

YOUTUBE YORUMLARI DUYGU ANALİZİ

HÜSEYİN ÇAĞLAR

İçindekiler

Çalışmanın Tanımı ve Amacı	4
Çalışmanın Tanımı	4
Çalışmanın Amacı	4
Literatür Bilgisi	5
Metin Sınıflandırma Teknikleri	5
Naive Bayes (NB):	5
Destek Vektör Makineleri (SVM):	5
Derin Öğrenme Modelleri:	5
Türkçe Metin Analizi	6
Tokenleştirme (Sayısallaştırma) ve Veri Hazırlama	6
Duygu Analizinde Yapay Sinir Ağları	6
Üretken Yapay Zeka Araçları	7
ChatGPT (OpenAI)	7
Google Gemini	7
GitHub Copilot	7
Eğitim Videolarından Çekilen CSV Dosyalarının İlk 20 Satırlık Tabloları	8
Yapılan Çalışma ve Seçilen Yöntemler	9
Arayüz Oluşturma	9
Arama Kutusu ve Kullanıcı Girdisi	9
Video Listeleme	9
Yorumların Getirilmesi (fetchComments)	9
CSV Dosyası İndirme	9
Stil ve Tasarım	10
Yüklenme Durumu ve Kullanıcı Deneyimi	10
Kullanılan Teknikler ve Araçlar	10
Arayüz Görselleri	10
Veri Seti Hazırlama	12
Veri Seti Oluşturma	12
Etiketleme Süreci	12
Kodların Detaylı Açıklaması	12
Model Eğitimi	13
Modelin Oluşturulması	13
Modelin Derlenmesi	14
Modelin Eğitilmesi	15
Modelin Test Verileri Üzerindeki Performansı	15
Eğitim ve Doğrulama Grafikleri	15
Model Analizi ve Çıkarımlar	16
Yapay Sinir Ağlarında Oluşturulan Ağın Şekli	16

Ağ Topolojisi ve Varsayımlar.....	17
Sonuç ve Değerlendirme	17
Kaynakça.....	18

Çalışmanın Tanımı ve Amacı

Çalışmanın Tanımı

Bu çalışma, YouTube eğitim videolarına yapılan yorumların duygu analizini yapmayı amaçlamaktadır. YouTube, dünya çapında en popüler video paylaşım platformlarından biri olup, içerik üreticilerinin eğitim, eğlence, teknoloji, yemek tarifleri gibi geniş bir yelpazede içerikler ürettikleri bir platformdur. Bu videolara yapılan yorumlar, videoların içeriği hakkında kullanıcıların düşüncelerini, duygularını ve geri bildirimlerini yansıtır. Bu yorumların pozitif, negatif veya nötr olarak sınıflandırılması, içeriğin kullanıcılar tarafından nasıl algılandığını anlamada önemli bir araçtır.

Bu çalışmada, Türkçe YouTube yorumları üzerinde duygu analizi yapılarak, yorumların duygusal yönleri belirlenmeye çalışılacaktır. Yorumları pozitif, negatif veya nötr olarak sınıflandırmak, içerik üreticilerinin videolarına dair geri bildirimleri anlamalarına yardımcı olabilir. Ayrıca bu sınıflandırma, içerik öneri sistemlerinin iyileştirilmesinde ve kullanıcı etkileşiminin artırılmasında kullanılabilir.

Çalışmanın Amacı

Türkçe Yorumlarda Duygu Analizi: Bu çalışmanın temel amacı, Türkçe yorumlarda duygu analizi yaparak, her bir yorumun duygusal içeriğini sınıflandırmaktır. Bu sayede, YouTube kullanıcılarının yorumları pozitif, negatif veya nötr olarak kategorize edilerek, içerik üreticilerine geri bildirim sağlanabilir.

Yapay Sinir Ağı Modeli Eğitimi: Yorumların duygusal içeriğini sınıflandırmak için Yapay Sinir Ağı (YSA) kullanılacaktır. YSA, metin verisini işleyerek, metinlerdeki duygu durumlarını öğrenen ve sınıflandıran bir modeldir. Bu modelin eğitilmesiyle, gelen yorumları otomatik olarak doğru sınıfa ayırmak mümkün olacaktır.

Literatür Bilgisi

Duygu analizi (sentiment analysis), doğal dil işleme (NLP) teknikleri kullanılarak metinlerdeki duygusal ifadeleri anlamaya ve sınıflandırmaya yönelik bir alandır. Bu alandaki çalışmalar genellikle metinleri pozitif, negatif veya nötr kategorilerine ayırmayı hedefler. Özellikle sosyal medya, müşteri yorumları ve içerik değerlendirme platformlarında bu tekniklerin kullanımı yaygınlaşmıştır. YouTube gibi platformlardaki yorumlar, geniş çaplı kullanıcı görüşleri içerdiği için bu tür analizler için ideal bir veri kaynağıdır.

Metin Sınıflandırma Teknikleri

Metin sınıflandırma, bir metni önceden tanımlanmış kategorilere ayırmayı amaçlar. Duygu analizi, bu sınıflandırmanın özel bir alt kümesidir. Metin sınıflandırma için kullanılan yöntemler şunlardır:

Naive Bayes (NB):

Basit ve hızlı bir yöntem olup, özellikle küçük ölçekli veri setlerinde etkilidir.

Kelimelerin bağımsız olduğu varsayımı ile çalışır.

Türkçe yorumlar gibi doğal dildeki karmaşıklıklar için yeterince esnek olmayabilir.

Destek Vektör Makineleri (SVM):

Özellikle metin sınıflandırmada sıklıkla tercih edilen güçlü bir algoritmadır.

Hangi sınıfa ait olduğu belirsiz verileri ayırmak için ideal hiper düzlemi bulur.

Ancak metinlerin uzun dönem bağımlılıklarını öğrenme konusunda sınırlıdır.

Derin Öğrenme Modelleri:

Yapay Sinir Ağları (YSA): Verilerin karmaşık ve çok boyutlu yapısını öğrenmede etkilidir. Özellikle yorum analizi gibi metin odaklı çalışmalarda öne çıkar.

LSTM (Long Short-Term Memory): Metinlerdeki uzun dönem bağımlılıkları öğrenmek için kullanılan bir tür tekrarlayan sinir ağıdır (RNN). Türkçe gibi dil bilgisi açısından zengin dillerde bağlamı anlamada başarılıdır.

Transformers (BERT, GPT): Son yıllarda popülerleşen modeller, metinlerin bağlamını anlama ve karmaşık sınıflandırma işlemleri yapmada oldukça etkilidir.

Türkçe Metin Analizi

Türkçe dilinde duygu analizi, İngilizce gibi daha yaygın dillerden farklı zorluklar içermektedir:

Ekleme Yapı: Türkçe, eklemeli bir dil olduğu için kelimeler farklı eklerle çok çeşitli anlamlar kazanabilir.

Zengin Morfoloji: Türkçe kelime yapısı ve cümle içindeki değişiklikler, yorumların doğru sınıflandırılmasını zorlaştırabilir.

Türkçe metin analizi üzerine yapılan çalışmalardan bazıları:

Türkçe İçin Gömülü Modeller (Word Embeddings):

Word2Vec ve FastText gibi yöntemler, Türkçe yorumlarda kelimelerin anlamını daha iyi temsil etmek için kullanılır.

Bu yöntemler, kelimeleri vektörlere dönüştürerek sinir ağlarının işleyebileceği formatta giriş sağlar.

Türkçe Veri Setleri:

Yıldız Teknik Üniversitesi'nin [Kemik Platformu](#), Türkçe metin veri setleri sunarak araştırmacılara büyük kolaylık sağlar.

Kaggle ve GitHub gibi platformlarda da Türkçe metin sınıflandırma projelerine yönelik veri setleri bulunmaktadır.

Tokenleştirme (Sayısallaştırma) ve Veri Hazırlama

Yapay sinir ağları gibi derin öğrenme modelleri, metinleri doğrudan işleyemez. Bu nedenle, metinlerin sayısallaştırılması (tokenleştirme) gereklidir:

Bag of Words (BOW): Kelimelerin sıklık temelli bir vektör temsili.

TF-IDF (Term Frequency-Inverse Document Frequency): Belirli bir kelimenin önemini vurgulayan bir sayısallaştırma yöntemi.

Word2Vec: Kelimeleri anlamlarına göre matematiksel vektörlere dönüştürür.

Text2Arff: Metinlerin ARFF formatına dönüştürülmesi, özellikle Weka gibi araçlarla kullanılabilirliği artırır.

Duygu Analizinde Yapay Sinir Ağları

Yapay sinir ağları, metinlerdeki duygu ifadelerini öğrenmek ve sınıflandırmak için güçlü araçlardır. Python, bu tür derin öğrenme uygulamalarını gerçekleştirmek için popüler ve esnek bir platformdur. TensorFlow ve PyTorch gibi kütüphaneler, sinir ağlarını oluşturmak, eğitmek ve değerlendirmek için yaygın olarak kullanılır. Bu projede Python kullanılarak duygu analizi yapılacaktır.

Üretken Yapay Zeka Araçları

Üretken Yapay Zeka (Generative AI), verilen girdilerden anlamlı ve yaratıcı sonuçlar üretebilen bir yapay zeka türüdür. Bu araçlar, doğal dil işleme (NLP), görüntü oluşturma, ses sentezi ve daha birçok alanda kullanılarak çeşitli projelerde büyük kolaylık sağlar. Projelerde hem zaman tasarrufu sağlamak hem de kaliteli sonuçlar elde etmek için bu araçlar oldukça etkilidir. Aşağıda yaygın olarak kullanılan bazı üretken yapay zeka araçlarının tanımları ve özellikleri verilmiştir:

ChatGPT (OpenAI)

ChatGPT, OpenAI tarafından geliştirilen bir dil modeli olup, kullanıcıların girdilerine doğal dilde yanıtlar üretir. Metin tabanlı projelerde bilgi sağlama, fikir önerme ve yazılım geliştirme gibi işlemler için sıkça kullanılır.

Özellikler:

- Kodlama, hata ayıklama ve algoritma oluşturma konusunda destek sağlar.
- İnsan benzeri yanıtlarla metin oluşturma ve düzenleme yapabilir.
- Kullanıcı dostu ve hızlı yanıt süreleri sunar.

Avantajlar:

- Özellikle duygu analizi gibi metin sınıflandırma projelerinde açıklamalar sunabilir.
- Eğitim verilerini anlamlandırma ve model tasarımı önerileri geliştirme konusunda yardımcıdır.

Google Gemini

Gemini, Google'ın dil modeli teknolojisini kullanan ve kullanıcıların doğal dildeki sorularına ileri düzeyde yanıtlar veren bir yapay zeka platformudur.

Özellikler:

- Geniş veri tabanı ve güçlü dil işleme yetenekleri.
- Farklı dillerde metin işleme ve anlamlandırma.
- Gelişmiş araştırma ve öneri sistemleri sunar.

Avantajlar:

- Metin analizi projelerinde, karmaşık veri kümelerini anlamlandırma konusunda güçlüdür.
- Kullanıcıların ihtiyaçlarına göre yaratıcı içerikler oluşturur.

GitHub Copilot

Copilot, GitHub ve OpenAI iş birliğiyle geliştirilen bir kod öneri aracıdır. Kullanıcıların yazdığı kodları analiz ederek, devamını otomatik olarak tamamlar ve kodlama sürecini hızlandırır.

Özellikler:

- Çeşitli programlama dillerinde kod önerisi yapar.
- Kodun bağlamını anlar ve hata önleyici önerilerde bulunur.
- Geliştirici araçlarıyla entegrasyon sağlar.

Avantajlar:

- Python ve TensorFlow projelerinde kodun hızlı bir şekilde yazılmasını destekler.
- Hata ayıklama ve tekrarlayan görevleri otomatikleştirerek zaman kazandırır.

Eğitim Videolarından Çekilen CSV Dosyalarının İlk 20 Satırlık Tabloları

	A	B	C	D	E	F	G	H
1	VideoBaşlık	YayınlanmaTarihi	KanalAdı	Yazar	Tarih	Yorum	Beğeni	Yanıtlar
2	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@KavaxVideo	2023-12-01T18:06:07Z	Arkadaşlar çok sorulmuş, Büyük eşitt	173	24
3	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@kaanerenkaradeniz395	2024-11-16T22:37:53Z	Abi çok yararlı çok teşekkür ederim	0	0
4	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@zonfir	2024-11-16T14:17:23Z	Moby Dick en az 600 sayfa:))) Üstelik 19!	0	0
5	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@YTslcx8	2024-11-16T08:54:58Z	ben indirdim ama nereden açacağım	0	0
6	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@GalaxyS8-v5b	2024-11-12T19:50:20Z	Benim bilgisayarım Asus E205SA pycharr	0	0
7	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@keriz6525	2024-11-11T12:22:51Z	hocam keşke "for i in" kalıbı	0	0
8	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@muhammetalisardogan1251	2024-11-11T07:54:14Z	Birazdaha önemli konulara değine bilirmi	0	0
9	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@muhammetalisardogan1251	2024-11-11T07:24:46Z	Bir sayının karekökünü bulmak için 0.50	0	0
10	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@Zinkoii	2024-11-10T23:54:54Z	değişkenleri çözemiyorum..	0	0
11	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@fikretakbass	2024-11-10T09:52:31Z	1 Saatte python öğrenebileceğini sanan 5	0	2
12	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@Shadow_3457	2024-11-09T19:35:26Z	Hayatımda harcadığım en verimli ve zevk	0	0
13	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@asyaabilova4912	2024-11-07T20:18:47Z	👍👍👍	0	0
14	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@serefabdu3042	2024-11-07T13:09:18Z	youtubede izlediğim en verimli paython	0	0
15	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@ayberkofpc	2024-11-04T16:58:37Z	Hocam linux için editör ne önerirsiniz	0	0
16	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@mustafayildirim7308	2024-11-03T20:53:51Z	👍 çok güzel bir anlatım olmuş elinize sa	0	0
17	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@dilanuluk5895	2024-11-03T15:39:50Z	Pythonu kurarken isten şifre ne oluyor	0	0
18	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@zack9199	2024-11-03T12:34:29Z	gerçekten mi?	0	0
19	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@l.ee_mi	2024-11-02T17:21:44Z	cok iyi izah etdiniz tesekkurler.🙏bir sey	0	0
20	Python Dersleri - 1	2023-09-01T10:00:25Z	Kavax	@ibrahimorcan289	2024-11-01T21:18:57Z	.	0	0

Video linki:

<https://www.youtube.com/watch?v=wZUNiGtKcw&t=34s>

(dosyalar .csv formatında indiriliyor. Fakat yukarıdaki gibi başlıklar ve içerikler belli olsun diye .xlsx formatına dönüştürerek ekran resmi aldım.)

1	VideoBaşlık	YayınlanmaTarihi	KanalAdı	Yazar	Tarih	Yorum	Beğeni	Yanıtlar
2	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@Travelwithusintheworld	2024-10-29T20:05:04Z	Hocam emeginize saglik ben szin v	0	0
3	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@sezginyildirim91	2024-10-29T01:07:07Z	Geçenlerde Sadık hocamızın React	0	0
4	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@sebledemiroglux	2024-10-25T19:10:43Z	Emeğinize sağlık , hayırlı uğurlu ols	0	0
5	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@ferdisimsek	2024-10-19T20:17:40Z	Hocam yeni eğitiminiz bizleri çok s	1	0
6	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@huseyineskan	2024-10-17T20:27:51Z	Emeğinize sağlık. Güncel bir React l	0	0
7	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@yavuz7744	2024-10-17T13:44:59Z	Springboot sonrası fe için react öğr	0	0
8	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@burakcavundur	2024-10-12T11:41:21Z	Selamlar. Bu React derslerini oynat	0	1
9	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@yakupuygur98	2024-10-12T09:31:37Z	Teşekkürler	0	0
10	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@elifdurukan9022	2024-10-10T14:28:37Z	Hocam emeğinize sağlık, devamlar	0	0
11	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@nostalgiamusic9947	2024-10-09T17:14:30Z	7iyuk	0	0
12	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@_owog_-5505	2024-10-08T05:42:42Z	Merhaba hocam çok alakasız olaca	1	2
13	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@annatar8718	2024-10-06T18:46:28Z	güncel bi react kursu arıyordum, re	0	0
14	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@aktifkopyalamamerkezi3495	2024-10-05T14:34:04Z	Çok şükür hocam gözler seni aradı,	0	0
15	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@aytcunai	2024-10-04T19:06:40Z	Sesiniz özlenmiş hocam 😊	0	0
16	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@pnaraliogullari6863	2024-10-03T10:27:54Z	teşekkürler hocam	0	0
17	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@ahmetkaya-de	2024-10-01T16:01:18Z	React.js mi Vue.js mi hocam?	0	1
18	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@enginb3818	2024-09-30T20:18:32Z	harikasınız hocam. sabirsizlikle bek	0	0
19	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@aminaliyev4615	2024-09-30T10:56:03Z	Devamını bekliyoruz 🙏	0	0
20	React Dersleri: #1-React Nedir	2024-09-26T18:00:31Z	Sadık Turan	@fatihcece3359	2024-09-30T07:16:05Z	Harika bir baslangic. Reacti bilenler	0	0

Video linki:

<https://www.youtube.com/watch?v=z3SEfSeFN5k>

Yapılan Çalışma ve Seçilen Yöntemler

Arayüz Oluşturma

React ile YouTube API'si kullanarak videoları aramak ve videolara ait yorumları görüntülemek için arayüz oluşturdum. Ayrıca, bu yorumları ve video meta verilerini bir CSV dosyası olarak indirme özelliği ekledim.

Arama Kutusu ve Kullanıcı Girdisi

State Yönetimi:

useState kullanarak query (arama terimi) ve diğer durumları (videos, selectedVideo, comments, loading) yönetiyorum.

Kullanıcı, arama kutusuna bir terim girdiğinde setQuery ile durum güncelleniyor.

Arama İşlevi (searchVideos):

axios kütüphanesi ile YouTube API'ye istek gönderilir.

Arama sonuçları (videolar), videos durumuna kaydedilir.

Maksimum 10 sonuç döndürülür (maxResults: 10).

Video Listeleme

Videoların Gösterimi:

Videolar bir liste halinde gösteriliyor. Her video bir buton olarak görüntüleniyor.

Bir butona tıklandığında:

setSelectedVideo ile seçili video durumu güncellenir.

fetchComments işlevi ile seçilen videonun yorumları alınır.

Yorumların Getirilmesi (fetchComments)

YouTube API İstekleri:

API'nin commentThreads endpoint'i kullanılarak yorumlar alınır.

nextPageToken kullanılarak tüm yorumlar sayfalananak çekilir.

Yorum Verilerinin İşlenmesi:

Alınan yorumlar birleştirilerek comments durumuna kaydedilir.

CSV Dosyası İndirme

CSV Formatına Dönüştürme (downloadCSV):

Yorum verileri, Papa.unparse kullanılarak CSV formatına dönüştürülür.

Her yorum için şu bilgiler CSV'ye eklenir:

Video başlığı, yayınlanma tarihi, kanal adı.

Yorum yazar adı, tarihi, metni, beğeni sayısı, cevap sayısı.

İndirme İşlevi:

JavaScript Blob nesnesi ile bir CSV dosyası oluşturulur.

Kullanıcının dosyayı indirmesi için bir <a> ögesi dinamik olarak oluşturulup simüle edilir.

Stil ve Tasarım

Inline CSS:

styles nesnesi içinde, uygulamanın görsel tasarımı tanımlanır. Girdi kutusu, düğmeler, yorumlar ve video listesi gibi bileşenlerin stilleri ayarlanmıştır.

Yüklenme Durumu ve Kullanıcı Deneyimi

Yüklenme Göstergesi:

API istekleri sırasında loading durumunu true yaparak "Yükleniyor..." mesajı gösterilir.

Hata Yönetimi:

API hataları için try-catch blokları kullanılmıştır. Hata durumunda console.error ile loglama yapılıyor.

Kullanılan Teknikler ve Araçlar

React ve Hooks (useState):

Uygulamanın durumlarını yönetmek için kullanıldı.

Axios:

YouTube API'ye HTTP istekleri yapmak için kullanıldı.

PapaParse:

Yorumları CSV formatına dönüştürmek için kullanıldı.

YouTube API:

Video arama ve yorum alma işlemleri için kullanıldı.

CSS ile Stil:

Uygulamanın kullanıcı dostu bir arayüz sunması için basit stil yönetimi yapıldı.

Arayüz Görselleri

YouTube Video Arama ve Yorumlar

Ara

Videolar

React mı Angular mı öğrenmeliyim | Özet anlatım.

1) Nedir Bu React ? | REACT Dersleri

React Tutorial for Beginners

REACT vs ANGULAR vs VUE

React Dersleri: #1-React Nedir ve Neden React Öğrenmeliyim?

Sıfırdan İleri Seviyeye REACT Eğitimi

React Full Course for free (2024)

3) React Projesi Oluşturma | Proje Yapısı İncelemesi | REACT Dersleri

Build Full Stack Instagram Clone with React JS | ReactJS, Tailwind CSS, shadCN UI, React Query

Yeni Başlayanlar İçin, 2.5 Saatte REACT Eğitimi

Video: React Dersleri: #1-React Nedir ve Neden React Öğrenmeliyim?

Yayınlanma Tarihi: 2024-09-26T18:00:31Z

Yorumlar

@Travelwithusintheworld - 2024-10-29T20:05:04Z

Hocam emeginiza saglik ben szin videolardan somaliadan izliyorum ve ogreniyorum tessekur ederim hocam

Beğeni Sayısı: 0

@sezginyildirim91 - 2024-10-29T01:07:07Z

Geçenlerde Sadık hocamızın React derslerini aramıştım, hemen bu seriye başlıyorum çünkü bir tek Sadık hocamın anlatımından çok net anlayabiliyorum. Emekleriniz için çok teşekkür ederim.

Beğeni Sayısı: 0

@sebledemiroglux - 2024-10-25T19:10:43Z

Emeğinize sağlık , hayırlı uğurlu olsun 🙏

Beğeni Sayısı: 0

@ferdisimsek - 2024-10-19T20:17:40Z

@annatar8718 - 2024-10-06T18:46:28Z

güncel bi react kursu arıyordum, reacte yeniden başlıyorum çok iyi oldu hocam devamını bekliyorum

Beğeni Sayısı: 0

@aktifkopyalamamerkezi3495 - 2024-10-05T14:34:04Z

Çok şükür hocam gözler seni aradı, BTK dersleri bitirdim youtube bitirdik, sıra ne öğrenebilirim diyordum karşıma çıktınız devamını bekliyor ve takip ediyor olacağız... 🙏❤

Beğeni Sayısı: 0

@aytcunal - 2024-10-04T19:06:40Z

Sesiniz özlenmiş hocam 😊

Beğeni Sayısı: 0

@pnaraliogullari6863 - 2024-10-03T10:27:54Z

teşekkürler hocam

Beğeni Sayısı: 0

Veri Seti Hazırlama

Model eğitmek için veri setini kendim oluşturmak istedim. Farklı videolardan yorumlar çekerek küçük bir veri seti oluşturdum daha sonrasında ise hazır bir model kullanarak bu verileri etiketledim. Bu şekilde modelimi oluşturmak için hazır bir veri seti elde etmiş oldum.

Veri Seti Oluşturma

Farklı YouTube videolarından yorumlar çekerek özgün bir veri seti oluşturdum.

Çekilen yorumları, mevcut bir CSV dosyasına eklemek veya yeni bir dosya oluşturmak için Pandas kütüphanesini kullandım.

Bu süreçte, Selenium kullanılarak otomasyon sağladım ve yorumların belirli bir sayıya ulaşana kadar dinamik olarak çekilmesini sağladım.

Etiketleme Süreci

Çekilen yorumları, Zero-Shot Learning yöntemini kullanarak etiketledim.

"facebook/bart-large-mnli" modeli ile yorumların duygu durumlarını ("Positive", "Negative", "Neutral") olarak sınıflandırdım.

Hazırladığım bu veri setini, daha sonra model eğitimi için kullanılabilir hale getirdim.

Kodların Detaylı Açıklaması

Yorum Çekme Kodu

Kullanılan Teknikler:

Selenium: Dinamik bir web sayfasındaki öğelere erişmek ve yorumları çekmek için kullanıldı.

Pandas: Çekilen yorumları düzenlemek ve CSV formatında kaydetmek için kullanıldı.

OS Modülü: Daha önce oluşturulmuş bir CSV dosyasını kontrol edip, varsa dosyaya yeni yorumlar eklemek için kullanıldı.

Aşamalar:

Selenium ile YouTube Video Yorumlarını Alma:

Selenium WebDriver kullanılarak Firefox tarayıcısı başlatıldı.

Web sayfası aşağı kaydırılarak daha fazla yorumun yüklenmesi sağlandı.

Yorumlar, #content-text CSS seçicisi ile çekildi.

Verilerin Saklanması:

Çekilen yorumlar bir listeye alındı.

Eğer daha önce bir CSV dosyası varsa, yeni yorumlar mevcut dosyaya eklenip tekrarlı olanlar kaldırıldı.

Yorumlar, CSV formatında kaydedildi.

Yorumlara Etiket Verme

Kullanılan Teknikler:

Hugging Face Transformers: "facebook/bart-large-mnli" modeli kullanılarak yorumlar duygu durumlarına (sentiment) göre sınıflandırıldı.

Zero-Shot Classification: Model, belirtilen aday etiketler arasından en uygun etiketi seçti.

Pandas: Çekilen yorumlara duygu durumu etiketlerini eklemek ve sonuçları bir CSV dosyasına kaydetmek için kullanıldı.

Aşamalar:

Sentiment Analizi:

Yorumlar birer birer işlenerek, model tarafından etiketlendi.

En yüksek olasılık eşiği (%45) altında olan durumlar "Neutral" olarak işaretlendi.

Sonuçların Saklanması:

Etiketlenmiş yorumlar, yeni bir CSV dosyasına kaydedildi.

Her yorum için duygu durumu etiketi ve model tarafından verilen olasılık skoru da eklendi.

Genel Süreç ve Kullanılan Araçlar

Selenium: Web sayfası otomasyonu ve dinamik içerik çekimi.

Pandas: Verilerin işlenmesi, birleştirilmesi ve saklanması.

Transformers ve Hugging Face: Etiketleme için hazır bir modelin kullanımı.

CSV İşlemleri: Veri seti oluşturma ve güncelleme.

Model Eğitimi

Modelin Oluşturulması

Modelin eğitimi için, derin öğrenme tabanlı bir yapay sinir ağı (YSA) oluşturdum. Model, TensorFlow Keras API'si kullanarak tasarladım. YSA'nın mimarisi üç ana bölümden oluşmaktadır: giriş katmanı, gizli katmanlar ve çıkış katmanı.

Giriş Katmanı:

Giriş verisi, Word2Vec kullanılarak her bir yorumun vektör temsili olan 100 boyutlu sayısal vektörlerden oluşmaktadır. Yani her bir yorum, 100 boyutlu bir vektörle ifade edilmiştir. Bu veriler, modelin ilk katmanına iletilmektedir.

İlk Gizli Katman:

İlk gizli katman, 128 nörondan oluşmaktadır ve her nöron, giriş verisindeki 100 boyutlu vektöre göre ağırlıklandırılıyor. Aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit) kullanılmıştır. ReLU fonksiyonu, negatif değerleri sıfırlayarak daha hızlı öğrenmeyi sağlar ve derin ağlarda yaygın olarak kullanılır. Bu katman, verilerin daha derin anlamlarını keşfetmek için önemli bir rol oynar.

Dropout Katmanı:

Gizli katmandan sonra, aşırı öğrenmeyi (overfitting) önlemek amacıyla bir Dropout katmanı ekledim. Dropout katmanı, her eğitim adımında belirli nöronları rastgele devre dışı bırakır, böylece modelin daha genel özellikler öğrenmesini sağlar. Burada %30 dropout oranı kullanılmıştır.

İkinci Gizli Katman:

İkinci gizli katman ise 64 nörondan oluşmaktadır. Bu katman, önceki katmandan gelen verileri işler ve yine ReLU aktivasyon fonksiyonu ile doğrusal olmayan özellikleri öğrenmeye devam eder.

Çıkış Katmanı:

Çıkış katmanı, 3 nöron içerir. Çünkü modelimiz, üç sınıf (pozitif, negatif ve nötr) arasında sınıflandırma yapacaktır. Çıkış katmanında softmax aktivasyon fonksiyonu kullanılmıştır. Softmax, her sınıf için bir olasılık değeri döndürür ve bu olasılıklar toplamı 1 olur, bu da sınıflandırma yaparken modelin hangi sınıfa daha fazla güven duyduğunu belirler.

Modelin Derlenmesi

Modelin eğitilmeden önce derlenmesi gerekmektedir. Derleme işlemi, modelin öğrenme sürecinde hangi optimizasyon yöntemlerinin kullanılacağını ve hangi kayıp fonksiyonlarının kullanılacağını belirler.

Optimizasyon Algoritması:

Adam optimizasyon algoritması kullanılmıştır. Adam algoritması, dinamik öğrenme oranlarıyla çalışan, genellikle hızlı ve etkili sonuçlar veren bir algoritmadır. Derin öğrenme uygulamalarında sıkça tercih edilir.

Kayıp Fonksiyonu:

Modelin kaybı, Sparse Categorical Crossentropy olarak belirlenmiştir. Bu fonksiyon, çok sınıflı sınıflandırma problemleri için uygundur. Her sınıf için tahmin edilen olasılık ve gerçek etiket arasındaki farkı minimize eder.

Başarı Ölçütü (Metric):

Modelin başarısını ölçmek için accuracy (doğruluk) metriği kullanılmıştır. Bu metrik, modelin doğru sınıflandırdığı örneklerin oranını verir.

Modelin Eğitilmesi

Model, eğitim verileriyle 20 epoch (dönem) boyunca eğitilmiştir. Her epoch, tüm eğitim verileri üzerinde yapılan bir tam geçişi temsil eder. Eğitim sırasında, model ağırlıklarını güncelleyerek daha iyi tahminlerde bulunabilmek için sürekli olarak öğrenir.

Batch Size: 32 olarak belirlenmiştir. Batch size, her adımda modelin işlediği örnek sayısını belirler. 32, öğrenme sürecinin verimli bir şekilde yapılabilmesi için iyi bir değerdir.

Modelin Test Verileri Üzerindeki Performansı

Eğitimden sonra, modelin test verileri üzerindeki doğruluğu hesaplanır. Bu, modelin bilinmeyen veriler üzerinde nasıl performans gösterdiğini görmek için önemli bir adımdır. Modelin tahminleriyle gerçek etiketler karşılaştırılır ve doğruluk oranı hesaplanır.

Test doğruluğu, test verilerinde %46.27 olarak ölçülmüştür. Bu, modelin doğru tahminlerde bulunduğu oranı ifade eder. Test doğruluğu, modelin öğrenme kapasitesini ve genelleme yeteneğini belirler. %46.27 doğruluk oranı, modelin hâlâ gelişmeye ihtiyaç duyduğunu gösteriyor. Geliştirme için modelin yapısının daha fazla optimize edilmesi veya daha fazla eğitim verisi eklenmesi gerekebilir.

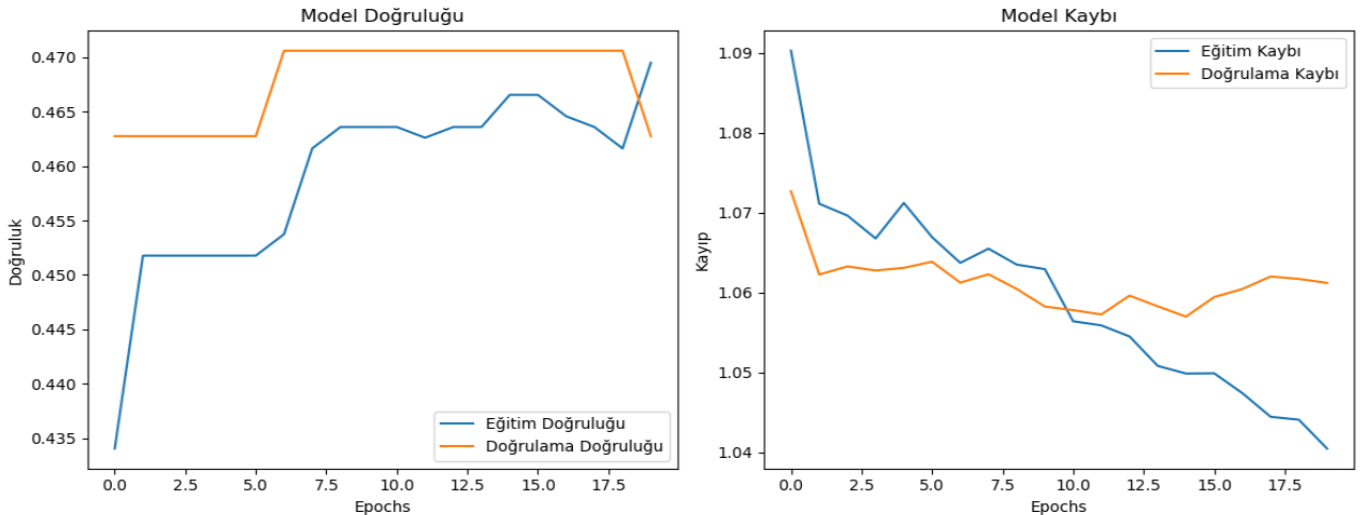
Eğitim ve Doğrulama Grafikleri

Modelin eğitim ve doğrulama sırasında elde edilen doğruluk ve kayıp değerleri görselleştirildi. Bu, modelin öğrenme sürecini daha iyi anlamamıza yardımcı olur. Grafikler, her epoch sonunda modelin doğruluğunun arttığını ve kayıp değerlerinin azaldığını göstermektedir.

Eğitim Doğruluğu: Modelin eğitim sırasında ne kadar doğru tahminlerde bulunduğunu gösterir.

Doğrulama Doğruluğu: Modelin doğrulama verileri üzerindeki doğruluğunu gösterir. Eğer eğitim doğruluğu yüksek fakat doğrulama doğruluğu düşükse, bu aşırı öğrenmeye (overfitting) işaret edebilir.

8/8 [=====] - 0s 1ms/step
Test Doğruluğu: 46.27%



Model Analizi ve Çıkarımlar

Yapay Sinir Ağlarında Oluşturulan Ağın Şekli

Modelin yapısı, derin öğrenme tabanlı bir yapay sinir ağı (YSA) mimarisi kullanılarak oluşturulmuştur. Model, üç ana katmandan oluşmaktadır:

Giriş Katmanı: Modelin ilk katmanı, 100 boyutlu vektörlerle beslenen bir yoğun (dense) katmandan oluşmaktadır. Bu katman, her bir yorumun Word2Vec ile dönüştürülmüş olan sayısal temsiline giriş yaparak modele veri aktarır. Burada kullanılan 100 boyutlu vektör, kelimelerin anlamını temsil etmek için yeterli bir boyuttadır ve modelin anlamlı özellikler öğrenmesine olanak tanır.

Gizli Katmanlar: Modelde iki adet gizli katman bulunuyor:

İlk gizli katman, 128 nöron içerir ve ReLU aktivasyon fonksiyonu kullanarak doğrusal olmayan özelliklerin öğrenilmesini sağlar. Bu katman, modelin daha karmaşık ilişkileri öğrenmesini ve sınıflandırma yapabilmesini sağlar.

İkinci gizli katman ise 64 nöron içerir ve benzer şekilde ReLU aktivasyon fonksiyonunu kullanarak daha detaylı ve soyut özellikleri keşfeder.

Çıkış Katmanı: Modelin çıkış katmanı 3 nörondan oluşur ve bu nöronlar, üç farklı sınıfa (pozitif, negatif, nötr) ait olasılıkları temsil eder. Çıkış katmanında softmax aktivasyon fonksiyonu kullanılır. Softmax, her sınıf için bir olasılık değeri döndürür ve modelin hangi sınıfa daha fazla güven duyduğunu gösterir.

Toplam Parametre Sayısı: Modelin toplam parametre sayısı 21,379'dur. Bu, modelin öğrenme sürecinde güncellenebilecek ağırlıkların toplam sayısını ifade eder. Parametre sayısı, modelin kapasitesini ve karmaşıklığını belirler. 21,379 parametre, modelin yeterli öğrenme kapasitesine sahip olduğu ancak daha fazla parametre ile daha karmaşık ilişkileri öğrenebileceği anlamına gelir.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 128)	12928
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 3)	195

Total params: 21379 (83.51 KB)
Trainable params: 21379 (83.51 KB)
Non-trainable params: 0 (0.00 Byte)

Ağ Topolojisi ve Varsayımlar

Modelin topolojisi, performansı artıracak şekilde tasarlanmıştır. Modelin temel bileşenleri ve yapısal özellikleri şu şekilde belirlenmiştir:

İki Gizli Katman: Model, iki gizli katmandan oluşmaktadır. İlk katman 128 nöron içerirken, ikinci katman 64 nörondan oluşmaktadır. Bu yapı, modelin öğrenme kapasitesini artırır ve veri setinin daha soyut özelliklerini öğrenmesini sağlar. Daha fazla katman, modelin karmaşıklığını artırarak daha iyi genelleme yapmasına olanak tanıyabilir. Ancak katman sayısının çok fazla olması, aşırı öğrenmeye yol açabilir.

Dropout Katmanları: Modelin her iki gizli katmanından sonra Dropout katmanları eklenmiştir. Dropout, eğitim sırasında bazı nöronları rastgele devre dışı bırakır. Bu sayede modelin aşırı öğrenmesini (overfitting) engellemek amaçlanmıştır. Dropout oranı %30 olarak belirlenmiştir, bu oran eğitim sürecinde modelin daha sağlam ve genelleştirilebilir özellikler öğrenmesini sağlar. Aksi takdirde, model sadece eğitim verilerine özgü özellikleri öğrenebilir ve doğrulama verileri üzerinde düşük performans gösterebilir.

Genel Yapı: Modelin toplamda 3 katmandan oluşan yapısı, sınıflandırma görevine yönelik doğru ve verimli sonuçlar elde edilmesini amaçlamaktadır. İki gizli katman ve çıkış katmanı, modelin karmaşık veri ilişkilerini öğrenmesine olanak tanır. Dropout katmanları, modelin genelleme gücünü artırarak yeni verilere daha iyi uyum sağlamasına yardımcı olur.

Sonuç ve Değerlendirme

Modelin eğitiminden sonra, test verileri üzerinde elde edilen doğruluk oranı **%46.27** olarak ölçülmüştür. Bu oran, modelin hala geliştirilmeye ihtiyaç duyduğunu gösteriyor. %46.27'lik doğruluk, modelin test verisi üzerinde %46.27'lik bir doğrulukla tahminlerde bulunduğunu gösterir; bu da modelin bazı sınıflar arasında net bir ayrım yapamadığı anlamına gelir. Bu doğruluk oranı, modelin henüz çok iyi genelleme yapmadığını ve daha fazla iyileştirme yapılması gerektiğini belirtiyor.

Modelin İyileştirilmesi İçin Potansiyel Adımlar:

Modelin Mimari Yapısının Geliştirilmesi: Modelde kullanılan gizli katman sayısı ve her katmandaki nöron sayısı artırılarak daha derin bir model oluşturulabilir.

Hiperparametre Optimizasyonu: Modelin öğrenme oranı, batch size gibi hiperparametreler optimize edilerek modelin daha hızlı ve verimli bir şekilde öğrenmesi sağlanabilir. Bu parametreler, modelin başarısını doğrudan etkileyebilir.

Daha Fazla Eğitim Verisi Kullanımı: Modelin doğruluğunu artırmak için daha fazla etiketli eğitim verisi kullanılabilir. Eğitim verilerinin çeşitliliği arttıkça modelin genelleme yeteneği de artar.

Özellik Mühendisliği: Girdi verileri üzerinde yapılan iyileştirmeler, özellikle yorumların daha etkili bir şekilde temsil edilmesi, modelin doğruluğunu artırabilir. Örneğin, farklı kelime gömme yöntemleri (word embedding) kullanılabilir veya verilerdeki anlamsal ilişkiler daha etkili bir şekilde yakalanabilir.

Sonuç olarak, modelin şu anki doğruluğu %46.27 seviyesinde olsa da, bu oran daha fazla optimizasyon, iyileştirme ve genişletme ile artırılabilir. Modelin performansının artırılması için sürekli olarak test edilmesi, hiperparametrelerin ayarlanması ve yeni veri kaynaklarının eklenmesi gerekiyor.

Kaynakça

Selenium Kütüphanesi :

Selenium Contributors. (n.d.). *Selenium WebDriver documentation*. Selenium. Retrieved November 17, 2024, from <https://www.selenium.dev/documentation/>

Pandas Kütüphanesi:

Pandas Development Team. (n.d.). *Pandas documentation*. Pandas. Retrieved November 17, 2024, from <https://pandas.pydata.org/docs/>

Time Modülü:

Python Software Foundation. (2020). time — Time access and conversions. <https://docs.python.org/3/library/time.html>

NumPy Kütüphanesi:

NumPy Developers. (n.d.). *NumPy v2.1 documentation*. NumPy. Retrieved November 17, 2024, from <https://numpy.org/doc/2.1/>

NLTK Kütüphanesi:

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media. Retrieved November 17, 2024, from <https://www.nltk.org/>

TensorFlow Kütüphanesi:

TensorFlow Contributors. (n.d.). *TensorFlow: An end-to-end open source machine learning platform*. TensorFlow. Retrieved November 17, 2024, from <https://www.tensorflow.org/?hl=tr>

Word2Vec:

TensorFlow Contributors. (n.d.). *Word2Vec tutorial*. TensorFlow. Retrieved November 17, 2024, from <https://www.tensorflow.org/text/tutorials/word2vec>

Scikit-learn:

Scikit-learn Developers. (n.d.). *Scikit-learn: Machine learning in Python*. Retrieved November 17, 2024, from <https://scikit-learn.org/stable/>

Matplotlib ve Grafikler:

Hunter, J. D. (2007). *Matplotlib: A 2D graphics environment*. Retrieved November 17, 2024, from <https://matplotlib.org/>

Transformers Kütüphanesi:

Wolf, T., & Hugging Face Team. (2020). *Transformers: State-of-the-art Natural Language Processing*. Hugging Face. Retrieved November 17, 2024, from <https://huggingface.co/docs/transformers/index>

BART Model (facebook/bart-large-mnli):

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., & Parikh, D. (2020). *BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension* (facebook/bart-large-mnli). Hugging Face. Retrieved November 17, 2024, from <https://huggingface.co/facebook/bart-large-mnli>

React Kütüphanesi:

React Team. (2021). React: A JavaScript library for building user interfaces. <https://reactjs.org/>

Axios Kütüphanesi:

Axios Contributors. (2021). Axios: Promise based HTTP client for the browser and Node.js. <https://axios-http.com/>

Papa Parse Kütüphanesi:

Papaparse Contributors. (2021). Papa Parse: A powerful CSV library. <https://www.papaparse.com/>

YouTube Data API v3:

Google Developers. (2021). YouTube Data API v3. <https://developers.google.com/youtube/v3>