



# Futbol Müsabakaları ile İlgili Tweetlerin Anlık Duygu Analizi

Rıza Korkusuz<sup>1\*</sup>, Aydın Carus<sup>2</sup>

<sup>1</sup> Trakya Üniversitesi, Mühendislik Fakültesi, Fen Bilimleri Enstitüsü, Edirne, Türkiye (ORCID: 0000-0003-2742-886X)

<sup>2</sup> Trakya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Edirne, Türkiye (ORCID: 0000-0003-3370-5974)

(International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2020 – 22-24 Ekim 2020)

(DOI: 10.31590/ejosat.821200)

**ATIF/REFERENCE:** Korkusuz, R. & Carus, A. (2020). Futbol Müsabakaları ile İlgili Tweetlerin Anlık Duygu Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (Özel Sayı), 386-396.

## Özet

Sosyal medya, insanların kendilerini ifade edebildikleri ortamlar olarak çok kullanılmaktadır. Bu sebeple Facebook, Instagram ve Twitter gibi sosyal medya ortamlarının kullanıcı sayıları giderek artmaktadır. Sosyal medya kullanıcılarının paylaşımları analiz edilerek ilgili konu hakkındaki duyguları ortaya çıkarılabilmektedir. Sosyal medya platformu olan Twitter da politikacılar, spor kulüpleri, şirketler, aktivistler kısaca neredeyse tüm bireyler ve kurumlar için kendilerini ifade etme ortamı haline gelmiştir. Hemen hemen tüm futbol takımlarının taraftarları için haberlerini paylaştıkları Twitter hesapları vardır. Kullanıcılar Twitter'da futbol müsabakası esnasında ve sonrasında da mesaj paylaşabilmektedir. Bu çalışmanın konusu, Twitter ortamında futbol takımları ve futbol müsabakaları hakkında Twitter kullanıcılarının paylaşımlarının duygu analizi ile ilgilidir. Bu çalışmada Twitter'da futbol müsabakaları hakkında paylaşılan 30.000 Türkçe tweet ile anlık duygu analizi yapılmıştır. Eğitim setlerindeki sınıflandırma hatalarını en aza indirmek için toplanmış olan tweetler el ile etiketlendikten sonra yine aynı kişi tarafından farklı zamanlarda beşer defa kontrol edilmiştir. Bu etiketlemede olumlu, olumsuz, tarafsız ve alakasız olarak 4 duygu sınıfı kullanılmıştır. Bu etiketlenmiş tweetlerden farklı özniteliklere sahip 12 farklı eğitim seti oluşturulmuştur. Oluşturulan bu eğitim setleri kullanılarak farklı sınıflama algoritmaları ile modeller çıkarılmış ve bu modellerin çapraz doğrulama ile sınıflama başarımları bulunmuştur. Farklı özniteliklere sahip olarak oluşturduğumuz eğitim setleri, belirlenen sınıflandırma algoritmaları ile test edilmiş ve algoritmalarının sınıflandırma doğrulukları; Naive Bayes algoritması için %84.30, K-En Yakın Komşu (KNN- K-Nearest Neighborhood) algoritması için %87.73, C4.5 algoritması için %89.60, Destek Vektör Makinesi (SVM- Support Vector Machine) algoritması için %92.30 olarak bulunmuştur. Çalışmada zemberek kütüphanesi kullanılarak Türkçe tweetlerde yaklaşık 48.000 kelimenin düzeltilmesi, sınıflandırma başarımına olumlu katkı sağlamıştır. Ayrıca futbol müsabakaları hakkında paylaşılan tweetleri anlık olarak toplayıp, oluşturulan modeller ile bu tweetlerin sınıfını tespit ederek sınıflara ait sonuçları gerçek zamanlı görselleştiren bir uygulama geliştirilmiştir.

**Anahtar Kelimeler:** Anlık Duygu Analizi, Makine Öğrenmesi, Twitter, Veri Madenciliği.

## Real Time Sentiment Analysis of Tweets about Football Matches

### Abstract

Social media is frequently used as a platform where people can express themselves. For this reason, the number of users of social media platforms such as Facebook, Instagram and Twitter has been gradually increasing. By analyzing the shares of social media users, their feelings about the relevant subject can be revealed. Twitter, as a social media platform, has become a medium for politicians, sports clubs, companies, activists, briefly for almost all individuals and institutions, to express themselves. Almost all football teams have Twitter accounts where they share their news with their fans. Users can share messages on Twitter during and after the football matches. This study is about the sentiment analysis of Twitter users' posts about football teams and football matches on Twitter. Within the scope of this study, a real-time sentiment analysis was made by using 30,000 Turkish tweets shared about football matches on Twitter. In order to minimize the classification errors in the training sets, the collected tweets were manually tagged and then checked five times by the same annotator at different times. In the labeling process, 4 sentiment classes were used as positive, negative, neutral, and irrelevant. 12 different training sets with different attributes were created from these tagged tweets. Using these training sets, models were created with different classification algorithms and classification performance of these models

\* Sorumlu Yazar: Trakya Üniversitesi, Mühendislik Fakültesi, Fen Bilimleri Enstitüsü, Edirne, Türkiye, ORCID: 0000-0003-2742-886X,  
[rizakorkusuz@gmail.com](mailto:rizakorkusuz@gmail.com)

with cross-validation was evaluated. The training sets that we have created with different attributes have been tested with the specified classification algorithms and the classification accuracies of the algorithms are 84.30% for Naive Bayes, 87.73% for K-Nearest Neighborhood (KNN- K-Nearest Neighborhood), 89.60% for C4.5, and 92.30% Support Vector Machine (SVM- Support Vector Machine). In this study, the correction of approximately 48,000 words in Turkish tweets by using the Zemberek library contributed positively to the classification performance. In addition, an application was developed that collects the tweets shared about football matches in real-time and it determines the class of these tweets with the models created and visualizes the class results.

**Keywords:** Real Time Sentiment Analysis, Machine Learning, Twitter, Data Mining.

## 1. Giriş

Duygu, insanların kişiliğini ve davranışsal özelliklerini yansıtan karmaşık ve çok boyutlu bir özelliktir. İnsanlar farklı konular, olaylar, kişiler hakkında günlük yaşamlarında duygularını ifade ederler. Diğer insanlara duygularını konuşarak ve yüz ifadeleriyle gösterirler. Ancak, teknolojinin ve sosyal medyanın yaygınlaşması ile insanlar duygularını, sosyal medya gönderileri kullanarak ifade etmektedir. Sosyal medya platformları, ürünler hakkında yorum, herhangi bir konu hakkında kullanıcı görüşleri gibi bilgiye dönüştürülebilecek birçok ham veriye sahiptir. Kullanıcılar, çok sık olarak sosyal medya ortamlarında, metin, resim, ses gibi farklı biçimlerde paylaşım yapmaktadırlar. Bir dakikada sosyal medya ortamlarından Twitter'da 481.000 tweet paylaşmakta, YouTube'da 4,5 milyon video izlenmekte, Facebook'a 1 milyon kullanıcı giriş yapmakta ve Instagram'da 347.222 fotoğrafa bakılmaktadır [1]. Bu ortamlarda kullanıcılar birbirleriyle etkileşime girmektedir. Kullanıcıların birbirleriyle en çok etkileşime girdiği sosyal medya ortamı olan Twitter'da ortalama olarak her saniye yaklaşık 6.000 tweet atılıyor, bu da dakikada 350.000'den fazla tweet'e, günde 500 milyon tweet'e ve yılda yaklaşık 200 milyar tweet'e karşılık geliyor [2]. Twitter 330 milyon aylık aktif kullanıcıya, 152 milyon günlük aktif kullanıcıya sahiptir [3]. Twitter'da kullanıcılar düşüncelerini ve duygularını daha az sözcükle paylaşmaktadır. Kullanıcılar tek bir tweette en fazla 280 karakter ile düşüncelerini ve duygularını ifade edebilmektedir. Bu da tweetlerin okunmasını ve yayılmasını kolaylaştırmaktadır.

Sosyal medya ortamlarından Twitter, kullanıcıların takipçileri ile paylaşım yapma ve iletişim kurma konusunda en önemli ortamlardan biri haline gelmiştir. Twitter, kullanıcılarına sportif, kültürel, ekonomik, politik ve diğer birçok alandaki olaya ilişkin görüşlerini, fikirlerini, yargılarını, kaygılarını ve tutumlarını ifade etme fırsatı sunmaktadır [4]. İnsanlar filmler, ürünler, markalar ve diğer çeşitli konularda tweetler atmaktadır. Bu tweetlerdeki en önemli ve değerli veriler kullanıcıların bir konu hakkındaki duyarlılığıdır. İnsanlar bir tweet paylaştıklarında konuştukları konuya, memnuniyetini ya da memnuniyetsizliğini, olumlu duygusunu ya da olumsuz duygusunu katmaktadır. Twitter verilerinin kullanıldığı bazı alanlar “Duygu Analizi”, “Dilbilimsel Analiz”, “Verilerin Karşılaştırılması”, “Etkileyenlerin Belirlenmesi”, “Afet Yönetiminde Öngörü Yapma”, “Bilgi Sistemi”, “İnsan Üzerine Etki”, “İlgili Konunun Belirlenmesi”, “Bilgisayar İletişimi”, “Siber Suç Tespiti”, “Kamuoyu Sağlık Hizmeti”, “Hastalık Yönetimi”, “Gelecekteki Etkinlik Tahmini” ve “Tıbbi Şikayetler” gibi alanlardır [5]. Twitter'da, Hastalık salgınları, borsadaki hareketlilik ve seçim sonuçları tahminleri araştırılmıştır [6]. Otomobillerle ilgili tüketici şikayetlerini tespit etmek için twitter verisi kullanılmıştır [7]. Londra 2012 Olimpiyat Oyunları sırasında turizm ürün ve hizmetlerinin algılanan kalitesini değerlendirmek için Twitter takip edilmiştir [8]. Paylaşılan tweetlerden, filmlerin gişe gelirlerini tahmin etmek için yapılmış çalışma mevcuttur [9]. Yine benzer şekilde H1N1 gibi salgın hastalıkların yayılımını takip etmek için yapılmış çalışmalar da vardır [10]. Benzer şekilde paylaşılan tweetler üzerinden meydana gelen bir depremin etkilerini izlemek için bir çalışma yapılmıştır [11]. Herkese açık olan gerçek zamanlı paylaşımlar ile kullanıcıların duygu analizi yapılabilmektedir. Duygu analizi yazılı metinde ifade edilmiş duyguların, düşüncelerin tespiti ve analizi ile de ilgilenir. Duygu analizleri sayesinde kişiler, olaylar, ürünler, markalar, müsabakalar vb. gibi daha birçok konuda paylaşım yapanların görüşleri hakkında bilgi sahibi olunabilir. Bazı firmalar ilgili oldukları sektör ile ilgili Twitter paylaşımlarını takip ederek piyasa analizi yapar ve firmalarına buradan ele ettikleri bilgilere göre yön verirler [12]. Firmalar, tüketicilerinin, ürünleri hakkındaki görüşlerini bilmek istemekte ve ürünleri hakkında görüşleri öğrenmek için duygu analizleri yapmaktadır [13]. Birçok firma müşterilerine ürünlerini tanıtmak ve kampanyalarını duyurmak için Twitter ortamını kullanmaktadır [14]. Bazı duygu analizi çalışmalarında, hisse senedi fiyatları, sinema gişesi gelirleri ve siyasi seçimler gibi sosyal olayların, sosyal medyadan etkilendiği iddia edilmektedir [15]. Twitter üzerinden yapılan anketler ile ifade edilen görüşlere dayanarak kamuoyunun fikrini dolaylı olarak anlamak için yapılmış çalışmalar mevcuttur [16]. Tweet paylaşımları üzerinden duygu analizi için tek tek kelimeler veya kısa metinlere dayalı etiketleme yapılır. Bu duygu etiketlemesi dil, etki alanı ve çoğu zaman konuya özeldir [17]. Duygu analizi, insanların fikirlerini, görüşlerini, değerlendirmelerini, tutumlarını ve duygularını yazılı dilden analiz edilen çalışma alanıdır [18]. Amacı, konulara yönelik olarak görüş, duygu veya tutumları ortaya çıkarmaktır [19].

Duygu analizinin farklı birçok alan yanında birçok toplumun hayatında önemli bir yere sahip olan futbol hakkında da çalışmalar mevcuttur. Ancak anlık olarak Türkçe futbol tweetleri üzerine duygu analizi çalışmasına rastlanmamıştır. Futbol müsabakalarında da ile ilgili tweetlerde heyecan, hayal kırıklığı, üzüntü, neşe, kızgınlık, kardeşlik vb. duygular ifade edilir. Twitter'da Dünya Kupası 2014 ile ilgili [20], #FIFA, #Fotball, #Worldcup ve #Soccer gibi etiketleri içeren tweetler toplanmış ve öfke, sevinç, hüzn, nefret ve sürpriz gibi 8 duygu durumu ölçülmüştür. Futbol müsabakaları ile ilgili yorumlar, serbest vuruşlar, penaltılar, goller, fauller ve hakem kararları gibi birçok faktör oyun sırasında taraftarların duygusal tepkilerini değiştirebilir. Bu çalışma ile Twitter'da futbol müsabakaları hakkında paylaşılan Türkçe tweetler üzerinden, taraftarların duyguları anlık olarak tespit edilmeye çalışılmıştır. Ayrıca anlık olarak tahmin edilen duyguların görselleştirilmesi için bir uygulama geliştirilmiştir.

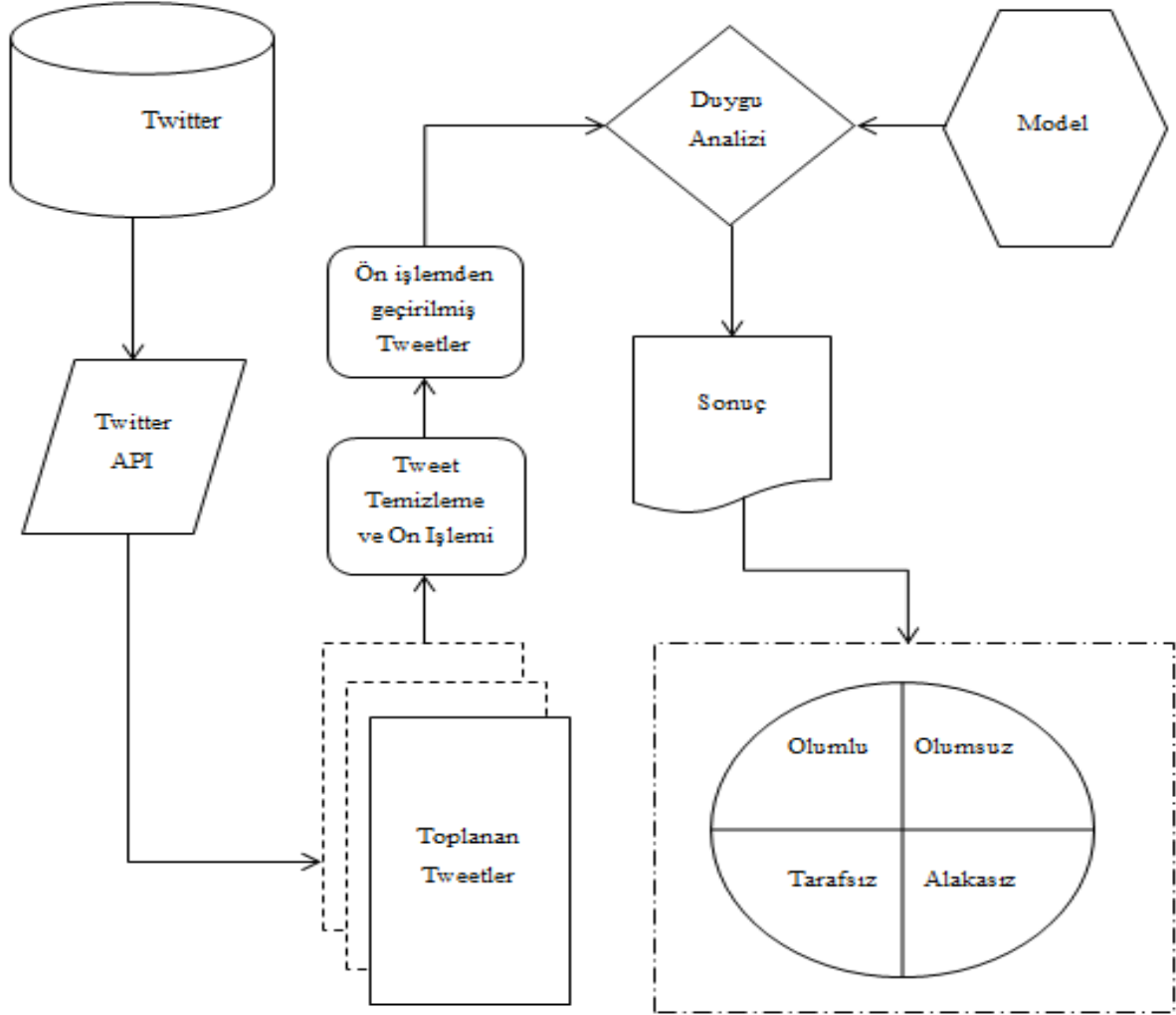
Bu makalenin giriş bölümünde, sosyal medya platformlarından olan twitter paylaşımlarından makine öğrenmesi algoritmaları kullanılarak yapılan duygu analizi çalışmasının amacı anlatılmıştır. İkinci bölümde, Twitter platformundan anlık veri toplayan ve bu verileri kaydetmek, temizlemek, etiketlemek gibi çeşitli ön işlem ve sonrası işlemleri yerine getirmek için geliştirilen uygulama anlatılmıştır. Üçüncü bölümde, toplanan tweetlerin etiketlendikten sonra öznitelik çıkarımı, oluşturulan veri setlerinin yapıları ve birbirlerinden farklarından bahsedilmiştir. Dördüncü bölümde, makine öğrenmesi için geliştirilmiş olan Weka yazılım aracı

kullanılarak oluşturulan eğitim setlerinin makine öğrenmesi algoritmalarına bağlı sınıflama başarımları analiz edilmiş ve elde edilen sonuçlar birbiriyle kıyaslanmıştır. Çalışmanın son bölümünde sonuçlar değerlendirilmiştir.

## 2. Materyal ve Metot

Bu çalışmada Twitter’da paylaşılan tweetler anlık toplanarak tweetlerin gerçek zamanlı duygu analizi için model bir uygulama geliştirilmiştir. Farklı özneliklere sahip eğitim setlerine ait modellerin oluşturulması için; tweetlerin toplanması ve ön işlemden geçirilmesi, tweetlerin el ile etiketlenmesi, belirlenmiş algoritmalar kullanılarak sınıflayıcı modellerin oluşturulup, modellerin çapraz doğrulama ile testlerinin yapılması işlemleri yapılmıştır. Tweetinvi<sup>2</sup> kütüphanesinin yardımıyla Twitter Streaming API kullanılarak C#.NET programlama dili ile geliştirilen uygulama üzerinden; Twitter’den belirlenen “Fenerbahçe, Galatasaray, Beşiktaş, Trabzonspor, maç, gol, galibiyet, mağlubiyet, futbol, hakem, penaltı” vb. gibi spora ait belirlenmiş anahtar kelimelerine göre tweetler filtrelenerek toplanmıştır. Toplanan tweetler 2.2. veri ön işleme bölümünde anlatıldığı gibi ön işlemden geçirilmiştir. Ön işlemden sonra, etiketleme için geliştirilmiş olan uygulama yardımıyla bu tweetler olumlu, olumsuz, tarafsız ve alakasız olarak el ile etiketlenmiştir. Etiketlenen tweetler üzerinden öznelik belirleme işlemleri yapılmıştır ve 12 farklı eğitim seti oluşturulmuştur. Bu eğitim setleri Weka aracı kullanılarak makine öğrenmesi algoritmaları ile test edilmiştir ve modeller oluşturulmuştur.

Oluşturulan bu modelleri kullanarak anlık olarak toplanan tweetleri sınıflandıran uygulama geliştirilmiş ve bu uygulama ile tweetler olumlu, olumsuz, tarafsız ve alakasız olarak sınıflandırılarak tüm sınıflara ait tweet sayıları anlık değişen grafik olarak gösterilmiştir. Anlık duygu analizi için sistem mimarisi Şekil 1’de yer almaktadır.



Şekil 1. Anlık duygu analizi sistem mimarisi

<sup>2</sup> <https://www.nuget.org/packages/TweetinviAPI/>

## 2.1. Veri Seti

Eğitim veri setlerini oluşturmak için, futbol sporuna ait belirlenen kelimeler ile filtreleme yaparak tweetler toplanmış ve bu tweetlerden farklı özneliklere sahip eğitim veri setleri oluşturulmuştur. Eğitim veri setlerini oluşturmak için farklı zamanlarda farklı futbol müsabakaları oynanırken atılan tweetleri depolamak için yerel sunucu bilgisayar kullanılmıştır. Bu amaçla toplanan 30.000 tweetten el ile etiketleme yapılarak farklı özneliklere sahip eğitim setleri oluşturulmuştur. Bir sınıf lehine yanlılığın engellenmesi için her bir sınıftan eşit sayıda (olumlu, olumsuz, tarafsız, alakasız) etiketli 7.500'er tweet ile dengeli eğitim veri setleri oluşturulmuştur. Anlık olarak paylaşılan tweetlere erişmek için geliştirilen model uygulama arayüzü Şekil 2'de gösterilmektedir.

Eğitim veri setlerini oluşturmak için filtrelenmiş olarak tweetler toplanırken futbol müsabakası ile ilgili olmayan reklam içeren veya farklı konuları içinde barındıran birçok tweet de toplanmıştır. Anlık olarak sınıflandırma yapılırken de bu alakasız tweetlerin olabileceği düşüncesiyle bu tür tweetler de alakasız olarak etiketlenmiş ve değerlendirilmeye alınmıştır.

- X

### Anlık Duygu Analizi İşlem Ekranı

Twitter Arama Veri Etiketleme

Anlık Tweet Toplama

Filtre Kelimeler

Toplanan Tweet

Twitter API

Kredi Limiti  Kalan Arama Kredisini  Arama Kredisini Tekrar Yenilenecek

Temizlenmiş Tweetler

kadroda yok, Fenerbahçe maçının iyilerindendi.  
Fenerbahçe Beko'nun konuğu Beşiktaş Sompö Sigorta  
Yalnız bütün bu olan bitenlere rağmen an itibarıyla ligde en çok galibiyeti olan tkm yine de Fenerbahçe  
Büyük Fenerbahçe Taraftarı! OmuzOmuzŞampiyonluğa  
Eee haksiz deęiller yani  
İyi geceler Büyük Fenerbahçe Taraftarı!  
Çünkü Fenerbahçe ve Galatasaraya gol atıyor  
Fenerbahçe sol bek, kanat ve maddi imkana göre forvet alacak. Zajc ve Jailson için gelecek teklifler deęerlendirilecek.  
Bunlara bundan sonra böyle her yerde böyle olmalı ZTK maçlarında stada gelecekleler o zaman iyi bişeyler yapmak lazım  
Bakın ben su içiyorum, Fenerbahçe de böyle bir bardak su içsin  
Günün teklisi Fenerbahçe u19 Kasımpaşa u19 2.5 Üst 1.60 Üyelik için ;  
Şimdi Fenerbahçe de oynamak istiyorum diyor. İnsan gelme de diyemiyor  
Fenerbahçe Beko'nun konuğu Beşiktaş Sompö Sigorta  
Fenerbahçe penaltı kazanıyor Sizlerden bu paylaşımı beni seneler önce engelleyen Fenerbahçe maşasına iletmenizdir. Neden stadyumda tiyatro olmaz maç oynandı Fenerbahçe kötü oynadı kazandı tarafsız bakarsanız yüzde 100 penaltılar doğru hakem iyi maç yönetti geçen hafta iyi oynadı 2 penaltısı verilmedi tarafı olan hiç ses çıkarmadı kaleci çok iyiydi lütfen artık başkanlar hocalar seyirci Hafta içi Cuma akşamı Kadıköy'de 38 bin Ataşehir de 9 bin taraftar işinize gelmeye bilir AMA SİZ Not edin BUNU Bu Fenerbahçe bu Fenerbahçe alsin bunu vutrlisında kousularda derece nettir.

Şekil 2. Tweet toplamak için geliştirilen uygulama arayüzü

## 2.2. Veri Ön İşleme

Parlar yaptığı çalışmada [21] Türkçe yorumların analizinde, belirli noktalama işaretlerinin ve etkisiz kelimelerin olumlu yönde katkıda bulunduğu sonucuna varmıştır. Noktalama işaretlerinin duygu ifade ederken katkı sağladığı düşüncesiyle eğitim seti için el ile etiketleme işlemi yapıldıktan sonra noktalama işaretleri ve etkisiz kelimeler kaldırılmıştır. Çalışmada retweet yapan kullanıcıların tweeti paylaşan kullanıcı ile aynı fikirde fakat eğitim setindeki diğer öznelikleri farklı olduğu düşünülerek retweetler göz ardı edilmemiştir. Tweetlerde, gereksiz veriler temizlenip, temizlenmiş tweetlerden eğitim setleri oluşturulmuştur. Eğitim seti oluşturmak için yapılan veri ön işlemleri, etiketleme işleminden önce ve sonra olarak 2 aşamada yapılmıştır. Tablo 1'de eğitim veri setlerinin oluşturulması için el ile etiketleme işleminden önce ve etiketleme işleminden sonra yapılmış veri ön işleme işlemleri gösterilmiştir.

Tablo 1. Etiketleme işleminden önce ve sonra yapılan veri ön işleme işlemleri

El İle Etiketleme İşleminin Önce		El İle Etiketleme İşleminin Sonra	
İçerik	Ön İşleme	İçerik	Ön İşleme
#kelime (#Fenerbahçe, #TSvFB gibi)	Temizlendi.	Noktalama işaretleri (! ? , . " ' ; )	Temizlendi.
@kullanıcıadı kullanıcıların tweetlerine verilen yanıtlar ('@' ile başlayanlar)	Temizlendi	Büyük harfler	Tüm içerik küçük harfe çevrildi.
URL'ler, küçük URL'ler ve Link'ler	Temizlendi.	Sayılar	Temizlendi.
Alfabede olmayan karakterler	Temizlendi.	Kelime	Kelimelerin kökleri çıkartıldı.
Emojiler	Temizlendi.	Tekrar eden kelimeler	Tekillendi.
Birden fazla boşluklar	Temizlendi.	Etkisiz Kelimeler	Temizlendi.

Bu şekilde 2 aşama olarak yapılan veri ön işleme işlemlerinden önce tweetin ham hali ve tweetin temizlenmiş haline örnek Tablo 2'de gösterilmektedir.

Tablo 2. Etiketleme işlemi öncesi ham tweet ve temizlenmiş tweet örneği

Ham Tweet	Temizlenmiş Tweet
RT @futbolarena: Trabzonspor, Fenerbahçe kaleisine gönderdiği 26. şutta 2. golu buldu. Fenerbahçe çok kötü oynuyor... #TSLive #TSvFB 😊 😊 😊 https://t.co/t7fOLB1LYd	Trabzonspor, Fenerbahçe kaleisine gönderdiği 26. şutta 2. golu buldu. Fenerbahçe çok kötü oynuyor...

### 2.3. Türkçe Doğal Dil İşleme Kütüphanesi

Zemberek [22] Kütüphanesi, Türkçe için açık kaynaklı bir doğal dil işleme kütüphanesidir. Kaplan yaptığı çalışmasında [23], tweetlerdeki tüm yazım hatalarını gidererek doğru bir şekilde sınıflandırmak için Zemberek kütüphanesini kullanmıştır. Bu çalışmada da bazı eğitim setlerinin oluşturulmasında Zemberek Kütüphanesi kullanılarak yazım denetimi, kelime köklerini bulma, yanlış yazılmış kelimeleri düzeltme ve kelime önerme gibi işlemler yapılmıştır. Zemberek kütüphanesi kullanılarak yapılan yazım denetimine örnek Tablo 3'de görüldüğü gibi hem imlaya göre kelime düzeltmesi yapılmış hem de anlama göre kelime düzeltmesi yapılmıştır.

Tablo 3. Zemberek kütüphanesi kullanılarak yazım denetimi

Ham Tweet	Zemberek İmlaya Göre Kelime Düzeltmesi Yapılmış ve Temizlenmiş Tweet
RT @futbolarena: Trabzonspor, Fenerbahçe kaleisine gönderdiği 26. şutta 2. golu buldu. Fenerbahçe..... çok kötü oynuyor... #TSLive #TSvFB 😊 😊 😊 https://t.co/t7fOLB1LYd	trabzonspor fenerbahçe kalesine gönderdiği şutta golu buldu fenerbahçe çok kötü oynuyor
Ham Tweet	Zemberek Anlama Göre Kelime Önerisi Yapılmış ve Temizlenmiş Tweet
RT @futbolarena: Trabzonspor, Fenerbahçe kaleisine gönderdiği 26. şutta 2. golu buldu. Fenerbahçe..... çok kötü oynuyor... #TSLive #TSvFB 😊 😊 😊 https://t.co/t7fOLB1LYd	trabzonspor fenerbahçe kalesine gönderdiği şutta golu buldu fenerbahçe çok kötü oynuyor
Ham Tweet	Zemberek Yazım Denetimi Yapılmış Tweet
RT @futbolarena: Trabzonspor, Fenerbahçe kaleisine gönderdiği 26. şutta 2. golu buldu. Fenerbahçe çok kötü oynuyor... #TSLive #TSvFB 😊 😊 😊 https://t.co/t7fOLB1LYd	trabzonspor fenerbahçe kalesine gönderdiği şutta golu buldu fenerbahçe çok kötü oynuyor

## 2.4. Eğitim Veri Setlerini El ile Etiketleme

Eğitim setleri için el ile etiketleme aşamasında Şekil 3'te verilen uygulama arayüzü üzerinden tweetler 4 sınıf olarak etiketlenmiştir. Eğer tweet içinde birden fazla duygu ifade ediliyorsa, tweette ifade edilen güçlü olan duygu etiketlenmiştir. Yapılan etiketleme işlemi sırasında farklı özniteliklere sahip eğitim dosyaları arff dosya formatında oluşturulmuştur.

Anlık Duygu Analizi İşlem Ekranı

Twitter Arama Veri Etiketleme

Veri Etiketleme Ekranı

Arff Dosya Formatı

Arff Dosyalarını Oluştur

Etiketli Verileri Aktar

Ve tam bir bomba mac!!! kırmızı kartlık ve penaltı pozisyonu veriyo izleyenlere, güzel gol attı güzel gol hediye etti...

Android,Uzunluk3,False,Hashtags0,False,Perşembe,Dilim12,22,58,2017,2,4,3,5,13,False,False,False,'KonumYok'

Geri İleri Başla Olumlu Olumsuz Tarafsız Alakasız

Veri Ön İşlemleri

Zemberek Kelime Düzeltici ve Önerici

Türkçe Kelime Denetimi 1

Temizlenmiş Durak Kelime Sayısı 5

Temizlenmiş Tekrar Eden Kelime Sayısı 1

Tweetdeki Kelimelerin Kökleri 11

Şekil 3. Veri etiketleme arayüzü

## 2.5. Oluşturulan Veri Setleri

Eğitim veri setlerinde yer alan veriler veri ön işleme aşamasından geçerek oluşturulmuştur. Tablo 4'te el ile etiketleme işleminden önce ham tweet ve tweetin veri ön işleme yapılmış şekli verilmiştir.

Tablo 4. Ham tweet ve etiketleme işleminden önce veri ön işleme yapılan tweet

<b>Ham Tweet</b>	Maç özetleri ve Maç yayınları ligin 20.haftasını Tüm Maçları birazdan izlyin <a href="https://bc.vc/THuohfi">bc.vc/THuohfi</a> <a href="#">#maçyayını</a> <a href="#">#yayın</a> <a href="#">#link</a> <a href="#">#Galatasaray</a> <a href="#">#Beşiktaş</a>
<b>Etiketlemeden önce işlenmiş Tweet</b>	Maç özetleri ve Maç yayınları ligin 20.haftasını Tüm Maçları birazdan izlyin

Eğitim setleri; özniteliklerine ve tweet metninde yapılan işlemlere göre 12 farklı eğitim seti oluşturulmuştur. Eğitim setlerinin temel öznitelikleri olarak paylaşım yapılan cihazın işletim sistemi, tweet uzunluğu, tweetin uzunluğunu kategorize ettik, tweetin devam edip etmeme durumu, tweetdeki hashtag sayısı, tweetin hassas içerik içermeme durumu, tweetin atıldığı gün, tweet haftanın hangi günü atıldı, tweetin atıldığı zaman aralığı, haftanın hangi günü atıldı, tweet atılma saati, tweet atılma dakikası, tweet atanın hesap açılış yılı, tweet atanın hesap açılış ayı, tweet atanın toplam attığı tweet sayısı, tweet atanın takipçilerinin sayısı, tweet atanın takip ettiklerinin sayısı, tweet atanın beğendiği tweet sayısı, tweet atan hesabın onaylı rozetli olup olmama durumu, tweet atılan hesabın korumalı olup olmama durumu, tweet atılan hesabın konumunun açık kapalı olma durumu, tweetin atıldığı konum bilgisi olarak belirlenmiştir. Tablo 5’de eğitim setlerinin nasıl oluşturulduğu ile ilgili açıklamalara yer verilmiştir.

Tablo 5. Eğitim setleri içerik ve öznitelikleri

EğitimSeti-1	Sadece belirlenen temel öznitelikleri içermektedir.
EğitimSeti-2	Türkçe kelime denetimi yapılmıştır ve belirlenen özniteliklere ayrıca Türkçe kelime denetimi sonunda düzeltilen kelime sayıları da eklenmiştir.
EğitimSeti-3	Türkçe kelime denetimi yapılmış ve belirlenen öznitelikler eklenmiştir.
EğitimSeti-4	Etkisiz kelimeler temizlenmiş ve belirlenen özniteliklere temizlenen etkisiz kelime sayısı eklenmiştir.
EğitimSeti-5	Etkisiz kelimeler temizlenmiş ve belirlenen öznitelikler eklenmiştir.
EğitimSeti-6	Kelimelerinin kelime kökleri elde edilip belirlenen özniteliklere, kelimelerinin kökleri ayrıca kaç tane kelime kökü olduğunun sayısı eklenmiştir.
EğitimSeti-7	Kelimelerin kelime kökleri elde edilmiş ve belirlenen özniteliklere kelime kökleri eklenmiştir.
EğitimSeti-8	Tekrar eden kelimeler varsa tekrar eden kelimeler çıkartılmış ve belirlenen öznitelikler eklenmiştir.
EğitimSeti-9	Tekrar eden kelimeler varsa tekrar eden kelimeler çıkartılmış ve belirlenen özniteliklere ayrıca tekrar eden kelime sayısı eklenmiştir.
EğitimSeti-10	Zemberek kütüphanesi kullanılarak Türkçe kelime denetimi yapılmış ve kelimeler hem imlaya hemde anlama göre düzeltilmiş ve belirlenen öznitelikler eklenmiştir.
EğitimSeti-11	Tweet ve sadece duygu durumu eklenmiştir.
EğitimSeti-12	Tweette sadece url, hashtag temizliği yapılmıştır. Sayılar, sayısal değerler temizlenmemiş ve belirlenen öznitelikler eklenmiştir.

Etiketleme işleminden sonra oluşturulan dosya formatındaki 12 farklı eğitim seti içerikleri aşağıdaki Tablo 6’da görüldüğü gibi oluşturulmuştur.

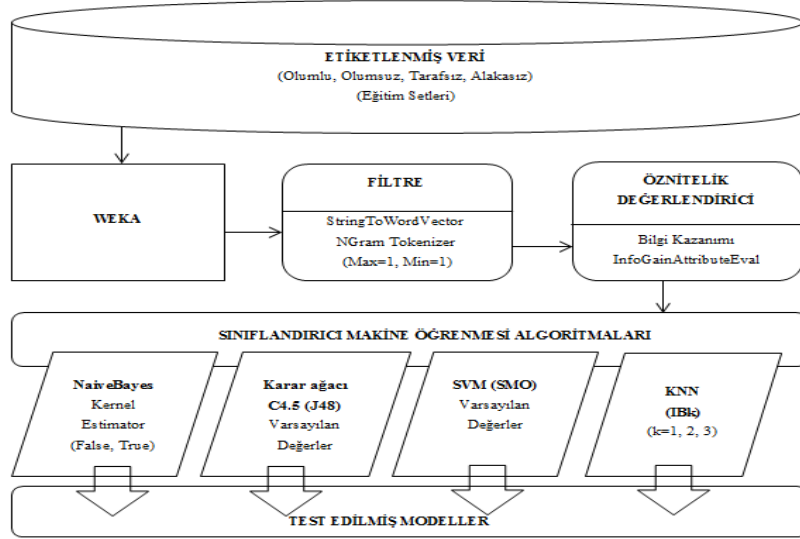
Tablo 6. Etiketleme işleminden sonra dosya isimleri ve işlenmiş tweet içerikleri

Dosya Adı	Tweet
EğitimSeti-1	'maç özetleri ve maç yayınları ligin haftasını tüm maçları birzadan izlyin',
EğitimSeti-2	'maç özetleri ve maç yayınları ligin haftasını tüm maçları birzadan izlyin'
EğitimSeti-3	'maç özetleri ve maç yayınları ligin haftasını tüm maçları birzadan izlyin'
EğitimSeti-4	'maç özetleri maç yayınları ligin haftasını maçları birzadan izlyin'
EğitimSeti-5	'maç özetleri maç yayınları ligin haftasını maçları birzadan izlyin'
EğitimSeti-6	'maç özet yayın lig maç'
EğitimSeti-7	'maç özet yayın lig maç'
EğitimSeti-8	'maç özetleri yayınları ligin haftasını maçları birzadan izlyin'
EğitimSeti-9	'maç özetleri yayınları ligin haftasını maçları birzadan izlyin'
EğitimSeti-10	'maç özetleri ve maç yayınları ligin haftasının tüm maçları birzadan izleyin'
EğitimSeti-11	'maç özetleri ve maç yayınları ligin haftasını tüm maçları birzadan izlyin'
EğitimSeti-12	'maç özetleri ve maç yayınları ligin 20 haftasını tüm maçları birzadan izlyin'

## 2.6. Makine Öğrenmesi Algoritmaları ve Model Oluşturma

Mevcut veriler ile eğitilerek yeni veriler hakkında kestirim yapabilmek Makine Öğrenmesi algoritmalarının başarılı olduğu bir alandır. Makine Öğrenmesi; veriyi ayrıştırmak, veriden öğrenmek ve ardından dünyadaki bir şey hakkında bir tahmin yapmak için algoritmalar kullanma uygulamasıdır [24]. Duygu analizinde makine öğrenmesi; sınıflandırıcı algoritmalar kullanılarak yani denetimli öğrenme yapılarak veriden olumlu, olumsuz, tarafsız, alakasız vb. gibi duygu sınıflarını belirlemek için kullanılır. En çok kullanılan denetimli makine öğrenmesi algoritmaları [25], veri setine uygun olması göz önünde bulundurularak çalışmada kullanılmıştır. Çalışma için farklı çalışma prensiplerine dayanan denetimli makine öğrenmesi algoritmaları olarak; C4.5 Algoritması, Destek Vektör Makinesi (SVM- Support Vector Machine), Naive Bayes Algoritması, K-En Yakın Komşu Algoritması (KNN- K-Nearest Neighbors) algoritmaları seçilmiştir. Bu çalışma için hazırlanan farklı özneliklere sahip 12 farklı eğitim veri seti ile Naive Bayes, C4.5, SVM, KNN algoritmalarının farklı parametreler ile denemeleri yapılmış ve sınıflandırıcı modeller oluşturulmuştur. Şekil 4 modelleri oluşturmak ve için yapılan işlem basamaklarını göstermektedir. İlk başta eğitim veri seti arff formatlı metin dosyası biçiminde verilmiştir. Weka yazılım aracı ve arff formatı hakkında açıklayıcı bilgi [26] ilgili bağlantıda verilmiştir.





Şekil 4. Model Oluşturma işlem basamakları

### 3. Araştırma Sonuçları ve Tartışma

#### 3.1. Makine Öğrenmesi Algoritmalarının Başarımı

Çalışmada tüm eğitim setleri için Naive Bayes, SVM, C4.5 ve KNN Makine Öğrenmesi algoritmaları 10 kat çapraz doğrulama yapılarak modeller oluşturulmuştur. Sınıflandırma algoritmalarının başarımlarını değerlendirme doğruluk ölçütü kullanılarak yapılmış, algoritma ve eğitim setleri için elde edilen modellerin başarımları Tablo 7’de gösterilmiştir.

Tablo 7. Sınıflandırma algoritmaları doğruluk başarımları

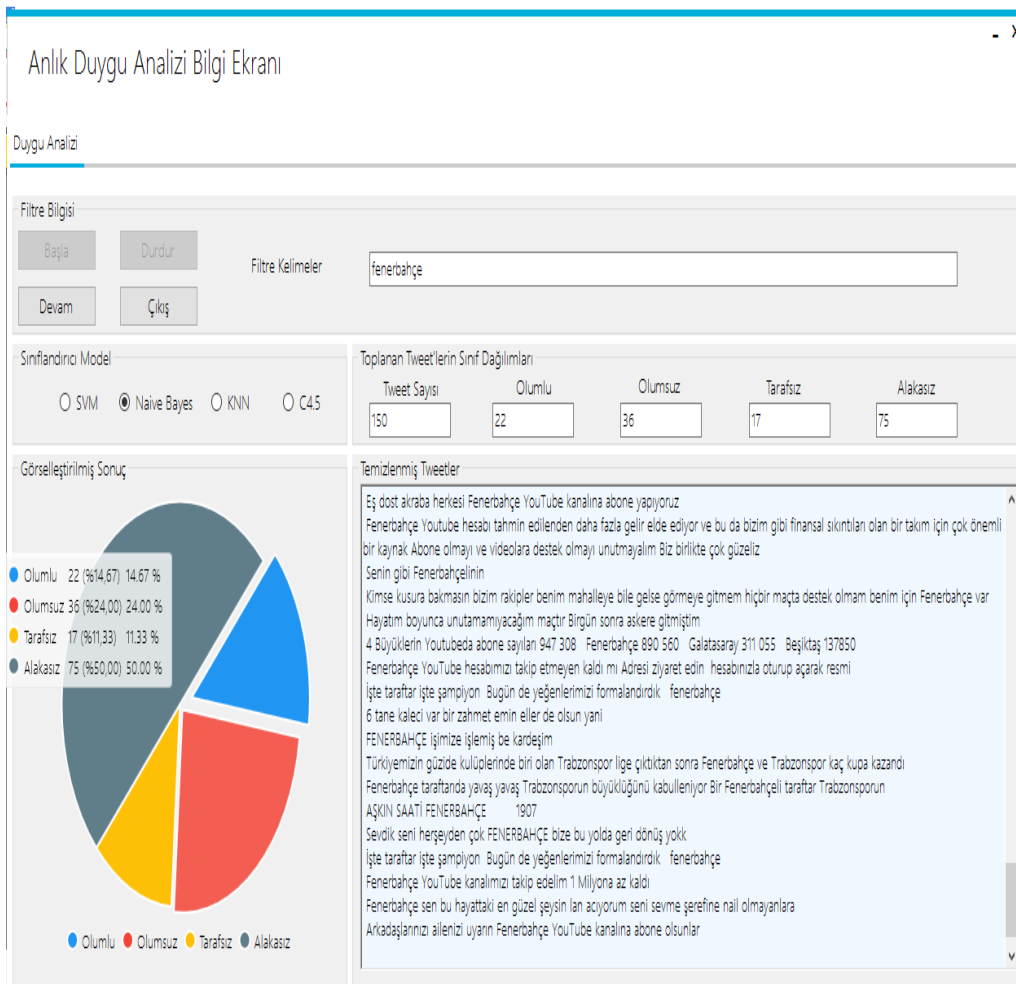
Dosya Adı	Naive Bayes (kernel Estimator = false) %	Naive Bayes (kernel Estimator = true) %	SVM %	C4.5 %	KNN (k=1) %	KNN (k=3) %
EğitimSeti-1	75,75	83,91	92,2	<b>89,6</b>	85,68	84,66
EğitimSeti-2	75,81	83,87	<b>92,3</b>	89,3	85,59	84,71
EğitimSeti-3	75,31	83,31	91,9	89,3	85,4	84,74
EğitimSeti-4	<b>76,82</b>	83,6	92,1	88,6	85,39	85,03
EğitimSeti-5	75,38	82,9	91,6	89	85,14	84,8
EğitimSeti-6	75,98	83,28	90,8	88,1	82,42	81,01
EğitimSeti-7	75,41	83,02	90,4	87,6	82,14	80,38
EğitimSeti-8	75,5	82,95	91,6	88,8	85,11	84,69
EğitimSeti-9	75,21	82,92	91,6	88,6	84,89	84,73
EğitimSeti-10	76,32	<b>84,3</b>	92,2	89,3	86,22	<b>85,04</b>
EğitimSeti-11	74,09	83,99	91,6	88,4	<b>87,73</b>	84,57
EğitimSeti-12	76,03	83,87	92,2	89,5	85,45	84,49

Yapılan çalışmadan elde edilen sonuçlar değerlendirildiğinde EğitimSeti-10 için verilen sonuç satırı değerlendirildiğinde Zemberek kütüphanesini kullanarak hem Türkçe imla denetimi hem de Türkçe anlam denetimi yapmak tüm algoritmaların başarımlarını arttırdığı görülmektedir. EğitimSeti-2 satırı bakıldığında SVM makine öğrenmesi algoritması ile en yüksek doğruluk başarımları elde edilmiştir. Ayrıca KNN algoritmasına öznelik olarak sadece tweet verildiğinde sınıflama performansının diğer öznelikli olan eğitim

setlerine göre sınıflama başarımının artmış olduğu görülmektedir. C4.5 algoritmasında tweet içindeki sayısal değerleri temizlemenin algoritmanın başarımına katkı sağladığı ortaya çıkmıştır. Tweet içinden Etkisiz kelimeleri temizlemenin Naive Bayes algoritmasının performansını artırdığı tespit edilmiştir.

### 3.2. Anlık Twitter Duygu Analizi

Anlık olarak Twitter'dan paylaşılan tweetlerin duygularını belirlemek için oluşturulan hazır modellerden seçilen modelleri kullanan bir model uygulama geliştirilmiş ve kullanılmıştır. Model uygulama, anlık olarak gelen yeni tweet paylaşımının olumlu, olumsuz, tarafsız ve alakasız olarak sınıfının ne olduğunu tahmin etmiştir. Şekil 5'te görüldüğü gibi anlık duygu analizi yapabilen model uygulama ekranı verilmiştir.



Şekil 5. Anlık duygu analizi sonuçları

## 4. Sonuç

Bu çalışma, futbol taraftarlarının attıkları tweetlerden futbol maçlarına nasıl tepkiler verdiğini anlamak için yapılmıştır. Futbol maçlarında; hakemler, futbolcular, teknik direktörler, goller, penaltılar, kontra ataklar, köşe vuruşları, kaleye şutlar ve ilk yarı istatistikleri, fauller, gösterilen kartlar, topla oynama yüzdesi gibi birçok durum futbol taraftarların tepkilerine neden olabilmektedir. Maç izlerken taraftarların duygularını ifade etmesinin bir yolu da sosyal medyadır. Taraftarlar maç sırasında sosyal medyayı kullanarak kendi aralarında daha fazla etkileşime girmektedir ve duygularını paylaşmaktadır. Bu çalışmada maç oynanırken taraftarların tuttukları futbol takımının performansına bağlı olarak attıkları tweetler incelenmiştir. Bu paylaşımlar taraftarların duygularına göre maçı değerlendirmek için kullanılmıştır. Böylece futbol taraftarlarının tweetlerindeki duygusal tepkileriyle, futbol maçlarının gerçek zamanlı ilişkili olduğu gösterilmiştir.

Bu çalışmada tweetleri doğru şekilde sınıflandırmak için birkaç farklı makine öğrenmesi algoritması kullanılmıştır. Twitter'da anahtar kelime tabanlı bir arama yapılarak veri setleri oluşturulmuştur. Veri setlerinin el ile duygu sınıfı etiketlenmesine kolaylık sağlaması için geliştirilen uygulama yardımıyla tweetler tek tek okunarak etiketlenmiştir. Farklı özneliliklerin ve ön işlemlerin etkisi araştırılmış ve denemeleri yapılmıştır. Bu denemelerden, sınıflandırma başarımına etkisi olmadığı belirlenen öznelilik ve ön işlemler çalışmada dikkate alınmamıştır. Sınıflandırma amacıyla farklı algoritmaların, sınıflandırma başarımını görmek için SVM, KNN, C4.5, Naive Bayes algoritmaları ile denemeler yapılmıştır. Sınıflandırma başarımının, sınıflandırma için kullanılan algoritmalara bağlı

olmasının yanı sıra, ön işlem aşamasında yapılan etkisiz kelime, emoji, noktalama işareti vb. gibi temizleme işlemlerine ve ayrıca bilgi kazancı olan öznitelikleri seçmeye de bağlı olduğu görülmüştür.

Bu çalışmayla hem Türkçe imla denetimi hem de Türkçe anlam denetimi yapmanın sınıflandırma başarımını artırabileceği görülmüştür. SVM algoritmasının, belirlenmiş olan öznitelikleri içeren eğitim setlerinde, tweet metni sınıflandırmada daha yüksek başarımla çalıştığı görülmüştür. Ayrıca KNN algoritmasının başarımının, öznitelik sayısının azalması ile arttığı görülmüştür. Bunun sebebi öznitelik sayısının artışı göstermesiyle boyut artmakta ve ilişkisiz özniteliklerin artması da ayırt ediciliği yüksek olan öznitelikleri etkisiz hale getirmesidir. Bu durum KNN algoritmasının başarımını düşürmektedir. C4.5 algoritmasında sayısal değerleri çıkarmanın algoritma başarımına pozitif katkı sağladığı belirlenmiştir. Naive Bayes algoritması, tweet metnindeki duygularla ilgili kelimelerin sıklığına dayalı sınıflandırma yaptığı için etkisiz kelimelerin temizlenmiş olduğu veri setinde en yüksek başarımla elde etmiştir.

## **Kaynakça**

- [1] “What Happens in an Internet Minute in 2019?” [Çevrimiçi]. Available: <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>. [Erişildi: 2020 09 07].
- [2] LiveStats, I. (2018). Twitter usage statistics - Internet live stats. [www.internetlivestats.com/twitter-statistics/](http://www.internetlivestats.com/twitter-statistics/). [Erişildi: 2019 09 08].
- [3] Omnicore (2018). Twitter by the numbers: Stats, demographics & fun facts. <https://www.omnicoreagency.com/twitter-statistics/>. [Erişildi: 2019 09 09].
- [4] Khritantsev, M., Shehadeh, M., A Thesis in the Field of Finance For the Degree of Master of Science in Finance Lund University, 2018.
- [5] Sentiment Analysis. [Çevrimiçi]. Available: <https://www.lexalytics.com/technology/sentiment> [Erişildi: 2020 06 12].
- [6] Evangelos Kalampokis, Efthimios Tambouris, Konstantinos. Tarabanis, Understanding the predictive power of social media, *Internet Res.* 23 (5)(2013) 544–559.
- [7] A.S. Abrahams, J. Jiao, W. Fan, G.A. Wang, Z. Zhang, What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings, *Decis. Support Syst.* 55 (4) (2013) 871–882.
- [8] Augusto José Waszczynskyj Antunes das Neves, Patricia Zeni Marchiori, Qualidade percebida em produtos e serviços em eventos: técnicas eferramentas para análise de conteúdo Twitter, *Turismo Desenvolvimento* 2(21/22) (2014) 173–182.
- [9] Sitaram Asur and Bernardo A. Huberman. Predicting the future with socialmedia. *IEEE/WIC/ACM International Conference on Web Intelligence and In-telligent Agent Technology*, 2010.
- [10] Cynthia Chew and Gunther Eysenbach. Pandemics in the age of twitter: Contentanalysis of tweets during the 2009 h1n1 outbreak. *PLoS ONE*, 5(11), November2011.
- [11] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitterusers: Real-time event detection by social sensors. *World Wide Web Conference(WWW)*, 2010.
- [12] Ficamos, P., Liu, Y., A Topic based Approach for Sentiment Analysis on Twitter Data, (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 12, pp.201-205, 2016.
- [13] Kethavath, S., Classification of Sentiment Analysis on Tweets using Machine Learning Techniques, Department of Computer Science and Engineering National Institute of Technology Rourkela Rourkela, India, 2015.
- [14] Liu, I. L. B. , Cheung, C. M. K., Lee, M. K. O., “Understanding Twitter Usage: What drive people Continue to Tweet,” in *Proceedings of the Pacific Asia Conference on Information Systems*, pp. 928–939, 2010.
- [15] Bollen, J., Mao, H., Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1–8.
- [16] O’Connor, B., Balasubramanyan, R., Routledge, B. R., Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 122–129. AAAI Press.
- [17] Mozetič, I., Grčar, H., Smailović, J., Multilingual Twitter Sentiment Classification: The Role of Human Annotators, 2016. (<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155036>)
- [18] Liu, B., Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies.*, Vol. 5, No. 1, Pages 1-167, 2012 (<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>).
- [19] Liu, B., *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2015.
- [20] Y. Yu and X. Wang, “World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans tweets,” *Computers in Human Behavior*, vol. 48, pp. 392–400, 2015
- [21] T. Parlar, Feature Selection for Sentiment Analysis in Turkish Texts, Published Ph.D. Thesis, Adana, 2016.
- [22] Akın, A. A. ve Akın, M. D. (2007). Zemberek, an open source nlp framework for turkish languages. *Structure*, 10:1–5.
- [23] B. A. Kaplan, Sentiment Analysis on Turkish Twitter Messages by Using Data Mining, Published Master's Thesis, İstanbul, 2016.
- [24] Nvidia. (2020). [Çevrimiçi]. Available: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>. [Erişildi: 2020 08 17].
- [25] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., Steinberg, D. (2007). *Top 10 algorithms in data mining*. Springer-Verlag.
- [26] Waikato, D. o. (2020). Weka 3: Data Mining Software in Java. [Çevrimiçi]. Available: <http://www.cs.waikato.ac.nz/~ml/weka/>. [Erişildi: 2020 07 27].