IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

LOAN PREDICTION SYSTEM USING MACHINE LEARNING

Preeti joshi [#1], Vinayak Tumbagi[#2], Gaurav Belakar [#3] , Sauradh Mahalankar [#4] , Aditya Wakankar [#5]

[#]Information Technology Department, Marathwada Mitra Mandal's College of Engineering, Pune, India

[1]preetijoshi@mmcoe.edu.in

[2]vinayaktumbagi2019.it@mmcoe.edu.in

[3]gauravbelakar2019.it@mmcoe.edu.in

[4]sauradhmahalankar019.it@mmcoe.edu.in

[5]Adityawakankar2019.it@mmcoe.edu.in

Abstract

The proposed system will use a machine learning algorithm to predict loan status. DNN is the algorithm used in this model. The user will be prompted for information and as a result, results will be generated upon loan approval. The data provided by the user will be compared with the data set of the train to be provided on the status of the loan. We have implemented a Loan Credibility Prediction System that helps organizations make the right decision to approve or deny customers' loan application. This will certainly help the banking sector to open efficient distribution channels. In this model, the DNN algorithm is used for prediction. Integration of other techniques that exceed the performance of popular mining models should be implemented and tested for the field.

Keywords: DNN, Machine Learning, Loan Prediction.

I. INTRODUCTION

People all around the world are some way or another ward on banks to loan them credits for different motivations to assist them with beating their monetary requirements and accomplish a few individual objectives. Because of the consistently changing economy and steadily expanding rivalry in the monetary world, the matter of applying for a new line of credit has become unavoidable. What's more, little and huge scope banking organizations rely upon the loaning industry to procure benefits from maintaining their business and to work flawlessly in the midst of monetary requirements. A credit is the fundamental type of revenue for banks in the area as well as the primary wellspring of monetary gamble for banks. A very remarkable bank's resources come straightforwardly from revenue procured on advances. The credit scoring process expected specialists as well as factual calculations to precisely foresee a candidate's reliability. In any case, as of late, specialists and banking specialists have selected preparing classifiers in light of different AI and profound learning calculations to consequently foresee a candidate's financial assessment in view of their record as a consumer. Furthermore, other verifiable information and incredibly work on the most common way of choosing reasonable competitors before

credit endorsement. Accordingly, by tending to the previously mentioned situation, the target of this article is to talk about the utilization of various AI models in the loaning system and to propose the best methodology for a monetary establishment to precisely distinguish to whom loan and help banks. Recognize defaulting indebted individuals for exceptionally low credit risk. The classifiers we used to construct the model are the DNNs. They will be utilized independently to examine the dataset and recognize designs in the dataset and gain from them. In light of this examination, foresee if another candidate is probably going to default on a credit.

II. BACKGROUND

The recent advancement of machine learning and data mining techniques has sparked interest in their use in a variety of sectors. The banking sector is no exception, and the growing pressure on financial organizations to have effective risk management has sparked interest in improving current risk estimating approaches. Machine learning techniques could potentially lead to a better estimate of the financial risks that banks are exposed to. The Basel accords, which establish frameworks for regulatory standards and risk management procedures as a guideline for banks to manage and quantify their risks, have been continuously developed in the credit risk domain. The standardized approach and the internal ratings-based approach are two approaches offered by Basel II for measuring the minimum capital requirement (IRB). Different risk metrics are taken into account by banks in order to predict the potential loss they may face in the future. The expected loss (EL) a bank would bear in the event of a defaulted customer is one of these metrics. The probability of a particular client defaulting is one of the components involved in EL-estimation. Customers who are in default have failed to meet their contractual responsibilities and may be unable to repay their loans.

III. MOTIVATION

There are unsolved fraudulent behaviors in society's financial processes, such as bank credit administration, that demand a solution through intelligent technology. Existing fraud detection techniques in bank credit administration have not fulfilled the requisite accuracy and avoidance of false alarms, and none have specifically targeted fraud in bank credit default. Also affecting forecast accuracy are fraudulent duplicates, missing data, and undefined fraud situations. A financial fraud is any criminal conduct committed by humans or robots that results in personal benefit at the expense of institutions or legitimate human benefactors, although an error should not be mistaken for a fraud. It is referred to as economic sabotage when considering the entire effect of financial scams. Money laundering, bank credit fraud, pension fraud, co-operative society fraud, tax evasion, telecommunications fraud, credit card fraud, inflated contracts, financial reports fraud, health insurance fraud, automobile insurance fraud, and mortgage insurance fraud are just a few examples of financial fraud. According to, credit card fraud, telecommunication fraud, computer intrusion, bankruptcy fraud, theft fraud or counterfeit fraud, and application fraud are all examples of fraud. The impact of fraud on a country's economy is real, and various measures have been tried, yet they all have flaws. Machine learning, on the other hand, has proven to be more dependable. Machine learning employs data mining techniques to uncover hidden patterns in big, volatile, and diverse datasets, allowing users to make informed decisions based on the information gained.

IV. LITERATURE SURVEY

Yu Li et.al [1] Credit scoring which assists with assessing the capacity of reimbursement of clients is perhaps the main issues for advance organization. Because of the fast advancement of AI procedures in software engineering region and other related subjects, different groupings strategies have been proposed to work with the execution of AI techniques for describing the reimbursement conduct of clients. In this exploration, they utilize the XGBoost calculation to distinguish the awful clients who don't repay cash from the great clients.

Lin Zhu ET.AL[2] Recently, with the development of electronic trade and large information innovation, P2P internet loaning stages have carried open doors to financial specialists, and yet, they are additionally confronted with the gamble of client advance default, which is connected with the manageable and solid improvement of stages. Consequently, in light of the Random Forest calculation, this paper fabricates an advance default expectation model considering this present reality client advance information on Lending Club. The SMOTE strategy is embraced to adapt to the issue of irregularity class in the dataset, and afterward a progression of tasks, for example, information it are completed to clean and dimensionality decrease. The test results show that: Random Forest calculation.

Mayank Anand et.al [3] Given advance default forecast to a great extent affects profit; it is one of the most powerful variables using a credit card score that banks and other monetary associations face. There have been a few conventional techniques for mining data about a credit application and some new AI strategies for which, the majority of these techniques seem, by all accounts, to be coming up short, as the quantity of defaults in advances has expanded. For advance default expectation, an assortment of strategies like Multiple Logistic Regression, Decision Tree, Random Forests, Gaussian Naive Bayes, Support Vector Machines, and other gathering techniques are introduced in this examination work. The expectation depends on advance information from different web sources like Kaggle, as well as informational collections from the candidate's advance application. Huge assessment measures including Confusion Matrix, Accuracy, Recall, Precision, F1-Score, ROC examination region and Feature Importance has been determined and displayed in the outcomes segment.

Jing Zhou et.al [4] lately, another Internet-based unstable credit model, distributed (P2P) loaning, is thriving and has turned into an effective supplement to the customary credit business. Be that as it may, credit risk stays inescapable. A key test is making a default forecast model that can successfully and precisely foresee the default likelihood of each credit for a P2P loaning stage. Because of the qualities of P2P loaning acknowledges information, for example, high aspect and class unevenness, regular factual models and AI calculations can't successfully and precisely foresee default likelihood. To resolve this issue, a choice tree model-based heterogeneous gathering default expectation model is proposed in this paper for exact forecast of client default in P2P loaning. Angle supporting choice trees (GBDT), outrageous inclination helping (XGBoost) and light slope supporting machine (LightGBM) are utilized as individual classifiers to make a heterogeneous gathering learning-based default forecast model.

Yuelin Wang et.al[5] a few procedures, (for example, factual methods and AI strategies) have been produced for assessing individual credit data to conclude whether the individual meets the rules of credit funding, and the interaction is known as credit scoring. This paper basically centers around the relative appraisal of the exhibitions of five famous classifiers associated with AI utilized for credit scoring: Naive Bayesian Model, Logistic Regression Analysis, Random Forest, Decision Tree, and K-Nearest Neighbor Classifier.

Vincenzo Moscato et.al [6] Credit risk appraisal assumes a vital part for accurately supporting monetary organizations in characterizing their bank approaches and business systems. Throughout the last ten years, the arising of social loaning stages has upset customary administrations for credit risk appraisal. Through these stages, moneylenders and borrowers can undoubtedly associate among them with practically no inclusion of monetary organizations. Specifically, they support borrowers in the raising money process, empowering the cooperation of any number and size of loan specialists. In any case, the absence of loan specialists' insight and absent or unsure data about borrower's record can increment gambles in friendly loaning stages, requiring an exact credit risk scoring. To beat such issues, the credit risk evaluation issue of monetary tasks is normally displayed as a parallel issue based on obligation's reimbursement and legitimate AI methods can be therefore taken advantage of. In this paper, they propose a benchmarking investigation of the absolute most utilized credit risk scoring models to anticipate assuming an advance will be reimbursed in a P2P stage. They manage a class awkwardness issue and influence a few classifiers among the most utilized in the writing, which depend on various inspecting procedures.

P. Maheswari et.al [7] with the advancement of innovation and execution of Data Science in banking, changes the substance of banking industry. The majority of the banking, monetary areas and social loaning stages are effectively contributing on loaning. In any case, monetary organizations could confront gigantic capital misfortune assuming that they

endorsed the advance without having any earlier appraisal of default risk. Monetary organizations generally need a more exact prescient framework for different purposes. Foreseeing credit defaulters is a significant assignment for the financial business. Banks have gigantically huge measure of information like client's information, exchange conduct, and so forth.

E. Chandra Blessie et.al [8] Extending credits to corporates and people for the smooth working of developing economies like India is inescapable. As expanding number of clients apply for advances in the banks and non-banking monetary organizations (NBFC), it is truly trying for banks and NBFCs with restricted cash-flow to gadget a standard goal and safe strategy to loan cash to its borrowers for their monetary necessities. Also, as of late NBFC inventories have experienced a huge defeat with regards to the stock cost. It has added to a disease that has additionally spread to other monetary stocks, unfavorably influencing the benchmark lately. In this paper, an endeavor is made to consolidate the gamble engaged with choosing the reasonable individual who could reimburse the credit on time along these lines keeping the bank's non-performing resources (NPA) on the hold.

Anthony Bellotti et.al [9] They look at the presentation of a wide arrangement of relapse strategies and AI calculations for foreseeing recuperation rates on non-performing advances, utilizing a private data set from an European obligation assortment office. They find that standard based calculations like Cubist, supported trees, and arbitrary timberlands perform fundamentally better compared to different methodologies. Notwithstanding credit contract specificities, indicators that allude to the bank recuperation process - before the portfolio's deal to an obligation authority - are additionally displayed to improve guaging execution.

I O Eweoya et.al [10] The rate at which banks loses assets to credit recipients because of advance default is disturbing. This pattern has prompted the conclusion of many banks, potential recipients denied of admittance to advance; and numerous laborers losing their positions in the banks and different areas. This work utilizes past advance records in light of the work of AI to foresee misrepresentation in bank advance organization and therefore stay away from advance default that manual examination by a credit official could not have possibly found. In any case, such secret examples are uncovered by AI. Factual and ordinary methodologies toward this path are limited in their exactness capacities.

Omprakash Yadav et.al [11] Data mining techniques are used in a variety of other fields, including the telecommunications business, retail industry, biological data analysis, intrusion detection, and a variety of other scientific applications. Data mining techniques can also be employed in the banking business to assist them compete with other competitors in the market. We have proposed an effective prediction model for bankers in this research, which will assist them in predicting the creditworthiness of consumers who have asked for a loan..

Hafiz Ilyas Tariq et.al [12] The goal of this project is to conduct a thorough investigation and construct a model for predicting loan defaults. This type of approach is unavoidable because bad loans are a major problem in the financial sector, particularly in microfinance institutions in both developing and developed countries. To address this issue, a comprehensive literature analysis was conducted to investigate the main components that contribute to this problem..

YA-QI CHEN et.al [13] This is a significant issue for banking and finance firms. The goal of this study is to address the data imbalance problem in order to improve the loan default prediction performance. The method used in this research is a hybrid under sampling strategy that incorporates clustering, stochastic sensitivity, and radial basis function neural networks. To validate the performance of our technique, we used real loan default data from a Chinese P2P company. The outcomes of our experiments show that our method produces higher generalization results.

SRISHTI SRIVASTAVA et.al [14] The key goal achieved here is that by employing this technology, we will be able to quickly identify reliable consumers, which would otherwise require a lengthy procedure by bank personnel in the current environment.

Dr. P.G Sivagaminathan et.al [15] The data for the applications is stored in multiple relational databases that are linked together using primary and foreign keys. Many people are seeking for bank loans today for a variety of reasons. However, not all of them are trustworthy. Some people fail to repay their loans, and as a result, they lose money. As a result, there is a significant risk connected with making a loan sanction decision. This presentation proposes a new framework for loan prediction based on multi-relational financial data from the PKDD'99 conference.

Sr. No	Title	Accuracy	Publication	Method
1	Credit Risk Prediction Based on Machine Learning Methods	79.2%	2019	XGBoost algorithm with logistic regression
2	A study on predicting loan default based on the random forest algorithm	81%	2019	Random forest algorithm
3	Prediction of Loan Behaviour with Machine Learning Models for Secure Banking	83.33%	2019	DNN
4	Default prediction in P2P lending from high-dimensional data based on machine learning	79.2%	2021	Gradient boosting decision trees (GBDT), extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM)
5	A Comparative Assessment of Credit Risk Model Based on Machine Learning——a case study of bank loan data	81%	2019	Random Forest Algorithm
6	A benchmark of machine learning approaches for credit score prediction	83.33%	2021	DNN
7	Predictions of Loan Defaulter - A Data Science Perspective	79.2%	2020 IEEE	Logistic Regression, Random Forest, KNN, LDA
8	Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process	85%	2019	SVM, Decision Tree, Naïve Bayes
9	Forecasting recovery rates on non-performing loans with machine learning	67.53%	2020	Forecasting Recovery

V. PROBLEM STATEMENT

Numerous researches in the fields of finance and banking have been conducted using data mining approaches to anticipate loan risk management. Predict loan approval using Machine Learning Algorithm (DNN) and know if an applicant is eligible to get a loan or not. The goal of loan default prediction is to anticipate whether or not the borrower will miss a payment. This is a significant issue for banking and finance firms. The model established thus will provide a better loan risk assessment, which could lead to a better capital allocation for the bank.

VI. PROPOSED SYSTEM

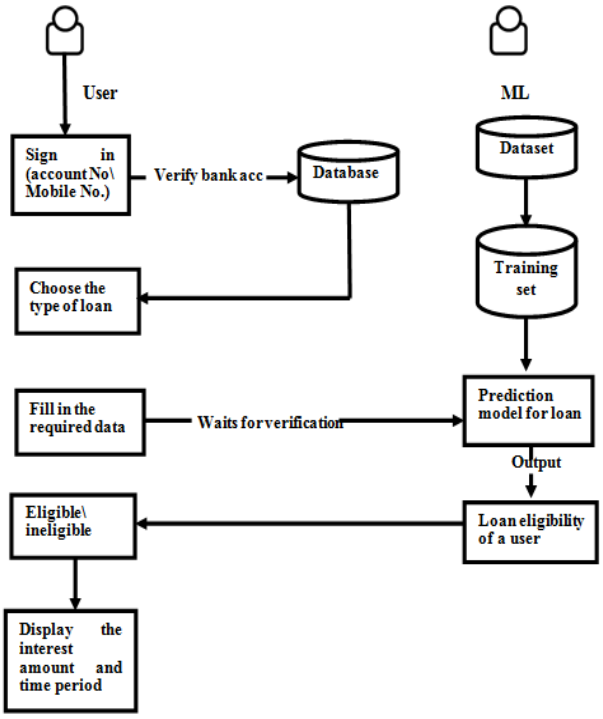

Figure 6.1: block diagram of system

Working of proposed system:

The proposed model spotlights on foreseeing client validity for credit reimbursements by examining their way of behaving. The contribution for the model is the gathered client conduct which incorporates boundaries, for example, candidate pay, co-candidate pay, orientation, wards, instruction, advance sum, credit add up to term, and so forth Depending on the result of the categorizer, you can choose whether to support or deny the client's solicitation. The DNN calculation is utilized to produce the applicable properties and furthermore settle on the choice in the model. On the off chance that the model supports the solicitation, the client will see the interest on the credit sum as well as the term.

Problem Understanding: - The information mining model is started by gathering insights about the financial business and existing credit handling strategies. This gives a superior comprehension of the difficulties and key dangers related with credit endorsement/disavowal in banking.

Data Understanding: - In the information understanding stage, the financial informational collection of client subtleties, which is important for information mining, is gathered and acclimated. Different essential ascribes are likewise contemplated.

Data Filtering: - Ascribes from the bank informational collection are separated and pertinent characteristics required for guaging are chosen. Therefore, fragmented and uproarious records from the dataset are taken out and ready for extraction.

System Modeling: - Now, the framework is grown proficiently and naturally so that even clients with less specialized information can utilize it easily. The framework gives the most significant characteristics that assist with deciding if to endorse or dismiss the advance solicitation. to anticipate the validity of future clients.

System Evaluation: - In the last stage, the planned framework is tried with test sets and the exhibition is guaranteed.

DNN

Deep learning (also known as deep structured learning) is part of a larger family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised. Deep learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks, and convolutional neural networks have been applied to areas such as vision computer-based, speech recognition, natural language processing, machine translation, bioinformatics, pharmaceutical product design, medical image analysis, climate science, materials inspection and board games, where they produced results comparable to, and in some cases superior to, the performance of human experts.

Artificial neural networks (ANN) were inspired by the distributed information processing and communication nodes in biological systems. Artificial neural networks have various differences from biological brains. Specifically, artificial neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog. The adjective "deep" in deep learning refers to the use of multiple layers in the network. Early work showed that a linear perceptron cannot be a universal classifier, but a network without polynomial activation function with a hidden layer of unlimited width can. Deep learning is a modern variant involving an unlimited number of layers of limited size, which allows practical application and optimized implementation, while maintaining theoretical universality under mild conditions. In deep learning, strata can also be heterogeneous and deviate widely from biologically informed connectionist models, for the sake of efficiency, formability and understandability, hence the “structured” part.

VII. RESULT AND DISCUSSION

Hence, we had implemented loan credibility prediction system that helps the organizations in making the right decision to approve or reject the loan request of the customers. This will definitely help the banking industry to open up efficient delivery channels. In this model, Logistic Regression algorithm is used for the prediction. Incorporation of other techniques that outperform the performance of popular data mining models has to be implemented and tested for the domain.

The DNN Classifier provided us with an accuracy of 95% which is way better than other algorithms like the Decision Tree, Random forest and Logistic regression. Hence, the DNN model appears to be a better option for such kind of data.

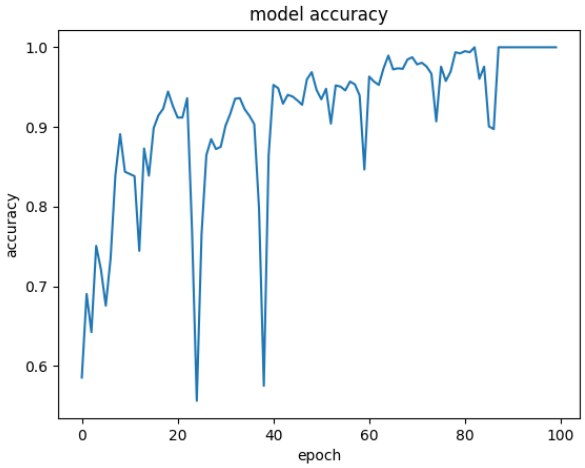

Figure 7.1: Accuracy Graph Using DNN

VIII. CONCLUSION

In our model, using the logistic regression model, we predict whether the loan is approved or not. To implement this, various input variables were used to get the output. Whenever the program takes input data, it provides the output as a binary i.e., 0 or 1. If the output is 1, '1' will be displayed, indicating the loan has been approved. If the output is 0, '0' will be displayed, indicating the loan has not been approved. Gave us 80% accuracy while the decision tree method gave us 73% accuracy. Therefore, DNN seems to be a better option for this type of data. As the algorithm places some of the non-defaulters in the default class, we may want to dig deeper into this issue to help the model accurately predict capable borrowers.

REFERENCES

[1] Yu Li. "Credit Risk Prediction Based on Machine Learning Methods". The 14th International Conference on Computer Science & Education (ICCSE 2019) August 19-21, 2019. Toronto, Canada

[2] Lin Zhua, Dafeng Qiua, Daji Ergua,*, Cai Yinga, Kuiyi Liu. "A study on predicting loan default based on the random forest algorithm". 7th International Conference on Information Technology and Quantitative Management (ITQM 2019)

[3] Mayank Anand1*, Arun Velu2, Pawan Whig. "Prediction of Loan Behaviour with Machine Learning Models for Secure Banking". Journal of Computer Science an Engineering (JCSE) e-ISSN 2721-0251 Vol. 3, No. 1, February 2022.

[4] Jing Zhou a, Wei Li b,c,d, Jiaxin Wangc, Shuai Ding c,d, Chengyi Xia. "Default prediction in P2P lending from high-dimensional data based on machine learning". 0378-4371/© 2019 Elsevier B.V. All rights reserved.

[5] Yuelin Wanga, Yihan Zhanga, Yan Lua, Xinran Yua. "A Comparative Assessment of Credit Risk Model Based on Machine Learning——a case study of bank loan data" 2019 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2019)

[6] Vincenzo Moscato, Antonio Picariello, Giancarlo Sperlí. "A benchmark of machine learning approaches for credit score prediction". Expert Systems With Applications 165 (2021) 113986

[7] P. Maheswari, CH. V. Narayana." Predictions of Loan Defaulter - A Data Science Perspective" 978-1-7281-9180-5/20/$31.00 ©2020 IEEE

[8] E. Chandra Blessie, R. Rekha." Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November, 2019

[9] Anthony Bellotti a,b, Damiano Brigo a, Paolo Gambetti c,∗, Frédéric Vrins. "Forecasting recovery rates on non-performing loans with machine learning" 0169-2070/© 2020 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

[10] I O Eweoya1, A Adebiyi1,2, A Azeta1 and Angela E Azeta. "Fraud prediction in bank loan administration using decision tree" 3rd International Conference on Science and Sustainable Development (ICSSD 2019)

[11] Omprakash Yadav#1 Chandan Soni#2 , Saikumar Kandakatla#3 , Shantanu Sawant." Loan Prediction System Using Decision Tree" IEEE 2019.

[12] Hafiz Ilyas Tariq, Asim Sohail, Uzair Aslam, and Nowshath Kadhar Batcha." Loan Default Prediction Model Using Sample, Explore, Modify, Model, and Assess (SEMMA)" 2019.

[13] ya-qi chen, jianjun zhang, wing w. y. ng." loan default prediction using diversified sensitivity undersampling" ieee 2018.

[14] Srishti Srivastava, Ayush Garg, Arpit Sehgal & Ashok Kumar." analysis and comparison of loan sanction prediction model using python" ieee 2018.

[15] Dr. P.G Sivagaminathan 1, Dr. C.R.Vijayalakshmi 2, Dr.M.Thangaraj ." A New Framework for Loan Prediction Using Multi relational Fuzzy Classifier" International Journal of Scientific Research in Computer Science Applications and Management Studies, 2018.