



Amazon Ürün Değerlendirmeleri Üzerinde Derin Öğrenme/Makine Öğrenmesi Tabanlı Duygu Analizi Yapılması

Yazılım Mühendisliği Ana Bilim Dalı
Tezsiz Yüksek Lisans

Neşe Gök
ORCID 0000-0002-8592-9614

Tez Danışmanı: Doç. Dr. Aytuğ Onan

Ocak 2023

Amazon Ürün Deęerlendirmeleri Üzerinde Derin Öęrenme/Makine Öęrenmesi Tabanlı Duygu Analizi Yapılması

Özet

Bu çalışmanın asıl amacı kaggle veri seti üzerinden çekilmiş olan amazon ürün yorumlarını makine öęrenmesi yöntemleri ve duygu analizi sınıflandırması ile analiz etmektedir. Yapılan çalışmada birden beşe kadar verilen duygu oranlar analiz edilmiş ve en son olarak duygu içeriyorsa pozitif, negatif ve nötr olma durumları saptanmıştır. Python programa dili kullanılmış olup, Kaggle'ın kendi idesi üzerinden 8 farklı makine öęrenmesi algoritması ile veriler analiz edilmiştir. Kullanılan algoritmalar arasında Naive Bayes (NB), Destek Vektör Makineleri(SVM), Karar Ağaçları(DT), Lojistik Regresyon(LR-tf-idf), Lojistik Regresyon(LR- w2v), Rastgele Orman (RF), K-En Yakın Komşu (KN), Bagging Algoritması (BG) gösterilebilir.

Anahtar Kelimeler: Amazon Duygu analizi, Kaggle, Makine öęrenmesi

Sentiment Analysis with Deep Learning/Machine Learning Techniques on Amazon Products Review

Abstract

The main purpose of this study is to analyze amazon product reviews drawn from the kaggle data set with machine learning methods and sentiment analysis classification. In the study, the ratios of emotions given from one to five were analyzed and finally, if they contain emotions, positive, negative and neutral situations were determined. Python programming language was used and data were analyzed with 8 different machine learning algorithms on Kaggle's own idea. Among the algorithms used are Naive Bayes (NB), Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression (LR- w2v), Logistic Regression (LR-tf-idf), Random Forest (RF), K-Nearest Neighbor (KN), Bagging Algorithm (BG).

Keywords: Amazon, Sentiment Analysis, Kaggle, Machine learning

İçindekiler

| | |
|--|----|
| Özet | i |
| Abstract | ii |
| Giriş..... | 1 |
| 1. Materyal ve Metot | 2 |
| 1.1. Materyal | 2 |
| 1.1.1. Veri Seti | 2 |
| 1.1.2. Veri Özellikleri..... | 2 |
| 1.2. Metod..... | 3 |
| 1.2.1. Veri Önışleme | 5 |
| 1.2.2. Kullanılan Sınıflandırma Algoritmaları..... | 5 |
| 1.2.2.1. Naive Bayes (NB) | 5 |
| 1.2.2.2. Destek Vektör Makineleri(SVM)..... | 5 |
| 1.2.2.3. Karar Ağaçları(DT)..... | 5 |
| 1.2.2.4. Lojistik Regresyon(LR)..... | 6 |
| 1.2.2.5. Rastgele Orman (RF)..... | 6 |
| 1.2.2.6. K-En Yakın Komşu (KN) | 6 |
| 1.2.2.7. Bagging Algoritması (BG)..... | 6 |
| Sonuç..... | 8 |
| Kaynakça..... | 9 |

Giriş

Yapay zekâ ve makine öğrenmesi konusunda önemli bir dönüm noktası olan 1960'li yıllar makine öğrenmesinin miladını oluşturmaktadır. IBM tarafından ilk defa kullanılan makine öğrenmesi o günden günümüze kadar çok fazla gelişme göstermiş ve yapay zeka başta olmak üzere bir çok alanda kullanıla gelmiştir. Bu nedenle makine öğrenmesi IBM'e göre *“insanların öğrenme şekillerini taklit etmek için veri ve algoritmaların kullanımına odaklanıp doğruluğunu kademeli olarak artıran bir yapay zeka (AI) ve bilgisayar bilimi dalıdır.”*[1]

Microsoft, Azure'ye göre ise *“Makine Öğrenmesi bir bilgisayarın doğrudan yönergeler olmadan öğrenmesine yardımcı olmak için matematiksel modelleri kullanma işlemidir.”* Microsoft, Azure tarafından, yapay zekanın (AI) bir alt kümesi olarak tanımlanan Makine Öğrenmesi algoritmalar yardımı ile verilerdeki kalıpları belirler ve veriyi işleyerek tahmin yapabilen bir model elde edebilmek için belirlenen bu kalıpları kullanır. Makine öğrenmesi uyarlanabilirlik özelliği sayesinde sürekli değişen ve teknoloji ağırlıklı olan günümüz dünyası için çözüm odaklı muhteşem bir seçenektir.[2]

Esasında Makine Öğrenmesi, belli görevleri nasıl yerine getireceğini keşfetme işlevi olarak tanımlamak mümkündür. Sağlanan veriler ile belirli görevlerin yerine getirilmesini sağlayan Makine Öğrenmesi eldeki veriler ile kendi algoritmasını sağlayarak belli görevleri yerine getirmesini sağlamaktadır. [3, ss. 13–18]

Duygu analizi, farklı yöntemler ile elde edilen duygu verilerinin genellikle bir yazı parçasının olumlu, olumsuz veya nötr (kararsız) olup olmadığını çeşitli veri işleme yöntemleri kullanarak belirleme yöntemidir. Duygu analizi konuşmacının fikrini veya tutumunu inceleyen fikir madenciliği olarak da bilinmektedir. Duygu analizinin en önemli özelliği ve öncelikli hedefi verisi elde edilen insan topluluğunun belirli bir konu hakkında hissettiklerini keşfetmektir. Bu nedenle fikir madenciliği yada başka bir deyişle duygu analizi metin analizi, bilişimsel dilbilim, biyometrelerin ve doğal dil işleme duygusal durumları ve öznel bilgileri sistematik olarak tanımlamak, ölçmek ve incelemek için kullanılmasını ifade eder. En kısa hali ile duygu analizi elde edilen verinin konusu olan bir yazar, bir konuşmacının veya bir konunun başka bir konuya olan duygusal tepkilerini ve

etkilerini belirlemeyi hedeflemektedir.[4, s. 361]

1. Materyal ve Metot

1.1. Materyal

Çevrimiçi bir veri bilimcileri ve makine öğrenimi uygulayıcıları topluluğu ve Google'ın bir alt şirketi olan Kaggle[5] veri tabanında bulunan veriler gerekli olan veri işleme ve sınıflandırma algoritmaları işlemlerden geçirildikten sonra Amazon müşterilerinin ürün satın alırken göstermiş oldukları duyguların analizi yapılarak, ortaya çıkan sonuçları karşılaştırılmıştır.

Bu verilerin toplanmasında özellikle Kaggle'ın oynanmış olduğu rol çok büyük önem arz etmektedir. . Kaggle, kullanıcıların veri kümelerini bulmasına ve yayınlamasına, web tabanlı bir veri bilimi ortamında modelleri keşfetmesine ve oluşturmasına, diğer veri bilimciler ve makine öğrenimi mühendisleriyle çalışmasına ve veri bilimi zorluklarını çözmek için yarışmalara katılmasına olanak tanır.

1.1.1 Veri Seti

Bu çalışma kapsamında Amazon müşterilerinin alışveriş yaparken göstermiş oldukları duyguları analiz edebilmek adına Kaggle[5] veri tabanında bulunan Amazon Ürünlerinin Tüketici İncelemeleri

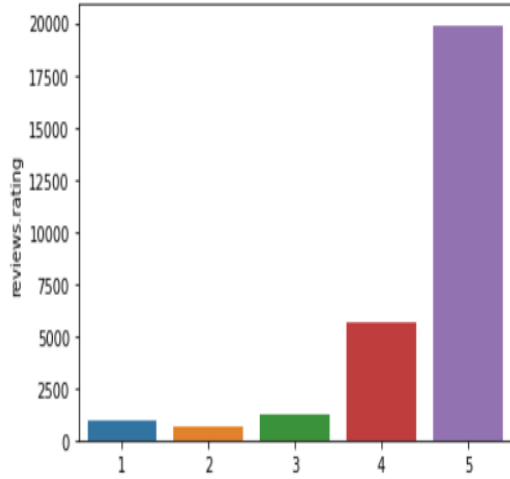
(Consumer Reviews of Amazon Products) veri setinden yola çıkılmış ve çalışmamızda veri setinin en güncel hali kullanılmıştır.

1.1.2. Veri Özellikleri

Bu çalışma kapsamında Kaggle veri tabanında bulunan 28331 satır ve 24 adet sütun veri seti dikkate alınmıştır. Şekil 1'de gözükeceği gibi bu veri setinde yorumlara verilen duygu dereceleri (reviews rating) dikkate alınarak karşılaştırma yapılmıştır. Aşağıdaki grafikte görüldüğü gibi Amazon müşterilerinin alışveriş sitesinde ürün satın alırken yorumlara vermiş oldukları derecelendirme oyları karşılaştırılmıştır. En yüksekte en düşüğe doğru gidildiğinde 19.897 kişi 5 verirken, 5648 kişi 4, 1206 kişi 3, 965 kişi 1 ve 616 kişi ise 2 şeklinde bir derecelendirme oyu vermiştir.

| | |
|---|-------|
| 5 | 19897 |
| 4 | 5648 |
| 3 | 1206 |
| 1 | 965 |
| 2 | 616 |

Name: reviews.rating, dtype: int64



1.2. Metod

Bu çalışma kapsamında Kaggle [5] veri tabanından indirilen 17MB'lik veri seti kullanılmıştır. Aynı şekilde bu çalışma ile birlikte verinin ön işlenmesi ve analizi, duygu sınıflandırması ve görselleştirmesi ele alınmıştır.

Nisan 2010 tarihinde Anthony John Goldbloom tarafından kurulan Kaggle şirketi genel itibari ile Veri Bilimi ve Makine Öğrenmesi konuları üzerinde uzmanlaşmış bir Amerikan şirkettir. 2011 yılında Nicholas Gruen'in yerine Max Levchin geçmiş ve şirketin öz sermayesini 25 milyon dolar değerinde artırmıştır. 8 Mart 2017 tarihinde Google çatısı altına giren Kaggle şirketi, veri bilimi alanında dünyanın en öncü şirketlerinden biridir. [6][7]

Yine aynı şekilde bu proje kapsamında kodlaması için Kaggle'ın

kendi web sayfasında bulunan kodlama bölümü idesi kullanılmıştır.

Bu araştırma kapsamında veri bilimi alanında en yaygın olarak kullanılan programlama olan Python tercih edilmiştir. 1990'lı yıllında Guido van Rossum tarafından geliştirilen Python programlama dili nesne yönelimli olmasının yanında etkileşimli ve yüksek seviyeli yani big data(büyük veri) için elverişli olan bir programlama dili olma özelliğini kendisinde barındırmaktadır. [8]

Python isimi genel kanının aksine bilinen yılan türünden değil, daha çok Rossum'un en çok sevmiş olduğu ve 1970'lerin en popüler komedi grubu olan Monty Python isimli altı kişiden oluşan gruptan almıştır. Python programlama dili hali hazırda Python Yazılım Vakfı'nın çabaları ile bir araya gelen gönüllerin emekleri neticesinde çalışmalarını sürdürmektedir. 1994 tarihinde Python 1.0 sürümü, 16 Ekim 2000 tarihinde Python 2.0 sürümü piyasa sürülmüştür. 2008 yılı itibari ile Python 3.x serisi piyasa sürülmeye başlanmıştır ama Python 3.x serisi Python 2.x serisi ile uyum içinde değildir.[9][10]

Bu proje kapsamında Pandas, Numpy, Matplotlib ve NLTK kütüphaneleri başta olmak üzere bir çok kütüphaneden yararlanılmıştır.

Bir yazılım kütüphanesi olan Pandas Python dilinde yazılmıştır. Esasında ismini bir ekonometri terimi olan veri panelinden alan Pandas Kütüphanesi, bir veri paneli birçok zaman aralığı içinde farklı gözlemlerin işlenebildiği yapıyı tarif etmektedir.[11] Bu kütüphane yardımı ile veri bilimci sayısal tablolar ile işlem yapabildiği gibi zaman etiketli seriler ile de veri yapısı oluşturabilmekte. Pandas kütüphanesi aracılığı ile veri bilimci oluşturuş olduğu veri yapısı üzerinde çok çeşitli işlemler yapabilmektedir. İlk yayınlandığı 11 Ocak 2008 tarihinden itibaren büyük bir kesim tarafından kullanılan Pandas Kütüphanesi ücretiz bir kütüphane olmasının ötesinde BSD lisansına sahiptir.[12]

Bir Python kütüphanesi olan NumPy, esasında büyük ve çok boyutlu dizileri ve matrisleri destekleyerek mevcut diziler üzerinde işlem yaparak üst düzey matematiksel işlevler ekleyen bir nevi kitaplıktır. [13]

Jim Fulton bir matris paketinin bir uygulamasını tamamladıktan sonra Jim Hugunin bu matris ve uygulamayı geliştirmeyi başardı ve Numeric olarak isimlendirdi. Numerical Python extensions (Sayısal Python uzantıları) veya "NumPy" [14]olarak bilinen bu kütüphaneye daha sonra David Ascher, Konrad Hinsen ve

Travis Oliphant gibi isimler çok büyük katkılarda bulundular. [15]

Yine bir Python kütüphanesi olan Matplotlib sayısal, matematiksel uzantısı NumPy için bir nevi çizim kitaplığıdır. Tkinter, wxPython, Qt veya GTK gibi genel amaçlı GUI araç takımlarını kullanarak grafikleri uygulamalara gömmek için nesne yönelimli bir API sağlar. [16]

Matplotlib aslen John D. Hunter tarafından yazılmıştır. O zamandan beri aktif bir geliştirme topluluğu var ve BSD tarzı bir lisans altında dağıtılıyor. Michael Droettboom, John Hunter'ın Ağustos 2012'deki ölümünden kısa bir süre önce matplotlib'in baş geliştiricisi olarak aday gösterildi ve buna Thomas Caswell da katıldı.[17][18][19]

Natural Language Toolkit veya daha yaygın olarak kullanılan kısaltması olan NLTK, Python programlama dili ile kodlanmış İngilizce için sembolik ve istatistiksel doğal dil işleme (NLP) için bir kitaplık ve program paketidir. Pennsylvania Üniversitesi Bilgisayar ve Bilişim Bilimleri Bölümü'nde Steven Bird ve Edward Loper tarafından geliştirilmiştir. NLTK, grafik gösterimleri ve örnek verileri içerir. Araç takımı tarafından desteklenen dil işleme görevlerinin altında

yatan kavramları açıklayan bir kitap ve bir yemek kitabı eşlik eder.[20][21]

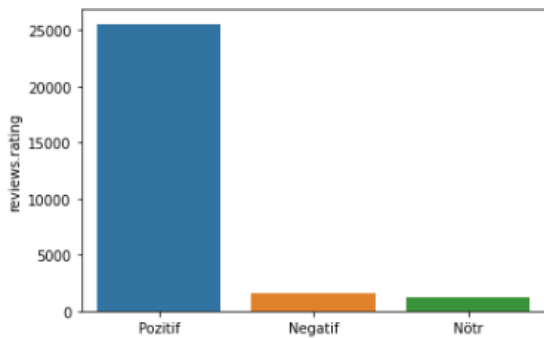
1.2.1. Veri Önışleme

Yorumların ve ürünlere verilen duygu oranlarının analiz edilebilmesi ve sınıflandırılabilmesi adına yorumlar ve duygu oranları çok farklı veri önışleme aşamalarından ve işlemlerinden geçirilmiştir.

Bu önışleme aşamalarında tüm harfleri küçük harfe dönüştürme, kök belirleme, belirteçleştirme ve etiketleme gibi işlemler ele alınmıştır.

Müşterilerin ürünleri satın alırken 1 ile 5 arasında duygu oranlarına göre oy vermişlerdir. 1 ile 2 dahil şekilde negatif, 3 nötr ve 4 ile 5 dahil şekilde pozitif şeklinde verilen işlenmiş ve yapılan hesaplama göre verilen yeni bir kolon ve veri görselleştirme kodu ile sonuçlar kayıt altına alınmıştır.

```
Pozitif      25545
Negatif      1581
Nötr         1206
Name: reviews.rating, dtype: int64
```



1.2.2 Kullanılan Sınıflandırma Algoritmaları

1.2.2.1. Naive Bayes (NB)

İsmi ünlü Bayes teoreminden alan alan Naive ve Bayes makine öğrenmesi algoritması özellikle metin sınıflandırma sürecinde anlaşılabilir ve kolaylıkla uygulanabilir algoritmalarından biridir. Bu algoritma ile birlikte sisteme sunulan verilerin sınıfını yani bir nevi kategorisini tespit etmeyi amaçlamaktadır.[22]

1.2.2.2. Destek Vektör Makineleri(SVM)

Destek Vektör Makinesi (SVM) algoritması sınıflandırma veya regresyon problemlerini ele almak için kullanılabilen denetimli bir makine öğrenmesi yöntemidir. Bununla birlikte, çoğunlukla sınıflandırma problemlerinde kullanılır. Destek Vektör Makinesi (SVM) algoritması esasında değişkenler arasındaki örüntülerin bilinmediği durumlarda veri setlerinde regresyon, örüntü tanıma ve sınıflama sorunları için kullanılan istatistiksel bir öğrenme teorisi olarak tanımlamak mümkündür. [23]

1.2.2.3. Karar Ağaçları(DT)

Kendi içinde anlaşılır ve basit bir yapıyı barındıran Karar Ağaçları algoritması böl ve yönet stratejisini kullanarak problemlere çözüm getirmeyi hedeflemektedir. Kullandığı yöntem itibari

ile parametre gerektirmeden regresyon ve sınıflandırma gerçekleştirerek hiyerarşik bir gözetimli öğrenme modeli oluşturmayı hedeflemektedir.[24]

1.2.2.4. Lojistik Regresyon(LR)

İki farklı veri faktörü arasındaki bağlantı ilişkisini elde edebilmek için matematikten faydalanan bir veri analizi metodu ve tekniği olan Lojistik regresyon, ilerleyen süreçlerde diğerine dayalı işlev görmüş olan bu faktörlerden birinin mevcut değerini tahmin etmek için ikisi arasında bulunan bu tür ilişkiyi kullanır. İşleme alınan tahmin sürecinin sonucunda genellikle iki sonuçtan biri olan “Evet” ya da “Hayır” gibi kısıtlı sayıdaki netice elde edilmiş olur. Makine öğrenmesi ve yapay zeka alanlarında çok önemli bir teknik olan bu algoritma ile herhangi bir insan müdahalesine gerek kalmadan meydana gelen karmaşık veri işleme işlevini yapabilmesi için eğitmeye olanak veren bir tür yazılım ve makine öğrenmesi programlarıdır.[25]

1.2.2.5. Rastgele Orman (RF)

Esasında birden çok karar ağacını meydana getiren bir sınıflandırma makine öğrenmesi algoritması olan Rastgele Orman algoritması özellikle eğitim için ayrılmış veri setini oluşturan verilerden meydana gelmektedir. Rastgele Orman algoritması birçok karar ağacının

oluşturmuş olduğu sınıflandırma sonuçlarından faydalanarak mevcut sistemde var olan verilerin test süreçlerinde girdinin içinde bulunmuş olduğu sınıfa göre çoğunluğun baskın olmuş olduğu oy neticesinde son kararı vermektedir. Bu nedenden ötürü literatürde genellikle topluluk öğrenme yöntemi ismi ile anılagelmiştir. [26]

1.2.2.6. K-En Yakın Komşu (KN)

Nesnelerin birbirlerine yakınlık ilişkilerini esas alarak sınıflandırma işlemini gerçekleştiren K-En Yakın Komşu (KNN) algoritması K (komşu sayısı) değerini ele alarak eleman sayısını belirlemeye çalışmaktadır. Bu algoritma yapısında herhangi bir veri ile işleme girdiğinde K'ye olan mesafe sıralanmakta, hesaplanmakta ve algoritmanın isminin de almış olduğu en küçük uzaklığa bağlı olarak oluşmuş olan yeni değere bağlı olarak en yakın komşuların mevcut şekilde bulunmuş olduğu kümeye eklemektedir. [26]

1.2.2.7. Bagging Algoritması (BG)

Kalifornia Üniversitesi, Berkeley'de bir istatistikçi olan Leo Breiman tarafından önerilmiş bir makine öğrenme algoritmasıdır. Bagging Algoritması'nın burada görmüş olduğu işlev mevcut eğitim setinde bulunan farklı örneklemeler aracılığıyla ve ilişkisiyle veri

üzerinde eğitilmiş olan temel öğrenme algoritmalarının bir nevi birleştirilmesi vasıtasıyla sınıflandırıcı toplumu meydana getirmek amacıyla esasında bir topluluk öğrenmesi metodudur. Bu nedenden dolayı bu makine öğrenmesi algoritmasının en çok öne çıkan özelliği ve kendisini diğer algoritmalarından farklı kılan özelliği “topluluğu oluşturan her bir temel öğrenme algoritmasının, farklı eğitim setleri üzerinde eğitilmesi ile çeşitliliğin sağlanması ilkesine dayanması” olarak açıklamak mümkündür. [27]

Sonuç

Bu çalışma kapsamında kaggle veri seti üzerinden çekilmiş olan amazon ürün yorumlarını makine öğrenmesi yöntemleri ve duygu analizi sınıflandırması ile analiz edilmiştir. Kullanılan yöntemler içinde Rastgele Orman yöntemi %87 doğruluk oranı ve %69 F-puanıyla yüksek başarımla elde etmiştir. Diğer sınıflandırıcılara göre Lojistik Regression (w2v) %72 ile daha düşük başarımla elde etmiştir.

| | Sınıflandırma Algoritmaları | | | | | | | |
|------------|-----------------------------|------------|---------|------------|------------|-------------|-------------|------------|
| | LR(tf-idf) | NB(tf-idf) | LR(w2v) | BG(tf-idf) | DT(tf-idf) | KN (tf-idf) | SVC(tf-idf) | RF(tf-idf) |
| Doğruluk | 0.78 | 0.73 | 0.72 | 0.73 | 0.82 | 0.73 | 0.78 | 0.87 |
| Kesinlik | 0.67 | 0.48 | 0.36 | 0.53 | 0.66 | 0.53 | 0.69 | 0.88 |
| Duyarlılık | 0.50 | 0.22 | 0.26 | 0.30 | 0.62 | 0.30 | 0.50 | 0.59 |
| F1-Puanı | 0.55 | 0.20 | 0.26 | 0.34 | 0.64 | 0.34 | 0.56 | 0.69 |

Kaynakça

- [1] IBM, “Makine öğrenmesi nedir?” <https://www.ibm.com/tr-tr/cloud/learn/machine-learning>.
- [2] Azure, “Makine öğrenmesi nedir?,” [Online]. Available: <https://azure.microsoft.com/tr-tr/resources/cloud-computing-dictionary/what-is-machine-learning-platform/>.
- [3] E. Alpaydin, *Introduction to Machine Learning*, Forth. The MIT Press, 2020.
- [4] P. KIRCI and E. GÜLBAK, “Instagram Verileri ile Duygu Analizi,” *Eur. J. Sci. Technol.*, no. August, pp. 360–364, 2020, doi: 10.31590/ejosat.780129.
- [5] “Consumer Reviews of Amazon Products | Kaggle.” <https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products> (accessed Dec. 29, 2022).
- [6] E. Moyer, “Google buys Kaggle and its gaggle of AI geeks - CNET,” 2017. <https://www.cnet.com/science/google-buys-kaggle-and-its-gaggle-of-ai-geeks/> (accessed Dec. 29, 2022).
- [7] M. L. M. Frederic Lardinois, “Google is acquiring data science community Kaggle | TechCrunch,” Mar. 08, 2017. <https://techcrunch.com/2017/03/07/google-is-acquiring-data-science-community-kaggle/> (accessed Dec. 29, 2022).
- [8] “General Python FAQ — Python 3.11.1 documentation.” <https://docs.python.org/3/faq/general.html> (accessed Dec. 29, 2022).
- [9] A.M. Kuchling and Moshe Zadka, “What’s New in Python 2.0 — Python v2.7.3 documentation.” <https://web.archive.org/web/20121030072329/https://docs.python.org/2/whatsnew/2.0.html> (accessed Dec. 29, 2022).
- [10] Guido van Rossum, “What’s New In Python 3.0 — Python 3.11.1 documentation.” <https://docs.python.org/3/whatsnew/3.0.html> (accessed Dec. 29, 2022).
- [11] Wes McKinney, “Pandas: a Foundational Python Library for Data Analysis and Statistics,” *Int. J. RF Microw. Comput. Eng.*, vol. 19, no. 5, pp. 583–591, 2009, doi: 10.1002/mmce.20381.
- [12] Pandas, “Package overview.” https://pandas.pydata.org/pandas-docs/stable/getting_started/overview.html.
- [13] C. R. Harris *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, 2020, doi: 10.1038/s41586-020-2649-2.
- [14] K. Jarrod Millman and Michael Aivazis, “Python for Scientists and Engineers,” *Science (80-.)*, vol. 184, no. 4141, p. 1028, 1974, doi: 10.1126/science.184.4141.1028-a.

- [15] P. F. Dubois, K. Hinsien, and J. Hugunin, “Numerical Python,” *Comput. Phys.*, vol. 10, no. 3, p. 262, 1996, doi: 10.1063/1.4822400.
- [16] Matplotlib, “API Reference.” <https://matplotlib.org/stable/api/index.html#module-pylab>.
- [17] G. W. Amy Brown, *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks*. California: ulu.com, 2008.
- [18] “Credits — Matplotlib 3.6.2 documentation.” <https://matplotlib.org/stable/users/project/credits.html> (accessed Dec. 29, 2022).
- [19] “Matplotlib Lead Developer Explains Why He Can’t Fix the Docs—But You Can - NumFOCUS.” <https://numfocus.org/blog/matplotlib-lead-developer-explains-why-he-cant-fix-the-docs-but-you-can> (accessed Dec. 29, 2022).
- [20] NLTK, “NLTK : Documentation ‘Natural Language Toolkit.’” <https://www.nltk.org/>.
- [21] “Preface,” *nltk.org*. <https://www.nltk.org/book/ch00.html> (accessed Dec. 29, 2022).
- [22] O. Kaynar, M. Yildiz, Y. Görmez, and A. Albayrak, “Makine Öğrenmesi Yöntemleri ile Duygu Analizi,” no. September, pp. 234–241, 2016.
- [23] E. ŞAHİN, “MAKİNE ÖĞRENME YÖNTEMLERİ VE KELİME KÜMESİ TEKNİĞİ İLE İSTENMEYEN E-POSTA / E-POSTA SINIFLAMASI,” Hacettepe Üniversitesi, 2018.
- [24] T. E. Kalaycı, “Kimlik hırsızları web sitelerinin sınıflandırılması için makine öğrenmesi yöntemlerinin karşılaştırılması,” *Pamukkale Univ. J. Eng. Sci.*, vol. 24, no. 5, pp. 870–878, 2018, doi: 10.5505/pajes.2018.10846.
- [25] A. AWS, “Lojistik regresyon nedir?” <https://aws.amazon.com/tr/what-is/logistic-regression/>.
- [26] T. A. Çelik, Esra, Deniz Dal, “Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması,” *Avrupa Bilim ve Teknol. Derg.*, no. 27, pp. 880–889, 2021, doi: 10.31590/ejosat.905259.
- [27] A. Onan, “Firma Başarısızlığının Tahmin Edilmesi İçin Kümelemeye Dayalı Bir Sınıflandırıcı Topluluğu Yaklaşımı,” *Alphanumeric J.*, vol. 6, no. 2, 2018, doi: 10.17093/alphanumeric.333785.