



NAME OF THE PROJECT

Micro Credit Defaulter Project

Submitted by:

Simran Panigrahi

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

INTRODUCTION

The purpose of this study is to provide a comprehensive research and to develop a model to predict the loan defaults. This kind of models becomes inevitable as the issue of bad loans are very much critical in the financial sector especially in micro financing banks of various underdeveloped and developed countries. To cope up with this problem a comprehensive literature review was done to study the significant factor that leads to this issue. Moreover, these reviewed studies were critically focused towards applying data mining techniques for the prediction and classification of the loan defaults. While in the experimentation phase, three different data mining techniques were applied for the proposed model and their performances were evaluated on various parameters. Based on these parameters, the best method was selected, explained and suggested because of its significant characteristics regarding the prediction of the loan defaults in the financial sector.

Without a doubt, financial lending services hold a great amount of significance for any individual, business or enterprise. As such services are required by an individual or a business to achieve or accomplish their goals and to compete with the giants of their fields. Financial loans are a major part of the primary source of capital not only in the emerging economies but also in the developed capital markets by both individuals and enterprises. As the lending growth by the financial firms and the banks are considered as the key factor for inflation level and interest rate of any country which drives its economic growth and depicts its economic condition. According to the mission statement of the study in the role of financial services, the economic growth of the real economy is the primary role of the financial firms. With such great importance and benefits of financial lending come some major issues and bottleneck problems. The most common and substantial issue in the domain of financial lending is the fair and successful lending of loans while keeping the ratio of loan defaulters to the least minimal value. In the financial lending, the risk of loan defaulters can never be neutralized but can be minimized. According to the study, during the year of 2017 in India, only the bad loans crossed the threshold of about 207 billion dollars equalling the percentage of about 9.6 for loan *Author to whom correspondence should be addressed. Defaulter ratio. While the greatest number of loan defaulter cases were registered in Italy making the total of 16.4% for the loan defaulters to cope up with the issue of high ratio loan defaulters' lot of work has been done.

In the more recent years, different researchers have also employed different data mining techniques for the loan defaulter predictions. The main objective of this research is to study the previous work regarding the field of financial lending and evaluate the different approaches which have been used for the prediction of loan defaulters. Aiming to the objective, the purpose of this study is to employ various data mining techniques on the collected dataset and to predict the successful financial lending. Finally, the proposed predicted model will be evaluated by benchmarking its performance against other modelling techniques.

Problem Statement

Loan default is one of the leading causes of MFI failures in many. The general business problem was that some MFI leaders failed to achieve MFI sustainability because the loan default rate exceeded the sustainability thresholds. The 3 specific business problems was that some MFI leaders lacked strategies to reduce loan default in the BOP market. Implicit guarantee of ready access to future loans if present loans are repaid fully and promptly. Microfinance is seen as a catalyst for poverty alleviation, delivered in innovative and sustainable ways to assist the underserved poor.

Conceptual Background of the Domain Problem

For banks, it is always an interesting and challenging problem to predict how likely a client is going to default the loan when they only have a handful of information. In the modern era, the data science teams in the banks build predictive models using machine learning. The datasets used by them are most likely to be proprietary and are usually collected internally through their daily businesses. In other words, there are not many real-world datasets that we can use if we want to work on such financial projects.

Review of Literature

Banks loan money to companies in exchange for the promise of repayment. Some will default on the loans, being unable to repay them for some reason. The bank maintains insurance to reduce their risk of loss in the event of default. The insured amount may cover all or just some part of the loan amount.

For this assignment, the micro finance wants to predict default on their loans based on their financial information. Dataset provided consists of whether the user paid back the credit amount within 5 days of issuing the loan, average payback time in 30 days, average payback time in 90 days, Total amount of loans taken by user in 90 days, total amount of loans taken by user in 30 days, maximum amount of loan taken by user in 30 days and 90 days, number of loans taken by user in 30 days and 90 days.

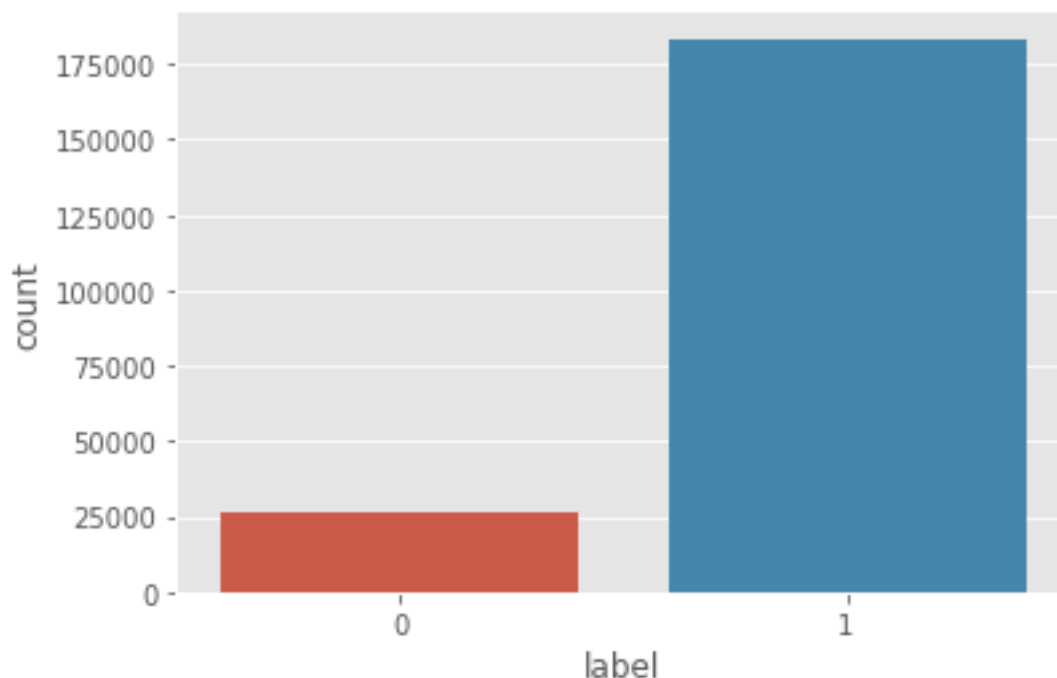
Objective behind the Project

To predict regarding the defaulter of the loan, I tried machine learning algorithms: Logistic Regression. The instruction of this assignment was to use accuracy as the evaluation metric. However, the precision would be essential in this scenario as we would like to minimize the potential of loan defaults. Predicting default rates is a significant part of money-lending because lenders must predict whether giving out a loan will result in profit or loss. Normally, loans are profitable because of interest, but sometimes a borrower will default, which is both a betrayal of the moneylender's trust and a hazard to the moneylender's business. Thus, it is important that the lender is able to gauge the likelihood of a borrower defaulting before making a loan to him/her. Given the high number of factors that might affect borrower default rate, it may be infeasible to come up with good estimates heuristically or by hand. The goal of this project is to explore whether or not we can employ statistical and machine learning models to better predict the risk of borrower default. By analyzing variables that describe loans and the financial situations of their borrowers, we may determine key relationships between default rates and a few other variables. Along the way, we will look into key relationships between loan default chances, loan characteristics, and buyer behaviours.

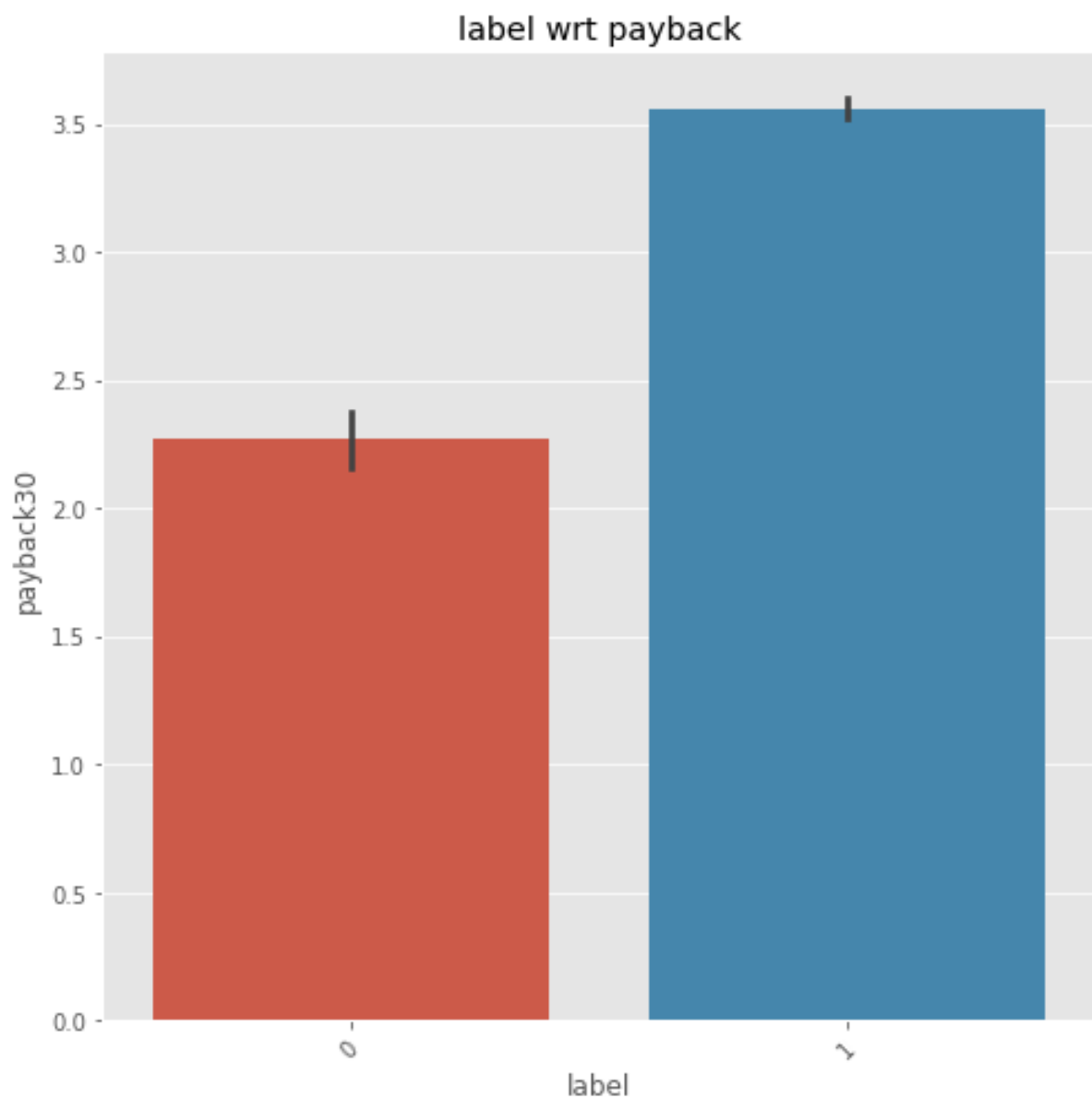
Analytical Problem Framing

We will first look into the distributions of and relationship between some of the characteristics of loan and the borrower of loan. This will help us determine which predictor variables may have interesting patterns and where we should be concerned about multicollinearity, which is when the model breaks down because multiple variables are too correlated.

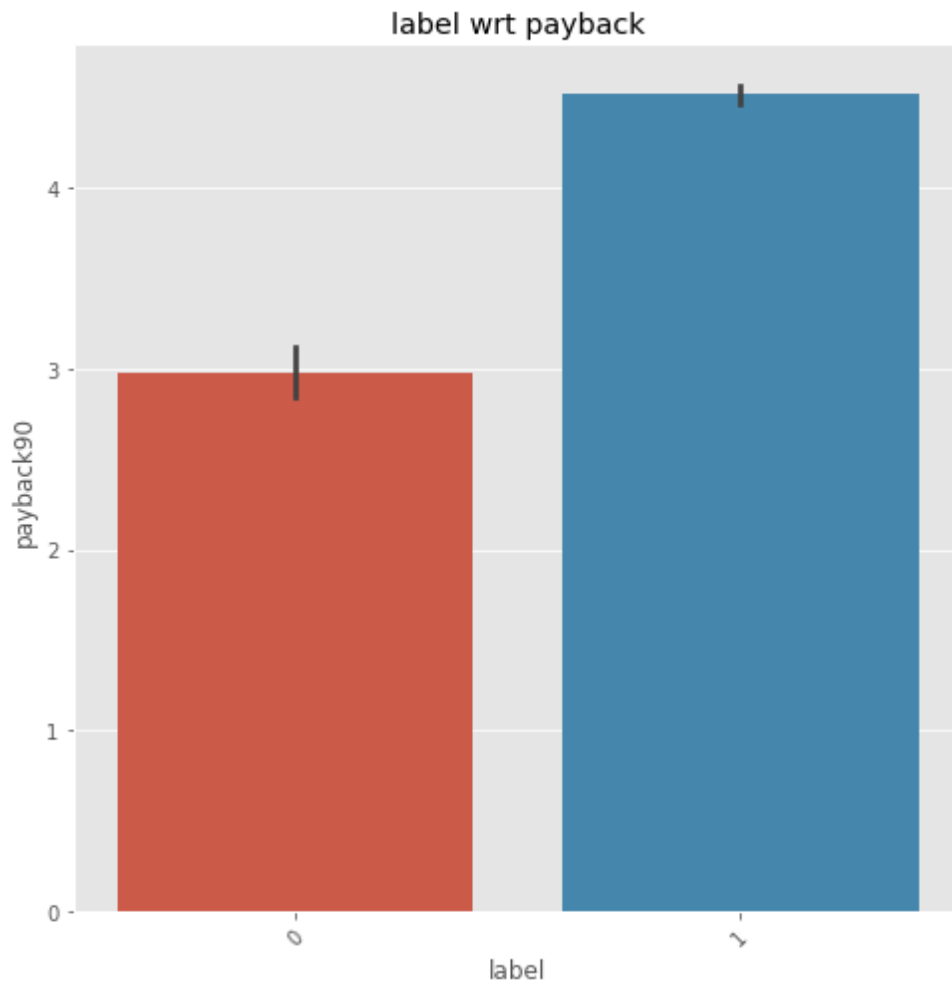
- **Label:** This flag indicates regarding whether the user paid back within 5 days or not of issuing the loan. From the above analysis we can observe that most users paid back the loan within time limit of 5 days. Only few people were not able to pay loan within the time limit as is mentioned 0 represents defaulter user and 1 represents non defaulter users. And it is a good observation as users are able to pay the loan.



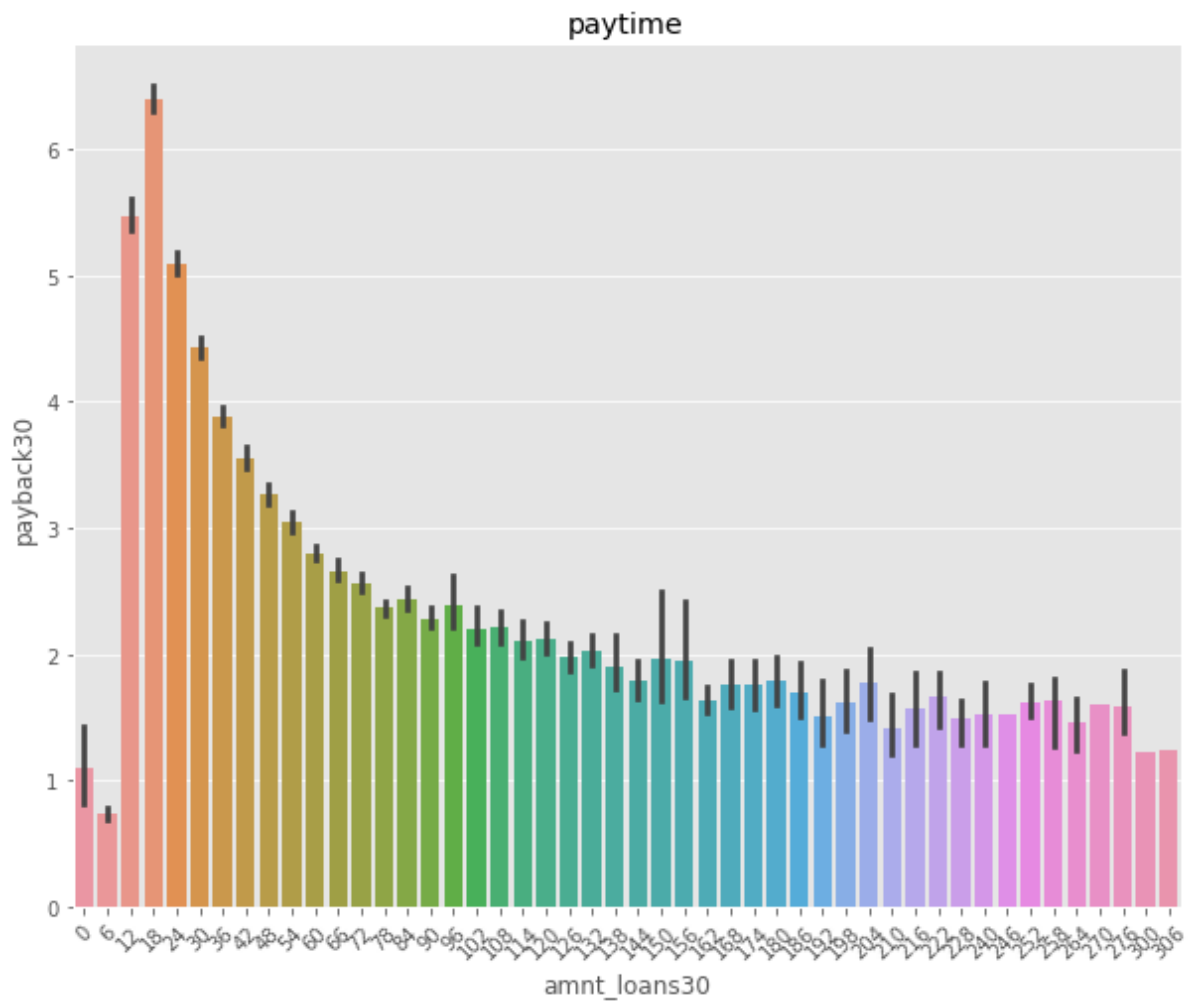
- Relationship between payback period 30 days and label i.e. user paid back within 5 days or not: From the picture plotting it represents the relationship between average payback time within 30 days and user paid within 5 days of issuing the loan. From the below observation we can observe that most users were successful to pay the loan .



- Relationship between payback period 90 days and label i.e. user paid back within 5 days or not: From the picture plotting it represents the relationship between average payback time within 90 days and user paid within 5 days of issuing the loan. From there we can observe that most users were successful to pay the loan .



- Relationship between total amount of loan taken by user in 30 days and average payback time by the user over last 20 days: When the loan is issued the main focus is on regarding the payback time .From the below observation we observed that people with total amount taken is less has paid the amount very fast in compared to higher amount of loan taken by the user.



- Relationship between total amount of loan taken by user in 90 days and average payback time by the user over last 90 days: When the loan is issued the main focus is on regarding the payback time .From the below observation we observed that people with total amount taken is less has paid the amount very fast in compared to higher amount of loan taken by the user.

