**VERİ BİLİMİ ÖZET 1**

**A) Analitik türleri**

**Betimleyici Analitik :** Verileri basit bir görsel haline getirebilmek (Grafikselleştirme)

**Teşhis / Tanı Analitik :** Nasıl Olmuş , Neden olmuş gibi sorulara yanıt arar

**Tahminsel Analitik :** Ne Olacak sorusuna yanıt arar gelecekte olacaklara yönelik tahminlerde bulunulur

**Yönergeli Analitik :** Ne , nasıl Olmalı sorusuna yanıt aranır Ne yaparsam bu gelirler artar gibi sorulara cevap aranır

**B) CRİSP-DM Adımları**

**1)** **Veriyi Anlamak :** Problemi çözmek için değişkenlerin ortalamalarını bulup çözümlemek

**2)** **Verinin Hazırlanması :** Veri modellemesine geçilmeden önce veride ki bozuklukların düzeltilmesi

**Aykırı değer :** Normal verilerin anormal derecede dışarda olan verilere denir

**3)** **Modelleme :** veri setinde ki değişkenlerin yönünü ve şiddetinin bulunmasıdır ve o değişkenlerle beraber bir model oluşturmaktır

**4)** **Değerlendirme :** oluşturulan modelin test edilmesi eğer uygunsa kullanıma sunulur eğer uygun değilse tekrardan aşamalar ile gözden geçirilip hataların düzeltilmesidir

5) Kullanıma Sokma : Ürettiğimiz modelin eğer hatası vb. şeyleri yoksa internet sitesi vb. gibi alanlardan insanların kullanımına sokulmasıdır

**C) Veri Okuryazarlığı** **:** Her türden veri tipini , değişken , ve ölçek türlerini tanımlayabilme , betimsel istatistikleri ve istatiksel grafikleri kullanarak veri değerlendirebilme yeteneğidir.

* **Veri Okuryazarlığı Hakkındaki Kavramlar :**

**1)** **Popülasyon ve Örneklem :**

**a)** **Popülasyon :** İlgilendiğimiz hedef kitle

**b)** **Örneklem :** popülasyondan seçilen bir alt kümedir. Popülasyonu temsil etmesi gerekmektedir.

**2)** **Gözlem Birimi :** Örneklemde incelediğimiz her birey her bir satır vb. şeyler “Gözlem Birimi” olarak adlandırılır

**3) Değişken Ve Değişken Türleri :**

**Değişken :** Birimden birime farklı değerler alan niceliktir.

**A) Sayısal Değişkenler :** nicel , sayısal değişkenlerdir

**B) Kategorik Değişkenler :** nitel ,0 ve 1 ile belirtilen değişkenlerdir

**4) Ölçek Türleri :**

**A) Sayısal değişkenler için :**

* **Aralık (İnterval) :** Başlangıç noktası sıfır olmayan sayısal değişkenlerin ölçek türüne denir
* **Oran (Ratio) :** Başlangıç noktası sıfır olan sayısal değişkenlerin türüne denir

**B) Kategorik değişkenler için :**

* **Nominal :** Kategoriler arasında fark olmayan değişken türüne denir
* **Ordinal :** Kategoriler arasında fark olan değişken türüne denir

**5) Merkezi Eğilim Ölçüleri :**

**A) Aritmetik Ortalama :** Bir seride (değişkende) yer alan tüm değerlerin toplanarak birim sayısına bölünerek elde edilen istatistiktir.

**B) Medyan :** bir seriyi küçükten büyüğe veya büyükten küçüğe sıraladığımızda tam orta noktadan seriyi iki eşit parçaya ayıran değere denir

N tek ise : (n+1)/2 . Terim , N çift ise : [ (n/2) . Terim + (n/2 +1 ) terim ] / 2

***Not1 :*** Aritmetik ortalama , seri dağılımının simetrik olduğu bilindiğinde kullanılabilir . Medyan ise simetrik olmasa bile kullanılabilinir .

***Not2 :*** Aritmetik ortalama ve medyan birbirine yakın ise bu doğru bir dağıtım homojen olduğunu gösterir

**C) Mod :** Bir seride en çok tekrar eden değere denir.

**D) Kartiller :** Küçükten büyüğe sıralanan bir seriyi 4 eşit parçaya ayıran değerlere denir.

Q1 = (n+1) / 4 = ¼ x (n+1) . Terim

Q3 =(n+1).3/4 = ¾ x (n+1) . Terim

Q2 = Q3 – Q1

**6 ) Dağılım Ölçüleri :**

**A) Değişim Aralığı :** Veri kümesinde ki “En Büyük” değerden “En Küçük” değerin çıkartılması ile bulunur

Değişim Aralığı = Maksimum Değer – Minimum Değer

**B) Standart Sapma :** Ortalamadan olan sapmaların genel bir ölçüsüdür.Her bir terimin ortalamadan çıkarılıp karesinin alındıktan sonra toplanıp birim sayısına bölünüp karekökünün alınması ile bulunur

**C) Varyans :** Standart sapmanın karesinin alınmasıdır.

**D) Çarpıklık :** Değişken dağılımının simetrik olamayışıdır.

**Pearson Çarpıklık Katsayısı (PÇK) :**

3 x ( Ortalama – Medyan) / Standart Sapma formülü ile bulunur

Eğer ;

PÇK < 0 ise Negatif Çarpık (Soldan)

PÇK > 0 ise Pozitif Çarpık (Sağdan)

PÇK = 0 ise Simetriktir.

**E) Basıklık :** Değişken dağılımın sivriliği / basıklığını gösteren istatistiktir.

Basıklık Katsayısı = m4 / S üstü 4 ile bulunur

m = ortalamaya göre 4. Moment demektir ve her bir birimden ortalamayı çıkarıp 4. Kuvvetini alıp toplamlarının terim sayısına bölünmesi ile bulunur

S üstü 4 = Standart sapmanın 4. Kuvvetinin alınması ile bulunur .

Basıklık katsayısı eğer ;

3’e eşit (BK = 3) ise dağılım standart normal dağılıma uygundur .

3’ten büyük (BK > 3) ise dağılım sivridir .

3’ten küçük (BK < 3 ) ise dağılım basıktır .

**7) İstatistiksel Düşünce :**

**İstatistiksel Düşünce Modelleri :** Veri okuryazarlığından veri analitiğine giden yolu modelleyen yol göstericidir.

**Mooney Düşünce Modeli :**  4 basamaktan oluşur bunlar ;

**1) Verinin Tanımlanması**

**2) Verinin Organize Edilmesi Ve İndirgenmesi**

**3) Veri Gösterimi**

**4) Verinin Analiz Edilmesi Ve Yorumlanması**

**İstatistiksel Düşünce Düzeyleri :**

**\***Kişiye Özgülük (Seviye 1)

\*Geçici(Seviye 2)

\*Nicel(Seviye 3)

\*Analitik(Seviye 4)

**1) Verinin Tanımlanması :** Elimizde olan veriden ölçekleri kullanarak işe yarayacak bilgi çıkarma aşamasıdır.

**2) Verinin Organize Edilmesi Ve İndirgenmesi :** Elimizde olan verileri anlaşılabilir ve yorumlanabilir hale getirme aşamasıdır.

**3) Verinin Gösterimi :**  Elimizde olan verileri kullanıcıya sunum halinde (Grafikselleştirerek) sunma aşamasıdır .

**4) Verinin Analiz Edilmesi Ve Yorumlanması** : Elimizde olan sunum vb. şekilde hazırlanmış verilerin değerlendirilip en iyisinin hangisinin olup karar verme aşamasıdır.