

**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ**

**TWİTTER DUYGU ANALİZİ ÜZERİNE BİR UYGULAMA
(YEMEKSEPETİ ÖRNEĞİ)**

**MERVE TUNALI
YÜKSEK LİSANS TEZİ
İŞLETME ANABİLİM DALI**

**GEBZE
2023**

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

TWİTTER DUYGU ANALİZİ ÜZERİNE BİR
UYGULAMA (YEMEKSEPETİ ÖRNEĞİ)

MERVE TUNALI
YÜKSEK LİSANS TEZİ
İŞLETME ANABİLİM DALI

DANIŞMANI
PROF. DR. HAKAN KİTAPÇI

GEBZE
2023



YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun 08/02/2023 tarih ve 2023/06 sayılı kararıyla oluşturulan jüri tarafından 20/02/2023 tarihinde tez savunma sınavı yapılan Merve TUNALI' nın tez çalışması İşletme Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI)

: Prof. Dr. Hakan KİTAPÇI

ÜYE

: Doç. Dr. M. Şahin GÖK

ÜYE

: Prof. Dr. Ayşe GÜNSEL

ONAY

Gebze Teknik Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR

ÖZET

Sosyal medya pazarlama faaliyetlerini yürütmek ya da bir konuya ilişkin olarak kamuoyunun fikrini almak için etkin olarak kullanılan mecralardır. Buralarda özellikle şirketlerin hedef kitlelerine kolaylıkla ulaşabilmeleri ve onlarla etkileşim kurabilmeleri mümkündür. Sosyal medya üzerinden toplanan bilgiler işletmelerin gelecek projeksiyonları oluşturmalarına da yardımcı olur. Ancak buralarda çok fazla veri vardır ve önemli olan ihtiyaç duyulan doğru veriyi toplayabilmektir. Hem çok fazla veri üretilmesi hem de özellikle yazım esnasında sosyal medyada gramer kurallarına uyulmaması, işletmelerin veri toplamalarını ve yorumlamalarını engellemektedir. İşte bu noktada, makine öğrenmesi, veri madenciliği ve analizler devreye girmektedir. Özellikle duygu analizi, insanların sosyal medyada, belli bir konu hakkında hangi duygular ile paylaşım yaptıklarının anlaşılması konusunda karar vericilere destek sağlamaktadır. Bu çalışmada da veri madenciliği ve duygu analizi ile Yemeksepeti isimli şirkete ilişkin olarak atılan tweetler incelenmiştir. Çalışmada Destek Vektör Makinesi, Random Forest KNN ve Naive Bayes yöntemleri ile duygu analizi yapılmıştır. Elde edilen bilgilere göre genel olarak kullanıcıların olumlu duyguları mevcuttur. Ancak her yöntem ayrı istatistik sonuç vermiştir. Gelecekteki çalışmalarda kullanıcılara gerçek duyguları sorularak farklı metotlar ile yapılan duygu analizlerinin sonuçlarının karşılaştırılması ve hangi yöntemin daha doğru sonuç verdiği karşılaştırılması mümkündür.

Anahtar Kelimeler: Duygu analizi, Sosyal medya, Veri madenciliği, Yemeksepeti

SUMMARY

Social media are the channels that are used effectively to carry out marketing activities or to get the opinion of the public on a subject. Here, it is possible for companies to reach their target audiences easily and interact with them. Information gathered through social media also helps businesses create future projections. However, there is a lot of data here and the important thing is to collect the right data needed. Both the production of too much data and the non-compliance with grammatical rules in social media, especially during writing, prevent businesses from collecting and interpreting data. This is where machine learning, data mining and analytics come into play. In particular, sentiment analysis provides support to decision makers in understanding what feelings people share about a certain issue on social media. In this study, tweets about the company named Yemeksepeti were examined with data mining and sentiment analysis. In the study, sentiment analysis was performed with Support Vector Machine, Random Forest KNN and Naive Bayes methods. According to the information obtained, users generally have positive feelings. However, each method gave different statistical results. In future studies, it is possible to compare the results of sentiment analysis with different methods by asking users about their real feelings and comparing which method gives more accurate results.

Keywords: Sentiment analysis, Social media, Data mining, Yemeksepeti

TEŐEKKÜR

BaŐta, yksek lisans eđitimimde ve akademik hayatımda desteđini ve yardımlarını hićbir zaman esirgemeyip bilgisi ile bu ćalıŐmanın oluŐmasının yolunu aćan danıŐmanım Prof. Dr. Hakan KİTAPćI' ya,

Btn ćalıŐmam boyunca beni cesaretlendiren, her daim yanımda olan, gstermiŐ olduđu desteklerinden dolayı aileme ve bu yolda yanımdan hić ayrılmayan, niversite yolunu beraber yrdđm canım kızım Zeynep Asya TUNALI' ya en ićten teŐekkrlerimi sunarım.



İÇİNDEKİLER

ÖZET	iii
SUMMARY	iv
TEŞEKKÜR	v
SİMGELER ve KISALTMALAR DİZİNİ	viii
ŞEKİLLER DİZİNİ	ix
TABLolar DİZİNİ	ix
1. GİRİŞ	1
1.1. Çalışmanın Amacı	1
1.2. Çalışmanın Önemi	2
1.3. Çalışmanın Sınırlılıkları	2
1.4. Çalışmanın Evren ve Örneklemi	2
2. TEMEL KAVRAMLAR	3
2.1. Sosyal Medya Analitiği	3
2.2. Duygu Analizi	4
2.2.1. Duygu Analizi Seviyeleri	8
2.2.2. Duygu Analizi Süreci	9
2.2.3. Duygu Analizinin Önündeki Engeller	10
2.3. Veri Madenciliği ve Makine Öğrenmesi	10
2.3.1. Makine Öğrenmesi Teknikleri	11
2.3.1.1. Denetimli Öğrenme	11
2.3.1.2. Denetimsiz Öğrenme	12
2.3.2. Veri Madenciliği	13
2.3.2.1. Veri Madenciliği Teknikleri	15
2.3.2.2. Veri Ambarı	18
2.3.2.3. Veri Madenciliği ve Sınıflandırma	19

3. YÖNTEM	22
3.1. Duygu Analizi Adımları	22
3.2. Bulgular	30
4. SONUÇ	35
KAYNAKLAR	37
ÖZGEÇMİŞ	40
EKLER	41



SİMGELER ve KISALTMALAR DİZİNİ

Simgeler ve Açıklamalar

Kısaltmalar

TF : Terim Frekansı

IDF : Ters Metin Frekansı

API : Uygulama Programlama Arayüzü

Rt : Retweet

ŞEKİLLER DİZİNİ

Şekil 1. Duygu Analizi Seviyeleri	8
Şekil 2. Duygu Analizi Süreci	9
Şekil 3. Disiplinler Arası Veri Madenciliği	14
Şekil 4. Veri Madenciliği Süreçleri	15
Şekil 5. Veri Madenciliği Modelleri	16
Şekil 6. Regresyon Modeli	18



TABLolar DİZİNİ

Tablo 1. Kod Anlatımı	22
Tablo 2. Veri Ön İşleme Süreci	23
Tablo 3. Veri Temizleme İşlemi	23
Tablo 4. Özellik Seçimi	24
Tablo 5. Eğitim ve Test	24
Tablo 6. Naive Bayes	25
Tablo 7. Random Forest	26
Tablo 8. KNN	26
Tablo 9. Destek Vektör Makinesi (SVM)	27
Tablo 10. API Veri Çekimi	28
Tablo 11. Aranacak Kelime ve Sayının Girişi	28
Tablo 12. Dataframe	29
Tablo 13. Veri Setinin Görselleştirilmesi	30
Tablo 14. Naive Bayes Veri Setinin Grafiğe Dönüştürülmesi	30
Tablo 15. KNN Veri Setinin Grafiğe Dönüştürülmesi	31
Tablo 16. Random Forest Veri Setinin Grafiğe Dönüştürülmesi	31
Tablo 17. Destek Vektör Veri Setinin Grafiğe Dönüştürülmesi	32
Tablo 18. Model Sonuçlarının Karşılaştırılması	32
Tablo 19. Oluşturulan Veri Setinin Vektörlerinin Alınması	33
Tablo 20. Destek Vektör Makinesi Grafiği	33

1. GİRİŞ

Sosyal medya günümüzde insanların diğer insanlar ile etkileşim içine girdiği, bilgi alışverişi gerçekleştirdiği, gerçek ve tüzel kişiler ile doğrudan temas kurdukları, eğlendikleri ve hatta alışveriş, eğitim ya da fikir belirtme gibi çeşitli aktivitelerde buldukları ortamlardır. Sanal olmasına karşın özellikle işletmeler için insanların fikirlerini anlamak konusunda gerçek verilerin saklandığı karmaşık yapılar olan sosyal medya ortamları, günümüz dünyasının en önemli kaynaklarından biri olan bilgiyi doğrudan sunmaktadır. Ancak gerek veri kalabalıklığı gerekse sosyal medya mecralarının neredeyse sonsuz olduğu seçenekleri nedeni ile bilgiyi doğru şekilde elde etmek güçleşmektedir. Dolayısıyla bilginin ayıklanması ve en uygun ve doğru olanının alınması ve işlenmesi gerekliliği doğmaktadır. İşte bu noktada, dijital dünya özelinde kullanılacak olan çeşitli araştırma yöntemleri devreye girmektedir.

Sosyal medya mecraları bilgi ve iletişim teknolojisinin ürünüdür. Burada gerçek ve tüzel kişilerin yazılı, sözlü, görsel olarak ya da bunların bir karışımı şeklinde paylaşım yapmaları olanağı vardır. Üstelik bir tek birey dahi milyonlarca veri üretebilmektedir. Bugün sosyal medya içerisinde doğrudan yer almasa bile bireyin buralardan etkilendiğini söylemek yanlış olmayacaktır. Zira politikaların oluşturulması, kanunların çıkarılması ya da kamuoyu yaratılması gibi bireyi de doğrudan ve dolaylı olarak etkileyen hemen her unsur sosyal medyanın da etkisi altındadır. İşte bu nedenle özellikle toplum ile doğrudan etkileşim içinde olan kar amaçlı kurumlar, buralardan sürekli olarak bilgi toplama ihtiyacı duymaktadırlar.

1.1. Çalışmanın Amacı

Bu çalışmanın amacı, sosyal medyada yer alan paylaşımların incelenerek bir şirketin gerçekleştirdiği pazarlama faaliyetlerine ve aldığı kararlara yön verilip verilemeyeceğinin anlaşılmasıdır.

1.2. Çalışmanın Önemi

Nihai olarak kar amaçlayan şirketlerin, toplumun dinamiklerini doğru anlaması ve kendi politikalarına da buna göre yön vermesi gerekmektedir. Sosyal medyanın karmaşık dünyasında doğru bilginin elde edilebilmesi için duygu analizi uygulamaları yapmak gerekmektedir. Bu analizlerin sonucunda da şirketlerin karar vericilerine topluma ve onun düşüncelerine ilişkin olarak bilgi akışı sağlanmaktadır. Bu çalışmada uygulanan analiz ile seçilen şirketin gerçekleştirdiği faaliyetler neticesinde insanların nasıl tepki verdikleri ve bu tepkilerin altında hangi duyguların yattığı bilgisine ulaşılmaya çalışılacaktır. Duyguları doğru anlamlandırmanın şirketlerin geleceklerini doğrudan etkilediği varsayımı ile yapılan çalışmada gerçekleştirilen örnek uygulama; yeni pazarlar bulmak isteyen ve bu pazarda tüketicilerin duygu, tutum ve eğilimlerini anlamaya çalışan şirketler, toplumun duyarlılıklarını anlamlandırmak isteyen sivil toplum kuruluşları, bireyler ve diğer tüzel kişiler için bir yol gösterici niteliktedir.

1.3. Çalışmanın Sınırlılıkları

Bu çalışma yalnızca Twitter üzerinde gerçekleştirilen paylaşımlar özelinde hazırlanmıştır. Diğer sosyal medya mecralarında yer alan kullanıcıların paylaşımlarının analiz edilmemesi bir sınırlılıktır. Ayrıca sosyal medyada yer alan şirketler arasında oldukça aktif olarak burada paylaşım yapan ve etkileşime sahip olan Yemeksepeti firması seçilmiştir. Firma oldukça geniş bir takipçi sayısına sahip olmasına karşın bu şirket özelindeki bir çalışmanın farklı sektörlerdeki şirketlere ilişkin olarak uyarılama yapılamayacağını belirtmek gerekir. Çalışma kapsamında belirlenen tarih aralığı ve anahtar kelime uygulamaları da birer sınırlılık olarak kabul edilmektedir.

1.4. Çalışmanın Evren ve Örneklemi

Çalışmanın evreni tüm tarihleri, tüm sosyal medya platformlarını ve buralarda yer alan tüm şirketleri kapsamaktadır. Örneklem olarak ise Twitter isimli sosyal medya platformunda, Yemeksepeti isimli şirkete ilişkin olarak, sosyal medya kullanıcılarının duygu analizi gerçekleştirilmiştir.

2. TEMEL KAVRAMLAR

2.1. Sosyal Medya Analitiđi

İnsanlar, dođal olarak sosyal olan varlıklardır. Çeşitli ihtiyaçlarını karşılamak için diđerlerine ihtiyaç duyan insan özellikle dijital teknolojilerin gelişmesi ile birlikte bilişim sistemleri aracılığı ile de diđerleri ile sosyalleşmeye başlamıştır. Dijital iletişim hem anonimlik hem de mahremiyet kazandırabilmesinin yanı sıra, ucuz ve hızlı, bireyler arası ya da birey- topluluk iletişiminin de imkanını oluşturmaktadır. Bu avantajlar, sosyal medya adı verilen ve dijital ortamda sosyalleşmeyi olanaklı kılan sistemlere karşı talebin her geçen gün artmasını da beraberinde getirmiştir. Davranış bilimlerine göre insanlar sosyal medya platformlarında, aidiyet, öz sunum ve izlenim ihtiyaçlarını karşılayabilmektedirler. Kuşkusuz psikolojik faktörler ile teknolojinin bir bileşkesi olan sosyal medya platformları, insanların ne düşündüğünü ve hissettiğini anlamak ve bunları toplumun yapısına göre analiz etmek için de olanaklı olan ve web teknolojisine dayalı bulunan alanlardır (Çelik ve ark., 2021: 881).

Sosyal medya platformlarını çeşitli alt kategorilere bölmek mümkündür. Bunlar özellikle yapılan paylaşım biçimlerine göre sınıflandırılmaktadır (Cebeci, 2020. 192):

i. İşbirlikçi Platformlar:

Wiki olarak adlandırılan iş birlik platformları vardır. Burada içerik üretimi için kullanıcının desteđi alınmaktadır. Ülkemizde de en yaygın örneklerden biri Ekşi Sözlük'tür. Buralarda kişiler bir konu hakkında sahip oldukları bilgiyi diđerleri ile paylaşmaktadırlar.

ii. Blog ve Miniblog Platformları:

Yine kişilerin doğrudan içerik üretebildikleri alanlardır. Kişi doğrudan kendi adına ya da anonim olarak paylaşım yapabilmektedir. Ancak Twitter gibi pek çok blog platformunda paylaşım yaparken, üretilen içerik üzerinde karakter sınırlamaları mevcuttur.

iii. İçerik Topluluğu:

Youtube en yaygın örnektir. Birey oluşturduğu videolu ya da sesli içeriği yayınlar ve bunlar bireyin aboneleri tarafından doğrudan görüntülenebilir. Buralarda ayrıca konuya göre arama yapılabilen ve o konu ile ilgili içeriklere aynı topluluk içinde olunmasa bile ulaşılabilir.

iv. Sosyal Ağlar:

Kişilerin doğrudan kendi profillerini oluşturdukları, tamamen açık ya da kendi ağları özelinde paylaşım yapabildikleri ya da diğer paylaşımlar ile etkileşime girebildikleri alanlardır. Buralarda resimler, yazılar, web sitesi içerikleri, videolar gibi çok çeşitli paylaşımlar yapılma imkanının olmasının yanı sıra birey ya da grup olarak mesajlaşma olanakları da bulunmaktadır.

Sosyal medya platformları yukarıda sayılanlar ile sınırlı değildir. Örneğin Meta gibi sanal dünyalar yaratan platformlar da vardır. Yine kimi oyunlar içerisinde kullanıcıların birbirleri ile iletişime girebilmeleri ve oyun oynarken burayı sosyalleşme ortamı olarak kullanabilmeleri de mümkündür. Tıpkı diğer yönetsel süreçlerde iş içi ve dışındaki koşulları analiz eden şirketler, sosyal medya analitiği sayesinde potansiyel müşterilerinin ve/veya rakiplerinin de neler yaptığını, ne düşündüğünü ya da hissettiğini anlamaya çalışmaktadırlar. Literatürde sosyal medya madenciliği ile daha çok kast edilen, sosyal medya içerikleri üzerine yapılan metin madenciliği eksensiz çalışmalardır. Metin madenciliğinden farklı olarak değerlendirilmesi altında yatan temel neden ise sosyal medya mesajlarının (özellikle tweetler) içerdiği özel jargon (kısaltmalar) ve resimler (emoji) ile düz metinlere göre yapısal olarak değişkenlik göstermesidir.

2.2. Duygu Analizi

İşletmeler ya da politik karar vericiler gibi doğrudan halk ile ilişki içinde olan kişi ve kurumlar için insanların ne düşündüğünü anlamak oldukça önemlidir. Özellikle ticari faaliyetlerde rekabetin ön plana çıktığı 2000 sonrası dönemde, insanların duygularını anlama konusunda kar amacı güden kurumların yüksek bir çabası olmaya başlamıştır (Ak, 2019: 4). Sosyal medya kullanımı tüketicilerin duygularını anlayabilmek için kar amacı

güden kuruluşlara geniş bir imkan tanımıştır. İnsanların sosyal medya ortamlarında tükettikleri mal ve hizmetler için yorum yapmaları buna olanak tanımaktadır. Bugün şirketlerin, müşterilerinin duygu ve düşüncelerini anlamak için onlara ulaşmalarına ve onlar ile doğrudan temas kurmalarına gerek yoktur. Çünkü bireyler, görüşlerini gönüllü olarak sosyal medya ortamlarında paylaşmaktadırlar. Burada önemli hale gelen sosyal medya içeriklerinin incelenmesi ve doğru yorumlanmasıdır ve işte bu noktada duygu analizi olgusu devreye girmektedir (Cebeci, 2020: 194).

Adından da anlaşılacağı üzere duygu analizi, insanların gerçekleştirdiği paylaşımların hangi duyguları yansıttığını anlamaya yönelik olarak yapılan uygulamalardır. Sosyal medya analitiği, duygu analizinin gerçekleştirilebilmesi için analizcilere önemli veriler sunmaktadır. Duygu analizleri ile metin, ses, görüntü verilerinden bilişim destekli olarak otomatik ve yarı otomatik olarak insan davranış, duygu ve inanışlarını çekmek ve yorumlamak mümkün hale gelmektedir. Her ne kadar doğal dil işleme çalışmaları 1950 ve sonrasında başlamış olsa da duyguların anlaşılması zaman almıştır. Duygu analizi yalnızca söyleneni anlamakla kalmayıp, insanların mesajlarını hangi duygular sonucunda yolladıklarını da anlamlandırmaya yaramaktadır.

Duygu analizinin daha etkin yapılabilmesi için makine öğrenmesi, yapay zeka ve veri madenciliği gibi uzmanlık alanlarına da ihtiyaç doğmuştur. Bu bağlamda duygu analizi sürecinin çok disiplinli bir yapısının olduğunu söylemek mümkündür. Duygu analizine ilişkin olarak alan yazınında genellikle doğal dil işleme ve semantik biliminin etkin olduğu söylene de analizin yalnızca metinler ile yapılması sınırlılığının ortadan kalkması, duygu analizinin çalışma alanlarını da genişletmiştir (Singh ve ark., 2014: 13927).

Duygu analizleri esnasında farklı disiplinlerden yardım almanın temelinde, analizi yapılacak olan verinin büyüklüğü vardır. Ayrıca sosyal medyada paylaşımlar oldukça çeşitlidir. Metinlerin yanı sıra görseller ve videolar da paylaşılmaktadır. Emojiler de yine duygu belirten ifadeler olarak hayatlarımızda yer edinmeye başlamıştır. Bunlara ek olarak sosyal medya paylaşımlarında gramerden bağımsız bir yazım diline rastlamak oldukça yaygındır. Anlatım bozuklukları da sıklıkla karşılaşılan diğer unsurlardır ve bunların her biri duygu analizini güçleştiren olgulardır. Yapay zeka başta olmak üzere veri madenciliği

ve makine öğrenmesi, analiz sonuçlarının daha gerçekçi olmasına olanak tanıyan teknoloji yapılarıdır (Varma ve ark., 2017: 139).

Duygu analizinde temel veri kaynağı daha önce ifade edildiği üzere sosyal medyadır. Buna ek olarak yine internet üzerinde yer alan forumlar, inceleme siteleri, bloglar, müşteri geri bildirim siteleri ve metin formatında bilgi talep eden bütün siteler duygu içeren sübjektif dokümanlar içermektedir. Bütün bu metinler için analize konu olduklarında bazı özellikleri taşımaları beklenmektedirler. Aşağıda bu özellikler kısaca tanımlanmıştır (Cebeci, 2020: 194):

i. Duygu: Bir fikir ile doğrudan ilişkili olan his, davranış, değerlendirme gibi sübjektif unsurlardır. Duygular genel kabul görecektir şekilde rasyonel olabildiği gibi kişiye özel göreceli de olabilir. Pratikte duygu analizinin konu alanının kişiye özel duyguların olduğu söylenebilir.

ii. Duygu Hedefi: Metin içerisinde ifade edilen bir duygunun yöneldiği nesneye karşılık gelir. Metin içerisindeki her bir duygu ifadesinin bir nesneye bağlanması analiz sürecinin başarısı için hayati önem taşımaktadır.

iii. Varlık: Duygunun ifade edildiği bir ürün, hizmet, konu, organizasyon veya insan olabilir.

iv. Polarite: Bir duygu ifadesinin yönünü vermektedir. Duygu analizi sürecinde negatif, nötr ve pozitif olarak üçlü polarite kullanılabilir gibi pozitiflik ve negatiflik derecesi de (5 yıldız gibi) tercih edilebilir.

Fikir madenciliği olarak da adlandırılan duygu analizi, bir yazarın metinde hangi duyguları ifade ettiğini otomatik olarak belirleyen hesaplamalı bir görevdir. Bir metnin duygusu genellikle ikili sınıflarla pozitif ve negatif olarak temsil edilir. Ayrıca çok pozitif, pozitif, nötr, negatif, çok negatif vb. çoklu sınıfları kullanarak daha ayrıntılı bir şekilde analiz edilmesi de mümkündür (Engüllü, 2018: 14).

Duygu analizi insanların gerçek duygularını bütünüyle anlamak için mükemmel bir çözüm değildir, ancak analiz edilecek çok fazla metin olduğunda bu metinler hakkında çok hızlı bir şekilde genel bir fikir vermek çok yararlıdır. Özellikle kamuoyunun bir konu

hakkındaki genel düşüncelerini kavramak adına metin analizi yöntemi sıklıkla kullanılmaktadır.

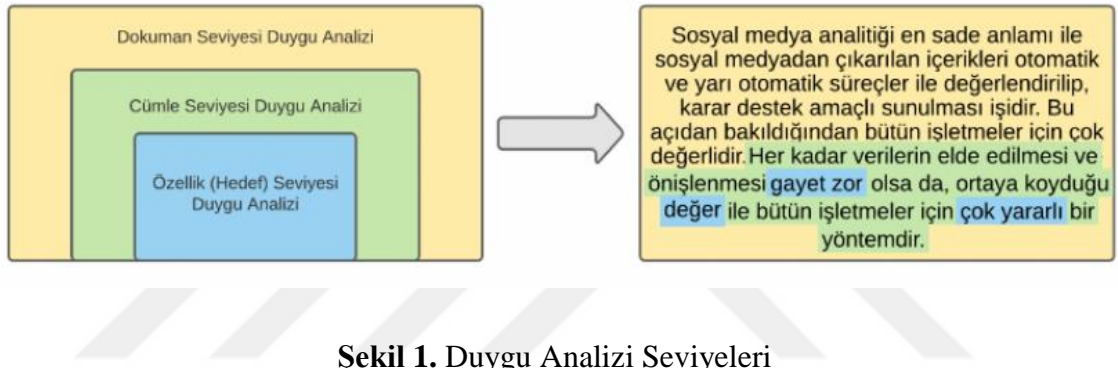
Duygu Analizi, akademik ve ticari çalışmalara dahil edilmektedir. Özellikle 2000'li yıllar öncesinde metaforun yorumlanması, duygu sıfatları, öznellik, algısallık ve etki ile ilgili bazı çalışmalar (Kamruzzaman ve ark., 2005: 3) olmasına rağmen, internet kullanımının artması ile çalışmaların sayısı ve içerikleri de artmaya başlamıştır (Yurtalan, 2016: 5).

Duygu analizi, görüş çıkarma ve semantik sınıflandırma terimleri ilk kez 2003 yılında Nasukawa ve Yi tarafından yapılan çalışmada ortaya çıkmıştır (Nasukawa ve Yi, 2003: 24). Bu çalışma, araştırmacılara duyuların konuşma tahsisinin bir parçası ve bilgi alma teknikleri üzerindeki sınıflandırıcılar hakkında rehberlik etmesi bakımından önemlidir (Yurtalan, 2016: 5).

Alan yazınında genel olarak duygu analizi çalışmaları denetimli öğrenme ve denetimsiz öğrenme teknikleri kullanılarak farklı düzeylerde (belge düzeyi, cümle düzeyi ve ifade düzeyi) duygu sınıflandırmalarına odaklanmışlardır. Denetimli öğrenme, makine öğrenimi tabanlı çalışmalardır ve denetimsiz öğrenme, duygu analizi için LB (Lexion Based) çalışmalarıdır. Duyarlılık sınıflandırmasının ardından geliştirilen algoritmalar ile sonuçlar üretilir. Polarite sınıflandırma çalışmaları, belge veya cümlelerin pozitif veya negatif kutuplarını belirlemek için yapılır. Belgelerin veya cümlelerin öznel mi yoksa nesnel mi olduğunu sınıflandırmak için kullanılan öznellik sınıflandırması üzerine birçok çalışma bulunmaktadır. Polarite sınıflandırma çalışmaları, belge veya cümlelerin pozitif veya negatif kutuplarını belirlemek için yapılır. Belgelerin veya cümlelerin öznel mi yoksa nesnel mi olduğunu sınıflandırmak için kullanılan öznellik sınıflandırması temeline dayanmaktadır. Riloff ve arkadaşlarına göre sübjektif ifadeler görüşleri, rantları, iddiaları, suçlamaları, şüpheleri ve spekülasyonları içerir. Yazarlar, öznellik ifadeleri için dil açısından zengin çıkarma kalıplarının tanımlanması gerektiğinin de altını çizmişlerdir (Riloff ve ark., 2003: 26-27). Kalıplar daha sonra bir cümlenin öznel mi yoksa nesnel mi olduğunu otomatik olarak belirlemek için kullanılmaktadır.

2.2.1. Duygu Analizi Seviyeleri

Duygu analizi çalışmaları incelenen derlemin özellikleri, çalışma alanı ve analist tercihinin bağlı olarak farklı seviyelerde gerçekleştirilebilir. Analiz bir dokümanın tamamının değerlendirilebileceği doküman (yorum, tweet, post) seviyesinde gerçekleştirilebileceği gibi, dokümanların cümlelere ayrılması ile cümle seviyesinde de yürütülebilir. Hatta cümle içerisindeki duyguların belirlenmesi ile varlık (entity) veya özellik (aspect) seviyesinde de yapılabilir. Tamamen derlem dediğimiz analize konu veri kaynağına olarak değerlendirilebilecek bu durum için evrensel doğru bir yaklaşım söz konusu değildir.



Şekil 1. Duygu Analizi Seviyeleri

Kaynak: Cebeci, 2020: 195

Doküman Seviyesi Duygu Analizi (DS-DA) yaklaşımlarında incelenen içerik bir bütün olarak duygu polaritesi açısından değerlendirilir. Burada doküman ile kast edilen mesaj, tweet, post, müşteri yorumu veya geri bildirim olabilir. DS-DA bütün dokümanın tek bir varlık ile ilgili olduğunu varsayar. Bir müşteri e-ticaret sitesindeki spesifik bir cep telefonu modeli ile ilgili bir yorum yazdığında, DS-DA bu yorumun polaritesini hesapladıktan sonra bu cep telefonu modeline bağlar. Fakat birden fazla model ile ilgili karşılaştırmalı bir yorum söz konusu ise analiz bu durumu dikkate almadan sonuç üreteceğinden başarıyı konusunda tereddütler olacaktır.

i. Cümle Seviyesi Duygu Analizi (CS-DA): Bir dokümanın cümlelere bölünerek her bir cümle için değerlendirmelerin yapılabileceği DA türüdür. Yine DS-DA yaklaşımındaki tereddütler burada da karşımıza çıkmaktadır. CS-DA her bir cümlenin tek bir varlık hakkındaki fikirleri ifade ettiğini varsayar, bu durum her

ne kadar DS-DA kadar problemlili olmasa da yine de analiz başarımını etkileyecektir.

ii. Özellik Seviyesi Duygu Analizi (OS-DA): İnce Taneli (Fine-Grained) DA olarak ta adlandırılan bu analiz seviyesinde her bir dokümandaki her bir duygu belirlendikten sonra ve duygu hedefi ve ilgili varlık tespit edilir. Bu şekilde farklı fikirler içeren doküman ve cümlelerde karşılaşılan başarım problemleri elemine edilecektir. Fakat anlaşılacağı üzere gayet zorlu bir analiz süreci gerektirir.

2.2.2. Duygu Analizi Süreci



Şekil 2. Duygu Analizi Süreci

Kaynak: Cebeci, 2020: 196

2.2.3. Duygu Analizinin Önündeki Engeller

Duygu analizi gerçekleştirmek kimi koşullarda oldukça güçtür. Bunun nedeni duyguların çok bireysel olması ve herkes tarafından aynı duygunun aynı şekilde yansıtılmamasıdır. Kimi koşullarda ise duyguların yansıtılmasından çekinme ya da karmaşık duygular yansıtma gibi durumlar söz konusu olabilmektedir. Bu durumda, doğru duygunun hangisi olduğunun ya da baskın olan duygunun ne olduğunun anlaşılması da güçleşmektedir.

Duygu analizi sistemleri, müşteri mesajı belirsiz olduğunda veya karışık duygular içerdiğinde, duyarlılığı doğru bir şekilde belirlemede genellikle zorluk yaşamaktadırlar. Kimi durumlarda ise bağlamın anlaşılması güç olmaktadır. Örneğin “kahveden nefret ediyorum” şeklindeki bir mesajda içecek olan kahve ya da renk olan kahve söz konusu olabilecektir ve makine sistemlerinin bağlamı anlamakta yanılması söz konusu olabilir. Benzer şekilde deyimlerde de bağlam kaymaları olabilir. Bir dilde olumlu bir duyguyu yansıtmakta olan ifadelerin deyimler içinde kullanılması halinde olumsuzluğa dönüşmesi ihtimali vardır ki bu tür durumlarda duygu analizi sistemleri yanlış sonuçlar verebilmektedir (Özdemir Akcan, 2021: 417).

2.3. Veri Madenciliği ve Makine Öğrenmesi

Veri madenciliği ve makine öğrenimi, bilgisayar biliminin ilgili alanlarıdır. Veri madenciliği, sınıflandırma, kümeleme, regresyon ve ilişkilendirme gibi yöntemleri içeren büyük veri kümelerindeki örüntüleri keşfetme sürecidir. Makine öğrenimi, sistemlere açıkça programlanmadan otomatik olarak öğrenme ve deneyimlerden gelişme yeteneği sağlayan bir yapay zeka (AI) uygulamasıdır. Makine öğrenimi algoritmaları, minimum insan müdahalesi ile veriye dayalı tahminler veya kararlar almak için örnek girdilerden bir model oluşturur. Makine öğrenimi, tahmine dayalı analitik, davranış kalıpları ve eğilimleri ve diğer içgörüler için verileri analiz etmek üzere kullanılabilir (Ersöz ve Çınar, 2021: 398).

2.3.1. Makine Öğrenmesi Teknikleri

Alan yazınında pek çok makine öğrenmesi tekniğine rastlamak mümkündür ve teknoloji geliştikçe kuşkusuz yeni teknikler de ortaya çıkacaktır. Ancak burada en yaygın olarak kullanılan iki temel teknik açıklanmıştır. Bunlardan ilki denetimli öğrenme, diğeri ise denetimsiz öğrenme olarak adlandırılmaktadırlar.

2.3.1.1. Denetimli Öğrenme

Denetimli öğrenme, görünmeyen veya gelecekteki veriler hakkında tahminler yapmak için etiketli eğitim verilerini kullanan bir tür makine öğrenimi algoritmasıdır. En yaygın kullanılan makine öğrenimi algoritmalarından biridir ve genellikle yüz tanıma, görüntü işleme, el yazısı tanıma, belge analizi, kredi puanlama ve dolandırıcılık tespiti gibi çeşitli alanlarda kullanılmaktadır.

Denetimli öğrenmede, etiketli eğitim verileri kullanılarak bir algoritma eğitilir. Bu etiketli veriler daha sonra görünmeyen verilerin sonucunu tahmin edecek bir model üretmek için kullanılır. Eğitim süreci, algoritmayı etiketlenmiş verilerle beslemeyi, yani ona bir dizi girdi (özellikler) ve bunlara karşılık gelen çıktıları (etiketler) sağlamayı içerir. Algoritma daha sonra özellik girdileri ile bunların çıktı etiketleri arasındaki temel ilişkileri öğrenmeye çalışır. Eğitim tamamlandıktan sonra model, gelecekteki veri noktalarını tahmin etmek için kullanılır. İki tür denetimli öğrenme algoritması vardır (Çelik ve ark., 2021: 884):

i. Sınıflandırma:

Sınıflandırma, çıktı değişkeninin kategorik olduğu denetimli bir öğrenme problemidir. Amaç, bir veri kümesini analiz etmek ve verilerin hangi kategorilere veya sınıflara ait olduğunu tahmin etmektir. Sınıflandırma algoritmalarına örnek olarak Naive Bayes, Karar Ağaçları ve Destek Vektör Makineleri (SVM'ler) dahildir.

ii. Regresyon:

Regresyon, çıktı deęişkeninin sayısal olduęu denetimli bir öğrenme problemidir. Amaç, bir dizi veriyi analiz etmek ve çıktı deęişkeninin nasıl deęişeceğini tahmin etmektir. Regresyon algoritmalarına örnek olarak Destek Vektörü Regresyonu (SVR) ve Lineer Regresyon verilebilir.

2.3.1.2. Denetimsiz Öğrenme

Denetimsiz öğrenme, etiketlenmemiş ve yapılandırılmamış verileri kullanan bir tür makine öğrenimi algoritmasıdır. Denetimsiz öğrenmede algoritma, herhangi bir etiket veya çıktı deęeri verilmeden verilerdeki kalıpları ve ilişkileri bulmaya çalışır. Sonuç olarak, hem tanımlayıcı hem de üretken görevler için kullanılabilir. Tanımlayıcı görevler, eldeki verilere dayalı olarak modeller veya kurallar oluşturmayı içeren görevlerdir. Bu tür görevlerin amacı, farklı deęişkenler arasındaki anlamlı kalıpları veya korelasyonları belirlemektir. Örneğin, denetimsiz öğrenme, bir veri kümesindeki benzer veri noktalarının kümelerini tanımlamak için kullanılabilir.

Üretken görevler, mevcut verilere dayalı olarak yeni veriler oluşturmayı içerir. Örnekler arasında, eksik deęerleri doldurmak için veri atama veya makine öğrenimi modellerinin performansını iyileştirmek için sentetik veri kümeleri oluşturma yer alır. İki tür denetimsiz öğrenme algoritması vardır (Çelik ve ark., 2021: 885):

i. Kümeleme:

Kümeleme, bir dizi nesneyi farklı kümeler veya gruplar halinde gruplandıran denetimsiz bir öğrenme tekniğidir. Belirli bir veri kümesindeki örüntüleri ve benzerlikleri belirlemek için kullanılabilen bir tür veri analizi tekniğidir. Kümelemenin amacı, birlikte gruplanan ve nesnelerin doğal kümelerini oluşturan benzer veri noktalarını bulmaktır. Kümeleme algoritmalarına örnek olarak k-ortalamlar, hiyerarşik kümeleme ve yoğunluk tabanlı kümeleme sayılabilir.

ii. Boyut Azaltma:

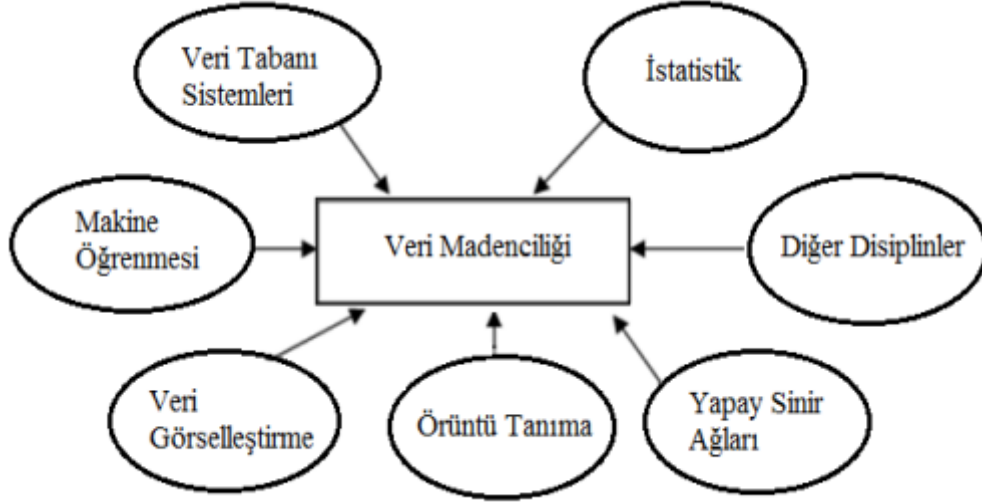
Boyut azaltma, ilgili kalıpları ve bilgileri korurken bir veri kümesinin boyut sayısını azaltan denetimsiz bir makine öğrenimi tekniğidir. Verileri basitleştirmek ve onunla ilişkili değişken ve parametre sayısını azaltarak veri kümelerinin karmaşıklığını azaltmak için kullanılır. Boyut azaltma için en popüler teknikler Temel Bileşen Analizi (PCA), Tekil Değer Ayrışımı (SVD) ve Bağımsız Bileşen Analizidir (ICA).

2.3.2. Veri Madenciliği

Teknolojinin gelişmesi ve kişi ile bilgisayarlar arasında bir etkileşimin ortaya çıkmış oluşu; devamlı şekilde yeni verilerin oluşmasına sebep olmaktadır. Dijital sistemlerin kullanımından dolayı her saniye yapılan milyonlarca işlem kayıt altına alınmaktadır. Bu kayıtlar veri tabanında tutulmaktadır. Kuşkusuz bu depolama, çok büyük ve karmaşık bir veri varlığını ve çok geniş depolama alanlarını da beraberinde getirmektedir. Bu veri depolarına kolaylıkla ulaşılabilmesi ve her işlemin kaydedilmesi, veri karmaşasına neden olmaktadır. İnsanlar, ellerindeki çok büyük veri kaynakları içerisinde kendilerine gerekli olan nitelikli ve öz veriye erişmek için yeni sistemlere ihtiyaç duymaya başlamışlardır. Gereksiz, karmaşık, doğru ve yanlışın bir arada bulunduğu veri depolarından faydalı veriyi elde etmek için veri işleme teknikleri üretilmiş ve bunlar neticesinde veriler analiz edilmeye başlanmıştır. Bu sürece de veri madenciliği adı verilmektedir (Uzun ve ark., 2021: 2).

Veri madenciliği, günümüzde tüm sektörlerde ve tüm organizasyonlarda, karar vericilerin en önemli kaynakları olan bilgiyi sağlamalarına yardımcı olmaktadır. Çok kısa süre öncesine kadar veri kıtlığı yaşanan dünyada bugün veri fazlalığı ve veri kirliliği bulunmaktadır. Bilgiye erişim kolaylaşmasına karşın, bilginin güvenilirliği, ayıklanması, kullanılabilir bilgiye erişim ve bilginin güncel olması gibi endişeler söz konusu olmaya başlamıştır. 1990'lı yıllardan bugüne dek veri madenciliği kullanılmasına karşın, konuya ilişkin bilinç henüz oldukça yenidir. Veri madenciliğinin kullanılabilirliğinin artması için bununla ilgili olarak tanımlamalar yapılmaya başlanmıştır. Bugün alan yazınında geçen tanıma göre veri madenciliği, “büyük ölçekli veriler arasından değeri olan, daha önceden bilinmeyen bir bilgiyi elde etme işidir. Bir diğer ifadeyle veri madenciliği, çoğunlukla

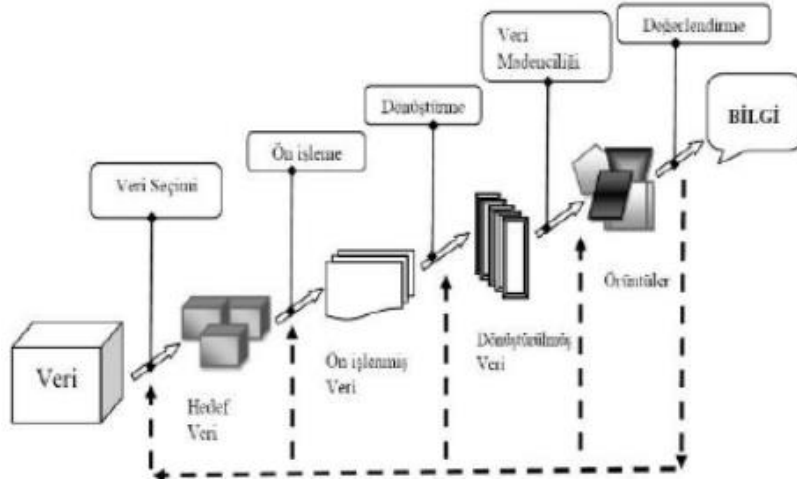
bilgisayarlar aracılığı ile otomatik biçimde toplanan verilerden bilgi elde edilmesi için gerçekleştirilen işlemlere verilen isimdir (Karacan ve Yeşilbudak, 2010: 18; Savaş ve ark., 2012: 3).



Şekil 3. Disiplinler Arası Veri Madenciliği

Kaynak: Şeker Elmas, 2019: 15

Veri madenciliği, veri tabanlarından veya veri ambarlarından elde edilen, çeşitli biçimlerde depolanan, kullanma hazır durumda bulunan büyük ölçekli verilerin içerisindeki gizli kalmış, önceden bilinmeyen ve kullanışlı olduğu düşünülen anlamlı bilgilerin keşfedilmesi sürecidir. Bilginin önceden elde edilmemiş – bilinmeyen – olması, veri madenciliğini çok daha değerli kılmaktadır. Günümüzde pek çok disiplinde, şekil 3'te de görüldüğü üzere, bilgiye erişim için veri madenciliğinden faydalanılır. Burada bilginin önceden bilinmemesi ifadesi, aslında madencilik ile elde edilecek olanın daha önceden tahmin edilmiyor olmasıdır (Çelik, 2015: 9).



Şekil 4. Veri Madenciliği Süreçleri

Kaynak: Uzun ve ark., 2021: 2

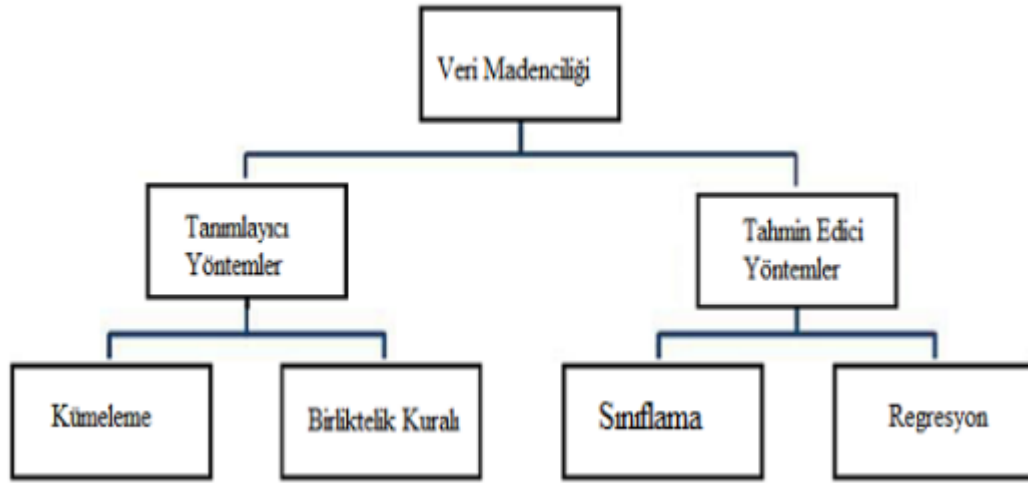
Şekil 4'ten görüleceği üzere, veri madenciliği sürecinde büyük veri içerisinde önce ham veri elde edilmektedir. Bu aşamaya veri seçim aşaması adı verilmektedir. Ardından Ön işleme aşamasına geçilmekte ve ön işlenmiş veri ortaya çıkarılmaktadır. Sonrasında dönüştürme aşaması başlamakta ve dönüştürülmüş veri çıktısı elde edilmektedir. Dönüştürülmüş ve üzerinde veri madenciliği aşamasına geçilmekte ve örüntülerin elde edilmesi sağlanmaktadır. Örüntülerin değerlendirilmesi ile de bilgiye ulaşılmaktadır (Uzun ve ark., 2021: 2).

Veri madenciliği işlemi gerçekleştirilirken yapılan faaliyetlerde, tündengelim yerine tümevarım bakış açısı söz konusudur. Eldeki büyük ölçekli veri ile genellemelerin yapılması esası söz konusudur. Veriler arasındaki ilişkilere odaklanılarak bilgiye ilişkin kurallar belirlenmektedir (Şeker Elmas, 2019: 15).

2.3.2.1. Veri Madenciliği Teknikleri

Veri madenciliği tekniklerinin alan yazınında, “tanımlayıcı” ve “tahmin edici” şeklinde iki farklı gruba ayrıldığı görülmektedir. Aslında modellerin adlandırılmasının temelinde veri madenciliğinin karar verme süreçlerinde etkilerinden yola çıkılmıştır. Örneğin tamamlayıcı teknikler ile hangi verilerin karar verirken kullanılacağına karar verme imkanı oluşmaktadır. Tahmin edici tekniklerde ise çoğunlukla keşif anlayışı hakimdir. Sonuçları bilinen veriler sayesinde, sonuçları bilinmeyen verilerden elde

edilmesi olası olan bilginin ne olduğu tahmin edilmeye çalışılmaktadır (Karacan ve Yeşilbudak, 2010: 18).



Şekil 5. Veri Madenciliği Modelleri

Kaynak: Şeker Elmas, 2019: 17

Yukarıdaki şekilde de görüleceği üzere, veri madenciliğini gerçekleştirmek için kullanılan iki temel model vardır. Bunlar da kendi aralarında Kümeleme, Birliktelik Kuralı, Sınıflama ve Regresyon olarak alt modellere sahiptir. Burada bu modelleri kısaca açıklamak yerinde olacaktır.

i. Kümeleme

Adından da anlaşılacağı üzere, birbirine benzerlik gösteren verilerin kümelendirilmesi işlemidir. Kümelendirme sayesinde verilerin benzer yanları ve birbirlerinden ayrılan özellikleri arasında sınırlar belirlenebilmektedir. Kümelendirme ile oluşturulmuş olan veri kümelerinin her birinin bir anlamı mevcuttur. Kümeleme algoritmaları küme içerisindeki benzer kayıtların maksimize edildiği ve küme dışındaki benzer kayıtların minimize edildiği, ilişkili homojen alt gruplar ya da kümelerin tüm veri setindeki parçalarını araştırmaktadır (Karaöz, 2018: 31). Kümeleme işleminde, kümelerin hangi özellikleri taşıyan verileri içereceği önceden belirlenmediği için, bir küme adlandırması söz konusu değildir. Bu durum kümelendirmeyi sınıflandırmadan ayıran en belirgin özelliktir. Bir başka önemli fark da eğitimidir. Kümelendirme işlemi esnasında veriler

modelin oluşturulması için eğitilmemektedir. Buradan hareketle, kümeleme tekniğinin diğer tekniklere göre iki büyük avantajının olduğunu söylemek yerinde olacaktır. Birinci avantaj kümelemenin sınıflama tekniğince desteklenen yapıyı gösterebilme yeteneğine sahip olmasıdır. Diğeri de test kümesinin, sınıflama doğruluk skorları oluşturulan kümelerin kalitesine ilişkin ek bilgi sağlamasıdır (Şeker Elmas, 2019: 17;

ii. Birliktelik Kuralı

Sıklıkla yan yana görülen veya birlikte hareket ettiği tespit edilen verilerin ilişkilerinin belirlenmesi temeline dayanan bir modeldir. Veriler arasında var olan ilişkilerin keşfedilmesi ile bir veriden ya da onun oluşturduğu bilgiden yola çıkarak, benzer ya da birlikte olan verinin de ortaya çıkaracağı bilgiye erişmeye odaklanılmaktadır. Özellikle pazarlama alanında bir müşterinin davranışları arasındaki ilişkileri inceleyerek ona yönelik pazarlama faaliyetlerinin yürütülebilmesi için bu teknikten fayda sağlanabilecektir (Özçakır ve Çamurcu, 2007: 23). Bu kural müşterileri tanıma ve anlama alanında çokça fayda sağladığından özellikle Pazar sepeti analizlerinde kendisinden faydalanılmaktadır. Örneğin bir müşterinin satın aldığı bir ürünle birlikte aldığı bir başka ürün arasında ilişki kurulabildiğinde, bir başka müşteriye de ilişkili ürünlerin pazarlaması yapılabilmektedir (Alanlar, 2021: 46).

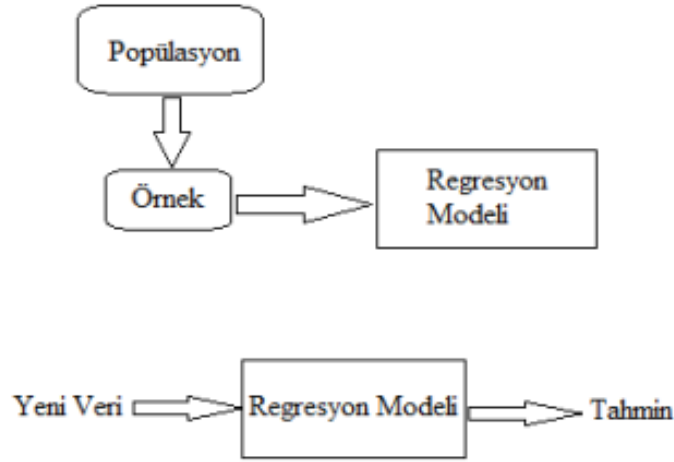
iii. Sınıflama

Sınıflama yöntemi de yine veri madenciliğinin popüler metotlarından biridir. Bunda da adından anlaşılacağı üzere veriler belirli sınıflara ayrılmaktadırlar. Dağınık biçimde bulunan veriler, belirli özelliklerine göre, önceden yapılmış olan sınıfların altına dahil edilirler. Sınıflandırmanın etkin biçimde yapılabilmesi için veri tabanının belli bir bölümü öncelikle eğitilmektedir. Yani verilerin neye göre hangi sınıfa gideceğine ilişkin bir makine öğrenmesi gerçekleşmektedir. Bu öğrenim sınıflandırmanın temel kurallarının belirlenmesini sağlamaktadır. O halde sınıflandırma tekniğinin bir öngörü tekniği olduğunu söylemek de yanlış olmayacaktır. Sınıflama tekniğinde iki farklı algoritmanın kullanılması

mümkündür. Bunlar; karar değişkeni ile sınıflama ve örnek ile sınıflamadır (Karaöz, 2018: 17).

iv. Regresyon

Regresyon analizi süreklilik gösteren değerlerin tahmin edilmesinde kullanılmaktadır. Burada bağımlı bir değişken; bir ya da daha fazla bağımsız değişkenle ilişkiye girmektedir. Bu ilişki bir fonksiyona dönüştürülmek istenmektedir. Yine verilerde yer alan değişkenlerin birbirleri ile ilişkisinin anlaşılması amacına hizmet etmektedir. İstatistiksel bir uygulamadır. Kimi koşullarda keşfedilen modelin doğrusal olmayacağını da eklemek gerekir. Doğrusal regresyon iki veya daha fazla değişken için doğrusal ilişki ve doğrusal veri modelini belirlemek için kullanılan istatistiksel yöntemlere aittir. Basit bir doğrusal regresyon, doğrusal ilişki ve iki genel rastgele değişken arasında doğrusal bir model saptayan en yaygın tekniktir (Çelik 2015: 13).



Şekil 6. Regresyon Modeli

Kaynak: Şeker Elmas, 2019: 19

2.3.2.2. Veri Ambarı

Veri ambarı, sorgulama ve analiz için kullanılmak üzere tasarlanmış ilişkisel bir veri tabanına verilen isimdir. Veri ambarının temel hedefi, çok büyük verilerin analiz edilebilmesi için uygun koşulları oluşturmaktır. Veri ambarında pek çok kaynaktan

ulaşan veriler bir araya getirilmektedir. Bunlar doğrudan ve oldukça hızlı olarak analiz edilmektedir. Veri ambarları ayrıca birbiriyle ilişkisi olmayan veri kaynaklarından aldığı verileri birleştirip bunları karar destek uygulamalarında kullanılmak üzere oluşturulan çok boyutlu gösterim işidir. Geleneksel veri tabanı sistemleri, kullanıcı hareketlerine bağlı günlük işlemleri desteklemek için tasarlanmıştır ve bu sistemler işletimsel ya da hareketsel sistemler olarak adlandırılır. İşletimsel sistemler hareket ya da işlem yönlendirmeli, veri ambarları ise konu yönlendirmelidir.

2.3.2.3. Veri Madenciliği ve Sınıflandırma

Veri madenciliği, bilgisayar sistemine bir öğretme sürecini ifade etmektedir. Önemli olan karmaşık ve çok büyük sayıdaki verinin içinden uygun koşulları taşıyanları ayıklayarak bunların kullanılabilir hale gelmesini sağlamaktır. Veri madenciliği, karar verilmesi gereken durumlarda kümeleme ve birliktelik kuralı gibi yöntemleri kullanırken, tahmin etmeye dayalı durumlarda sınıflandırma ve veri kümesindeki değişkenler arasındaki ilişkileri inceleyen regresyon üzerine yoğunlaşmaktadır. Ancak hangi yoldan gidilecek olunursa olunsun, öğrenme sürecinde öncelikle model oluşturulmalıdır. Model oluşturulmasıyla birlikte, makineye aslında hangi yöntem ile öğrenmesi gerektiği bilgisi verilmiş olmaktadır. Farklı öğrenme modellerinin varlığı doğru veri seti için doğru makine öğrenmesinin sağlanmasını mümkün kılmaktadır. Makine öğrenmesinde kullanılan öğrenme modelleri denetimli ve denetimsiz adı altında iki temel gruba ayrılmaktadır. Denetimli öğrenme söz konusuysa, girişleri ve çıkışları içeren bir veri kümesi algoritmaya girdi olarak verilmektedir. Algoritma ise bu girişlere ve çıkışlara nasıl ulaşılacağını belirleyen bir yöntemi bulmaktadır. Denetimsiz öğrenmede ise var olan girdilere ait bir çıktı (etiket) bulunmadığından (sınıflar önceden belli olmadığından) algoritma mevcut verileri analiz ederek ilişkileri kendisi belirlemektedir (Çelik ve ark., 2021: 884). Oluşturulan model, verilerin sınıflandırılması için de önemli bir adımdır. Aşağıda bu çalışmanın uygulama aşamasında da kullanılan sınıflandırma türlerinin neler olduğu açıklanmıştır:

i. KNN (K-En Yakın Komşu Algoritması - K-Nearest Neighbor Algorithm)

En yakın komşu (NN), hem sınıflandırma hem de regresyon için kullanılan denetimli bir öğrenme algoritmasıdır. En yakın komşu algoritmasının arkasındaki

temel fikir, yeni bir veri noktası ile en yakın komşularının tümü arasındaki mesafeyi ölçmektir. En yakın komşuların etiketlerine dayanarak, yeni veri noktasının etiketi tahmin edilebilir. Naif bir yöntemdir çünkü ayrımcı bir model oluşturmaz; bunun yerine, bir veri noktasını sınıflandırmak için yerel benzerliğe dayanır. En yakın komşu algoritması, bilinmeyen bir nesneyi bilinen en benzer nesneye olan benzerliğine göre sınıflandırır. Bu, önce yeni nesne ile tüm eğitim nesneleri arasındaki mesafeleri hesaplayarak ve ardından mesafeler listesinden yeni nesneye en çok benzeyen eğitim nesnesini seçerek yapılır. En Yakın Komşu algoritmaları genellikle öneri sistemleri, görüntü tanıma ve tıbbi teşhis için kullanılır (Nizam ve Akın, 2019: 3).

ii. Naive Bayes

Naive Bayes, sınıflandırma problemleri için kullanılan bir makine öğrenme algoritmasıdır. Bayes Teoremine dayanır ve tahminlerde bulunmak için kullandığı özellikler arasında bağımsızlık olduğunu varsayar. Naive Bayes denetimli bir algoritmadır ve genellikle metin sınıflandırmasında, tıbbi teşhiste ve spam filtrelemede kullanılır. Hızlı bir algoritmadır ve daha gelişmiş tekniklerin aksine saymaya dayanır. Uygulaması da basittir ve doğru tahminler yapmak için yalnızca az miktarda eğitim verisi gerektiren küçük veri kümelerini işleyebilir (Şahinaslan ve ark., 2022: 222).

iii. Destek Vektör Makinesi

Destek Vektör Makinesi (DVM), hem sınıflandırma hem de regresyon görevleri için kullanılabilen bir tür denetimli makine öğrenimi algoritmasıdır. Bilinmeyen veri noktalarını, önce tam yapılarını belirlemek zorunda kalmadan otomatik olarak sınıflandırabilen bir tür denetimli öğrenme algoritmasıdır. Noktaları sınıflandırmak için DVM, veri kümelerini ayıran bir "hiper düzlem" oluşturur. Daha sonra bu iki kümeyi, bilinmeyen noktaların sınıfını veya özelliğini doğru bir şekilde tahmin edebilen bir model oluşturmak için kullanır. DVM, diğer sınıflandırma algoritmalarından farklıdır, çünkü yalnızca kümeler arasındaki sınırlara bakmaz, aynı zamanda kümeler arasındaki mesafeyi en üst düzeye çıkarması muhtemel olan en önemli sınırları da tanımlar. DVM ayrıca doğrusal

olmama sorununu da çözebilir ve çok sınıflı sorunları lojistik regresyondan daha iyi çözebilir (Kasaba ve Yıldıztepe, 2016: 2).

iv. Random Forest (Rastgele Orman)

Rastgele orman, sınıflandırma ve regresyon görevleri için kullanılan denetimli bir makine öğrenimi algoritmasıdır. Belirli bir veri noktasının sonucunu tahmin etmek için birden fazla karar ağacı kullanan bir topluluk algoritmasıdır. Rastgele ormanlar, özellikleri ve gözlemleri rastgele seçerek ve ardından bu özelliklere dayalı bir sınıflandırıcı ağaç oluşturarak karar ağaçları oluşturur. Birden fazla karar ağacının sonuçlarını toplayan rastgele ormanlar, tek bir karar ağacından daha doğru ve kararlıdır. Aykırı değerlere ve verilerdeki gürültüye karşı daha dayanıklıdır, çünkü bireysel karar ağaçlarında büyük hatalar olabilir, ancak genel rastgele orman tahmini yine de doğrudur. Rastgele ormanlar, görüntü tanıma, tıbbi teşhis ve dolandırıcılık tespiti dahil olmak üzere çeşitli uygulamalarda kullanılır (Nizam ve Akın, 2019: 3)

3. YÖNTEM

3.1. Duygu Analizi Adımları

Çalışma kapsamında öncelikle internette kullanılan etiketlerin neler olduğunun tespit edilmesine çalışılmıştır. Etiketler kullanılarak atılan tweetlerin dağılımları gerçekleştirilmiştir. Bu tweetler ayrıca nötr, negatif ve pozitif duygulara göre ayrılmış ve program tarafından verilen verilerin sayı değerlerinin uygunluğu kontrol edilmiştir.

```
In [1]: import pandas as pd
import numpy as np
from warnings import filterwarnings
filterwarnings('ignore')
```

```
In [2]: df1=pd.read_excel('turkce_tweet.xlsx')
df2=pd.read_csv('turkce_duygu.csv', encoding = "utf-16")
df1.head() # 0 nötr 1 negatif 2 pozitif
```

Out[2]:

	Tweets	Duygu	Unnamed: 2
0	sayın cumhurbaşkanım turizm için ertelendiğin...	1	NaN
1	bir kere ya bir kere gencinizi dinleyin biz y...	1	NaN
2	o kadar çok mağduruz ki neye üzüleceğime şaşır...	1	NaN
3	bu sınav tarihini değiştirirken pedaloglara v...	1	NaN
4	turizm yılı gençlik geleceğinizi kurtarır	2	NaN

```
In [3]: df1.drop('Unnamed: 2',axis=1, inplace=True)
df1.head()
```

Out[3]:

	Tweets	Duygu
0	sayın cumhurbaşkanım turizm için ertelendiğin...	1
1	bir kere ya bir kere gencinizi dinleyin biz y...	1
2	o kadar çok mağduruz ki neye üzüleceğime şaşır...	1
3	bu sınav tarihini değiştirirken pedaloglara v...	1
4	turizm yılı gençlik geleceğinizi kurtarır	2

Tablo 1. Kod Anlatımı

Pandas ve numpy kütüphanelerinin tanımlanması, veri setlerinin okunması ve 1. Veri setinin ekranda gösterimi sağlanmış; veri setindeki fazla sınıfın drop ile silinmesi işlemleri gerçekleştirilmiştir.

```
In [4]: df2.head()
```

```
Out[4]:
```

	Tweets	Duygu
0	ses kalitesi ve ergonomisi rezalet, sony oldu...	1
1	hızlı teslimat teşekkürler	0
2	ses olayı süper...gece çalıştır sıkıntı yok....	2
3	geldi bugün kullandık hemen bozıldı hiçtavsıye...	1
4	Kulaklığın sesi kaliteli falan değil. Aleti öv...	1

```
In [5]: df=pd.concat([df1, df2])
```

Veri Ön İşleme Süreci

```
In [6]: df['Tweets'] = df['Tweets'].astype('str')
df.head()
```

```
Out[6]:
```

	Tweets	Duygu
0	sayın cumhurbaşkanım turizm için ertelendiğin...	1
1	bir kere ya bir kere gencinizi dinleyin biz y...	1
2	o kadar çok mağdumuz ki neye üzüleceğime şaşır...	1
3	bu sınav tarihini değiştirirken pedaloglara v...	1
4	turizm yılı gençlik geleceğinizi kurtarır	2

Tablo 2. Veri Ön İşleme Süreci

Çalışmada elde edilen veri sayı değerlerinin uygunluğunun kontrol edilmesinin ardından, eksik veriler için analiz gerçekleştirilmiştir. Veri setinin ekranda gösterimi, Concat ile iki veri setinin birleştirilmesi ve veri setindeki tweets sütununun string yapılması işlemleri sonrasında, veri setleri birleştirilmiştir.

Veri temizleme işlemi

```
In [7]: from nltk.corpus import stopwords
from textblob import Word
sw=stopwords.words("turkish") # anlamsız kelimelerin tanımı
gereksizler= "@\s+|https?:\s+|http?:\s+|[\^a-zA-Z0-9ğüşöçİİĞÜŞÖÇ]+"
```

```
In [8]: def data_cleaning(df,choice):
df = df.applymap(lambda s: s.lower() if type(s) == str else s) # Tüm metinler küçük harf formatına dönüştürülür
df["Tweets"]=df["Tweets"].str.replace("[^\w\s]", "") # noktalama işaretlerinin silinmesi
df["Tweets"]=df["Tweets"].str.replace("\d", "") # sayıların silinmesi
df["Tweets"]=df["Tweets"].apply(lambda x: " ".join(x for x in x.split() if x not in gereksizler)) # tanımlananlar silinir
df["Tweets"]=df["Tweets"].apply(lambda x: " ".join(x for x in x.split() if x not in sw)) # stopwords kelimelerin silinmesi
if choice == 1:
df["Tweets"]=df["Tweets"].apply(lambda x: " ".join(Word(i).lemmatize() for i in x.split())) # Lemmatization kitaplık
elif choice == 2:
df["Tweets"]=df["Tweets"].apply(lambda x: " ".join(Word(i).stem() for i in x.split())) # stemming kitap
return df['Tweets']
```

```
In [9]: df['Tweets']=data_cleaning(df, 1)
```

Tablo 3. Veri Temizleme İşlemi

Çalışmada kullanılacak gerçek verilerin elde edilebilmesi için gereksiz verilerin temizlenmesi gerekliliği bulunmaktadır. Veri temizleme işlemi için gerekli kütüphaneler tanımlanmıştır. Sanal ortamda genellikle insanların yazım dili kurallarına uygunsuz

içerikler üretmesi bu temizleme işleminin ana sebebidir. Bu temizleme işlemi esnasında öncelikle veri silinmesi için gerekli fonksiyon açılmıştır. Ardından, aranması istenen tüm kelimeler küçük harfe dönüştürülmüştür. Noktalama işaretleri ve sayılar silinmiştir. Bunlara ek olarak tanımlanan gereksiz karakterler de silinmiştir. Bu işlemler aracılığı ile Türkçede anlamsız kelimelerin ayıklanması için kütüphane oluşturulmuş ve bunlar değişkene atanmıştır. Veri temizleme işlemi esnasında yukarıda sayılan ve analiz esnasında kullanılmasını istemediğimiz gereksiz karakterlerin de değişkene atanması da yapılmıştır. Stopwords kelimelerin silinmesi ve Lematizasyon veya steaming seçilmesi gerçekleştirilmiştir.

Özellik seçimi

```
In [10]: from sklearn.feature_extraction.text import TfidfVectorizer

In [11]: def ozellik_secimi(df,max_features):
          print("TF-IDF yöntemi seçildi")
          vectorizer = TfidfVectorizer(lowercase=False,ngram_range=(1,2),dtype=np.byte,max_features=1000)
          X = vectorizer.fit_transform(df['Tweets'].tolist())
          return X,vectorizer

In [12]: X,vectorizer=ozellik_secimi(df,None)
TF-IDF yöntemi seçildi
```

Tablo 4. Özellik Seçimi

Oluşturulan veri setine gerekli ön işleme süreçleri yaptım. Bunlar küçük harf, anlamsız kelimelerin çıkarılması, imojilerin noktalama işaretlerinin silinmesi gibi adımları devam ettirdim. Özellik seçimi konusunda “Tfidf” yöntemi uygulanmıştır. Tfidf kütüphanesinin tanımlanmasının ardından, özellik seçimi için veri seti ve “max features” seçilme işlemi yapılmıştır. Tf idf için n gram 1,2 belirlenmiştir. Veri seti için tweets sütununu liste yapılmış ve vektörü alınmıştır. Ardından, oluşturulan fonksiyon kullanımı gerçekleştirilmiştir.

Eğitim ve Test

```
In [13]: from sklearn.model_selection import train_test_split

In [14]: X_train,X_test,y_train,y_test = train_test_split(X,df['Duygu'],random_state=1,test_size=0.25)

In [15]: from sklearn.metrics import accuracy_score, classification_report,confusion_matrix
```

Tablo 5. Eğitim ve Test

Veri setinin eğitim ve test olarak ayırmak için “train test split” kütüphanesi eklenmiştir. Kütüphane kullanılarak eldeki veriler %75 eğitim %25 test olarak ayrılmıştır. Kelime sayısı 1000 olarak belirlenmiştir.

Naive Bayes

```
In [16]: from sklearn import naive_bayes
nb = naive_bayes.MultinomialNB()
nb_model = nb.fit(X_train,y_train)
prediction_nb=nb_model.predict(X_test)
print("Eğitim verisi doğruluk skoru: ",nb.score(X_train,y_train))
print("test verisi doğruluk skoru: ",nb.score(X_test,y_test))
print(confusion_matrix(y_test, prediction_nb))
print(classification_report(y_test, prediction_nb))
print("Accuracy Score: ",accuracy_score(y_test, prediction_nb))
```

```
Eğitim verisi doğruluk skoru: 0.6738036338821121
test verisi doğruluk skoru: 0.6368986693961105
[[ 340  387  305]
 [ 158 1106  173]
 [ 133  263 1043]]
      precision    recall  f1-score   support

0         0.54        0.33        0.41        1032
1         0.63        0.77        0.69        1437
2         0.69        0.72        0.70        1439

 accuracy          0.64        3908
 macro avg         0.62        0.61        0.60        3908
weighted avg         0.63        0.64        0.62        3908

Accuracy Score: 0.6368986693961105
```

Tablo 6. Naive Bayes

Naive Bayes, KNN, Random forest, Destek vektör makinaları gibi algoritmalarda veriler test edilmiştir. Naive Bayes kütüphanesi çağrılarak model tanımlaması yapılmıştır. Eğitim verisi fit edildikten sonra test verisi tahmin edilmiş ve eğitim ve test veri setlerinin doğruluk skorları ekrana gösterilmiştir. Confusion matrisi ekrana verilmiş; Classification report kullanarak test ve tahminlerin özelliklerini ekrana gösterilmiştir. Ardından doğruluk skoruna ulaşılmıştır. Bu işlemlerden sonra doğruluk skorunun % 63 olduğu görülmüştür.

Random Forest

```
In [17]: from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf_model = rf.fit(X_train,y_train)
prediction_rf=rf_model.predict(X_test)
print("Eğitim verisi doğruluk skoru: ",rf.score(X_train,y_train))
print("test verisi doğruluk skoru: ",rf.score(X_test,y_test))
print(confusion_matrix(y_test, prediction_rf))
print(classification_report(y_test, prediction_rf))
print("Accuracy Score: ",accuracy_score(y_test, prediction_rf))

Eğitim verisi doğruluk skoru: 0.9648554124370895
test verisi doğruluk skoru: 0.6371545547594678
[[ 306  335  391]
 [ 135 1083  219]
 [ 147  191 1101]]
      precision    recall  f1-score   support

0         0.52      0.30      0.38      1032
1         0.67      0.75      0.71      1437
2         0.64      0.77      0.70      1439

 accuracy          0.64      3908
macro avg          0.61      0.61      0.60      3908
weighted avg          0.62      0.64      0.62      3908

Accuracy Score: 0.6371545547594678
```

Tablo 7. Random Forest

Bir sonraki adımda Random Forest kütüphanesi çağrılarak model tanımlaması yapılmıştır. Eğitim verisi fit edildikten sonra test verisi tahmin edilmiş ve eğitim ve test veri setlerinin doğruluk skorları ekrana gösterilmiştir. Confusion matrisi ekrana verilmiş; Classification report kullanarak test ve tahminlerin özelliklerini ekrana gösterilmiştir. Ardından doğruluk skoruna ulaşılmıştır. Bu işlemlerden sonra doğruluk skorunun burada da % 63 olduğu görülmüştür.

KNN

```
In [18]: from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=1)
knn_model = knn.fit(X_train,y_train)
prediction_knn=knn_model.predict(X_test)
print("Eğitim verisi doğruluk skoru: ",knn.score(X_train,y_train))
print("test verisi doğruluk skoru: ",knn.score(X_test,y_test))
print(confusion_matrix(y_test, prediction_knn))
print(classification_report(y_test, prediction_knn))
print("Accuracy Score: ",accuracy_score(y_test, prediction_knn))

Eğitim verisi doğruluk skoru: 0.9617845261451847
test verisi doğruluk skoru: 0.47824974411463667
[[ 213  201  618]
 [ 158  580  699]
 [ 227  136 1076]]
      precision    recall  f1-score   support

0         0.36      0.21      0.26      1032
1         0.63      0.40      0.49      1437
2         0.45      0.75      0.56      1439

 accuracy          0.48      3908
macro avg          0.48      0.45      0.44      3908
weighted avg          0.49      0.48      0.46      3908

Accuracy Score: 0.47824974411463667
```

Tablo 8. KNN

Yeni aşamada KNN kütüphanesi çağrılarak model tanımlaması yapılmıştır. Eğitim verisi fit edildikten sonra test verisi tahmin edilmiş ve eğitim ve test veri setlerinin doğruluk skorları ekrana gösterilmiştir. Confusion matrisi ekrana verilmiş; Classification report kullanarak test ve tahminlerin özelliklerini ekrana gösterilmiştir. Ardından doğruluk skoruna ulaşılmıştır. Bu işlemlerden sonra doğruluk skorunun burada yüzde 47 olduğu görülmüştür.

Destek Vektör Makinası

```
In [19]: from sklearn import svm
sv = svm.SVC()
sv_model = sv.fit(X_train,y_train)
prediction_sv=sv_model.predict(X_test)
print("Eğitim verisi doğruluk skoru: ",sv.score(X_train,y_train))
print("test verisi doğruluk skoru: ",sv.score(X_test,y_test))
print(confusion_matrix(y_test, prediction_sv))
print(classification_report(y_test, prediction_sv))
print("Accuracy Score: ",accuracy_score(y_test, prediction_sv))

Eğitim verisi doğruluk skoru: 0.8737524524439136
test verisi doğruluk skoru: 0.6583930399181167
[[ 382  346  304]
 [ 166 1094  177]
 [ 162  180 1097]]
      precision    recall  f1-score   support

     0       0.54      0.37      0.44      1032
     1       0.68      0.76      0.72      1437
     2       0.70      0.76      0.73      1439

 accuracy          0.64          0.63          0.66      3908
 macro avg         0.64          0.63          0.63      3908
 weighted avg      0.65          0.66          0.65      3908

Accuracy Score: 0.6583930399181167
```

Tablo 9. Destek Vektör Makinesi (SVM)

Ardından, SVM kütüphanesi çağrılarak model tanımlaması yapılmıştır. Eğitim verisi fit edildikten sonra test verisi tahmin edilmiş ve eğitim ve test veri setlerinin doğruluk skorları ekrana gösterilmiştir. Confusion matrisi ekrana verilmiş; Classification report kullanarak test ve tahminlerin özelliklerini ekrana gösterilmiştir. Ardından doğruluk skoruna ulaşılmıştır. Bu işlemlerden sonra doğruluk skorunun burada yüzde 65 olduğu görülmüştür.

Twitter API ile veri çekimi

```
In [20]: import tweepy, codecs

In [21]: def hashtag_df(key, count):
    consumer_key = 'bIQ3GqTgdNIS9qtYATf1f2kw3'
    consumer_secret = 'HuF1e9E6CRT1QmvOm130s3cUnX3FxSbSzkweMM6nztddGGprU26'
    access_token = '100390695-WeMfH504k2zeXyozN3EAqpM93Qxjts6sff3BKjkh'
    access_token_secret = 'VFGMaqkkRbtUhnmmuYr55MzTtwgOPTNuyB50vh00gbIss'
    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_token_secret)
    API = tweepy.API(auth)
    tweetler=API.search_tweets(q=key, lang="tr", result_type="recent", count=count)
    df2 = pd.DataFrame()
    df2["Tweets"] = [tweet.text for tweet in tweetler]
    return df2
```

Tablo 10. API Veri Çekimi

Twitter API sayesinde belirlenen etikete göre en son atılan anlık olarak veriler çekilmiştir. Tweepy ve codecs kütüphanelerinin tanımlanmasının sonrasında Fonksiyon tanımlaması gerçekleştirilmiş; API için gerekli olan key, secret, token ve token secret şifreleri girilmiştir. Auth için oluşturulan şifrelerin de tanımlanmasının sonrasında, araştırma için çekilecek olan tweetler ile ilgili kelime, dil, türü ve adet belirlenmiştir.

```
In [22]: key = input('Input a key:')
Input a key:yemeksepeti

In [24]: df2 = hashtag_df(key, count)

In [25]: from datetime import datetime
df2.head(5)
an = datetime.now()
print(an)
2022-12-12 06:14:21.889647

In [26]: df2.Tweets[1]
Out[26]: "RT @yemeksepeti: Yemeksepeti'nden ilk siparişine 50 TL indirim!"

In [27]: df3 = df2.copy()
df2.Tweets = data_cleaning(df2,1)
df2.Tweets = df2.Tweets.str.replace("rt", "")
df2

Out[27]:
```

	Tweets
0	dan basladi sizin ben reklaminizi ayri morunuz ...
1	yemeksepeti yemeksepetinden ilk siparişine tl...
2	yemeksepeti a_llolitta dolandırıcı arkadaşlar ...
3	ensuperlig yemeksepeti trabzonspor
4	ensuperlig yemeksepeti trabzonspor
...	...
95	protalitygg hivex_esc büyük finallere gümbür ...
96	borracar basketmerkez yemeksepeti var mesela
97	ensuperlig yemeksepeti için sefiri trabzonspor
98	protalitygg xgrifin yıldızlaştığı maçta hive...

Tablo 11. Aranacak Kelime ve Sayının Girişi

Burada API'nın standart olmasından dolayı yalnızca 3 çeşit olmak üzere ve toplam 100 tweet çekilebilmiştir. Sonrasında çekilen tweetler dataframe olarak atanmış ve her bir tweet bu dataframe üzerine eklenmiştir.

```
In [28]: test_data = vectorizer.transform(df2.Tweets)

In [29]: nb=nb_model.predict(test_data)
print('naive bayes tahminleri:',nb_model.predict(test_data))
knn=knn_model.predict(test_data)
print('knn tahminleri:',knn_model.predict(test_data))
rf=rf_model.predict(test_data)
print('random forest tahminleri:',rf_model.predict(test_data))
sv=sv_model.predict(test_data)
print('destek vektör makinaları tahminleri',sv_model.predict(test_data))

Logistic Regression tahminleri: [1 1 1 1 1 1 1 1 2 1 1 1 0 1 1 1 1 2 1 1 1 1 1 1 0 2 1 2 1 2 1 2 1 1 1 1 1
1 1 1 1 1 2 2 1 1 1 2 1 2 1 2 2 2 1 1 1 2 1 1 2 1 1 2 2 1 2 1 1 2 2 1 2 1 1 1 1 1 1 2
2 2 1 1 1 1 1 1 1 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1]
yapay sinir ağı tahminleri: [2 1 1 2 2 2 1 2 0 1 0 2 0 2 1 2 2 0 2 2 2 2 2 1 0 2 1 2 2 1 2 2 2 2 2 2
2 2 2 2 2 1 2 1 2 0 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 1 2 1 1 2 2 1
1 2 2 1 0 2 2 1 1 2 2 2 2 1 2 2 1 1 2 1 0 1 1 2 1 2]
Naive bayes tahminleri: [1 1 1 1 2 2 2 1 2 1 1 2 1 0 1 1 2 0 0 2 2 1 2 0 1 2 1 2 2 2 2 2 2 2 0 2 1
2 1 2 2 2 2 2 1 2 0 2 2 2 1 2 2 2 2 1 1 1 2 1 1 2 2 2 1 2 0 1 1 2 2 2 2 1
1 1 2 1 1 2 2 1 2 2 2 2 2 1 2 2 1 1 2 1 1 2 0 2 0 0]
Karar ağaçları tahminleri [0 1 1 2 2 2 1 2 1 1 1 1 0 1 0 2 1 2 1 1 1 2 1 0 2 1 2 2 1 2 1 2 2 2 1 1 1
2 1 2 2 2 1 2 1 2 1 2 1 2 1 2 1 1 2 0 0 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 2 2 1
2 2 1 1 0 2 2 1 1 1 2 2 2 1 2 2 1 1 2 1 1 0 1 2 1 1]
```

Tablo 12. Dataframe

En son atılan tweetlerin çekilmesinden dolayı sisteme tweetlerin çekilme tarih ve zamanı da girilmiştir. Ardından çekilen tweetlerin kopyası oluşturulmuş ve bunlar ön işleme sürecine sokulmuştur. Tweetlerden daha doğru duygu analizi gerçekleştirebilmek için önlerinde yer alan “rt” yazıları silinmiş ve veri seti yeniden ekrana yansıtılmıştır. Oluşturulan veri setinin ön işleme sürecinin bitiminde bunların vektörleri alınmıştır. Bütün modeller ile test veri seti tahminlenmiştir.

```

In [31]: import matplotlib.pyplot as plt

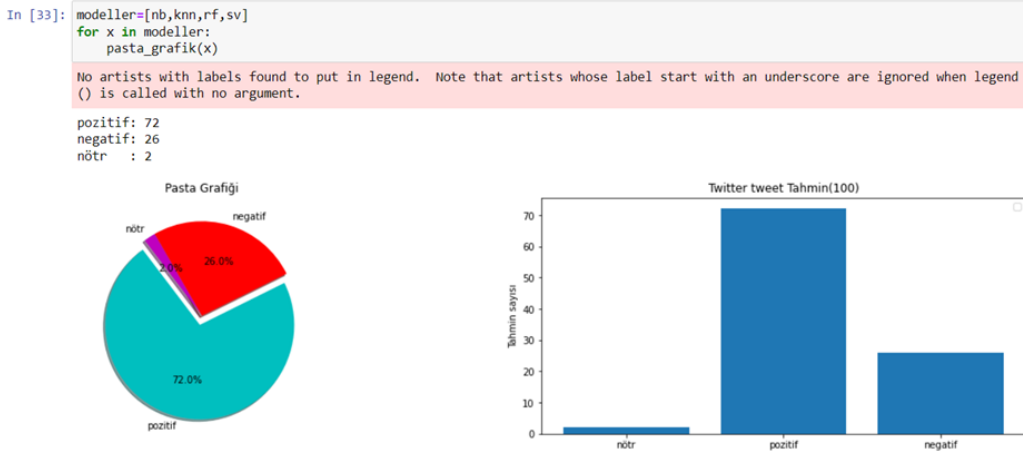
In [32]: def pasta_grafik(model_adi):
    pozitif=0
    negatif=0
    nötr=0
    for x in range(len(model_adi)):
        if model_adi[x]== 1:
            pozitif+=1
        if model_adi[x]==2:
            negatif+=1
        if model_adi[x]==0:
            nötr+=1
    print("pozitif:",pozitif)
    print("negatif:",negatif)
    print("nötr : ",nötr)
    plt.figure(figsize=(20,10))
    dilimler=[nötr,pozitif,negatif]
    isimler=["nötr","pozitif","negatif"]
    renkler=["m","c","r"]
    plt.subplot(2,2,1)
    plt.pie(dilimler,
            labels=isimler,
            colors=renkler,
            startangle=120,
            shadow= True,
            explode=(0,0.1,0),
            autopct='%1.1f%%')
    plt.title("Pasta Grafiği")
    plt.subplot(2,2,2)
    plt.bar(isimler,dilimler)
    plt.legend()
    plt.ylabel('Tahmin sayısı')
    plt.title('Twitter tweet Tahmin('+count+')')
    plt.show()

```

Tablo 13. Veri Setinin Görselleştirilmesi

3.2. Bulgular

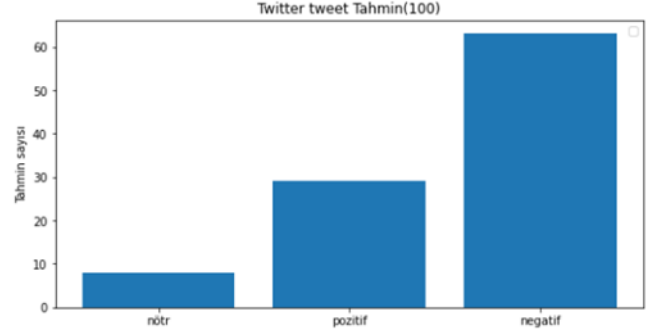
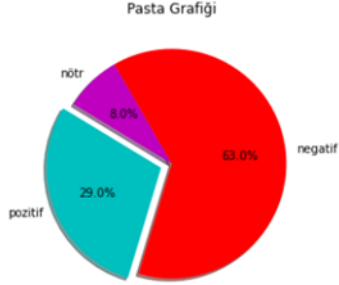
Elde edilen tüm bilginin daha anlaşılabilir olması ve yorumlanabilmesi için verilerin grafik haline getirilmesi sağlanmıştır. Bu aşamada duygu ayrımı yapabilmek ve veri tekrarını önlemek için “for” fonksiyonundan faydalanılmıştır. Grafikler ile ayrıca tweetlerden çekilen, içinde “yemeksepeti” kelimesi geçen tweetlerin pozitif, negatif ve nötr olmak üzere yüzdeler dilimlerinin anlaşılmasına da olanak tanımıştır.



Tablo 14. Naive Bayes Veri Setinin Grafiğe Dönüştürülmesi

Naive Bayes verilerine göre çekilen tweetlerin 72 adeti pozitif, 26'sı negatif ve 2 nötr duygu yansıtmaktadır.

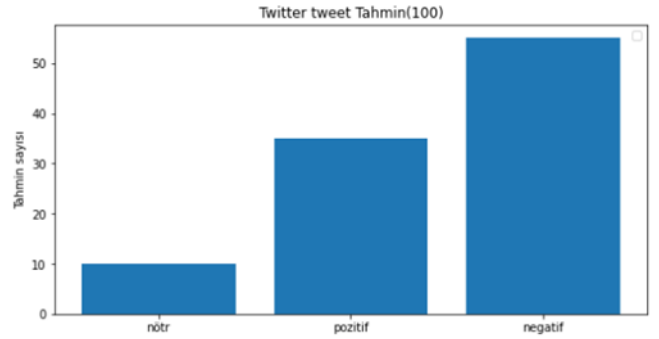
pozitif: 29
negatif: 63
nötr : 8



Tablo 15. KNN Veri Setinin Grafiğe Dönüştürülmesi

KNN verilerine göre çekilen tweetlerin 29 adeti pozitif, 63'ü negatif ve 8'i ise nötr duygu yansıtmaktadır.

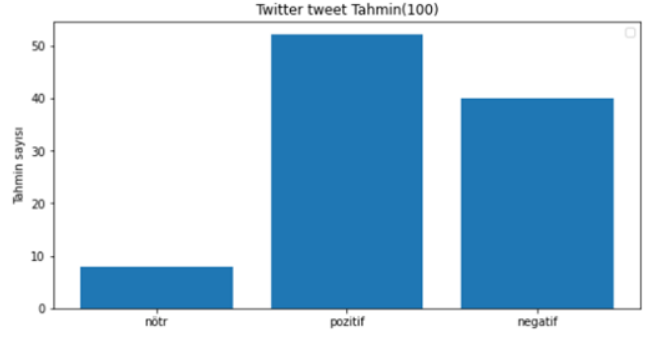
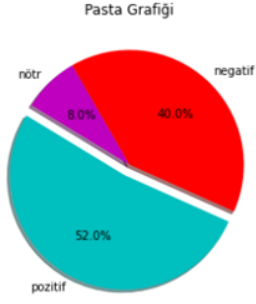
pozitif: 35
negatif: 55
nötr : 10



Tablo 16. Random Forest Veri Setinin Grafiğe Dönüştürülmesi

Random Forest verilerine göre çekilen tweetlerin 35'i pozitif, 55'i negatif ve 10'u nötr duygu yansıtmaktadır.

```
pozitif: 52
negatif: 40
nötr : 8
```



Tablo 17. Destek Vektör Veri Setinin Grafiğe Dönüştürülmesi

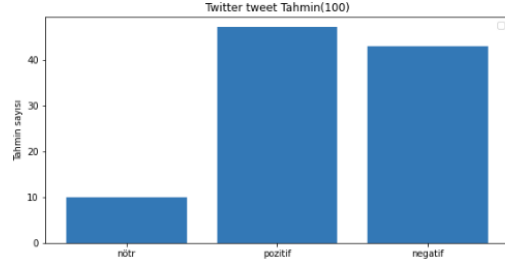
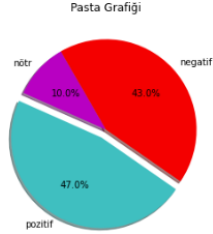
Destek vektör makinası verilerine göre çekilen tweetlerin 52'si pozitif, 40'ı negatif ve 8'i nötr duygu yansıtmaktadır.

```
In [45]: sonucdizisi=[]
for x in range(100):
    y=statistics.mode([nb[x],knn[x],rf[x],sv[x]])
    sonucdizisi.append(y)
```

```
In [104]: pasta_grafik(sonucdizisi)
```

No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend () is called with no argument.

```
pozitif: 47
negatif: 43
nötr : 10
```



```
In [50]: dizi=pd.Series(sonucdizisi)
dizidf=pd.DataFrame(dizi,columns=['Duygu'])
```

```
In [51]: df = pd.concat([df2,dizidf],axis=1)
```

```
In [53]: df ## 0 nötr 1 negatif 2 pozitif
```

Tablo 18. Model Sonuçlarının Karşılaştırılması

Çalışmada her farklı model ile farklı duygu sonuçlarına ulaşıldığı için bunları karşılaştırmak gerekliliği ortaya çıkmıştır. Elde edilen sonuçlar grafiklere dönüştürülmüştür. Ardından diziye göre dataframe atanmıştır ve bunlar birleştirilmiştir.

```
In [63]: X,vectorizer=ozellik_secimi(df,100)
TF-IDF yöntemi seçildi

In [100]: X_train,X_test,y_train,y_test = train_test_split(X,df['Duygu'],random_state=1,test_size=0.30)

In [102]: sv = svm.SVC()
sv_model = sv.fit(X_train,y_train)
prediction_sv=sv_model.predict(X_test)
print("Eğitim verisi doğruluk skoru: ",sv.score(X_train,y_train))
print("test verisi doğruluk skoru: ",sv.score(X_test,y_test))
cv_score=cross_val_score(sv, X, df['Duygu'], cv=5, scoring='accuracy').mean()
print(' Cross Validation Skoru (K=5): {} \n'.format(cv_score))
print(confusion_matrix(y_test, prediction_sv))
print(classification_report(y_test, prediction_sv))
print("Accuracy Score: ",accuracy_score(y_test, prediction_sv))

Eğitim verisi doğruluk skoru: 0.7428571428571429
test verisi doğruluk skoru: 0.6
Cross Validation Skoru (K=5): 0.62

[[ 2  0  0]
 [ 0 14  0]
 [ 0 12  2]]
      precision    recall  f1-score   support

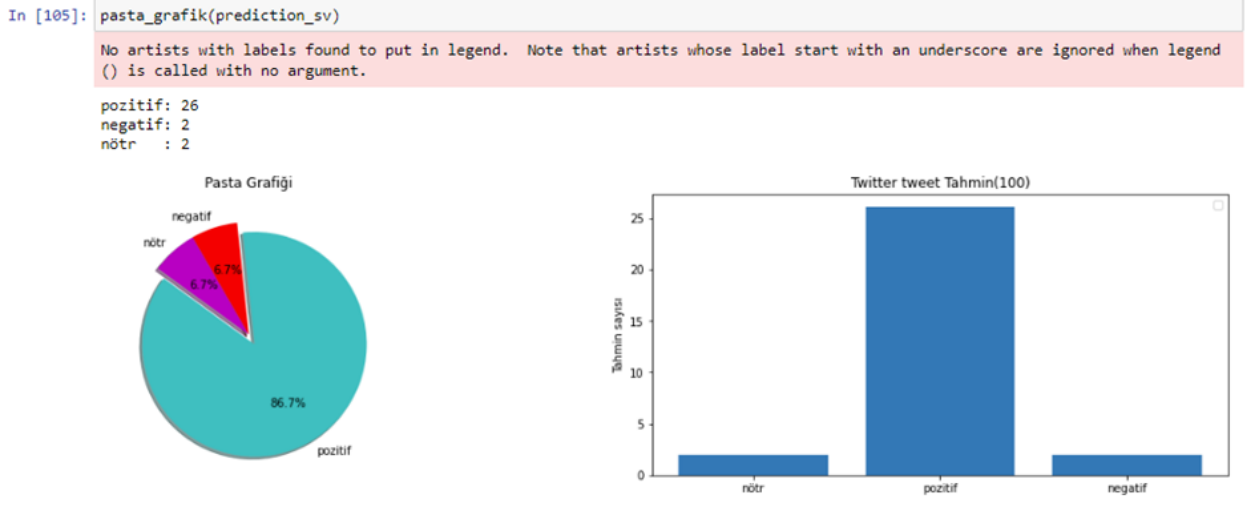
     0       1.00      1.00      1.00         2
     1       0.54      1.00      0.70        14
     2       1.00      0.14      0.25        14

 accuracy          0.60          0.60          0.60         30
 macro avg          0.85          0.71          0.65         30
 weighted avg          0.78          0.60          0.51         30

Accuracy Score: 0.6
```

Tablo 19. Oluşturulan Veri Setinin Vektörlerinin Alınması

Bir sonraki aşamada oluşturulan yeni veri setlerinin tf idf yöntemi ile vektörleri alınmıştır. Veri seti yüzde 30 test, yüzde 70 eğitim olarak bölünmüştür. Destek vektör Makinaları modeli üzerinde yeni oluşturulan tweetler analiz edilmiştir. Cross validation değeri %62; Accuracy skoru %60 olarak tespit edilmiştir.



Tablo 20. Destek Vektör Makinesi Grafiği

Son elde edilen grafikte görüldüğü gibi tüm yöntemlerden elde edilen bilgilerin harmanlanması sonucunda, yemeksepeti için atılan tweetlerin yüzde 86,7'si pozitif duygular içermektedir. Yüzde 6,7 oranla ise negatif duygular içeren tweetler söz konusudur.



4. SONUÇ

Rekabetin sürekli olarak daha yoğun hale geldiği dünyada özellikle teknoloji sayesinde işletmeler bilgi toplayarak müşterilerinin ne düşündüğünü ve ne istediğini anlamaya çalışmaktadırlar. Bu sayede onların beklentilerini daha fazla karşılamaları ve rekabette üstünlük elde etmeleri de mümkün olabilmektedir. Sosyal medya siteleri ve uygulamaları, müşteriler, pazardaki rakipler ve hatta diğer pek çok alanda bilgiye erişmek için etkili kaynaklardır.

Nitekim sosyal medyadan bilgi toplamak güçtür. Burada oldukça fazla insan pek çok konuda fikir beyan etmekte, karmaşık bir kütüphane ortaya çıkmaktadır. Dahası paylaşılan duyguların ifade edilme biçimi belli bir kurala bağlı değildir. O halde işletmelerin doğru veriyi elde ederek bilgiye dönüştürmeleri ve kullanmaları için etkin bir araca ihtiyaçları vardır.

Veri madenciliği yöntemleri sayesinde şirketler, eğilimleri ortaya çıkarmak, korelasyonlar oluşturmak, müşteri duyarlılığını analiz etmek ve müşterilerinin davranışlarını anlamak için sosyal medyayı kullanabilir. Bu tür süreçler, şirketlerin hedef kitleleri belirlemesine, pazarlama stratejileri geliştirmesine, kampanyalarının etkinliğini ölçmesine ve müşteri bağlılığını geliştirmesine yardımcı olabilir. Ek olarak, sosyal medyadaki veri madenciliği, müşteri hizmetlerini iyileştirmek ve şirketlerin daha iyi ürün ve hizmetler geliştirmesine yardımcı olmak için de kullanılabilir.

Veri madenciliği ve makine öğrenmesi sayesinde çeşitli kelimeleri ve hedef grupları seçerek, onların gramer hatası olan yazılarındaki duyguları anlamlandırmak ve buradan hareketle markaya, paylaşıma ya da herhangi bir gündeme ilişkin halkın genel tepkisini anlamak mümkündür. Bu çalışmada duygu analizi gerçekleştirilerek Yemeksepeti şirketine ilişkin Twitter kullanıcılarının duyguları incelenmiştir. Çalışma kapsamında KNN, Rastgele Orman, Destek vektör makinası ve Naive Bayes algoritmaları kullanılmış, ardından bu algoritmaların bir bileşkesi oluşturularak nihai sonuca ulaşılmıştır. Elde edilen verilere göre ilgili şirkete ilişkin atılan tweetlerin büyük çoğunluğu olumlu duygular ile oluşturulmuştur. Ancak belirgin bir tarih aralığı ya da ikinci bir kelime girerek, ilgili şirketin, ilişkili kelime konusunda gerçekleştirdiği faaliyetleri hakkında insanların duygularına erişmek daha mümkün olabilecektir. Ek

olarak gelecekteki alıřmalarda kullanıcılara gerek duyguları sorularak farklı metotlar ile yapılan duygu analizlerinin sonuçlarının karşılařtırılması ve hangi yöntemin daha doğru sonuç verdiğinin karşılařtırılması mümkündür.



KAYNAKLAR

Çelik, Davut. (2015). Veri Madenciliği Kullanarak Akıllı Reklam/Anket Uygulaması, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, Bilgisayar Mühendisliği Bilim Dalı, Yüksek Lisans Tezi

Çelik, Esra; Dal, Deniz ve Aydın, Tolga. (2021). Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması, Avrupa Bilim ve Teknoloji Dergisi, 27, 880-889

Karaöz, Burcu. (2018). Büyük Veri ve İşletme Analitiği: Sosyal Medya ve Duygu Analizi ile Bir Öngörü Modeli, T.C. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, Sayısal Yöntemler Bilim Dalı, Doktora Tezi

Savaş, Serkan; Topaloğlu, Nurettin ve Yılmaz, Mithat. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 11(21): 1-23

Karacan, Hacer ve Yeşilbudak, Mehmet. (2010). Kullanıcı Merkezli İnteraktif Veri Madenciliği: Bir Literatür Taraması, Bilişim Teknolojileri Dergisi, 3(1), 17-22

Alanlar, Ezgi. (2021). Pazar Sepeti Analizi ile Birliktelik Kurallarının Belirlenmesi: Perakende Sektöründe COVID-19 Etkisi, Karabük Üniversitesi, Lisansüstü Eğitim Enstitüsü, Endüstri Mühendisliği Anabilim Dalında, Yüksek Lisans Tezi

Özçakır, F. Cemal ve Çamurcu, A. Yılmaz. (2007). Birliktelik Kuralı Yöntemi İçin Bir Veri Madenciliği Yazılımı Tasarımı ve Uygulaması, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 6(12), 21-37

Çelik, Esra; Dal, Deniz ve Aydın, Tolga. (2021). Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması, Avrupa Bilim ve Teknoloji Dergisi, 27, 880-889

Cebeci, H. İbrahim. (2020). Sosyal Medya Verileri ile Duygu Analizi, Mühendislikte Yapay Zeka Uygulamaları Kitabı, Sakarya: SAÜ Mezunlar Derneği Yayınları

Uzun, Yusuf; Uzun, F. Nur ve Çakar, Esra. (2021). Veri Madenciliği ve Kullanım Alanları, Uluslararası Mühendislik, Doğa ve Sosyal Bilimler Sempozyumu "Enerji ve Toplum", 25- 28 Kasım 2021, Batman Üniversitesi, https://www.researchgate.net/publication/356819774_VERI_MADENCILIGI_VE_KULLANIM_ALANLARI_DATA_MINING_AND_AREAS_OF_USE

Riloff, Ellen. (1996). Automatically Generating Extraction Patterns from Untagged Text". In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1044–1049, <https://dl.acm.org/doi/10.5555/1864519.1864542>

Kamruzzaman, S.M., Haider, Farhana, Hasan, A. Ryadh, (2005). Text Classification Using Data Mining, <https://arxiv.org/ftp/arxiv/papers/1009/1009.4987.pdf>

Nasukawa, Tetsuya ve Yi, Jeonghee. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing, Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003), October 23-25, 2003, Sanibel Island, FL, USA

Ak, Dilek. (2019). Pazarlama Alanındaki Uluslararası Akademik Dergilerin Metin Madenciliği Yöntemi ile Değerlendirilmesi, T.C. Sakarya Üniversitesi, İşletme Enstitüsü, Yüksek Lisans Tezi

Kasaba, Evrim ve Yıldıztepe, Engin. (2016). Destek Vektör Makinesi Yöntemi ile Sosyal Medya Verileri Üzerine Bir Duygu Çözümleme Uygulaması, <https://ab.org.tr/ab16/bildiri/247.pdf>

Şahinaslan, Önder; Dalyan, Hüseyin ve Şahinaslan, Ender. (2022). Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi, Bilişim Teknolojileri Dergisi, 15(2), 221-229

Nizam, Hatice ve Akın, S. Sıla. (2019). Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, <http://inet-tr.org.tr/inetconf19/bildiri/10.pdf>

Ersöz, Filiz ve Çınar, Yasemin. (2021). Veri Madenciliği ve Makine Öğrenimi Yaklaşımlarının Karşılaştırılması: Tekstil Sektöründe bir Uygulama, Avrupa Bilim ve Teknoloji Dergisi, 29, 397-414



ÖZGEÇMİŞ

2008 yılında başladığı Darıca Neşet Yalçın Anadolu lisesinden 2012 yılında mezun oldu. 2012 yılında Düzce Üniversitesi Bilgisayar Mühendisliği Bölümüne başladı ve 2017 yılında başarıyla tamamladı. Gebze Teknik Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı yüksek lisans eğitimine 2018 yılında başladı. 2017 yılından bu yana Darıca Belediyesi Bilgi İşlem Müdürlüğünde Bilgisayar Mühendisi olarak çalışmaktadır.



EKLER

Ek A: Tez Çalışması Kapsamında Yapılan Yayınlar

Tunalı M. (2023), “Twitter Duygu Analiz Üzerine Bir Uygulama(Yemeksepeti Örneği)”, Participated in the 8 th International ‘BAŞKENT’ Congress On Humanities and Social Sciences, 20-40.

