**T.C**

**KOCAELİ SAĞLIK VE TEKNOLOJİ ÜNİVERSİTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ**

**YAZILIM LAB DERSİ**

**VERİ BİLİMİ PROJESİ**

**HAZIRLAYANLAR**

**Selime Selin CAN – 210501005**

**Zeynep İrem AKYALÇIN - 210501008**

**DERS SORUMLUSU**

**DR. ÖĞR. ÜYESİ NUR BANU ALBAYRAK**

**26 MAYIS 2024**

İÇİNDEKİLER

[1. AMAÇ 3](#_Toc167649578)

[2.1 Arayüz Gereksinimleri 3](#_Toc167649579)

[2.2 Fonksiyonel Gereksinimler 3](#_Toc167649580)

[3.1 Mimari Tasarım 4](#_Toc167649581)

[3.2 Kullanılacak Teknolojiler 4](#_Toc167649582)

[3.3 Veritabanı Tasarımı 4](#_Toc167649583)

[3.4 Kullanıcı Arayüzü Tasarımı 4](#_Toc167649584)

[4.1 Kodlanan Bileşenlerin Açıklamaları 5](#_Toc167649585)

[4.2 Görev Dağılımı 5](#_Toc167649586)

[4.3 Karşılaşılan Zorluklar ve Çözüm Yöntemleri 5](#_Toc167649587)

[4.4 Proje İsterlerine Göre Eksik Yönler 5](#_Toc167649588)

[SVM Modelinin Performans Metrikleri 6](#_Toc167649589)

[BERT Modelinin Performans Metrikleri 9](#_Toc167649590)

[**Epoch Sonuçları:** 9](#_Toc167649591)

[**Test Sonucu:** 9](#_Toc167649592)

[**BERT ve SVM Modellerinin Karşılaştırılması:** 11](#_Toc167649593)

[**Açıklama:** 11](#_Toc167649594)

[AlexNet Modelinin Performans Metrikleri 13](#_Toc167649595)

[AlexNet Modeli Hata Analizi 13](#_Toc167649596)

[Model Eğitim Yöntemleri 14](#_Toc167649597)

[Performans Değerlendirmesi ve Karşılaştırma 14](#_Toc167649598)

# AMAÇ

Projenin amacı, kullanıcıdan alınan bir cümlenin pozitif, negatif veya nötr duygu içerdiğini belirleyen bir yapay zeka tabanlı duygu analizi programı geliştirmektir. Bu program, farklı makine öğrenmesi ve derin öğrenme modellerini kullanarak duygu analizi yapabilmelidir.

* Kullanıcıdan cümle alarak duygu analizini gerçekleştiren bir arayüz oluşturulması.
* Üç farklı model (Support Vector Machine, AlexNet ve BERT) kullanarak duygu analizi yapılması.
* Modellerin eğitiminde "Twitter Sentiment Analysis" veri setinin kullanılması.
* Modellerin performanslarının doğruluk, kesinlik, duyarlılık ve F-measure metrikleri ile değerlendirilmesi.

1. **GEREKSİNİM ANALİZİ**

### 2.1 Arayüz Gereksinimleri

* **Kullanıcı Arayüzü Gereksinimleri:**
  + Kullanıcıdan cümle alabilen bir giriş alanı.
  + Kullanıcının duygu analizi için model seçimini yapabileceği bir seçenek alanı.
  + Analiz sonuçlarını gösteren bir çıktı alanı.
* **Donanım Arayüzü Gereksinimleri:**
  + Donanım gereksinimi yoktur; yazılım tamamen yazılım tabanlı çalışacaktır.

### 2.2 Fonksiyonel Gereksinimler

* Kullanıcıdan cümle alma.
* Kullanıcının model seçimi yapabilmesi.
* Seçilen modelin eğitilmesi ve kullanıcıdan alınan cümlenin analizi.
* Analiz sonuçlarının kullanıcıya gösterilmesi.
* Modellerin eğitiminde ve testinde kullanılan metriklerin raporlanması (doğruluk, kesinlik, duyarlılık ve F-measure).

1. **TASARIM**

### 3.1 Mimari Tasarım

* Kullanıcıdan cümle alma ve model seçimi arayüzü.
* SVM, AlexNet ve BERT modellerinin eğitimi ve tahmin fonksiyonları.
* Performans değerlendirme ve raporlama modülü.

### 3.2 Kullanılacak Teknolojiler

* **Yazılım Dili:** Python
* **Harici Kütüphaneler:**
  + scikit-learn (SVM ve metrikler için)
  + transformers (BERT modeli için)
  + tensorflow (BERT ve AlexNet için)
  + numpy ve pandas (veri işlemleri için)
* **Diğer Teknolojiler:**
  + Komut satırı arayüzü

### 3.3 Veritabanı Tasarımı

* Veri tabanı kullanılmamaktadır; eğitim verisi dosya tabanlıdır.

### 3.4 Kullanıcı Arayüzü Tasarımı

* Kullanıcıdan cümle almayı ve model seçimini sağlayan basit bir komut satırı arayüzü geliştirilmiştir.
* Programın başlatılması ve cümlelerin analiz edilmesi komut satırından yapılmaktadır.

1. **UYGULAMA**

### 4.1 Kodlanan Bileşenlerin Açıklamaları

**okuma()**: Veriyi CSV dosyasından okur.

**svm\_model()**: SVM modelini seçer ve eğitimini yapar.

**svm\_model\_egit()**: SVM modelini eğitir ve metrikleri hesaplar.

**cümle\_tahmin\_et\_SVM()**: SVM modeli ile cümle tahmini yapar.

**large\_model\_egit()**: BERT modelini eğitir ve metrikleri hesaplar.

**cümle\_tahmin\_et\_Large()**: BERT modeli ile cümle tahmini yapar.

**alexnet\_model\_egit()**: AlexNet modelini eğitir ve metrikleri hesaplar.

**cümle\_tahmin\_et\_alexnet()**: AlexNet modeli ile cümle tahmini yapar.

**menu()**: Kullanıcı menüsünü gösterir.

**main()**: Programın ana fonksiyonudur.

### 4.2 Görev Dağılımı

Projenin her aşaması ortaklaşa yapılmıştır.

### 4.3 Karşılaşılan Zorluklar ve Çözüm Yöntemleri

* **Model Eğitimi**: Büyük veri setleri ile çalışırken model eğitimi süresi uzayabilmektedir. Çözüm olarak eğitim süresi optimize edilmiştir.
* **Metrik Hesaplama**: Farklı metriklerin doğru hesaplanabilmesi için sklearn kütüphanesi kullanılmıştır.

### ****4.4 Proje İsterlerine Göre Eksik Yönler****

* AlexNet, aslında görüntü işlemede kullanılan bir modeldir. Metin verileri üzerinde kullanımı uygun değildir ve bu proje için alternatif bir model kullanılmalıdır.

#### 5. TEST VE DOĞRULAMA

**5.1 Yazılımın Test Süreci**

* **Test Kapsamı**: Tüm modeller için eğitim ve tahmin fonksiyonları test edilmiştir.
* **Test Metodolojisi**: Her modelin performansı doğruluk, kesinlik, duyarlılık ve F-measure metrikleri ile değerlendirilmiştir.

**5.2 Yazılımın Doğrulanması**

### SVM Modelinin Performans Metrikleri

Support Vector Machine (SVM) modeli, çeşitli performans metriklerine göre değerlendirildi. Aşağıdaki çubuk grafik, performans skorlarını göstermektedir:

* **Model Doğruluğu (Accuracy):** 0.9388
  + Bu değer, modelin doğru tahminlerinin toplam tahminlere oranıdır. %93.88 doğruluk oranına sahiptir.
* **Precision:** 0.9411
  + Bu değer, modelin pozitif olarak tahmin ettiği örneklerden kaçının gerçekten pozitif olduğunu gösterir. Yani, yanlış pozitiflerin oranını belirtir. %94.11 precision oranına sahiptir.
* **Recall:** 0.9388
  + Bu değer, gerçek pozitif örneklerin model tarafından ne kadarının doğru tahmin edildiğini gösterir. Yani, yanlış negatiflerin oranını belirtir. %93.88 recall oranına sahiptir.
* **F-measure (F1 Skoru):** 0.9374
  + Precision ve recall'un harmonik ortalamasıdır. Bu metrik, modelin genel performansını değerlendirmek için kullanılır. %93.74 F1 skoru oranına sahiptir.

ekran görüntüsü, metin, renklilik, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bu grafiği oluşturmak için Python dilinde yazdığımız kod aşağıdaki görselde bulunmaktadır.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bu metrikler, SVM modelinin farklı değerlendirme kriterlerinde tutarlı bir şekilde iyi performans gösterdiğini göstermektedir.

SVM modelinin sonuçlarının kullanıcı arayüzünde nasıl gözüktüğü aşağıdaki görselde bulunmaktadır.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

### BERT Modelinin Performans Metrikleri

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Görselde bir BERT modelinin eğitim ve doğrulama sonuçları yer alıyor. Bu sonuçlar modelin eğitim sürecindeki performansını gösteriyor. Aşağıda her bir epoch için kayıp (loss) ve doğruluk (accuracy) değerleri verilmiş. Model 3 epoch boyunca eğitilmiş ve test doğruluğu da verilmiş.

**Epoch Sonuçları:**

1. Epoch 1:
   * Eğitim kaybı: 1.0949
   * Eğitim doğruluğu: 0.5400
   * Doğrulama kaybı: 0.6840
   * Doğrulama doğruluğu: 0.7857
2. Epoch 2:
   * Eğitim kaybı: 0.3199
   * Eğitim doğruluğu: 0.9082
   * Doğrulama kaybı: 0.2768
   * Doğrulama doğruluğu: 0.9218
3. Epoch 3:
   * Eğitim kaybı: 0.0666
   * Eğitim doğruluğu: 0.9821
   * Doğrulama kaybı: 0.1948
   * Doğrulama doğruluğu: 0.9490

**Test Sonucu:**

* Test doğruluğu: 0.9489796161651611

Bu sonuçlar modelin epoch'lar boyunca giderek daha iyi performans gösterdiğini ve doğrulama doğruluğunun her epoch'da arttığını göstermektedir. Son epoch'da modelin doğrulama doğruluğu %94.90'a ulaşmıştır.

Aşağıdaki görsellerde modelin grafik görseli ve grafiği oluşturmak için yazdığımız kod bulunmaktadır.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**BERT ve SVM Modellerinin Karşılaştırılması:**

**BERT Modeli:**

* **Doğruluk (Accuracy):** 0.9490
* **Precision:** Bu metrik görselde yer almıyor.
* **Recall:** Bu metrik görselde yer almıyor.
* **F-measure:** Bu metrik görselde yer almıyor.

**SVM Modeli:**

* **Doğruluk (Accuracy):** 0.9388
* **Precision:** 0.9411
* **Recall:** 0.9388
* **F-measure:** 0.9374

**Açıklama:**

* **Doğruluk (Accuracy):** BERT modeli (%94.90) SVM modeline (%93.88) göre daha yüksek doğruluk oranına sahiptir. Bu, BERT modelinin genel olarak daha fazla doğru tahmin yaptığını gösterir.
* **Precision:** SVM modelinin precision değeri %94.11. Bu metrik BERT modeli için mevcut değil, bu yüzden karşılaştırma yapılamaz.
* **Recall:** SVM modelinin recall değeri %93.88. Bu metrik BERT modeli için mevcut değil, bu yüzden karşılaştırma yapılamaz.
* **F-measure:** SVM modelinin F-measure değeri %93.74. Bu metrik BERT modeli için mevcut değil, bu yüzden karşılaştırma yapılamaz.

Görsellerdeki verilere göre doğruluk (accuracy) metriklerini karşılaştırabiliriz. Ancak diğer metrikler (precision, recall, f-measure) BERT modeli için mevcut olmadığından dolayı tam bir karşılaştırma yapmak mümkün değil.

ekran görüntüsü, metin, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu



### AlexNet Modelinin Performans Metrikleri

### AlexNet Modeli Hata Analizi

AlexNet modelinin eğitimi sırasında aşağıdaki ekranda gösterildiği gibi bir hata ile karşılaşıldı:

metin, ekran görüntüsü, yazı tipi, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Hata mesajı, Tokenizer sınıfının tanımlanmadığını göstermektedir. Bu, eksik bir import ifadesi veya kodda bir yazım hatası olduğunu işaret eder. Bu sorunu çözmek için, doğru tokenizörün içe aktarıldığından ve model eğitim sürecinde kullanılmadan önce tanımlandığından emin olunmalıdır.

**Eksik ve Hatalı Çalışan Bileşenler:**

* AlexNet modelinin metin verileri üzerinde kullanımı uygun olmadığından, bu modelde hatalı sonuçlar gözlemlenmiştir.

**Tam ve Doğru Çalışan Bileşenler:**

* SVM ve BERT modelleri beklenen şekilde çalışmakta ve doğru sonuçlar üretmektedir.

### Model Eğitim Yöntemleri

**Support Vector Machine (SVM)**

SVM, denetimli bir makine öğrenme algoritmasıdır ve sınıflandırma problemleri için sıklıkla kullanılır. SVM, veriyi yüksek boyutlu bir uzaya dönüştürerek en iyi ayrıştırma hattını bulmaya çalışır. Bu model, sınıflandırma problemlerinde doğruluğu yüksek sonuçlar verebilir ve küçük veri setleri ile de etkili çalışabilir.

**BERT (Bidirectional Encoder Representations from Transformers)**

BERT, Google tarafından geliştirilmiş bir büyük dil modelidir. Doğal dil işleme görevlerinde önceden eğitilmiş bir modeldir ve özellikle bağlamı anlamada güçlüdür. BERT, cümle içindeki kelimelerin çift yönlü ilişkilerini öğrenerek anlamlandırır. Bu proje kapsamında, BERT modeli kullanılarak duygu analizi yapılmıştır. BERT, metin sınıflandırma ve duygu analizi gibi NLP görevlerinde oldukça yüksek performans gösterir.

**AlexNet**

AlexNet, derin öğrenme modelidir ve özellikle görüntü tanıma alanında kullanılır. Convolutional Neural Networks (CNN) tabanlıdır ve görsel veriyi işlemek için tasarlanmıştır. Ancak, bu projede metin verisi üzerinde kullanımı uygun değildir. Metin verilerini işlemek için daha uygun modeller (örneğin, LSTM veya GRU gibi RNN tabanlı modeller) tercih edilmelidir.

### Performans Değerlendirmesi ve Karşılaştırma

* **Accuracy (Doğruluk)**: Modelin doğru sınıflandırdığı örneklerin tüm örneklere oranı.
* **Precision (Kesinlik)**: Pozitif olarak tahmin edilen sonuçlardan kaç tanesinin gerçekten pozitif olduğunu ölçer.
* **Recall (Duyarlılık)**: Gerçek pozitiflerin ne kadarının doğru tahmin edildiğini gösterir.
* **F-measure**: Precision ve Recall'un harmonik ortalaması.

#### SONUÇLAR

Bu proje, çeşitli makine öğrenimi ve derin öğrenme modelleri kullanarak metin sınıflandırma problemini ele almayı amaçlamıştır. Proje kapsamında üç farklı model incelenmiştir: Support Vector Machine (SVM), AlexNet ve BERT (Bidirectional Encoder Representations from Transformers). Her bir modelin performansı doğruluk, precision, recall ve F-measure gibi metriklerle değerlendirilmiştir. Sonuçlar, her modelin metin sınıflandırma görevinde nasıl performans gösterdiğini ve hangi modelin en iyi sonuçları verdiğini karşılaştırmalı olarak sunmaktadır.

**Support Vector Machine (SVM) Modeli**

SVM modeli, metin sınıflandırma görevinde genellikle iyi performans gösteren bir makine öğrenimi algoritmasıdır. Bu projede, SVM modeli TF-IDF vektörizasyon yöntemi kullanılarak eğitilmiştir.

SVM modeli, eğitim ve test verileri üzerinde hızlı ve etkili bir şekilde çalışmıştır. Model, sınıflandırma görevinde yüksek doğruluk oranları elde etmiş ve sınıf dengesizliklerine karşı dayanıklı olmuştur. Ancak, metin verilerinin daha karmaşık özelliklerini yakalamada sınırlamalar göstermiştir.

**AlexNet Modeli**

AlexNet modeli, genellikle görüntü sınıflandırma görevlerinde kullanılan bir derin öğrenme modelidir. Bu projede, AlexNet modelinin metin sınıflandırma görevine uyarlanmış hali kullanılmıştır.

AlexNet modeli, metin verilerinin karmaşıklığını yakalamada bazı zorluklar yaşamıştır. Modelin doğruluk oranı, SVM ve BERT modellerine kıyasla daha düşük olmuştur. Bununla birlikte, modelin derin öğrenme mimarisi, metin verilerindeki belirli desenleri ve özellikleri öğrenme potansiyeline sahiptir.

**BERT Modeli**

BERT modeli, doğal dil işleme görevlerinde son derece başarılı sonuçlar veren bir derin öğrenme modelidir. Bu projede, BERT modeli metin sınıflandırma görevinde kullanılmıştır.

BERT modeli, metin sınıflandırma görevinde en yüksek performansı göstermiştir. Model, dilin karmaşık yapısını ve bağlamını anlamada üstün yeteneklere sahiptir. BERT'in çift yönlü kodlayıcı temsilleri, metin verilerindeki ince nüansları ve ilişkileri yakalamada büyük avantaj sağlamıştır. Model, sınıflandırma görevinde yüksek doğruluk oranları elde etmiş ve metin verilerindeki anlamlı özellikleri başarıyla öğrenmiştir.

**Karşılaştırma ve Genel Değerlendirme**

Bu projede incelenen üç model arasında en yüksek performansı BERT modeli göstermiştir. BERT modeli, metin sınıflandırma görevinde en yüksek doğruluk, precision, recall ve F-measure değerlerine ulaşmıştır. SVM modeli de oldukça iyi performans göstermiştir ve özellikle sınıflandırma hızında avantaj sağlamıştır. AlexNet modeli ise diğer iki modele kıyasla daha düşük performans göstermiştir, ancak derin öğrenme mimarisi sayesinde belirli özellikleri öğrenme potansiyeline sahiptir.

Proje kapsamında elde edilen sonuçlar, metin sınıflandırma görevlerinde hangi modellerin daha etkili olduğunu ve her modelin güçlü ve zayıf yönlerini ortaya koymaktadır. Bu değerlendirmeler, gelecekteki çalışmalarda ve uygulamalarda doğru model seçiminde yol gösterici olacaktır. Ayrıca, bu projede kullanılan yöntemlerin ve modellerin daha büyük ve çeşitli veri setleri üzerinde denenmesi, sonuçların genellenebilirliğini artıracaktır. Gelecekte, bu modellerin optimizasyonu ve farklı derin öğrenme mimarileri ile karşılaştırılması, metin sınıflandırma görevlerinde daha iyi sonuçlar elde etmeye katkı sağlayacaktır.

#### KAYNAKÇA

**Veri Seti:**

Twitter Entity Sentiment Analysis Veri Seti: <https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis/data>

**Ek Kaynaklar:**

<https://www.unite.ai/tr/Makine-%C3%B6%C4%9Frenimi-yapay-zekas%C4%B1-i%C3%A7in-en-iyi-10-python-kitapl%C4%B1%C4%9F%C4%B1/>

<https://miracozturk.com/python-kutuphaneleri-ve-ozellikleri/>

<https://github.com/deeplearningturkiye/turkce-yapay-zeka-kaynaklari>

Formun Altı