



**T.C.**  
**FIRAT ÜNİVERSİTESİ**  
**TEKNOLOJİ FAKÜLTESİ**

**YAZILIM MÜHENDİSLİĞİ**  
**YAZILIM MÜHENDİSLİĞİNDE**  
**GÜNCEL KONULAR**  
**4.Aşama 1.Rapor**

**GİZEM ÇOBAN**

**175541307**

**PROJE DANIŞMANI**

**DOÇ. DR. FATİH ÖZKAYNAK**

**2019-2020**

## 1. Giriş

Bu hafta 4. Aşamada daha önceden belirlemiş olduğumuz veri seti için model araştırılması gerçekleştirildi. Araştırılan modeller hakkında bilgi toplanıp ne gibi işlemler yapıldığına bakıldı.

## 2. Yapılan Çalışmalar

Proje uygun olduğunu düşünülen modellerimiz şu şekildedir;

### **DecisionTreeClassifier**

Bir Karar Ağacı, örnekleri sınıflandırmak için basit bir temsildir. Verilerin belirli bir parametreye göre sürekli olarak bölündüğü bir Makine Öğrenimidir. Karar ağacında yani DecisionTreeClassifier algoritmasında öncelikle bir ağaç oluşturulur

### **KNeighborsClassifier (KNN)**

K En Yakın Komşu (KNN) çok basit, anlaşılması kolay, çok yönlü ve en üst düzey makine öğrenme algoritmalarından biridir. KNN algoritması hem sınıflandırma hem de regresyon problemleri için kullanılır. Özellik benzerliği yaklaşımına dayalı bir algoritmadır. KNN finans, sağlık, siyaset bilimi, el yazısı algılama, görüntü tanıma ve video tanıma gibi çeşitli uygulamalarda kullanılır.

### **GaussianNB (Naive Bayes)**

Naive Bayes sınıflandırıcısının temeli Bayes teoremine dayanır. lazy ( tembel ) bir öğrenme algoritmasıdır aynı zamanda dengesiz veri kümelerinde de çalışabilir. Algoritmanın çalışma şekli bir eleman için her durumun olasılığını hesaplar ve olasılık değeri en yüksek olana göre sınıflandırır. Az bir eğitim verisiyle çok başarılı işler çıkartabilir. Test kümesindeki bir değerin eğitim kümesinde gözlemlenemeyen bir değeri varsa olasılık değeri olarak 0 verir yani tahmin yapamaz. Bu durum genellikle Zero Frequency ( Sıfır Frekans ) adıyla bilinir. Bu durumu çözmek için düzeltme teknikleri kullanılabilir. En basit düzeltme tekniklerinden biri Laplace tahmini olarak bilinir. Kullanım alanlarına örnek olarak gerçek zamanlı tahmin, çok sınıflı tahmin, metin sınıflandırması, spam filtreleme, duyarlılık analizi ve öneri sistemleri verilebilir

### **RandomForestClassifier**

Karar ağaçlarının en büyük problemlerinden biri aşırı öğrenme-veriyi ezberlemedir (overfitting). Random Forest modeli bu problemi çözmek için hem veri setinden hem de öz nitelik setinden Rastgele olarak 10'larca 100'lerce farklı alt-setler seçiyor ve bunları eğitiyor. Bu yöntemle 100'lerce karar ağacı oluşturuluyor ve her bir karar ağacı bireysel olarak tahminde bulunuyor.

## **SVC**

Support Vector Machine sınıflandırma için kullanılan yöntemlerden birisidir. Temel olarak iki sınıfı bir doğru veya düzlem ile birbirinden ayırmaya çalışır. Bu ayırmayı da sınırdaki elemanlara göre yapar. SVC sınıfı scikit-learn kütüphanesi SVM modülünün altından projeye dahil edilmektedir.

### **3. Sonuç**

Önümüzdeki hafta ise bu modelleri kendi data setimizde kullanıp sonuçlarını karşılaştıracğız. En yüksek başarıyı veren modeli seçerek projemizde ilerleyen aşamalarda o modeli kullanarak çalışmaya devam edeceğiz.