RESEARCH PAPER ON LOAN APPROVAL PREDICTION USING MACHING LEARNING

NIKHIL KUMAR TIWAR
(DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING)
GALGOTIAS UNIVERSITY,GREATER NOIDA.
EMAIL:-
nikhiltiwar98560@gmail.com

ADITYA RANJAN
(DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING)
GALGOTIAS UNIVERSITY,GREATER NOIDA.
EMAIL: -
Adityabittu85@gmail.com

Abstract

Now a days, a lot of people are applying for bank loans for education, house construction, buying cars, and many more expenditures. Some of them are not able to return back the money because of some reasons.

In such scenarios, bank faces a loss. Also, bank has limited amount of assets which they want to spend only on those customers which can repay the loan amount. Thus, finding the right customers for the banks is one of the most challenging task.

So, we try to reduce the losses faced by banks by selecting the right customers who will repay the loan. We have collected some old data of customers from thebanks and using these data and some machine learning algorithms, we will trainthe machine and then will test on some test cases.

We will be focusing on the classification problem of "whether the person should be provided the loan or not?". We have divided the paper in some under-named parts:-

- Collection and distribution of Data into train and test part
- Different results obtained from already available models
- Designing more accurate model
- Results of our model

Introduction

The main business for the banks are the loan. They earn a large amount of their money from the interests on the loan they provide. There is a 90% system, according to which, a bank can give loan amount upto the 90% of the capital they have. Thus, they have to choose the right candidate for providing loan so that they does not face loss.

The banks have a very high level of validation checking procedures even though there is no guarantee that whether the chosen person is the right among the remaining ones.

Loan prediction is very important for the banks as well the customers. It will efficiently predict whether the should provide the loan to a particular candidate or not. It will consider some parameter and on the weightage of those parameters, it will output the result. The parameters which will affect the loan are selected by consulting banks and reading some articles on it. Some of the parameters which will be considered are:

- **Self Employed:** Whether the person is employed or not.
- **Educational Qualification:** Whether the person is educated or not.
- **Income:** How much the candidate earns.
- **Dependents:** Number of family members which are dependent on thecandidate.
- **Loan Amount:** How much amount the candidate demands.
- **LoanAmountPeriod:** Loan tenure in months
- **Credit History:** Credit history of the applicant
- **Property Area:** Area of the total property the applicant holds

Corpus

We create and release a corpus that aligns descriptions of financial events with changes in stock prices. We describe below these two components of the dataset. The entire corpus is available for download here:

http://nlp.stanford.edu/pubs/stock-event.html.

Basic Working

To evaluate the effectiveness of the models, a comparison is made between the two techniques on five different sector companies namely, JP Morgan, Nike, Johnson and Johnson, Goldman Sachs and Pfizer using both ANN and RF models.

Basic Design\Working

The historical data for the five companies has been collected from Yahoo Finance . The dataset includes 10 year data from 4/5/2009 to 4/5/2019 of Nike, Goldman Sachs, Johnson and Johnson, Pfizer and JP Morgan Chase and Co. The data contains information about the stock such as High, Low, Open, Close, Adjacent close and Volume. Only the day-wise closing price of the stock has been extracted.

Basic Concepts

We have done the project using google collab which is a free open system software designed by google to execute tasks using cloud processor.
Colab is one of the most used tools by data scientists for executing machine learning and deep learning projects. It uses Jupyter Notebook on the back-end torun ML projects.
Some of the libraries used in the projects are:

- **Pandas**
- **Numpy**
- **Mathplotlib**
- **Seaborn**
- **Sklearn**

We have used **Pandas** for importing and getting some statistical values like mean and deviation of the data. Pre-processing has been done using “pandas” and “numpy”.

We have used the visuals at most of the place in our project to understand the data better and most of these graphs like count of employed participant, loan amount-density, loan amount term are made using the powerful “**sea born**” library. Some of the plots were also made by using “**Mathplotlib**”. The “**skylearn**” library contains all the machine learning algorithms like ”gradient Boosting Classifier”, “Random forest classifier”, “Decision tree classifier” and K Neighbours Classifier”. We have used these algorithms in our project to get the desired precision.

Problem Statement

Account firms and banks need to automatize the credit qualification activity (continuously) essentially dependent on data given by customers when rounding out an online structure. Sex, Marital Status, Education, Number of Dependents, Salary, Loan Amount, Credit History, and different subtleties are incorporated. To digitize this interaction, they made an issue to group the client sections that can apply for a credit sum, permitting them to focus on these clients explicitly. They have introduced a fractional informational collection for this situation. “Approval of Loan is a very common real-life problem that every company faces in their lending operations. If the loan approval process is automated, it can save a lot of man hours and improve the speed of service to the customers.

The increase in customer satisfaction and savings in operational costs are significant”. “However, the rewards can only be realised if the bank has a sturdy model in place to accurately forecast which client's loans it should accept and which it should reject, in order to reduce potential risk.

Literature Survey

Vaidya had suggested a method for approving loan forecasts using logistic regression”“Logistic Regression is one of the most popular and very useful classification based algorithm” The purpose or the importance of using Logistic Regression was that it uses the concept of predictive analysis which was suitable enough for describing the data “M. Bayraktar et al. proposed a method for credit risk analysis using machine learning. Boltzman machine was used to make the analysis for risk calculation of loan “Y. Shi and P. Song proposed a method for evaluating project loans using risk analysis.

The method evaluated the risk involved in loans of commercial banks”“V. C. T. Chan et al. proposed a credit approval system using web-services. For clients loan as approved by the system. The consumer provides extra relevant information with the credit application. This information’s are processed by Credit Approval System which finally give credit score to the applicant. The paper developed a web services based solution of this problem After going through this, it is found that loan approval prediction task is very crucial for banking system. Machine learning algorithm are very helpful in predicting outcomes even when data is huge in size.

Machine Learning Algorithm Used:

Random Forest is a prominent learning method in Supervised Machine Learning, and it is effective for Regression & Classification tasks in ML

1. It generates random forests and then uses these random forests to seek thesolutions.
2. It is an ensemble learning in which significant number of classifiers are used to solve a complex problem.
3. Random forest analyzes each tree for prediction rather than just one to avoidoverfitting issues.
4. The greater the number of trees, the greater the accuracy in problem solving.

B.SUPPORT VECTOR MACHINE (SVM) SVM is a well known Supervised Machine Learning algorithm. Currently, the SVM classifier is the most common classifier 1.SVM has demonstrated a wide range of exceptional abilities, particularly in classification problems.

C.DECISION TREE Decision Tree is a non-parametric supervised machine learning algorithm. 1. It can also be used for both classification and regression, although it is more widely used for classification technique. 2. It works with categorical and continuous variables.

Loan Rejection Overview

A personal loan is a versatile unsecured loan and it can be rejected due to several reasons. Multiple personal loan rejections could increase your hard enquiries, ultimately affecting your credit score negatively. Hard enquiries are a sign that you are hungry for more credit and the lenders might assume you are a risky borrower. If your recent personal loan application has been rejected, you may want to find out the reasons for the denial. Read on to know more why your personal loan application could be rejected.

Reasons for Personal Loan Rejection:-

The reasons for a personal loan rejection can be determined broadly under six reasons, however these vary from one lender to another, each lender have their own internal policy to determine the creditworthiness and risk rating.

- Scant credit score
- Unstable employment
- Insufficient income
- Working in a delisted private company
- Not falling into the age limit
- Living in a location where defaults are high

Personal Loan Rejection Criteria:-

Every bank has its own eligibility criteria for approval of personal loans. Likewise, the reasons for personal loan rejection could vary with the rejection criteria of the other banks. The following could be the main reasons why your personal loan application could be rejected.

Personal Loan Rejection and Its Impact:-

The answer to this question depends on the level of your personal loan application being processed. The borrower generally applies for a personal loan after checking the eligibility criteria and proceed with document submission. If there are difficulties in furnishing the relevant documents, the personal loan might get rejected.

If all your documents are verified, the lenders will check your credit score to assess your creditworthiness. It is called hard enquiry. In worst case scenario, if your credit score is low and the personal loan gets rejected, the hard enquiry will have an impact on your credit score further in a small way. Hence, if you are faced with multiple hard enquiries, your credit score is sure to take a hit with it.

Conclusion

In this paper, machine learning was used to predict loan acceptance. The prediction method begins with data pre-processing, filling the missing values, experimental data analysis. After evaluating model on test dataset, each of thesealgorithms obtained a precision rate between 70% and 80%. Although here it can be concluded with certainty that the Support Vector Machine model is very efficient and produces superior results than other models.

The comparative analysis based on RMSE, MAPE and MBE values clearly indicate that ANN gives better prediction of stock prices as compared to RF. Results show

For future work, deep learning models could be developed which consider financial news articles along with financial parameters such as a closing price, traded volume, profit and loss statements etc., for possibly better results.

References

[1] Ashwini S. Kadam, Shraddha R Nikam, Ankita A. Aher, Gayatri V. Shelke, Amar S. Chandgude. "Prediction for Loan Approval using Machine Learning Algorithm", April 2021 International Research Journal of Engineeringand Technology (IRJET).

[2] Mohammad Ahmad Sheikh, Amit Kumar Goel,Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm",2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020

[3] X.FrencisJensy, V.P.Sumathi,Janani Shiva Shri, "An exploratory Data Analysis for Loan Predict ion based on nature of clients", International Journal of Recent Technology and Engineering (IJRTE),Volume-7 Issue-4S, November2018

[4] J. Tejaswini1, T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni Venkata Rao Maddumala. "ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH" Vol 11, www. jespublication.com, page 523 Issue 4, April/ 2020 ISSN NO: 0377-9254