Overview

* The Super Market Sales Analytics is used to analyze the sales of different products in a super market located in different cities.
* Follow the Data Science Methodology Analyze the data set, investigate and evaluate the result and predict the overall performance.

Methodology

* The 2 classifiers used

1] Random Forest Classifier

2]Association Rule Mining

* Hyper-parameter tuning:

1]For Association Rule mining we have used following parameters

min\_support=0.0045, min\_confidence=0.5, min\_lift=3, min\_length=3

2]For Random Forest Classifier we have used following parameters

n\_estimators=100, max\_depth=None, min\_samples\_split=2,min\_samples\_leaf=1

Dataset

* How many features->11
* Size of the dataset->1000
* Multiple files -> No
* What kind of data – numerical or character -> Both
* Balanced or imbalanced – what is the distribution-> Balanced Dataset (Observed column: Rating)
* Distribution of Training set, validation set, testing set->

Train Data – 75% and Test Data – 25%

* Missing data and Preprocessing challenges-> As per observation there is no missing data in the data set

Results

* Table for the evaluation metric for each ML technique used

The evaluation matrix used for classification technique is classification report and confusion matrix.

* Plot of the curves

The output of matrix is shown in the python file.

Feature Engineering Techniques

* Features removed-> invoice\_id,tax\_,payment\_type,time,date
* Feature creation->Year, Month, Day
* Class imbalance treatment-> We have balanced data set