

miuul

Data Scientist Workshop

Uber Yol Fiyatlarının Tahminlenmesi Projesi



Hande Küçükbulut

Ajanda

- 1** Veri Biliminin Kapsamlı İncelenmesi
- 2** Veri Bilimi Projesi İlk Adımlar
- 3** Veri Bilimi Projesinde Modelleme

1.1 Veri Bilimi Nedir?

Veri bilimi, yapılandırılmış ve yapılandırılmamış verilerden bilgi ve öngörü elde etmek için bilimsel yöntemleri, süreçleri, algoritmaları ve sistemleri kullanan çok disiplinli bir alandır.

Veri bilimi, bir dizi ilkeyi, çeşitli algoritmaları, olayları ve büyük veri kümelerinden gelen kullanışlı kalıpları ayıklamak için gerekli süreçleri kapsamaktadır. Bununla birlikte veri bilimi, bu süreçlerde; veri analizini, istatistikleri, makine öğrenmesi ve veri madenciliği gibi alanları ve bunlarla ilgili birçok yöntemi birleştirmek için kullanılan bir kavram olarak belirtilir.

1.1 Veri Bilimi Nedir?



$\text{İstatistik} + \text{Bilgisayar Bilimleri} = \text{Makina Öğrenmesi}$

1.2 Veri Bilimi Neden Gereklidir?

Veri bilimi, işletmelere kuruluş çapında dönüşüm yaratma potansiyeline sahip yeni düzen ve ilişkileri keşfetme olanağı tanır. Kâr marjlarını önemli ölçüde etkileyebilecek düşük maliyetli kaynak yönetimi değişikliklerini ortaya çıkarabilir.

Örneğin, bir e-ticaret şirketi, veri bilimini kullanarak müşterilerin çalışma saatlerinden sonra çok sayıda soru sorduğunu keşfediyor. Yapılan araştırmalar, sorduğu sorunun yanıtını ertesi iş günü yerine kısa süre içinde alan müşterilerin alışveriş yapma olasılığının daha yüksek olduğunu ortaya çıkarıyor. Bunun üzerine 7/24 müşteri hizmeti sunmaya başlayan işletme, bu sayede gelirini %30 artırıyor.

1.2 Veri Bilimi Neden Gereklidir?

Data Science

Advantages	Disadvantages
<ul style="list-style-type: none">A It's in DemandB Abundance PositionsC Highly Paid CareerD Highly PrestigiousE Versatile	<ul style="list-style-type: none">A It's in Blurry TermB Mastering Data Science is near to impossibleC Large amount of domain knowledge requiredD Arbitrary Data May Yield Unexpected ResultsE Problem of Data Privacy

1.3 Veri Bilimi Süreçleri

- **O** - Obtain data (Verileri edinme)Veri toplama
- **S** - Scrub data (Verileri ovma)
- **E** - Explore data (Verileri keşfetme)
- **M** - Model data (Verileri modelleme)
- **N** - Interpret results (Sonuçları yorumlama)

1.3 Veri Bilimi Süreçleri



Veriseti Hikayesi

Uber, San Francisco, Kaliforniya menşeli Amerikan uluslararası ulaşım ağı şirketidir. Hizmetlerine, araç çağırma, yemek teslimi, paket teslimi, kuryecilik, yük taşımacılığı dahildir. Uber mobil uygulaması geliştirir, pazarlar ve işletir.

Veriler Kasım 2018'de fiyatlar için her 5 dakikada, hava durumu için 1 saatte toplanmıştır. UBER verilerini açık olarak sunmadığı için veriler test olarak toplanmıştır.

İlk verimiz uzaklık ve fiyat bazlı bir veri ve ikinci verimiz hava durumunu içeren veridir. İkinci veriseti zamansal bazda hava durumu için kullanılabilecektir.

Uygulama | Veriyi Anlama ve Hazırlama

Adım 1: Projede kullanılacak dataları database'den alınız.

Adım 2: Verisetlerini tek bir dataframe olacak şekilde birleştiriniz.

Adım 3: İndirilen datalarının detaylı analizini yapınız.

Adım 4: Dataframelerdeki eksik değerler için imputation işlemlerini yapınız.

Adım 5: Dataframelerdeki aykırı değerler için kontrol yapınız.

Adım 6: 'time_stamp' değişkenini datetime tipine çeviriniz.

Adım 7: . Feature Engineering için target açısından değerlendirerek değişken oluşturmaya çalış.

Adım 8: Standartlaştırma yap.

Uygulama | Veriyi Anlama ve Hazırlama

Adım 9: Modellerini kur.

Adım 10: Model başarısını değerlendir.

Adım 11: Modeli daha iyi hale getir.

Adım 12: Özellik önemlerine bak.

Adım Sonsuz: Yeni projelere dal

BONUS : uber_pipeline

miuul

MIUUL™ Copyright© Miuul, Inc. All Rights Reserved

www.miuul.com