

[**Mühendislik ve Mimarlık Fakültesi**](https://www.konya.edu.tr/muhendislikvemimarlik)

[**Bilgisayar Mühendisliği Bölümü**](https://www.konya.edu.tr/bilgisayarmuhendisligi)

**Bilgisayar Mühendisliği Uygulamaları 2 Dersi Proje Formu**

|  |
| --- |
| **Proje Başlığı** |
| TAHMAK Makine Öğrenmesi |

|  |  |
| --- | --- |
| **Öğrenci Bilgileri** | |
| **Öğrenci No** | 18010011014 |
| **Öğrenci Ad Soyad** | Ömer Faruk Kolcu |

|  |
| --- |
| **Dr. Öğr. Üyesi**  **Murat Karakoyun** |

**Haziran 2023**

**Konya**

**İçindekiler**

1. Proje Konusunun Önemi ve Özgün Değeri 3

2. Projenin Amacı ve Hedef 3

3. Projenin iş-zaman çizelgesi 4

4. Projede kullanılan donanımlar ve yazılımlar ile ilgili bilgileri 5

4.1. Donanımlar 5

4.2. Yazılımlar 5

5. Projenin yapım aşamaları 7

Kaynaklar 26

# Proje Konusunun Önemi ve Özgün Değeri

**Özgün Değer**

Makine öğrenmesi algoritmalarını kullanarak paylaşılacak YouTube videolarının kaç sayıda izlenme, beğeni ve beğenememe alacağını öngören bir web uygulaması tasarlamak.

# **Projenin Amacı ve Hedefi**

Kaggle üzerinden alınan örnek veri setleri ile makine öğrenmesini sağlamak. Bu makine öğrenmesi kullanıcıların paylaşacağı videoların kaç izlenme, beğeni ve beğenememe alacağını öngörerek videolarınız için sponsorluk anlaşmaları yapmanızı sağlayabilir. Proje de Python kütüphaneleri ile makine öğrenmesi yapıldıktan sonra bu makineyi Django ile bir web uygulamasında kullanıma sunmak.

# Projenin iş-zaman çizelgesi

İş-zaman çizelgesi

|  |  |  |  |
| --- | --- | --- | --- |
| İP.  No. | İş Paketlerinin Adı ve Hedefleri | Zaman Aralığı (Haftalık) | Başarı Ölçütü |
| 1 | Makine öğrenme modelinin oluşturulması. | 1., 2. ve 3. haftalar | Kullanıcıların etiketler kullanarak paylaştıkları videoların izlenme, beğenme ve beğenmeme sayısına bağlı olarak bir makine öğrenmesi gerçekleştirmek. Başarı ölçütü %30. |
| 2 | Abone sayısının izlenmeye etkileri. | 4. ve 5. haftalar | Sadece etiketlerin değil abone sayısının da izlenme üzerindeki etkilerini araştırılarak veriler üzerinde uygulanması.%20 |
| 3 | Web uygulama arayüzünün oluşturulması. | 6. ve 7. haftalar | Html, Css ve Bootstrap5 ile arayüz oluşturulması. Başarı ölçütü %15. |
| 4 | Veri setlerinin filtrelenmesi ve makine öğreniminin sağlanması. | 8., 9. ve 10. haftalar | Kullanıcının ilgi alanına göre değerlendirmeler yapmak için tüm verilerimi içeren veri setine filtreler uygulamak ve makine öğrenimini gerçekleştirmek. Başarı ölçütü %25. |
| 6 | Uygulamanın tamamlanması. | 11. ,12.,13. ve 14. haftalar | Makine öğrenmesi sonucunda oluşturulan veri setlerinin uygulamaya dahil edilmesi ardından istenilen sonuçların döndürülmesi. Başarı ölçütü %20. |

# Projede kullanılacak yöntem, donanımlar ve yazılımlar ile ilgili bilgiler

Programlama dili olarak: Python

Python kütüphanesi ve framework olarak: NumPy, Pandas, Matplotlip ve framework olarak django kullanacağım.

Web programlama için HTML, CSS, Boostrap

HTML: Web tasarım konusunda araştırma yapan hemen herkesin karşına çıkan temel kavram HTML’dir. Web sayfalarını oluşturma aşamasında kullanılan standart bir metin işaret dili olan HTML açılımı “Hyper Text Markup Language” olarak bilinir. Daha açık anlatmak gerekirse, Chrome, Firefox, Yandex gibi tarayıcıların okuyup anlamlandırdığı dil HTML dilidir.[7]

CSS: Web sitelerinin görsel olarak şekillendirilmesine olanak tanıyan ve kendine has kuralları olan bir tanım dilidir şeklinde ifade edilebilir. HTML ve JavaScript ile en temel web teknolojileri arasında bulunan CSS, web sayfaları üzerinde oldukça fazla görsel denetim sunar.[8]

metin, küçük resim, döşeli, boyanmış içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 1.0**

Boostrap: Bootstrap, kullanılabilir kod parçalarından oluşan açık kaynaklı ve ücretsiz bir web uygulaması geliştirme araç takımıdır. HTML, CSS, Less, Sass ve JavaScript ile yazılmış olan Bootstrap, tamamen etkileşimli ve duyarlı web uygulamaları geliştirmek için kullanılabilecek öğrenmesi kolay bir alternatiftir.[

9]

takvim içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 1.1**

NumPy: NumPy, doğrusal cebir, fourier dönüşümü ve matrisler alanında çalışmak için gerekli işlevlere sahiptir. Açık kaynak kodlu bir projedir ve özgürce kullanabilirsiniz. NumPy, Numeriacal Python (Sayısal Python)’un kısaltmasıdır.[10][11]

logo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 1.2**

Pandas:Veri işlemesi ve analizi için Python programlama dilinde yazılmış olan bir yazılım kütüphanesidir. Bu kütüphane temel olarak zaman etiketli serileri ve sayısal tabloları işlemek için bir veri yapısı oluşturur ve bu şekilde çeşitli işlemler bu veri yapısı üzerinde gerçekleştirilebilir olur.[12]

çizelge içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 1.3**

Matplotlib: Matplotlib; veri görselleştirmesinde kullandığımız temel python kütüphanesidir. 2 ve 3 boyutlu çizimler yapmamızı sağlar. Matplotlib genelde 2 boyutlu çizimlerde kullanılırken, 3 boyutlu çizimlerde başka kütüphanelerden yararlanılır.[13]

logo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 1.4**

Django: Python Programlama Dili için hazırlanmış ve BSD lisansı ile lisanslanmış yüksek seviyeli bir web çatısıdır(framework). Basit kurulumu ve kullanımı, detaylı hata raporu sayfaları ve sunduğu yepyeni arayüz ile diğer sunucu yazılımı ve frameworklerden kendini ayrıştırmaktadır. İsmi, caz gitaristi Django Reinhardt’tan gelmektedir. [14]

web sitesi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 1.5**

**Yapım Aşamaları:**

1. **Hafta**:

Danışman hocamız ile birlikte proje hakkında görüşmelerde bulunduk. Önceki dönemden kalan projemizin eksikliklerinin giderilmesi ve projenin geliştirilmesi konusunda neler yapacağımıza karar verdik. Bu durumda kullanılan makine öğrenmesi tekniklerinde birkaç tanesinin karşılaştırılmasına karar verdik. Oluşan sonuçlardan hangisi öğrenme yönteminin sonuçları daha doğru sonuçlar veriyor ise onu kullanmaya karar verdik.

1. **Hafta**:

Bu hafta veri setimi aldığım kaggle sitesinde insanların, veri seti üzerinde yazdıkları kodları ve yaptıkları değerlendirmeleri inceledim. Bir kullanıcı veriler üzerinde normalizasyon, eksik veri tamamlama gibi veri önişleme işlemlerini gerçekleştirdiği kodları paylaşmış. Yine aynı kullanıcı veri seti üzerinde makine öğrenmesi methodları gerçekleştirmiş. Ancak bu veri seti kullandığım veri setinden farklı olmalı ki gizli bir şekilde paylaşmış. Bu çalışmada rassal ağaç makine öğrenme yönteminin diğer yöntemlere göre daha doğru sonuçlar çıkardığını açıklıyordu. Bende bu durumu değerlerndirmek için makine öğrenmesi yöntemlerinden doğrusal regresyon yöntemini, karar ağacı yöntemini ve son olarak da rassal ağaç yöntemini test etmeye karar verdim. Bu durumda hangisinin daha doğru sonuçlar verdiğine göre uygulamamda kullanacağım yöntemi tercih edeceğim. Makine öğrenme yöntemlerinden doğrusal regresyon yöntemi ile veri setimi eğitmek için çalışmalara başladım. Doğrusal makine öğrenme yönteminde iki farklı değişken kullanılarak öğrenme gerçekleştirilir. Mesela etiket verisi üzerinden beğenme tahmini gerçekleştirilirebilir. Bu durumda etiket verisi üzerinden beğenmeyi, beğenmemeyi ve izlenme verileri ayrı ayrı tahmin ettikten sonra bu tahminleri bir araya getirmek şemasını oluşturdum. Bu yüzden verilerin kullanacağım sütunlarını pandas kütüphanesinin drop methodunu kullanarak verileri ayırdım.

1. **Hafta**:

Geçen hafta oluşturduğum şemanın uygulanmasında bir hata ile karşılaştım. Bu hatada verilerimin eğitilmesi sırasında oluşan bir hataydı. Verilerimin (1,-1) arasından düzenlenmesi hatanın giderilmesi için yeterli olduğunu söylüyordu. Bende verilerimi birer numpy dizisine çevirdikten sonra reshape methodunu kullanarak verilerimi düzenledim.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 1. İlk karşılaşılan hata.

metin, kara tahta, levha, plaka içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 2. Verileri 2 boyutlu dizi haline getirme.

Ancak bu sefer de eğitim verisi sayısı hakkında bir hata alıyorduk. Bu hata da eğitimi gerçekleşmiş makine öğrenmesini kullanmak istediğim zaman gerçekleşiyordu. Makine öğrenmesinde veri setindeki verilerin kaç tanesinin öğrenmeye ayrılacağını belirlememe rağmen, öğrenmeyi test ederken test verisinde beklenen sayının gelmediğini söyleyen bir hata alıyordum. Bu hatayı araştırdım ancak tam olarak bir sonuca varamadım.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 3. Verilerin 2 boyutlu dizi haline getirdikten sonra alınan hata.

1. **Hafta**:

Bu hafta danışman hocam, yaptığımız görüşmede hatanın veriyi uygun şekilde düzenlemediğim için olmuş olabileceğini söyledi. Bende Btk Akademi üzerindeki eğitimi açarak orada verilerin nasıl hazırlandığını tekrardan inceleme yaptım. Verileri iloc ile ayırmayı denemiştim zaten ancak doğru sonuca ulaşamamıştım. En baştan eğitim sürecinde kullanılan veri seti ile birkaç örnek gerçekleştirdikten sonra hatamı tam olarak tespit edememek ile birlikte hataya çözüm buldum. Bu süreçte veri setlerini ayarlamak için verileri iloc komutu ile ayırdım.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 8. Verilerin doğru şekilde ayrılması.

Bu hafta hatalarımın hepsini düzeltmiş bir şekilde makine öğrenmelerini tekrar gerçekleştirmeye başladım. Bu hafta içerisinde doğrusal regresyon ile oluşan sonuçların çıktısını aldık. Bu öğrenmede verilerimin 0.33 oranında eğitim verisi olarak ayarladım ve eğitimi gerçekleştirdim. Verilerin eğitim ve test olarak ayrılmasında sklearn.model\_selection modülünden train\_test\_split fonksiyonunu kullandım. . Ardından bir doğrusal regresyon sınıfı tanımladım. Eğitim verilerimin üzerinden eğitimi gerçekleştirdim. Bu öğrenmede y\_test verilerinin x\_test verilerine, öğrenimin uygulanmış sonuçları ile karşılaştırma yapınca sonuçların birbirinden çok farklı olduklarını gözlemledim. Eğitimi gerçekleştirirken ise sklearn.linear\_model modülünden LinearRegression fonksiyonunu kullandım.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, dış mekan, levha, plaka içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 4. Şekil 5.

Şekil 1’de x\_test verilerinin, makine öğrenmesine göre sonuçları gösterilirken, Şekil 2’de ise y\_test verileri gösterilmiştir. Şekil 2’de 1. sütunda bulunan veriler etiket verisinin label encoder edilmiş halidir. 2. sütünda yer alan veriler ise o etiketlerin gerçekte ne kadar beğenildiği verileridir. Birbiri ile aralarındaki farklar çok fazla olduğu için bu makine öğrenmesini yönteminin benim uygulamamın doğruluğunu çok düşüreceği için kullanmaktan vazgeçiyorum.

1. **Hafta**:

Bu hafta daha doğru sonuçlar alacağımı düşündüğüm karar ağacı öğrenmesini gerçekleştireceğim. Karar ağacı öğrenmesi bir sınıflandırma öğrenmesidir. Verilerimi tekrar düzenledikten sonra sklearn.tree kütüphanesinden DecisionTreeRegressor modülünü programıma dahil ediyorum. Bu modül üzerinden bir r\_dt adında sınıf oluşturuyorum. Bu sınıfı fit methodu ile öğrenme gerçekleştiriyorum. Öğrenmeyi label encoder edilmiş etiket verisinden beğenmeyi tahmin etme şeklinde gerçekleştiriyorum. Test gerçekleştirdiğim zaman veri setindeki herhangi bir veri için gerçeğe yakın sonuçlar elde ediyorum.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 6. Verilerin ayrılması.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 7. Karak ağacı uygulaması.

Karar ağacı uygulamasının değerlendirilmesinde veri seti dışarısından bir etiket verisinin gelmesi halinde, etiketin herhangi bir sayısallaştırmasının yapılmayacı için bir problem ile karşılaşıyoruz. Burada veri setinde ki label encoder oluşturduğumuz en son veriden, sayısal olarak daha büyük birkaç veri girersek hepsi için aynı sonuçları alıyoruz. Bu durumda algoritmamız istediğimiz sonuçları göstermekte yetersiz kalıyor.

Doğrusal regresyon kulanılırken de aynı sorun ile karşılaşıyoruz. Sayısal olarak üretilen verilerin gerçekten bir etikteti temsil edebiliyor olsa bile bu etiketin kullanılması bizlere doğru sonuçlar vermeyecektir. Bu durumu danışmanım ile görüştüm.

1. **Hafta**:

Danışmanım ile yaptığımız görüşmede fikir üzerinde birkaç değişiklik yapılmasını önerdim. Kendisi de bir fikir geliştirip uygulayabilirsem bunu konuşalım dedi. Bende YouTube üzerinde paylaşılan videoların izlenmelerinin aslında etiketlerden ziyade abone sayısı ile alakası oluduğunu fark ettim. Ardından kaggle üzerinde bir araştırma yaptım. Bu araştırmada YouTube üzerindeki popüler kanalların sahiplerini, abone sayısını, toplam video görüntülenme sayısını, toplam video sayısını, kategorisini ve kanalın ne zaman açıldığını gösteren bir veri seti buldum[15]. Bu veri seti içerisindeki sayısal verilerin değerleri çok yüksek olduğu için veri tiplerini değiştirdim.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 8. Veri tipleri int64 ile değiştirildi.

Niteliklerde ki eksik verileri ve değeri 0 girilen verileri, niteliğin medyan değeri ile doldurdum. Total izlenme ve toplam video sayısı nitelikleri üzerinden kanalın bir videosunun ortalama ne kadar izlendiğini hesapladım ve veri setine bir nitelik olarak ekledim.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 9. Ortalama izlenme hesaplaması ve veri setine eklenmesi.

Ardından doğrusal regresyon ve karar ağacı algoritmalarını uyguladım. Burada oluşturduğum fikir ise kullanıcının içerik üretmekte kararsız kaldığı zamanlarda abone sayısı üzerinden kaç izlenme alacağını tahmin eden bir makine sağlamak. Bu tahmin sonucunda ise aynı izlenmeleri alan birkaç farklı kanalların içeriklerinin neler olduğunu öneri şeklinde sunan bir program oluşturmak. Aynı zamanda kaç beğeni ve beğenmeme alacağınıda kullanıcıya etiket ve konu önerilerinin yanında bir fikir olarak sunmak. Ardından danışmanım ile tekrar görüştük ve fikirin uygulanabilir olduğu konusunda anlaştık. Ancak vize haftasına az bir süre kaldığı için mobil uygulama yerine web uygulaması yapmaya karar verdik.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 10. Doğrusal regresyon ve karar ağacı uygulaması.

Karar ağacı uygulaması veri seti dışındaki veriler için doğru sonuçlar oluşturamadığı için, kullanıcının abone sayısının, veri setindeki en büyük ve en küçük abone sayılarının arasında olup olmamasına göre farklı algoritmaları uyguladım.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 11. Algoritmaların uygulanması.

Eğer abone sayısı, veri seti içerisindeki abone sayısı niteliğinde var olabilirse karar ağacı algoritmasını, değilse doğrusal regresyon uygulamasını gerçekleştirdim. Ancak bu algoritmalarında doğruluk değerleri oldukça düşük. Bir uygulama sunabilmek için bu algoritmaları kullandım. Ayrıca abone sayısına ulaştığım veri setindeki videoların popüler videolar olmasından kaynaklı büyük bir fark oluşuyor. Daha verimli çalışacağım bir veri seti içi araştırmalar gerçekleştiriyorum. Bu farktan dolayı oluşan tahmin sonuçlarının diğer veri setinde bir karşığı olmuyor. Bende bu durumda izlenme tahmin sonucunu ile gerçekleştireceğim aralığı ayarlamakta zorluk çektim.



Şekil 12. Diğer veri setinde arama yapabilmem için bir aralık belirliyorum.

Hala tam olarak doğru sonuçlar üretemiyorum ancak uygun veri setleri ile bu sorunun çözüleceğine eminim.

1. Hafta:

Django ile web uygulamamı gerçekleştirildi. PyCharm idesinde terminal üzerinden, Django-admin startproject bitirme komutu ile bitirme adında bir proje oluşturuyorum. Ardından startapp komutu ile pages isminde bir uygulama oluşturuyorum. Bu uygulamayı settings.py dosyasında ana projeye tanıtıyorum. Bunun için installed\_app kısmına pages uygulamasının apps.py dosyasındaki sınıfı tanıtıyoruz.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 13. Uygulamamızı settings.py da tanımladık.

Ardından pages uygulaması içerisinde views.py dosyasında hangi html sayfasını göstereceğimizi belirtiyoruz. Bu metotu çalıştırmak için uygulamanın urls.py dosyasında belirtmemiz gerekiyor. Bu dosyayı biz oluşturuyoruz.

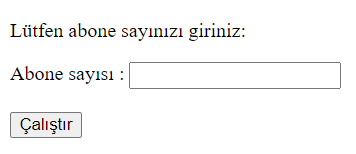
metin içeren bir resim

Açıklama otomatik olarak oluşturuldu metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

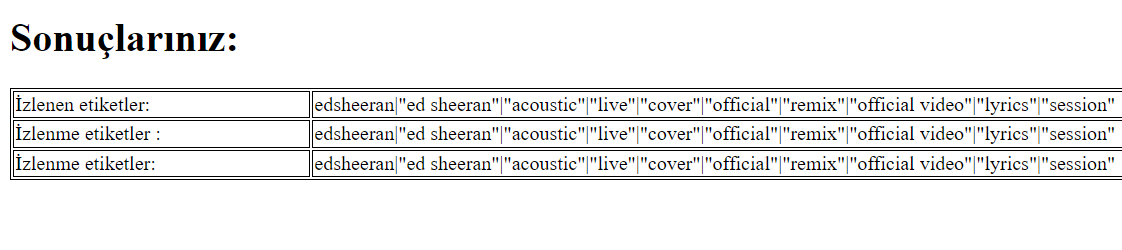
Şekil 14. Views.py şekil 15. Urls.py

Urls.py sayfasında Django.urls kütühanesinden path modülünü import ediyoruz. Böylece kullanıcının hangi uzantı eklentisiyle hangi sayfayı açacağını belirtiyoruz. Ardından projeye urls.py’ı tanıtıyoruz. Ardından iki web sayfası tasarlıyorum. Bu web sayfaları geçen dönemki web sayfalarım ile aynı. İlk oluşturduğum sayfada kullanıcıdan abone sayısını girmesini istiyorum. Ardından bir çalıştır butonu ekliyorum.



Şekil 16. İndex.html sayfası.

Ardından çalıştır butonuna basılınca gidilecek web sayfasını hazırlıyorum. Bu sayfada kaç tane öneri sunacağımı belirliyorum ve ona göre bir tablo oluşturuyorum. Bu tabloda tahmin edilen izlenme sayısına en yakın değere sahip verilerin etiketlerini göstereceğim.



Şekil 17. About.html sayfası.

Ardından views.py dosyasında tüm kodlarımı birleştiriyorum. Kullanıcının girdiği abone sayısını bir değişkende tutuyorum. Ardından makine öğrenmesi algoritmalarını gerçekleştiriyorum. Kullanıcının tahmin edilen izlenme sayısına bağlı olarak başka bir veri setindeki o izlenmeye yakın izlenmeleri bir veri setinde tutuyorum. Ardından veri setindeki değişkenlerin etiket özelliklerini bir liste değişkeninde tutuyorum. Verileri web sayfasına gönderebilmek için bir sözlük değişkeni oluşturuyorum. Bu değişken içerisinde üç farklı eleman oluşturup her birinde liste değişkenimdeki ilk elemanı tutuyorum. Ardından about.html sayfasını render ederken sözlük değişkenimi de render ediyorum. Böylece html kodları içerisinde kullanabilir hale getiriyorum.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 18. Sözcük değikeninin oluşturulması ve render edilmesi.

metin içeren bir resim

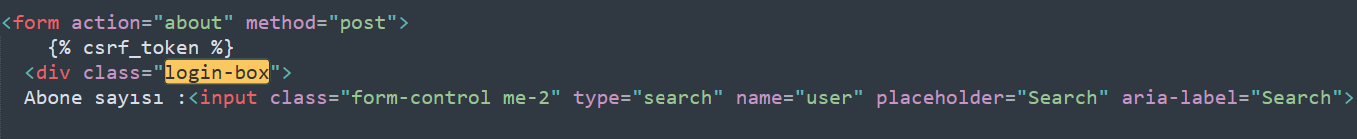
Açıklama otomatik olarak oluşturuldu

Şekil 19. HTML içerisinde, render edilen değişkenlerin kullanılması.

Böylelikle sonuçlarımı web sayfamda görüntülüyorum.

1. Hafta:

Bu hafta tasarımda güncellemeler yapmaya kara verdim. Bu güncellemeler HTML, CSS ve Bootstrap5 kullanılarak gerçekleştirilecek. Öncelikle sayfanın nasıl bir görüntüsü olmasına karar verildi ve bu basit bir çizimle gerçekleştirildi. Önceki tasarımda abone sayısı girdisini aldığımız form yapısını sayfanın ortasına almak istiyorum. Ancak öncesinde daha güzel görünen bir yapı oluşturulması gerekiyor. Bu kısımda Bootstrap5’ten faydalanılacak. Aynı zaman da CSS ile düzenlemelerde gerçekleştirilecek.



Şekil 20. Html kodumuz.

Şekil 20 de görüldüğü gibi form yapısının içerisine bir div etiketi açıldı. Ardından içerisine Bootstrap5’in elemanlarından olan “form-control me-2” sınıfına ait eleman ile girdi kısmı oluşturuldu.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 21. CSS kodumuz.

Şekil 21 de görüldüğü gibi login-box etiketi kullandığımız eleman CSS kodları ile düzenlendi.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 22. Güncel form yapısı.

1. Hafta:

Bu hafta girdi yapısını oluşturduktan sonra aynı renklere sahip çıktı sayfası da oluşturulması planlanıyor. Sayfanın orta üst kısmına tahmin sonucu döndürülecek.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 23. İzlenme tahmin sonucunun yazdırılması.

Burada bir konteynır oluşturup bu konteynır, sayfanın üst kısmına varsayılan olarak yerleştirildi. Bunun ardından metin etiketi ile metin eklendi. Metnin sonuna da tahmin değişkeni yazdırıldı.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 24. Tahmin sonucu CSS düzenlemesi.

CSS ile düzenlemeler gerçekleştirildi. Bu düzenlemeler ile konteynırın köşelerinin oval bir görüntüde olmasını, margin değerini, padding değerini ve rengi ayarlandı.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 25. Tahmin sonucu çıktısı.

Tahmin sonucununa yakın olan izlenmeye sahip videoların belirli özelliklerinin çıktısını ise sayfanının ortasına tablo şeklinde çıktı olarak verilecek.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 26. Tablonun HTML kodu.

Tabloyu oluştururken Bootstrap5 kütüphanesinden faydalanıldı. Tablonun öncelikle hangi değerleri döndüreceğine göre nitelikler belirlendi. Ardından niteliğe göre gelmesi gereken değişken isimleri teker teker ayarlandı.

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

Açıklama otomatik olarak oluşturuldu

Şekil 27. Tablo çıktısı.

Ardından tasarımı daha da güzelleştirmek için fikir araştırması gerçekleştirildi. Bu araştırmalar ile beraber arka plan rengini beyazdan farklı bir renk yapmaya karar verdim. Aynı zamanda abone sayısı bilgisini alan form yapısını sayfanın ortasına yerleştireceğim. Renkler olarak ise üniversitemizin logosunda bulunan renkleri tercih edebilirim.

1. Hafta:

Bir önceki hafta hocamla yaptığımız görüşmede kullanıcının sadece abone sayısından bir tahmin gerçekleştirmenin yeterli sonuçlar vermeyeceğine karar verdik. Bu karar doğrultusunda kullanıcıdan bir ilgi alanı ile ilgili bilgi de almam gerektiğine karar verdim. Sonuç olarak ise yine kullanıcıdan alınan abone sayısına bağlı olarak makine öğrenmesi gerçekleştirilecek, çıktı olarak ise hem tahmin edilen izlenme sayısında izlenilen videoları hemde ilgi alanı ile ilgili, tahmin edilen izlenmeye yakın izlenmeleri çıktı olarak verecek.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 28. İlgi alanı girdisi.

Kullanıcıdan ilgi alanı ile ilgili bilgileri almak için bir girdi daha ekledim. Pythonda kullanabilmek için ise “alan” ile isimlendirdim.



Şekil 29. İlgi alanı alma, HTML kodu.

Ardından burada alınan bilgi pythonda bir değişkende tutuldu.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 30. Abone sayısı ve ilgi alanı verilerinin pythonda tutulması.

1. Hafta:

Bu hafta çıktı olarak var olan tablonun ardından başka bir tablo daha yapıp ilgi alanı ile ilgili çıktıları bu tabloda sundum. Bu ilgi alanı içeriğine girilen girdiyi, abone sayısı gibi bir değişkende tuttum. Ardından yakın izlenmelere sahip olan videoların içerisinde etiketler kısmında bu girilen ilgi alanını arattım.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 31. Tahmin edilen izlenmeye yakın izlenmedeki videoların bulunması.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 32. İlgi alanı ile ilgili eşleşen verilerin ilgili listesinde tutulması.

Çıkan sonuçları ise web sayfasına render ettim. Aynı zamanda abone sayısını aldığım veri seti popüler olan kanalların abone sayılarını verdiği için sonuçlar bazen gerçekçi değerler sunmuyordu. Bu yüzden veri seti araştırmasını da devam ettiriyorum.

1. Hafta:

Bu hafta veri seti araştırmasına devam ettim ancak istediğim veri setini hala bulamadım. Bu yüzden kendi veri setimi oluşturmaya karar verdim. Ardından aradığım veri setini buldum [16]. Bu veri setinde 4547 veri bulunuyor. Sütun olarak ise video\_id, last\_trending\_date, publish\_date, publish\_hour, categoy\_id, channel\_title, views, likes, dislikes, comment\_count, comment\_disabled, tag\_appered\_in\_title\_count, tag\_appered\_in\_title, title, tags, description, trend\_day\_count, trend\_publish\_diff, trend\_tags\_highest, trend\_tag\_total, tags\_count ve subscriber niteliklerini bulunduruyor. Bu niteliklerden benim en çok kullanacağım nitelikler abone sayısı ve izlenme nitelikleridir. Bu nitelikler üzerinden tahmin makinesini gerçekleştiriyorum. Bunu yapmaya çalıştığım zaman bir hata ile karşılaştım. Bu hata verilerimin düzenlenmesini gerektiğini söylüyordu. Bu yüzden öğrenme gerçkeleştireceğim verileri numpy dizilerine çevirip (-1,1) arasında yeniden şekillendirme gerçekleştirip hatayı düzeltiyorum. Bunun ardından çıktı olarak neler vermem gerektiğine karar veriyorum.

1. Hafta:

Bu hafta danışmanım ile yaptığım görüşmede ilgi alanları ile anlatmak istediği seçeneği yanlış anladığım ortaya çıktı. Kendisi bana, aynı abone sayısına sahip farklı içerikler üreten kanalların aynı izlenmere sahip olamayacağından bahsetmiş ancak ben kendisini yanlış anlamışım. Bunun üzerine veri setim üzerinde videoların tags(etiketler) niteliği üzerinde filtreleme yapmaya karar verdim. Bu filtreleme işlemini yaparken öncelikle abone sayısının ortalamasını aldım. Bu ortalama ile abone sayısındaki eksik verileri tamamladım.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 34. Değişkenlerin ayarlanması. Eksik verinin tamamlanması.

Şekil 34’te görüldüğü gibi mean() fonksiyonu ile ortalama alındı. Ardından fillna() fonksiyonu ile abone sayısında bulunan eksik veriler ortalama ile dolduruldu. Ardından videoların tags(etiketler) niteliği bir değişkende tutuldu. Şekil 34’te bunu görebiliriz. Sonrasında ilgi alanları ile ilgili veri setleri oluşturuldu. Bu işlemi gerçekleştirmek için yedi farklı kategori oluşturuldu. Bu kategoriler müzik, yemek, eğlence, siyasi, eğitim, spor ve oyun kategorileridir. Kategori sayısını daha da arttırabiliriz. Ancak kategori sayısını az seçmemdeki sebep makine öğrenmesi sonuçlarının daha gerçekçi olmasını sağlayabilmektir.

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

Açıklama otomatik olarak oluşturuldu

Şekil 35.

Şekil 35’te veri setinin oluşturulmasının oyun içeriği için olan örneğini görüyoruz. Ardından ana veri setindeki veri sayısı kadar bir döngü oluşturuldu. Bu döngü başlangıcında tags(etiketler) niteliğinde eksik veri olması halinde, eksik veri o videonun açıklaması ile tamamlandı.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 36. Eksik etiket verisinin tamamlanması.

Ardından filtreleme işlemleri gerçkeleştirildi. Bu işlemde belirli anahtar kelimeler ile tags(etiketler) niteliği içerisinde arama yapıldı. Bu arama işlemi her kategori için ayrı ayrı yapıldı. Eğer eşleşen sonuçlar bulunursa, eşleşen verilerin tamamını daha önceden ayarlanmış olan ilgili veri setine dahil edildi. Burada verilerin ayrılması için iloc komutu, birleştirilmesi için ise concat komutu kullanıldı.

metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 37. Filtreleme işlemleri.

Bu işlemlerin ardından veri setleri csv dosyalarına kaydedildi.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 38. Dosyalara verilerin yazılması.

1. Hafta:

Bu hafta girdi sayfasında değişiklik yapıldı. Kullanıcıdan ilgi alanı ile ilgili veriler alınırken HTML select etiketi ile bir seçenek kutusu oluşturuldu. Bu kutuya bir önceki hafta belirlenen kategoriler eklendi.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 39. HTML kodu

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 40. Web sayfasındaki görüntü.

Şekil 39’da görüldüğü üzere select etiketinde seçilen seçenek ‘Education’ olurken geri dönülen değer ‘egitim’ değişkenidir. Bu değişkeni ‘value’ özelliği ile ayarlıyoruz. Buradan dönen değere göre kontrol sağlandı. Bu kontrol ile kullanıcının video içeriği üretmek istediği kategorinin makine öğrenmesi gerçekleştirildi. Makine öğrenmesi algoritmasında bir değişiklik yapılmadı.

metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 41. Kategori kontrolü ve doğrusal regresyon.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 42. Karar ağacı yapısı ve hangi algoritmanın uygulanacağının kontrolü.

Ancak verilerin render edilmesinde değişiklikler yapıldı. Çıktı sayfasında bulunan tablo sayısı bir adet yapıldı. Aynı zamanda bu tablonun çıktı sayısı üç ile sınırlandırıldı.

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

Açıklama otomatik olarak oluşturuldu

Şekil 43. Tablo çıktısı.

Makine öğrenmesine bağlı olarak tahmin edilen izlenme sayısındaki içeriklerin sayısına bağlı olarak, render edilen verilerin sayısı sabit kalmakta ancak verilerin içeriği değiştirilmektedir.

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

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, yazı tipi, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 44. Şekil 45.

Şekil 44’te tahmin edilen izlenmeye yakın olarak bulunan içerik sayısı bir adet bulunmuş ise render edilen veriyi, şekil 45’te ise tahmin edilen izlenmeye yakın olarak bulunan içerik sayısı sıfır adet bulunmuş ise render edilen veriyi göstermektedir. Eşleşen veri sayısı sıfır ise çıktı olarak eşleşen veri bulunamadı olarak çıktı sağlıyoruz. Ancak tahmin edilen izlenme sayısı yine çıktı olarak sağlanıyor. Bu durumu şekil 45’te görebiliriz. Ardından sistemdeki hata olasılıkları giderildi. Burada kullanıcının bir veri girmeden tahmin gerçekleştirmeye çalışma hatası engelleniyor.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 46. Try Catch yapısı.

Ardından kullanıcı kategori seçiminde varsayılan olarak eğitim kategorisi seçiliyor. Böylece kullanıcı kategori seçimi yapma durumunda bırakılıyor.

Çıktı olarak, videoların kaç gün trend oldukları da eklendi. Gerekli değişkenler ayarladıktan sonra çıktımızı sağlıyoruz. Tabloya kaç adet eşleşen veri geleceğine göre bu çıktının sayısıda ayarlandı. Sonrasında tasarımsal düzenlemeler yapıldı. Bu düzenlemeler çerçevesinde bir icon görseli oluşturuldu.

mavi, meneviş mavisi, açık mavi, Majorelle Blue içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 47. Tahmak makine öğrenmesi logosu.

Logo iconunu sayfanın üst kısmına yerleştirdikten sonra, alt kısmına girdilerimiz yerleştirildi. Sayfanın ve logo renkleri belirlenirken yine üniversitemizin logosunda kullanılan renklerden esinlenildi.

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

Açıklama otomatik olarak oluşturuldu

Şekil 48. Girdi sayfasının görüntüsü.

Sonrasında çıktı sayfasında görsel güncellemeler gerçekleştirildi. Bu güncellemeler sadece renk değişiklikleri ve tablonun büyüklüğünün değişiklikleridir.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 49. Çıktı sayfasının güncel görünümü.

Çıktılarımızın daha doğru sonuçlar elde etmesini sağlamak için daha fazla veri seti kullanılabilir ve daha güncel veri setleri oluşturulup kullanılabilir. Daha fazla veri içeren veri setleri vardı. Ancak kanalın abone sayısını bir nitelik olarak içeren veri setleri hem az hem de güncel değildi. İncelediğiniz için teşekkürler.

**KAYNAKÇA:**

1-https://wmaraci.com/nedir/dart

2-https://www.webtekno.com/flutter-nedir-nasil-kullanilir-h115673.html

3-<https://numpy.org/doc/stable/user/whatisnumpy.html>

4-<https://kerteriz.net/python-numpy-kullanimi-nedir-ve-nasil-kullanilir/>

5-https://tr.wikipedia.org/wiki/Pandas#:~:text=pandas%2C%20veri%20i%C5%9Flemesi%20ve%20analizi,veri%20yap%C4%B1s%C4%B1%20%C3%BCzerinde%20ger%C3%A7ekle%C5%9Ftirilebilir%20olur.

6-<https://medium.com/datarunner/matplotlibkutuphanesi-1-99087692102b>

7-https://www.vargonen.com/blog/html-nedir-hmtl-kodlama/

8-https://www.vargonen.com/blog/css-nedir-css-kodlari-nelerdir/

9-https://www.innova.com.tr/tr/blog/dijital-donusum-blog/bootstrap-nedir-ne-ise-yarar

10- <https://numpy.org/doc/stable/user/whatisnumpy.html>

11- <https://kerteriz.net/python-numpy-kullanimi-nedir-ve-nasil-kullanilir/>

12-https://tr.wikipedia.org/wiki/Pandas#:~:text=pandas%2C%20veri%20i%C5%9Flemesi%20ve%20analizi,veri%20yap%C4%B1s%C4%B1%20%C3%BCzerinde%20ger%C3%A7ekle%C5%9Ftirilebilir%20olur.

13- <https://medium.com/datarunner/matplotlibkutuphanesi-1-99087692102b>

14-https://halilozel1903.medium.com/django-nedir-52b29ebb7298

15-https://www.kaggle.com/datasets/surajjha101/top-youtube-channels-data?resource=download

16-https://www.kaggle.com/datasets/sgonkaggle/youtube-trend-with-subscriber?resource=download