Loan Prediciton Using Machine Learning Algorithm

[1]Tharun Puppala, [2]Sohan Reddy Baddam, [3]Akhil R, [4]Sandeep Reddy

[1,2,3,4]Bharath institute of higher education and research

Abstract: Technology has improved the existence of humankind and the way of life they live. Each and every day we are planning to create something new and different. This project makes to filter the list of people who are eligible for taking the loans with clean background. There are positive bank rules bank must keep in mind while selecting a candidate for mutual approval, based on these parameters the bank employee have to decide which person should be tested maximum. In this project we are going to propose a application based software that used to know the complete approach that predicts a person who is completely and certified to get a loan based on his previous file of having clean records of paying the interest amount or the loan amount within the time. The main purpose is this project is to predict whether a person is eligible to take a loan according to this background verification.

1. INTRODUCTION

As fact is increasing every day because of digitization in the banking process, humans need to apply for on line loans. Artificial intelligence (AI) as an average approach of information studies is attracting an increasing number of interest. People from numerous agencies are using AI computing to remedy problems related to their data enterprise. Banks face a first-rate challenge in mutual approval. Every day there are such a lot of applications that it is difficult to deal with banking services, and there is also a excessive possibility of some mistakes. Most banks make a make the most of loans, however deciding on sincere clients from packages is volatile. One mistake can reason a big loss to the bank. Lending is the center commercial enterprise of almost all banks. This program pursuits to offer loans to all eligible applicants. An efficient and streamlined machine that reduces the financial institution's time entails checking every applicant on a concern basis. The financial institution management well timed fulfills all of the different formalities of the patron, which has a nice effect on the clients. The exceptional part is that it is effective for each the bank and the candidates. This device lets in you to ship specific applications that deserve approval on a concern foundation. There are some predictive functions which include Gender, Spouse, Dependents, Education, Employment, Applicant Income, Co-Applicant Income, Loan Amount, Loan_Amount_Term,Credit_History,"Property_area", "Loan_status".

2. LITERATURE SURVEY

In recent years, there has been a lot of interest in using machine learning to estimate medical insurance premiums. By using massive datasets and complicated models, machine learning techniques provide a potential strategy for reliably predicting medical insurance prices. Many research have been undertaken to explore the usefulness of machine learning models for predicting medical insurance premium.

[1] The objective of recent year research is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing. They used these Algorithms Decision Trees, Random Forest, Support Vector Machine (SVM), Linear Models (LM), Neural Network and Adaboost. [2]This is achieved by feeding the past records of the customer who acquired loans from the bank into a trained machine learning model which could yield an accurate result. The prime focus of the paper is to determine whether or not it will be safe to allocate the loan to a particular person. This paper has the follwing sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting**.** They used these algorithms Logistic Regression, SVM, Decision Tree, Naïve Bayes.[3]The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting. They used these Algorithms Logistic Regression, Decision Tree, Random Forest, XGBoost.[4]The main focus of this research is to determine whether the loan given to a particular person or an organization shall be approved or not.They used these Algorithms Decision tree, Random forest. [5]The proposed work in data mining focused on using data from banking sector to predict the status of loans. Two classification algorithms Naïve Bayes and Support Vector Machine are discussed here. The results have been proven that the speed and accuracy has been improved when combined these two techniques. They used these Algorithms Naïve Bayes and Support Vector Machine

3. SYSTEM ANALYSIS
3.1 EXISITNG SYSTEM

"Adyan Nur Alfiatin, Hilman Taufik and their project are running on predicting residence expenses. They use regression analysis and particle system optimization (PSO) to expect home fee. Another comparable work turned into overseen through Mohamed El Mohadab, Belaid Bouikhalene and Said Safi for a research paper pronouncing the usage of scholarship. Kumar Arun, Garg Ishan and Kaur Sanmeet are operating at the financial institution mortgage forecast in which the financial institution will approve the loan.

Y. Shi and P. Song proposed a method for evaluating project loans using risk analysis. The method evaluate the risk involved in loans of commercial banks.R. ZhangandD. 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. Franket al. used machine learning in prediction of smoking status. Different machine learning

approaches were applied and investigated for finding the smoking status. From the results it was ensured that logistic algorithm performs better. R. Lopeset 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.

3.2 PROPOSED SYSTEM

Two incidents, it needs to be a little homework. System of authorized or rejected loan applications. Credit reimbursement is the primary parameter that influences the economic statements of the bank. It could be very difficult to expect whether or not a patron will repay the mortgage. Recently, many researchers have worked on the aforementioned techniques of loan approval. Machine getting to know (ML) techniques are very useful for predicting consequences for huge amounts of records. In this text, device learning algorithms are used to are expecting client pride. Experimental effects show that the accuracy of the gadget getting to know set of rules is better as compared to the logistic regression and random forest gadget mastering tactics.

This proposed model will characterize the behavior of customers on the Basis of their record. These records are taken from the customers, and create a data set. With the help of these data sets and training machine learning model, we predict that the customer's loan will pass or not.

Here we can use the Support Vector Machine Logistic Regression, Decision Tree Machine Learning Algorithms for doing this project.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.

4. IMPLEMENTATION

This system implementation starts with the Training and testing the data, the date collected will be trained and tested in this process and then the trained data will send to the future process that is known as Data cleaning and processing here that trained data gets completely processed according to the type of data given after completion of the process machine learning algorithm will be applied, the algorithm used here are Random forest algorithm (RFA), Support vector machine (SVM) and linear regression is applied for the data, and then the best data will be sent for further process and it initiates to predict the best model after the prediction gets completed according to the input details the result will be provided based on the applied algorithm for the trained data.
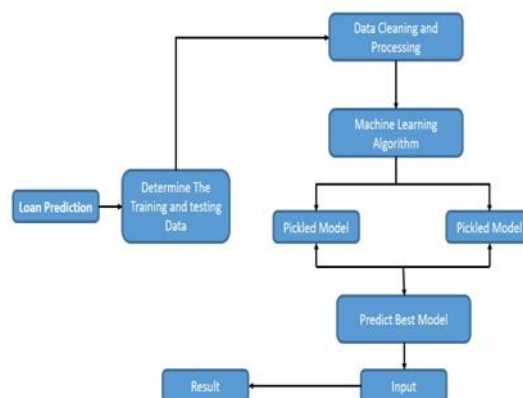

Fig. 4.1 Implementation

4.1 Training and testing the data

Firstly we enter in to our website then we get logged using credentials, then after the data is to be uploaded as show in the below diagram.
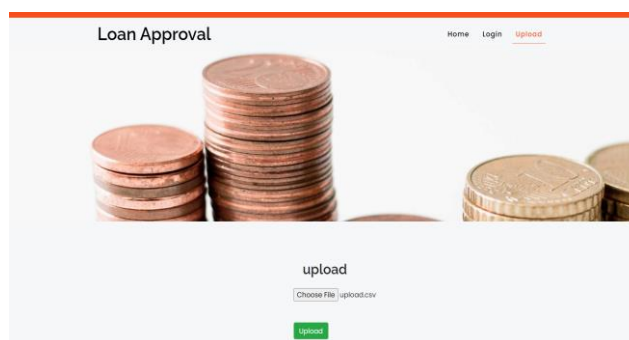

Fig. 4.2 uploading dataset

Then the data is uploaded and get trained using he machine learning model, updating the data using a html button in the website as show in below.


Fig. 4.3 Training the dataset

5. Results

Here are our results firstly the admin can enter the details about the person, then using machine learning algorithms the prediction is calculated. The approval rate and non approval rate of the person.


Confuison matrix is calculated by using a formula,the formula is precision = TP/TP + FP,for recall = TP/TP + FN
Where TP is True Positive, FP is False Positive and FN is False Negative
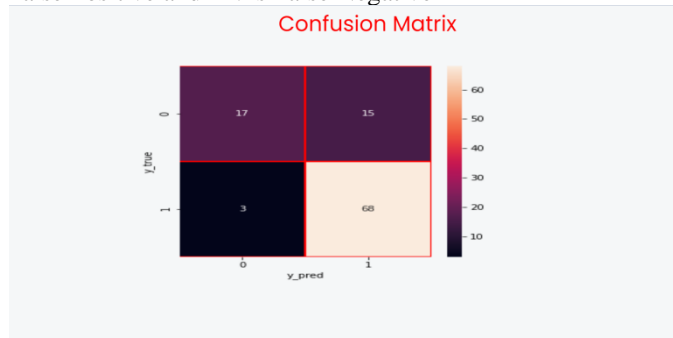
6. SYSTEM ARCHITECTURE

The description of the not unusual capabilities of this system has a deep meaning for the definition of the requirements and the installation. In the architectural layout, the diverse pages and their relationships are diagnosed and designed. Major software program components are identified and broken down into processing techniques and conceptual data systems, and relationships among modules are diagnosed. The proposed machine consists of those modules.

Firstly applicant will fill the loan application form in online or offline then the data is captured and customers data will be reviews then this details are required for bank use . The data which we have captured that will be Pre-processed and it is nothing but it is divided Train data and test data set so these both data will be collected and processed the final processed data will capture the review of total process and then it will analysis and predict whether the loan is approved or rejected then this decision will be sent Via email
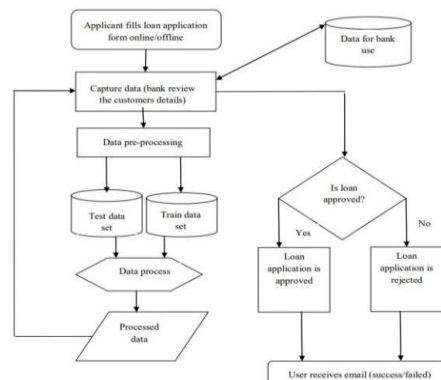
7. CONCLUSION

From a proper analysis of positive points and constraints on the member, it can be safely concluded that the product is a consider. This use is working duly and meeting to all Banker requisites. This member can be freely plugged in numerous other systems. There have been mathematics cases of computer glitches, violations in content and most important weight of features is fixed in automated prophecy system, so in the near future the so – called software could be made more secure, trustworthy and dynamic weight conformation. In near future this module of prophecy can be integrated with the module of automated processing system. The system is trained on old training dataset in future software can be made resembling that new testing date should also take part in training data after some fix time

8. REFERENCES

1. Kumar Arun, Garg Ishan, Kaur Sanmeet, May-Jun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE)
2. Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent DecisionMaking, Ministry of Education, Hefei University of Technology, Hefei 2009, China
3. Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C.Sealand on july 20, 2018.
4. Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, Dr. K. Kavitha, International Journal of Advanced Research in Computer Science and Software Engineering.
5. Amruta S. Aphale and R. Prof. Dr. Sandeep. R Shinde, "Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval", International Journal of Engineering Trends and Applications (IJETA), vol. 9, issue 8, 2020)
6. Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016
7. Exploratory data analysis https://en.wikipedia.org/wiki/Exploratory_data_analysis
8. PandasLibraryhttps://pandas.pydata.org/pandas-docs/stable/
9. MeanDecreaseAccuracy https://dinsdalelab.sdsu.edu/metag.stats/code/randomforest.html
10. J. R. Quinlan. Induction of Decision Tree. Machine Learning, Vol. 1, No. 1. pp. 81-106., 1086.
11. A. Goyal and R. Kaur, "A survey on Ensemble Model for Loan Prediction", International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.
12. G. Shaath, "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R".
13. A. Goyal and R. Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models".
14. Hsieh, N. C., & Hung, L. P. (2010). A data driven ensemble classifier for credit scoring analysis. Expert systems with Applications, 37(1), 534-545.
15. https://en.wikipedia.org/wiki/Exploratory_data_analysis