Prediction of Modernized Loan Approval System Based on Machine Learning Approach

Mr. S. J Vivekanandan[1] , Mrs. K. Arul Jothi[2], Kalavakuru Sujareddy[3],
Sangana Praveena[4], Dandu Roshini[5]

[1,2]Assistant Professor, Dhanalakshmi College of Engineering, Chennai, India
[3,4,5]Students, Dhanalakshmi College of Engineering, Chennai, India

Abstract: *To determine and to understand the working of loan systems for the cause of Loan Prediction using the demographic information of various factors that combine to form the nature of the approval using algorithms and concepts of Machine Learning and ultimately deploying this model on Cloud Based Platforms. Machine learning being aided by Cloud services are progressively seeing immense growth in the industry as they have benefits of Scalability, Affordability and easy use of models on systems as and when required. Therefore datasets are designed, automated and put under testing and training. The major aim of this project is to predict which of the customers will have their loan paid or not using prominent algorithms like Decision Tree, Logistic Regression and Random Forest. Logistic Regression Confusion matrix analysis is relatively in accordance to Decision Tree and Random Forest algorithm helping us attain an accuracy of 86% with minimum error.*

Keywords: Loan Prediction, Machine Learning, Cloud services, testing and Training, Logistic Regression.

I. INTRODUCTION

The impact of emerging computing techniques together with an increasing dimension of large datasets and the availability of more and more performing and accessible computing resources is transforming many research areas [1]. A huge amount of our lives is dominated and driven by machine learning algorithms. These algorithms are now used in nearly every field to build models that can predict future events with an compelling level of reliability [2], [3]. Most of those algorithms and models are cloud-based.

Artificial intelligence and machine learning algorithms [4], [5] are the forces driving social media, marketing, customer support, fraud detection, business intelligence, and pretty much every movie and music recommendation we see every day.

We're collecting and accumulating data at massive and unmanageable rates. Businesses collect website clicks, social media interactions, credit card transactions, GPS trails, and on and on and on. But it's almost impossible for most companies to process all of this information and use it in any kind of meaningful way. For a long time, machine learning models [6] were simply out of reach for most businesses. The costs alone made them prohibitive. Even if a business could afford to implement one, it probably didn't have anyone on hand who could design a model and interpret the results. Cloud-based machine learning solutions have changed that. Relatively speaking, they're cheap to operate and often come with prebuilt solutions to complex problems.

One of the biggest advantages of cloud-based machine learning [7], [8] is that it gives organizations access to highperformance infrastructure that they couldn't afford (or properly use) on their own. ML applications require a ton of processing power. That's traditionally been very expensive! Now, many organizations use systems that rely on GPUs to handle ML workloads. It's much more affordable to rent access to these systems in the cloud than to purchase them outright.

A. Objectives and Methodology

The idea of this project is to construct understand and implement the workings of a machine learning application on a Cloud based platform. In this project, different common language preparing strategies and AI calculation to group

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/568

ISSN
2581-9429
IJARSCT

225

credit information utilizing the library from python are utilized and is actualized utilizing python programming language. Perusing the train, test and approval information records and playing out some pre-handling. Preparing information will be pre-handled and includes are closed. These pre-handled preparing information and highlights are exposed to AI calculation, from which the best appropriate calculation is utilized to fabricate the model like choice tree, strategic relapse. Last characterization model is manufacture and likelihood of bogus and truth can be identified in the yield.

II. LITERATURE REVIEW

There are numerous fields where data mining can be used in financial industries such as customer segmentation and profitability, applicants for high-risk loans, predicting default payments, promotions, collateral monitoring, rating assets, irregular sales, managing stock holdings, cash management and estimating activities, most profitable clients with credit cards and cross-selling [9]. Data mining plays a vital role in the management of transaction data and customer profile in banking. From that, a user can make an accurate decision using data mining techniques [10]. A data mining technique produces a lot of patterns and rules, typically. Among those rules created should be selected beneficial and interesting [11]. One great example is the credit scoring which aims to divide the applicants into two categories-good credit applicants and poor credit applicants. The former class has excellent chances of repaying financial commitments, and the others have high defaulting risks [12]. At the same time, the return engagement with the right judgment about credit risk is high. It would also be particularly beneficial to see some progress in creating a credible distinction between those who are willing to repay the loan and those who are not [13]. The general principle of credit appraisal is to associate a customer's attributes or qualities with other past borrowers, whose debts they have already paid off. And credit rating is also used to evaluate a list of previous clients and differentiate current and potential credit clients [14]. People who cannot pay back the amount of their loan are considered defaulters. Unforeseen economic conditions may increase the number of defaulters for such years, which in turn may increase the financial institutions' losses. This has a negative impact not only on the institutes but also on the credit record of the client and potential financial stability. Given the increase in default rates, keeping track of the vulnerability of default is critical for both the good of banks and customers [15].

One of the common tools used today in data mining is Weka. The Weka facilitates these traditional data mining functions in the pre-processing, classification, clustering, association, regression and feature selection phases [16]. It also provides a visualization tool [17]. And it is useful in learning basic machine learning principles with various options and in analyzing the output being generated [18].

Classification algorithm should also be considered, and it is used to forecast any result depends on a given data. The algorithm generates a training set that includes a collection of attributes and the corresponding outcome, usually referred to as the target or prediction attribute. The algorithm aims to discover relationships between the attributes that would require predicting the result [19].

III. ARCHITECTURE OF LOAN PREDICTION

The architecture of loan prediction model is shown in figure

It consists of following main blocks.

- INPUT: The principle feature of this credit validity *B.* forecast framework is that it utilizes choice tree, calculated relapse enlistment information mining calculation to sift through the advance solicitations. A choice tree is created by preforming information mining on a current bank dataset containing 4520 records and 17 properties.

- DATA PREPROCESSING: At first the ascribes which are basic to make a credit believability forecast is related to data gain as the property evaluator and ranker as the hunt technique. Manual pre-handling is likewise performed.

- DATA FILTERING: Last dataset after pre-handling is partitioned so that there is 66% preparing set and 34% test set. Test set is utilized to approve the conclusive outcome of the performed.

- DECISION TREE ALGORITHM: [20], [21] A productive choice tree is detailed with choice tree enlistment calculation. It delivers a model with the most pertinent 6 property. Characteristic with rank-1 is put as the root

hub of the choice tree, different credits from rank-2 to rank-6 comprise the moderate hubs. A choice is made at every hub and the leaf hub gives us the conclusive outcome. That is, on the off chance that the client have the base advance reimbursement limit, at that point the future dangers can be maintained a strategic distance from. The principle advantage of applying information mining is that we can generally depend on the aftereffect of the calculation to acknowledge or dismiss the advance application.

- LOGISTIC REGRESSION: [22]–[24] This is a measurable model that in its essential structure utilizes a strategic capacity to show a double reliant variable, albeit a lot increasingly complex expansion exists. In relapse examination, strategic relapse is assessing the parameters of a calculated model.
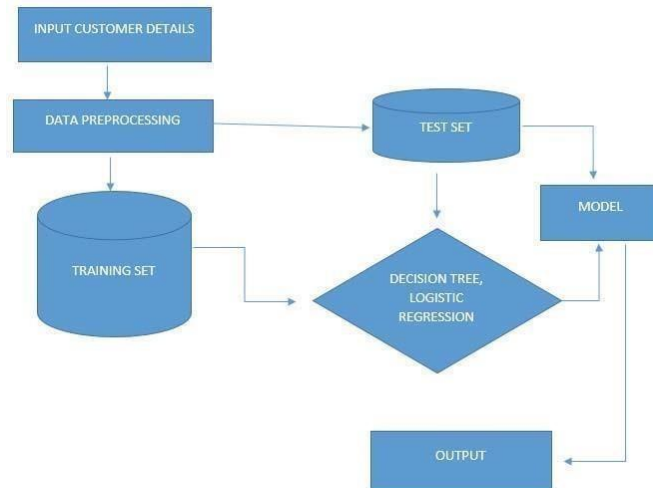

Fig. 1. Architecture of loan Prediction

A Flowchart

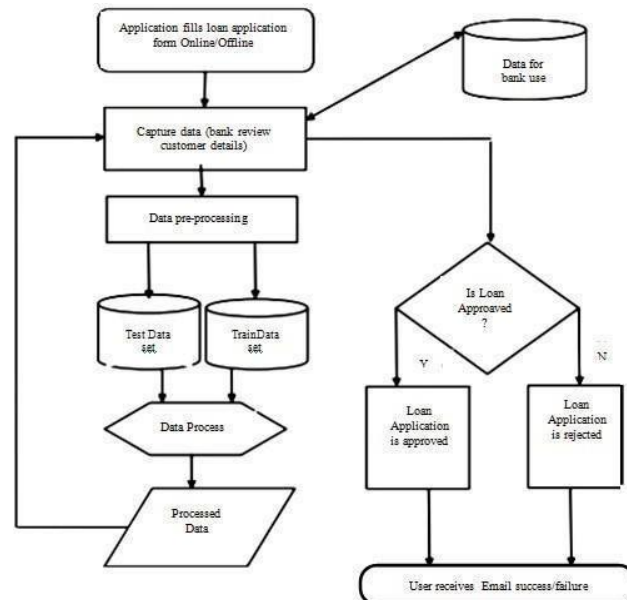Flowchart of the loan prediction model is shown in figure 2

Fig.2. Flowchart of loan Prediction model

Algorithm

Step1 : Start

Step2 : Recording the loan data(ID, married, loan amount, education Etc.)

Step3: Collecting the dataset from the loan recorded. Step4: the data is trained.

Step 5: Decision tree (taken to predict the target node), then from the target node we are predicting the loan approval based on the logistic regression (this is a basic form uses a logistic function to model a binary dependent variable).

Step 6 : the data is tested

IV. RESULTS AND DISCUSSION

In this project we are dividing the given data into train and test data. Using machine learning algorithm like decision tree and logistic regression we are training the data. Based on the train data, the test data is tested, were this predicts the loan status for all the given dataset

Based on the accuracy score the model [25] is build in such a way that it predicts for the individual data. The output is in the form of 0 and 1(binary form), were '1' indicates that loan is approved and '0' for loan rejected

```
nb 0.789130 0.036441
0.8246753246753247
[[ 23  25]
 [  2 104]]
              precision    recall  f1-score   support

           0       0.92      0.48      0.63        48
           1       0.81      0.98      0.89       106

    accuracy                           0.82       154
   macro avg       0.86      0.73      0.76       154
weighted avg       0.84      0.82      0.81       154
```
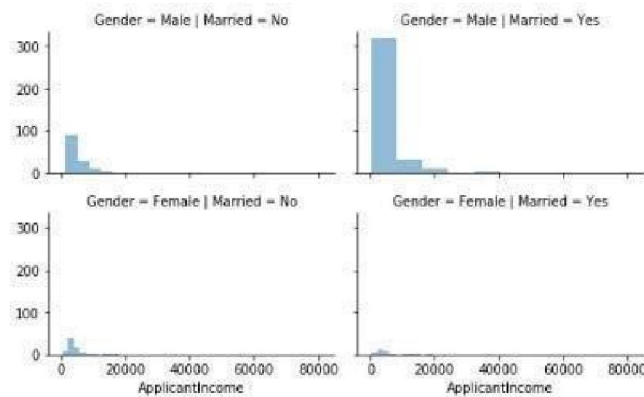
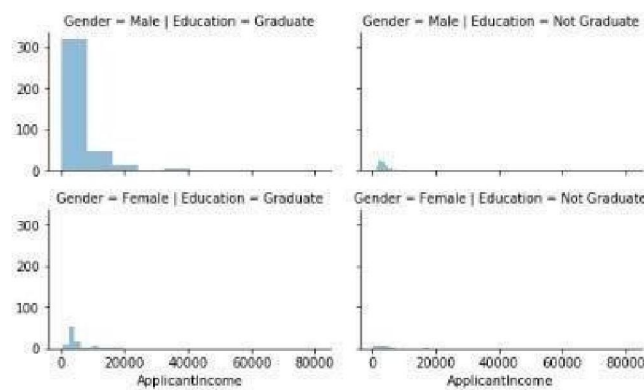Fig. 3. Accuracy score

Fig.4.Graph betweeen Gender and Married

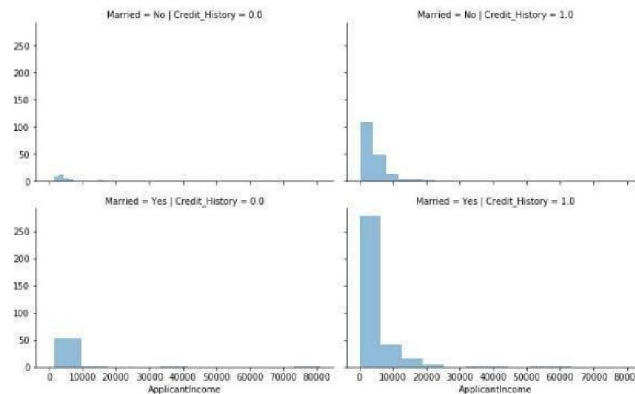Fig. 5. Graph between Gender and Education
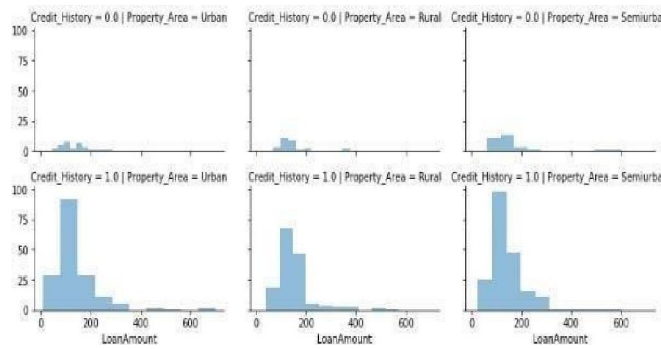
Fig. 6. Graph between Married and Credit_History

Fig. 7. Graph between Credit_History and Property_Area

Fig. 8. Loan Status for all the Dataset

Fig. 9. Loan Status for an Individual Data

V. CONCLUSION

This application can help banks in anticipating the fate of credit and its status and relies upon that they can make a move in introductory long periods of advance. Utilizing this application banks can diminish the quantity of awful advances from bringing about cut off misfortunes. A few AI calculations and bundles were utilized to set up the information and to fabricate the arrangement model. AI bundle libraries help in fruitful information examination and highlight determination. Utilizing this technique bank can without much of a stretch distinguish the necessary data from immense measure of informational collections and aides in fruitful advance forecast to diminish the quantity of awful credit issues. Information mining strategies are helpful to the financial part for better focusing on and procuring new clients, most significant client maintenance, programmed credit endorsement which is utilized for extortion avoidance,

misrepresentation identification progressively, giving section based item, investigation of the client, exchange designs after some time for better maintenance and relationship, hazard the executives and showcasing.

ACKNOWLEDGEMENT

REFERENCES

[1]. Á. López García et al., " A Cloud - Based Framework for Machine Learning Workloads and Applications, " in IEEE Access, vol. 8, p. p. 18681 - 18692, 2020, doi: 10.1109/ACCESS.2020.2964386.

[2]. Websiteurl:https://towardsdatascience.com/simple-way-todeploymachine-learning-models-to-cloud-fd58b771fdcf , accessed on 20th sept., 2020

[3]. Website url: https://info.cloudquant.com/2019/07/aiml20190715, accessed on 21st sept., 2020

[4]. Ongsulee, Pariwat. " Artificial intelligence, machine learning and deep learning. " 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE). IEEE, 2017.

[5]. Tyagi, Amit Kumar, and Poonam Chahal. " Artificial Intelligence and Machine Learning Algorithms." Challenges and Applications for Implementing Machine Learning in Computer Vision. IGI Global, 2020. 188 - 219.

[6]. https://cloud.google.com/ai-platform/prediction/docs/deploying-models, accessed on 22nd sept., 2020

[7]. Pakdel, Rezvan, and John Herbert. " Adaptive Cost Efficient Framework for Cloud - Based Machine Learning." 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). Vol. 2. IEEE, 2017.

[8]. Miller, Shane, Kevin Curran, and Tom Lunney. " Cloud - based machine learning for the detection of anonymous web proxies." 2016 27th Irish Signals and Systems Conference (ISSC). IEEE, 2016.

[9]. Navada, Arundhati, et al. " Overview of use of decision tree algorithms in machine learning." 2011 IEEE control and system graduate research colloquium. IEEE, 2011.

[10]. Hamid, Aboobyda & Ahmed, Tarig. (2016). " Developing Prediction Model of Loan Risk in Banks Using Data Mining. Machine Learning and Applications". An International Journal. 3. 1 - 9.10.5121/mlaij.2016.3101.

[11]. Preethi, M. (2017). " Data Mining In Banking Sector " Ikizler, Nazli & Güvenir, Halil Altay. (2001). " Mining Interesting Rules in Bank Loans Data "

[12]. Chen, Weimin & Xiang, Guocheng & Liu, Youjin & Wang, Kexi. (2012). " Credit risk Evaluation by hybrid data mining technique ". Systems Engineering Procedia. 3. 194 -200.10.1016/j.sepro.2011.10.029.

[13]. Zurada J. (2002) " Data Mining Techniques in Predicting Default Rates on Customer Loans ". In: Haav HM., Kalja A. (eds) Databases and Information Systems II. Springer,Dordrecht.

[14]. Dzelihodzic, Adnan & Donko, Dzenana. (2013). " Data Mining Techniques for Credit Risk Assessment Task ".

[15]. Akcura, Korhan & Chhibber, Appan. (2018). " Design and Comparison of Data Mining Techniques for Predicting Probability of Default on a Loan"

[16]. S., Singaravelan & Arun, R. & Shunmugam, D. & Soundar, K. & Mayakrishnan, R. & Murugan, D.. (2018). " ANALYSIS OF CLASSIFICATION ALGORITHMS ON DIFFERENT DATASETS ". Review of Innovation and Competitiveness. 4. 41 - 54. 10.32728/ric.2018.42/3.

[17]. Jothikumar, R. & Sivabalan, R.V. & Sivarajan, E.. (2015). " Accuracies of J48 weka classifier with different supervised weka filters for predicting heart diseases ". 10. 7788 - 7793.

[18]. Desai, Aaditya & Rai, Sunil. (2013). " Analysis of Machine Learning Algorithms using Weka ".

[19]. Shivangi Gupta & Neeta Verma. (2016). Comparative Analysis of classification Algorithms using WEKA tool.International Journal of Scientific & Engineering Research, Volume 7, Issue 8, August 2016. ISSN 2229 - 5518

[20]. Ranka, Sanjay, and V. Singh. " CLOUDS: A decision tree classifier for large datasets." Proceedings of the 4th Knowledge Discovery and Data Mining Conference. Vol. 2. No. 8. 1998.

[21]. Zhu, Xu Dong, Hui Li, and Feng Hua Li. " Privacy - preserving logistic regression outsourcing in cloud computing." International Journal of Grid and Utility Computing 4.2 - 3 (2013): 144 - 150.

[22]. R. Ramesh, " Predictive analytics for banking user data using AWS Machine Learning cloud service," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, 2017,

[23]. p. p. 210 - 215, doi: 10.1109/ICCCT2.2017.7972282.

[24]. Motwani, A., P. Chaurasiya, and G. Bajaj. " Predicting Credit Worthiness of Bank Customer with Machine Learning Over Cloud." International journal of computer sciences and engineering 6 (2018): 7.

[25]. Ravuri, Suman, and Oriol Vinyals. " Classification accuracy score for conditional generative models." Advances in Neural Information Processing Systems. 2019.

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/568

ISSN
2581-9429
IJARSCT

231