Report on loan approval prediction

Name: Naga Venkata Ashok

Contact: 9398995500

mail id: [akiashok987@gmail.com](mailto:akiashok987@gmail.com)

Abstract:

The cost of assets is increasing day by day and the capital required to purchase an entire asset is very high. So, purchasing it out of your savings is not possible. The easiest way to get the required funds is to apply for a loan. But taking a loan is a very time consuming process. The application has to go through a lot of stages and it’s still not necessary that it will be approved. To decrease the approval time and to decrease the risk associated with the loan many loan prediction models were introduced. The aim of this project was to compare the various Loan Prediction Models and show which is the best one with the least amount of error and could be used by banks in real world to predict if the loan should be approved or not taking the risk factor in mind. After comparing and analysing the models, it was found that the prediction model based on Random Forest proved to be the most accurate and fitting of them all. This can be useful in reducing the time and manpower required to approve loans and filter out the perfect candidates for providing loans.

Introduction:

A Prediction Model uses data mining, statistics and probability to forecast an outcome. Every model has some variables known as predictors that are likely to influence future results. The data that was collected from various resources then a statistical model is made. It can use a simple linear equation or a sophisticated neural network mapped using a complex software. As more data becomes available the model becomes more refined and the error decreases meaning then it’ll be able to predict with the least risk and consuming as less time as it can. The Prediction Model helps the banks by minimizing the risk associated with the loan approval system and helps the applicants by decreasing the time taken in the process.

The main objective of the Project is to compare the Loan Prediction Models made implemented using various algorithms and choose the best one out of them that can shorten the loan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person on the basis of various parameters like credit score, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters

In the present scenario, a loan needs to be approved manually by a representative of the bank which means that person will be responsible for whether the person is eligible for the loan or not and also calculating the risk associated with it. As it is done by a human it is a time consuming process and is susceptible to errors. If the loan is not repaid, then it accounts as a loss to the bank and banks earn most of their profits by the interest paid to them. If the banks lose too much money, then it will result in a banking crisis. These banking crisis affects the economy of the country. So it is very important that the loan should be approved with the least amount of error in risk calculation while taking up as the least time possible. So a loan prediction model is required that can predict quickly whether the loan can be passed or not with the least amount of risk possible.

Literature survey:

* The author, Vaidya, Ashlesha uses logistic regression as a machine learning tool in paper and shows how predictive approaches can be used in real world loan approval problems. His paper uses a statistical model (Logistic Regression) to predict whether the loan should be approved or not for a set of records of an applicant. Logistic regression can even work with power terms and nonlinear effect. Some limitations of this model are that it requires independent variables for estimation and a large sample is required for parameter estimation.
* A work by Amin, Rafik Khairul and Yuliant Sibaroni was referenced which used Decision tree algorithm called C4.5 to implement a predictive model. This algorithm creates a decision tree that generally gives a high accuracy in decision making problems. Dataset of 1000 cases is used in which 70% is approved and rest is rejected. This paper shows C4.5 algorithm performance in recognizing the eligibility of the applicant to repay his/her loan. From the conducted tests, it is found that the highest precision value is 78.08% which was found using data partition of 90:10. The greatest recall value is 96.4% and was reached with data partition of 80:20. Partition of 80:20 is considered to be best since it has a high recall and the highest accuracy.
* Yet another approach was used by Hassan, Amira Kamil Ibrahim and Ajith Abraham wherein they used German dataset and built a prediction model working basically on backpropagation and implemented with three different back propagation algorithms. They also used two different methods for two filtering functions for the attributes which resulted in DS2 giving highest accuracy using PLsFi filtering function

Methodology:

The paper will be comparing different prediction models and deduce their limitations as well as advantages. Since all the research papers used different sets of data to infer the accuracy and for cross validation of data, the authors have used the same data for all the models which will give a clearer view on their performance and lead to a better comparison of the same. On the basis of the results, a modified prediction model will be created to ensure maximum accuracy and performance.

1.Data collection and preprocessing

2.Filling the missing values

3.exploary data analysis

4.ploting graph to infer more results

5.Buliding predictive model using logistic regression

6.Train and test the model

Conclusion:

Loan approval prediction involves the analysis of various factors, such as the applicant’s financial history, income, credit rating, employment status, and other relevant attributes. By leveraging historical loan data and applying machine learning algorithms, businesses can build models to determine loan approvals for new applicants

References:

* Vaidya and Ashlesha, Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval, 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017.
* Amin, Rafik Khairul and Yuliant Sibaroni, Implementation of decision tree using C4. 5 algorithms in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region), 2015 3rd International Conference on Information and Communication Technology (ICoICT). IEEE, 2015.
* Arora, Nisha and Pankaj Deep Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment, Applied Soft Computing 86 (2020), 105936.