715296..
- [29] Ö. Tonkal, H. Polat, "Traffic Classification and Comparative Analysis with Machine Learning Algorithms in Software Defined Networks", *Gazi University Journal of Science Part C: Design and Technology*, 9 (1) , 71-83 . doi: 10.29109/gujsc.869418.
- [30] S. A. Arpacı ve O. Kalıpsız, "A Comparison of Different Naive Bayes Techniques for Software Defect Classaification", *Omer Halisdemir University Journal of Engineering Sciences*, c.7(1), 1-13, 2018. doi: 10.28948/ngumuh.383709.
- [31] H. Deng, Y. Sun, Y. Chang ve J. Han, 2014, Probabilistic models for classification. in CC Aggarwal (ed.), *Data Classification: Algorithms and Applications*. CRC Press, 65-86.2014, doi:10.1201/b17320.
- [32] Frank, E., Bouckaert, R. R., "Naive bayes for text classification with unbalanced classes", **In European Conference on Principles of Data Mining and Knowledge Discovery**, Springer, Berlin, Heidelberg, 503-510, (2006).
- [33] E. Çelik, D. Dal ve T. Aydın , "Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması", *Avrupa Bilim ve Teknoloji Dergisi*, 27, 880-889, 2021, doi:10.31590/ejosat.905259.