

Data Mining CSE 385

**Final Project**

Submitted To: **Engineer Ahmed Hesham**

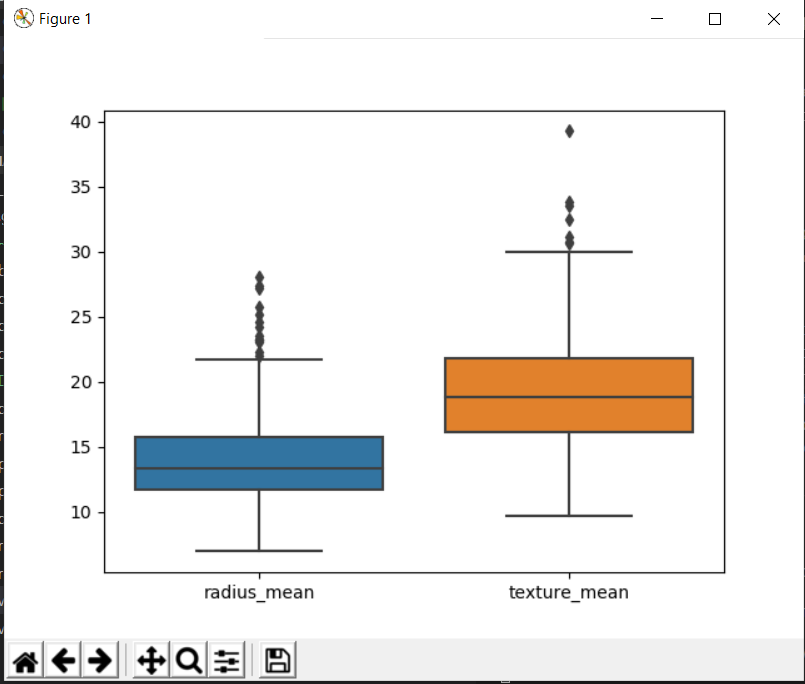
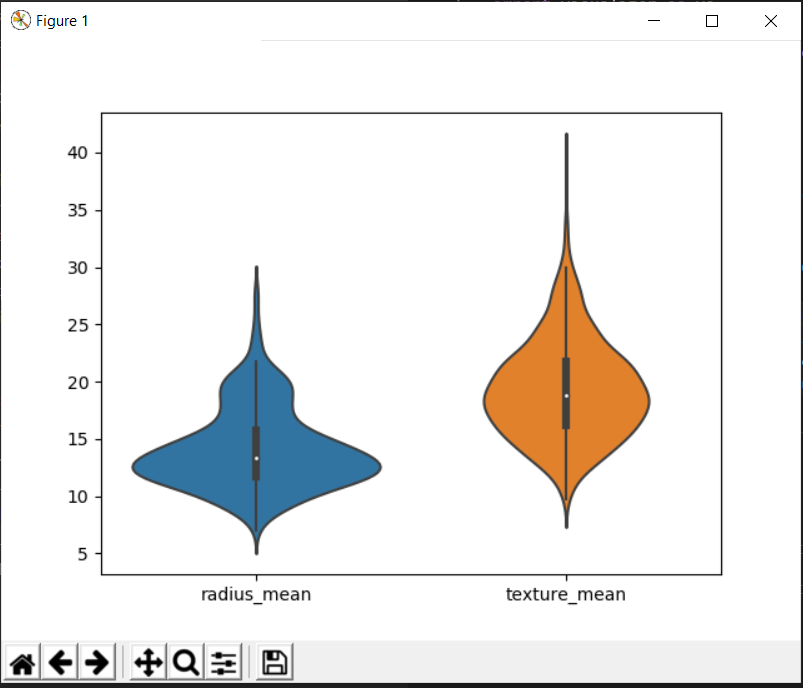
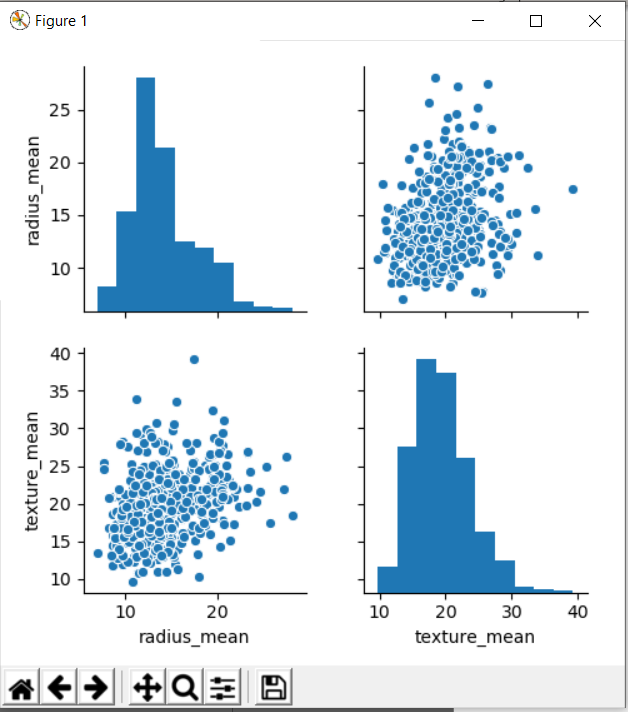
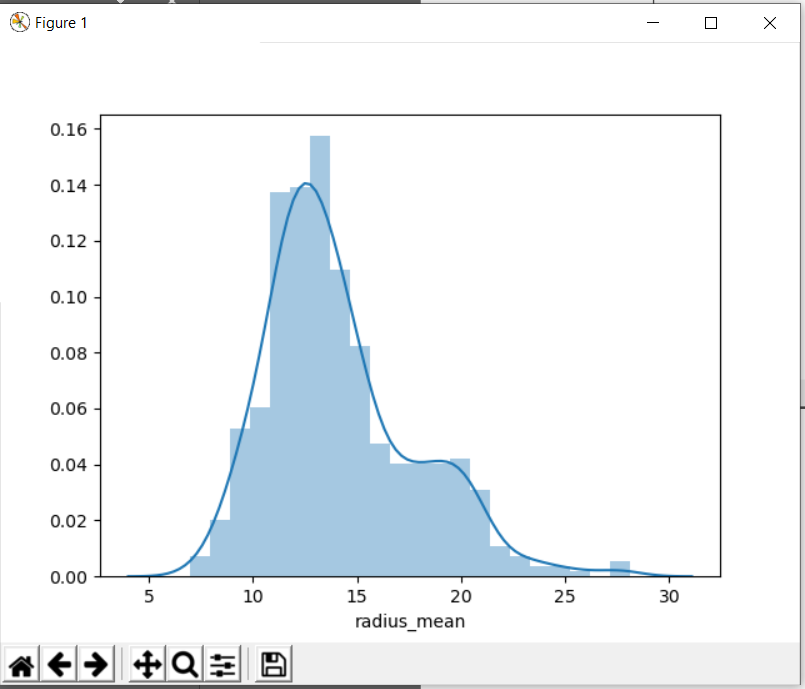
Submitted By:

**Mohamed Sameh 16p3061**

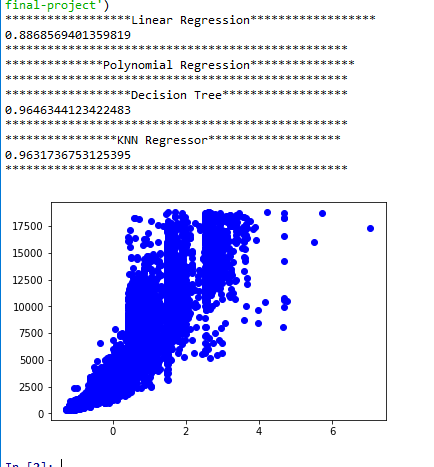
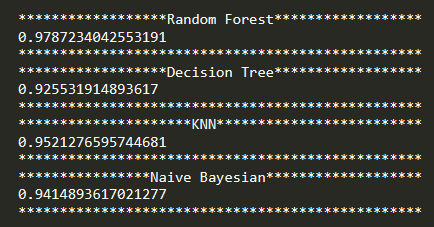
**Nour E-din Osama 16p6043**

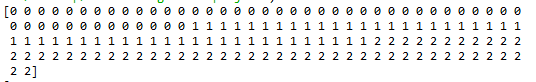
# Bonus Features

We have implemented the following bonus features to our project

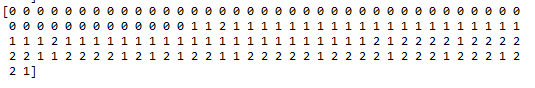
1. **Visualization:** (box plot, violin plot, pair plot, histogram)  
     
     
     
     
     
     
   
2. **Preprocessing:** We have added two extra options to **handling missing values**, insert mean and insert mode.
3. **Regressor:** We have implemented an extra type called Decision Tree Regressor
4. **Classifier:** We have implemented the **Random Forest** classifier, which is the same as decision trees except it uses multiple decision trees as to avoid the overfitting problem of a normal decision tree.

# Results

1. **Regressors:** for the regressors we used the Diamonds dataset from Kaggle and got the following results.  
   
2. **Classifiers:** for the classifiers we used the Wisconsin Breast Cancer dataset from Kaggle **after** removing the irrelevant columns like the ID and got the following results.  
   
3. **Clustering:** for the clustering we used the IRIS dataset after removing the class label and got the following results.

**Real Clustering after encoding**  


**Output Clustering**



# Implementation