LOAN APPROVAL SYSTEM USING MACHINE LEARNING

Pratiksha Bhosle, Amruta Pawar, Priti Patil, Priya Jogdankar,
Prof. Supriya More
Information Technology
Parvatibai Genba Sopanrao Moze College of Engineering, Wagholi, Pune

***Abstract*—In today's world, taking loans from financial institutions has become a very common phenomenon. Everyday a large number of people make application for loans, for a variety of purposes. But all these applicants are not reliable and everyone cannot be approved. Every year, we read about a number of cases where people do not repay bulk of the loan amount to the banks due to which they suffers huge losses. The risk associated with making a decision on loan approval is immense. So the idea of this project is to gather loan data from multiple data sources and use various machine learning algorithms on this data to extract important information. This model can be used by the organizations in making the right decision to approve or reject the loan request of the customers. In this paper, we examine a real bank credit data and conduct several machine learning algorithms on the data for that determine credit worthiness of customers in order to formulate bank risk automated system**

Keywords*— *Machine learning, bank credit, classification, confusion matrix, predictive analysis

I. IN TROD UCTION

Bank plays a vital role in market economy. The success or failure of organization largely depends on the industry's ability to evaluate credit risk. Before giving the credit loan to borrowers, bank decides whether the borrower is bad (defaulter) or good (non defaulter) The prediction of borrower status i.e. in future borrower will be defaulter or non defaulter is a challenging task for any organization or bank. Basically the loan defaulter prediction is a binary classification problem Loan amount; costumer's history governs his credit ability for receiving loan. The problem is to classify borrower as defaulter or non defaulter. However developing such a model is a very challenging task due to increasing in demands for loans. Prototypes of the model which can be used by the organizations for making the correct or right decision for approve or reject the request for loan of the customers. This work includes the construction of an ensemble model by combining different machine learning models. Banks struggle a lot to get an upper hand over each other to enhance overall business due to tight competition. Credit Risk assessment is a crucial issue faced by Banks nowadays which helps them to evaluate if a loan applicant can be a defaulter at a later stage so that they can go ahead and grant the loan or not. This helps the banks to minimize the possible losses and can increase the volume of credits Labeled data is known. in the literature

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 some special advantages to the bank

II. LITERATURE REVIEW

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers.

Amira Kamil Ibrahim Hassan, Ajith Abraham (2008)
[1] uses a prediction model which is constructed using three different training algorithms to train a supervised twolayer feed-forward network. The results show that the training algorithm improves the design of loan default prediction model.

Angelini (2008)
[2] used a neural network with standard topology and a feed-forward neural network with ad hoc connections. Neural network can be used for prediction model. This paper shows that the above two models give optimum results with less error.

Sarwesh Site, Dr. Sadhna K.Mishra (2013)
[3] proposed a method in which two or more classifiers are combined together to produce an ensemble model for the better prediction. They used the bagging and boosting techniques and then used random forest technique.

Akkoç (2012)
[4] used a model namely hybrid Adaptive Neuro-Fuzzy Inference model, grouping of statistics and Neuro-Fuzzy network. A 10-fold cross validation is used for better results and a comparison with other models.

Alaraj M, Abbod M (2015)
[5] introduced a model that are based on homogenous and heterogeneous classifiers. Ensemble model based on three classifiers that are logistic artificial neural network, logistic regression and support vector machine.

to be suitable for classification algorithms. The dataset used in this paper is a labeled data and is, therefore, suitable for doing classification analysis. And thus, we employed various classification algorithms described comprehensively in Section II-B. Some of the algorithms are implemented in MATLAB §R and some taken from the Python scikit-learn package to predict the creditworthiness of bank customers with regards to their ability to pay their credit or otherwise within a given time frame.

III. METHODOLOGY

A. Naïve Bayes

This classification technique is based on Bayes' theorem that assumes independence between predictors, thus, the presence of a particular feature in a class is independence of another feature in a another class. Naive Bayes classification is therefore, based on estimating P(X Y), the probability or probability density of features X given class Y

B. KNN Predicted Loan Eligibility

It is a non-parametric method used for making predictions. In this, the predicted value is a class membership. The first step of the K-NN algorithm is to identify the k nearest neighbors for each incoming new instance. The instance is classified by a majority vote of these neighbors. In the second step, depending on the label sets of the k neighbors, a label is predicted for the new instance. The K-NN algorithm is a non-parametric algorithm, which means it makes no assumptions about the underlying results. It's also known as a lazy learner algorithm because it doesn't learn from the training set immediately instead, it saves the dataset and performs an operation on it when it comes time to classify it. During the training process, the KNN algorithm simply stores the dataset, and as it receives new data, it classifies it into a group that is somewhat close to the new data.

C. Linear Regression:

It is used to estimate real values based on continuous variable(s). In linear regression, a relationship is established between independent and dependent variables by fitting the best line. This best fit line is known as regression line and represented by a linear equation Y = a X + b, where Y is the dependent variable, a is the slope, X is the independent variable and b is the intercept. The coefficients a and b are derived based on minimizing the sum of squared difference of distance between data points and Resource Allocation Network. Their experimental results showed that the algorithm has a comparable performance as that of the original online sequential extreme learning machine but with a reduced number of hidden layers.

D. Discriminant Analysis:

The discriminant analysis is based on the assumption that different classes of data are generated by using different Gaussian distributions. The main types of discriminant algorithms used for classification are the linear and the quadratic discriminant. We used the quadratic discriminant classifier in this paper

E. Neural Networks

The neural network supports both classification and regression algorithms and therefore, is very appropriate for studying the classification problem in this paper

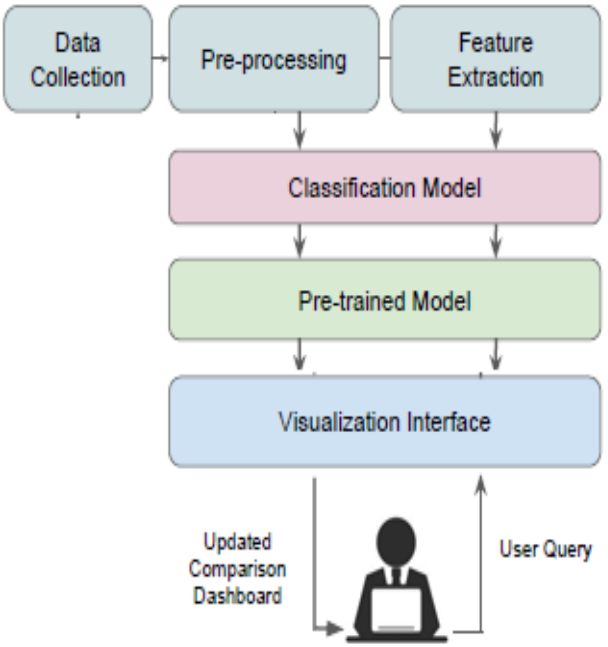
F. Proposed System

Attributes:
Our loan sanctioning process predicts whether the loan amount requested by the customer will be granted to him or not. To arrive at the conclusion we use the following attributes

i. Age
ii. Profession
iii. Total Income
iv. Existing Loan
v. Loan Tenure
vi. Loan Amount
vii. Loan Approved

Taking into account the values of these attributes, the system uses the Naive Bayes classifier to classify the given case into one of the pre-defined classes. Based on this classification we give the result as Yes (suggesting that loan will be sanctioned) or No (loan will not be sanctioned). Components of the System:
A. Pre-Processing
B. Classification
C. Database Updating

G. Architecture



IV. RESULTS

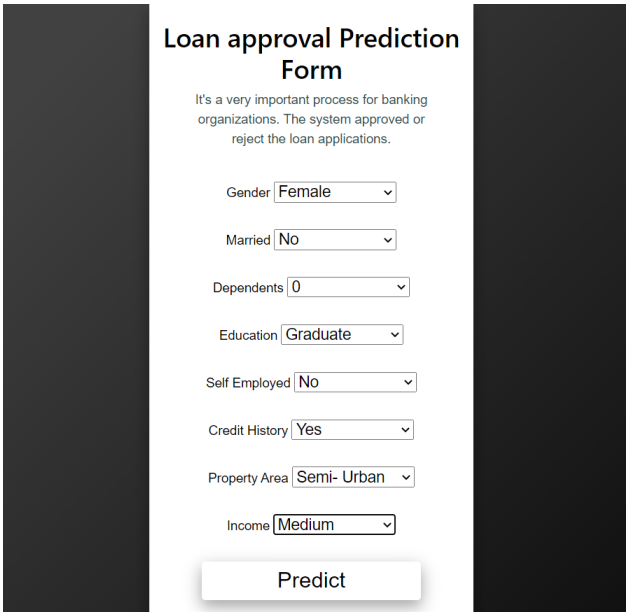


Fig.2 Entering Data sets

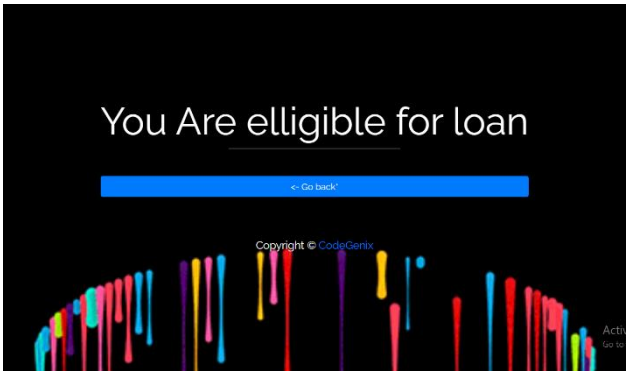


Fig.3 Predicted Loan Eligibility (approved)

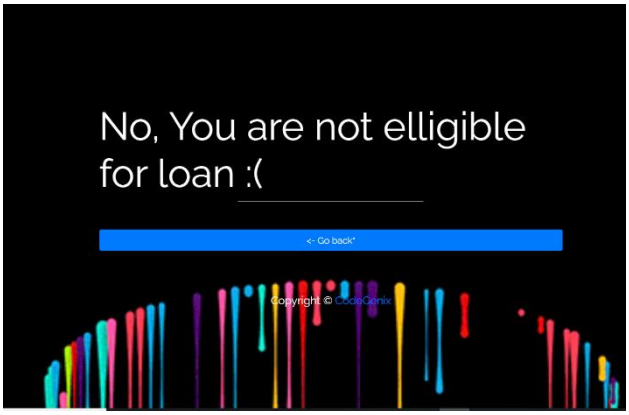


Fig. Predicted Loan Eligibility (Not Approved)

V. CONCLUSION

In this research paper, we have used machine learning algorithms to predict the eligibility of an applicant for loan. The data is pre-processed and fed to different regression models to determine the best model and with the help of both K-Fold Cross-Validation and different classification metrics we compared different model. So, according to above results we came to know that Logistic Regression is the most effective model with maximum accuracy and can be used as an effective model for predicting weather an applicant is eligible for loan or not, which should help banks to skip the tedious process of loan eligibility.

VI. Future Work

In future, this model can be used to compare various machine learning algorithm generated prediction models and the model which will give higher accuracy will be chosen as the prediction model. This paper work can be extended to higher level in future. Predictive model for loans that uses machine learning algorithms, where the result from each graph of the paper can be taken as individual criteria for the machine learning algorithm

VII. ACKNOWLEDGEMEN T

We take this opportunity to thank our Project Guide Prof. Supriya More and Head of the Department Prof. A. A. Shaikh for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of Parvatibai genba sopanrao moze college of Engineering, wagholi, Pune for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

VIII. REFER ENCES

[1] Amira Kamil Ibrahim Hassan and Ajith Abraham, ―Modeling Consumer Loan Default Prediction Using Ensemble Neural Networks‖, International Conference on Computing, Electrical and Electronics Engineering, pp. 719 – 724, August 2013.
[2] E. Angelini, A. Roli, and G. di Tollo, ―A neural network approach for credit risk evaluation‖ elsevier, The Quarterly Review of Economics and Finance, Vol. 48, Issue 4, pp. 733–755, November 2008.
[3] Sarwesh Site, Dr. Sadhna K. Mishra, ― A Review of Ensemble Technique for Improving Majority Voting for Classifier‖, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 1, pp. 177- 180, January 2013.
[4] S. Akkoç, ―An empirical comparison of conventional techniques, neural networks and the three stage hybrFId Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis : The case of Turkish credit card data‖, Elsevier Europezan