Automated Credit Card Approval

-Laisha Wadhwa

Predicting if credit card request will be approved for a given applicant given certain attributes.

Motivation

1. Banking industries receive so many applications for credit card request.
2. Going through each request manually can be very time consuming,also prone to human errors.
3. However, using the historical data to build a model which can shortlist the candidates for approval is what a bank would need.

Libraries Used

1. Sklearn: Model prototyping
2. Pandas: Data analysis
3. Numpy: For mathematical computations
4. seaborn
5. Matplotlib
6. Keras

To install the above packages using pip, use command

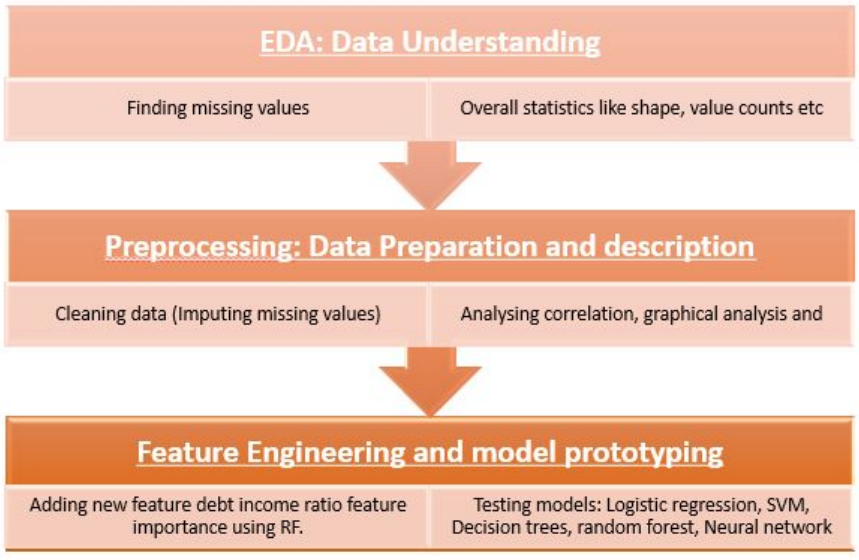
pip install package-name

Usage

Run the jupyter notebook CreditCardApproval_Amex.ipynb

Code Walkthrough :

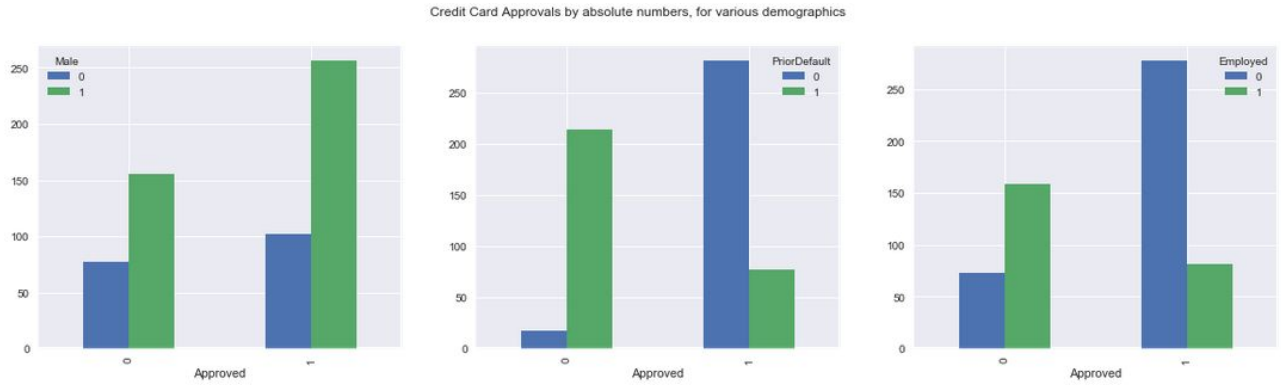
The following flowchart shows the series of steps followed while analysing and model prototyping.



1. **Data Understanding:**
    a. The dataset is loaded in dataframe.
    b. df.describe and similar functions are used to understand the distribution of data.
    c. Missing values are found using isna.sum().

2. **Data Preparation**
    a. **Data Cleaning**:
        - Imputing with mode for categorical values.
        - Imputing with mean for numerical variables.

    b. **Data description and distribution:**
        A. Correlation Matrix for all the features.
        B. DistPlot for analysing the distribution of univariate variables
        C. Scatter plot for 'Age', 'Income', 'CreditScore', 'Debt', 'YearsEmployed' for analysing the effect of each of the features on other features
        D. Barplots and pd.groupby are used to analyse the effect of features like education level and prior default on the predicting variable Approved.



3. Feature Engineering

New feature Income to debt ratio is added which ranks higher in feature importance than already existing features.

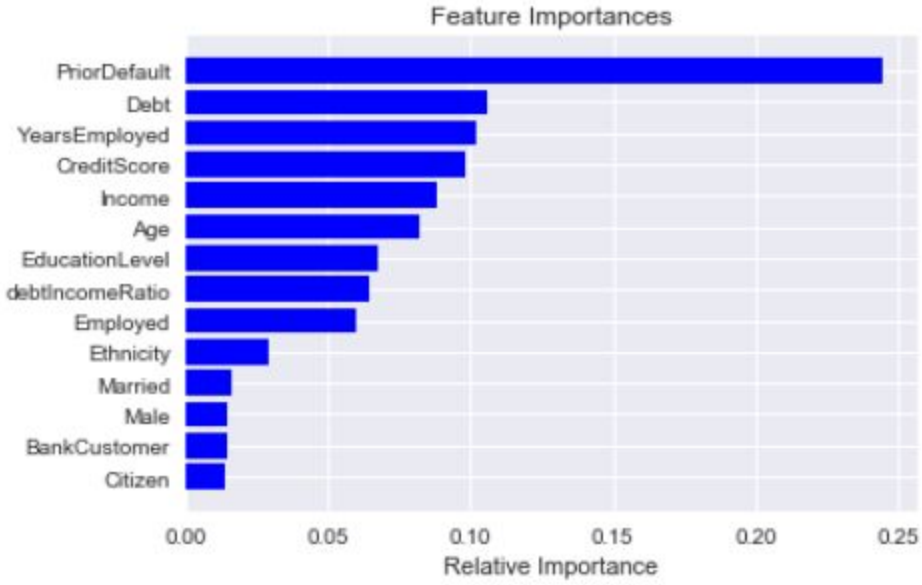
The following image shows the feature importance ranked from highest to lowest.

Feature Engineering: Adding new feature (Debt to income Ratio)

In [247]:
```
CC_data['debtIncomeRatio'] = CC_data['Debt']/ CC_data['Income']

#CC_temp = CC_data.copy()
CC_data = CC_data.replace([np.inf, -np.inf], np.nan)
CC_data.fillna(CC_data.mean(), inplace=True)

plotDistPlot(CC_data['debtIncomeRatio'])
```

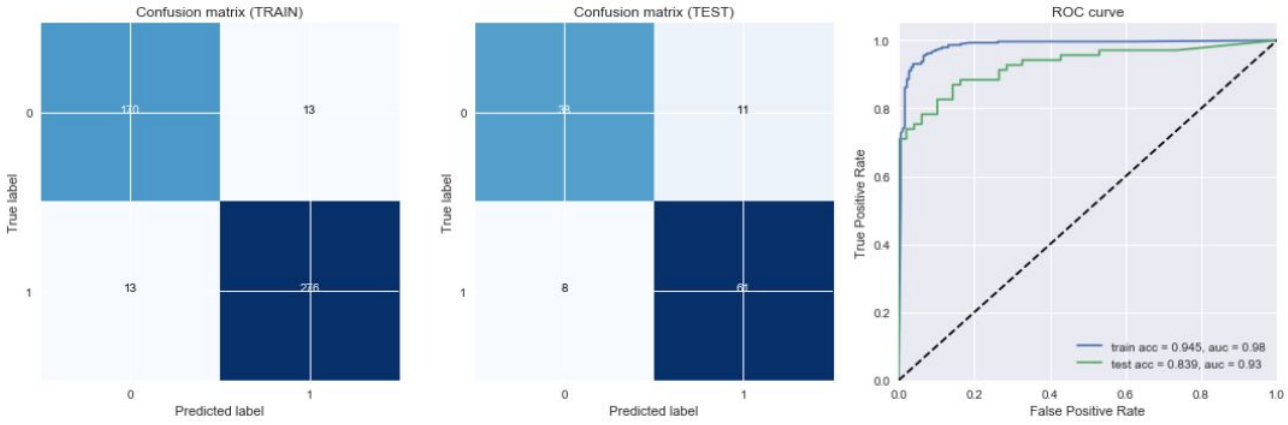


---

4. Model Prototyping:

For every model Confusion matrix and AUC curve are plotted to understand the results and performance.
Sklearn libraries are used to build models.
The jupyter notebook gives an elaborate study of the performance along with graphs like these:

This image shows the confusion matrix for train and test for XGbost along with AUC curve



Results:

Model	Accuracy Score(%)
Logistic regression	84.74
Support Vector Machine (SVM)	83.9%
Boosted Decision trees (XGBoost)	88.13%
Random Forest	88.983
Neural Network	88.984%

Summary

In this project, I have tried to find out the factors that are most important for getting an approval for the credit card by leveraging data analysis and Machine Learning.
Though the accuracy achieved is 88.984%, I also tried to check if we can improve the performance further and tried grid searchCV and neural networks.
However, 89% is the best I could get from this data using all the models.