Smart Lender - Applicant Credibility Prediction for Loan Approval Using IBM Watson

20BCE1385 Vaibhav Thalanki
19BCE10079 Disha Rajesh Ghosh
20BCE10529 Jayesh Mehta
19BCE10185 Naman Jain

1. INTRODUCTION

1.1 Overview

One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. "As we know credit risk evaluation is very crucial, there are a variety of techniques used for risk level calculation. In addition, credit risk is one of the main functions of the banking community.

The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement. This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

We will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

1.2 Purpose

All the technical aspects that students would get if they complete this project.

(i) Knowledge of Machine Learning Algorithms.

(ii) Knowledge of Python Language with Machine Learning.

(iii) You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.

(iv) You will be able to know how to pre-process/clean the data using different data pre-processing techniques.

(v) Applying different algorithms according to the dataset and based on visualization.

(vi) Real-Time Analysis of Project

(vii) Building ease of User Interface (UI)

(viii) Navigation of ideas towards other projects(creativeness)

(ix) Knowledge of building ML models.

(x) How to build web applications using the Flask framework.

2. LITERATURE SURVEY

2.1 Existing problem

Amira Kamil Ibrahim Hassan, Ajith Abraham (2008) uses a prediction model which is constructed using three different training algorithms to train a supervised two-layer feedforward network. The results show that the training algorithm improves the design of loan default prediction model.

Angelini (2008) used a neural network with standard topology and a feed-forward neural network with ad hoc connections. Neural network can be used for prediction model. This paper shows that the above two models give optimum results with less error.

Ngai (2009) uses the classification model for predicting the future behaviour of costumers in CRM. In CRM domain, the mostly used model is neural network. He recognized eighty seven articles associated to data mining applications and techniques between 2000 and 2006.

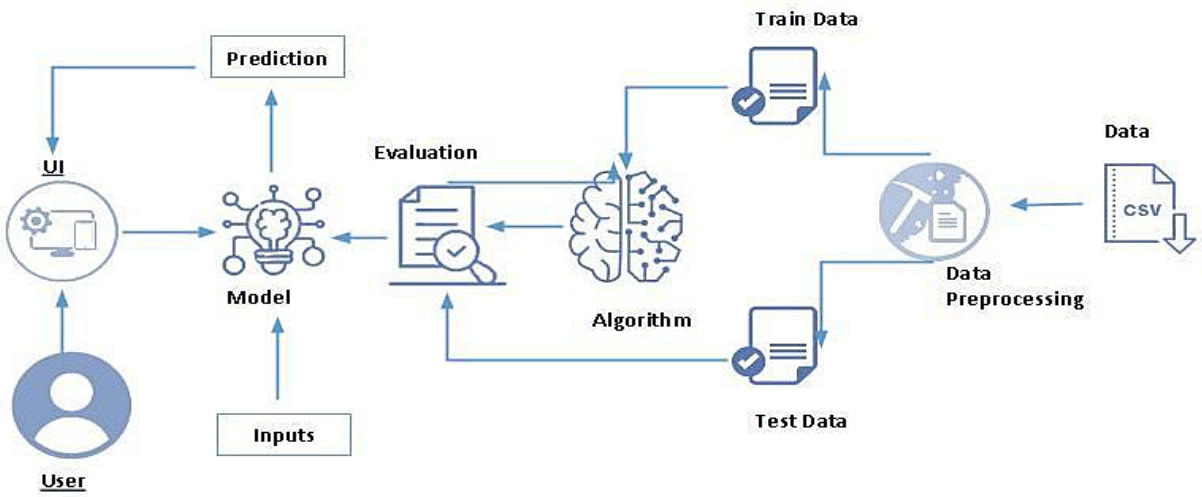
Dr. A. Chitra and S. Uma (2010) introduced a ensemble learning method for prediction of time series based on Radial Basis Function networks (RBF), K - Nearest Neighbor (KNN) and Self Organizing Map (SOM). They proposed a model namely PAPEM which perform better than individual model.
Akkoç (2012) used a model namely hybrid Adaptive NeuroFuzzy Inference model, grouping of statistics and NeuroFuzzy network. A 10-fold cross validation is used for better results and a comparison with other models.
Sarwesh Site, Dr. Sadhna K. Mishra (2013) proposed a method in which two or more classifiers are combined together to produce an ensemble model for the better prediction. They used the bagging and boosting techniques and then used random forest technique.

2.2 Proposed solution

We would compare machine learning models like decision tree, random forest classifier, KNeighbours and XGBoost algorithm. Out of which the model with the best accuracy will be chosen and export the respective model.

3. THEORETICAL ANALYSIS

3.1 Technical Architecture



3.2 Hardware / Software designing

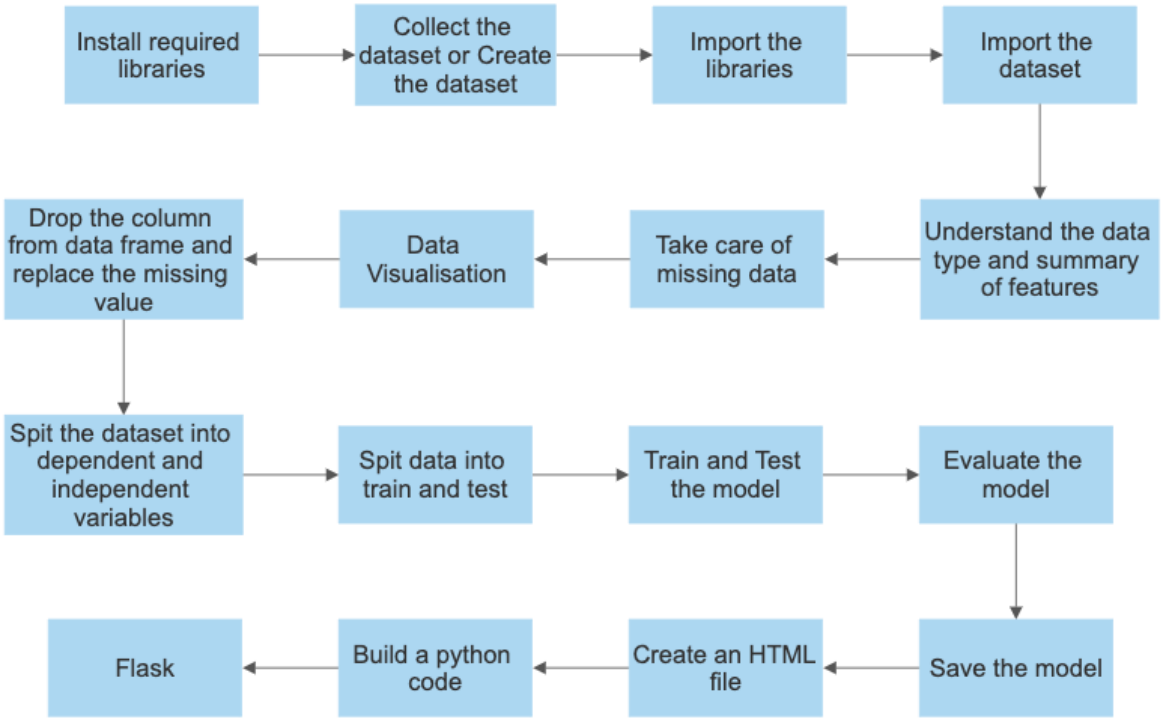
To complete this project, you must require the following software, concepts, and packages.

1. Anaconda navigator:
2. Python packages:
    - Open anaconda prompt as administrator
    - Type “pip install numpy” and click enter.
    - Type “pip install pandas” and click enter.
    - Type “pip install scikit-learn” and click enter.
    - Type ”pip install matplotlib” and click enter.
    - Type ”pip install pickle-mixin” and click enter.
    - Type ”pip install seaborn” and click enter.
    - Type “pip install Flask” and click enter.

4. EXPERIMENTAL INVESTIGATION

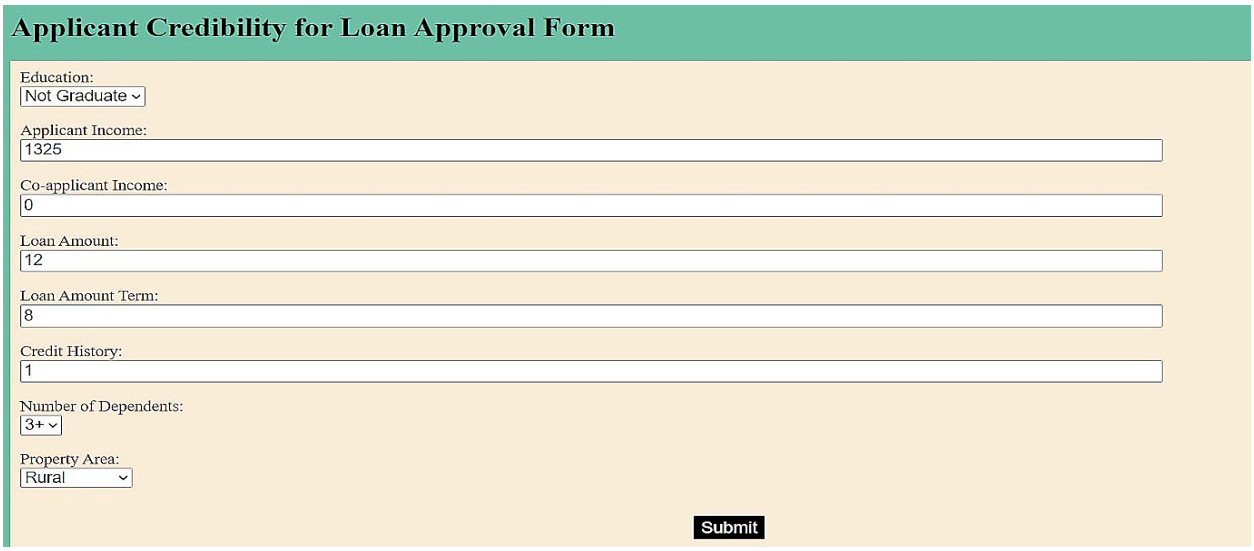
There were four machine learning algorithms used to- Decision Tree, Random Forest Classifier, KNeighbours and XGBoost. Out of which XGBoost was the best machine learning model with 78% accuracy.

5. FLOW CHART



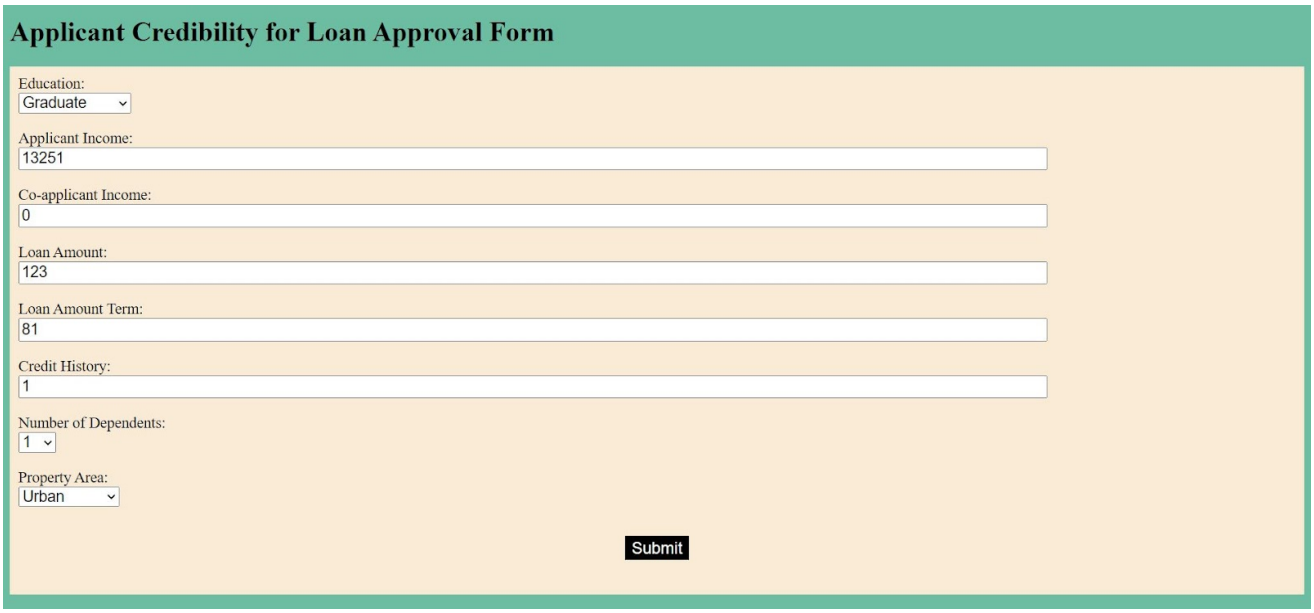
6. RESULT

Testcase 1: Applicant was a non-graduate from rural background with income 1325 units. The co-applicant's income was 0. Loan Amount was 12 units and the term was 8 terms. The applicant does have credit history and number of dependents were 3.



The model's prediction on this applicant was that they were rejected from the loan approval.

Testcase 2: Applicant was a graduate from urban background with income 13251 units. The co-applicant's income was 0. Loan Amount was 123 units and the term was 81 terms. The applicant does have credit history and the number of dependents were only 1.



The model's prediction on this applicant was that they were accepted for the loan.

7. ADVANTAGES AND DISADVANTAGES

7.1 Advantages:

(i) Keep Control of the Company:

A bank loans money to a business based on the value of the business and its perceived ability to service the loan by making payments on time and in full. Unlike with equity finance where the business issues shares, banks do not take any ownership position in businesses. Bank personnel also do not get involved in any aspect of running a business to which a bank grants a loan. This means you get to retain full management and control of your business with no external interference.

(ii) Bank loan is temporary:

Once a business borrower has paid off a loan, there is no more obligation to or involvement with the bank lender unless the borrower wishes to take out a subsequent loan. Compare this with equity finance, where the company may be paying out dividends to shareholders for as along as the business exists.

(iii) Interest is tax deductible:

The interest on business bank loans is tax-deductible. In addition, especially with fixed-rate loans, in which the interest rate does not change during the course of a loan, loan servicing payments remain the same throughout the life of the loan. This makes it easy for businesses to budget and plan for monthly loan payments. Even if the loan is an adjustable-rate loan, business owners can use a simple spreadsheet to compute future payments in the event of a change in rates.

7.2 Disadvantages:

(i) Tough to qualify:

One of the greatest disadvantages to bank loans is that they are very difficult to obtain unless a small business has a substantial track record or valuable collateral such as real estate. Banks are careful to lend only to businesses that can clearly repay their loans, and they also make sure that they are able to cover losses in the event of default. Business borrowers can be required to provide personal guarantees, which means the borrower's personal assets can be seized in the event

the business fails and is unable to repay all or part of a loan.

(ii) High Interest rates:

Interest rates for small-business loans from banks can be quite high, and the amount of bank funding for which a business qualifies is often not sufficient to completely meet its needs. The high interest rate for the funding a business does receive often stunts its expansion, because the business needs to not only service the loan but also deal with additional funding to cover funds not provided by the bank. Loans guaranteed by the U.S. Small Business Administration offer better terms than other loans, but the requirements to qualify for these subsidized bank loans are very strict.

8. APPLICATION

If you have money to invest for the short term, you can consider a new option in the debt segment other than traditional debt instruments such as debentures and bonds - smart lending, which has emerged as an attractive avenue for people who don't mind taking some additional risks for extra returns. This involves lending money to individuals or businesses through online services that match lenders with borrowers.

9. CONCLUSION

Smart lending revolutionized the field of lending across the world as well as in India. It is powered by technology and new business models. It is starting to emerge as a new alternative to traditional lending due to its cost effective, less time consuming and inclusive approaches. Digital lending is in fact an alternative to traditional lending because even the financially excluded people can access digital lending firms to avail loans and digital lending firms provide loans, mostly small and medium ticket loans, to such excluded people and their organizations using alternative credit scoring methodology.

10. FUTURE SCOPE

(i) Paperless transactions and online loans are the future of banking. The fintech market is fueled by a relaxed regulatory environment and the latest cutting-edge technologies.

(ii) Digitization of end-to-end lending processes eliminates the possibility of human error. The processing of applications along with credit underwriting is overseen using technology.

(iii) Again, the absence of any mediators makes online, paperless loan processing much more convenient and dynamic for borrowers with urgent financial requirements.

(iv) Moreover, the new-age digital lending processes do not have to go through the hassle of slow, bureaucratic methods.

(v) With online loans and alternative lending platforms, borrowers can apply for a loan anywhere, any time. Borrowers can get the loan irrespective of the lending bank being in their geography.

11. BIBILIOGRAPHY

IBM Deployment:
https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=functions-deployment-overview

Gradient Boosting model:
https://colab.research.google.com/github/lewtun/hepml/blob/master/notebooks/lesson04_intro-to-gradient-boosting.ipynb

Data Preprocessing, ML models theory:
SmartInternz Applied Data Science Course

APPENDIX

Source Code:

https://github.com/smartinternz02/SI-GuidedProject-48926-1652694452

(Or)

https://github.com/Vaibhav-Thalanki/SmartLender