**Proje Raporu**

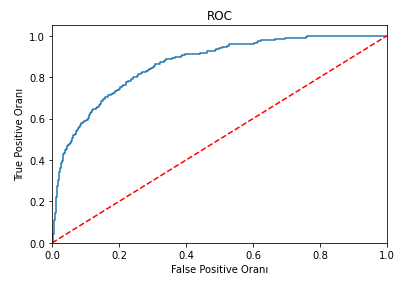
**Giriş**

Seçtiğimiz veri, Customer Personality Analysis. Veri, şirketin müşterilerinin ayrıntılı analizine yarıyor. Şirketin; müşterilerinin davranışlarını, ihtiyaçlarını anlamlandırmaya ve ürünlerini buna göre dizayn ettirmesini kolaylaştırır.

Müşteri kişilik analizi sayesinde her müşteriye yeni ürün pazarlamak yerine hangi müşterinin hangi ürünleri satın alabileceğini analiz ederek o ürünü o müşteriye pazarlayabilir. Bu sayede hem maddi olarak hem de zamandan kar etmiş olur. 2240 veri ve 29 attribute içeriyor. Datada insanların özellikleri, ürünlerin özellikleri, müşterilerin mağazayı nerede ziyaret ettikleri ve kampanyaların kategorize edilmiş şekilde öznitelikleri bulunuyor. Bu veriyi gerekli kategorize işlemleri yaparak Python ( Jupyter Notebook) ile analiz edeceğiz. Bu analizdeki amacımız bazı müşterilerin bu kampanyaya verdiği yanıtları başka birinin tepkilerini tahmin etmek için kullanabilir miyiz?

**Gelişme**

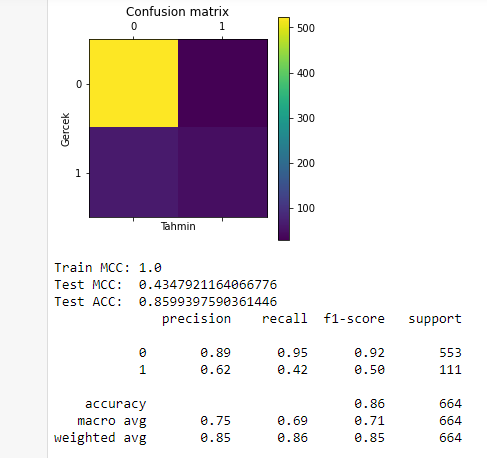
Veri ön işleme teknikleri olarak veri temizleme, veri dönüştürme ve veri azaltmayı kullandık. Incomeda null değerleri olmasından dolayı dropna komutu ile datayı 2216ya düşürdük. Daha sonra işimize yaramayan bazı attributeları sildik. Education ve Marital attributelarının değerleri fazla geldiginden sadeleştirdik. Müşterinin yaptığı harcama tutarını tek attribute oluşturarak toparladık. Age ve Incomeda bazı ayrık verileri grafikte tespit edip bu değerleri ayıklamak için belirli aralık dışındaki verileri sildik. Elimizdeki özelliklerden Education, Dt\_Customer ve Marital kategorik değişkenler olduğu için bu değişkenleri sayısal değişkenlere for döngüşü içerisinde LE.fit\_transform komutu ile çevirdik.

 Ödevde kullandığımız algoritmalar ; logistic regression, confusion matrix,ROC eğrisi, cross validation, KNN, XGBClassifier, Karar ağaçları, Xgboost, Gradient Based Boosting. Logistic kullanırken başta Y değerine Response özelliğini atadık. X değerleri ise Response haricindeki tüm özellikler. Accuracy score değeri ise 0.88. Daha sonra %25 test olarak ayırıp tekrar algoritmayı çalıştırdığımda accuracy score sonucunu 0.87 olarak aldık.

Cross validationu ise k yı 10 seçerek çalıştırdığımızda 0.84 sonucunu aldık. KNN algoritmasını çalıştırmadan önce testi %30 olarak belirledik. Random\_state kısmını araştırma yaptığımızda genel olarak 42 seçildiğini gördük ve bunu seçtik. Burada da accuracy sonucumuz 0.81 çıktı. KNNyi grid search ile optimum vaziyete getirmek için tekrar çalıştırdığımızda accuracy 0.86e yükseldi. XGB algoritmasını uyguladığımızda accuracy 0.87 sonucunu verdi. Daha sonra %30 test, %70 train ile Decision Tree algoritmasını çalıştırdığımızda test setinde 0.83 başarı yakaladık. Ada Boost Classifier algoritması ile confusion matrix oluşturduk. Burada da 0.86 sonucunu yakaladık ve matrixi yazdırdık.

**Sonuç**

Sonuç olarak Modelin genel test dogrulugu 0.86'dır. Model negatif örneklerde iyi sonuc verirken pozitif orneklerde iyi sonuc vermiyor. Test MCC'si 0.43'tür. Bu da modelin test setinde pozitif örnekler bulmada iyi olmayabileceğini gösterir. Bu sonuç veri setinde kullandığımız tahmin edicilerin 'Response'u çok iyi tahmin edemeyebileceğini gösterebilir.



**İsmail Gülbay 18052032**

**Serhat Özdemir 17052009**