

Deciding Bank Client's Credit Limit and Home Loan Using Machine Learning

Ganga A^{a,1}, Vijayalakshmi S^b, Ms Kalpana R A^b,

^aAssistant Professor, Dept of CSE, Sri Sairam Engineering College, Chennai, India

^bUG Scholar, Dept of CSE, Sri Sairam Engineering College, Chennai, India

Abstract. Bank customer's profile helps the bankers to whom the banking facilities may be given and what credit limit to produce. The profile includes both demographic and transaction data. By the best technique, the exactness might be accomplished and furthermore the banks can get higher gainfulness by fulfilling the purchasers. The various data analytic techniques are used for the credit limit prediction and the loan amount for that credit limit is also determined. The regression algorithms are employed for improving high accuracy in predicting the credit card limit.

Keywords. Regression, Demographic data, Transaction Data

1. Introduction

A Credit Card may seem like simply one more device which assists you with making buys, however it without a doubt can accomplish such a great deal more. Whenever utilized dependably, a Credit Card can help you fabricate a decent record as a consumer, additionally permitting you to get advances at good Interest Rates. It also helps you earn reward points on your everyday purchases which can be redeemed for Gifts, vouchers, flight tickets etc. Credit Card also has host of offers on various categories like Travel, Dining, Shopping, Bill payments etc. Credit Card is safer for Online & Store purchase as the transactions are authenticated using PIN or OTP. While assessing a home advance application, banks or loan specialists, figure a candidate's getting capacity dependent on his/her credit limit. Fundamentally, when you apply for an advance, the bank takes a gander at different subtleties to check in the event that you have the capacity to reimburse. Banks will likewise attempt to break down the level of hazard implied in giving you a credit. The components considered to evaluate reimbursement capacity incorporate compensation, securities, age and record as a consumer. With regards to financial record, your charge card assumes an immense part. On the off chance that your record as a consumer is acceptable, you will be considered a 'acceptable danger' and will be conceded an advance. Although, some studies revealed statistical information about the home loan approval via credit card limits.

¹Ganga. A, Assistant Professor, Dept of CSE, Sri Sairam Engineering College, Chennai, India;
E-mail: ganga.cse@sairam.edu.in

The methods like Regression, AdaBoost algorithm, XGBoost algorithm, Stochastic Gradient Descent algorithm are intricated to develop the project to its higher level of accuracy. The proposed framework can improve bank customer's behavior using machine learning algorithms such as regression and classification. Here the credit card limit with house loan is predicted and a series of pre-processing for converting the raw data to useful data is done, so that it will increase the execution time as well as the accuracy of the model. The algorithm analysis for both the model with regression and classification algorithm is being done and found the best fittest algorithm for the model.

2. Related Works

Using different approaches and datasets, a number of experts are tackling the problem of profiling bank clients. In 2015, Using two stages and k-implies clustering algorithms, a characterization model for the dataset of Sepah Bank Branches in Tehran was introduced by Sharahi and Aligholi [2]. Division of 60 organizations, which were clients of Sepah Bank was a sort of segment and conduct division and it helped to distinguish the loyal customers. In 2016, Ayoubi [3] clarified a client division model dependent on the two-venture calculation and Kohonen neural organization. In 2017, Palaniappan et al. [4] introduced a profiling model for the clients of a Portuguese retail bank inside the duration of five years (2008 to 2013). This paper zeroed in on assisting manages an account with expanding the exactness of their client profiling through characterization just as distinguishing a gathering of clients who had a high likelihood to buy in a long haul store. Three grouping calculations were utilized which were Naïve Bayes, Random Forest and Decision Tree. In 2017, Bansalet al. [5] introduced an alteration in a bunching model of the k-implies calculation. In 2017, Patil and Dharwadkar [6] produce forecast and characterization model for two datasets of bank client's information. [7-8]

3. Existing System

The primary thought of our proposed model is to improve profiling bank client's conduct utilizing distinctive AI procedures. This model beginnings with the dataset, which acquired from the UCI AI vault. At that point information goes through the progression of information preprocessing. From that point forward, the AI procedures are applied to fabricate the client profile. In AI, the profiling stage perceives the things in a gathering and places them under target classes. In this paper, the precision pace of strategies is assessed through Gini co-effective for the solo methods at that point utilized the outcomes as contribution for regulated procedure (Artificial Neural Network) (ANN) at that point assesses there results to contrast them with get the best strategy.

4. Proposed System

The proposed framework is used to improve bank client's conduct utilizing AI calculations such as regression and classification. In this model the credit card limit and house loan is predicted for the particular location in India. A series of pre-processing is done for converting the raw data to useful data. The various regression algorithms like linear regression ,decision tree, random forest regression and SVM is used in credit limit model and the accuracy of the algorithms is found. Similarly, the classification algorithms like logistic regression decision tree, naïve bayes classification, svm are used in loan prediction model and accuracy of the algorithms.

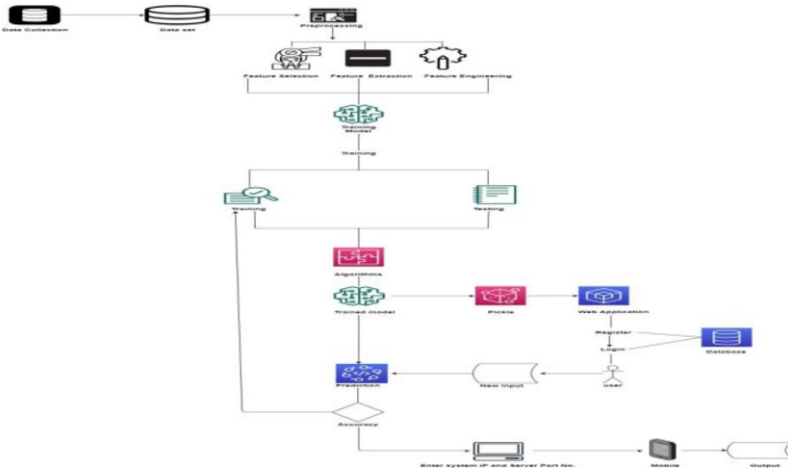


Figure 1. The proposed model for profiling bank customers

5. Methods Used

5.1. Naive bayes theorem

This learning model applies Bayes rules through autonomous highlights. Each example of information D is assigned to the class of most noteworthy resulting likelihood. In AI, it is frequently keen on choosing the best speculation (h) given information (d). The theory (h) may be the class to relegate for another knowledge case in a grouping problem (d). Perhaps the simplest method of selecting the most likely hypothesis provided the information available, which can be used as previous information about the question. The Bayes' Theorem explains how to calculate the probability of a hypothesis based on previous data. $P(h|d) = \frac{P(d|h) * P(h)}{P(d)}$ is Bayes' Theorem's formula (d) P(h|d) denotes the probability of speculating h given the knowledge d. The back probability is a term used to describe this situation.

5.2. Decision tree Algorithm

A decision tree is developed by recursive parceling Starting from the root, the information is part on the element that outcomes in the biggest Information Gain (IG)

(clarified in more detail beneath). In an iterative interaction, we at that point rehash this parting technique at every youngster hub until the leaves are unadulterated — for example tests at every hub all have a place with a similar class. Here, our target work is to augment the data acquire at each split, which we characterize as s :

$$IG(D_p, f) = I(D_p) - \left(\frac{N_{left}}{N_p} I(D_{left}) + \frac{N_{right}}{N_p} I(D_{right}) \right)$$

5.3. Random Forest

Random forest, also known as random decision forests, is a group learning technique for order, relapse, and other tasks that involves creating a large number of option trees over a period of time and determining the class that is the method of the classes (characterization) or mean expectation (relapse) of the individual trees.

5.4. Support Vector Machine

A SVM model is a representation of the models as focuses in space, with the instances of the various classes separated by an unmistakable hole as large as is reasonably possible. Planned new models with same space and are expected to have a designation based on which side of the hole they fall on. In addition to performing direct characterization, SVMs may also play out a non-direct arrangement using what is known as the component stunt, effectively designing their contributions to high-dimensional element spaces. When unlabeled information is there, regulated learning is impossible, so a solo learning method is needed, in which the goal is to discover common groupings of information and then map new information to these framed groups.

5.5. Linear Regression

Linear Regression is an AI calculation dependent on managed learning. Direct relapse depicts the task of predicting the value of a reliant variable (y) in light of a provided autonomous variable (x). In this way, this relapse protocol discovers a direct link between x (input) and y (output). The name will now be Linear Regression.. In the graph, X (input) is the work insight and Y (yield) is the compensation of an individual. The relapse line is the best fit line for our model. $Y = \Theta_1 + \Theta_2 X$

5.6. Logistic Regression

The logistic potential, which is at the heart of the strategy, is named after logistic regression. Analysts developed the logistic capacity, also known as the sigmoid capacity, to represent properties of population growth in nature, such as rapid growth and maximization at the climate's transport limit.

6. Implementation

Module-1 Data collection and preprocessing

Data collection is the process of collecting details about the bank clients with credit transaction with personal details like Age, City, Segment, income, card type, etc.

Preprocessing is the process of removing the incomplete, inconsistent, incorrect and null data of the dataset. After the selection of the features we will use that data to train our model.

Module – 2 Credit Card limit Model

This module for credit card limit is used to predict the credit card limit of the bank clients using various regression algorithms and also the best accuracy of the algorithm is found in request to build up the model for prediction. After the preprocessing of the data ,the dataset is used for training. The dataset is split as features and label after then it is used for training the model. The different regression algorithm like random forest regression, decision tree , linear regression, and support vector regression algorithms are analyzed . Finally, which algorithm has greater accuracy will be used for final model and the result of this module is used, in order to predict the loan of the predicted card limit.

Module-3 Loan prediction Module

After creating the credit card limit modules ,the preprocessed loan dataset is used for training. The dataset is split as features and label, after that we will use that for training the model. Here ensemble modules are used in order to predict the result. In this module supervised learning is used for the classification of loan prediction thus, many classification algorithms are present, here we using logistic regression because comparing with other algorithms it gives more accuracy. Finally the algorithm with greater accuracy is used for developing the model for prediction of loan.

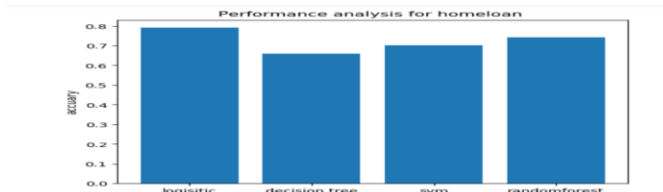


Figure 2. Module-4 Interface with UI

After the creation of model for loan and credit limit. The web application is created to see the prediction results but the machine learning and Web development is different domain so the pipeline for interacting machine learning and web application model using the pickle package is created. After that user can give input and can get the output results.

7. Future Scope

The executives of credit limits is quite possibly the most basic choices identified with Master card accounts. It influences various factors that have immediate or circuitous effect on the productivity. Supervised learning showed higher accuracy rate in determining credit limit and home loan. The accuracy in predicting the home loan can be improved by using unsupervised learning algorithm in future.

8. Conclusion

Profiling of customer details helped the banks to form an interactive relationship with customers through trust. Banks get high profitability through the customer satisfaction. Supervised learning is employed for producing higher accuracy in determining credit card limit. Decision tree regression produced higher accuracy rate when compared to other regression algorithms. Based on the credit limit home loan can be predicted. The classification algorithms are used for home loan prediction in which logistic regression produced higher accuracy rate.

References

- [1] M. Sharahi and M. Aligholi, Classify the data of bank customers using data mining and clustering techniques (case study: Sepah bank branches Tehran), *J. Appl. Environ. Biol. Sci.*, vol. 5, no. 5, pp. 458–464, 2015.
- [2] M. Ayoubi, Customer segmentation based on CLV model and neural network, *Int. J. Computer. Sci.*, vol. 13, no. 2, p. 31, 2016.
- [3] A. Arshad, S. Riaz, and L. Jiao, Semi-supervised deep fuzzy c-mean clustering for imbalanced multi-class classification, *IEEE Access*, vol. 7, pp. 28100–28112, 2019.
- [4] Ahram Online. Egypt's BanqueMisr Conditionally Suspends Card Usage Abroad Amid Currency Crisis. Accessed: Apr. 10, 2019.[Online].
- [5] Ambeth Kumar.V.D .Precautionary measures for accidents due to mobile phone using IOT. *Clinical eHealth*, Volume 1, Issue 1, March 2018, Pages 30-35.
- [6] Nanagasabapathy.K, G. Harish; O. I. Allen Sebastian; N. Sowrabh Chandra; V. D. Ambeth Kumar . Validation system using smartphone luminescence. *IEEE International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, Pages: 235 – 239, 6-7 July 2017, Kannur, India
- [7] Ambeth Kumar.V.D,V. D. Ashok Kumar; H. Divakar; R. Gokul. Cloud enabled media streaming using Amazon Web Services. *IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, Pages: 195 – 198, 2-4 Aug. 2017, Vel Tech University, Chennai, India (DOI: 10.1109/icstm.2017.8089150)
- [8] Aravindh.B, V.D.Ambeth Kumar,G. Harish, V. Siddarth. A novel graphical authentication system for secure banking systems. *IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, Pages: 177 – 183, 2-4 Aug. 2017, Vel Tech University, Chennai, India
- [9] K. Sabarinathan et.al ., “ Machine Maintenance Using Augmented Reality”, 3rd International Conference on Communication and Electronics Systems (ICCES), 2018. (DOI: 10.1109/CESYS.2018.8723900)
- [10] R. Subha Shini et.al., “ Recurrent Neural Network based Text Summarization Techniques by Word Sequence Generation”,*IEEE International Conference on Inventive Computation Technologies (ICICT)*, 2021, DOI: 10.1109/ICICT50816.2021.9358764