PAMUKKALE ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ

Kullanıcı Yorumları Üzerinde Makine Öğrenmesi Yoluyla Duygu Analizi

LİSANS TEZİ

Faruk GÜRLESİN
(19253062)

Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Meriç ÇETİN

OCAK 2023

Pamukkale Üniversitesi Bilgisayar Mühendisliği Bölümü, 19253062 numaralı Lisans Öğrencisi Faruk GÜRLESİN, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "Kullanıcı Yorumları Üzerinde Makine Öğrenmesi Yoluyla Duygu Analizi" başlıklı tezini aşağıdaki imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Meriç ÇETİN** .....

Jüri Üyeleri : .....
Pamukkale Üniversitesi

.....
Pamukkale Üniversitesi

Teslim Tarihi : **6 Ocak 2023**
Savunma Tarihi : **27 Aralık 2022**

Önsöz

Çalışmalarım sırasında bilgi ve tecrübelerini benden esirgemeyen degerli hocam ve tez danışmanım Sayın Doç. Dr. Meriç ÇETİN'e teşekkürlerimi sunarım. Kariyer çizgimde küçük dokunuşlarıyla büyük etkiler yaratan değerli hocam Sayın Dr. Elif HAYTAOĞLU'na teşekkürlerimi sunarım.

OCAK 2023 Faruk GÜRLESİN

İçindekiler

Sayfa

ÖNSÖZ ..... **v**
İÇİNDEKİLER ..... **vii**
KISALTMALAR ..... **ix**
SEMBOLLER ..... **xi**
ÇİZELGE LİSTESİ ..... **xiii**
ŞEKİL LİSTESİ ..... **xv**
ÖZET ..... **xvii**
SUMMARY ..... **xix**
1. GİRİŞ **1**
1.1 Tezin Amacı 1
1.2 Literatür Araştırması 2
2. BİLİMSEL ARKAPLAN **5**
2.1 Doğal Dil İşleme 5
2.2 Türkçe Metinler Üzerinde Duygu Analizi 5
2.3 Türkçe Metinler Üzerinde Kategori Analizi 5
2.4 Makine Öğrenmesi 5
2.5 Derin Öğrenme ve Yapay Sinir Ağları 7
2.5.1 Multilayer Perceptron (Çok Katmanlı Algılayıcı) 8
2.5.2 Convolutional Neural Network 8
2.5.3 Recurrent Neural Network 9
2.5.4 Long-Short Term Memory 9
3. METODOLOJİ **11**
3.1 Veri Seti 11
3.2 Ön İşleme 12
3.2.1 Gürültü Temizleme 12
3.2.2 Dönüşüm İşlemleri 12
3.2.3 Boyut Seçimi 12
3.2.4 Dil ve Emoji Tespiti 12
3.3 Model seçimi 13
3.3.1 Transformerlar 13
3.3.1.1 Transformer Mimarisi 13
3.3.2 BERT Modeli 14
4. UYGULAMA **15**
4.1 Web Projesi 15
4.1.1 Sentiment Paketi 15

4.1.2 Sub Category Paketi 16
4.1.3 TestDb Paketi 18
4.2 Database 19
5. SONUÇLAR VE DEĞERLENDİRME 21
5.1 Uygulama 21
5.2 Sonuçlar 21
5.2.1 Makine Öğrenimi Modeli Sonuçları 21
5.2.2 Web Projesi Çıktıları ve Analizler 23
5.3 Genel Değerlendirme 23
5.4 Tartışma 23
KAYNAKLAR 25
EKLER 27
ÖZGEÇMİŞ 29

KISALTMALAR

NB	**:** Naive Bayes
SVM	**:** Support-Vector Machine
POS	**:** Part Of Speech
RF	**:** Random Forest
KNN	**:** k-Nearest Neighbors
DTs	**:** Decision Trees
NLP	**:** Natural Language Processing
MSE	**:** Mean Square Error
MLP	**:** Multilayer Perceptron
CNN	**:** Convolutional Neural Network
RNN	**:** Recurrent Neural Network
LSTM	**:** Long-Short Term Memory
CV	**:** Computer Vision
GPT	**:** Generative Pre-trained Transformer
BERT	**:** Bidirectional Encoder Representations from Transformers

SEMBOLLER

Tablo Listesi

Sayfa

Tablo 3.1 : Duygu Analizi Veri kümesi 11
Tablo 3.2 : Kategori Analizi Veri kümesi 11

Şekil Listesi

Sayfa

Şekil 2.1 : Multilayer Perceptron 8
Şekil 2.2 : CNN Bağlantı Katmanı 9
Şekil 2.3 : LTSM Bağlantı Katmanı 9
Şekil 4.1 : Duygu Analizi Tahminleme Sayfası 15
Şekil 4.2 : Yorumların Duygulara Göre Dağılımı 16
Şekil 4.3 : Marka Bazında Duygu Analizi 16
Şekil 4.4 : Kategori Analizi Tahminleme Sayfası 17
Şekil 4.5 : Yorumların Kategorilere Göre Dağılımı 17
Şekil 4.6 : Marka Bazında Kategori Analizi 18
Şekil 4.7 : Ürünler ve Yorumlar Sayfası 18
Şekil 4.8 : Proje ile İlgili Genel Bir Projeksiyon 19
Şekil 4.9 : Projede Kullanılan İlişkisel Veritabanı ER Diyagramı 20
Şekil 5.1 : Train Loss Sonuçları 21
Şekil 5.2 : Kategori Analizi Precision, Recall, F1-Score Sonuçları 22
Şekil 5.3 : Kategori Analizi Precision, Recall, F1-Score Grafik 22
Şekil 5.4 : Duygu Analizi Precision, Recall, F1-Score Grafik 23

Kullanıcı Yorumları Üzerinde
Makine Öğrenmesi Yoluyla Duygu Analizi

Özet

Gelişen rekabetçi pazar araştırmaları içerisinde şirketlerin müşterilerle etkileşiminin kuvvetlendirilmesi; ürün geliştirme, pazarlama gibi alanlarda büyük faydalar sağlamaktadır. Günümüz teknolojisi, şirketlerin karlılığını artırmada geleneksel yöntemlerin hantallığının aksine hızlı ve kesin çözümler geliştirebilmektedir. Bu araştırma, kullanıcı yorumları ile derlenen bir veri seti üzerinde çeşitli makine öğrenmesi ve dil işleme teknikleri uygulanarak oluşturulacak modeli incelemiştir. Bu model, kullanıcı yorumlarını olumlu, olumusuz veya tarafsız olarak sınıflandırmaktadır. İkinci aşama olarak, yapılan yorumun hangi konuyla(Kargo, Satıcı, Ürün Performans vb.) ilgili olduğunu tespit etmektedir. Elde edilen modeller test edilerek başarı oranları, doğruluk oranları analiz edilmiştir. Kullanılan çeşitli algoritmalar performans yönünden karşılaştırılmıştır.

Anahtar Kelimeler : duygu analizi, makine öğrenmesi, doğal dil işleme

Sentiment Analysis on E-Commerce User Reviews via Machine Learning

SUMMARY

Strengthening the interaction of companies with customers in competitive markets provides great benefits in areas such as product development and marketing. Today's technology can offer fast and precise solutions that can save companies from the unwieldiness of traditional methods. This research examined the model to be created by applying various machine learning and language processing techniques on a data set compiled with user comments. This model classifies user reviews as positive, negative, or neutral. As the second step, it determines which subject (Shipping, Seller, Product Performance, etc.) the comment is related to. The obtained models were tested and their success rates and accuracy rates were analyzed. Various algorithms used were compared in terms of performance.

Keywords : sentiment analysis, machine learning, natural language processing

1. GİRİŞ

Günümüzde her sektörde ortaya çıkan memnuniyet ölçümü, rekabet halindeki firmalar için başarıya ulaşmada oldukça önemli bir faktördür. Müşteri memnuniyetini müşterinin aldığı hizmetten memnun kalma düzeyi olarak tanımlayabiliriz. Müşteriler satın aldıkları hizmet veya ürün hakkındaki memnuniyet durumlarını çevresindeki diğer tüketicilerle paylaşma ihtiyacı hissederler. Bu paylaşım ilk zamanlarda sözlü olarak tüketicilere ulaşırken; günümüzde e-ticaretin gelişmesiyle alışveriş sitelerindeki yorumlar üzerinden ilerlemektedir. Kullanıcılar satın aldıkları veya kullanma fırsatı elde ettikleri ürünler hakkındaki deneyimlerini ürün sayfasında paylaşabilirler. Bu paylaşım kimi zaman ürünle ilgili kimi zaman ürünün satıcısıyla ilgili olabilir.
2022 yılı verilerine baktığımız zaman internet kullanıcılarının yüzde 58.4'ü her hafta çevrimiçi bir alışveriş yapıyor. Bu oran düşünüldüğü zaman e-ticaret deneyimi elde eden kullanıcı sayısı, veri kaynağının miktarı ve çeşitliliği gün geçtikçe artmaktadır. Çeşitli kullanıcı yorumlarını, herhangi bir bilgi işletim sisteminin süzgecinden geçirmeden; takip etmek, analiz etmek ve ticari anlamda faydaya çevirmek mümkün değildir. Bu noktada müşteri memnuniyetini ölçmek için geleneksel pazarlama yöntemlerin yetersiz kaldığı alanlarda otomatize edilmiş programlara ihtiyaç vardır. Bu ihtiyacı karşılamak için; makine öğrenmesiyle eğitilmiş modeller, analiz sürecini daha az kaynakla daha hızlı bir şekilde tamamlayabilir.

1.1 Tezin Amacı

Bu çalışma e-ticaret sitelerindeki kullanıcı yorumları ile oluşturulacak veri setini hazırlamayı, doğal dil işleme ve makine öğrenmesi yöntemleriyle duygu analizi ve kategorizasyon tahminleri yapabilecek modeller geliştirmeyi ve bu modelleri karşılaştırmayı amaçlamaktadır. Tez çalışması süresince izlenecek adımlar aşağıdaki gibidir:

- Veri setini internet üzerinden elde etmek:

  - Scrapping

- Veri setinini işlemek:
  - Tokenizasyon
  - Lemmatizasyon
  - Word Correction
  - Anlam ifade etmeyen kelimelerin temizlenmesi

- Makine öğrenmesi modelini oluşturmak:
  - Duygu Analizi
  - Sınıflandırma

- Sonuçların incelenmesi

1.2 Literatür Araştırması

Duygu analizi oldukça popüler bir araştırma konusu olmasına rağmen yapılan çalışmaların ingilizce metin ağırlıklı olmasından dolayı, türkçe metinler üzerinde makine öğrenmesi ile birlikte kullanımı araştırmaya müsait bir alandır.
[1] Bu çalışmada veri seti olarak film yorumlarının bulunduğu Internet Movie Database arşivi kullanılmıştır. İlgili veri seti üzerinde unigram, bi-gram, ikili birleşimler, POS gibi öznitelik çıkarma yöntemlerini kullanarak model için gerekli vektör uzay modellerini oluşturulmuştur. Bu modelleri kullanarak oluşturulan veri seti ile NB ve SVM gibi algoritmalar kullanarak sınıflandırma işlemi yapılmıştır. Bu çalışmalar sonucunda yapılan analizlerde en iyi sonuç %82,9 oranıyla unigram ile hazırlanmış veri setinin SVM yöntemiyle eğitilmesi ile elde edilmiştir.
Go, Bhayani ve Huang, 2009 yılında yaptıkları çalışmalarında [2] veri seti olarak Twitter verilerini kullanarak denetimli öğrenme yöntemiyle duygu analizi yapmayı amaçlamışlardır. Çalışmadaki veri seti toplam 1 milyon 600 bin tweetten oluşmaktadır. Bu tweetlerin 800 bini olumlu 800 bini olumsuz olacak şekilde seçilmiştir. Oluşturulan veri seti bigram, unigram ve bigram-unigram yöntemleri bir arada kullanılarak analiz edilmiştir. Makine öğrenmesi algoritmalarından NB, SVM, Maximum Entropi sınıflandırma yöntemleri kullanılmıştır. Çalışma sonucunda NB ve bi-grama göre

%81,6 doğruluk oranı, SVM ve unigrama göre %82,2 doğruluk oranı elde edilmiştir. En iyi performans Maximum Entropi ve unigram ile bi-gramın birlikte kullanılması sonucu %83 doğruluk oranı ile elde edilmiştir.

Pazarlama sektöründe, müşterilerin firmalar hakkında görüşlerinin analizi amacıyla Twitter ve duygu analizi tekniklerinin beraber kullanıldığı çeşitli çalışmalar bulunmaktadır. Mostafa [3], telekomünikasyon, hava yolu ve sağlık sektörü gibi çeşitli sektörlerden firmaları için Twitter paylaşımlarında ilgili firma adlarının yanında en sık kullanılan kelimeleri belirlemeye çalışmıştır. He ve arkadaşları [4] belirli bir sektördeki üç firmanın Facebook ve Twitter üzerinden elde ettikleri veriler ile müşteri memnuniyetini ölçmeye çalışmışlardır. Takipçi, yorum sayıları, paylaşım sayıları, iletiler belirli yöntemler kullanılarak değerlendirilmiştir. Ghiassi [5] ise literatürdeki modellerden farklı olarak geliştirdikleri algoritmalar ile Twitter'dan bir marka hakkında müşteri yorumlarını derlemiş ve bu verilerle duygu analizi yapmışlardır.

Türkçe veri setleri üzerinde de duygu analiziyle ilgili olarak çalışmalar mevcuttur. Nizam ve Akın [6], sınıflar içerisindeki veri dağılımının sınıflandırma algoritmasındaki başarı oranına etkisini araştırmıştır. Twitter üzerinden toplanan veriler dengeli ve dengesiz olmak üzere 2 gruba ayrılmıştır. Bu 2 farklı veri seti olumlu, olumsuz ve tarafsız olmak üzere 3 ayrı sınıfa ayrılarak etiketlenmiştir. Gıda sektöründeki bazı firmaların farklı ürünlerine ait tweetlerinden Dengesiz veri seti için 1113 olumlu, 277 olumsuz ve 610 tarafsız veri olmak üzere toplam 2000, dengeli veri seti içinse 257 olumlu, 277 olumsuz ve 299 tarafsız veri olmak üzere toplam 824 tweetten içeren bir veri kümesi incelenmiştir. NB, RF, SVM, KNN ve DTs sınıflandırma algoritmaları kullanılarak doğruluk, kesinlik, duyarlılık ve kappa istatistiği sonuçlarına göre 2 veri kümesi karşılaştırılmıştır. Bu modellerin başarı oranlarına göre karşılaştırıldığında dengesiz veri seti dengeli veri setine göre daha kötü performans göstermiştir. Yani sınıflandırma algoritmalarının başarı oranını etkileyen unsurlardan birisi, incelenen sınıflardaki veri dağılımlarıdır. Bu çalışmada %72,33 doğruluk oranı ile en iyi performansı SVM sınıflandırma algoritması sağlamıştır.

Fatih Samet Çetin ve Gülşen Eryiğit'in yapmış oldukları çalışmada [7] veri seti olarak restoran yorumlarından oluşan Türkçe metinler üzerinde Türkçe hedef tabanlı duygu analizi yöntemi kullanılmıştır. Hedef tabanlı duygu analizinde kişilerin herhangi bir

ürün hakkındaki olumlu veya olumsuz düşüncesini tespit etmek yerine, o ürünün bir özelliğiyle ilgili memnuniyetini ölçmek amaçlanmaktadır. Hedef tabanlı duygu analizinde, bir tanımlama grubu kullanılır. Bu tanımlama grubu temel olarak dört bilgi barındırmaktadır: Hedef kategori belirleme,Hedef terim belirleme, Hedef kategori ve hedef terimin aynı anda belirlenmesi ve Duygu sınıfı belirleme. Hedef kategori, hedef terim ve ikisinin aynı anda belirlenmesi görevleri için, kelime vektörleri ve doğal dil işleme çıktıları kullanan koşullu rastgele alanlara (CRF) dayalı bir dizilim etiketleme algoritması tasarlanmış ve her üç görevi de tek aşamada çözebildiği gösterilmiştir. Elde edilen sonuçlar ile bu ilk üç görev için literatürdeki en yüksek başarımların elde edildiği görülmüştür: Hedef kategori belirlemede %66,7 F1-skoru, hedef terim belirleme %53,2 F1-skoru, hedef kategori ve hedef terimin aynı anda belirlenmesinde %46,7 F1-skoru. Bunun yanı sıra, duygu sınıfı belirleme için cümle analizi sonucunda hedef terime yakın kelimelerden özellik seçimine dayalı bir lineer sınıflandırma yöntemi sunulmuş ve literatürde sınırlı sistemler tarafından raporlanan en başarılı sonuç (%76,1 F1-skoru) elde edilmiştir. Bu çalışmada Türkçe hedef tabanlı duygu analizi çalışmalarında, hedef terim ve hedef kategori bulma görevlerinde, kelime vektörlerinin pozitif etkileri ve Türkçe'nin serbest dizilimli yapısı sebebiyle ortaya çıkan problemleri aşmak amacı ile tüm alt görevlerde bağlılık ayrıştırıcı kullanmanın faydaları ortaya çıkarılmıştır.

Meral ve Diri [8] , twitter üzerinden oluşturulan veri seti üzerinde 9 farklı kategori için algı analizi çalışması yapmıştır. Bu çalışmada doğal dil işleme ve NB, RF, SVM gibi makine öğrenmesi yöntemleri kullanılmıştır.2-gram, 3-gram ve kelime tabanlı yöntemler ile eğitilen sistemlerde dokuz farklı alan ve bütün alanları içeren veri setleri kullanılmıştır. Destek Vektör Makinesinin alanların çoğunda ve alan bağımsız veri setinde en iyi f-ölçüm değerlerini verdiği görülmüştür. Bazı alanlarda Naive Bayes başarılı olsa da Rasgele Orman ve diğer yöntemlerden daha az başarılıdır. Dokuz farklı alan ve birleşiminden oluşturulan veri setleri ile sistem eğitildiğinde kelime tabanlı yöntem kullanıldığında %89.5'lik başarı alınmıştır. N-gram yöntemi, alan bağımlı verilerde %90 gibi yüksek başarı vermesine rağmen, alan bağımsız verilerde aynı başarıyı gösterememiştir. Bu çalışmada kelime tabanlı bir doğal dil işleme sürecinden geçirilen Twitter verileri ile sınıflandırıcılar eğitilmiş ve %90'a yakın başarılı sınıflandırma performansı gösteren bir sistem geliştirilmiştir.

2. BİLİMSEL ARKAPLAN

2.1 Doğal Dil İşleme

Sosyal medyanın ve internetin yaygınlaşmasıyla beraber metinler üzerindeki duygunun ve fikrin araştırılması borsa, medya, politika gibi alanlarda pek çok yönde fayda sağlamıştır. Ürün kampanya ve reklamlarının kullanıcı üzerindeki etkisi, politikacıların kampanyalarına halkın tepkisi, ürün yorumlarının tedarik zinciri üzerindeki geri besleme yaratma potansiyeli doğal dil işleme modellerinin önemini ortaya çıkarıyor.

2.2 Türkçe Metinler Üzerinde Duygu Analizi

Türkçe metinler üzerinde duygu analizi son zamanlarda oldukça popülerleşen bir konu. Bu yönde, ingilizce dilinde geliştirilen pek çok algoritma ve kütüphane vardır. Bu çalışmada belirtilen eksikliğe ithafen alışveriş siteleri üzerindeki yorumlar ile bir makine öğrenmesi modeli eğitilecektir. Bu model ürün yorumlarının duygu analizini yapacaktır. Bu analiz sonuçları grafikler halinde bir uygulama ile gösterilecektir.

2.3 Türkçe Metinler Üzerinde Kategori Analizi

Metinleri duygu açısından sınıflamak her zaman istediğimiz analizleri yapmaya yetmez. Bu bağlamda metinlerin hangi kategori ile ilgili olduğunun belirlenmesi ile birlikte daha da anlamlanacaktır.

2.4 Makine Öğrenmesi

Makine öğrenmesi esas olarak 1950'li yıllarda bilgisayar biliminin yapay zekada sayısal öğrenme ve model tanıma çalışmalarından geliştirilmiş bir alt dalıdır. Makine öğrenmesi yapısal işlev olarak öğrenebilen ve veriler üzerinden tahmin yapabilen algoritmaların çalışma ve inşalarını araştıran bir sistemdir. Bu tür algoritmalar statik program talimatlarını harfiyen takip etmek yerine örnek girişlerden veri tabanlı

tahminleri ve kararları gerçekleştirebilmek amacıyla bir model inşa ederek çalışırlar [9].

Makine öğremesi öğrenme biçimine göre üç başlıkta toplanabilir. Bunlar;

- Gözetimli Öğrenme (Supervised Learning)
- Gözetimsiz Öğrenme (Unsupervised Learning)
- Pekiştirmeli Öğrenme (Reinforcement learning)

Makine Öğrenmesi ile çözülebilecek problem türleri;

- Metin ve Belge Sınıflandırması
- Doğal Dil İşleme
- Bilgisayarlı Görü
- Hesaplamalı Biyoloji

Makine öğrenmesinin birçok uygulama alanı var ve bunlara sürekli yenileri ekleniyor. İşletmelerin makine öğrenmesi projeleriyle elde ettiği en önemli avantajlardan bazıları şunlardır:

- İçgörüleri ortaya çıkarma : Makine öğrenmesi, hem yapılandırılmış hem de yapılandırılmamış verilerde desenin tanımlanmasına ve verilerin anlattığı hikayelerin belirlenmesine yardımcı olabilir.
- Veri bütünlüğünü geliştirme : Veri madenciliği konusunda oldukça iyi olan makine öğrenmesi, bunu bir adım ileriye taşıyarak zamanla özelliklerini geliştirebilir.
- Kullanıcı deneyimini geliştirme : Uyarlamalı arabirimler, hedefli içerikler, sohbet botları ve sesli sanal yardımcılar, makine öğrenmesinin müşteri deneyimini iyileştirmeye nasıl yardımcı olabileceğini gösteren örneklerdir.
- Risk azaltma : Dolandırıcılık taktikleri sürekli olarak değiştikçe, makine öğrenmesi de buna ayak uydurur. Makine öğrenmesi, dolandırıcılık denemeleri başarılı olmadan önce bunları yakalamak için yeni desenleri izler ve belirler.

- Müşteri davranışlarını tahmin etme : Makine öğrenmesi, desen ve davranışları belirlemenize yardımcı olmak için müşteriyle ilgili verileri araştırıp ürün önerilerini iyileştirmenize olanak tanır ve mümkün olan en iyi müşteri deneyimini sağlar.

- Maliyetleri azaltma : Bir makine öğrenmesi uygulaması, zamandan ve kaynaklardan tasarruf sağlayarak ekibinizin en önemli şeylere odaklanmasına olanak tanıyan süreç otomasyonudur. [10]

2.5 Derin Öğrenme ve Yapay Sinir Ağları

Derin öğrenme, dijital sistemlerin yapılandırılmamış, etiketlenmemiş verilere dayalı olarak öğrenmesini ve kararlar almasını sağlamak üzere yapay sinir ağlarını kullanan bir makine öğrenmesi türüdür. [11] Yapay Sinir Ağları uygulamaları en çok tahmin, sınıflandırma, veri ilişkilendirme, veri yorumlama ve veri filtreleme işlemlerinde kullanılmaktadır [12]. Bunları tanımlayacak olursak;

- Tahmin: Bu prensipte çalışan yapay sinir ağları girdi değerinden çıktıları tahmin etme üzerine çalışır, örneğin altın ons fiyatının tahmini.

- Veri Filtreleme: Bu doğrultuda kodlanan yapay ağlar toplanan veriler arasından en işe yarayan verileri kullanır.

- Sınıflandırma: Girdi değerlerini sınıflandırarak sistemin daha hızlı sonuca varmasına etkide bulunur.

- Veri Yorumlama: Önceden eğitilen ağ girdilerini analiz eder, bir olay hakkında bu girdiler sayesinde yeni yorumlamalar yapabilmektedir.

- Veri İlişkilendirme: Öğrendiği bilgilerle konuları ilişkilendirir ve bunun sonucunda ortaya çıkan eksik bilgileri tamamlar.

Derin öğrenme, çeşitli alanlarda kullanılır. Örneğin:

- Görüntü, konuşma ve duygu tanıma

- Kişisel dijital yardımcılar

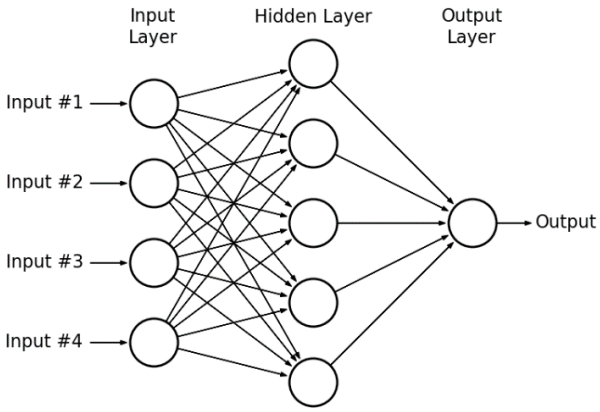
- Sohbet botları

- Sürücüsüz araçlar

Pek çok çeşit sinir ağı vardır.

2.5.1 Multilayer Perceptron (Çok Katmanlı Algılayıcı)

MLP, yapay sinir ağlarının en basit halidir ve sadece tek bir yapay sinir hücresinden oluşur. MLP ise 3 veya daha fazla katman ve çoklu perceptrona sahip olan modellerdir. Genellikle doğrusal olmayan aktivasyon fonksiyonlarını (Tanh ve ReLU gibi) kullanırlar. Kayıp oranını hesaplamak içinse MSE veya Logloss fonksiyonlarını kullanırlar. Bir katmandaki her düğüm, sonraki katmandaki her düğüme bağlanarak ağı tam olarak birbirine bağlar. [13]

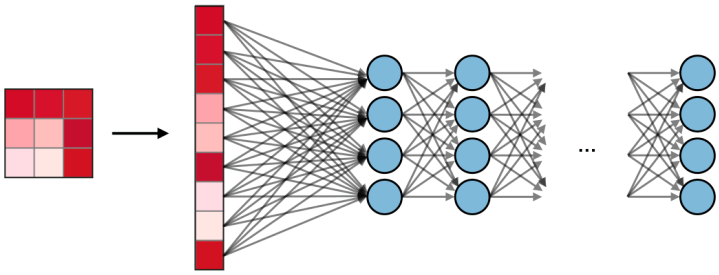


Şekil 2.1 : Multilayer Perceptron
[14]

2.5.2 Convolutional Neural Network

Bir evrişimli sinir ağı bir veya birden fazla evrişimli katman içerir. Evrişimli katmanlarda bir filtre görsel üzerinde gezdirilir. Bu sayede görseldeki belirli özellikler elde edilir. Ardından bir feature matrisi elde edilir. Bu matrise Max-Pooling yardımıyla havuzlama uygulanır ve görseller daha küçük matrislerde saklanır. Bu tekrarlayan katmanların ardından ise fully-connected katmanı sayesinde matrisimiz tek boyutlu bir vektöre dönüştürülür. Günümüzde Evrişimli Sinir Ağları genelde bilgisayarlı görme alanında kullanılır. Görüntü sınıflandırma, video işleme, otonom araçlar gibi pek çok yerde karşımıza çıkar. [13]

2.5.3 Recurrent Neural Network

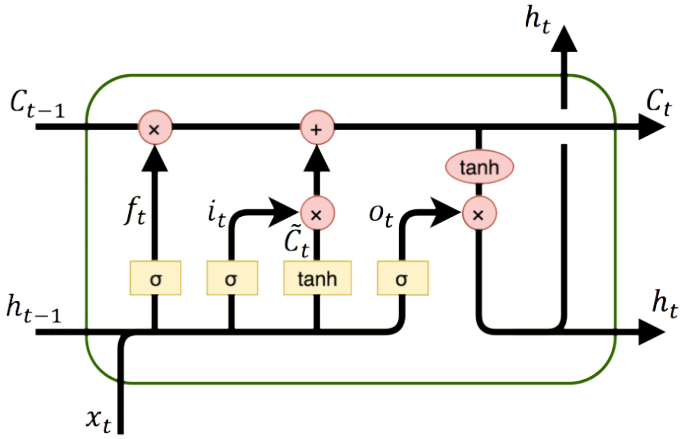


Şekil 2.2 : CNN Bağlantı Katmanı
[15]

RNN, ileri beslemeli bir sinir ağının aksine, nöronlar arasındaki bağlantıların yönlendirilmiş bir döngü oluşturduğu yinelemeli yapay sinir ağının bir çeşididir. Yinelenen sinir ağlarında çıktı sadece mevcut girdilere göre değil, aynı zamanda önceki adımın nöronunun durumuna da bağlıdır. Bu bellek, kullanıcıların el yazısını tanıma veya konuşma tanıma gibi doğal dil işleme sorunlarını çözme de kullanılır. [13]

2.5.4 Long-Short Term Memory

LSTM Ağları, geçici dizileri ve uzun menzilli bağımlılıklarını geleneksel yinelenen sinir ağlarından daha doğru modellemek için tasarlanmış, özel bir yinelenen sinir ağı mimarisidir. LSTM, tekrar eden bileşenleri içinde aktivasyon işlevini kullanmaz, saklanan değerleri değiştirmez. Genellikle, LSTM birkaç birimle “bloklar” halinde uygulanır. Bu bloklar 3 ya da 4 kapıya sahip olur. Örnek vermek gerekirse girdi kapısı, unutma kapısı ve çıkış kapısı. [13]



Şekil 2.3 : LTSM Bağlantı Katmanı
[16]

3. METODOLOJİ

Bu bölümde tez çalışmasında kullanılan veri seti, metotlar, teknolojiler hakkında detaylı bilgi verilmektedir.

3.1 Veri Seti

Çalılşmada üç farklı veri seti kullanılmıştır. Çalışmanın duygu analizi kısmı için internet üzerinden hazır dataset [17] ve e-ticaret sitelerinden scrapping edilerek toplanmış veriler derlenmiştir. Kategori analizi kısmı için ise aynı e-ticaret sitelerinden alınan veriler etiketlenmiş ve derlenmiştir. Bu veri seti içerisinde yaklaşık 80 bin adet veri vardır. Bu veri setleri içerisinden eğitim ve kontrol verileri gruplandırılmıştır. Gruplandırılan veri seti işlemeye hazır hale getirilmek için belirli yöntemlerle işlenmiş ve eğitime hazır hale getirilmiştir. Bu işlemler içerisinde gürültü temizleme, küçük harf dönüşümü, boyut belirleme ve vektörleştirme vardır. Bu işlemlerden sonra veri seti için duygu analizi için olan kısımda 2500 yorum, kategorileştirme kısmında 1350 yorum kullanılmıştır. Duygu analizi için kullanılan dataset içerisinde olumlu, olumsuz, nötral olarak üç sınıf belirlenmiştir. Kategori analizi için ise Performans, Paketleme ve Gizlilik, Servis, Satıcı, Fiyat, Kargo olmak üzere altı sınıf içermektedir. Bu kategoriler içerisinde doğruluğun artması için veri sayısı birbirine en yakın grup belirlenmiştir.

Tablo 3.1 : Duygu Analizi Veri kümesi

Duygu	Veri Sayısı	Eğitim Veri Sayısı
Olumlu	47546	1100
Nötral	11382	300
Olumsuz	5124	1100

Tablo 3.2 : Kategori Analizi Veri kümesi

Kategori	Veri Sayısı	Eğitim Veri Sayısı
Performans	49689	500
Kargo	13470	500
Fiyat	8136	500
Satıcı	5649	500
Servis	236	150
Paketleme ve Gizlilik	6939	500

3.2 Ön İşleme

Veri setinin işlenebilmesi veya bir model eğitilmesi için bir dizi işlemden geçirilmesi gerekir. Bu nedenle model üzerinde çalışılmadan önce bazı işlemler yapılması gereklidir.

3.2.1 Gürültü Temizleme

Bu aşamada metin içinde yer alan noktalama işaretleri metinden atılır. Metinden atılması gereken diğer grup ise İngilizcede "Stop Words" olarak bilinen eklerdir. Türkçede faydasız kelimeler olarak adlandırılan bağlaç gurubu bu ekler öğrenme modelinin eğitiminde etkisi olmadığı için işlem gücünün gereksiz kullanılmaması amacıyla metinden atılır.

3.2.2 Dönüşüm İşlemleri

Veri setinde büyük/küçük harfler oldukça yoğun bir şekilde mevcuttu. Verilerin eğitimden önce sayısallaştırılmasında hatalı gösterime neden olacağı ve modelin başarımını kötü yönde etkileyeceği için tüm metin genel olarak küçük harflere dönüştürülür.

3.2.3 Boyut Seçimi

Verilerin boyutları web sitelerinden alındığı şekliyle farklı kişiler tarafından yazıldığı için oldukça farklıydı. Bu verilerin sep token'ları ile belirlenmiş bir uzunluğa eşit olması sağlanır.

3.2.4 Dil ve Emoji Tespiti

Veriler içerisinde farklı dillerde girilen ve ya emojiler ile birlikte yapılan yorumlar modelin eğitiminde olumsuz etkilere sahip olabileceğinden temizlenerek dengeli ve anlamlı bir veri seti oluşturulur. Bu temizleme işlemi metnin dilini çözümleyebilen önceden hazırlanmış bir Dil Tanımlama modeli kullanılarak yapıldı. Emojiler ise metin içerisinde tespit edilip boş karakterle değiştirilmiştir.

3.3 Model seçimi

Literatürde, metin tahminleme problemi için kullanılan çeşitli algoritmalar bulunmaktadır. Bunlardan büyük çoğunluğu derin öğrenme ve yapay sinir ağı yöntemleridir. Bu yöntemlerle geliştirilen modellere alternatif olarak bu projede gerçekleştirilmek istenen modellerin gerek uygulama kolaylığı gerek başarım oranları sebebiyle Transformer yardımıyla geliştirilmesine karar verilmiştir. Bu nedenle Google tarafından açık kaynak kodlu olarak paylaşılan transformer modülü olan BERT kullanılmasına karar verilmiştir.

3.3.1 Transformerlar

Transformer, girdi verilerinin her bir bölümünün önemini farklı şekilde ağırlıklandıran, dikkat mekanizmasını benimseyen derin bir öğrenme modelidir. NLP ve CV alanlarında kullanılır. [18]

RNN gibi, transformatörler de çeviri ve metin özetleme gibi görevlere yönelik uygulamalarla doğal dil gibi sıralı giriş verilerini işlemek için tasarlanmıştır. Ancak, RNN'lerin aksine, transformatörler tüm girdiyi bir kerede işler. Dikkat mekanizması, giriş dizisindeki herhangi bir pozisyon için bağlam sağlar. Örneğin, giriş verileri doğal bir dil cümlesiyse, transformatörün her seferinde bir kelimeyi işlemesi gerekmez. Bu, RNN'lerden daha fazla paralelleşmeye izin verir ve bu nedenle eğitim sürelerini azaltır. [19]

Transformerlar, 2017 yılında Google Brain'deki bir ekip tarafından tanıtıldı ve LSTM gibi RNN modellerinin yerini alarak NLP sorunları için giderek daha fazla tercih edilen model haline geliyor. Ek eğitim paralelleştirmesi, daha büyük veri kümeleri üzerinde eğitime izin verir. Bu, Wikipedia Corpus ve Common Crawl gibi büyük dil veri kümeleriyle eğitilmiş ve belirli görevler için ince ayar yapabilen BERT ve GPT gibi önceden eğitilmiş sistemlerin geliştirilmesine yol açtı. [20] [21]

3.3.1.1 Transformer Mimarisi

Dil modelleri özetle verilen bir metinde maskelenmiş kelimeleri tahmin etmeye çalışan yapılardır. Maskelenen veri tek bir kelime olabileceği gibi bir cümle de olabilmektedir. 2017 yılında Vaswani tarafından [19] önerilen ve transformer olarak bilinen dikkat mekanizması dil modeli metin sınıflandırma, makine çevirisi ve metin özetleme gibi

doğal dil işleme uygulamalarında o ana kadar ki en yüksek puanları elde etmiştir. Transformer mimarisi genel olarak encoder ve decoder olmak üzere iki ayrı ağdan oluşur. Encoder katmanlarının çıkışları decoder katmanlarına giriş olarak verilir. Ağ n adet giriş alır ve n adet çıkış üretir. Giriş verisinden hangi veriye daha fazla dikkat edilmesi gerektiğini bulmaya çalışır. Mimari genel olarak şu şekilde çalışır: Giriş olarak kelime embeding vektörlerini alır ve kelimelerin cümle içindeki pozisyon bilgilerini oluşturur. Bir sonraki katmanda hangi kelimeye dikkatin artması gerektiğini hesaplayabilmek için çok başlıklı dikkat mekanizmasını kullanılır. Burada girişteki her bir kelime için query, key ve value vektörleri hesaplanır. Dikkat mekanizmasından sonra giriş vektörü eklenir ve normalizasyon işlemi yapılır. Sonraki katmanda tam bağlı sinir ağı ile çıkış üretilir. Bir önceki katmanda olduğu gibi bu katmanın girişi de çıkışına eklenir ve normalizasyon işlemi yapılır [22]

3.3.2 BERT Modeli

BERT, NLP için açık kaynaklı bir makine öğrenimi çerçevesidir. BERT, önceden eğitilmiş, açık kaynak kodlu bir NLP modeline dayanır. Her kelimenin cümledeki diğer kelimelerle olan ilişkisini anlamaya çalışır. Sağdan sola, soldan sağa giden yüzeysel çift yönlü bir dil işlemesinin tersine daha karmaşık maskeli dil modeli kullanır. Bu modelin eğitilmesi için çok büyük miktarlarda veriler kullanıldı. Bu verinin geliştirilmesi Transformer mimarisi sayesinde mümkün hale geldi.

Herhangi bir NLP tekniğinin amacı, insan dilini doğal olarak konuşulduğu gibi anlamaktır. Bunu yapmak için, modellerin tipik olarak büyük bir özel, etiketli eğitim verisi kullanarak eğitim almaları gerekir. Klasik yöntemler dilbilimciler tarafından verilerin tek tek etiketlemesini gerektirir.

BERT, etiketlenmemiş, düz metinli bir arşiv (Vikipedi'nin tamamı) kullanılarak önceden eğitilmiştir. Etiketlenmemiş metinden denetimsiz öğrenme ve pratik uygulamalarda (Google arama) kullanılsa bile gelişmeye devam ediyor. BERT, sürekli büyüyen aranabilir içerik ve sorgu modeline uyum sağlayabilir ve kullanıcının özellik ve isteklerine göre ince ayar yapabilir. Ön eğitimli Bert dil modelinin son katmanından sonra ek katmanlarla farklı görevler için ince ayar yapılabilmektedir. Örneğin 2'li bir sınıflandırma yapmak için tek hücreli bir sinir ağı veya soru cevap sistemi geliştirmek için cümle boyutunda bir katman eklenmesi yeterlidir. [22]

4. UYGULAMA

Çalışmanın analiz kısmında modellerin çıktılarını görebilmek ve kullanıcı tarafında anlaşılabilir veriler elde edebilmek için bir web projesi geliştirilmiştir. Bu web projesinde veriler grafikler halinde gösterilmiştir. Markalar bazında ürünlere alınan olumlu veya olumsuz yorumlar listelenmiştir. Ürünler hakkındaki yorumların kategorileri markalara göre grafikler halinde listelenmiştir.

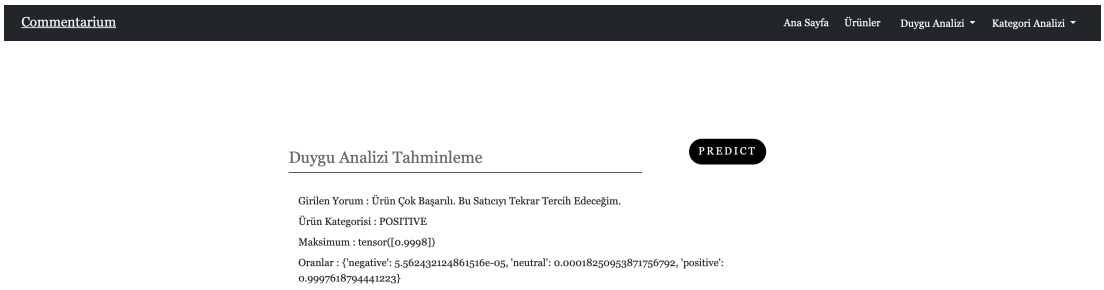
4.1 Web Projesi

Bu proje Python yazılım dilinin 3.10.7 versiyonunda bir web geliştirme kütüphanesi olan Django Framework üzerinde geliştirildi.

Proje üç alt paketten oluşuyor:

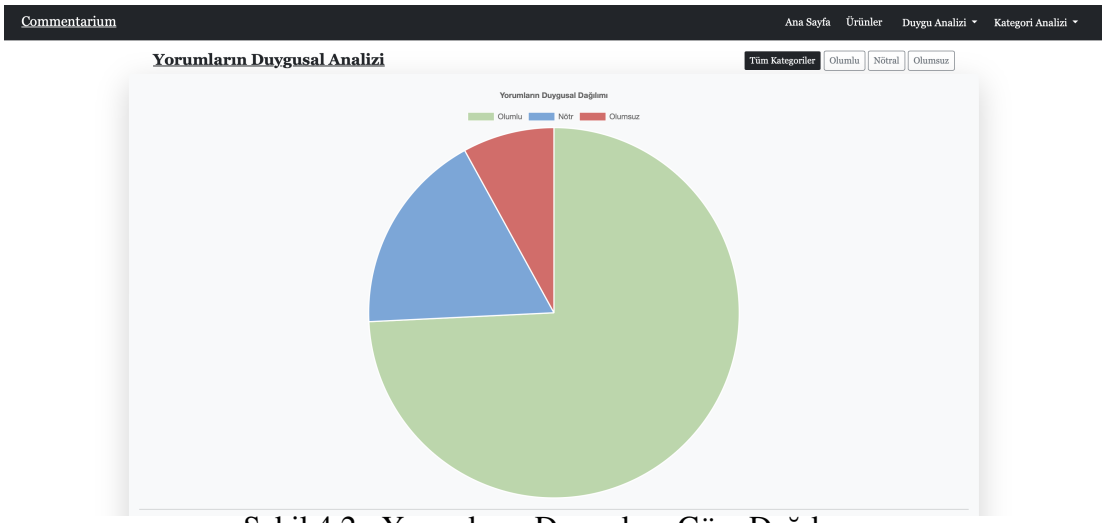
4.1.1 Sentiment Paketi

Yorumların duygu analizi tahminlerini yapan kodların bulunduğu paket burasıdır. Bir metnin duygu yorumlamasını yapmak için geliştirilen tahminleme sayfasında verilen alana girilen metnin model çıktıları yine aynı ekran üzerinde verilmektedir.



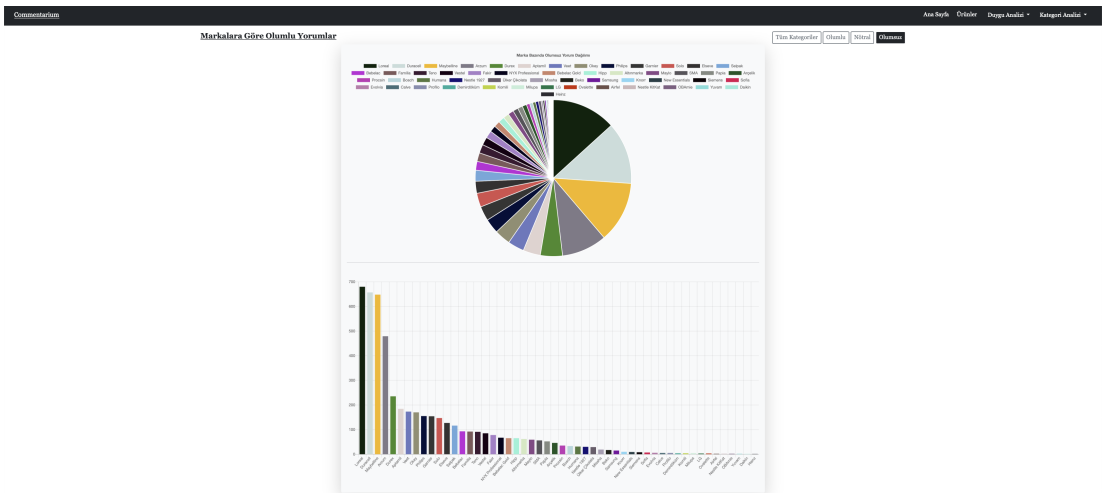
Şekil 4.1 : Duygu Analizi Tahminleme Sayfası

Tüm yorumlar içerisindeki yorumların Olumlu, Nötral ve Olumsuz olarak dağılımını gösteren analizler bu paket altında bulunuyor.



Şekil 4.2 : Yorumların Duygulara Göre Dağılımı

Duygu analizi yapılan yorumların marka bazında gösterimi yine bu paket altında yapılmıştır.

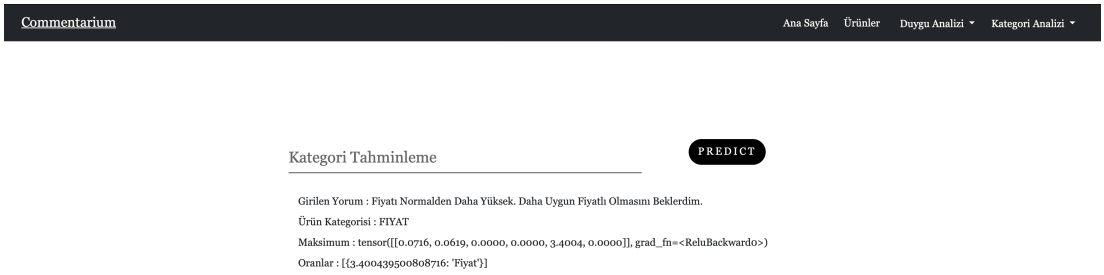


Şekil 4.3 : Marka Bazında Duygu Analizi

4.1.2 Sub Category Paketi

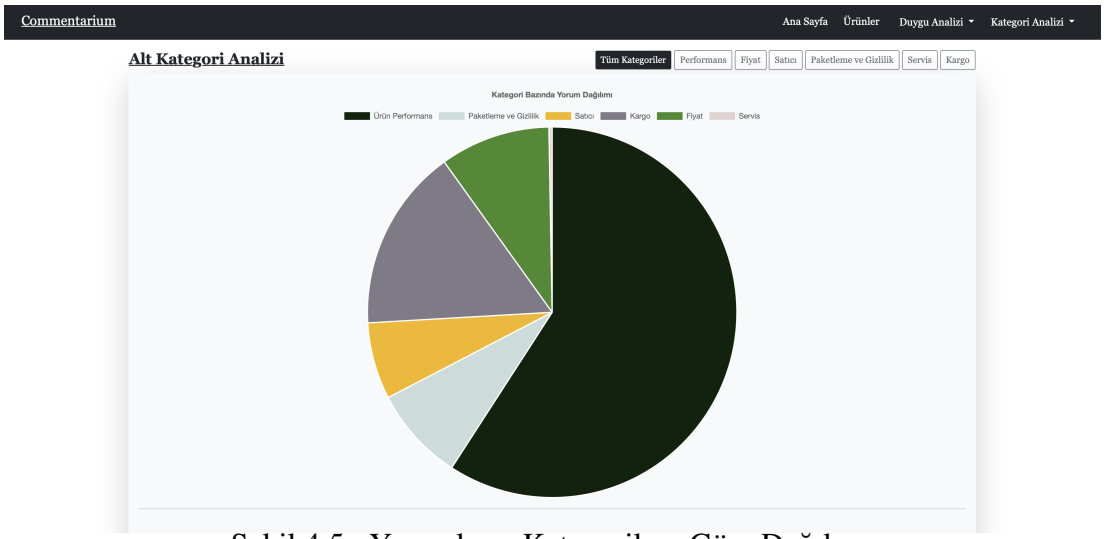
Yorumların kategori analizi tahminlerini yapan kodların bulunduğu paket Sub Category paketidir. Bir metnin kategorilendirmesini yapmak için geliştirilen tahminleme

sayfasında verilen alana girilen metnin model çıktıları yine aynı ekran üzerinde gösterilmektedir.



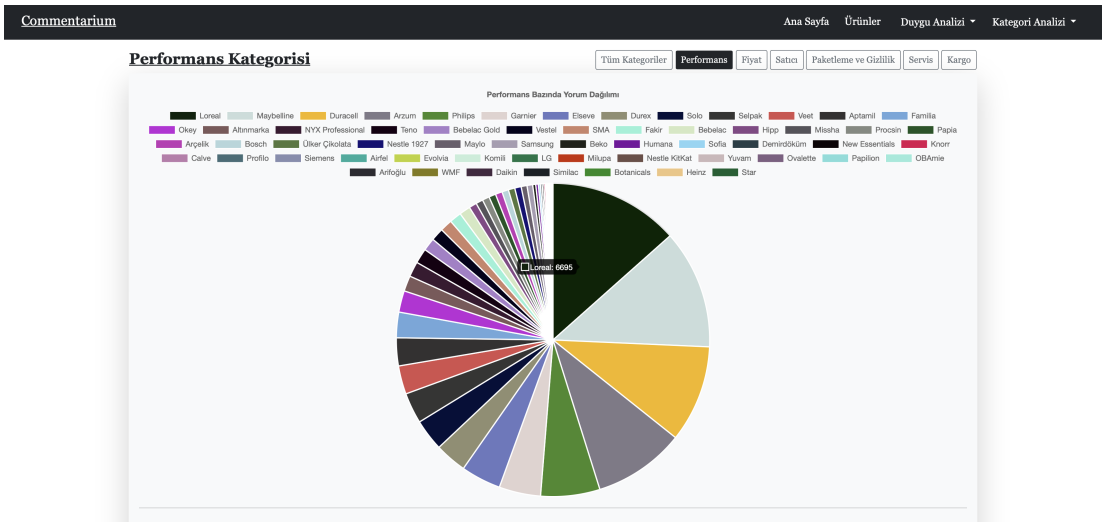
Şekil 4.4 : Kategori Analizi Tahminleme Sayfası

Tüm yorumlar içerisindeki yorumların dağılımını gösteren analizler bu paket altında bulunuyor.



Şekil 4.5 : Yorumların Kategorilere Göre Dağılımı

Duygu analizi yapılan yorumların marka bazında gösterimi yine bu paket altında yapılmıştır.



Şekil 4.6 : Marka Bazında Kategori Analizi

4.1.3 TestDb Paketi

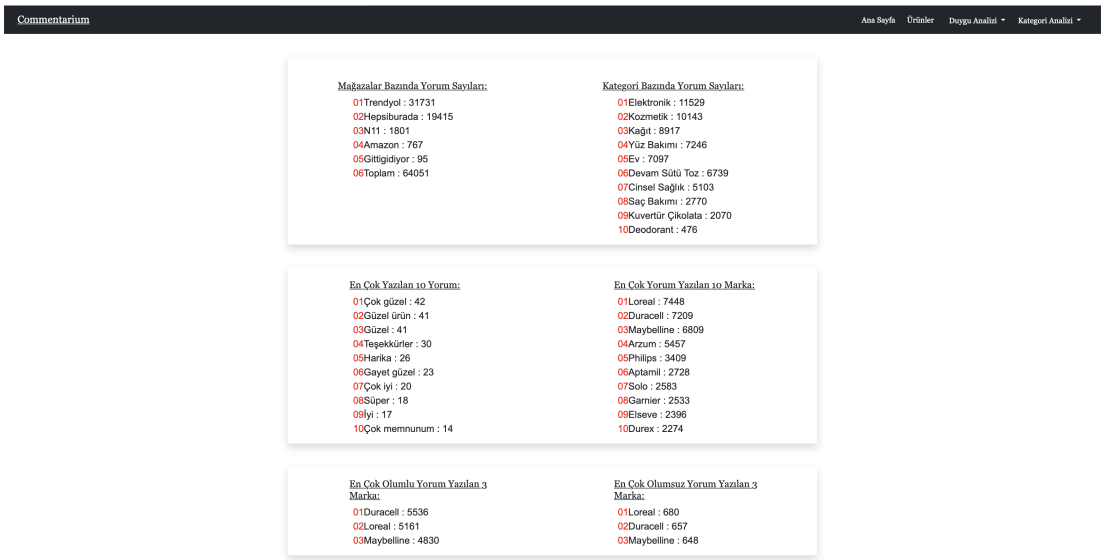
Projenin ana sayfası ve scrap edilen ürünlerin gösterildiği sayfalar bu pakette bulunuyor. Aynı zamanda database tablolarının django tarafında karşılığı olan modeller burada bulunuyor. Ayrıca database'e excel üzerinden veri kaydetmek ve web tarafından ürün yorumlarını çekmek için yine bu paket kullanılıyor. Kısaca projenin üzerine kurulduğu paket olarak kabul edebiliriz. Tüm yorumları ürün ismi, mağaza, marka, puan ve model tarafından doldurulan sentiment ve subcategory kolonları ile beraber bu sayfada görebilliriz.

Commentarium | Ana Sayfa | Ürünler | Duygu Analizi | Kategori Analizi

Id	Ürün İsmi	Mağaza	Puan	Yorum	Sentiment	Sub Category
144170	Selpak Pudralı Kokulu Tuvalet Kağıdı 32 adet	hepsiburada.com	1.0	Zoraki değerlendirme yaptığı için tek yıldız verdim. Yoksa ürün güzel	Nötr	Ürün Performans
144177	Veet Pure Yüz Sir Ağda Bandı 20 adet	Hepsiburada	5.0	Zincir mağazalarda 1kutusuna 69₺ demişlerdi, firmanın kartı varsa fiyatı. Aynı ürün, hiçbir farkı yok. Gayet memnunum.	Olumlu	Ürün Performans, Fiyat
144194	Loreal Değerli Çiçekler Tonik Hassas 400 ml	Trendyol	5.0	Zaten sürekli kullandığım bi ürün memnunum	Olumlu	Ürün Performans
144185	Veet Sir Ağda Bandı Hassas Ciltler 40+40 adet	Trendyol	5.0	zaten veet kalitesi biliniyor gayet güzel indirimden de uygun fiyata almıştım.	Olumlu	Ürün Performans, Fiyat
144223	Loreal Glow Cherie Işıltı Veren Renkli Nemlendirici Medium Glow	Hepsiburada	0.0	Zaten kullandigim bir üründü tekrar aldim cildinize parlaklik isilti veriyor. Satici da hemen kargo yapti tesekkurler ve paketlemesi cok iyiydi koruma bandi vardi uzerlerinde.	Olumlu	Kargo, Satıcı
144195	Solo Tuvalet Kağıdı 32 adet	hepsiburada.com	0.0	Zaten sürekli kullandıgım bir ürün az yaprakla daha hijyen elde parçalanıp dagılmıyor çevremdeki insanlara güvenle tavsiye ediyorum	Olumlu	Ürün Performans
144196	Solo Tuvalet Kağıdı 32 adet	hepsiburada.com	5.0	Zaten sürekli kullandıgım bir ürün az yaprakla daha hijyen elde parçalanıp dagılmıyor çevremdeki insanlara güvenle tavsiye ediyorum	Olumlu	Ürün Performans
144197	Solo Tuvalet Kağıdı 32 adet	hepsiburada.com	5.0	Zaten sürekli kullandıgım bir ürün az yaprakla daha hijyen elde parçalanıp dagılmıyor çevremdeki insanlara güvenle tavsiye ediyorum	Olumlu	Ürün Performans
144174	Maybelline Lash Sensational Yelpaze Etkili Suya Dayanıklı Siyah Maskara	Trendyol	5.0	Zor cikiyor gayet iyi	Olumlu	Ürün Performans
144184	Maybelline Affinitone Fondöten 24 Golden Beige	Trendyol	5.0	Zaten yıllardır kullandığım bir fondöten çok kaliteli ❤️	Olumlu	Ürün Performans
144189	Loreal Infaillible 24H Matte Cover Yüksek Kapatıcı Fondöten 110 Rose Vanilla	Trendyol	5.0	Zaten sürekli kullandığım üründü, hızlı kargo sorunsuz teslimat. Teşekkürler	Olumlu	Kargo, Paketleme ve Gizlilik, Teslimat ve Kurulum

Şekil 4.7 : Ürünler ve Yorumlar Sayfası

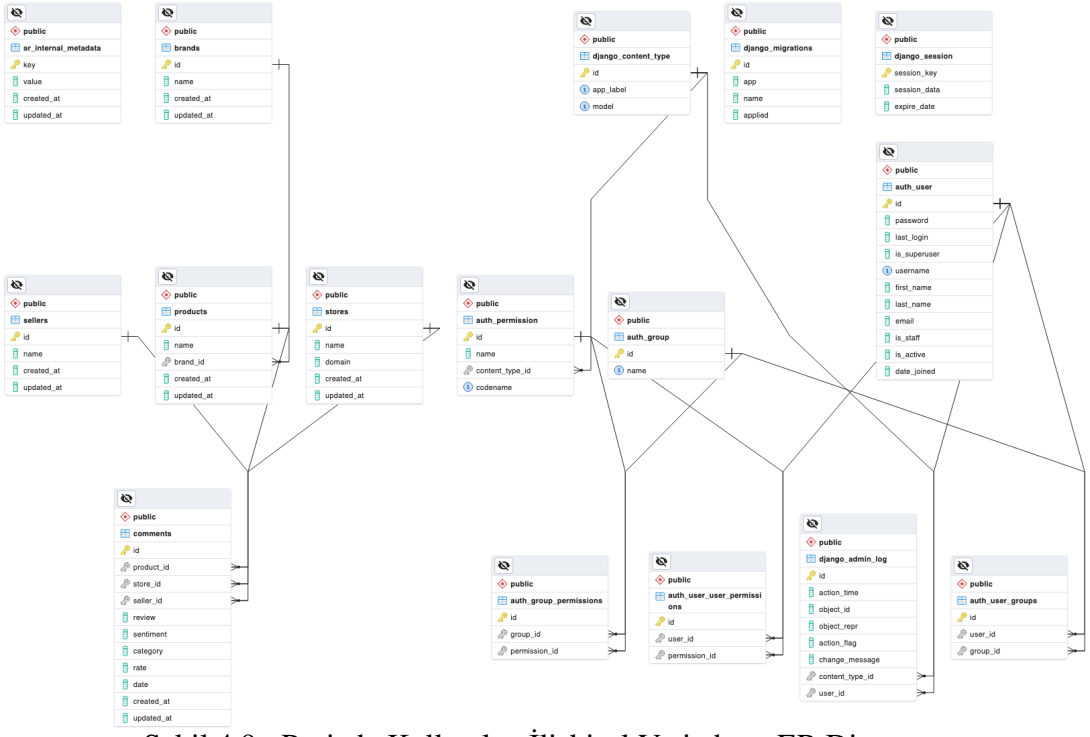
Projenin ana sayfasında projede elde edilen verilerle ilgili bir projeksiyon oluşturulmuştur. Bu sayfada e-ticaret platformları için toplanan ürün yorum sayıları, kategoriler bazında kaç adet yorum toplandığı, markalara bazında en çok yorum alan markalar, en çok olumlu ve olumsuz yorum alan markalar ve en çok yazılan yorumlar gösterilmiştir.



Şekil 4.8 : Proje ile İlgili Genel Bir Projeksiyon

4.2 Database

Veritabanı PostgreSQL üzerinde ilişkisel olarak geliştirildi. Pek çok tablodan oluşan database product, brand, seller, store, comment ve django için özelleştirilmiş tablolardan oluşmaktadır. E-ticaret sitelerinden toplanan veriler database talolarına ilişkisel olarak yazıldıktan sonra çalıştırılan job duygu ve kategori analizlerini yaptıktan sonra comment tablosundaki sentiment ve subcategory kolonlarını dolduruyor.



Şekil 4.9 : Projede Kullanılan İlişkisel Veritabanı ER Diyagramı

5. SONUÇLAR VE DEĞERLENDİRME

5.1 Uygulama

Çalışmada, çeşitli veri setleri kullanılarak türkçe metinler üzerinde NLP ve makine öğrenmesi yöntemleri test edilmiştir. Veri setini model eğitimine hazırlamak için gürültü temizleme, küçük harf dönüşümü, boyut belirleme ve vektörleştirme, verilerin normalizasyonu, yeni özniteliklerin eklenmesi gibi ön işlemler yapılmıştır. Bu işlemler sonucunda tahminleme yapabilecek modeller BERT Transformer kullanılarak eğitilmiştir. Eğitilmiş modeller geliştirilen bir web projesi üzerinde gerçek veriler ile test edilmiş ve gösterilmiştir.

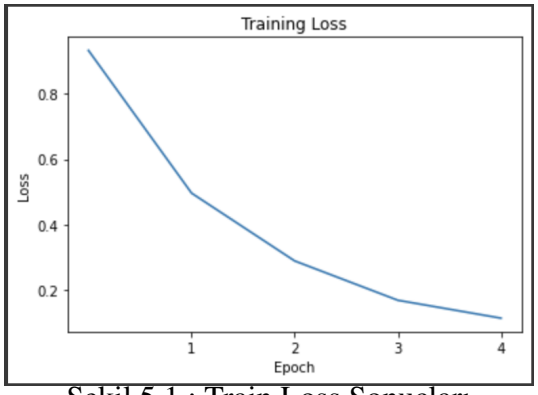
5.2 Sonuçlar

Çalışmanın sonuçları iki başlık altında incelenecektir. Bunlar;

- Makine Öğrenimi Modeli Sonuçları
- Web Projesi Çıktıları ve Analizler

5.2.1 Makine Öğrenimi Modeli Sonuçları

Yorumların duygu analizi tahminlerini yapan kodların bulunduğu paket burasıdır. Bir metnin duygu yorumlamasını yapmak için geliştirilen tahminleme sayfasında verilen alana girilen metnin model çıktıları yine aynı ekran üzerinde verilmektedir.

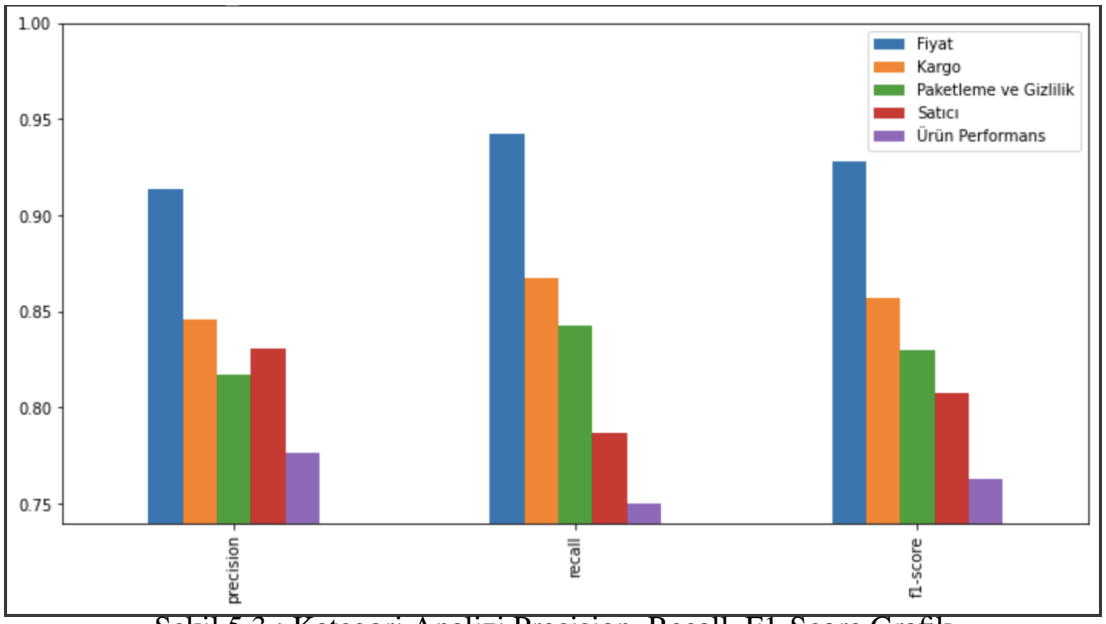


Şekil 5.1 : Train Loss Sonuçları

Şekil 5.1'de modele ait sonuçlar görülmektedir. Burada Train Loss azalışı bize modelin eğitim derecesini gösteriyor. Bu model için. Epoch sayısı 5 olarak belirlendi. Daha yüksek epoch sayısı ile çalışılan modellerde grafiğin bir süre sonra sıfırdan uzaklaştığını görürüz. Bu bize modelin öğrenmek yerine ezberlediğini gösterebilir. Ve bu durumda daha farklı veri setleri ile çalışıldığı zaman başarısız sonuçlarla karşılaşma olasılığı artar. Bu grafik bize epoch sayısını doğru seçmemiz için yardımcı olur.

	Fiyat	Kargo	Paketleme ve Gizlilik	Satıcı	Servis	Ürün Performans	accuracy	macro avg	weighted avg
precision	0.913907	0.845833	0.817276	0.830247	0.75	0.776471	0.841909	0.822289	0.841124
recall	0.941980	0.867521	0.842466	0.786550	0.75	0.750000	0.841909	0.823086	0.841909
f1-score	0.927731	0.856540	0.829680	0.807808	0.75	0.763006	0.841909	0.822461	0.841224
support	293.000000	234.000000	292.000000	342.000000	4.00	176.000000	0.841909	1341.000000	1341.000000

Şekil 5.2 : Kategori Analizi Precision, Recall, F1-Score Sonuçları



Şekil 5.3 : Kategori Analizi Precision, Recall, F1-Score Grafik

Şekil 5.4'de görüldüğü gibi duygu analizinde yüksek bir başarı oranı yakaladık. Bunun en önemli sebebi veri setinin eğitime uygun standartlarda olmasıdır. Veri setinin hazır olarak alınması ve BERT transformer başarımı ile birleşince yüksek başarımlı bir model ortaya çıkmıştır. Kategori analizinde veri setinin standartlara yeterinve uygun olmaması ve yeterince iyi etiketlenememiş olması başarımı

oldukça düşürmüştür. Fakat yine de fazla öğrenmeye kapılmadan kullanıcıyı tatmin edecek sonuçlar vermektedir.

	Negative	Positive	accuracy	macro avg	weighted avg
precision	0.92	0.96	0.94	0.94	0.94
recall	0.95	0.93	0.94	0.94	0.94
f1-score	0.94	0.94	0.94	0.94	0.94

Şekil 5.4 : Duygu Analizi Precision, Recall, F1-Score Grafik

5.2.2 Web Projesi Çıktıları ve Analizler

Geliştirilen web projesi üç alt paketten oluşmantadır. Python Django web geliştirme kütüphanesinde Model-View-Template yapısı üzerinden tüm alt paketlerin geliştirmesi ve PostgreSQL database ile entegre edilmesi sağlanmıştır. Hızlı ve son kullanıcıya hitap eden arayüz tasarımı geliştirilmiştir. Duygu analizi ve Kategori analizi modellerinin bulunduğu paketler geliştirilerek yeni ekranlar tasarlanıp son kullanıcının kullanımına sürülebilir. Ayrıca yeni paketlerin geliştirilmesi ve projeye yeni modüllerin eklenmesi Django framework yapısı nedeniyle çok kolay bir şekilde yapılabilmektedir.

5.3 Genel Değerlendirme

Analizler sonucunda E-Ticaret siteleri üzerindeki yorumların duygu analizi modeli ile tahminlenmesi $\sim$ %94 doğruluk oranıyla; kategori analizi modeli ile tahminlenmesi $\sim$ %84 doğruluk oranıyla gerçekleştirilmiştir. Bu proje geliştirilecek modüllerle beraber, müşteri memnuniyeti gözeten iş çevreleri için büyük bir rehber niteliği taşıyabilir.

5.4 Tartışma

Çalışma sonucunda, dil modellerinin eğitiminin Transformer'lar ile hızlı ve yüksek başarı oranları ile gerçekleştirilebildiği görülmüştür. Kısıtlı veri setleriyle yapılan bu çalışma, kullanıcı yorumlarını günlük olarak toplayıp model çıktılarını kullanıcıya anlık gösteren bir uygulama haline getirildiği zaman gerçek başarısını gösterecektir. Veri setinin seçildiği kategoriler dışındaki ürünlerin de olduğu göz önünde

bulundurulduğunda; daha dengeli ve kapsamlı bir veri seti ile yapılacak çalışmada modellerin günlük hayattaki başarımı artacaktır. Model belirli bir süre kullanılıp veriler alındıkça başarım oranları ve analiz verisi olarak kullanılma kapasitesi daha net anlaşılabilecektir.

Kaynakça

[1] **Pang, B.**, **Lee, L. ve Vaithyanathan, S.** (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques, *EMNLP*, *10*.

[2] **Go, A.**, **Bhayani, R. ve Huang, L.** (2009). Twitter sentiment classification using distant supervision, *Processing*, *150*.

[3] **Mohamed, M.** (2013). More Than Words: Social Networks' Text Mining for Consumer Brand Sentiments, *Expert Systems with Applications*, *40*, 4241–4251.

[4] **He, W.**, **Zha, S. ve li, L.** (2013). Social Media Competitive Analysis and Text mining: A Case Study in the Pizza Industry, *International Journal of Information Management*, *33*, 464–472.

[5] **Ghiassi, M.**, **Skinner, J. ve Zimbra, D.** (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network, *Expert Systems with Applications*, *40*, 6266–6282.

[6] **Nizam, H. ve Akın, S.S.** (2014). Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, *XIX. Türkiye'de İnternet Konferansı*.

[7] **Çetin, F.S. ve Eryiğit, G.** (2018). Türkçe Hedef Tabanlı Duygu Analizi İçin Alt Görevlerin İncelenmesi – Hedef Terim, Hedef Kategori Ve Duygu Sınıfı Belirleme, *Bilişim Teknolojileri Dergisi*, *11*(1), 43 – 56.

[8] **Meral, M. ve Diri, B.** (2014). Sentiment analysis on Twitter, *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, 690–693.

[9] **ALKAN, M.A.**, (2018), Makine öğrenimi Nedir?, `https://www.endustri40.com/makine-ogrenimi-nedir/`.

[10] Makine öğrenmesi Nedir?, `https://azure.microsoft.com/tr-tr/resources/cloud-computing-dictionary/what-is-machine-learning-platform`.

[11] Derin Öğrenme Nedir?, `https://azure.microsoft.com/tr-tr/resources/cloud-computing-dictionary/what-is-deep-learning/`.

[12] **Ağyar, Z.** Yapay Sinir Ağlarının Kullanım Alanları ve Bir Uygulama, *Mühendis ve Makina*, *662*, 22–23.

[13] **Doğan,** , (2020), Derin öğrenme Nedir? Yapay Sinir Ağları Ne işe Yarar?, https://teknoloji.org/derin-ogrenme-nedir-yapay-sinir-aglari-ne-ise-yarar/.

[14] **Hassan, H.**, **Negm, A.**, **Zahran, M. ve Saavedra, O.** (2015). ASSESSMENT OF ARTIFICIAL NEURAL NETWORK FOR BATHYMETRY ESTIMATION USING HIGH RESOLUTION SATELLITE IMAGERY IN SHALLOW LAKES: CASE STUDY EL BURULLUS LAKE., *International Water Technology Journal*, *5*.

[15] CNN, https://stanford.edu/~shervine/l/tr/teaching/cs-230/cheatsheet-convolutional-neural-networks.

[16] LTSM, https://apmonitor.com/do/index.php/Main/LSTMNetwork.

[17] Hugging Face - Dataaset, https://huggingface.co/datasets/winvoker/turkish-sentiment-analysis-dataset.

[18] (2019), Transformers, https://en.wikipedia.org/wiki/Transformer_(machine_learning_model).

[19] **Vaswani, A.**, **Shazeer, N.**, **Parmar, N.**, **Uszkoreit, J.**, **Jones, L.**, **Gomez, A.N.**, **Kaiser, L. ve Polosukhin, I.**, (2017), Attention Is All You Need, https://arxiv.org/abs/1706.03762.

[20] (2019), Derin öğrenme Nedir? Yapay Sinir Ağları Ne işe Yarar?, https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html.

[21] (2019), Daha İyi Dil Modelleri ve Etkileri, https://openai.com/blog/better-language-models/.

[22] **Hanbay, D.** (2021). Ön Eğitimli Dil Modelleri Kullanarak Türkçe Tweetlerden Cinsiyet Tespiti, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *33*(2), 675 – 684.

EKLER

EK A.1 : Çizelgeler
EK A.2 : Algoritmalar

ÖZGEÇMİŞ

Ad Soyad: Faruk GÜRLESİN

Doğum Tarihi ve Yeri: Nisan 1996 - DENİZLİ

E-Posta: fgurlesin19@posta.pau.edu.tr

ÖĞRENİM DURUMU: Lisans Öğrenci

- **Lise:** Denizli Anadolu Lisesi
- **Lisans:** Pamukkale Üniversitesi/Bilgisayar Mühendisliği

MESLEKİ DENEYİMLER VE ÖDÜLLER:

- Publins - Yazılım Geliştirici, Kasım 2020 - Nisan 2022
- Kapsül Bilişim Danışmanlık - Yazılım Uzman Yardımcısı, Nisan 2022 - Halen