A

Capstone Project

in

Data Science

BANK LOAN PAYOFF

By:
Ganesh Bhandari

@

January 2021

1. Problem, background and data:

A manager at a ABC bank is disturbed because of more and more loans are being collected. The manager would really appreciate it if someone could predict for them which loan is going to get collected.

I would like to help the manager by solving the problem. I talked with the manager and he sent me a data containing records of customers at the bank. The data has 8 features. The description of the features of the data is as follows:

Field	Description
Loan_status	Whether a loan is paid off on in collection
Principal	Basic principal loan amount
Terms	Origination terms which can be weekly (7 days), biweekly, and monthly payoff schedule
Effective_date	When the loan got originated and took effects
Due_date	Since it's one-time payoff schedule, each loan has one single due date
Age	Age of applicant
Education	Education of applicant
Gender	The gender of applicant

The original data, provided by the Manager contains data related to the record of 346 customers at the bank. After cleaning the data, it is of shape (336, 8) and our target variable is 'loan status'.

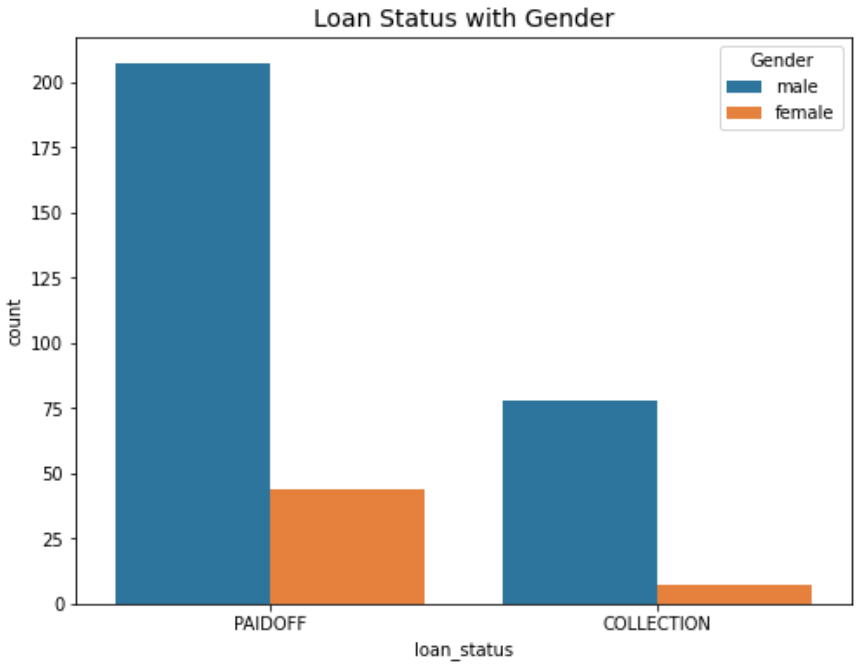
2. Methods used:

This problem is a classification problem. So to solve this problem I have created a model based on classification algorithms . In this project, I used classification algorithms like Logistic Regression, Naive Bayes, KNN, SVM, Decision Tree or Random Forest and chose the best model based on their accuracy performance. I used the best selected model to predict which loan is going to get collected. The steps that I used to get the result are as follows:

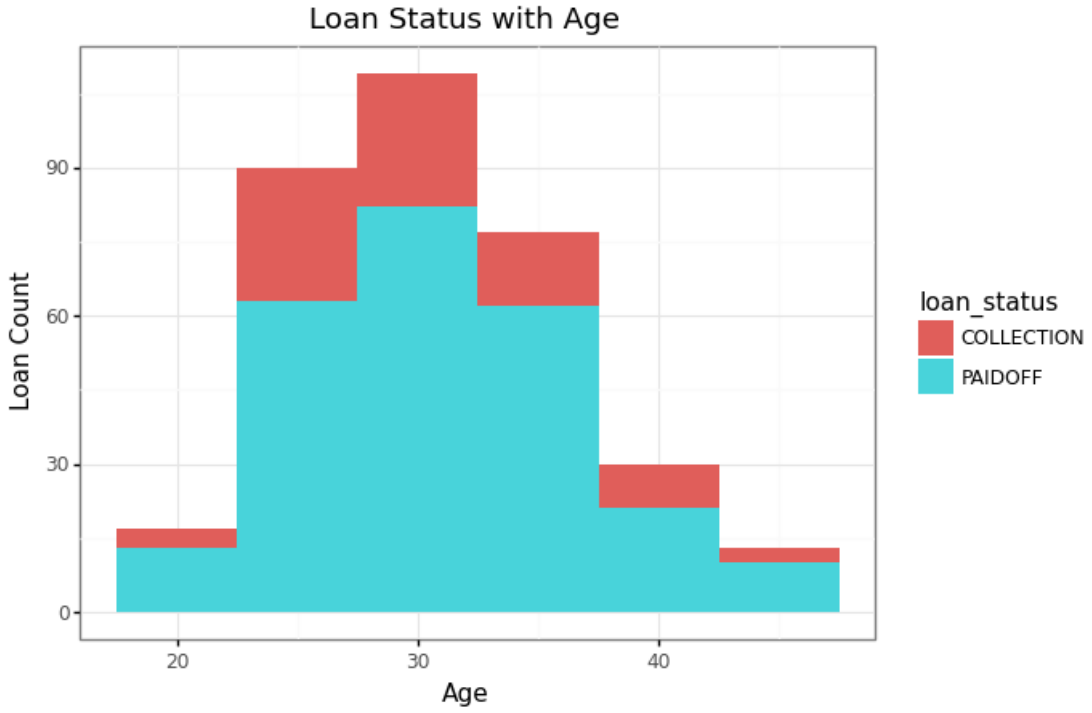
- Loading data
- Cleaning and organizing data
- Performing Exploratory Data Analysis (EDA) to understand the data and important features.
- Creating model and training the model
- Using suitable metric to check the performance of the model
- Deploying the model to get the results

3. Current Situation:

The data shows that the current situation of the bank is not good because of the number of the collected loan. I found 75% people have paid off the loan on time while 25% have gone into collection. Also there is ainteresting scenario gender wise. 86 % of female pay their loans while only 73 % of males pay their loan.



Another interesting finding is that people of age around 25-35 take more loan and have more collection as well.



4. Model:

In this project, I used classification algorithms like Logistic Regression, Naive Bayes, KNN, SVM, Decision Tree or Random Forest to create a suitable model. The performance report of these model on the data is as follows:

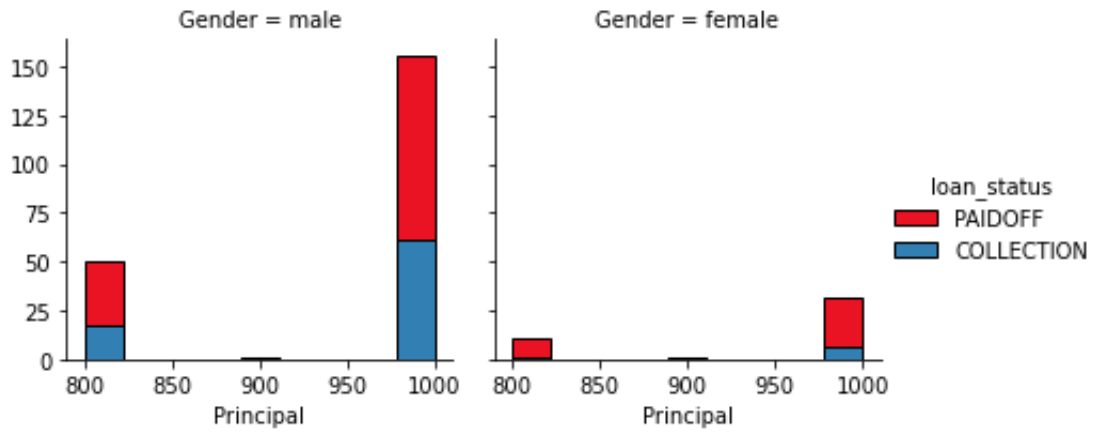
Algorithm	Accuracy Score	F1-score	LogLoss
Logistic Regression	0.76	0.71	0.51
Naive Bayes	0.74	0.63	8.95
KNN	0.63	0.62	1.74
SVM	0.72	0.67	NA
Decision Tree	0.78	0.77	2.3
Random Forest	0.72	0.73	1.85

The report table shows that, among all these algorithm, the Logistic Regression model 'LR' is more consistent and works better than the others. So, I chose the Logistic Regression model as the final model for my project.

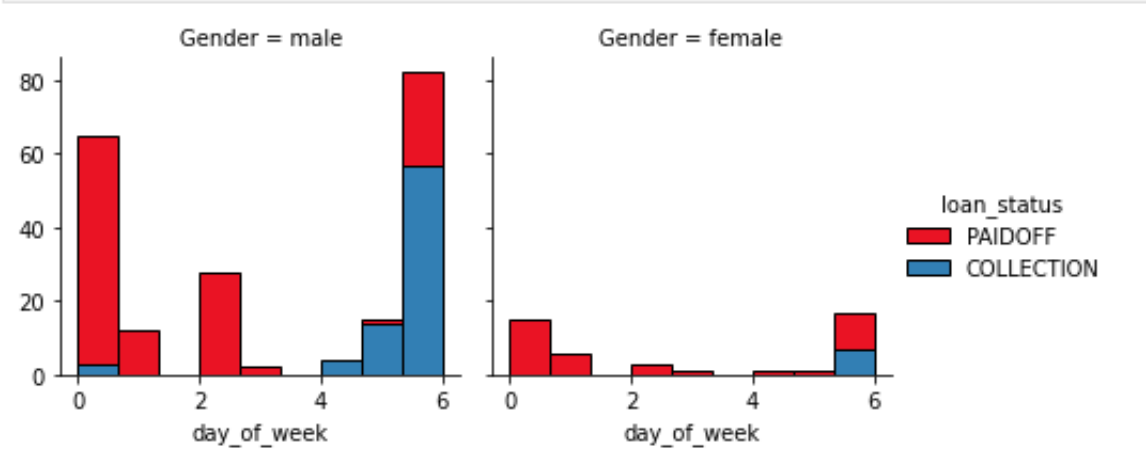
5. Findings:

I tried to do my best research on the data. After performing exploratory data analysis and deploying the model, I have some interesting findings on the data :

- More than 25% of the loan have gone into collection.
- Loan taken by males is in more risk than the loan taken by females.



- Loan of bigger amount has more chance of being collection especially for men.
- People of age around 25-35 have more collected loan.
- People who take loan at the end of the week mainly on Thursday and Friday do not pay it off.



6. Recommendations:

Based on the findings obtained during data analysis process and deploy of the model, I have some recommendations for the bank:

- Emphasize on providing loans for female customers than male.
- Do not provide loan of bigger amount for male unless good guarantee of the loan.
- Reduce the number loans for the people of age around 25-35.
- Reduce the number of loans providing at the end of the week mainly on Thursday and Friday.

7. Conclusion:

Apply all the credit check methods strictly, collect correct data and do good research about the costumers before providing any loans. Also apply the model for each customer's information to predict whether the loan for that customer will be paid off or collected.

8. Acknowledgments:

I am very grateful to my mentor for his valuable suggestions while completing this project. Also I am thankful to the bank authority for providing valuable information.