Homework 1

**1) Makine Öğrenimini nasıl tanımlarsınız?**

Makine Öğrenimi, yazılım dilleri yardımı ile makinelere net olan beklentilerin/sorunların öğrenme ve geliştirme yeteneği vererek davranışlarında ve kararlarında bizim beklentilerimize yakın çözümleri üretmesi haline getirme çalışmalarının tümüdür.

**2) Denetimli ve Denetimsiz Öğrenim arasındaki farklar nelerdir? Bunların her biri için örnek 3 algoritma yazınız ve nasıl çalıştıkları hakkında kısaca bilgi veriniz.**

Denetimli öğrenmede "kategoriler", "sınıflar" veya "etiketlerin" bilinmesidir. Denetimsiz öğrenmede, sınıflar ve etiketler bilinmez ve öğrenme süreci uygun "kategoriler" bulmaya çalışılır. Ayrıca denetimli bir öğrenme algoritması bilinen bir girdi verisi seti ve verilere bilinen yanıtları alır, ardından yeni verilere yanıt için makul tahminler oluşturmak üzere bir model eğitilir. Bilinen gözetimli öğrenme sınıflandırma algoritmaları, destek vektör makinesi (SVM), karar ağaçları, k-en yakın komşu, Naive Bayes algoritmalarıdır. Bununla birlikte regresyon yöntemleri de (diskriminant analizi, lojistik regresyon) gözetimli öğrenmedir. Yaygın regresyon algoritmaları arasında doğrusal model, doğrusal olmayan model, regülasyon, kademeli regresyon, karar ağaçları sayılabilir. Gözetimsiz (denetimsiz) öğrenme, modeli denetlemenize gerek olmayan bir makine öğrenme tekniğidir. Bunun yerine, modelin bilgileri keşfetmek için kendi başına çalışmasına izin vermeniz gerekir. Kümeleme yöntemi bir gözetimsiz öğrenme yöntemidir, veri kümesi içerisinde benzerlikleri olan verileri otomatik olarak gruplara ayırır. Hiyerarşik kümeleme, Veri noktalarınızı üst ve alt sınırları belirlenerek kümeler. Örneğin müşterilerinizi daha genç ve daha büyük yaşlara bölebilir ve ardından bu grupların her birini kendi bireysel kümelerine de bölebilirsiniz.

**3) Eğitim, test ve doğrulama seti nedir ve neden onları kullanmamız gerekir?**

Eğitim veri seti kullanacağımız verinin %70 kadar oranının modelin veri setini tanıması ve eğitilmesi amacı ile kullanılan alanıdır.

Test veri seti, eğitim veri seti ile eğitilen modele verilerek tahmin sonuçlarının alındığı veri kümesidir. %30 oranında kullanılması tavsiye edilir. Ama beklenti ve veri setine göre değişkenlik gösterebilir.

Doğrulama veri seti ise, tahminlerin doğrulunu ve model başarı oranının görüldüğü veri setidir.

**4) Temel ön işleme (pre-processing) adımları nelerdir? Bu adımları ayrıntılı olarak açıklayınız ve verilerimizi neden ön işleme sokmamız gerektiği hakkında bilgi veriniz.**

Veri ön işleme adımları şu şekilde sıralanabilir hatalı kayıtların düzeltilmesi, tekrarlayan (duplicate) verilerin silinmesi, interpolasyon (interpolation), model eğitiminde kullanılacak verilere karar verilmesi (feature selection), normalizasyon (normalization), verinin dönüştürülmesi (feature extraction), test için kullanılacak verinin ayrıştırılması (train test split), çeşitli kodlama yöntemleri (encoding), anomali yaratan verilerin tespiti (anomaly detection) adımları sayılabilir , veri ön işleme sebebimiz ise model veriye göre tahminleme yapacağından sapmalar ve hataların giderilmesi ve modelin performans ve doğruluğunu yüksek tutmaktır.

**5) Sürekli (continuous) ve ayrık (discrete) değişkenleri nasıl belirleyebilir ve analiz edebilirsiniz?**

Sürekli bir değişkenin alanı belli aralıktaki tüm gerçek değerlerden oluşurken, ayrı değişkenlerin alanı en fazla sayılabilir. Genellikle ayrık değişkenler sayılar olarak tanımlanır, ancak sürekli değişkenler ölçümler olarak tanımlanır.

**6) Makine öğrenimi içinde kullanılan temel istatiksel kavramlar nelerdir?**

MerkeziEğilimÖlçüleri

‐Ortalama(aritmetikortalama)

‐Mod (tepedeğeri)

‐Medyan(ortanca)

‐AğırlıklıOrtalama

Dağılım(Değişkenlik)Ölçüleri

‐Sapma

‐Ortalama Mutlak Sapma

‐Varyans ve Standart Sapma

‐Değişim Katsayısı

**7) Aşağıda verilen grafiği analiz ediniz. (Grafik ve değişken türü nedir, dağılımı nedir ve hangi ön işlemlerden geçmelidir?)**

Seaborn kütüphanesinden distplot grafiği kullanılmıştır. S4 şeklinde bir veri setinden alınan nominal dağılımı hatalı gibi gözükmektedir.