Project G

Koç University, QMBU 450, Final Project

Prepared by Gülnaz Baş , 64965

About the Project

- Customer Potential and Market Size are considered to be associated with the sales of product X.

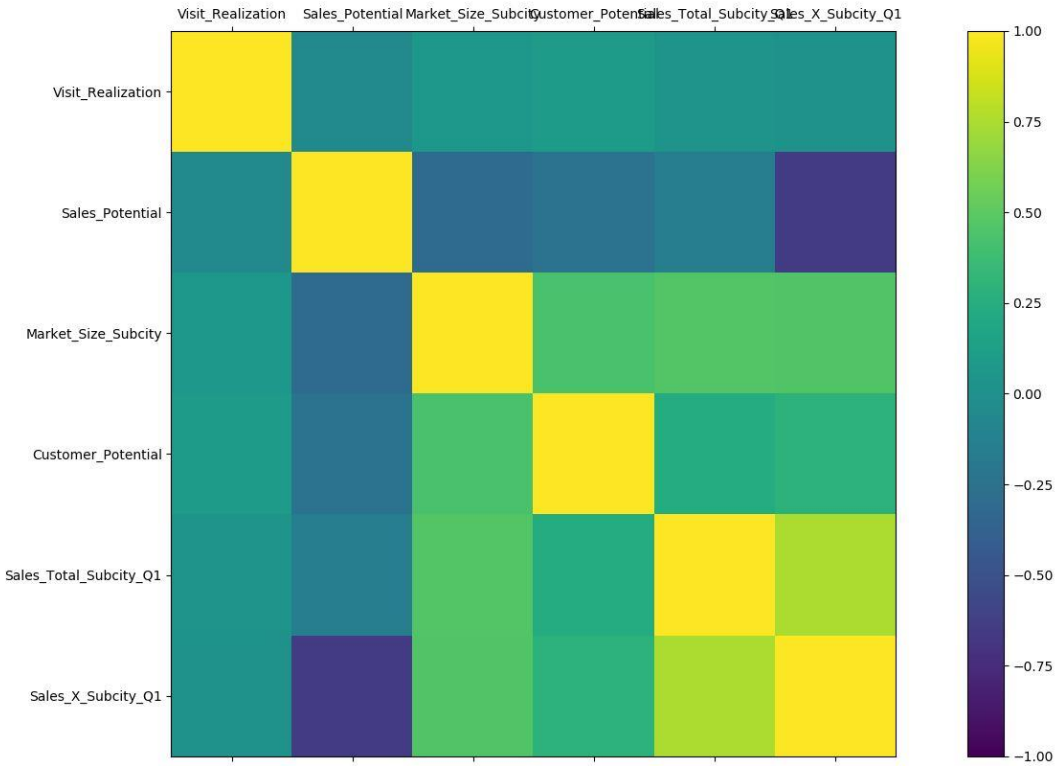
- In this project, the accuracy of this relationship and the estimation of how much sales can be made by using these inputs are questioned.

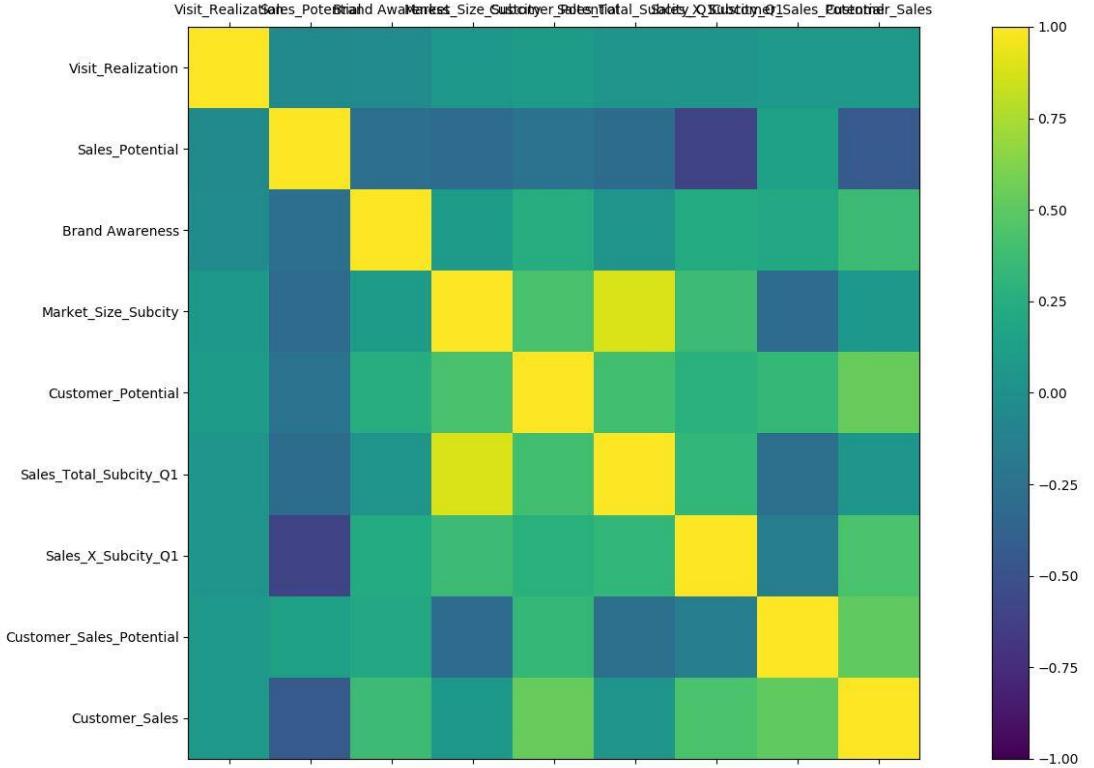
Overall Data Sets

CRM data of a wholesale company was used.

- CustomerID: Uniq Customer ID
- Product: The item offered for sale
- City: City where the customer lives
- SubCity: Region where the customer lives
- Store: the store that the customer is using (can only shop in one store.)
- Current Target: Sales target segment for the customer
- Visit_Target: Customer Communication target (Quarter)
- Actual_Visit: Number of communications (Quarter)
- Brand Awareness: Customer's brand loyalty (Calculated by other product usage numbers.)
- Customer Potential: Customer's shopping potential multiplier
- Market_Size_Subcity: Potential of the region where the customer lives
- Sales_Total_Subcity_Q1: Total sales of the region where the customer lives
- Sales_X_Subcity_Q1: Sales of X products in customer’s region

Correlation of Data





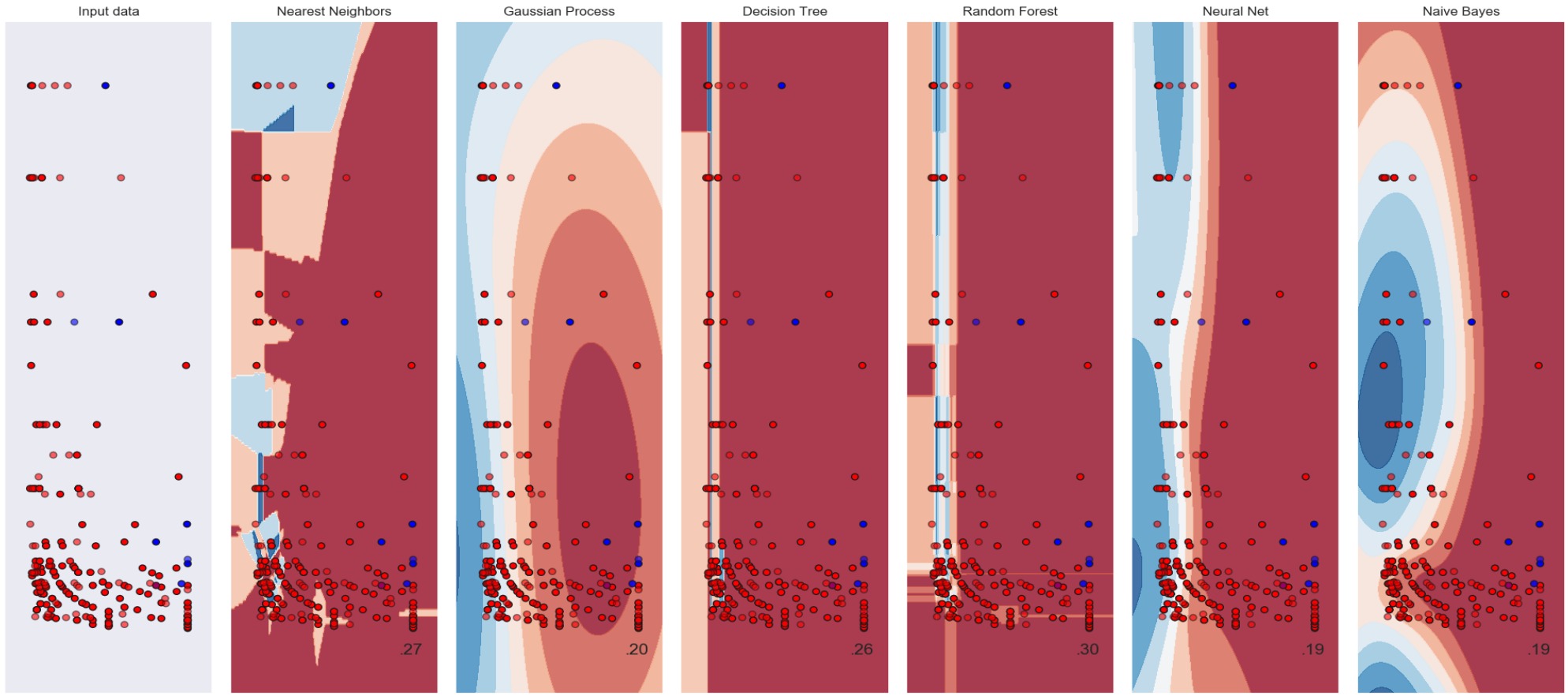
The Data for creation model

- There is a positive relation between Market Size Subcity – Customer Potential, Sales Total Subcity, Sales X Subcity
- Customer Sales and Sales Potantial arasında have a strong negative relation.

The Test Data

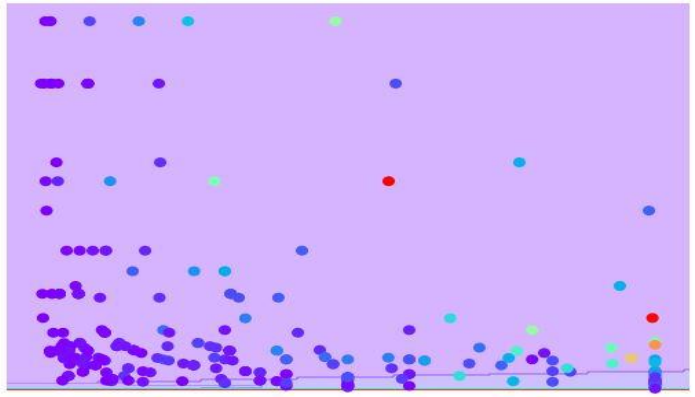
- Market Size Subcity – Customer Potential, Sales Total Subcity, Sales X Subcity have a positive relation.
- Sales X Subcity and Sales Potantial, have a strong negative relation.
- Customer Sales and Sales Potantial have a strong negative relation.

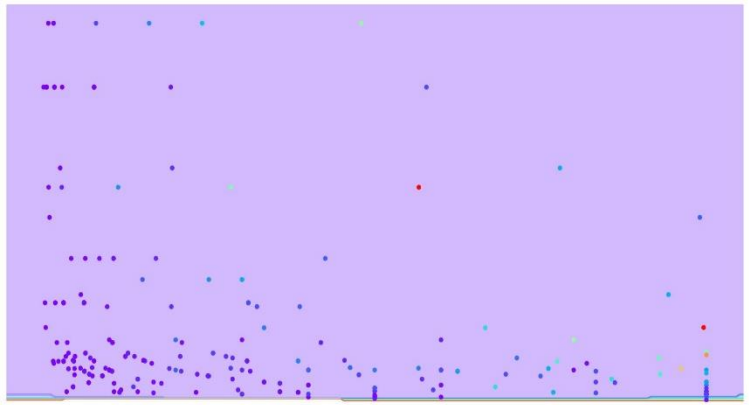
Selection of Model



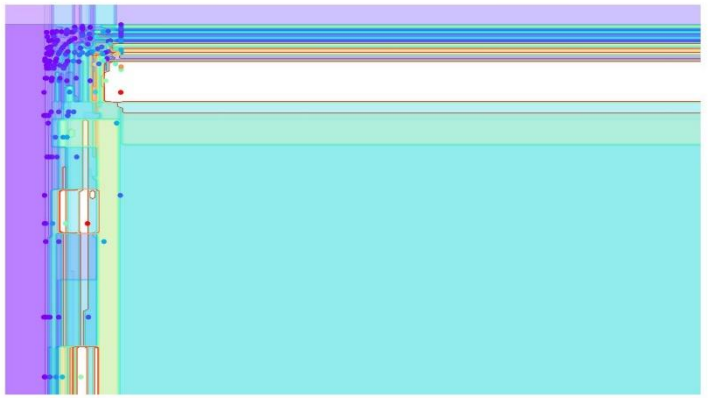
The most suitable model was Decisition tree and Random Forest. In detail, the Decision Tree Calssification (also BaggingClassifier) and MLPClassifier algorithms are studied.

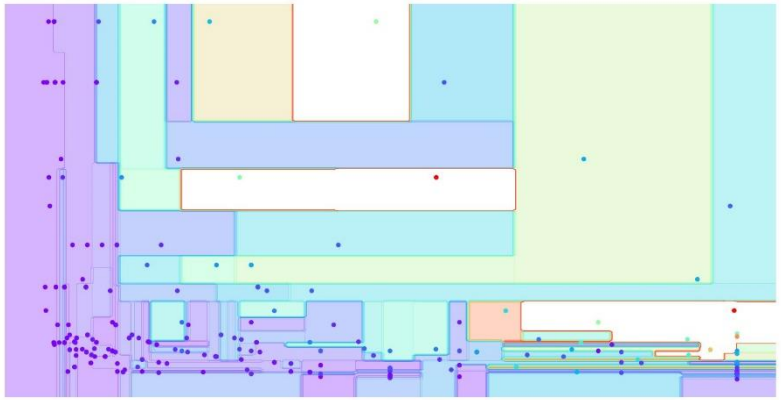
MLPClassifier



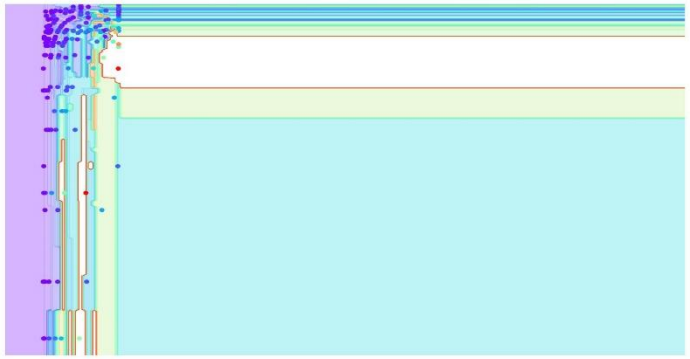


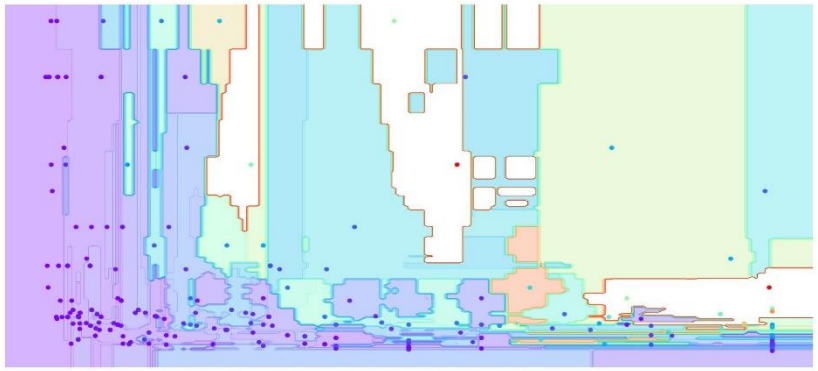
DecisionTreeClassifier





BaggingClassifier





Results

Data for Creation Model Prediction Result

MSE = $\sum_{k=1}^{n}$ (Y − Y_hat)^2 ~ 2

Test Data Model Prediction Result

MSE_test = $\sum_{k=1}^{n}$ (Y_t − Y_t_hat)^2 ~2

- The model is estimated with an error of 17% on average.