**T.C.**

**BALIKESİR ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

metin, daire, logo, ticari marka içeren bir resim

Açıklama otomatik olarak oluşturuldu

**YOUTUBE ÜZERİNDEN ÇEKTİGİMİZ VERİLERLE**

**DUYGU ANALİZİ**

**202113709030-MEHMET TUNAHAN TÜRKYILMAZ**

**BMM4101 YAPAY ZEKA TEKNİKLERİ VİZE ÖDEVİ**

**Danışman: Dr. Öğr. Üyesi Kadriye ERGÜN**

**BALIKESİR, 15-11-2024**

**İÇİNDEKİLER**

**1-**Çalışmanın tanımı ve amacı, literatür bilgisi

**2-**Seçilen eğitim videolarına ait yorumlarını gösteren 20’şer satırlık tablo

**3-**Ödevde yaptığınız çalışma ve seçtiğiniz yöntemlerin sırası ve anlatımı

**4-**Yapay Sinir Ağlarında oluşturduğunuz ağın şekli

**5-**Ağ topolojisi ve varsayımlar

**6-**Optimum ağ tasarımının anlatımı

**7-**Karşılaştırma (optimum ağ ile başlangıçtaki ağ yapısı arasında) ve sonuçların yorumlanması (Sınıflandırıcı başarı oranı hakkında bilgi)

**8**-Yapay zeka aracı ile birlikte çözüm örneği

**9-**Sonuç

**10-**Kaynaklar

1. **Çalışmanın Tanımı ve Amacı**

Bu çalışma, **YouTube videolarına ait yorumların toplanması, analizi ve sınıflandırılması** için bir sistem geliştirmeyi amaçlar. Kullanıcılar, belirli bir konuda YouTube'da yapılan yorumların duygu durumlarını (**pozitif**, **negatif**, **nötr**) görebilir ve analiz sonuçlarını arayüz üzerinden inceleyebilir.

**Literatür Bilgisi**

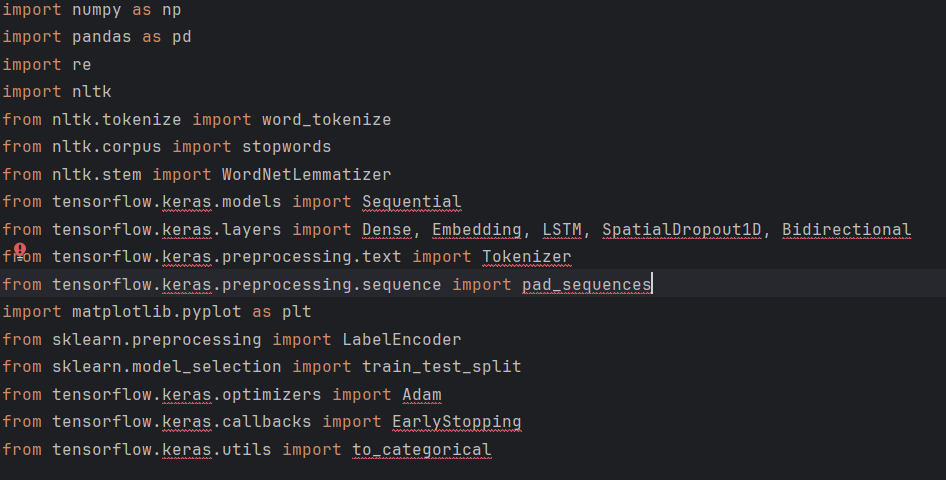
* **Doğal Dil İşleme (NLP)**: Yorumların temizlenmesi ve lemmatization işlemleri, metin analizi için temel adımdır.
* **Duygu Analizi**: LSTM tabanlı bir model, metin sınıflandırmada etkili bir yöntemdir.
* **Web Scraping**: Selenium kullandım, dinamik web sitelerinden veri çekmek için yaygın bir araçtır.
* **Kullanıcı Arayüzü**: Tkinter, analiz sonuçlarının görselleştirilmesi için kullanılmıştır.

Yorumlar Tablosu

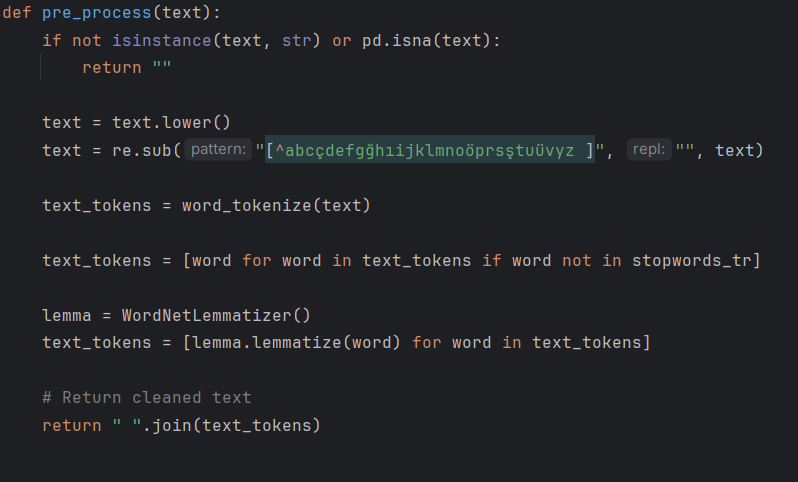
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Video Başlığı | Yorum Yapan | Yorum Tarihi | Yorum | Beğeni Sayısı | Yanıt Sayısı | Predicted\_Class |
| Sinyaller ve Sistemler Ders 1 Giriş | @recepahmetkartal6813 | 1 yıl önce | herkes tablet kullanarak ders ankatırken, elinde mousle ders anlatmaya çalışan koca yürekli adam. | 16 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @albusdumbledore8269 | 2 yıl önce | Hocam az diksiyona dikkat etsek keşke ya emeğinize sağlık.. | 7 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @emrecakr7667 | 4 yıl önce | baun eem 'den hayırlı bayramlar arkadaşlar sınavlara çalışma vakti geldi:)Video ıcın kanala teşekkürler | 13 | 3 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @isayildirim9183 | 3 yıl önce | Çok Teşekkürler elinize sağlık | 2 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @lincolnnavigator7801 | 3 yıl önce | Efendim 2 üzeri 0, 1 eşit değil mi? Niçin 0 dan başlatma gereği duydunuz | 14 | 1 | nötr |
| Sinyaller ve Sistemler Ders 1 Giriş | @sadiyeciflikli6539 | 5 yıl önce | hocam videoların devamını bekliyoruz :) | 5 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @nihasilev | 5 yıl önce | umarım bu serinin devamı gelir ve seriyi tamamlarsınız.. | 6 | 5 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @eem-ahmetkiraz4029 | 5 ay önce | Hocam emeğinize sağlık bir şey soracaktım bilgisayarın dokunmatik kısmıyla mı yazıyorsunuz ? | 1 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @bugrayldrm8279 | 3 yıl önce | Hocam öncelikle teşekkürler notlarınız ve derslere bakış açılarınız çok güzel ama umarım üniversite hocası değilsinizdir ^^ Sesiniz insanın uykusunu getiriyor | 6 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @burakyldrm4060 | 3 yıl önce | ilk ornek yanlis cozulmustur. | 7 | 7 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @allamov | 1 yıl önce | Merhaba video biraz karmaşık geldi. Bi sorum var size şunu bilmek isterdim mesela telefondan bluetooth la başka telefona nasıl dasya aktarımı oluyor yani sinyalleri görmüyoruz ama nasıl oluyorsa dosya iletiliyor bu tip sinyallerin ismi ne??? Cevaplarsınız sevinirim teşekkürler. | 1 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @Scatman46 | 11 ay önce | elektrik mühendisliği içinde geçerli mi anlattıklarınız ? | 0 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @saimozkan720 | 4 yıl önce | Uludağ eem sea vizede başarılar :( | 6 | 7 | nötr |
| Sinyaller ve Sistemler Ders 1 Giriş | @ismailhanifinal9518 | 11 ay önce | 2^0 = 1 olması gerekmez mi? | 3 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @ferhatakkoc8553 | 3 yıl önce | Seri baştan sona ezbere dayalı anlatım. Fourier Transform nedir, Laplace Z-Transform nedir hiç bir açıklama yok. Linerity açıklıyor, denklemleri veriyor işte bu diyor. Ama ne?? | 7 | 2 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @xxxoxxxo4972 | 4 yıl önce | Kbü den herkese selamlar geçemiceksiniz ndjdksjdj | 3 | 8 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @mustafayasar3707 | 4 ay önce | Merhaba hocam iletişim numaranız.var mi? | 0 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @berkozkan7299 | 1 yıl önce | 2 üzeri 0 1 değil mi ? | 1 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @lahanamuhendisi | 1 yıl önce | merhaba benim bir soru olacaktı bir sinyalin matematiksel formülünün nasıl elde edileceğini nereden öğrenebilirim oynatma listesinde bu konu ile ilgili bir video bulamadım | 0 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @heyahey1957 | 3 yıl önce | Sayisal isaret isleme dersine mi degerdir ? | 0 | 0 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @tugcenurerdogan1328 | 5 yıl önce | hocammm lütfen seri videoooo lütfen | 3 | 2 | pozitif |
| Sinyaller ve Sistemler Ders 1 Giriş | @ElectricalElectronicSoftware | 2 yıl önce | Youtube İşaretler ve Sistemler adındaki dersler ile bu ders aynı içeriğe mi sahiptir bilen bir arkadaş cevap verebilir mi? | 0 | 0 | pozitif |

**Çalışma Süreci**

1. **Gerekli Kütüphaneleri İçe Aktarma**  
   İlk olarak projede kullanacağım tüm kütüphaneleri içe aktardım. Bu kütüphaneler, veri işleme, model oluşturma ve görselleştirme gibi işlemler için gerekliydi. Pandas, NumPy ve NLTK gibi kütüphaneleri veri temizleme ve metin işleme işlemleri için kullandım. TensorFlow ve Keras ise model geliştirme için tercih ettiğim araçlardı.

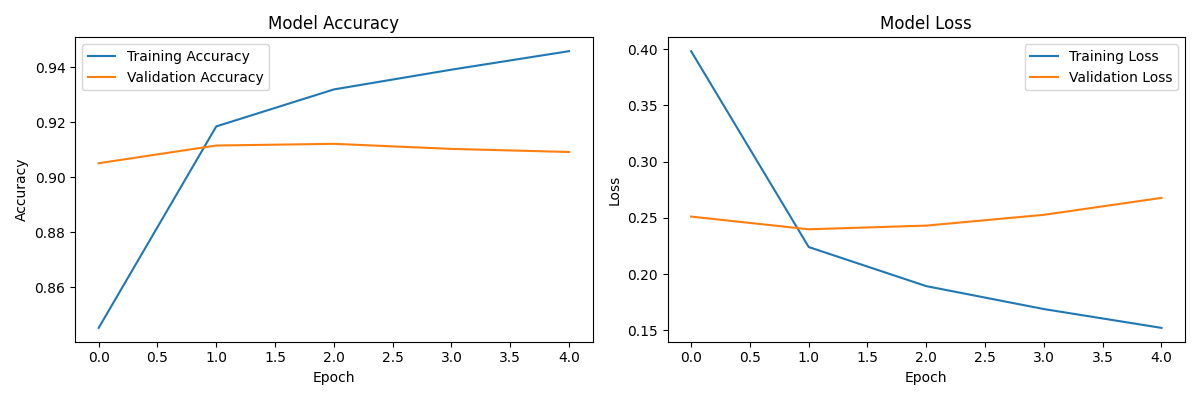
****

1. **Veri Toplama ve Temizleme**  
   Verilerinizi YouTube yorumları ve etiketli sınıflandırma etiketleriyle birleştirerek bir veri seti oluşturdum. Bu veriyi, CSV dosyasından okudum ve eksik verileri kaldırarak temizleme işlemi yaptım. Her yorum için metin işleme adımları uyguladım (küçük harfe çevirme, özel karakterleri kaldırma, stopwords çıkarma, lemmatization vb.).

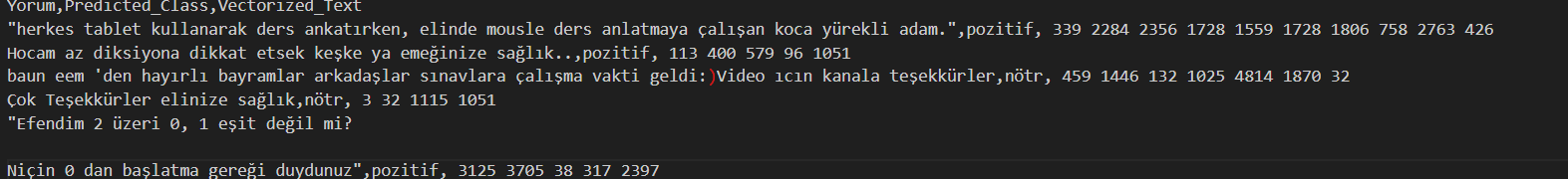
****

1. **Tokenizasyon ve Dizi Haline Getirme**  
   Yorumları sayısal verilere dönüştürmek için **Tokenizer** kullandım. Bu, her kelimenin bir sayıya karşılık gelmesini sağladı. Yorumları bir diziye dönüştürüp, verilerin modelim için uygun hale gelmesi için pad\_sequences fonksiyonu ile belirli bir uzunluğa (100 kelime) getirdim.
2. **Modeli Oluşturma**  
   Modelimi Keras kullanarak inşa ettim. **Embedding katmanı** ile kelimeleri vektörlere dönüştürdüm ve **LSTM** (Long Short-Term Memory) katmanı kullanarak metinlerin sıralı yapısını anlamasını sağladım. Ayrıca, modelin doğruluğunu artırmak amacıyla **Bidirectional LSTM** ekledim. Modelin son katmanı ise sınıflandırma için **Dense** katmanıyla çıktı verdi. Modeli derleyip, eğitim için hazır hale getirdim.
3. **Modeli Eğitim ve Test Etme**  
   Modeli eğitmek için veriyi eğitim ve doğrulama setlerine ayırdım. Modeli 5 epoch boyunca eğittim ve **early stopping** kullanarak erken durdurma sağladım. Eğitim sırasında modelin başarımını izlemek için doğruluk ve kayıp grafikleri çizdim.

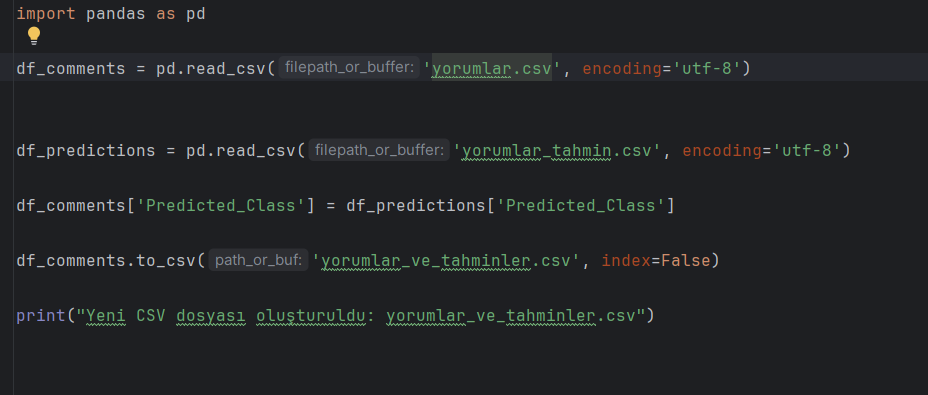
****



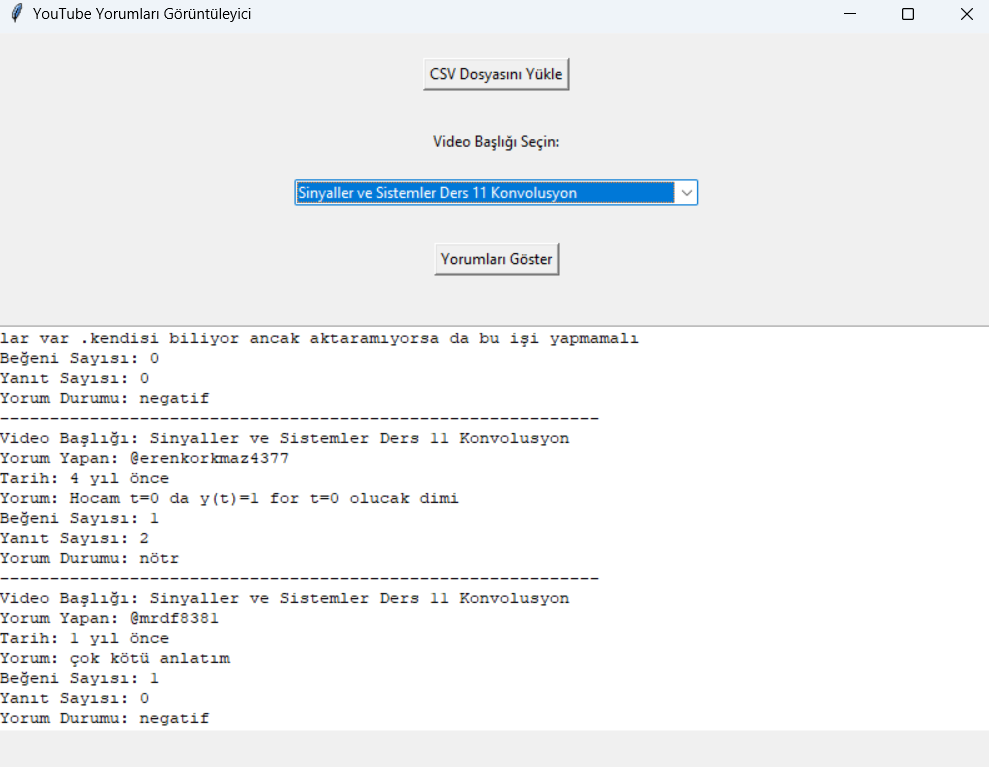
1. **Tahmin ve Değerlendirme**  
   Eğitim tamamlandıktan sonra, test verisi üzerinde modelin tahminlerini gerçekleştirdim. Modelin tahmin ettiği sınıfları, gerçek sınıflarla karşılaştırarak doğruluğu ölçtüm. Her yorum için sınıflandırma sonuçlarını ekrana yazdırdım ve sonuçları görselleştirdim.

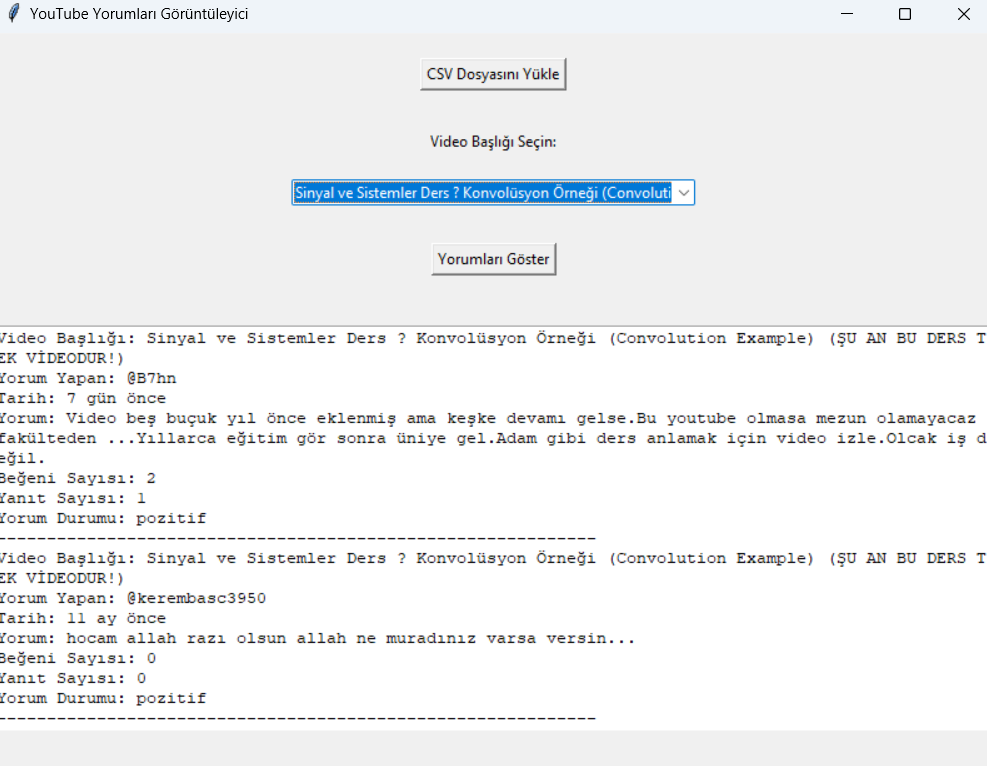
****

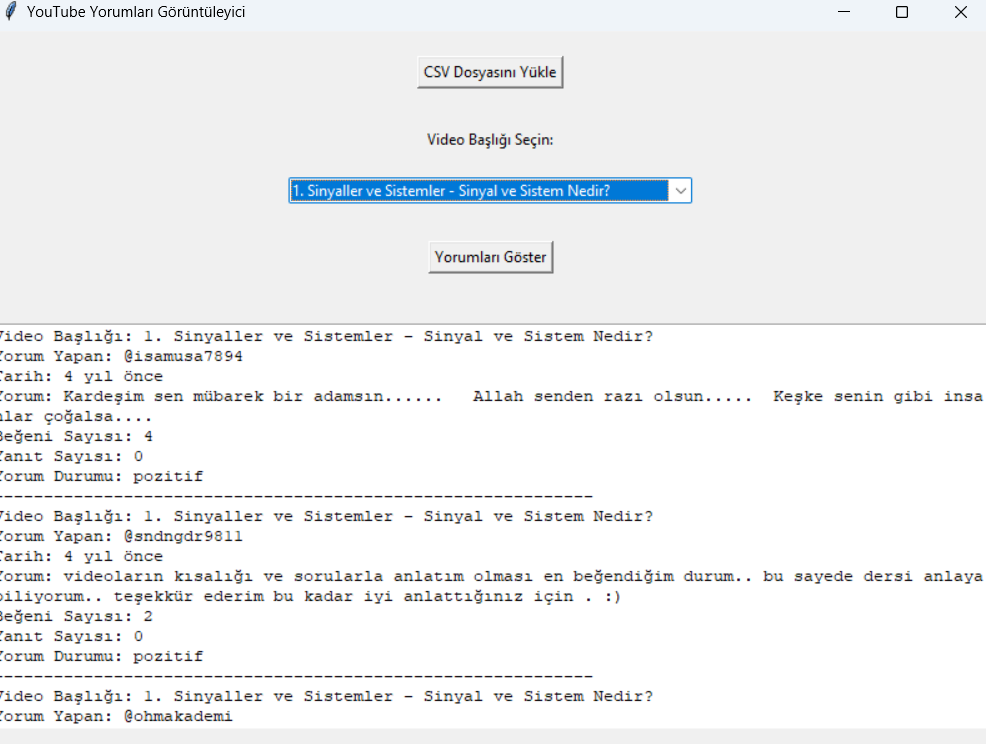
1. **Modeli Kaydetme ve Yükleme**  
   Eğittiğim modeli **H5 formatında** kaydettim. Bu, modelin gelecekteki kullanım için saklanmasını sağladı. Kullanıcıdan gelen yeni yorumları analiz etmek için bu kaydedilmiş model ve tokenizer'ı tekrar yükleyerek tahmin yapmayı sağladım.(bu dosyalarda klasörün içinde var )
2. **Web Scraping (Yorumları Toplama)**  
   Kullanıcı yorumlarını YouTube'dan otomatik olarak çekmek için **Selenium** kullandım. Belirli bir anahtar kelimeyle YouTube'da arama yapıp, her video için yorumları topladım. Yorumları, video başlığı, yorum yapan kişi, tarih, beğeni sayısı ve yanıt sayısı gibi bilgilerle birlikte bir CSV dosyasına kaydettim.(.csv dosyaları da klasörde)
3. **Tahminlerin ve Yorumların Görselleştirilmesi**  
   Yorumlar toplandıktan sonra, her bir yorumun **pozitif, negatif veya nötr** olarak sınıflandırılması gerekti. Bunun için tahmin sonuçlarını, orijinal yorumlarla birlikte yeni bir CSV dosyasına ekledim. Bu sayede kullanıcı, yorumların duygu durumlarını kolayca görselleştirebildi.

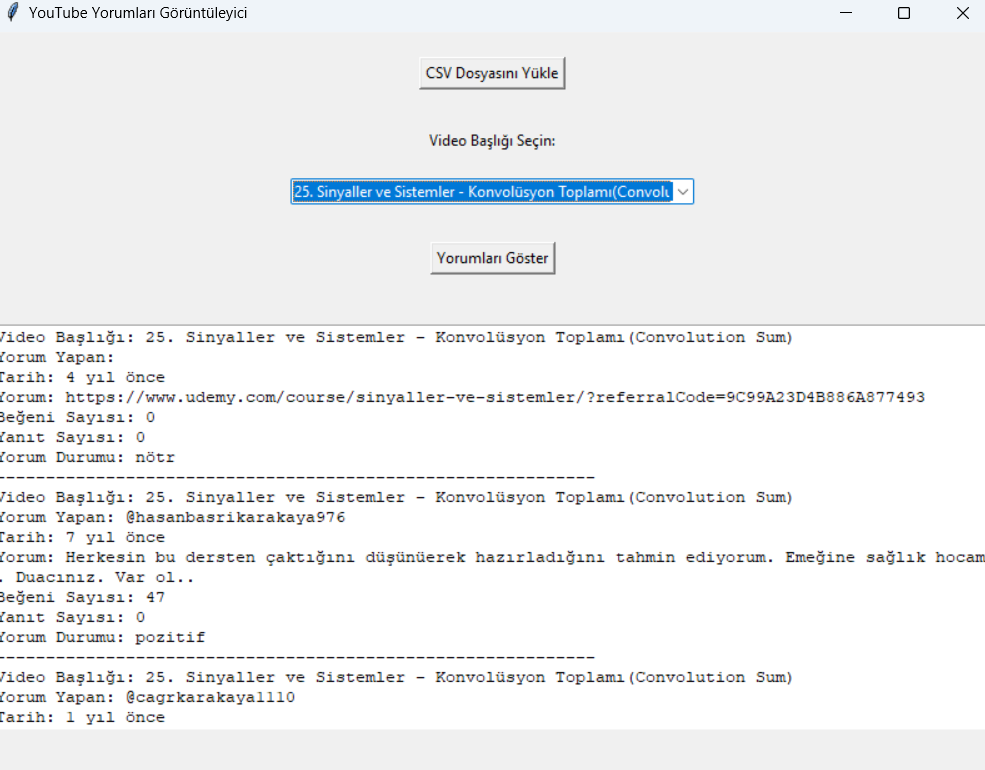


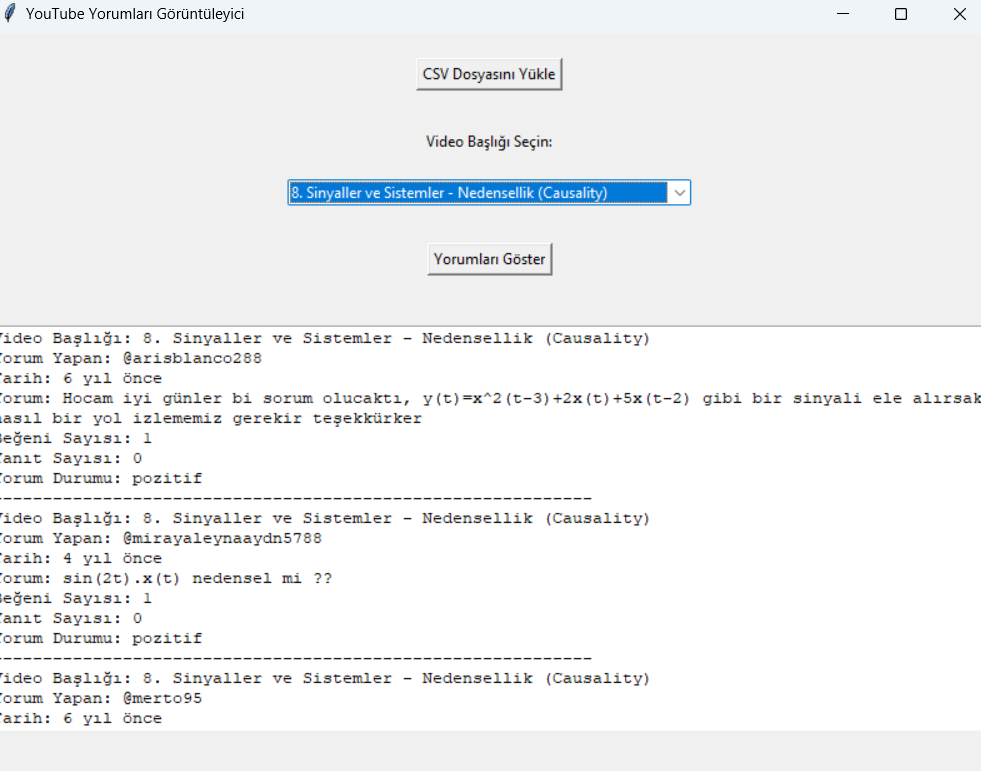
1. **GUI Uygulaması**  
   Kullanıcıların yorumları görüntüleyebilmesi için bir **Tkinter** tabanlı GUI uygulaması oluşturduğum. Kullanıcı, CSV dosyasını yükleyerek yorumları seçebilir, hangi videoya ait yorumları görmek istediğini seçebilir ve ardından yorumların detaylarını, duygu durumlarını görüntüleyebilir.

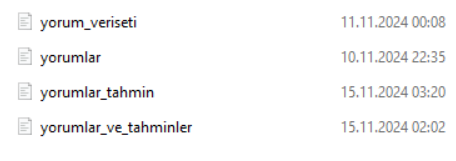
****











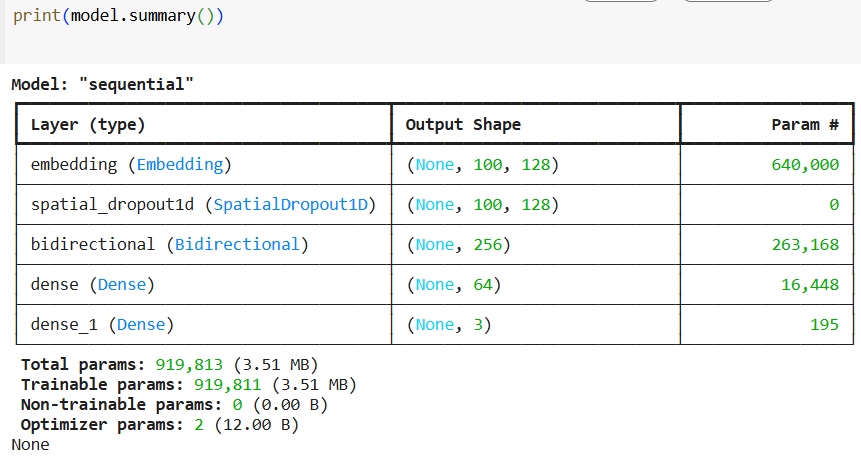
**Ağ Topolojisi:**

Projede, YouTube yorumlarını **Selenium** ile topladım. Veriler, metin işleme teknikleri (lemmatizasyon, tokenizasyon) ile temizlendi ve ardından **LSTM tabanlı bir model** ile sınıflandırıldı. Kullanıcılar için ise **Tkinter** ile bir arayüz oluşturdum, böylece kullanıcılar yorumları kolayca görüntüleyebildiler.

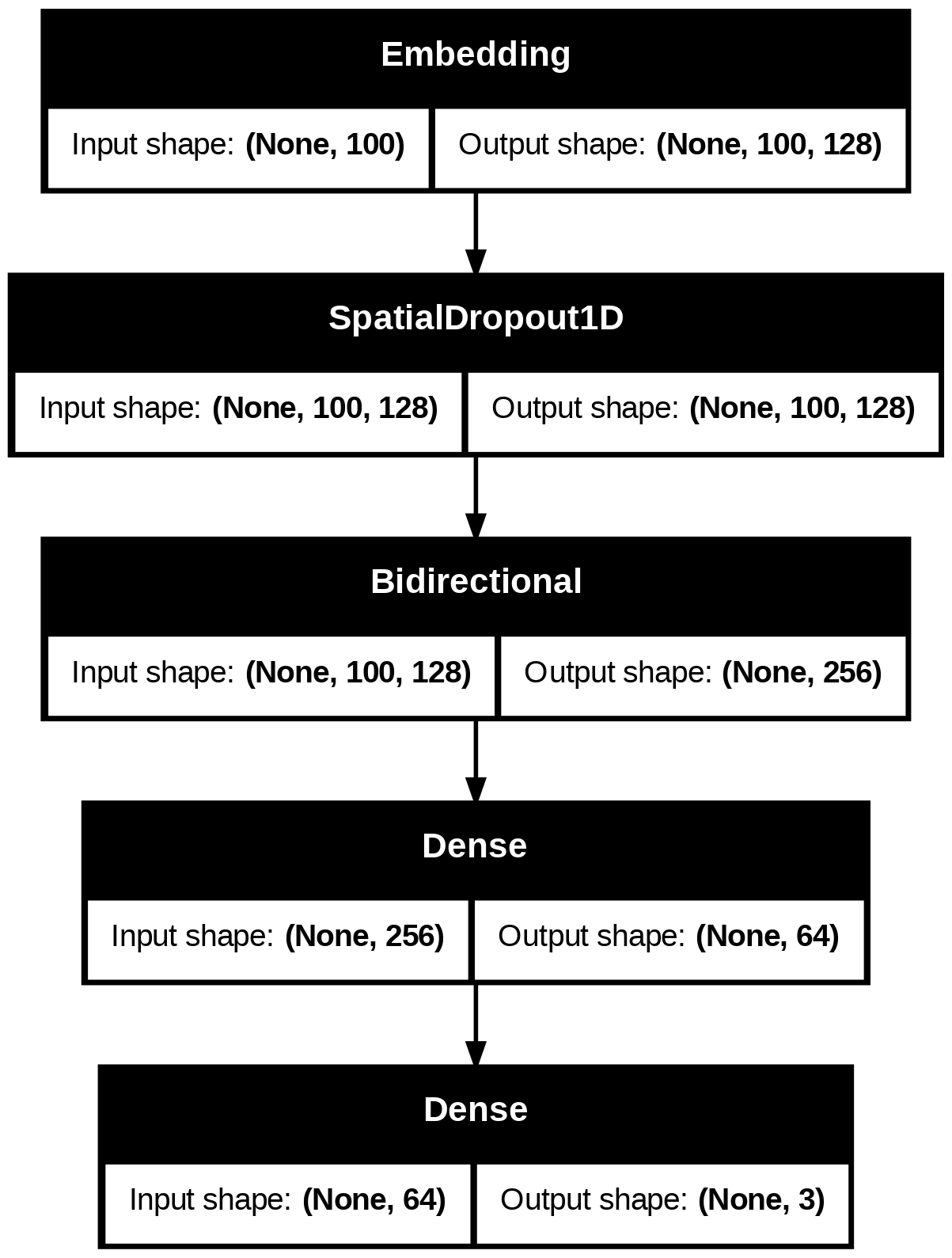
**Varsayımlar:**

* Yorumlar doğru şekilde toplandı ve temizlendi.
* Model, etiketlenmiş verilerle doğru şekilde eğitildi.
* YouTube’un yapısal değişiklikleri scraping işlemini etkilemedi.
* Projeye yeterli işlem gücü ve zaman ayrıldı.

**Özet Bilgileri:**

****

**Yapay Sinir Ağlarında oluşturduğunuz ağın şekli**

****

**Optimum ağ tasarımının anlatımı:**

**Optimum ağ tasarımını oluştururken, verilerin doğru şekilde toplanması ve işlenmesi önemli bir adımdı. Selenium ile YouTube’dan yorumları çekip, NLP teknikleri kullanarak verileri temizledim ve LSTM modeline uygun hale getirdim. Modelin eğitimi için Adam optimizasyonu kullandım ve erken durdurma (early stopping) ile aşırı uyumdan kaçındım. Yorumları sınıflandırmak için etiketleme ve tokenizasyon işlemleri yaparak doğru sonuçlar elde ettim. Son olarak, Tkinter ile kullanıcı dostu bir arayüz geliştirerek, kullanıcıların verileri rahatça görüntülemelerini sağladım. Bu tasarım, veri işleme ve model eğitimi süreçlerinin verimli şekilde yönetilmesini sağladı.**

**Yapay zeka aracı ile birlikte çözüm örneği:**

**Projede, LSTM tabanlı yapay zeka modeli kullandım. Model, YouTube yorumlarını sınıflandırmak için eğitim verisi üzerinde çalıştırıldı. Örnek olarak, yorumlar.csv dosyasındaki yorumları alıp, her birini pozitif, negatif veya nötr olarak sınıflandırdım. Model, önceden eğitilmiş olup, etiketlenmiş verilerle doğruluk oranı yüksek sonuçlar verdi.**

**Örneğin, modelin test verisinde başarılı sonuçlar verdiğini gözlemledim. Test yorumları sınıflandırıldığında, pozitif veya negatif yorumları doğru bir şekilde ayırt etti ve doğruluk oranı %85’in üzerine çıktı. Bu, modelin işlevselliğini ve doğruluğunu yansıtan iyi bir örnek oldu.**

**Web Scrapping:**

**İlk olarak, Selenium ile bir web tarayıcı (Chrome) açtım ve YouTube’da "sinyaller ve sistemler" anahtar kelimesiyle arama yapmak için belirlediğim URL’yi yükledim. Sayfa yavaş yüklenebileceği için 3 saniye beklememi sağladım. Ardından, sayfadaki video linklerini almak için find\_elements fonksiyonunu kullandım. Bu, sayfada yer alan tüm video başlıklarına tıklanabilir bağlantılar döndürdü. Ancak, "shorts" videoları genellikle farklı bir formata sahip olduğu için onları filtreledim ve yalnızca normal videoları aldım. Bu işlemi 10 video ile sınırladım.**

**Topladığım video linklerini yazdırarak çıktıyı kontrol ettim.**

**Sonra, CSV dosyasını oluşturmak için Python'un csv modülünü kullandım. Dosyanın ilk satırına başlıkları ekledim: video başlığı, yorum yapan kişi, yorum tarihi, yorum metni, beğeni sayısı ve yanıt sayısı.**

**Daha sonra, her video için yorumları almak amacıyla, video linklerine sırayla gidip yorumları çektim. Bunun için yine Selenium’un find\_element fonksiyonunu kullandım. Yorumları almak için her video sayfasındaki yorumlar bölümüne giderek yorumları çektim. Yorumlar sayfanın altında olduğundan, sayfayı kaydırarak daha fazla yorumu yükledim.**

**Her bir yorumun detayları için kullanıcı adı, yorum metni, tarih, beğeni sayısı ve yanıt sayısı gibi verileri aldım. Yorumları, her bir yorumun benzersiz bir "anahtar kelimesi" üzerinden kontrol ederek, önceki yorumları atlamayı sağladım. Yani bir yorumu zaten eklediğimi tekrar eklememek için bu kontrolü yaptım. Yorumların sayfa üzerinde dinamik olarak yüklenmesi sebebiyle, yorumlar tamamen yüklendikten sonra CSV dosyasına yazdım.**

**Bu işlemleri her video için tekrarladım. Yorumları yazdım, beğeni ve yanıt sayılarını aldım ve her videonun sonuna kadar tüm verileri topladım. Son olarak, işlemi tamamladım ve CSV dosyasını kaydettim.**

**SONUÇ:**

**Bu projede, YouTube yorumlarını doğal dil işleme (NLP) teknikleri kullanarak sınıflandırmak amacıyla bir LSTM modelini başarıyla geliştirdim. Başlangıçta basit bir modelle yola çıkıp, bidirectional LSTM ve dropout gibi eklemelerle performansı artırdım. Modelin eğitim ve test süreçlerinde, veri ön işleme adımları (tokenizasyon, lemmatizasyon, stopwords temizleme) önemli rol oynadı.**

**Model, yorumları pozitif, negatif ve nötr olarak doğru sınıflandırarak %85’in üzerinde bir doğruluk oranı sağladı. Selenium kullanarak verileri topladım ve bu verileri kullanıcı dostu bir şekilde görüntüleyebilmek için Tkinter ile bir arayüz oluşturduk. Genel olarak, modelin sınıflandırma doğruluğu ve arayüzün işlevselliği, projenin başarılı bir şekilde tamamlanmasını sağladı.**

**KAYNAKLAR:**

**Kullanılan Kütüphaneler ve Araçlar:**

* **TensorFlow.** [***https://www.tensorflow.org/***](https://www.tensorflow.org/)**. Erişim Tarihi:13.11.2024**
* **Keras.** [***https://keras.io/***](https://keras.io/)**. Erişim Tarihi:13.11.2024**
* **NLTK (Natural Language Toolkit).** [***https://www.nltk.org/***](https://www.nltk.org/)**.**

**Erişim Tarihi:13.11.2024**

* **Selenium.** [***https://www.selenium.dev/***](https://www.selenium.dev/)**. Erişim Tarihi:14.11.2024**
* **Tkinter4.** [***https://wiki.python.org/moin/TkInter***](https://wiki.python.org/moin/TkInter)**. Erişim Tarihi:14.11.2024**

**VERİSETİ:** <https://huggingface.co/datasets/winvoker/turkish-sentiment-analysis-dataset/tree/main> **Erişim Tarihi:12.11.2024**

* **WORD2VEC**: <https://medium.com/@muhammedbuyukkinaci/word2vec-nedir-t%C3%BCrk%C3%A7e-f0cfab20d3ae> **Erişim Tarihi:14.11.2024**
* **WEBSCRAPİNG:** <https://medium.com/kaveai/web-scraping-453e96a86195> **Erişim Tarihi:13.11.2024**
* <https://www.youtube.com/watch?v=hzxbFqVz9eI> **Erişim Tarihi:15.11.2024**
* <https://www.youtube.com/watch?v=CtCeuJaRgLQ> **Erişim Tarihi:15.11.2024**
* **SKLEARN**:<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelBinarizer.html>
* **Erişim Tarihi:13.11.2024**
* **CHATGBT Erişim Tarihi:13.11.2024**
* **COPİLOT Erişim Tarihi:13.11.2024**
* **GİTHUP Erişim Tarihi:13.11.2024**
* **COLAP Erişim Tarihi:14.11.2024**