**1.2) Veri Bilimi Süreci**

Bu dokümantasyonda “Veri Bilimi Süreci nedir, nasıl ve kim tarafından yapılır?” Sorularına cevap bulacağız.

Sorular ve cevaplar neticesinde daha derinlemesine kavram öğrenimleri gerçekleştireceğiz.

**Veri Bilimci (Data Scientist) Kimdir?**



**Veri bilimci**; veriler ışığında bilgi odaklı keşifler yapabilen, mevcut durumu tanımlayan ve olay hakkında gelecekle ilgili tahminleri yazılım sistemleri kullanarak yapabilen kişidir.

Farklı araçlar kullanarak elindeki veriyi analiz eder, betimler, olay hakkında çıkarımlar (tahminler) yapar.

Özetle; **Veri Bilimi Sürecini** yöneten Kişidir.

**Veri Bilimci Becerileri**

1. Teknik Beceriler

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

Açıklama otomatik olarak oluşturuldu

1. Sosyal Beceriler

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

Açıklama otomatik olarak oluşturuldu

**Veri Bilimi Süreci Nedir? (Veri Bilimi Nasıl Yapılır?)**

Veri Bilimi Süreci, belirlenen iş problemini anlamak ve gidermek için izlenen sürece verilen isimdir.

Problem tanımlandıktan sonra, veri bilimci bu sorunu OSEMN veri bilimi süreciyle şu şekilde yönetir:

diyagram, taslak, plan, teknik çizim içeren bir resim

Açıklama otomatik olarak oluşturuldu

***1) Veri Toplama (Obtain data)***

Problemin belirlenmesiyle beraber ihtiyaçlara yönelik veri, veri setlerinin toplanması gerekmektedir.

Bu tür veriler gerekli kurum/kuruluşlardan satın alınabilir, veri madenciliği yöntemleriyle elde edilebilir veya web verilerinden elde edilebilmektedir.

**Örneğin**: APİ Uygulamaları sayesinde ücretsiz Twitter, Facebook, Instagram, Spotify verilerine ulaşılabilmektedir.

***2)******Veri Temizleme (Scrub data)***

Veri temizleme adıyla karşımıza çıkan bu süreç, verileri önceden belirlenen bir formata göre standart hale getirme sürecidir.

Eksik verileri ele alma, veri hatalarını düzeltme ve aykırı verileri kaldırma adımlarını içerir. Bazı veri temizleme örnekleri şunlardır:

· Tüm tarih değerlerini ortak bir standarda dönüştürmek.

· İmla hatalarını veya gereksiz boşlukları düzeltmek.

· Matematik hatalarını düzeltmek veya büyük sayılardaki noktaları kaldırmak.

***3)******Verileri Analizi (Explore data)***

Veri keşfi, ileride uygulanacak veri modelleme (makine öğrenmesi) stratejilerini planlarken kullanılan ön veri analizidir.

Veri bilimciler, açıklayıcı istatistiklerden ve veri görselleştirme araçlarından yararlanarak verileri genel hatlarıyla anlamaya başlar.

Bundan sonra, üzerinde çalışılabilecek veya eyleme dönüştürülebilecek ilgi çekici düzenleri belirlemek için verileri analizleri mümkün olur.

***4)******Veri Modelleme (Model data) (Makine Öğrenmesi)***

Daha derin öngörüler edinmek, sonuçları tahmin etmek ve en iyi eylem tarzını belirlemek için yazılım ve makine öğrenimi algoritmaları kullanılır.

Eğitim veri kümesine ilişkilendirme, sınıflandırma ve kümeleme gibi makine öğrenimi teknikleri uygulanır.

İsabet oranını değerlendirmek amacıyla, model önceden belirlenen test verileriyle karşılaştırılarak test edilebilir.

Sonuçların iyileştirilmesi amacıyla, veri modeli üzerinde üst üste defalarca hassas ayar yapılabilir. (AB Test)

***5)******Yorumlama ve Raporlandırma (Interpret results)***

Veri bilimciler, veri öngörülerini eyleme dönüştürmek için analistlerle ve işletmelerle birlikte çalışır.

Yapılan tüm çalışmalar sonucunda Veri Analizleri görselleştirilir, raporlandırılır ve gerekli kişi/kişilere aktarılır.

Verilerin özetlenmesi, yöneticilerin (kişilerin) sonuçları etkili bir şekilde anlamasına ve uygulamasına yardımcı olmaktadır.

Bu yüzden Veri Bilimciler için İletişim yeteneği oldukça önemlidir.

**Veri Bilimi Teknikleri Nelerdir?**

Veri bilimi uzmanları, veri bilimi sürecini izlemek için bilgi işlem sistemleri oluşturur.

Veri bilimciler tarafından kullanılan başlıca teknikler şunlardır:

***Sınıflandırma***

Sınıflandırma, verilerin belirli grup veya kategorilere göre tasnif edilmesidir.

Bilgisayarlar, verileri belirlemek ve tasnif etmek üzere eğitilir. Bilinen veri kümeleri kullanılarak, bir bilgisayarda verileri hızlı bir şekilde işleyen ve kategorize eden karar algoritmaları oluşturulur.

Örneğin:

· Ürünleri popüler veya popüler değil olarak tasnif etme

· Sigorta başvurularını yüksek riskli veya düşük riskli olarak tasnif etme

· Sosyal medya yorumlarını olumlu, olumsuz veya nötr olarak tasnif etme.

***Kümeleme***

Kümeleme, düzenleri veya anormallikleri bulmak amacıyla, birbiriyle yakından ilişkili verileri gruplandırma yöntemidir.

Kümelemenin, sınıflandırmadan farkı, verilerin sabit kategorilere isabetli bir şekilde sınıflandırılamamasıdır.

Bu nedenle veriler, en olası ilişkileri göz önünde bulundurularak gruplandırılır. Kümeleme sayesinde yeni düzenler ve ilişkiler keşfedilebilir.

Örneğin:

· Daha iyi müşteri hizmeti sunmak amacıyla, benzer alışveriş davranışları sergileyen müşterileri gruplandırma.

· Günlük kullanım düzenlerini belirlemek ve bir ağ saldırısını daha hızlı tespit etmek için ağ trafiğini gruplandırma.

· Makaleleri farklı haber kategorileri altında kümeleme ve bu bilgileri asılsız haber içeriklerini bulmak üzere kullanma.

***Regresyon***

Regresyon, alakasız görünen iki veri noktası arasında bir ilişki bulma yöntemidir.

Bağlantı genellikle bir matematik formülü etrafında modellenir ve bir grafik ya da eğriler olarak temsil edilir.

Veri noktalarından birinin değeri bilindiğinde, diğer veri noktasını tahmin etmek için regresyon kullanılır.

Örneğin:

· Hava yoluyla bulaşan hastalıkların yayılma hızı.

· Müşteri memnuniyeti ile çalışan sayısı arasındaki ilişki.

· Belirli bir konumda itfaiye istasyonlarının sayısı ile yangından kaynaklanan yaralanma sayısı arasındaki ilişki.

**Korelasyon**

Korelasyon, iki rassal (random) değişken arasındaki doğrusal ilişkinin yönünü ve gücünü belirtir.

Korelasyon (ilişki), iki değişkenin birlikte değişiminin bir ölçüsüdür.

Örneğin:

-Boy uzunluğu ile kilo arasındaki ilişki,

-çalışma saati ve sınav puanı arasındaki ilişki,

-hava sıcaklığı ve doğalgaz tüketimi arasındaki ilişki

**Veri bilimi tekniklerinin ardındaki temel ilke**

Ayrıntılar değişiklik gösterebilmekle birlikte, bu tekniklerin altında yatan ilkeler şunlardır:

Bir makineye, bilinen bir veri kümesine dayalı olarak verileri nasıl tasnif edeceğini öğretmek.

— Örneğin, bilgisayara örnek anahtar sözcükler ve her birinin tasnif değeri verilir. “Mutluluk” olumlu, “Nefret” ise olumsuzdur.

Makineye bilinmeyen verileri vermek ve cihaza veri kümesini bağımsız bir şekilde tasnif ettirmek.

Sonuçlardan bazılarının isabetsiz olacağı bilinciyle hareket etmek ve sonucun olasılık faktörünü ele almak.

**Veri Biliminde Kullanılan Farklı Teknolojiler**

1. **Yapay Zekâ:**

Yapay Zekâ (AI), görevleri yerine getirmek için insan zekasını taklit eden ve topladıkları bilgilere (verilere) göre yinelemeli olarak kendilerini iyileştirebilen sistemler veya makineler anlamına gelir.

Tahmine dayalı ve kuralcı analiz için makine öğrenimi modelleri ve ilgili yazılımlar kullanılır.

Kuralcı ya da normatif analitik (Prescriptive Analysis), gelecekteki eylemleri önermek için geçmiş trendleri ve verileri kullanarak tahmine dayalı analitiği bir adım öteye taşır.

Bu tür veri analizi, kaynakları optimize etmede ve genişleme gibi yeni iş fırsatlarını belirlemede çok faydalıdır.

1. **Bulut Bilgi İşlem (Cloud Computing):**

Bulut bilişim, bilgisayarlar ve diğer cihazlar için, istendiği zaman kullanılabilen ve kullanıcılar arasında paylaşılan bilgisayar kaynakları sağlayan, internet tabanlı bilişim hizmetlerinin genel adıdır.

Bulut teknolojileri, veri bilimcilere gelişmiş veri analizleri için ihtiyaç duydukları esnekliği ve işlem gücünü vermiştir.

**3)** **IoT (Nesnelerin İnterneti):**

IoT, internete otomatik olarak bağlanabilen çeşitli cihazları ifade eder.

Bu cihazlar, veri bilimi girişimleri için veri toplar.

Bunlar, veri madenciliği ve veri ayıklama için kullanılabilen, devasa boyutta veriler oluşturmaktadır.

**4)** **Kuantum Bilgi İşlem:**

Kuantum bilgisayarları, karmaşık hesaplamaları yüksek hızla yapabilir. Gerekli becerilere sahip veri bilimciler, karmaşık nicel algoritmalar oluşturmak için bunları kullanabilir.