IBM

Title: Smart Lender - Applicant Credibility Prediction for Loan Approval

LITERATURE SURVEY:

M. Sheikh, A. Goel, T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," International Conference on Electronics and Sustainable Communication Systems [1] According to the authors, the forecasting process begins with data clean-up and processing, missing value substitution, data set experimental analysis, and modelling, and continues to model evaluation and test data testing. A logistic regression model has been executed. The highest accuracy obtained with the original dataset is 0.811. Models are compared based on performance measurements such as sensitivity and specificity. As a result of analysing, the following conclusions were drawn. However, other characteristics of customers that play a very important role in lending decisions and forecasting defaulters should also be evaluated. Some other traits, such as gender and marriage history, do not seem to be considered by the company

S. M S, R. Sunny T, "Loan Credibility Prediction System Based on Decision Tree Algorithm," International Journal of Engineering Research & Technology [2] A credit credibility soothsaying system that helps companies make the right opinions to authorize or reject the credit claims of guests. This helps the banking assiduity to open effective distribution channels. This means that if the customer has a minimum repayment capacity, their system can avoid future risks. Including other techniques (using the Weka tool) that are better than the general data mining model has been implemented and tested for domains

A. Kumar, I. Garg and S. Kaur, "Loan Approval Prediction based on Machine Learning Approach," IOSR Journal of Computer Engineering,[3] The author suggests that, a credit status model for predicting loan applicants as valid or standard customers. The proposed model shows a score of 75.08 when classifying loan aspirants using R-Package. Lenders can use this interpretation to make mortgage choices for mortgage operations. In addition, comparative studies were conducted at different iterative levels. The replication position is a 30- grounded

ANN model that offers a more advanced delicacy than other situations. This model can be used to avoid large losses in marketable banks

Dr K. Kavitha, "Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques," IJARCSSE - Volume 6, Issue 2 [4] Six machine learning classification models were used to predict Android applications. The model is available in open-source software R. This application works well and meets the requirements of all banks. The downside of this model is that it gives each element a different weight, but in reality, it may be possible to approve a loan only based on a single powerful element, which is not possible with this system. This component can be easily connected to many other systems. There are cases of computer failure, and the most important weights of content errors and features are fixed by the automatic prediction system, and soon, so-called software may be safer, more reliable and more

P. Dutta, "A STUDY ON MACHINE LEARNING ALGORITHM FOR ENHANCEMENT OF LOAN PREDICTION", International Research Journal of Modernization in Engineering Technology and Science [5]. Risk assessment and forecasting is an important task in the banking industry in determining whether a good and lazy loan applicant is applicable. To improve the accuracy of risk, risk assessments are conducted in primary and secondary education. Customer data is extracted and related attributes are selected using information gain theory. Rule forecasting is performed for each credit type based on predefined criteria. Approved and rejected applicants are considered "Applicable" and evaluated as "Not Applicable". Corresponding experimental results have shown that the method proposed predicts better accuracy and takes less time than existing methods

G. Arutjothi, Dr C. Senthamarai, "Prediction of Loan Status in Commercial Bank using Machine Learning Classifier," Proceedings of the International Conference on Intelligent Sustainable Systems [6] The main purpose of this design is to prognosticate which customers will be repaid with a loan because the lender needs to anticipate the problem that the borrower won't be suitable to repay the threat. Studies of three models show that logistic regression with a rating is superior to other models, random forests, and decision trees. Poor credit seekers aren't accepted, presumably because they have the option of not paying. In utmost cases, high-value appliers may be eligible for a reduction that may repay the loan. Certain sexual orientations and marriage status appear to be out of the reach of the company

P. Supriya, M. Pavani, N. Saisushma, N. Kumari and K. Vikas, "Loan Prediction by using Machine Learning Models," International Journal of Engineering and Techniques[7] they have used only one algorithm; there is no comparison of different algorithms. The algorithm used was Logistic Regression and the best accuracy they got was 81.11%. The final conclusion reached was only those who have a good credit score, high income and low loan amount requirement will get their loan approved. Comparison of two machine learning algorithms was made in

R. Salvi, R. Ghule, T. Sanadi, M. Bhajibhakare, "HOME LOAN DATA ANALYSIS AND VISUALIZATION," International Journal of Creative Research Thoughts [8] The two algorithms used were two class decision jungle and two class decision and their accuracy were 77.00% and 81.00% respectively. Along with these they also calculated parameters such as Precision, recall, F1 score and AUC. The shows a comparison of four algorithms. The algorithms used were Gradient Boosting, Logistic Regression, Random Forest and CatBoost Classifier. Logistic Regression gave a very low accuracy of 14.96%. Random forest gave a good accuracy of 83.51%. The best accuracy we got was from CatBoost Classifier of 84.04%. There was not much difference between Gradient Boosting and CatBoost Classifier in terms of accuracy. Accuracy of Gradient Boosting was 84.03%. Logistic Regression, Support Vector Machine, Random Forest and Extreme Gradient Boosting algorithms are used in [4]. The accuracy percentage didn't vary a lot between all the algorithms. But the support vector Machine gave the lowest variance.

B. Srinivasan, N. Gnanasambandam, S. Zhao, R. Minhas, "Domain-specific adaptation of a partial least squares regression model for loan defaults prediction," 11th IEEE International Conference on Data Mining Workshops[9] The less the variance, the less is the fluctuation of scores and the model will be more precise and stable. Only the K Nearest Neighbor Classifier is used in . The process of Min-Max Normalization is used. It is a process of decomposing the attributes values. The highest accuracy they got was 75.08% when the percentage of dataset split was 50-50% with k to be set as 30.

M. V. Reddy, Dr B. Kavitha, "Neural Networks for Prediction of Loan Default Using Attribute Relevance Analysis," International Conference on Signal Acquisition and Processing[10] uses a prediction model which is constructed using three different training algorithms to train a supervised twolayer feedforward network. The results show that the training algorithm improves the design of loan default prediction model. 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.

G. Chornous, I. Nikolskyi, "Business-Oriented Feature Selection for Hybrid Classification Model of Credit Scoring," IEEE Second International Conference on Data Stream Mining & Processing[11] uses the classification model for predicting the future behaviour of costumers in CRM. In CRM domain, the mostly used model is neural network. He recognized eighty seven articles associated to data mining applications and techniques between 2000 and 2006. used a model namely hybrid Adaptive NeuroFuzzy Inference model, grouping of statistics and NeuroFuzzy network. A 10-fold cross validation is used for better results and a comparison with other models.

M. Cary Collins Improving Information Quality in Loan Approval Processes for Fair Lending and Fair Pricing[12] Bank data management on loan approval processes has great room for improvements of information quality and data problems prevention especially with regards to fair lending and fair pricing practices. They first reviewed briefly typical data collection protocols deployed at many financial institutions for loan approval and loan pricing. Federal regulations mandate portions of these data protocols. While discussing the data capture and analysis for fair lending, they illustrated some initial key steps currently needed for improving information quality to all parties involved.

Sivasree M S, Rekha Sunny T Loan Credibility Prediction System Based on Decision Tree Algorithm[13] Data mining techniques are becoming very popular nowadays because of the wide availability of huge quantity of data and the need for transforming such data into knowledge. Data mining techniques are implemented in various domains such as retail industry, biological data analysis, intrusion detection, telecommunication industry and other scientific applications. Techniques of data mining are also be used in the banking industry which help them compete in the market well equipped. In this paper, they introduced a prediction model for the bankers that will help them predict the credible customers who have applied for a loan. Decision Tree Algorithm is being applied to predict the attributes relevant for credibility. A prototype of the model has been described in this paper which can be used by the organizations for making the right decisions to approve or reject the loan request from the customers.

Loan Approval Prediction based on Machine Learning Approach Kumar Arun, Garg Ishan, Kaur Sanmeet With the enhancement in the banking sector, lots of people apply for bank loans but the bank has its limited assets which it grants to only limited people , so finding out to whom the loan can be granted is a typical process for the banks. So, in this paper , they tried to reduce this risk by selecting the safe person so as to save lots of bank efforts and assets. It was done by mining the previous records of the people to whom the loan was granted before and on the basis of these records the machine was trained using the machine learning model which gave the most accurate result. The main goal of this paper is to predict if loan assignment to a specific person will be safe or not. This paper has into four sections (i) Collection of data (ii) Comparing the machine learning models on collected data (iii) Training the system on most promising model (iv) Testing the system.