

LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING

Manjunatha. N

Research Scholar, Department of
Computer Science & Engineering, Shri
JIT University, Rajasthan, India.

Dr. Prasadu Peddi

Assistant Professor, Department of
Computer Science & Engineering, Shri
JIT University, Rajasthan, India.

ABSTRACT

Modern financial landscape, efficient and accurate loan processes are critical for managing risk and enhancing customer satisfaction. This study explores the application in machine learning techniques to predict loan approval outcomes. Analyzing historical loan data, including applicant demographics, financial history, and other relevant features, various machine learning algorithms are logistic regression, decision trees, random forests, and gradient boosting are employed to develop predictive models. Performance of these models is evaluated using metrics like accuracy, precision, recall, and with a focus on identifying the most effective approach for predicting loan approvals. Machine learning models can significantly improve the precision of loan approval predictions, leading to more informed decision-making and optimized resource allocation. This research highlights the potential for machine learning to enhance financial decision processes and proposes future directions for integrating these techniques into practical loan approval systems.

KEYWORDS: Machine Learning, logistic regression, decision trees, random forests.

INTRODUCTION

Distribution of the loans is the core business part of almost every banks. The main portion the bank's assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approve loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is

the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it emphasizes different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through this system.

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight. A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Paper is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing. Result against particular Loan Id can be send to various department of banks so that they can take

appropriate action on application. This helps all others department to carried out other formalities

LITERATURE SURVEY

1. An introduction to logistic Regression Analysis and Reporting

The purpose of this article is to provide researchers, editors, and readers with a set of guidelines for what to expect in an article using logistic regression techniques. Tables, figures, and charts that should be included to comprehensively assess the results and assumptions to be verified are discussed. This article demonstrates the preferred pattern for the application of logistic methods with an illustration of logistic regression applied to a data set in testing a research hypothesis. Recommendations are also offered for appropriate reporting formats of logistic regression results and the minimum observation-to-predictor ratio. The authors evaluated the use and interpretation of logistic regression presented in 8 articles published in The Journal of Educational Research between 1990 and 2000. They found that all 8 studies met or exceeded recommended criteria.

2. An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain Park

The purpose of this article is twofold: 1) introducing logistic regression (LR), a multivariable method for modeling the relationship between multiple independent variables and a categorical dependent variable, and 2) examining use and

reporting of LR in the nursing literature. Methods: Text books on LR and research articles employing LR as main statistical analysis were reviewed. Twenty-three articles published between 2010 and 2011 in the Journal of Korean Academy of Nursing were analyzed for proper use and reporting of LR models. Results: Logistic regression from basic concepts such as odds, odds ratio, logit transformation and logistic curve, assumption, fitting, reporting and interpreting to cautions were presented. Substantial shortcomings were found in both use of LR and reporting of results. For many studies, sample size was not sufficiently large to call into question the accuracy of the regression model. Additionally, only one study reported validation analysis. Conclusion: Nursing researchers need to pay greater attention to guidelines concerning the use and reporting of LR models.

3. A comprehensive introduction to machine learning that uses probabilistic models and inference as a unifying approach.

Today's Web-enabled deluge of electronic data calls for automated methods of data analysis. Machine learning provides these, developing methods that can automatically detect patterns in data and then use the uncovered patterns to predict future data. This textbook offers a comprehensive and self-contained introduction to the field of machine learning, based on a unified, probabilistic approach.

The coverage combines breadth and depth, offering necessary background material on such topics as probability, optimization,

and linear algebra as well as discussion of recent developments in the field, including conditional random fields, L1 regularization, and deep learning. The book is written in an informal, accessible style, complete with pseudo-code for the most important algorithms. All topics are copiously illustrated with color images and worked examples drawn from such application domains as biology, text processing, computer vision, and robotics. Rather than providing a cookbook of different heuristic methods, the book stresses a principled model-based approach, often using the language of graphical models to specify models in a concise and intuitive way. Almost all the models described have been implemented in a MATLAB software package—PMTK (probabilistic modeling toolkit)—that is freely available online. The book is suitable for upper-level undergraduates with an introductory-level college math background and beginning graduate students.

4. Loan Approval Prediction based on Machine Learning Approach

The training data set is now supplied to machine learning model, on the basis of this data set the model is trained. Every new applicant detail filled at the time of application form acts as a test data set. After the operation of testing, model predict whether the new applicant is a fit case for approval of the loan or not based upon the inference it concludes on the basis of the training data sets.

Six machine learning classification models have been used for prediction of android applications. The models are available in R open source software. R is licensed under

GNU GPL. The brief details of each model are described below.

➤ Decision Trees (C5.0):

The basic algorithm of decision tree requires all attributes or features should be discretized. Feature selection is based on greatest information gain of features. The knowledge depicted in decision tree can be represented in the form of IF-THEN rules. This model is an extension of C4.5 classification algorithms described by Quinlan.

➤ Random Forest (RF):

Random forests are a group learning system for characterization (and relapse) that work by building a large number of Decision trees at preparing time and yielding the class that is the mode of the classes yield by individual trees.

➤ Support Vector Machine (SVM):

Support vector machines are administered learning models that use association learning algorithm which analyze features and identified pattern knowledge, utilized for application classification. SVM can productively perform a regression utilizing the kernel trick, verifiably mapping their inputs into high-dimensional feature spaces.

➤ Linear Models (LM):

The Linear Model is numerically indistinguishable to a various regression analysis yet burdens its suitability for both different qualitative and numerous quantitative variables.

➤ Neural Network (Nnet):

Neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs, to find patterns in data, or to capture the statistical structure in an unknown joint

probability distribution between observed variables.

➤ Adaboost (ADB):

Adaboost short for " Adaptive Boosting ". It is delicate to noisy information data and outliers. It is different from neural systems and SVM because Adaboost preparing methodology chooses just those peculiarities known to enhance the divining power of the model, decreasing dimensionality and conceivably enhancing execution time as potentially features don't have to be processed.

5. Prediction of Loan Approval using Machine Learning

Vaidya proposed a method for approval of loan prediction using logistic regression. Logical regression is a machine learning technique which is very useful in prediction system. The approval of loan is a very important process in banking system. Vaidya solves the problem by applying machine learning in a sample data set for loan approval applications. It also opens other areas on which machine learning is applicable. Li and Q. Sun find a method to calculate risk involved in loan approvals for SMEs. A concept of loan consuming radius was introduced which was based upon supply chain in consumer market. F. M. Isik et al. develop a loan approval system using Business Process Execution Language BPEL. The concept of BPEL is very useful in business firms. A reasoning engine was developed which removes some services from the BPEL process which are not necessary to complete a process. The system was applied on loan approval which involve many processes. V. C. T. Chan et al. proposed a credit approval system using web services.

The system approved credit for the customers. With credit application the customer submits some other useful information's. 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. J. Lohokare et al. proposed a system which automatically collect data for an applicant and decides the credit score. The system work on the social media to collect information about the user. R. Yang et al. analyzed that whether the credit default behavior of a SME depends upon credit features of its owner or not. The author concluded that features of the owner behave as valuable parameters to calculate risk of a loan for SMEs. 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. H. A. P. Pérez et al. introduced fuzzy model for calculation of credit score of the customer.

The information collected by the system for calculation of the credit score was converted into gradual values using fuzzy sets. The fuzzy based method performs better for calculation of the credit score of the applicants. S. Yadav and S. Thakur applied Big Data approach for loan analysis. The techniques of big data analysis were applied on customer data to calculate bank loan analysis. Hadoop based method was used in the loan analysis. Y. Lin analysis of the effect of the political approaches effect the loans of state banks. The paper investigated that in state owned banks, the political relationship plays a considerable role. Ruifen Zhao worked on

approval of college loans. Education loans are very common among students because of rise in the cost of education. The paper investigated the issues in loan approval of college students. M. Houshmand and M. D. Kakhki proposed an expert system which evaluates the loan approvals. The system used rule base approaches for loan approval decisions. L. Hui-ling analyze the relation between characteristics of the banks, firms and loans approval.

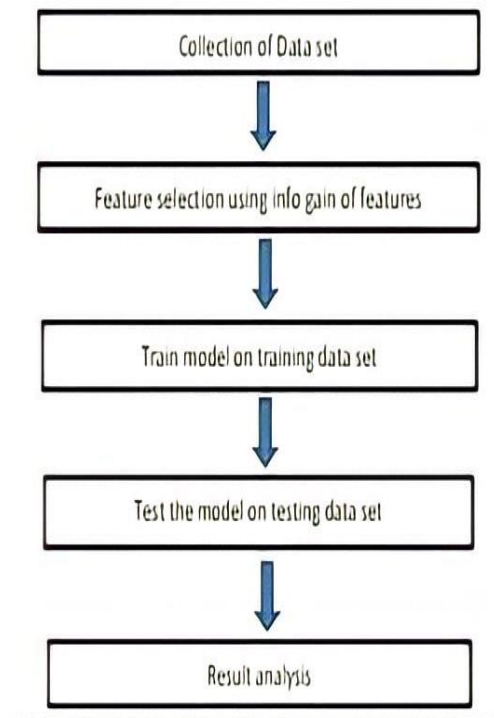
The paper investigated that there is a strong relationship between approval of loans and characteristics of business firm who apply the loan and characteristics of the bank. C. Yin apply fuzzy logic to calculate the bank loan risks. A new pattern recognition system using fuzzy logic was developed which evaluate the risks involves in the approval of bank loans for applicants. J. Ma and Y. Cheng proposed Markov Chain based model for risk management of bank loans. A. V. Gutierrez proposed a model for housing loan. The model was worked for green housing loans. J. Chen and W. Guo worked on loan limit of the loan applicants. The model worked on supply chain for financing decision making. G. Arutjothi and C. Senthamarai, used machine learning classifier for prediction of loan approval status in banks.

The machine learning based prediction system was applied on commercial banks. The paper conclude that the machine learning approach is very useful in loan status prediction. Y. Shi and P. Song proposed a method for evaluating project loans using risk analysis. The method evaluates the risk involved in loans of commercial banks. R. Zhang and D. Li used machine learning approached in prediction

systems. The machine learning approach was used for assessment of water quality. The paper concluded that machine learning is a very unimportant tool in prediction systems. C. Frank et al. It used machine learning in prediction of smoking status. Different machine learning approaches were applied and investigated for finding the smoking status. From the results its was ensured that logistic algorithm performs better. R. Lopes et al. applied machine learning approach for the prediction of credit recovery.

Credit recovery is very important issue for banking system. The prediction of credit recovery is a challenging task. Different machine learning approach was applied to predict the credit recovery and gradient expansion algorithms (GBM) outperformed the other machine learning approaches. After going through this literature, it is found that loan approval prediction problem is very important for banking system. Machine learning algorithm are very useful in predicting outcomes even when data is very big in size. This paper investigated some machine learning algorithms and applied ML on test data set of loan approvals.

METHODS



MACHINE LEARNING METHODS

Six machine learning classification models have been used for prediction of android applications. The models are available in R open-source software. R is licensed under GNU GPL. The brief details of each model are described below.

a) Support Vector Machine (SVM)

Support vector machines are administered learning models that uses association r learning algorithm which analyze features and identified pattern knowledge, utilized for application classification. SVM can productively perform a regression utilizing the kernel trick, verifiably mapping their inputs into high-dimensional feature spaces.

b) Linear Models (LM)

The Linear Model is numerically indistinguishable to a various regression

analysis yet burdens its suitability for both different qualitative and numerous quantitative variables.

CONCLUSION:

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so called software could be made more secure, reliable and dynamic weight adjustment. In near future this module of prediction can be integrate with the module of automated processing system. the system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time.

RFERENCES:

1. Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, —Developing Prediction Model of Loan Risk in Banks using Data Mining^l, *Machine Learning and Applications: An International Journal (MLAIJ)*, Vol.3, No.1, pp. 1-9, March 2016.
2. Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, Archana Gupta, —Loan Sanctioning Prediction System^l, *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 6, no. 4, pp. 50-53, 2016
3. Adewale O Adebayo, Aibangee Zandra, Lufadeju Oluwatosin & Maradesa Adepeju “Creation of Loan Automation Application”*G.J. E.D.T., Vol.3(1):52-61, 2014.*



4. T.Subbareddy, M. Meghana and K. Swathi "online banking loan services" *Intelligence and Sustainable Systems (ICAISS). IEEE, 2023.* Volume 2, Issue 7, July 2013.
5. K. Hanumantha Rao, G. Srinivas, A. Damodhar, M. Vikas Krishna: *Implementation of Anomaly Detection Technique Using Machine Learning Algorithms: Internatinal Journal of Computer Science and Telecommunications (Volume2, Issue3, June 2011).*
6. S.S. Keerthi and E.G. Gilbert. *Convergence of a generalize SMO algorithm for SVM classifier design. Machine Learning, Springer, 46(1):351–360, 2002.*
7. Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6. doi: 10.1109/ICCCNT.2017.8203946.
8. *Prediction of Loan Approval using Machine Learning - Rajiv Kumar*1, Vinod Jain2, PremSagar Sharma3, Shashank Awasthi4, Gopal Jha5 1,2,4,5GL Bajaj Institute of Technology & Management Greater Noida, Computer Science, 201306, India 3GL Bajaj Institute of Technology & Management Greater Noida, Computer Application, 201306, India.*
9. Verma, Abhigya, et al. "Advanced Regression Models for Accurate House Price Prediction: An Analysis of Performance and Interpretability." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.
10. Gupta, Amit, Shashi Kant Dargar, and Abha Dargar. "House prices prediction using machine learning regression models." 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2022.
11. Kalaivani, K. S., et al. "A Comparative Study of Regression Algorithms on House Sales Price Prediction." 2023 Second International Conference on Augmented