Dunynin en degerli kaynagi petrol degil veri

The Economist

Yapay zeka yeni elektrik

Andrew NG

Yapar Zeka:

Bir bilgisaray veya bilgisaray bagli robotun zeki canlilara benzer sekilde yerine getirme faliyeti

Yapay zeka ornek kullanimi:

• Arkadas onerileri

• Otomatik fotograf etiketleme

• Hedefli icerik pazalama

• Otomatik mesaj pazarlama

• Hedegli urun pazarlama

• Tavsiye sistemi

Farkli alanlarda ornek uygulamalar

• Musteri segmentasyonu

• Kanser/Hastalik teshisi

• Sirketlerin gelir tahminleri ile strateji belirlemesi

• Basvuru degerlendirme sistemleri

• Akilli portfoy yonetimi

• Dogal afet modelleme calismalari

• E-spor analitigi

Goruntu Isleme Temelli Uygulamalar:

• Otonom araclar

• Nesne tanima/takip uygulamalari

• Sahte videolar

• Eski resimlerin canlandirilmasi

• Algoritmalarin gelistirdigi resimler/var olmayan kisiler

Veri Bilimi :: Matematik Istatistik / Bilgisaray Bilimleri / Is-Sektor Bilgisi kesisimi

Veriden faydali bilgi cikarma sureci

Betimleyici Analitik (Ne olmus sorusuna yanit arar)

Veriyi betimelemk basit raporlarla resmini cekme Urun satis periyordunu tablolastirdigimizda betimlemeye girer

Teshis Tani Analitigi (Neden olmus nasil olmus sorusuna yanit arar)

Belli aylardaki dususun nedenini sorgulayan nedenini araran tani koyan

Tahminsel Analitik (Ne olacak sorusuna yanit arar)

Nihai sonucu bilinmeyen olaylar ile ilgili olaylar veya gelecen ile olgili tahmin yapar

Yonergeli Analitik (Nasil olmali ne olmali sorularina yanit arar)

Ileriki aylardaki satislarin ne olacagini tahmin ettik ama Ne yaparsak gelirler artar verilen yonergelere gore bagzi is aksiyon kararlari alarak gelirlerimi arttirmaya calisiyoruz

Bu adimlari kullanarak veriden faydali bilgi cikiyor

CRISP-DM

Veri bilimi sureclerinin is dunyasindaki ele alinisi bicimi

Veri Okuryazarligi

Gunluk hayatta veriyle temas ettigimiz ilk anlardaki basit veri yorumlama kabiliyetidir

Veri Okuyyazarligi her turden veri tipini degisken ve olcek turlerini tanimlayabilme betimsel istatistikleri ve istatstiksel grafikleri kullanarak veri degelnedirmebile yetenegidir

Olcek Turleri

Sayisal degiskenler icin Aralik ve Oran

Baslangic noktasi 0 olmayan sayisal olcek turu aralik

Baslangic noktasi 0 olan sayisal degiskenlerin olcek turu orandir

Kategorik degiskenler icin Nominal ve Ordinal

Siniflar arasi fark yok ise nominal

Siniflar arasi fark var ise ordinal (rutbe sistemi)

Merkezi Egilim Olculeri

Aritmetiksel Ortalama

Medyan

Mod

Kartiler

Dagilim Olculeri =Elimizdeki degiskenin deglerinin ne sekilde dagildigi

Degisim Araligi

Standard Sapma

Varyans

Carpiklik

Basiklik

Istatistiksel Dusunme Modellleri

Veri okuryazarligindan veri analitigine giden yolu modelleyen yol gosterilcilerdir

Istatistiksel Dusunme Modeli

MONNEY MODELI

Verinin Tanimlanmasi

Degisken nedir olcek nedir degisken turleri olcek turleri nelerdir ortlamasi standard sapmasi mod medyan gibi degerleri nedir merkezi egilimi nedir veriyi tanimlama ne oldugunu bilme islemleri

Verinin organize edilmesi ve indirgenmesi

Veri onumuze geldigine belli basil on islmerden gecirilmesi organize edilmesi

Veri Gosterimi

Verinin gorsellestirilmesi istatistiksel grafik youmlama ve benzeri ifadeleri temsil ediyor

Verinin Analiz Edilmesi ve Yorumlanmasi

Istatistiksel Dusunce Duzeyleri

Kuisiye Ozgu 1.seviye

Gecici 2.seviye

Nicel 3.seviye

Analitik 4.seviye

Veri Okuryazigi nedir: gunluk yasantimizda veriyle temas ettigimiz ilk anlardaki veriyi youmlayabilme kabiliyeti

Temel kavrmlar: popilasyon, orneklem, gozlem birimi degisken nedir altindaki olcek turleri nelerdir

sayisal degiskenler icin aralik ve oran

katagorik degiskenler icin nomonal ve ordinal olcek

merkezi egilim olculeri: mod medyan ortalama ve kartiller gibi bir degisken geldiginde ortalama ve merkezinin gozlemlenmesine yarar

simeterik ise ortalama verilir simetrik degilse medyan

Dagilim olculeri: standard sapma degisim araligi ve varyans

Bir degisken geldiginde yapisi dagilimi ortalama etrafinda nasi dagilmis

VERI MANIPULASYONU

Veri Manpulasyonun`da en cok kullandigimiz kutuphanelerden biri NumPy bu kutuphane sayesinde

veri olusturabilir (np.zeros, np.ones((3,5)), np.full((3,5),3), np.arange(0,31,3), np.linspace(0,1,10))

olusturlan bu verilere sekillendirilebilir (np.concatenate([x,y]), np.split(x,[3,5]))

bu verilere sagdan soldan alttan usten ayirma islemleri yapilabilir ( np.vsplit(m,[2]), np.arange(16).reshape(4,4), np.hsplit(m,[2]))

olturulan verileri siralayabilir (sort() , np.sort(m,axis=1), np.sort(m,axis=0)

Math kutuphaneinde kullandigimiz methodlarin hepsi bunda da var

NumPy python un klasik anlamda programcilik yetenkelerinin kisitli kaldigi yerlerde ortaya cikan matematiksel istatistiksel bilimsel anlamda hesaplar yapmak icin kullananiln bir kutuphanedir Data Science ile ilgili bircok kutuphane nin timeline NumPy barindirir

NumPy icerinsnde katagoik ve fix degiskenleri tutma kousunda cok basarili degil bundan dolayi ihtiyac duyulan ileri sevie manipulative islmlerde Pandas kullaniliyor

Pandas daha analitik ogrenme makine ogrenmesi alinda NumPy altyapi olarak kullanmasina ragmen daha ilidir

Loc : veirlen kurallara bagli secim yapilma ihtiyaci varsa

Iloc : isimlendirmelerden bagimsiz klasik index mantigiyla secim yapmaya ihtiyac arsa

VERI GORSELLESTIRME

Kesifci Veri Analizi: Veri Bilimcinin ozgurce calisabildigi soz konudu olmayan yrni bulgulara ulasabilecegi yeni is fikirleri yeni ise yarar sonuclara ulasabilecegi sorular sorarak hipotezler kurarak ilerledigi surectir

Python Veri Gorsellesitrme Kutuphaneleri

Matplotlib

Pandas

Seaborn

ggplot

Bokeh

Plot.ly

1. Başlangıç Zamanı Grafiği:

o Eşleşen Method: Lineplot

 Zaman içinde değişen verileri görselleştirdiği için, lineplot bu tür bir grafiği oluşturmak için kullanılır.

2. Çizgi Grafiği (Line Plot):

o Eşleşen Method: Lineplot

 Zaman serisi verilerini veya sürekli değişkenlerin eğilimlerini gösterir.

3. Isı Haritası (Heatmap):

o Eşleşen Method: Heatmap

 Verinin iki boyutlu gösterimi olarak, değişkenler arasındaki ilişkileri renkler ile temsil eder.

4. Korelasyon Grafiği (Correlation Plot):

o Eşleşen Method: Heatmap veya Pairplot

 Değişkenler arasındaki korelasyonu görsel olarak sunar. Heatmap, korelasyon matrislerini renkler ile görselleştirir. Pairplot, sayısal değişkenler arasındaki ilişkileri gösterir.

5. Violin Grafiği (Violin Plot):

o Eşleşen Method: Violinplot

 Verinin dağılımını ve yoğunluğunu gösterir.

6. Kutu Grafiği (Boxplot):

o Eşleşen Method: Boxplot

 Verinin dağılımını, medyanını, çeyreklerini ve uç değerleri gösterir.

7. Histogram:

o Eşleşen Method: Distplot veya Kdeplot

 Distplot, verinin histogramını ve yoğunluğunu gösterir. Kdeplot ise verinin yoğunluk fonksiyonunu görselleştirir.

8. Yoğunluk Grafiği (Density Plot):

o Eşleşen Method: Kdeplot

 Sürekli verilerin yoğunluk fonksiyonunu gösterir.

9. Sütun Grafiği (Bar Plot):

o Eşleşen Method: Barplot

 Kategorik verilerin karşılaştırılmasında kullanılır.

Diğer Yöntemler:

• Scatterplot:

o Eşleşen Grafik: Scatterplot

 İki sayısal değişken arasındaki ilişkiyi gösterir.

• Pairplot:

o Eşleşen Grafik: Pairplot

 Birden çok sayısal değişken arasındaki ilişkileri ve dağılımları gösterir.

• Lmplot:

o Eşleşen Grafik: Lmplot

 İki değişken arasındaki doğrusal ilişkiyi gösterir ve ayrıca regresyon çizgisi ekler.

• Catplot:

o Eşleşen Grafik: Catplot

 Kategorik verilerin farklı görselleştirmelerle (bar, violin, box vs.) analizini yapar.

• FacetGrid:

o Eşleşen Grafik: FacetGrid

 Çoklu grafiklerin aynı anda farklı alt gruplara göre düzenlenmesini sağlar.

Özetle:

• Lineplot: Çizgi grafiği ve başlangıç zamanı grafiği.

• Heatmap: Korelasyon grafiği ve ısı haritası.

• Pairplot: Korelasyon grafiği ve değişken ilişkilerini gösteren çoklu grafik.

• Lmplot: İki değişken arasındaki doğrusal ilişkiyi gösterir.

• Scatterplot: İki değişken arasındaki ilişkiyi noktalarla gösterir.

• Catplot: Kategorik verilerin farklı görselleştirmeleri.

• Boxplot: Kutu grafiği.

• FacetGrid: Çoklu grafik düzenleme.

• Kdeplot: Yoğunluk grafiği.

• Distplot: Histogram ve yoğunluk grafiği.

• Barplot: Sütun grafiği.

Veri Bilimi Icin Istatistik

1. Örnek Teorisi: Popülasyondan alınan örneklemlerle genellemeler yapmayı amaçlar. Örneklem dağılımı, örneklem istatistiklerinin dağılımıdır.

2. Betimsel İstatistik: Veriyi özetler. Ortalama, medyan, mod gibi ölçütlerle merkezi eğilim hesaplanır. Dağılım, histogram ve boxplot gibi grafiklerle gösterilir.

3. Güven Aralığı: Bir popülasyon parametresinin, belirli bir güven düzeyinde aralıkla tahmin edilmesidir. Örneklem ortalaması ± (kritik değer) \* (standart hata) ile hesaplanır.

4. Olasılığa Giriş ve Olasılık Dağılımları:

• Bernoulli Dağılımı: İki sonuçlu deneyler (başarı/başarısızlık).

• Büyük Sayılar Yasası: Örneklem büyüklüğü arttıkça, örneklem ortalaması popülasyon ortalamasına yaklaşır.

• Binom Dağılımı: Birçok bağımsız Bernoulli denemesinde başarı sayısı.

• Poisson Dağılımı: Nadiren gerçekleşen olayların sayısını modeller.

• Normal Dağılım: Çoğu doğa olayını modelleyen simetrik bir dağılımdır (çan eğrisi).

• Hipotez Testleri: İddiaların doğruluğunu test etmek için kullanılır. H0 ve H1 arasında seçim yapılır.

• Tek Örneklem T Testi: Bir örneklem ortalamasını, belirli bir popülasyon ortalamasıyla karşılaştırır.

• Tek Örneklem Oran Testi: Bir popülasyon oranını, belirli bir değere karşı test eder.

• Bağımsız İki Örneklem T Testi: İki bağımsız grubun ortalamaları arasındaki farkı test eder.

• Bağımlı İki Örneklem T Testi: Aynı grubun iki ölçümünü (örneğin, öncesi ve sonrası) karşılaştırır.

• İki Örneklem Oran Testi: İki bağımsız grubun oranlarını karşılaştırır.

• Varyans Analizi (ANOVA): Birden fazla grup arasındaki ortalama farklarını test eder.

• Korelasyon Analizi: İki değişken arasındaki ilişkiyi ölçer.

Betimsel Istatistik

Kovaryans Iki degisken arasindaki ilişkinin değişkenlik olcusudur

Korelasyon iki degisken arasdaki ilişkinin anlamli olip olmadigini ilişkinin siddetini ve yonunu ifade eden istatistiksel bir tekniktir

Veri On Isleme

Veri Ön İşleme

Veriyi analiz için hazırlama sürecidir.

1. Aykırı Gözlem Analizi

• Aykırı Değerler: Verideki sıra dışı değerlerdir.

• Tespit: Boxplot, Z-Score, IQR yöntemleriyle bulunur.

• İşlem: Silinebilir, değiştirilir veya olduğu gibi bırakılır.

2. Eksik Gözlem Analizi

• Eksik Değerler: Boş veya eksik hücreler.

• Tespit: isnull().sum() veya görselleştirme ile bulunur.

• Doldurma: Ortalama, medyan, tahmin modelleri ya da eksik veriyi silme.

3. Değişken Dönüşümleri

• Amaç: Veriyi analiz için uygun hale getirmek.

o Logaritmik dönüşüm, normalize etme, kategorik değişkenleri sayısallaştırma (Label veya One-Hot Encoding).

Sonuç

Doğru veri ön işleme, daha güvenilir analiz ve sonuçlar sağlar.

Makine Ogrenmesine Sik Kullanilan Kavramlar

Bagimli Degisken Bagimsiz Degisken

Ilgilenen proejede hedeflenilen odaklanilan ana degisken (fiyat)

Bu degiskeni olusturdugu varsayilan digger degiskenlere de bagimsiz degisken denir (km, durumu,vites)

Ogrenme Turleri

Gozetimli ve Gozetimsiz Ogrenme

Bagimli degisken ve bagimsiz degiskenler bir aradaysa bir duruma gozetimli ogrenme denir

Bagimli degiskenin calisma icerisinde olmadigi turlere de gozetimsiz ogrenme denir

Amac Gozlem birimlerini birbirine benzer ozelliklerine gire bir araya getirektir

Regresyon Bagimli degiskenin sayisal/surekli bir degisken olmalis