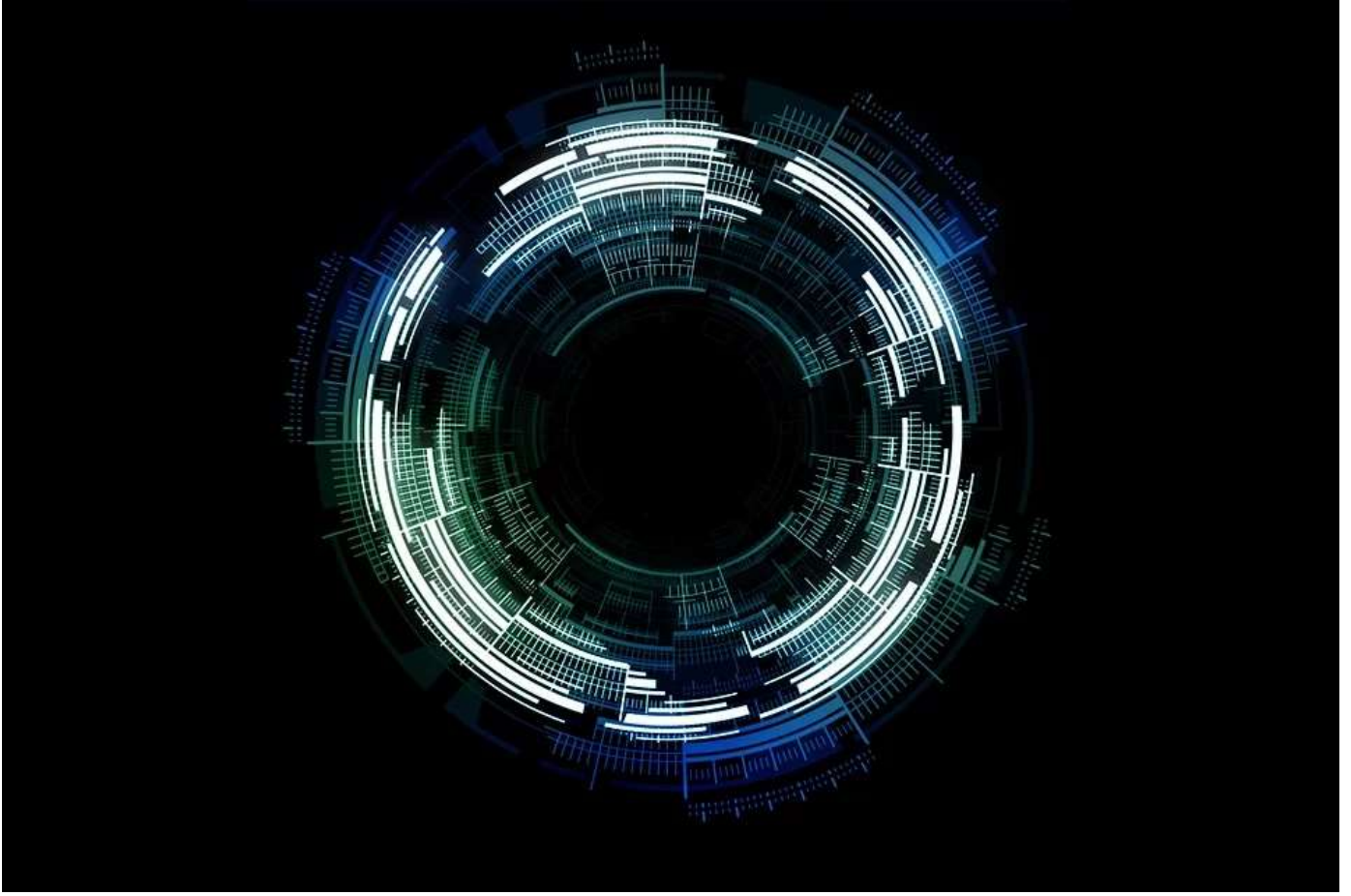


[🏠](#) ▶ [Makine Öğrenmesi](#) ▶ [Regresyon](#) ▶ [Doğrusal Regresyon](#) ▶ [Basit Doğrusal Regresyon Nedir?](#)

Basit Doğrusal Regresyon Nedir?

👤 Ecem Esra Öztürk 📅 12 Mayıs 2020 📖 Doğrusal Regresyon, Genel bir bakış, Makine Öğrenmesi, Regresyon, Yeni Başlayanlar 💬 2 👁 55720

🕒 2302 kelime - 15 dakika ❤️ 116

Bu yazıya puan ver ▶▶

★★★★★ (5 votes, average: 4,00 out of 5)

Bu yazımda Python'da basit doğrusal regresyon modeli oluşturacağız.

Regresyon analizinde, iki ya da daha çok değişkenin yer aldığı istatistiksel modellerde, genellikle neden-sonuç ilişkileri araştırılır. Yani değişkenlerden biri ya da birkaçının, diğer bir ya da birkaç değişkeni ne ölçüde etkilediği incelenir. Eğer değişkenler arasında ilişki varsa, ilişkinin derecesi matematiksel bir fonksiyon olarak ortaya konur. Bu fonksiyona regresyon fonksiyonu denir.



Privacy - Terms

Regresyon analizi; bağımsız değişkenler (X_1, X_2, \dots, X_n) ile bağımlı değişken (Y)'deki değişimi açıklamayı hedefler. Örneğin; bir öğrencinin başarısı ve çalışma saati arasındaki ilişki araştırıldığında; bağımlı değişken Y olarak tanımlanır ve çalışma saati bağımsız değişkeni X olarak tanımlanır.

Regresyon modeli iki ya da daha fazla değişken arasındaki ilişkinin fonksiyonel şeklini göstermekle kalmaz, değişkenlerden birinin değeri bilindiğinde diğeri hakkında tahmin yapılmasını da sağlar.

İki ya da daha çok değişken arasındaki ilişkinin matematiksel bağıntısı "Regresyon Analizi" ile ilişkinin yönü ve derecesi ise "Korelasyon Analizi" ile incelenir.

Regresyon analizi bir tahmin (öngörüs) analizi olup, bağımlı değişkenin bağımsız değişkenler yardımıyla tahmin edilmesini sağlar. Ayrıca bağımlı değişkeni etkileyen en önemli bağımsız değişken/değişkenlerin hangisini olduğunu ortaya çıkarır.

Regresyon analizi;

Basit doğrusal regresyon,

Çoklu değişkenli regresyon analizi,

Doğrusal olmayan regresyon analizi olarak sınıflandırılabilir.

Aralarında ilişki araştırılması istenen değişkenler sayılabilir veya ölçülebilir nitelikte olabilir. Üzerinde durulan değişkenlerden bağımlı değişken y, bağımsız değişken x ise, $y=f(x)$ şeklindeki fonksiyona regresyon modeli denir. $f(x)$ fonksiyonu farklı şekiller alabilir.

Bu bölümde; $f(x)$ fonksiyonu doğrusal olan bir bağımlı(y) ve bir bağımsız(x) değişkeninin yer aldığı "basit doğrusal regresyon" analizinden bahsedeceğim.

Basit doğrusal regresyon; bağımsız değişken (X) ile bağımlı değişken (Y)deki değişimi açıklamayı, bağımsız değişkendeki bir birimlik değişimin bağımlı değişken üzerindeki etkisini ölçmeyi amaçlar.

Temel amaç, bağımlı ve bağımsız değişken arasındaki ilişkiyi ifade eden doğrusal fonksiyonu bulmaktır.



Stokastik (Olasılıklı) bir model olan ve ana kütledeki ilişkiyi gösteren basit doğrusal regresyon denklemi aşağıdaki gibi ifade edilir.

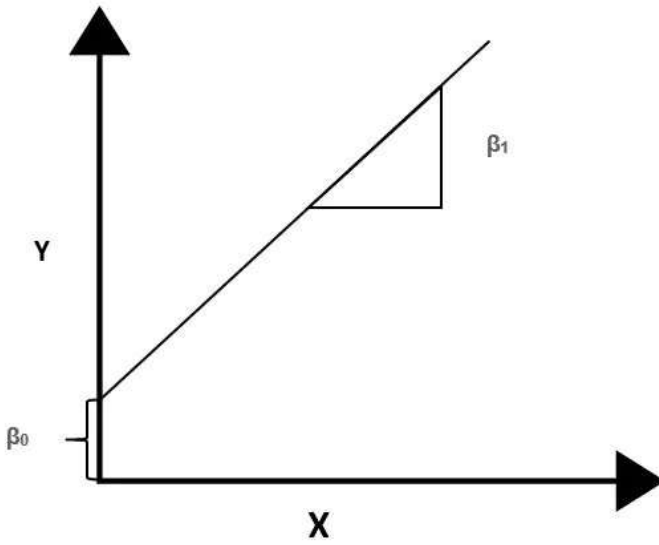
$$y = \beta_0 + \beta_1 X + \epsilon$$

Burada;

β_0 : Doğrunun y-eksenini kestiği yer ve regresyon sabitidir.

β_1 : Doğrunun eğimi veya regresyon katsayısıdır.

ϵ : Rastgele(Tesadüfi-Şans) hata değeridir.



Şekil-1: Regresyon Grafiği

R²: Regresyon Modelinin Performansı

Bağımsız değişken x'in, regresyon modeli ile bağımlı değişken y'i ne kadar açıkladığı yüzdesinin ölçütü olan belirtme katsayısı bir diğer adıyla determinasyon katsayısıdır ve R² ile gösterilir. Belirtme katsayısı, bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişkenler tarafından açıklanabildiğini gösterir.

R², "0" ile "1" arasında değerler alır(0<R²<1). Değişkenler arasında doğrusal bir ilişki olduğunda, R² değerinin 1'e yaklaşması; bağımlı değişkendeki değişimin büyük bir kısmının bağımsız değişkenler tarafından açıklandığını gösterir. Bağımlı ve bağımsız değişken arasındaki ilişkinin derecesi ve yönünü gösteren korelasyon katsayısının(r) karesi belirtme katsayısına (R²) eşittir. Belirtme katsayısı:

R²=Açıklanabilen Değişim/Toplam Değişim=RKT/YOAKT eşitliğiyle hesaplanır

Eşitlikte verilen RKT; regresyon kareler toplamı ve YOAKT; Y ortalama ayrış kareler toplamıdır.



Kurulan regresyon modelinin performansı, R^2 ile ölçülür. R^2 , 1'e ne kadar yakınsa, regresyon o kadar anlamlıdır ve belirleyicidir. Bu konuda bilimsel bir karar verebilmek için hipotez testi yapılır.

Düzeltilmiş R Kare

R-kare, lineer regresyon modeli için bağımsız değişkenlerimiz (X) tarafından açıklanan bağımlı değişkenimizdeki (Y) varyasyon oranını ölçer. Düzeltilmiş R-kare sadece gerçekte bağımlı değişkeni etkileyen bağımsız değişkenler tarafından açıklanan varyasyon oranını ölçer.

$$\text{Düzeltilmiş R kare} = 1 - (1 - R^2) * n - 1 / n - p - 1$$

n: Örneklemdeki veri sayısı

p: bağımsız değişken sayısı

Korelasyon Katsayısı(r)

Korelasyon katsayısı, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin gücünü gösteren bir katsayıdır. Örneğin; öğrencinin ders çalışma süresi ile aldığı istatistik notu arasında ilişki olup olmadığını veya borsada işlem gören bir hisse senedinin belli bir dönemdeki günlük getirisi (X) ile içinde yer aldığı bir endeksin günlük getirisi (Y) arasındaki ilişki korelasyon katsayısı ile incelenebilir. Korelasyon katsayısı değişkenlerin yönü ve etkileşimlerin nasıl olduğu hakkında bilgi verir.

Korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin ölçüsü olup incelenen değişkenlerin birimlerinden bağımsızdır ve $-1 \leq r \leq 1$ arasındadır. Yani 5 ya da -5 olamaz. Korelasyon katsayısının 0'a yaklaşması değişkenler arasında zayıf ilişkinin varlığını gösterir. Değişkenler birlikte artıyor veya azalıyorsa pozitif yönde, değişkenlerden biri artarken diğeri azalıyorsa ise negatif yönde bir ilişki vardır.

Korelasyon Katsayısı ve R Kare Farkları

1. Korelasyon katsayısı, iki değişken arasındaki doğrusal ilişkiyi temsil etmek için kullanılır. Regresyon katsayısı ise bir değişkeni başka bir değişken temelinde tahmin etmek için kullanılır.
2. Korelasyon değişkenler arasındaki ilişkinin gücünü gösterir. Regresyon katsayısı ise bağımsız değişkendeki birim değişikliğinin bağımlı değişken üzerindeki etkisini gösterir.



Regresyon Katsayısının Hipotez Testi

Kurulan regresyon modelinde verilerin (gözlemlerin) regresyon doğrusuna uyumunun araştırılması gereklidir. Regresyon doğrusunun anlamlılığını test etmek için β_1 regresyon katsayısının analizi yapılmalıdır. Çünkü β_0 , regresyon sabiti "0" olsa bile, regresyon modeli kurulabilir ($Y = \beta_0 + \beta_1 X$). Dolayısıyla β_1 katsayısının "0"dan farklı olması bir model kurmak için önemlidir β_1 katsayısının hipotez testi (önem testi) aşağıdaki gibi kurulur.

Bir regresyonun anlamlı olması aslında doğrunun eğimi olan β_1 'in sıfırdan farklı olması ile eş değerdir.

Kullanılacak Olan Bütün Formüller

X Ortalamadan Ayrılış Kareler Toplamı (XOAKT)

$$XOAKT = \sum X^2 - (\sum X)^2 / n$$

Y Ortalamadan Ayrılış Kareler Toplamı (YOAKT)

$$YOAKT = \sum y^2 - (\sum y)^2 / n$$

XY Ortalamadan Ayrılış Kareler Toplamı (XYOAÇT)

$$XYOAÇT = \sum xy - (\sum y * \sum x) / n$$

XY Ortalamadan Ayrılış Çarpımlar Toplamı (XYÇT)

$$XYÇT = \sum x * y - n * \bar{x} * \bar{y}$$

Regresyon Kareler Toplamı (RKT)

$$RKT = b_1 * XYOAÇT = B_1 * XYÇT$$

Regresyon Ayrılış Kareler Toplamı (RAKT)

$$RAKT = YOAKT - RKT$$

Regresyon Kareler Ortalaması (RKO)

$RKO = RKT / R_{SD} = RKT / 1$ formülüyle bulunur. Basit doğrusal regresyon analizinde regresyon serbestlik derecesi (R_{SD}) her zaman 1'dir.

Regresyon Ayrılış Kareler Ortalaması (RAKO)

$$RAKO = RAKT / R_{ASD} = RAKT / (n - 2) = s^2 \text{ formülüyle bulunabilir.}$$



Burada **1.adım**: H_0 (null) ve H_s (alternatif) hipotezler yazılır.

$H_0: \beta_1 = 0$ (Regresyon katsayısı önemsizdir, regresyon doğrusu önemsizdir.)

$H_s: \beta_1 \neq 0$ (Regresyon katsayısı önemlidir, regresyon doğrusu önemlidir.)

2.adım: Bir regresyonun anlamlı olup olmadığını anlamak için anlamlılık düzeyinde $t_T = t_{(\alpha/2; n-2)}$ ve $t_H = b_1 / S_{b_1}$ hesaplanır. t_T , t-tablosu yardımıyla bulunur. t_H , modelin test istatistiği hesap değeridir ve formül ile bulunur.

b_1 katsayısı = $XYO\Delta CT / XOAKT$

b_1 katsayısının hatası olan $S_{b_1} =$

$$S_{b_1} = \sqrt{(RAKO / XOAKT)}$$

formülüyle hesaplanır.

3.adım : t_H ve t_T karşılaştırılır.

Eğer $|t_H| > t_T$ ise H_0 hipotezi reddedilir. Yani $b_1 \neq 0$.

Eğer $|t_H| < t_T$ ise H_0 hipotezi kabul edilir.

4.adım: Regresyon modelinin anlamlı olup olmadığıyla ilgili istatistiksel karar verilir. $b_1 \neq 0$ ise kurulan regresyon modeli α anlamlılık düzeyinde anlamlıdır deriz.

Makine Öğrenmesinde Basit Doğrusal Regresyon Uygulaması

Öncelikle bir veri seti seçmemiz gerekiyor. Ben bu örnekte [reklam veri setini](#) kullanacağım.

Kullanacağımız kütüphaneleri içe aktaralım. Veri görselleştirme için gerekli kodları da girelim.

```
1. import pandas as pd
2. import numpy as np
3. import seaborn as sns
4. from matplotlib import pyplot as plt
```

Kütüphaneyi kodumuzda içeri aktardıktan sonra ilk işimiz işleyeceğimiz verilere erişim sağlamak olacak.

- CSV (Virgülle ayrılmış): read_csv



- Veri tabanından: `read_sql`
- JSON biçiminde dosyadan: `read_json`

`df.head()` koduyla verileri çağırdık. `iloc[]` kodu ile bir seçim işlemi yapıyoruz.

```
1. df = pd.read_csv("Desktop/reklam/reklam.csv")
2. df= df.iloc[:,1:len(df)]
3. df.head()
```

Elimizde TV, radyo, gazete için reklam harcamaları ve satış verileri var. Biz bu çalışmada TV ve satış değişkenleriyle ilgileneceğiz.

Amacımız TV-reklam harcamalarını modelleyebilmek ve çalışmalar yapmak.

TV 'yi bağımsız değişken(x), satışı da bağımlı değişken(y) olarak belirledik.

Kodlarımızı çalıştırarak X ve y'yi çağırıyoruz.

```
1. X= df[["TV"]]
2. X.head()
```

```
1. y=df[["satis"]]
2. y.head()
```



Regresyonu modelleyebilmek için bir model nesnesi oluşturalım. Bunun için scikit-learn kütüphanesinden yararlanalım.

```
1. reg =LinearRegression()
```

Modelimizi fit() metoduyla kuralım. Doğrusal bir fit en küçük kareler yöntemiyle bulunur.

```
1. model=reg.fit(X,y)
2. model
```



β_0 katsayısını çağırmak için:

```
1. model.intercept_
```

β_1 katsayısını çağırmak için:



```
1. model.coef_
```

Scikit-learn kütüphanesinde eğitim verisinden elde edilen parametrelerin sonuna alt tire(_) konur. Böylece bulunan parametreler ile kullanıcıdan alınan parametreler birbirinden ayrılır.

R² değeri

Yazımın başında R² den bahsetmiştim kısaca anlatmam gerekirse R² modelin açıklanabilirliği ile ilgili bir değerdir. R² elimizdeki bağımsız değişkenleri kullandığımızda bağımlı değişkendeki değişimin yüzde kaçını açıklayabiliyoruz bilgisini bize sunar. R²yi hesaplayabilmek için:

```
1. model.score(X, y)
```

Regresyon Grafiği

```
1. import seaborn as sns
2. import matplotlib.pyplot as plt
3. g= sns.regplot(df["TV"], df["satis"],ci=None,scatter_kws={'color':'r','s':
4. g.set_title("Reklama Göre Satış Tahmin Regresyon Model")
5. g.set_ylabel("Satış Sayısı")
6. g.set_xlabel("TV Harcamaları")
```



Yukarıdaki kodda regresyon modelimizi görselleştirme işlemini gerçekleştirdik. Tek tek açıklayacak olursak;

1.satır ve 2.satır: Veri görselleştirme için gerekli kodları girdik.

3.satır: TV ve satış değişkeni üzerinden bir regresyon görselleştirmesi yapmak için bu satırdaki kodu girdik. Bu satırdaki `ci=None` güven aralığı koymaması için girilmiş bir koddur.

4.satır: Regresyon modelimize bir başlık ekledik.

5.ve 6. satır: y-ekseni ve x-eksenini adlandırma işlemini yaptık.

Mavi çizgi basit bir fonksiyondur. Başta söylemiş olduğum temel amacımıza ulaşmış olduk. Yani bağımlı ve bağımsız değişken arasındaki ilişkiyi ifade eden doğrusal fonksiyonu bulmuş olduk.

Modelimizi eğitmiş olduk şimdi bu kurmuş olduğumuz model üzerinde tahmin işlemi yapalım.

Tahmin Yapma İşlemi



Eğitmiş olduğumuz model üzerinden tahminler yapmaya çalışalım. TV'ye ne kadar yatırım yaparsam ne kadar kazanç elde ederimin sonucunu bulmak için tahmin etmeye çalışalım.

Sckit-Learn aracılığıyla model nesnemizi kullanarak tahminde bulunalım.

```
1. model.intercept_+model.coef_*165
```

Harcamalarımız 5, 15, 30 olduğunda kazancımızın ne olacağını tahmin etmek için `model.predict()` metodundan yararlanalım.

```
1. model.predict([[165]])
```

```
1. yeni_veri=[[5],[15],[30]]
```

```
1. model.predict(yeni_veri)
```



Hatalar ve Hata Kareler Ortalaması

Amacımız gerçek değerler ile tahmin değerler arasındaki farkı minimum yapmak.

y mizin gerçek değerlerini çağıralım.

```
1. y.head()
```

Kurmuş olduğumuz modeli kullanarak bir tahmin edilen değeri çağıralım.

```
1. model.predict(X)[0:6]
```

```
1. gercek_y=y[0:10]
2. tahmin_edilen_y=pd.DataFrame(model.predict(X)[0:10])
3. hatalar=pd.concat([gercek_y,tahmin_edilen_y],axis=1)
4. hatalar
```



Bu tabloyu isimlendirmek için:

```
1. hatalar.columns=["gercek_y", "tahmin edilen y"]
2. hatalar
```

```
1. hatalar["hata"] = hatalar["gercek_y"] - hatalar["tahmin edilen y"]
2. hatalar
```



Kare alma işlemini pozitif ve negatif değerlerinin birbirini götürmesini engellemek için yaparız. Hata kareler ortalamasını bulmak için:

```
1. hatalar["hata_kareler"]=hatalar["hata"]**2
2. hatalar
```

```
1. np.mean(hatalar["hata_kareler"])
```



Hata kareler ortalaması, veriler üzerinde elde ettiğimiz ortalama hatamızıdır. Yani tahmin edilen değer ve gerçek değer arasındaki farkının karesini alıp ortalamasını aldığımızda hatamızın 9.2 olduğu sonucuna ulaştık.

Başka bir yazıda görüşmek dileğiyle hoşça kalın.

Kaynaklar:

1. <https://gelecegiyazanlar.turkcell.com.tr/konu/makine-ogrenmesi>
2. Prof. Dr. Filiz ERSÖZ & Dr. Taner ERSÖZ, İstatistik, Seçkin Kitabevi

Ecem Esra ÖZTÜRK

Related Posts:

- [MXNet ile Derin Öğrenme 1.2: Doğrusal \(Linear\)...](#)
- [Alternatif Regresyon Yöntemleri: Lasso Regresyon](#)
- [AWS Glue ile Basit Bir ETL Uygulaması](#)
- [MXNet ile Derin Öğrenme 2.1: Softmax Regresyon \(Teori\)](#)
- [MXNet ile Derin Öğrenme 2.2: Softmax Regresyon \(Uygulama\)](#)
- [P-değeri Nedir?](#)

Basit lineer regresyon örneği Doğrusal regresyon makine öğrenmesi Makine öğrenmesi teknikleri Python Python ile Doğrusal Regresyon Regresyon Regresyon modeli kurmak

Paylaş: [f](#) Paylaş [t](#) Tweetle [g](#) Gönder

Yazar Hakkında

Toplam 3 yazı

Ecem Esra Öztürk

Endüstri Mühendisliği, İşletme (ÇAP) ve Yönetim Bilişim Sistemleri öğrencisiyim.





Tüm yazılarını gör ►

Şunlar da ilginizi çekebilir

İlgili içerik

[dbt \(DataBuildTool\) ile Veri Analitiği Yolculuğunda Yeni Bir Dönem](#)



İlgili içerik

Oppenheimer Filmi ve Nükleer Bombaların Gölgesinde Veri Biliminin Önemi



◀ Önceki yazı

Canlı Yayın Etkinlikleri #4 – R ile Veri Analizi ve Veri Görselleştirme Uygulaması



[Sonraki yazı ►](#)

MATLAB SİMULİNK İLE GÖRÜNTÜ İŞLEME – 3

Yorumlar (2 yorum)

Pingback: [Linear Regression \(Doğrusal Regresyon\) | Yusuf AFŞİN](#)

mert ●

10 Aralık 2022 - 18:36

Yanıtla

veri setini indiremedikten sonra neye yarar? link ölmüş.

Bir yanıt yazın

E-posta adresiniz yayınlanmayacak. Gerekli alanlar * ile işaretlenmişlerdir



Yorumunuzu yazın *

Adınız *

E-posta adresiniz *

Website

☐

Daha sonraki yorumlarımda kullanılması için adım, e-posta adresim ve site adresim bu tarayıcıya kaydedilsin.

Yorum gönder

💬 Son Yorumlar

- [melih ▶](#) Microsoft Excel' de VBA Yardımıyla Otomatik Olarak E-posta Gönderimi
- [Hakan NAKIŞ ▶](#) Microsoft Excel' de VBA Yardımıyla Çalışma Sayfalarını Ayrı Ayrı PDF Olarak Kaydetme
- [kenan ▶](#) R ile Makine Öğrenmesi Uygulamaları: Doğrusal Regresyon
- [huseyin ▶](#) Seaborn ile Veri Görselleştirmesi
- [Dilara ▶](#) Windows 10 Üzerine Kafka Kurmak, Mesaj Okumak ve Yazmak

Ecem Esra Öztürk

Endüstri Mühendisliği, İşletme (ÇAP) ve Yönetim Bilişim Sistemleri öğrencisiyim.

💬 Son Yorumlar



melih ▶

Microsoft Excel' de VBA Yardımıyla Otomatik Olarak E-posta Gönderimi

Hakan NAKİŞ ▶

Microsoft Excel' de VBA Yardımıyla Çalışma Sayfalarını Ayrı Ayrı PDF Olarak Kaydetme

kenan ▶

R ile Makine Öğrenmesi Uygulamaları: Doğrusal Regresyon

huseyin ▶

Seaborn ile Veri Görselleştirmesi

Dilara ▶

Windows 10 Üzerine Kafka Kurmak, Mesaj Okumak ve Yazmak

★ Bunları Da Oku

1

3 Adımda CRM (Müşteri İlişkileri Yönetimi)

2

AB Testi Nedir ? İstatistiksel A/B Testleri Nasıl Yapılır?

3

Aşırı Öğrenme (Overfitting)

📄 VBO Ekibine Özel

VBO Dosya Merkezi

VBO Ekip Eğitimi

2022, VBO BLOG | Tüm Hakları Saklıdır. Tema, Veri Bilimi Okulu tarafından düzenlenmiştir.

[🏠 Anasayfa](#) [✍️ Yazar Olmak İstiyorum](#)