

**İSTANBUL GELİŞİM MESLEK YÜKSEKOKULU BİLGİSAYAR TEKNOLOJİLERİ BÖLÜMÜ**

**BİLGİSAYAR PROGRAMCILIĞI PROGRAMI**

**ÖDEVİN ADI**

**Gayri Safi Yurt İçi Hasılatı Hesaplama**

**FİNAL PROJE ÖDEVİ**

**Hazırlayan**

**210111307 – Furkan GÜR**

**210111335-Muhammet Can DEMİR**

**210111261-Uğur KAYKAF**

**210111381-Burak YEŞİLYURT**

**200111123-Serhat KILIÇ**

**ÖdevDanışmanı**

**Öğr.Gör.Tuğba Saray ÇETİNKAYA**

**İSTANBUL – 2023**

ÖDEV TANITIM FORMU

YAZAR ADI SOYADI :Uğur Kaykaf, Serhat Kılıç, Muhammet Can Demir, Burak Yeşilyurt, Furkan Gür

ÖDEVİN DİLİ : Türkçe

ÖDEVİN ADI : Gayri Safi Yurt İçi Hasılatı Hesaplama

BÖLÜM : Bilgisayar Teknolojileri

PROGRAM : Bilgisayar Programcılığı

ÖDEVİN TÜRÜ : Final Projesi

ÖDEVİN TES. TARİHİ : 02/06/2023

SAYFA SAYISI :

ÖDEV DANIŞMANI : Öğr. Gör.Tuğba Saray ÇETİNKAYA

# BEYAN

Bu ödevin/projenin hazırlanmasında bilimsel ahlak kurallarına uyulduğu, başkalarının ederlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğu, kullanılan verilerde herhangi tahrifat yapılmadığını, ödevin/projenin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir ödev/proje olarak sunulmadığını beyan eder, aksi durumda karşılaşacağım cezai ve/veya hukuki durumu kabul eder; ayrıca üniversitenin ilgili yasa, yönerge ve metinlerini okuduğumu beyan ederim.

02/06/2023

metin, beyaz tahta içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturulduMuhammed Can Demir Furkan Gür Burak Yeşilyurt

metin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, beyaz tahta içeren bir resim

Açıklama otomatik olarak oluşturuldu Serhat Kılıç Uğur Kaykaf

# KABUL VE ONAY SAYFASI

210111335,210111307,210111261, 210111381,200111123 numaralı Muhammed Can Demir, Furkan Gür, Burak Yeşilyurt, Uğur Kaykaf ve Serhat Kılıç adındaki çalışması, benim tarafımdan Final ödevi olarak kabul edilmiştir.

Öğretim Görevlisi

Tuğba Saray ÇETİNKAYA

# ÖZET

Projemiz , 179 farklı ülkenin veri setlerini kullanarak gayri safi yurt içi hasılat (GSYİH) hesaplama üzerine odaklanmaktadır. Bu veri setlerini yapay zekaya öğretmek suretiyle istenilen ülkenin ve yılın GSYİH'sini hesaplayabilen bir model geliştirmek amacıyla çalışmaktadır.

**İÇİNDEKİLER**

[BEYAN 3](#_Toc136631489)

[KABUL VE ONAY SAYFASI 4](#_Toc136631490)

[ÖZET 5](#_Toc136631491)

[ÖN SÖZ 7](#_Toc136631492)

[GİRİŞ 8](#_Toc136631493)

[Uygulama Çıktıları 10](#_Toc136631494)

[Kaynakça 12](#_Toc136631499)

# ÖN SÖZ

Yaptığımız çalışmanın büyük bir kısmı “Pyton” dilinde yazılmıştır. Çalışmayı yaparken bizi en zorlayan olaylardan birisi uygun ögrenme türünü ödevimize uygulamaktı. Yapmış olduğumuz anlatımda gayet basit ve açıklayıcı anlatımlara yer verilmiştir. Bu projede katkılarından dolayı Öğretim Görevlisi Tuğba Saray ÇETİNKAYA hocamıza şükranlarımı sunmaktayız.

Uğur Kaykaf

Serhat Kılıç

Burak Yeşilyurt

Muhammet Can Demir

Furkan Gür

# 

# GİRİŞ

* Konumuz istenilen ülkenin istenilen tarihte, ülkenin gayri safi yurt içi hasılatını hesaplar.
* Uygulamamıza uygun veri setleri için birden fazla ülkenin verisi lazımdır. Bu yüzden Kaggle, İbb veri portalı, Data.Gov sitelerini incelendi. Sitelerde en uygun olarak 1999-2022 yılları arasında 179 ülkenin verilerinin olduğu veri seti seçilmiştir.
* Veri setimiz kendine has bir veri setidir.Diğer veri setlerinde ülkelerin kendine ait verileri vardır. Tüm ülkelerin olduğu veri setleri vardır ancak verileri düzensizdir ve yılları uzun olduğu için eksik veriler yoğundur.
* Veri setimizde 1999 ile 2022 yılları arasında 179 ülkenini verileri vardır. Öznitelik olarak yıl ve veriler vardır. Pandas kütüphanesinde kullanılarak bir read\_csv fonksiyonu kullanılarak veri seti değişkene atandı sonra değişkende head fonksiyonu kullanılarak ekrana yazdırıldı.
* **Amaç:** Tahmin edilecek ülkenin gayri safi milli hasılasını belirli bir yıl için tahmin etmek.
* **Seçilen Algoritma:** Doğrusal Regresyon (Linear Regression)
* Doğrusal regresyon, bağımlı bir değişkenin bağımsız değişkenlerle olan ilişkisini modellemek için kullanılan bir istatistiksel yaklaşımdır. Bu algoritma, bağımlı değişken ile bağımsız değişkenler arasındaki doğrusal ilişkiyi kullanarak, veri kümesine en uygun doğruyu bulmaya çalışır.
* Belirlenen amaca yönelik olarak kullanılacak uygun makine öğrenmesi modeli doğrusal regresyon (Linear Regression) algoritmasıdır. Aşamaları ve parametreleri aşağıda açıklanmıştır:
  + Hazırlığı:
    - Veri seti, CSV formatında "veri-seti.csv" dosyasından okunur.
    - İstenen ülkenin ismi kullanıcıdan alınır ve veri setinde kontrol edilir.
    - Kullanıcının istediği tahmin yılı alınır ve hatalı girişler kontrol edilir.
    - Seçilen ülkeye ait veri seti ve yılların değerleri seçilir.
  + Eksik Veri Düzenleme:
    - Veri setindeki sıfır değerler eksik veriler olarak kabul edilir.
    - Eksik veriler, yakınsak değerlerle doldurulur. Önceki ve sonraki değerlerin ortalaması kullanılarak doldurma işlemi gerçekleştirilir.
  + Veri Ön İşleme:
    - Standartlaştırma işlemi için StandardScaler kullanılır.
    - Bağımsız değişken (X) olarak yıllar, bağımlı değişken (y) olarak gayri safi milli hasıla değerleri seçilir.
    - Veri seti, eğitim ve test veri setlerine ayrılır. Bu ayrım, train\_test\_split fonksiyonu kullanılarak yapılır.
    - Eğitim veri seti (X\_train ve y\_train) üzerinde fit\_transform işlemi uygulanır.
    - Test veri seti (X\_test) üzerinde ise transform işlemi uygulanır.
  + Model Oluşturma:
    - LinearRegression sınıfı kullanılarak doğrusal regresyon modeli oluşturulur.
    - Model, eğitim veri seti üzerinde fit işlemi ile eğitilir.
  + Tahmin ve Değerlendirme:
    - Test veri seti üzerinde tahminler yapılır.
    - Tahminlerin performansı, gerçek değerlerle karşılaştırılarak değerlendirilir.
    - Mean Squared Error (MSE) ve R^2 score metrikleri kullanılarak değerlendirme yapılır.
    - Tahmin edilen değerler, ters dönüşüm uygulanarak orijinal ölçekte ifade edilir.
  + Görselleştirme:
    - Veri setindeki yıllara göre gayri safi milli hasıla grafiği çizdirilir.
    - X ekseni yılları, Y ekseni ise artış değerlerini temsil eder.
  + Parametreler:
    - test\_size: train\_test\_split fonksiyonunda kullanılan bu parametre, veri setinin hangi oranının test veri seti olarak ayrılacağını belirler. Örneğin, test\_size=0.2 kullanılırsa, veri setinin %20'si test veri seti olarak ayrılır.
    - random\_state: train\_test\_split fonksiyonunda kullanılan bu parametre, veri setinin rastgele ayrılmasında kullanılan bir seed değeridir. Aynı random\_state değeri kullanıldığında, her çalıştırmada aynı veri seti bölünmesi elde edilir.
    - fit\_intercept: LinearRegression sınıfının bir parametresidir ve doğrusal regresyon modelinin bir kesme terimi (intercept) hesaplaması gerekip gerekmediğini belirler. Varsayılan olarak True olarak ayarlanır.
    - normalize: LinearRegression sınıfının bir parametresidir ve bağımsız değişkenlerin normalize edilip edilmeyeceğini belirler. Varsayılan olarak False olarak ayarlanır. Burada, veri ön işleme aşamasında StandardScaler kullanıldığı için normalize=False olarak kullanılmıştır.

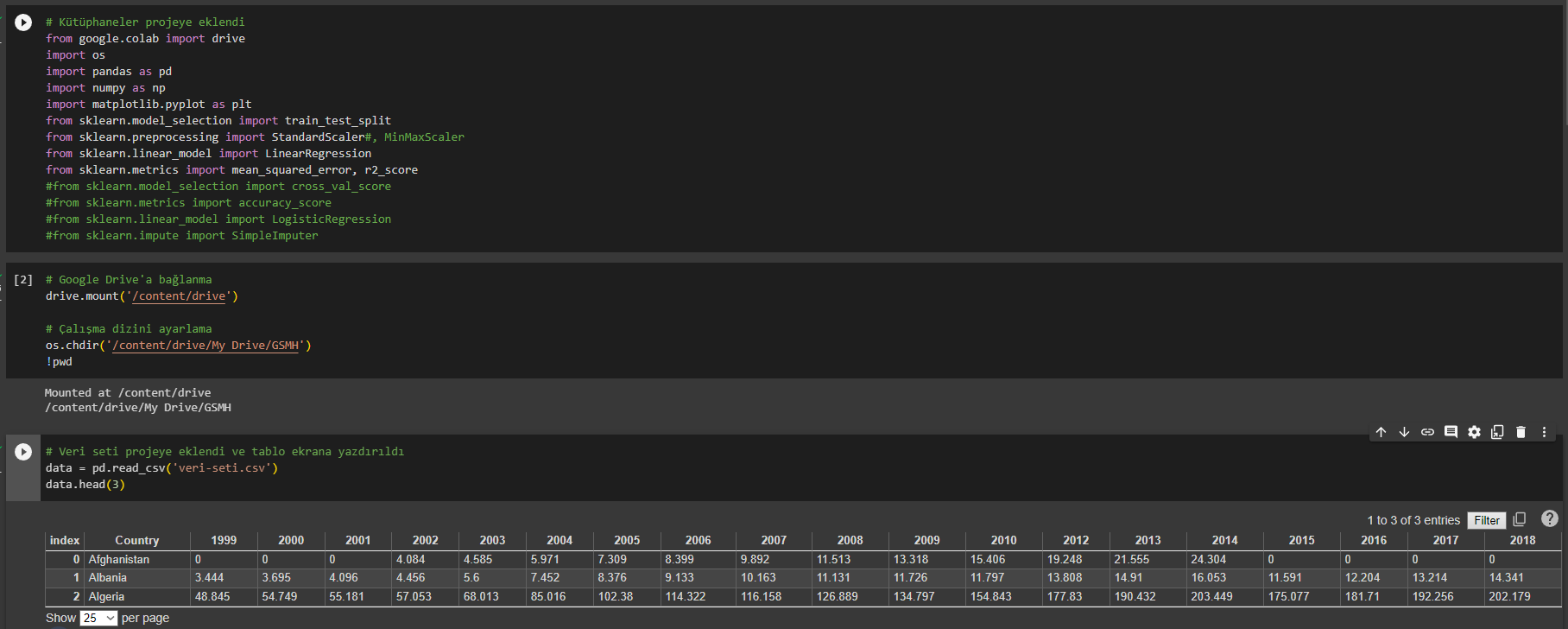
# Uygulama Çıktıları

# 

# 

# 

# 



# Kaynakça

# <https://www.kaggle.com/>

# <https://tr.wikipedia.org/wiki/Nominal_GSY%C4%B0H_de%C4%9Ferlerine_g%C3%B6re_%C3%BClkeler_listesi>

# <https://stackoverflow.com/>

# <https://medium.com/>

# Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. IEEE Intelligent Systems, 24(2),

# Varian, H. R. (2014). Big data: New tricks for econometrics. Journal of Economic Perspectives, 28(2),