**Machine Learning Predictions Desktop Application**

The application developed is to enable the analysis of business processes that needed by different departments within a company and to make predictions of planned and anticipated targets in the light of the data recorded in the database.

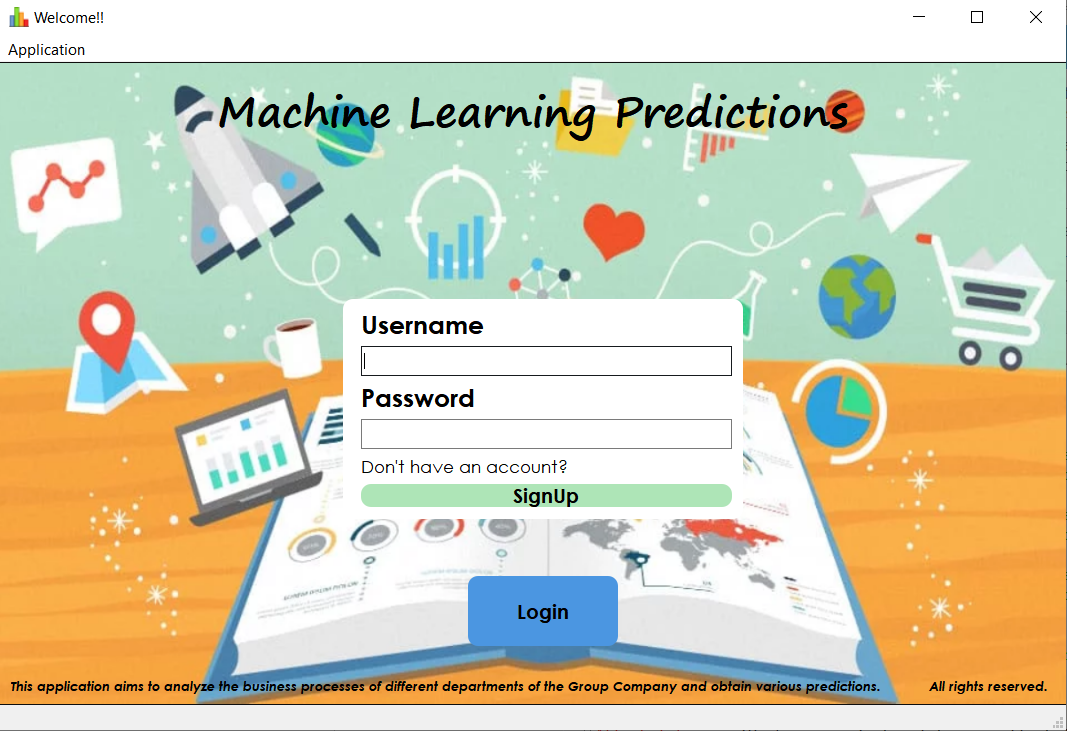
The application analyzes data and produces predictions by using supervised, unsupervised, and regression within the scope of machine learning methods. In this way, it is aimed to produce the most accurate estimates by minimizing the budget, production and shipment expenditure risks.

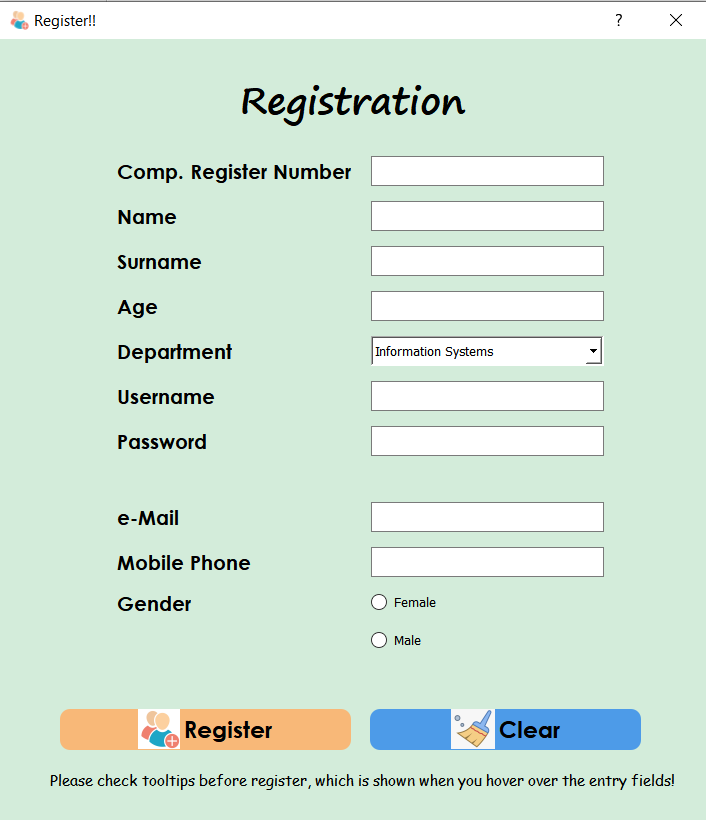
It will enable the company to obtain more effective and productive results in line with the predictions made in the competitive environment in the market.

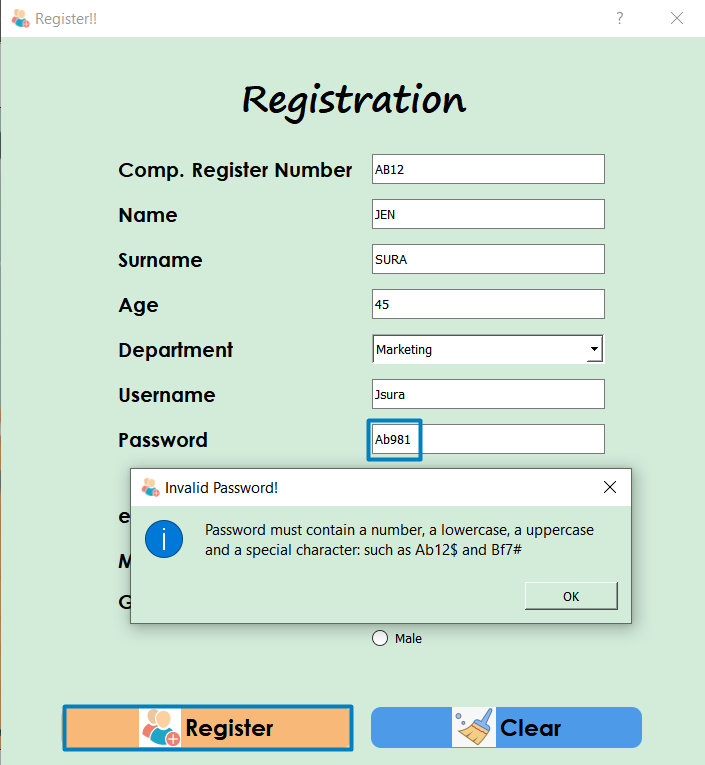
The resource planning for the future periods and optimum estimates closest to the correct result, it will achieve maximum profit with minimum budget and expenditure.

The fact that experiments and examinations can be carried out faster ensures that R & D can be carried out much faster and more effectively in many sectors and provides higher quality products.

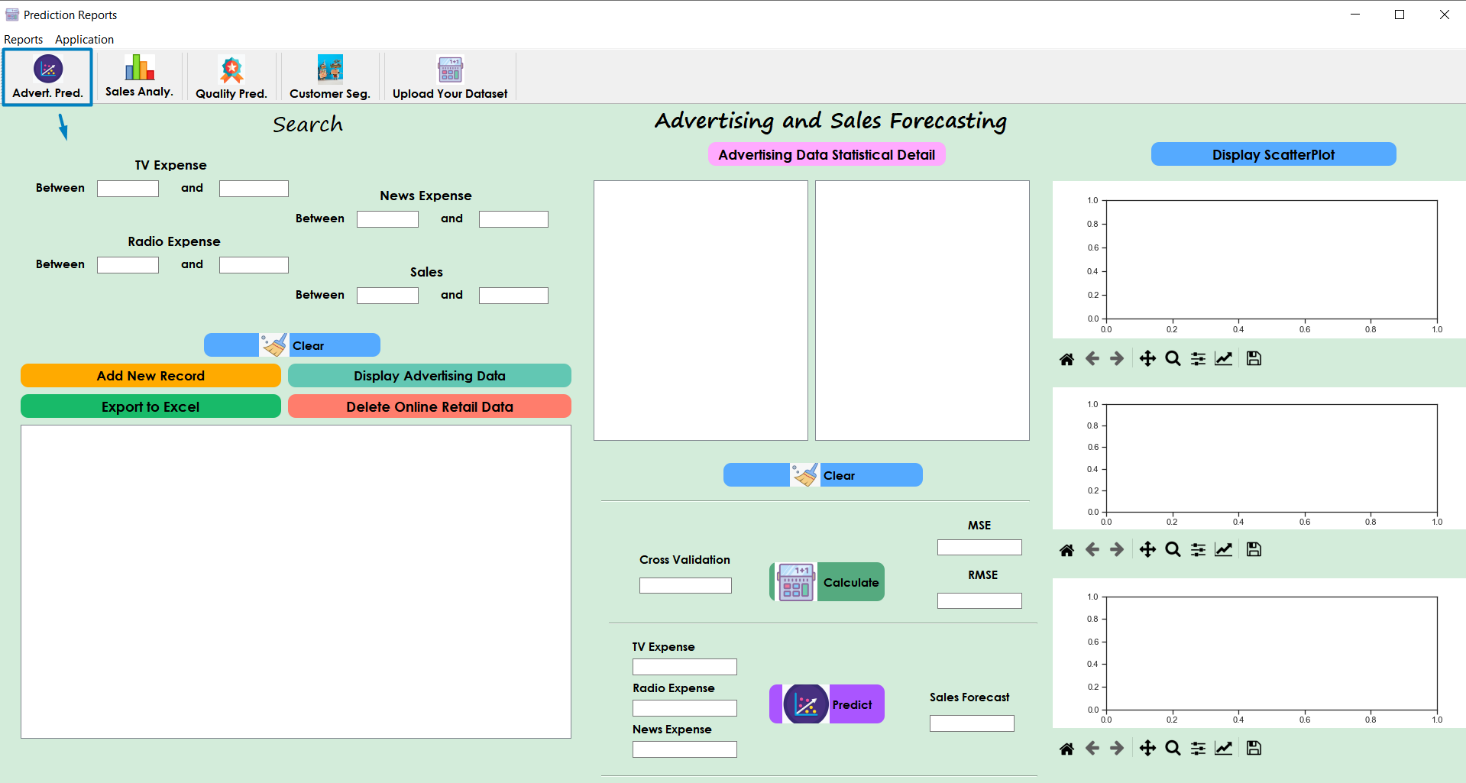
It plays an important role in product market analysis, determination of consumer behavior, determining the factors that determine efficiency and quality in production, determining the best sales, marketing and production methods for companies, and it plays an crucial role in the sustainability of the processes and enables them to be managed with predictable outcomes.

Home screen of the application is as follows. If the user has an account on the login screen, it can be Login with username and password information.

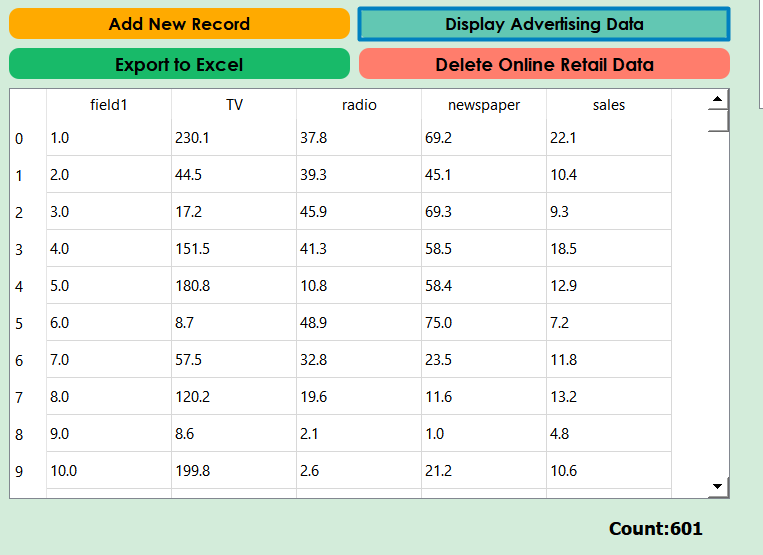
If the user does not have an account, a new user can be registered with the “**SignUp**” button and should fill the new user form below by checking the tooltips of fields otherwise the application will warn the user about the incorrect entry.



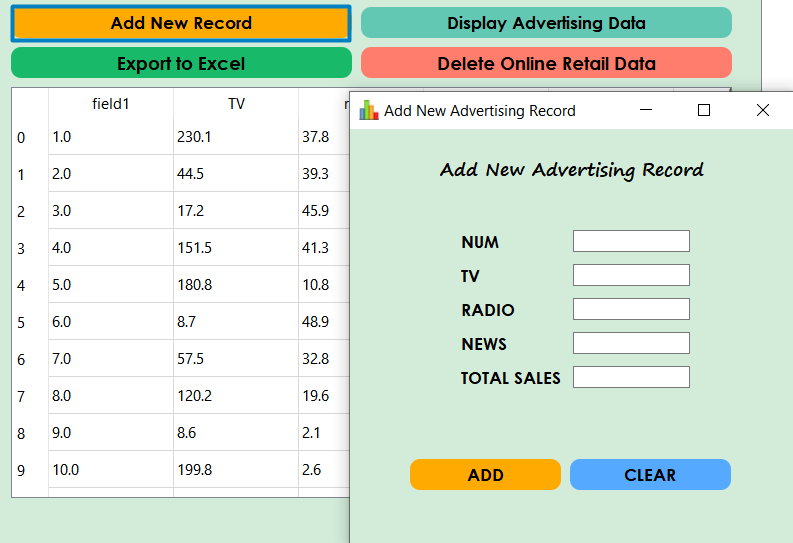
1. **Advertising and Sales Forecasting**

This page contains the advertising expenses data of the advert department in the company.

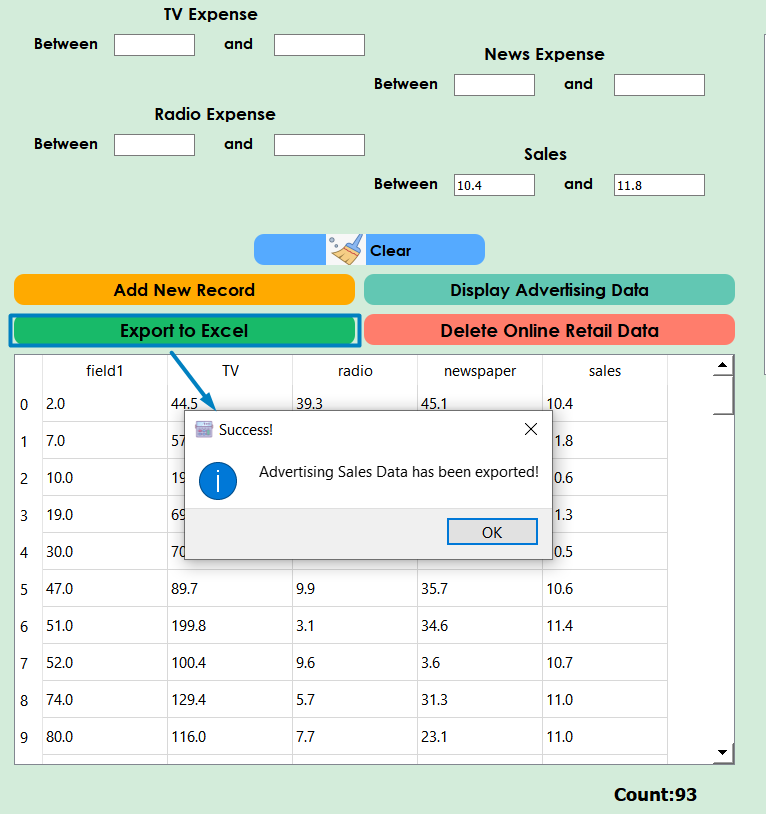
* All or filtered advertising data can display with the **“Display Advertising Data”** button.



* If the user wants to add a new record, can use the “**Add New Record button**” and fill the new advertising record form in the incoming window.

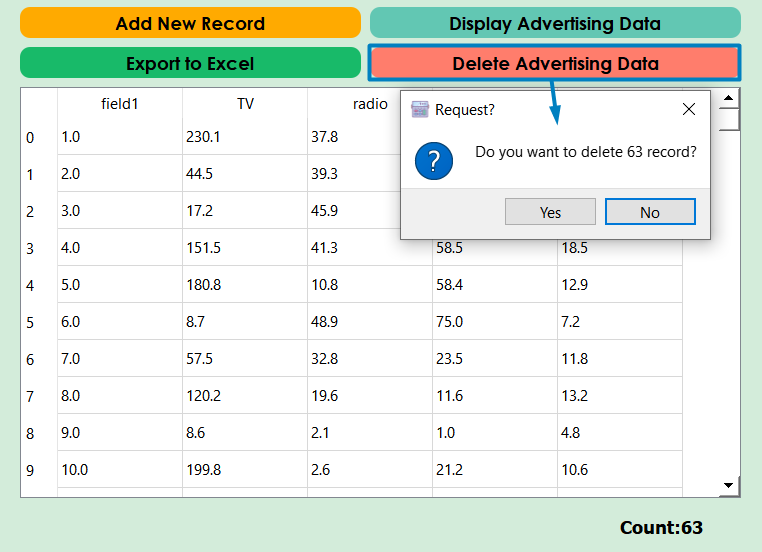


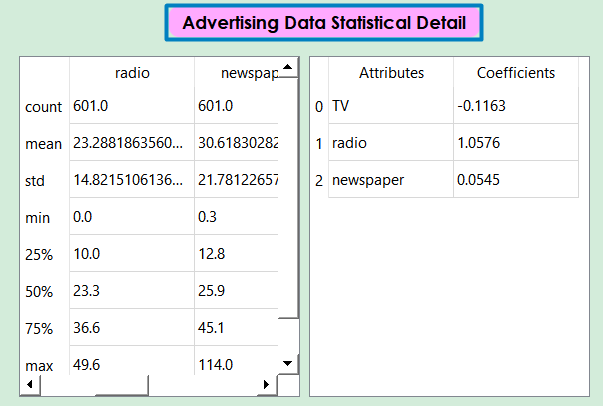
* If the user wants to export all or filtered data can use the “Export to Excel” and export displayed data to the excel file in the project file directory.





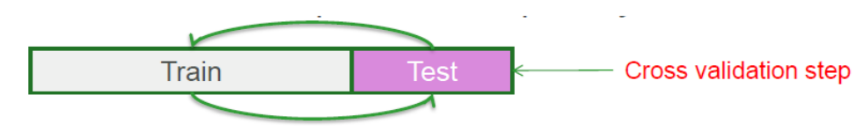
exports displayed data to the excel file in the project file directory.

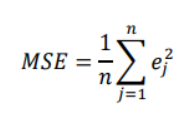
* If the user wants to delete the filtered data (allows only the filtered data to be deleted, not all), you can use the "**Delete Advertising Data**" button and the application will once again prompt the user for the number of lines to be deleted.
* “**Advertising Data Statistical Detail**” button displays the statistical properties of the attributes in the first table and contains the coefficients of attributes for the sales prediction of each of the targets.

****

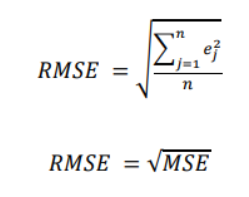
* “**Calculate**” button, calculates the **MSE** and **RMSE** values with the **Cross Validation** parameter.
* **Cross Validation**

In cross-validation, the original sample is split into two parts. One part is called the training (or derivation) sample, and the other part is called the validation (or validation + testing) sample. Modeling of the data uses one part only. The model selected for this part is then used to predict the values in the other part of the data. A valid model should show good predictive accuracy.



* **MSE (Main Squared Error)**

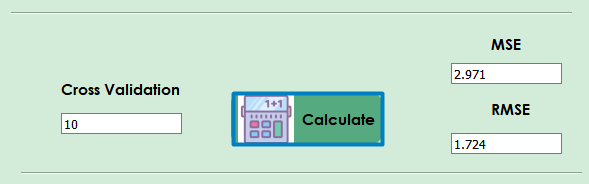
The mean square error indicates how close a regression curve is to a set of points. The MSE measures the performance of a machine learning model, the predictor is always positive and it accepts that predictors with an MSE value close to zero perform better.



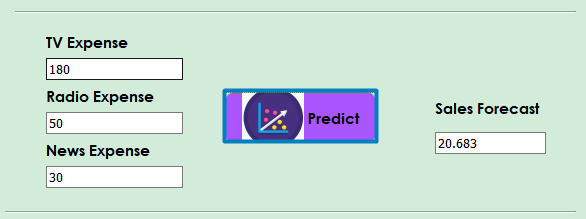
* **RMSE (Root Mean Square Error)**

The RMSE value is the standard deviation of the estimation errors. It is a measure of how far the regression line is from the data points; RMSE is a measure of how far these errors spread.

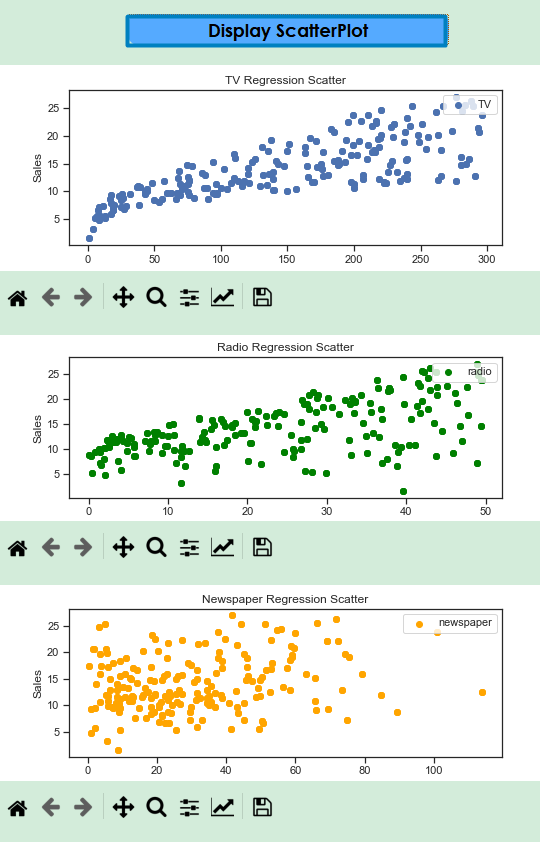
A zero RMSE value means the model made no mistakes. RMSE has more advantage of punishing big mistakes.



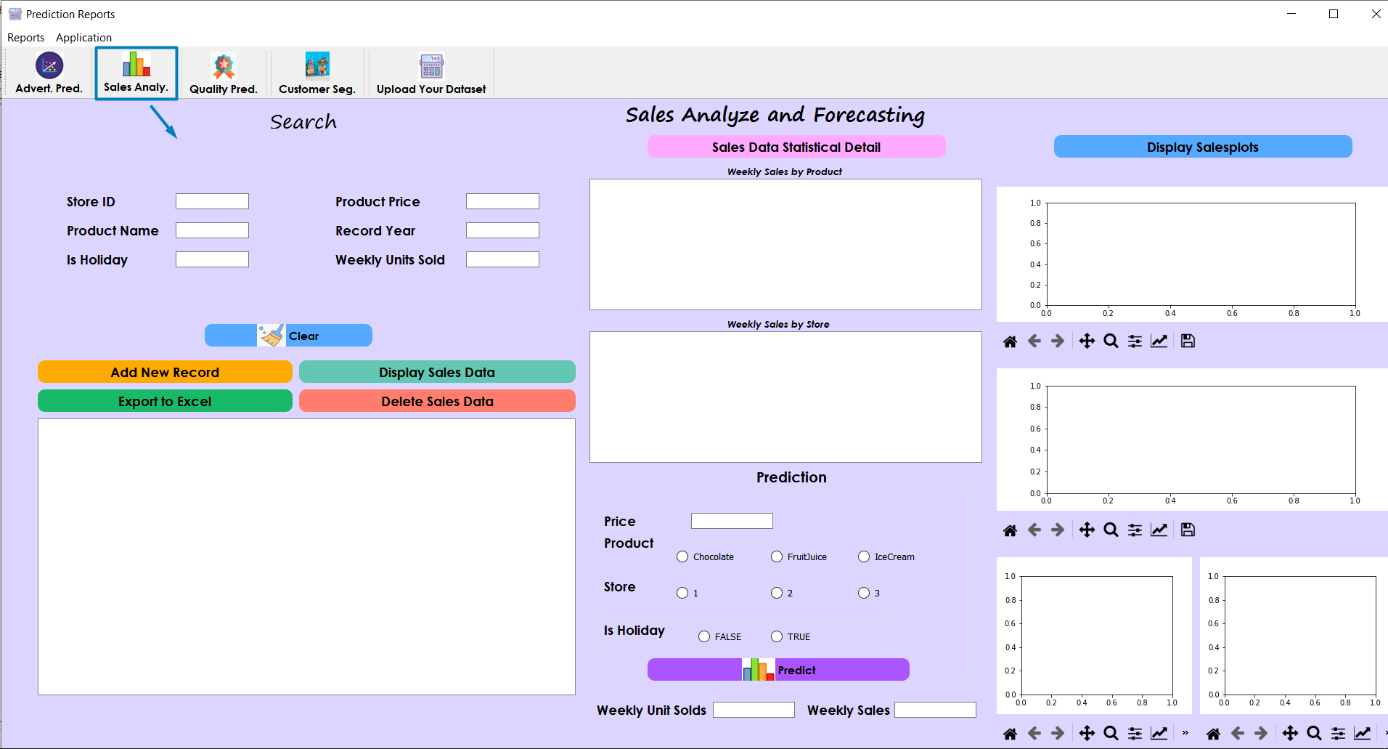
* **“Predict”** button gives the sales prediction by using Linear Regression which is provided with the appropriate expenditure values to be entered by the user.

****

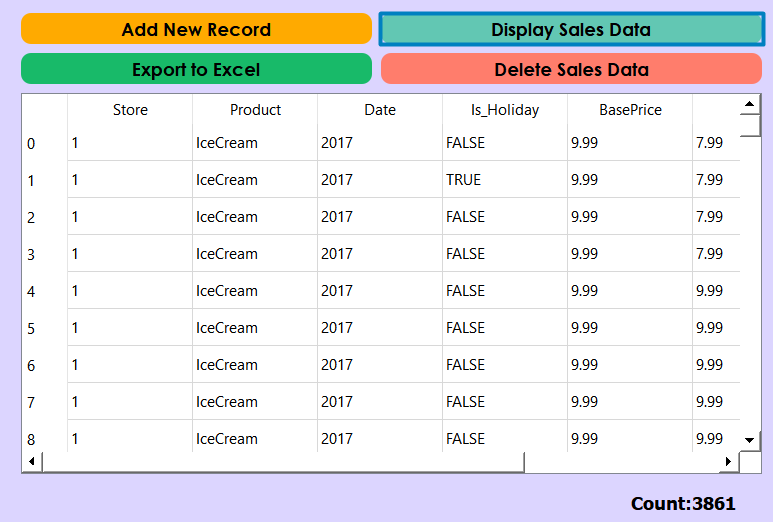
* “**Display ScatterPlot**” button displays the scatters of the advertising attributes (TV,Radio,News) to the sales data.

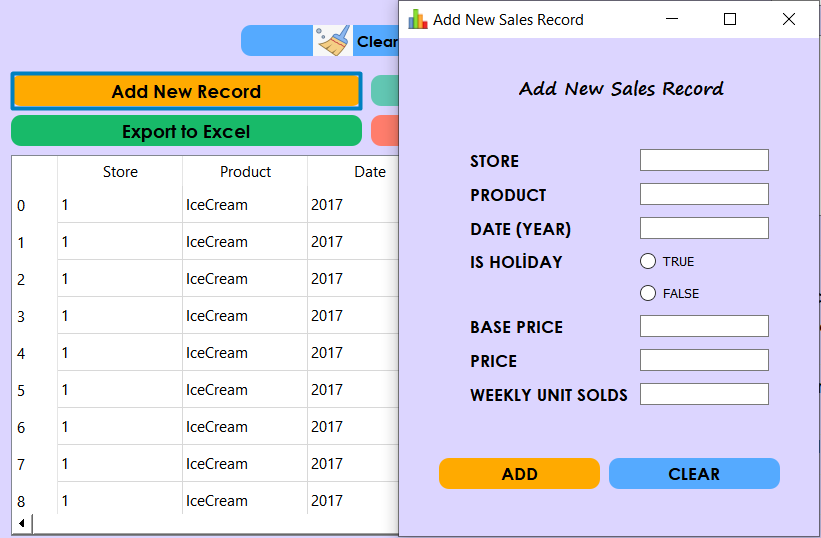


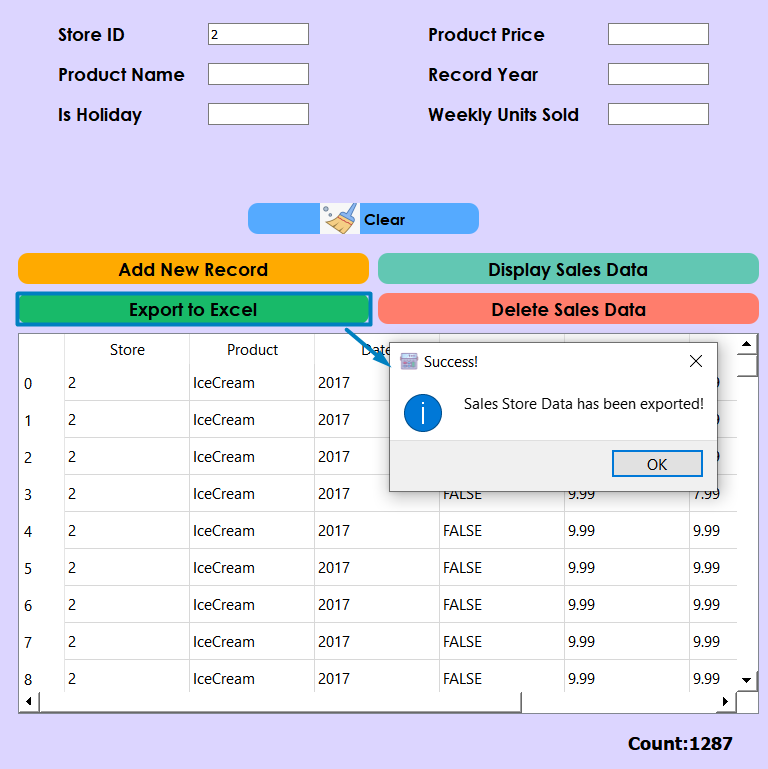
1. **Sales Analyze and Forecasting**

****This page contains the product sales data of the Sales Department in the company.

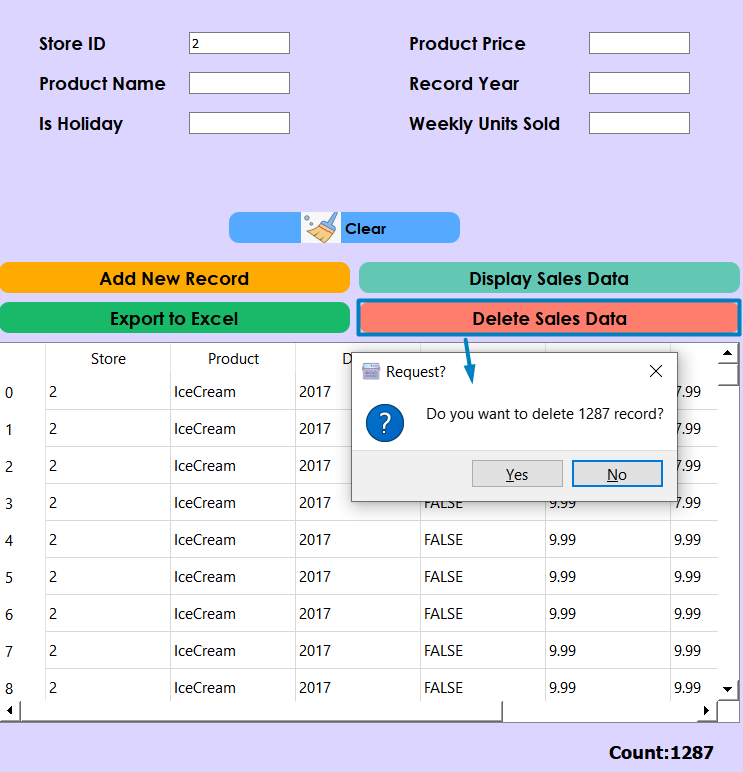
* All sales data can display with the **“Display Sales Data”** button.



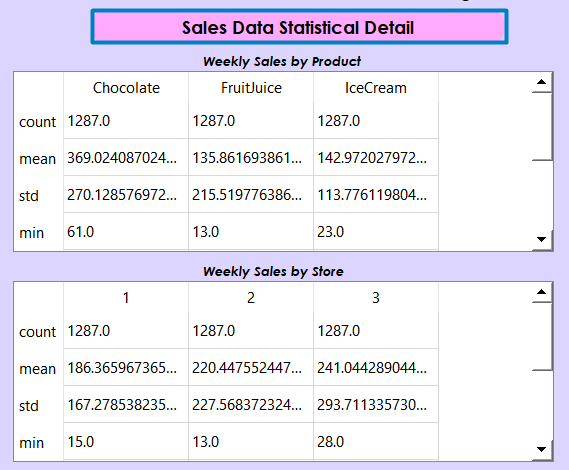
* If the user wants to add a new record, can use the “**Add New Record button**” and fill the new sales record form in the incoming window.
* If the user wants to export all or filtered data can use the “Export to Excel” and export displayed data to the excel file in the project file directory.



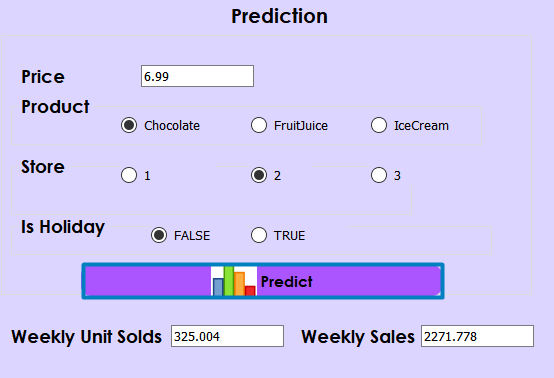
* If the user wants to delete the filtered data (allows only the filtered data to be deleted, not all), you can use the "**Delete Sales Data**" button and the application will once again prompt the user for the number of lines to be deleted.



* “**Sales Data Statistical Detail**” button displays the statistical properties of the Products in the first table and Stores in the second table according to the Weekly Unit Sold Totals.

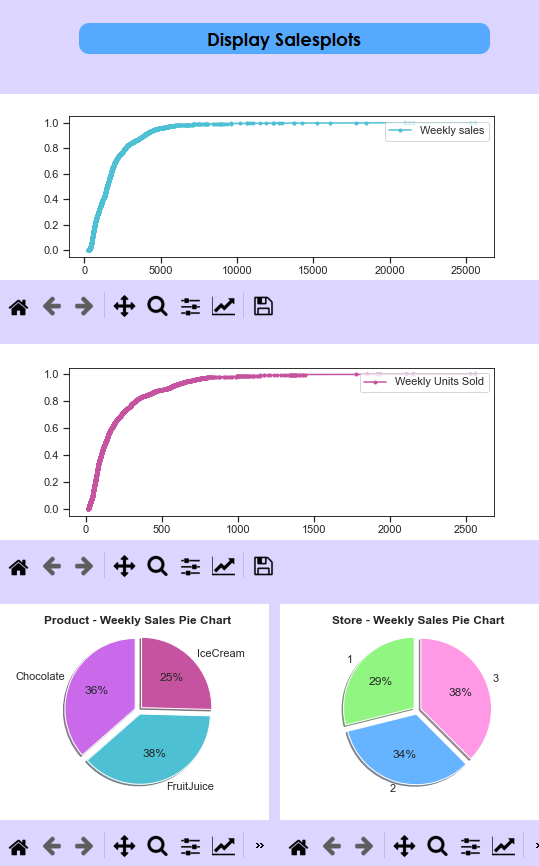


* The Price, Product, Store and Is Holiday values to be entered by the user and the "**Prediction**" button gives the provided Weekly Unit Sales and Weekly Sales forecast by using ElasticNet Regression.

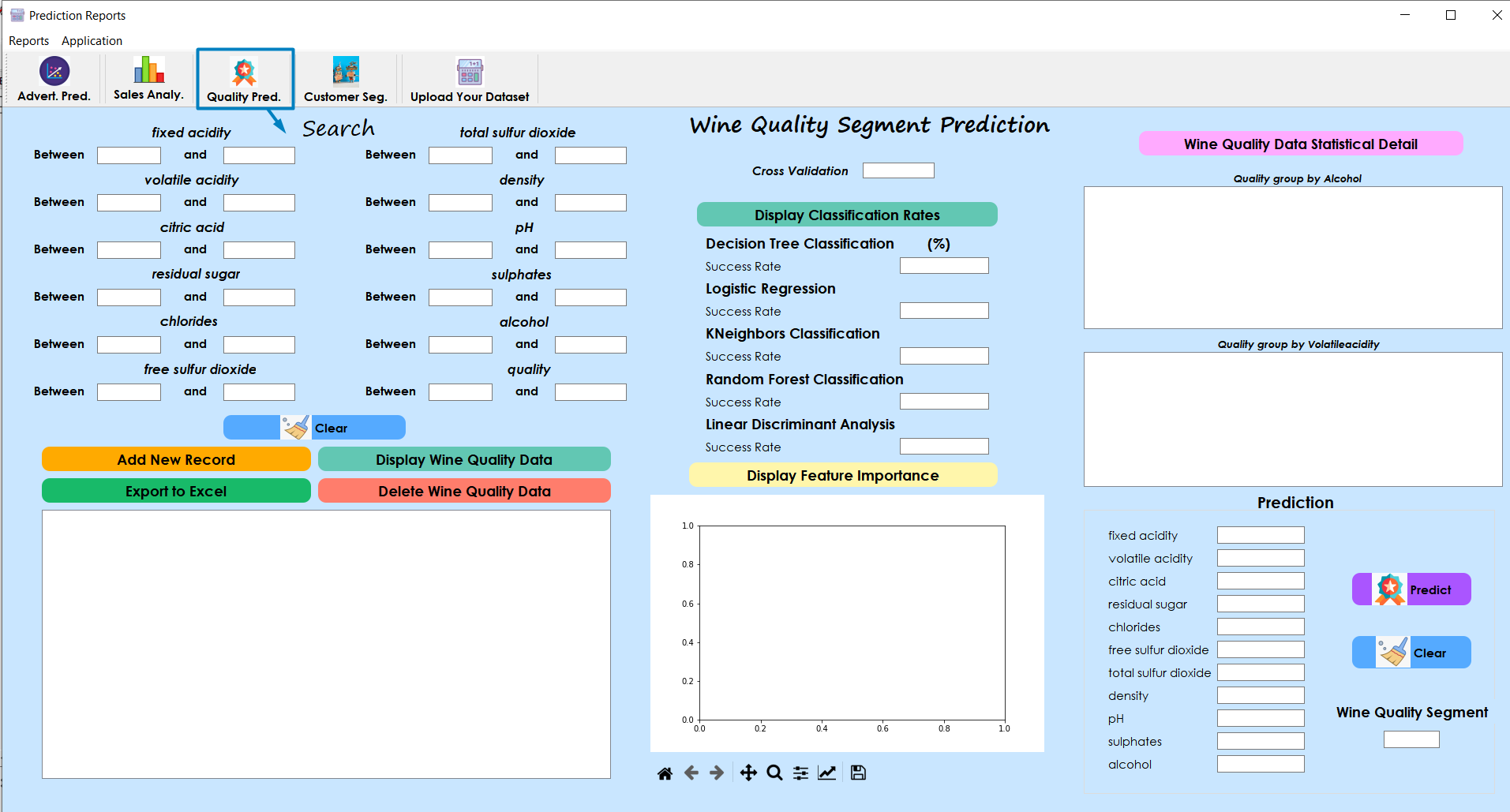
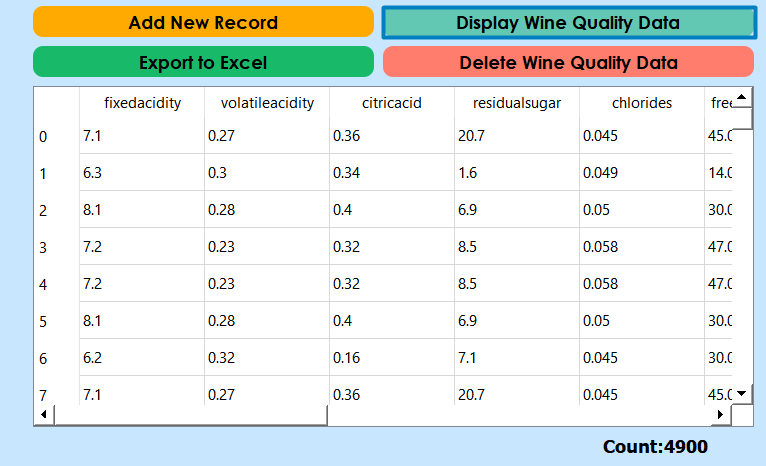


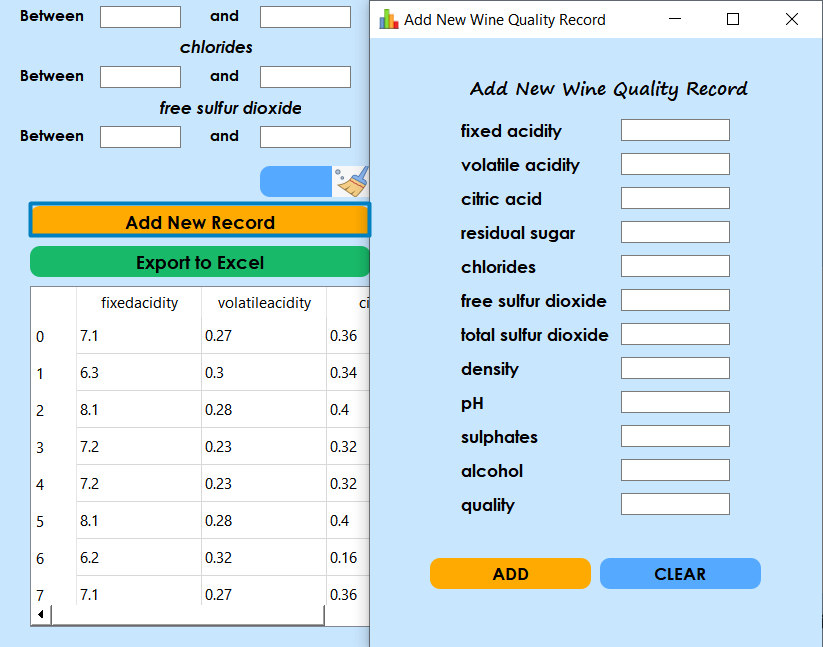
* **“Display Salesplots”** button displays the plot of weekly sales and weekly units sold quantities. It has been visualized by using ecdf (Empirical cumulative distribution function)

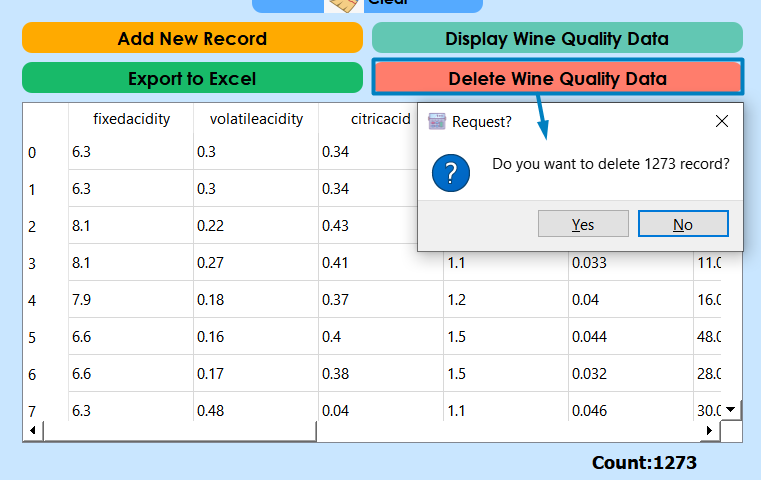
The **ECDF** essentially allows you to plot a feature of your data in order from least to greatest and see the whole feature as if is distributed across the data set.

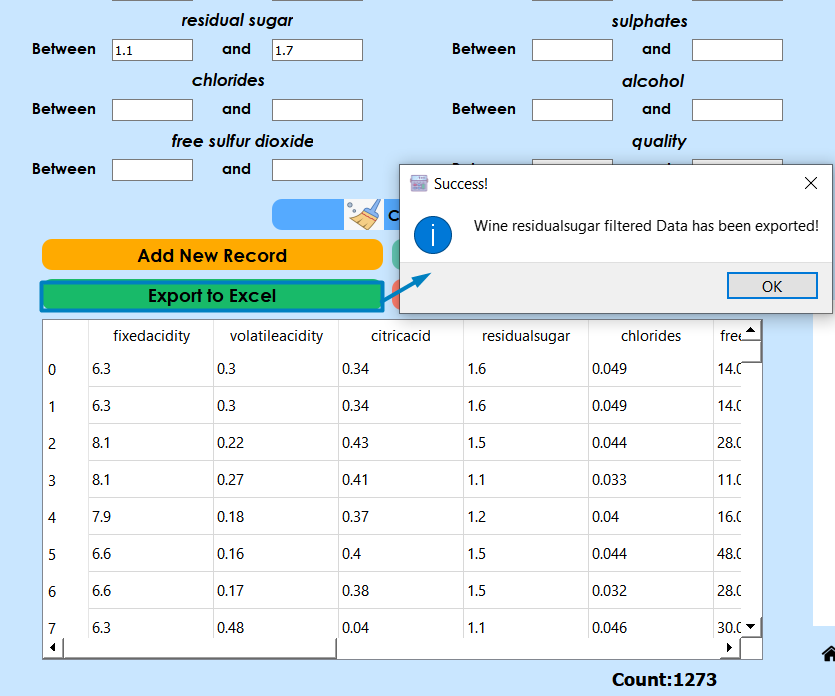
* ****Weekly sales are shown on pie charts by classifying by **product** and **store** percentages.

1. **Wine Quality Segment Prediction**

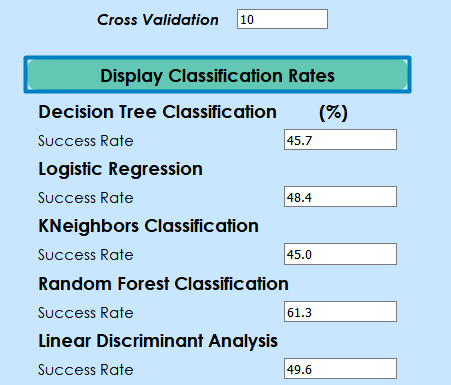
* This page contains the wine quality data of the R&D department in the company.
* All or filtered wine quality data can display with the **“Display Wine Quality Data”** button.
* If the user wants to add a new record, can use the “**Add New Record button**” and fill the new wine quality record form in the incoming window.



* If the user wants to export all or filtered data can use the “**Export to Excel**” button and it exports displayed data to the excel file in the project file directory.
* If the user wants to delete the filtered data (allows only the filtered data to be deleted, not all), you can use the "**Delete Wine Quality Data**" button and the application will once again prompt the user for the number of lines to be deleted.

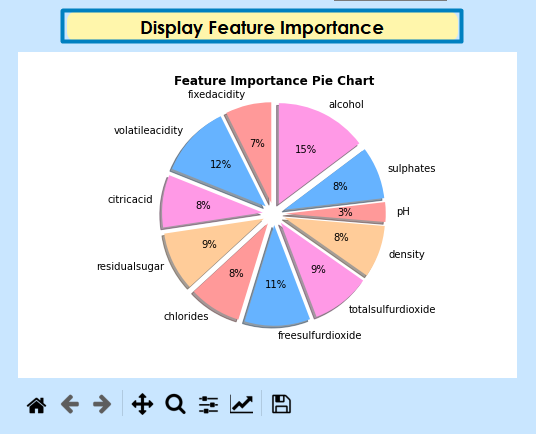


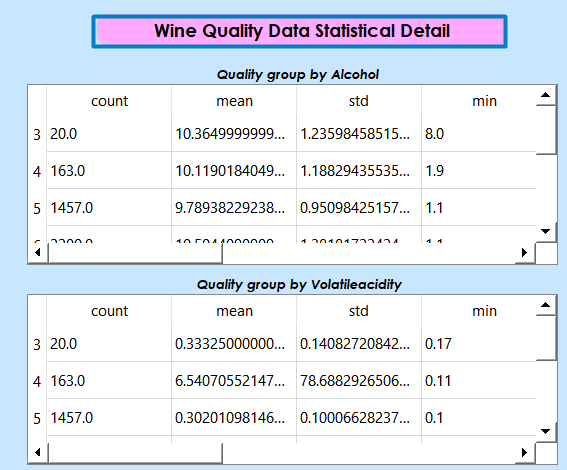
* “**Display Classification Rates**” button calculates the success rates of the supervised algorithms in the below list by using cross-validation parameter.

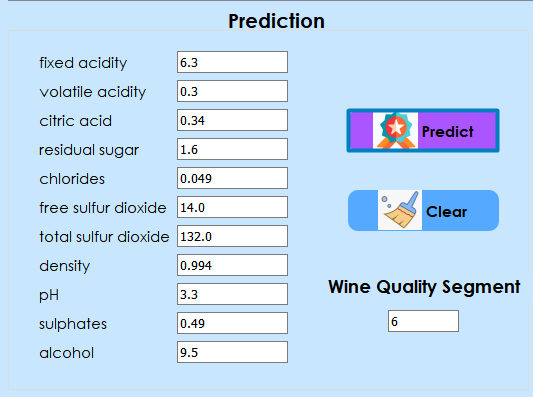


* “**Display Feature Importance”** button displays the Importance percentages of the features by using Random Forest Classification algorithm. Because maximum success rate provides through Random Forest Classification.

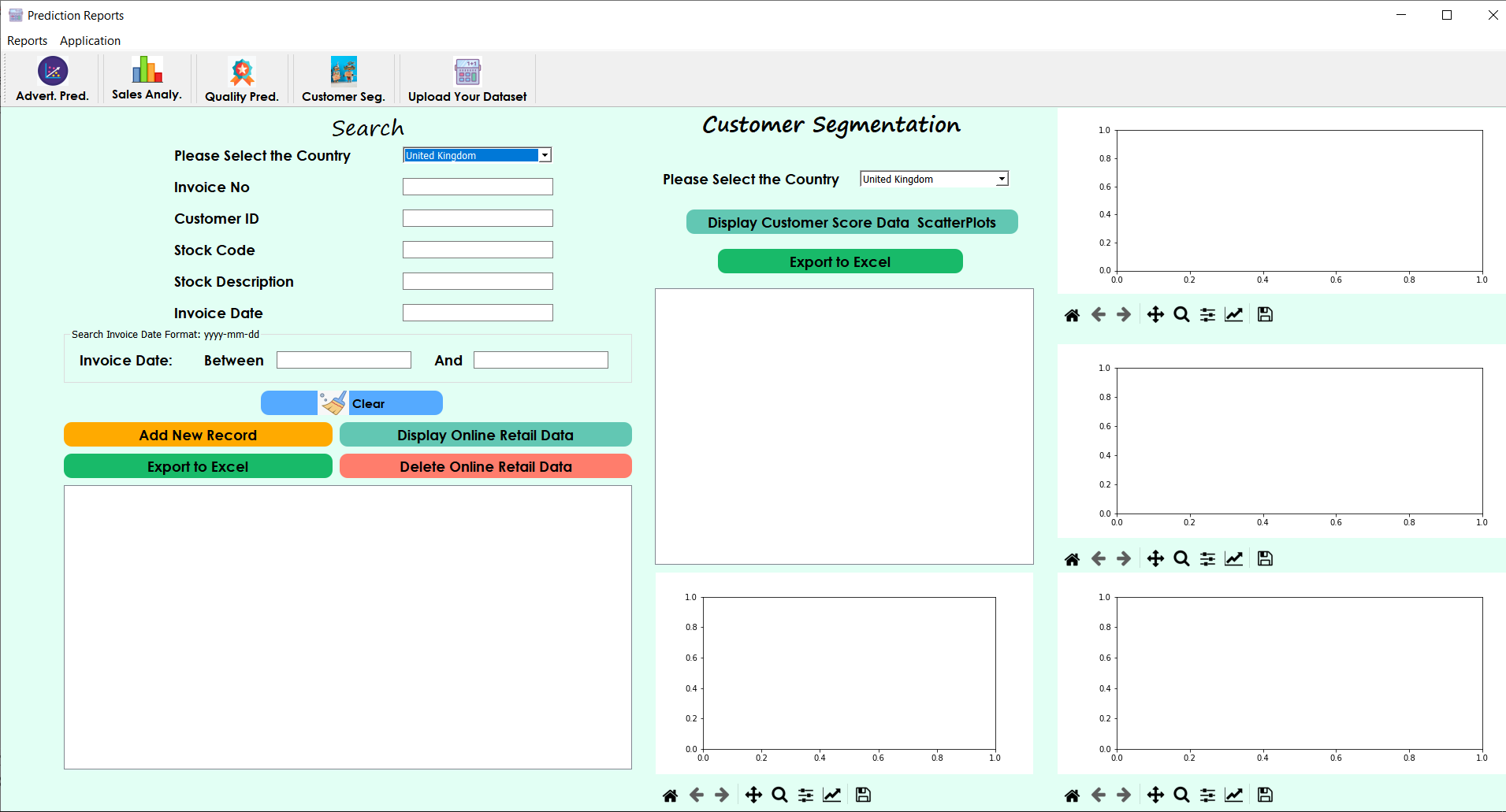
Feature importance refers to a class of techniques for determining scores to input features to a predictive model that indicates the relative importance of each feature when making a prediction.



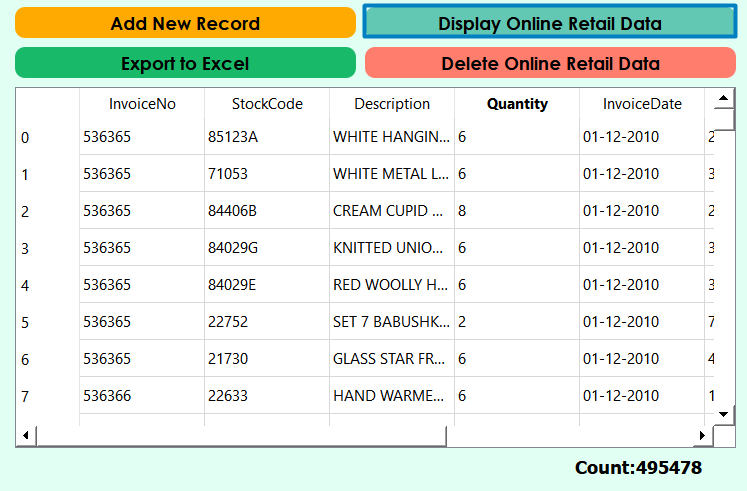
* “**Wine Quality Data Statistical Detail**” button displays the statistical outcomes of the **Alcohol** (Alcohol has **%15** Importance) in the first table and **Volatileacidity (**Volatileacidity has **%12** Importance) in the second table according to the Quality Score.
* **“Predict”** button gives the Wine Quality Segment prediction by using Random Forest Classification which is provided with the experimental values to be entered by the user.



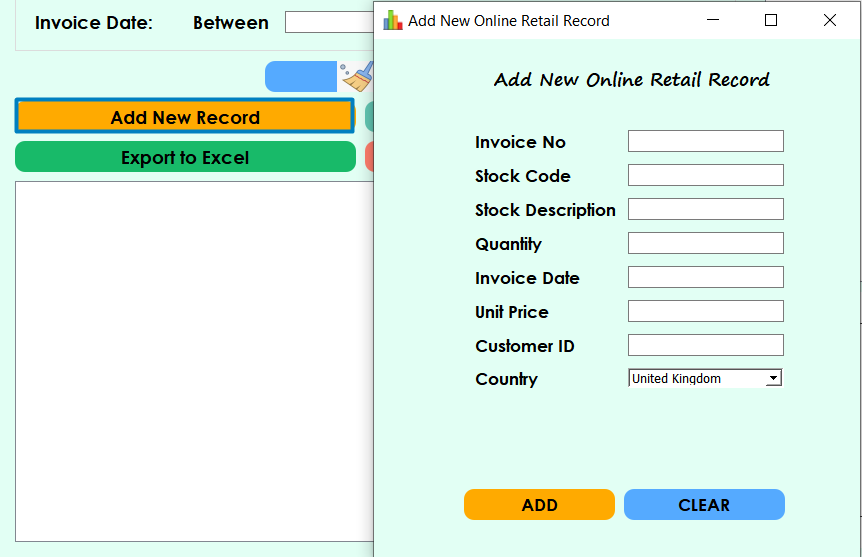
1. **Customer Segmentation**

This page contains the online retail data of the Sales Department in the company and divided customers into the segments.

* All or filtered online retail data can display with the **“Display Online Retail Data”** button.

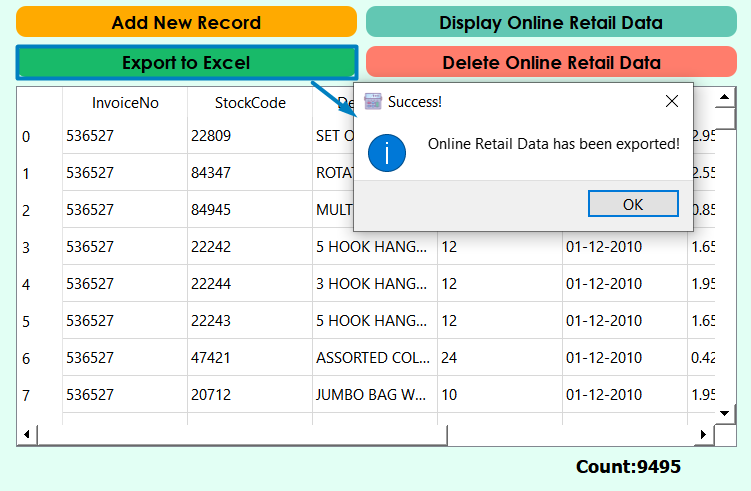


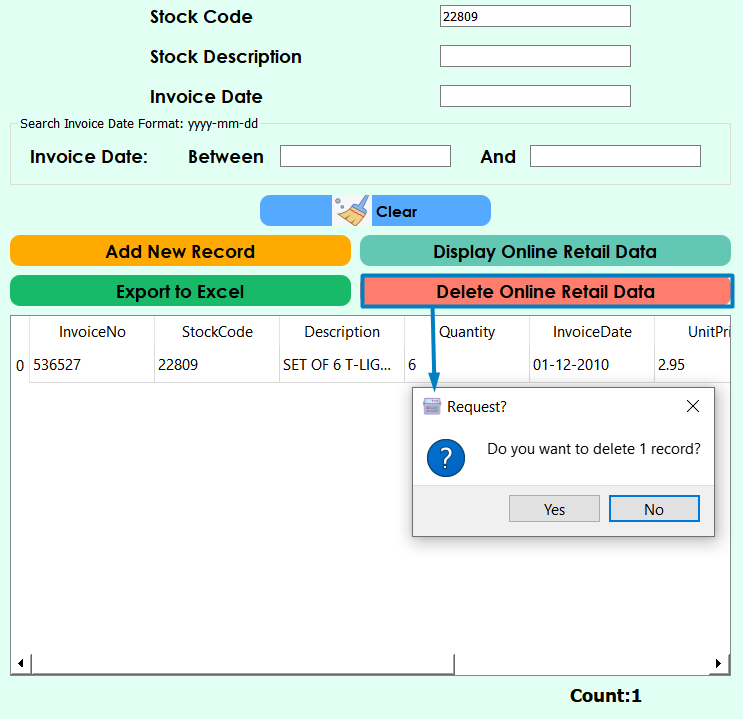
* If the user wants to add a new record, can use the “**Add New Record button**” and fill the new online retail record form in the incoming window.



* If the user wants to export all or filtered data can use the “**Export to Excel**” button the new online retail record form in the incoming window.

and it exports displayed online retail data to the excel file in the project file directory.



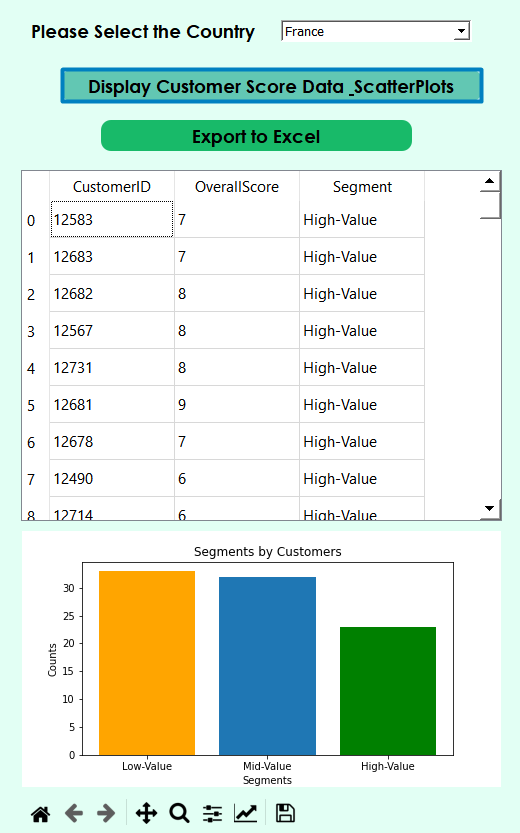
* If the user wants to delete the filtered data (allows only the filtered data to be deleted, not all), you can use the "**Delete Online Retail Data**" button and the application will once again prompt the user for the number of lines to be deleted.
* “**Display Customer Score Data\_ScatterPlots**” button segments customers according to their **recency**, **frequency** and **revenue** values.

The recency value will be calculated based on the most recent purchase date of each customer, Depends on how many days have passed since your last purchase. Unsupervised K-means clustering algorithm will be used to get the recency score of customers.

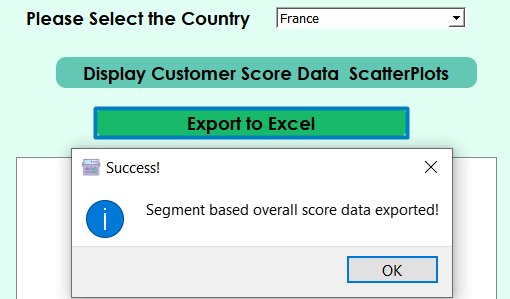
The frequency value is based on how many orders have been taken from customers and order total numbers for each customer.

The revenue value is calculated by multiplying Unit Price and Quantity for each order.

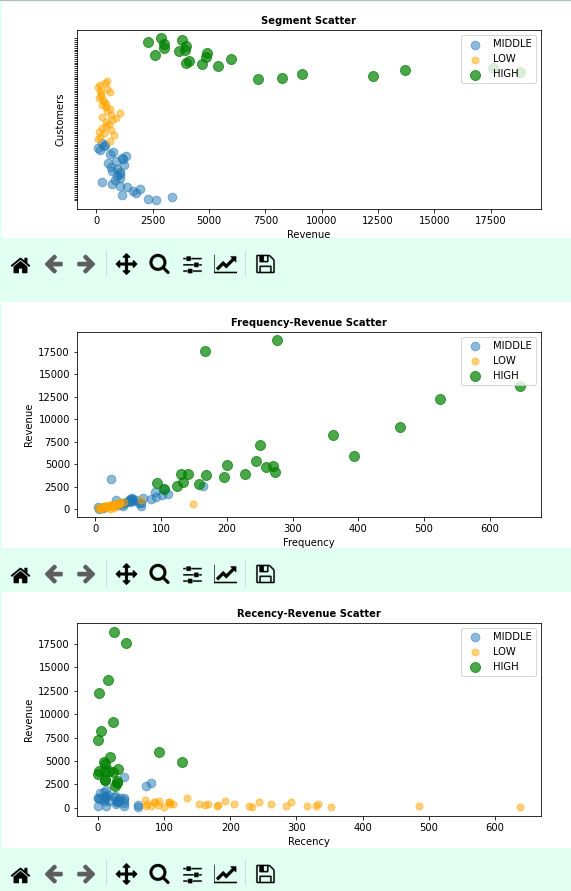
The overall score will be calculated with the average values of Recency, frequency, revenue values for each customer by using unsupervised **Kmeans** algorithm. Hereby, customers will be separated by the segments based on the overall score. Between 0 and 2 will be **Low Value**, between 3 and 4 will be **Mid Value**, greater than 5 will be **High Value**.



* If the user wants to export customer score data can use the “**Export to Excel**” and export displayed data to the excel file in the project file directory.



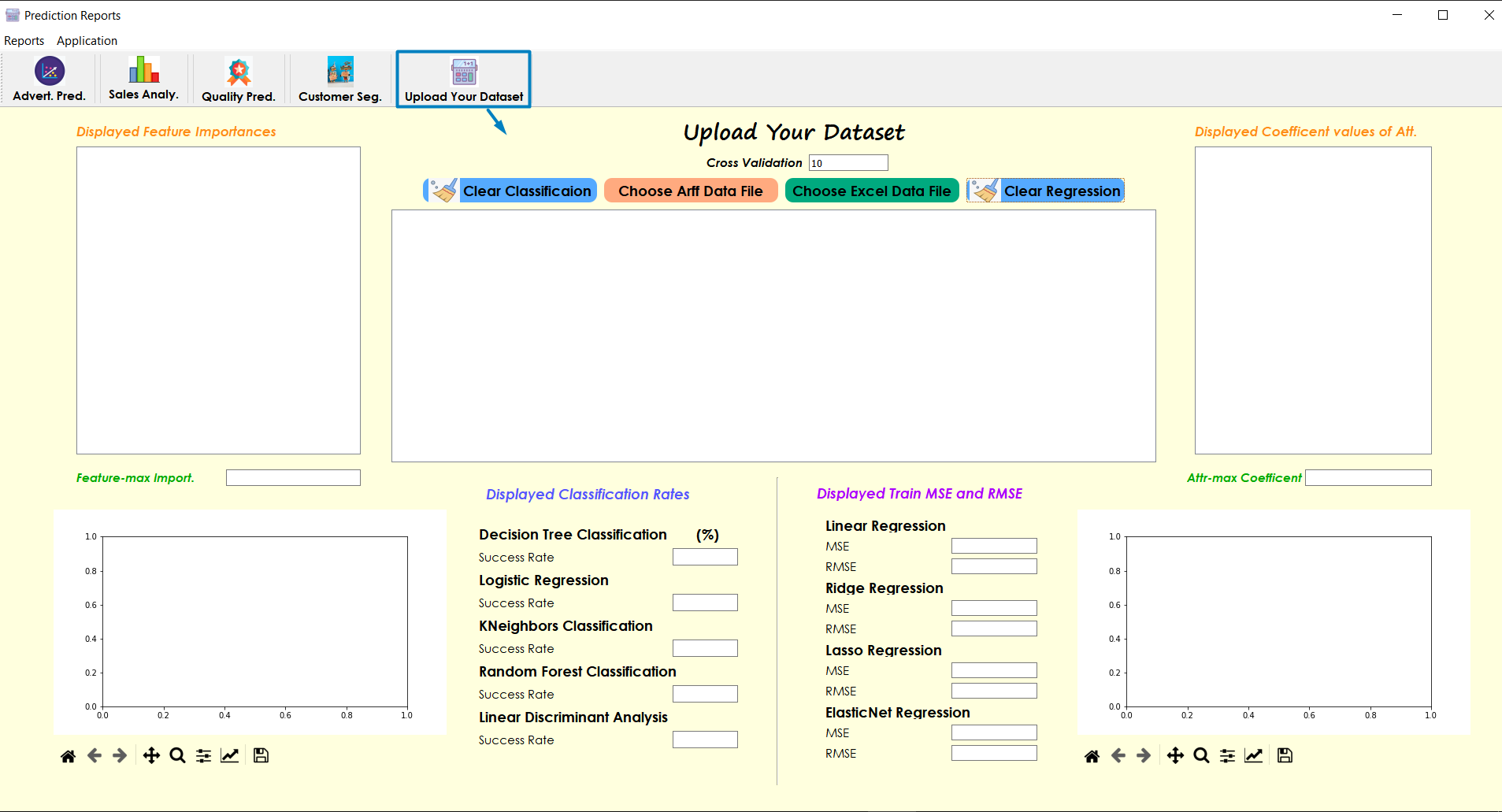
* “**Display Customer Score Data\_ScatterPlots**” displays scatter plots by visualizing the relational relationships between frequency, revenue and recency values.



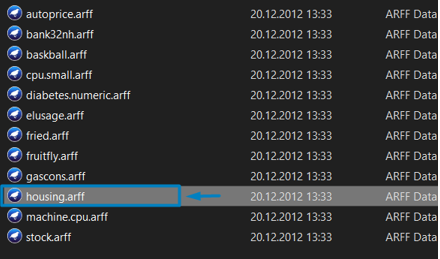
1. **Upload Your Dataset**

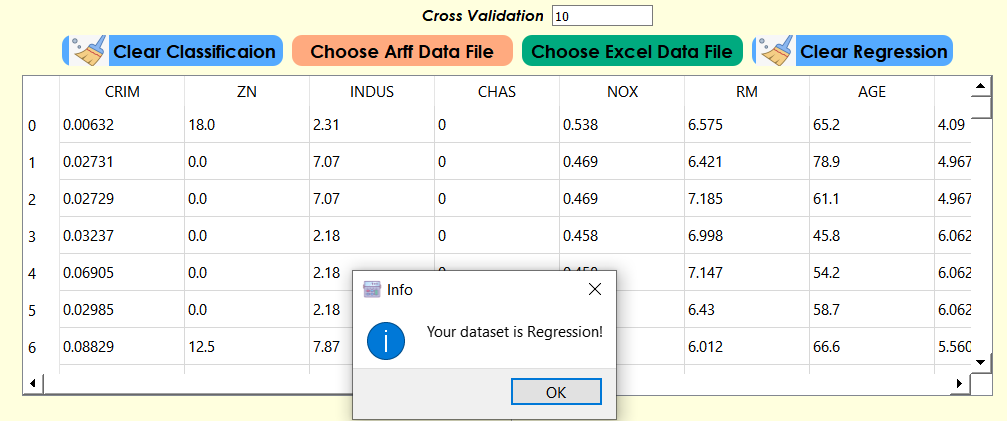
* This page evaluates whether the data to be loaded by the user is regression or classification.

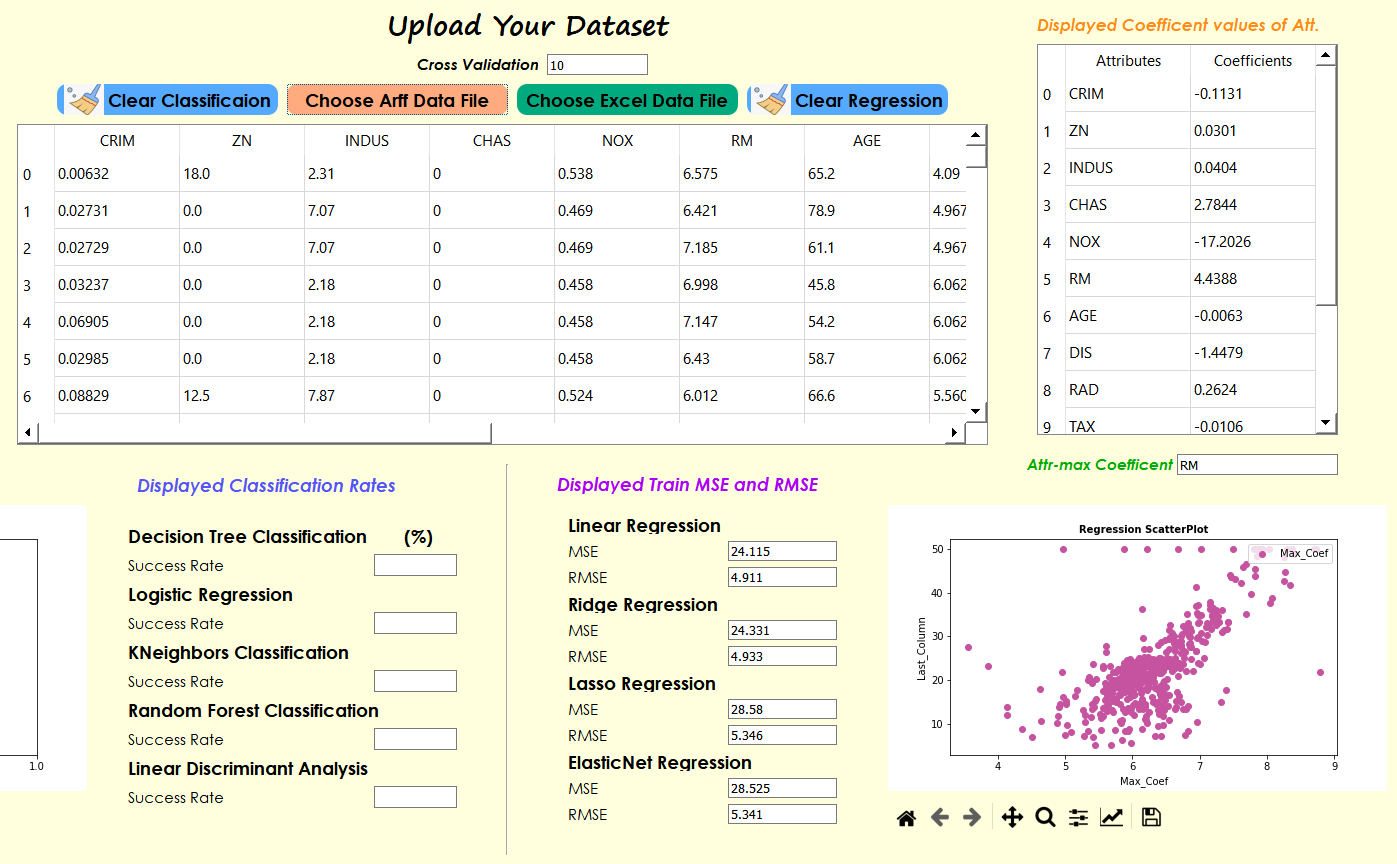
The evaluation function has been tested with over 300 data sets.



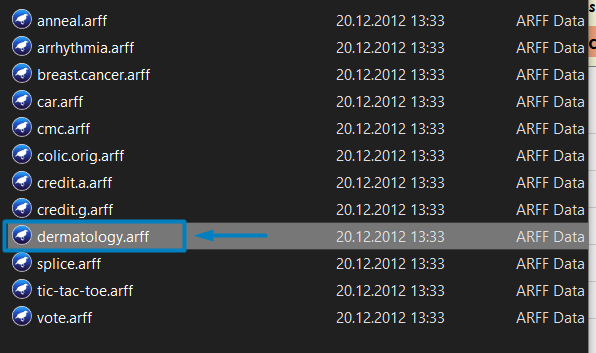
* **“Choose Arff Data File”** buttonallows to find in file explorer the arff data file to be uploded by the user and evaluates the uploded data as regression or classification.

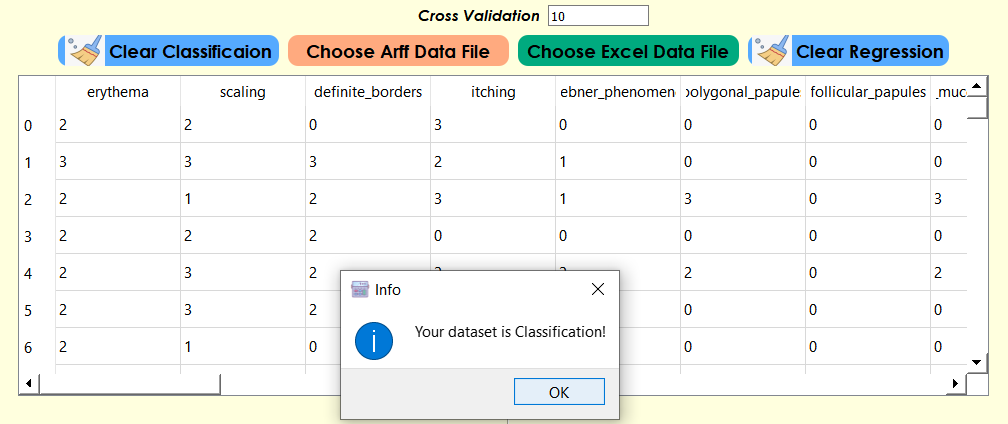


* After selecting the arff data file, the application will evaluate the loaded data and share information on whether it is a regression or a classification.
* After pressing the OK button; displayed Coefficient valus of Attributes, Displayed Train **MSE** and **RMSE** fields will be filled with the calculated values. Regression ScatterPlot will be visualized by denoting **max\_coefficent** on the horizontal axis and **last column** as variable on vertical the horizontal axis.

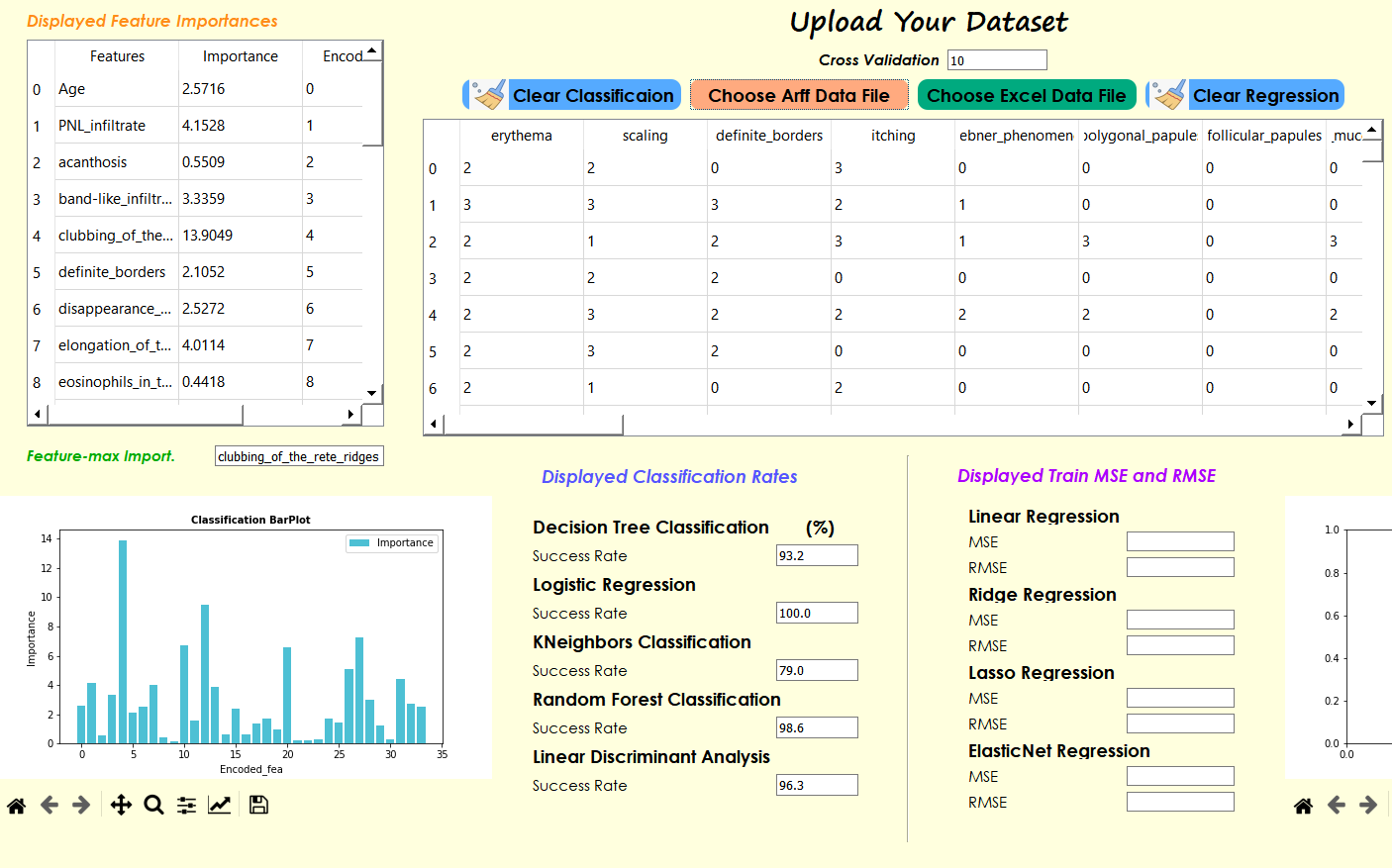


* When the user select an other arff data file by pressing “Choose Arff Data File”, aplication will evaluate the data file once again.





* If the data determined as Classification, , **Displayed Feature Importance,** **Displayed Classification Rates** fields will be filled with the calculated supervised algorithms success rates.
* Classification Bar Plot will be visualized by denoting encoded feature on the horizontal axis and Importance percentage on the vertical axis.



* **“Choose Excel Data File”** buttonallows to find in file explorer the .xls or .xslx data file to be uploded by the user and evaluates the uploded data as regression or classification.



