INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

Customer loan prediction analysis

B. Divakar , P.Purnachandrakala,G.Lokesh,P.Narendra Babu,I.Bhanu Vaibhav,
Assitant Professor, Priyadarshini Institute of Technology&Science,AP,India.
UG, CSE, Priyadarshini Institute of Technology & Science, AP, India.

Abstract:

Product reviews are valuable for upcoming buyers in helping them make decisions. To this end, different opinion mining techniques have been proposed, where judging a review sentence's orientation (e.g. positive or negative) is one of their key challenges. Recently, deep learning has emerged as an effective means for solving sentiment classification problems. A neural network intrinsically learns a useful representation automatically without human efforts. However, the success of deep learning highly relies on the availability of large-scale training data. I propose a novel deep learning framework for product review sentiment classification which employs prevalently available ratings as weak supervision signals. The framework consists of two steps: (1) learning a high level representation (an embedding space) which captures the general sentiment distribution of sentences through rating information; (2) adding a classification layer on top of the embedding layer and use labeled sentences for supervised fine-tuning. I explore two kinds of low level network structure for modeling review sentences, namely, convolution feature extractors and long short-term memory. To evaluate the proposed framework, I construct a dataset containing 1.1M weakly labeled review sentences and 11,754 labeled review sentences from Amazon. Experimental results show the efficacy of the proposed framework and its superiority over baselines.

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

1 INTRODUCTION

Overview

Circulation of the loans is that the core business a part of as good as each and every bank. The principle parcel the bank's resources are straightforwardly came from the benefit acquire from the advances distributed by the banks. The main goal in banking system is to invest their resources in safe hands wherever it's. Now a day's several banks/financial agencies approves loan after a relapse method of verification and validation however still there's no surety whether or not the chosen candidate is the worthy right candidate out of all candidates. Through this method we are able to predict whether that particular candidate is safe or not and the whole method of validation of attribute is automated by machine learning technique [8][6]. The disadvantage of this model is that it emphasizes completely different weights to every issue however in reality sometime loan can be approved on the premise of single strong part only, that isn't possible through this method. Loan Prediction is useful for member of staff of banks as well as for the candidate. The aim of this Paper is to apply quick, immediate and easy way to choose the worthy person [6]. It will give special gain to the bank. The Loan Prediction method can automatically compute the heaviness of each attribute taking part in loan processing and on new test data information same issues are prepared with regard to their comparable heaviness. A period breaking point can be set for the candidate to check regardless of whether his/her loan can be affirmed or not. Loan Prediction technique licenses bouncing to explicit candidates with the goal that it very well may be keep an eye on need premise. This Paper is completely overseeing the power of Bank/finance Company, entire procedure of prediction is done secretly no colleagues would have the option to caution the process. Result against specific Loan Id can be ship off different divisions of companies so that they can make a proper move on application. This aides all other divisions to different conventions.

ISSN: 2057-5688

Data Source we obtained customer loan dataset from kg [4][2]. The dataset consists of various values/variables such as sex, marital status, education, self employed, loan status, applicant income, co applicant income etc…Data Description the dataset has 614 rows and 13 columns. 1 out of 13 columns is the target attribute i.e., default one attribute is target value. The dataset split into train and test data having shape (614, 13) and (367, 12) respectively.

The two most pressing issues in the banking sector are: 1) How risky is the borrower? 2) Should we lend to the borrower given the risk? The response to the first question dictates the borrower's interest rate. Interest rate, among other things (such as time value of money), tests the riskiness of the borrower, i.e. the higher the interest rate, the riskier the borrower. We will then decide whether the applicant is suitable for the loan based on the interest rate. Lenders (investors) make loans to creditors in return for the guarantee of interest-bearing repayment. That is, the lender only makes a return (interest) if the borrower repays the loan. However, whether he or she does not repay the loan, the lender loses money. Banks make loans to customers in exchange for the guarantee of repayment. Some would default on their debts, unable to repay them for a number of reasons. The bank retains insurance to minimize the possibility of failure in the case of a default. The insured sum can cover the whole loan amount or just a portion of it. Banking processes use manual procedures to determine whether or not a borrower is suitable for a loan based on results. Manual procedures were mostly effective, but they were insufficient when there were a large number of loan applications. At that time, making a decision would take a long time. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies. This model extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status). These reports make a bank manager's job simpler and quicker. A monetary loan is when one or more persons, organizations, or other entities lend money to other people, organizations, or entities. The recipient (i.e., the borrower) incurs a debt for which he or she is generally responsible for paying interest until the loan is repaid along with the principal amount borrowed. Nowadays, sanctioning of

ISSN: 2057-5688

loans has become a significant function of the financial institutions/banking sector. Loans are also one of the significant sources of income for banks. Banks apply interests on loans which are then sanctioned to their customers (borrowers). While sanctioning a loan, the lender needs to have an assurance of earning their money back along with interest. Thus, identifying the creditworthiness of an individual/an organization is highly important before sanctioning the loan. In this project, we focus mainly on monetary loans. The project aims to thoroughly verify the borrower and perform a background check based on several variables like gender, income, employment status, etc., to ensure whether the borrower is creditworthy and can be sanctioned the loan or not. .

2 RELATED WORK

1. Two Step Credit Risk Assessment Model for Retail Bank Loan Applications

Authors : M.Sudhakar and C.V.K. Reddy

Nowadays, there are many risks related to bank loans, for the bank and for those who get the loans. The analysis of risk in bank loans need understanding what is the meaning of risk. In addition, the number of transactions in banking sector is rapidly growing and huge data values are available which represent the customer behavior and the risks around loan are increased. Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets. In this paper a new model for classifying loan risk in banking sector by using data mining. The model has been built using data from banking sector to predict the status of loans. Three algorithms have been used to build the proposed model: j48, bayas net and Naïve bayas , By using weka application, the model has been implemented and tested. The result has been discussed and a full comparison between algorithms was conducted. J48 was selected as best algorithm based on accuracy.

2. Developing Prediction Model of Loan Risk in Banks

Authors: J.H. A Boobyda, and M.A. Tarrig

Data mining techniques uses some key ideas for data classification and prediction. Clustering techniques is used to place data items in to similar groups without prior

ISSN: 2057-5688

knowledge of group definitions. Clustering provides efficient decision making by grouping large voluminous datasets in bank. Risk assessment is an important task of bank, as the increase and decrease of credit limits in bank depends largely to evaluate the risk properly. The key problem consists of identifying good and bad customer's status those who applied for loan. An improvised risk evaluation of Multidimensional Risk prediction clustering Algorithm is implemented to determine the good and bad loan applicants whether they are applicable or not. In order to increase the accuracy of risk, risk assessment is performed in primary and secondary levels. Hence for avoiding Redundancy, Association Rule is integrated. This method allows for finding the risk percentage to determine whether loan can be sanctioned to a customer or not. Finally it is proven that proposed method predicts the better accuracy and consumes less time than existing method.

3. Credit Risk Analysis and Prediction Modeling of Bank Loans Using R

Authors: A.B. Hussain, and F.K.E. Sharouq

Nowadays there are many risks related to bank loans, especially for the banks so as to reduce their capital loss. The analysis of risks and assessment of default becomes crucial thereafter. Banks hold huge volumes of customer behaviour related data from which they are unable to arrive at a judgment if an applicant can be defaulter or not. Data Mining is a promising area of data analysis which aims to extract useful knowledge from tremendous amount of complex data sets. In this paper we aim to design a model and prototype the same using a data set available in the UCI repository. The model is a decision tree based classification model that uses the functions available in the R Package. Prior to building the model, the dataset is pre-processed, reduced and made ready to provide efficient predictions. The final model is used for prediction with the test dataset and the experimental results prove the efficiency of the built model.

4. Application of Data Mining Tools in CRM for Selected Banks

Authors: Dileep B. Desai #1, Dr. R.V.Kulkarni *2

Today, Banks to survive and grow it becomes critical to manage customers, build and maintain a healthy relationship with customers. Data Mining in Banks can play a significant role for customer relationship Management. The areas in which Data

ISSN: 2057-5688

mining Tools can be used in the banking industry are customer segmentation, Banking profitability, credit scoring and approval, Predicting payment from Customers, Marketing, detecting fraud transactions, Cash management and forecasting operations, optimizing stock portfolios, and ranking investments. Various Data Mining techniques for data modeling are Association, Classification, Clustering, Forecasting, Regression, Sequence discovery Visualization etc. Some examples of some widely used data mining algorithms are Association rule, Decision tree, Genetic algorithm, neural networks, k means algorithm, and Linear/logistic regression. This paper reviews some Data Mining tools and its application in Banks for Customer Relationship Management.

3.METHODOLOGY

LOAN APPLICANT DATA ANALYSIS

Whenever the bank makes decision to give loan to any customers then it automatically exposes itself to several financial risks. It is necessary for the bank to be aware of the clients applying for the loan. This problem motivates to loan EDA on the given dataset and thus analysis the nature of the customer. The dataset that uses EDA undergoes the process of normalization, missing value treatment, choosing essential columns using filtering, deriving new columns identifying the target variables and visualization the data in the graphical format. Python is used for easy and efficient processing of data. This paper used the Pandas library available in Python to process and extract information from the given dataset. The processed data is converted into appropriate graphs for better visualization of the results and for better understanding. For obtaining the graph matplot library is used.

Exploratory Data Analysis (EDA)

At the start, the dataset was cleaned. Then exploratory data analysis and feature engineering were performed. Then a model was created which predicted whether the

ISSN: 2057-5688

applicant would repay the loan or not. Whenever the bank makes a decision to give loan to any customers then it automatically exposes itself to several financial risks. It is necessary for the bank to be aware of the clients applying for the loan. This problem motivates to do an EDA on the given dataset and thus analyzing the nature of the customer. The dataset that uses EDA undergoes the process of normalization, missing value treatment, choosing essential columns using filtering deriving new columns, identifying the target variables and visualizing the data in the graphical format. Python is used for easy and efficient processing of data. This paper used the Pandas library available in Python to process and extraction formation from the given dataset. The processed data is converted into appropriate graphs for better visualization of the results and for better understanding. For obtaining the graph matplot library is used.

Machine Learning Methods

Machine learning is a subset of AI that trains machines with vast volumes of data to think and act like humans without being explicity programmed. In this paper we are using supervised (Classification methods) methods

Five machine learning classification models have been used for prediction of android applications. The models are available in python open source software. The brief details of each model are described below.

Decision Trees algorithm the basic algorithm rule of call tree needs all attributes or options ought to be discredited. Feature choice relies on greatest info gain of options. The data pictured in call tree will delineate within the kind of IF-THEN rules. This model is associate degree extension of C4.5 classification algorithms represented by Quinlan. Random Forest

Random forests are a classifying learning framework for characterization (and backslide) that work by building a very large number of Decision trees at planning time and yielding the class that's the mode of the classes surrender by individual trees.

Support Vector Machine Used SVM to build and train a model prepare a demonstrate utilizing human cell records, and classify cells to whether the tests are benign (mild state) or dangerous (evil state).Support vector machines are managed learning models that utilize affiliation R-learning calculation which analyze attributes and

ISSN: 2057-5688

distinguished design information, utilized for application classification. SVM can beneficially perform a replace utilizing the kernel trick, verified mapping their inputs into high dimensional attribute spaces [8].

Logistic Regression Logistic regression is supervised learning classification algorithm (try to method connections and conditions between the target prediction output and input attributes) such that we are able to anticipate the yield values for new information based on those connections which it learned from the previous information sets

Determine the training and testing data:

Typically , Here the system separate a dataset into a training set and testing set ,most of the data use for training ,and a smaller portions of data is use for testing. after a system has been processed by using the training set, it makes the prediction against the test set.

Data cleaning and processing:

In Data cleaning the system detect and correct corrupt or inaccurate records from database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing , modifying or detecting the dirty or coarse data. In Data processing the system converted data from a given form to a much more usable and desired from i.e. make it more meaningful and informative.

4 RESULTS

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

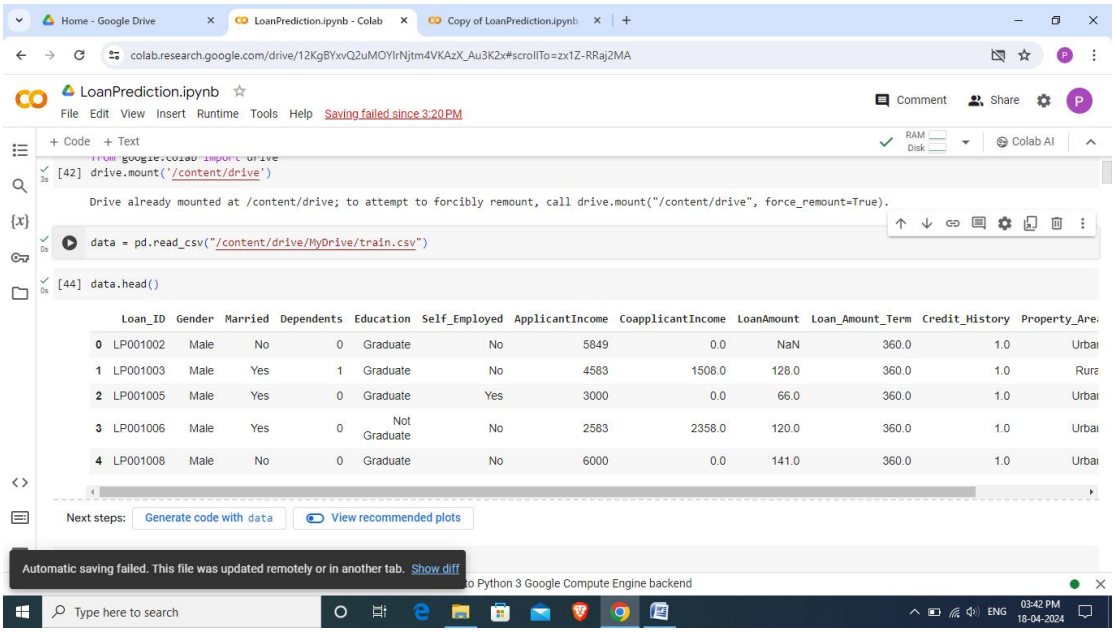

Fig1data set

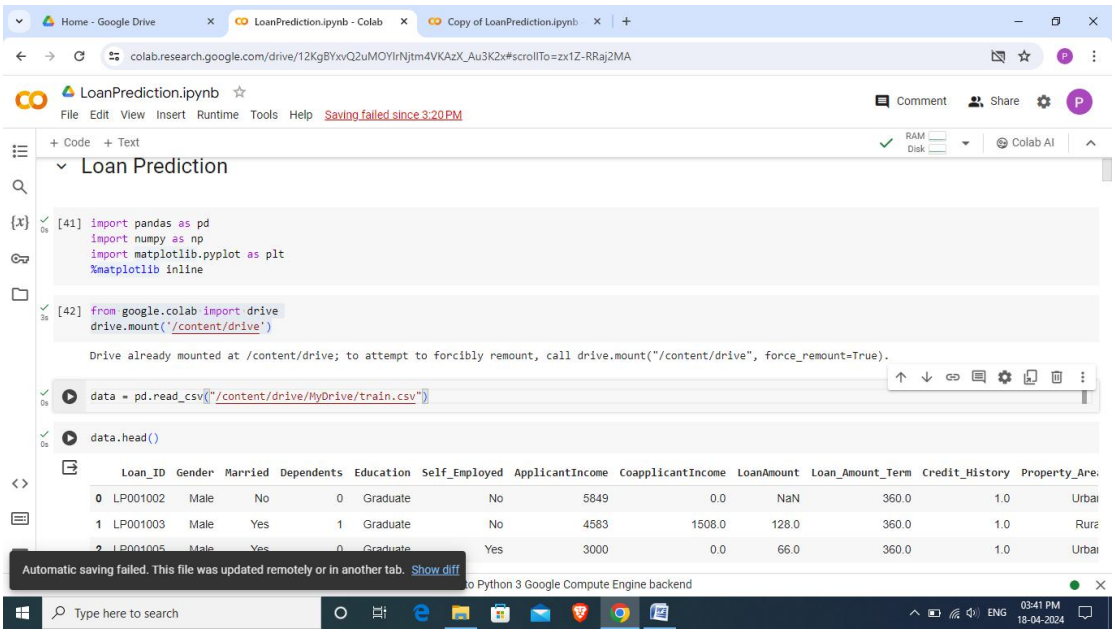

Fig2importModuels

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

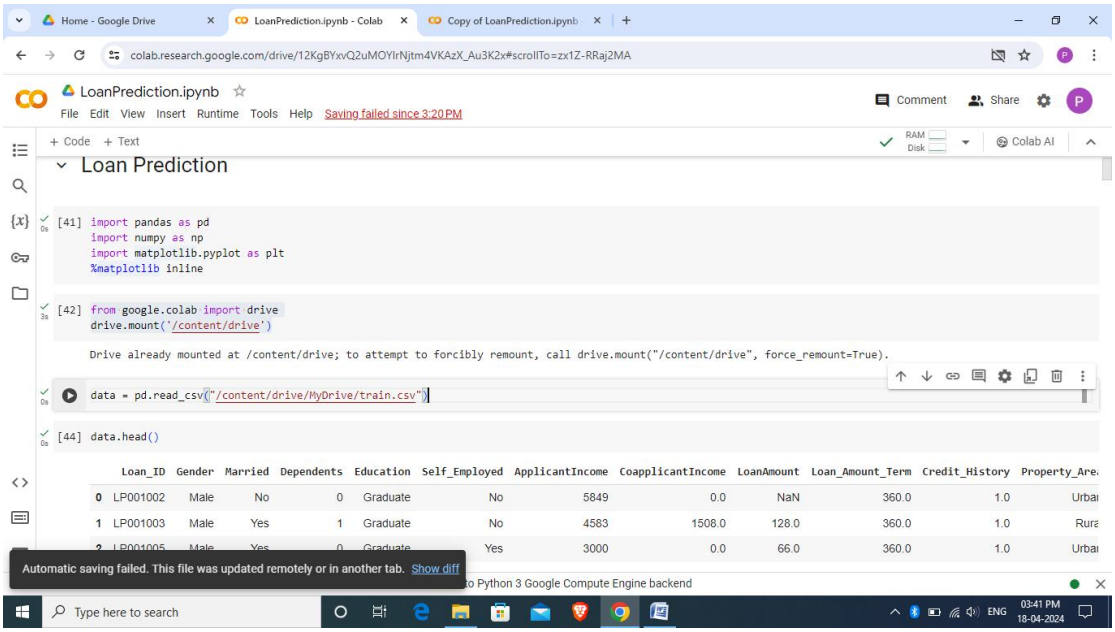

Fig 3: Final result

5.CONCLUSION

In this paper, we have proposed customer loan prediction using supervised learning techniques for loan candidate as a valid or fail to pay customer. In this paper, various algorithms were implemented to predict customer loan. Optimum results were obtained using Logistic Regression, Random Forest, KNN, and SVM, decision Tree Classifier. Compare these five algorithms random forest is the high accuracy. From a correct analysis of positive points and constraints on the part, it can be safely ended that the merchandise could be an extremely efficient part. This application is functioning properly and meeting to all or any Banker necessities. This part is often simply obstructed in several different systems. There are numbers cases of computer glitches, errors in content and most significant weight of option is mounted in machine-driven prediction system, therefore within the close of future the therefore called software system might be created more secure, reliable and dynamic weight adjustment. In close to future this module of prediction can be integrated with the module of machine-driven processing system.

6.REFERENCE

ISSN: 2057-5688

[1] Yu Jin and Yu Dan Zhu, “A data-driven approach to predict default risk of loan for online Peer-to-Peer (P2P) lending,” School of Information, Zhejiang University of Finance and Economics, 310018 Hangzhou, China.

[2] https://www.kaggle.com/telco-churn

[3] Bhoomi Patel, Harshal Patel, Jovita Hembram, Shree Jaswal “Loan default forecasting using data mining” Department of Information Technology, St. Francis Institute of Technology, Mumbai, India (2020)

[4] Octave Iradukunda, Haiying Che, Josiane Uwineza, Jean Yves Bayingana, Muhammad S Bin-Imam, Ibrahim Niyonzima ”Malaria Disease Prediction Based on Machine Learning” School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China (2019).

[5] G. Arutjothi, Dr. C. Senthamarai “Prediction of Loan Status in Commercial Bank using Machine Learning Classifier” department of computer applications government arts college (Autonomous) Salem, India (2017.)

[6] Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar “An Approach for Prediction of Loan Approval using Machine Learning Algorithm” School Of Computer Science And Engineering Galgotias University Greater Noida, India (2019).

[7] Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li “Overdue Prediction of Bank Loans Based on LSTM-SVM”Jiangsu Key Lab of Big Data and Security and Intelligent Processing Nanjing University of Posts and Telecommunications, Nanjing, 210023, China.

[8] Aakanksha, Tamara Denning, Vivek Srikumar, Sneha Kumar Kesera “secrets in source code: reducing false positives using ML” software engineering (Microsoft) school of computing, USA (2020)

[9] Arutjothi .G, Dr. C. Senthamarai. "Credit Risk Evaluation using Hybrid Feature Selection Method. Software engineering and technology (2017)

[10] Ch. Balayesu and S Narayana, “An Improved Algorithm for Efficient Mining of Frequent Item Sets on Large Uncertain Databases” in International Journal of Computer Applications, Volume 73, No. 12 July 2013, Page No. 8-15.

ISSN: 2057-5688

[11] Bala brahmeswara kadaru et al.”A novel ensemble decision tree classifier using hybrid feature selection measures for parkinson’s disease prediction”, Int. J. Data science (IJDS), ISSN: 2053-082X, Vol.3, No.4,2018.

[12] Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.