SMART LENDER - APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL

TEAM ID:

PNT2022TMID15942

TEAM MEMBERS:

1. RAJESH KUMAR A   (927619BEC4161)

2. NAVEEN B   (927619BEC4130)

3. NAVEEN K   (927619BEC4131)

4. SANJAY G   (927619BEC4171)

INDUSTRY MENTORS NAME:

1. SOWJANYA

2. SANDEEP DOODIGANI

FACULTY MENTOR NAME:

1. Dr. SIVANANDAM K

ABSTRACT:

As the needs of people are increasing, the demand for loans in banks is also frequently getting higher every day. Banks typically process an applicant's loan after screening and verifying the applicant's eligibility, which is a difficult and time-consuming process. In some cases, some applicants default and banks lose capital. The machine learning approach is ideal for reducing human effort and effective decision making in the loan approval process by implementing machine learning tools that use classification algorithms to predict eligible loan applicants. The credit risk is defined as the likelihood that borrowers will fail to meet their loan obligations. To predict whether the borrower will be good or bad is a very difficult task for any bank or organization. The banking system uses a manual process for checking whether a borrower is a defaulter or not. No doubt the manual process will be more accurate and effective, but this process cannot work when there are a large number of loan applications at the same time. To solve this problem, we have developed a system in which we can predict whether the applicant chosen will be a deserving applicant for approving the loan or not. This can be done with the help of machine learning algorithms.

LITERATURE REVIEW:

A recent development of machine learning techniques and data mining has led to an interest of implementing these techniques in various fields. The banking sector is no exclusion and the increasing requirements towards financial institutions to have robust risk management has led to an interest of developing current methods of risk estimation

AUTHOR: Steenackers and Goovaerts

DESCRIPTION: The key factors that may influence loan default are borrower's age, location, resident/work duration, owner of phone, monthly income, loan duration, whether or not applicant works in a public sector, house ownership and loan numbers. In addition to identifying factors that may influence loaned fault, there is also a need to build robust and effective machine learning models that can help capture important patterns in credit data. The choice of model so great importance as the chosen model plays a crucial role in determining accuracy, precision and efficiency of a prediction system. Numerous models have been used for loan default prediction and although there is no one optimal model, some models definitely do better than others.

AUTHOR: Vimala and Sharmili

DESCRIPTION: A loan prediction model using and Support Vector Machines (SVM)methods. Naïve Bayes, an independent speculation approach, encompasses probability theory regarding the data classification. On the other hand, SVM uses statistical learning model for classification of predictions. Dataset from UCI repository with 21 attributes was adopted to evaluate the proposed method. Experimentations concluded that, rather than individual performances of classifiers (NB and SVM), the integration of NB and SVM resulted in an efficient classification of loan prediction

AUTHOR: Jency, Sumathi and Shiva Sri

DESCRIPTION: An Exploratory Data Analysis (EDA) regarding the loan prediction procedure based on the client's nature and their requirements. The major factors concentrated during the data analysis were annual income versus loan purpose, customer 's trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in the current job, and chances for loan repayment versus the house ownership. Finally, the outcome of

the present work was to infer the constraints on the customer who are applying for the loan followed by the prediction regarding the repayment. Further, results showed that, the customers were interested more on availing short-tenure loans rather than long-tenure loans

AUTHOR: Supriya, Pavani, Saisushma

DESCRIPTION: They presented a ML based loan prediction model. Themodulesin the present approach were data collection and pre-processing, applying the ML models, training followed by testing the data. During the pre-processing stage, the detection and removal of outliers and imputation removal processing were carried out. In the present method, SVM, DT, KNN and gradient boosting models were employed to predict the possibilities of current status regarding the loan approval process. The conventional 80:20 rule was adopted to split the dataset into training and testing processes. Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models.

AUTHOR: Vimala Kumari and Vikas

DESCRIPTION: presented a loan prediction model using several Machine Learning (ML) algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co applicant 's income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. Various ML models adopted in the present method includes, Linear model, Decision Tree (DT), Neural Network (NN), Random Forest (RF), SVM, Extreme learning machines, Model tree, Multivariate Adaptive Regression Splines, Bagged Cart Model, NB and TGA. When evaluated these models using Environment in five runs, TGA resulted in better loan forecasting performance than the other methods.

AUTHOR: Aboobyda Jafar Hamid and Tarig Mohammed Ahmed

DESCRIPTION: presented a loan risk prediction model based on the data mining techniques, such as Decision Tree (J48), Naïve Bayes (NB) and BayseNet approaches. The procedure followed was training set preparation, building the model, Applying the model and finally. Evaluating the accuracy. This approach was implemented using Weka Tool and considered a dataset with eight attributes, namely, gender, job, age, credit amount, credit history, purpose, housing, and class. Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction approach resulted in better accuracy than the other methods.

AUTHOR: Kacheria, Shivakumar, Sawkar and Gupta

DESCRIPTION: A loan sanctioning prediction procedure based on NB approach integrated with K-Nearest Neighbor (KNN) and binning algorithms. The seven parameters considered were income, age, profession, existing loan with its tenure, amount and approval status. The sub-processes include, Preprocessing (handling the missing values with KNN and data refinement using binning algorithm), Classification using NB approach and updating the dataset frequently results in appropriate improvement in the loan prediction process. Experimentation put-forth the conclusion that, integration of KNN and binning algorithm with NB resulted in improved prediction of loan sanctioning process

AUTHOR: Goyal and Kaur

DESCRIPTION: An ensemble technique-based loan prediction procedure for the customers. The sub processes in the present method includes, data collection, filtering the data, feature extraction, applying the model, and finally analysis the results. The various loan prediction procedures implemented in the present method were Random Forest (RF), SVM and Tree model with Genetic Algorithm (TGA). The parameters considered for evaluating the models were accuracy, Gini

Coefficient, Area Under Curve (AUC), Receiver Operating Curve (ROC), Kolmogorov - Smirnov (KS) Chart, Minimum Cost - Weighted Error Rate, Minimum Error Rate, and K-Fold Cross Validation parameters. Experimentation outcome concluded that the integration of three methods (RF, SVM and TGA) resulted in improved loan - prediction results rather than individual method 's prediction. The various steps involved in extracting knowledge from raw data. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc..

AUTHOR: Sudhamathy

DESCRIPTION: A risk analysis method in sanctioning a loan for the customers using R package. The various modules include data selection, pre-processing, feature extraction and selection, building the model, prediction followed by the evaluation. The dataset used for evaluation in this method was adopted from UCI repository. To fine tune the prediction accuracy, the pre-processing operation includes the following sub-processes: detection, ranking and removal of outliers, removal of imputation, and balancing of dataset by proportional bifurcation regarding testing and training process. Further, feature selection process improves the prediction accuracy. When evaluated, the DT model resulted in 94.3% prediction accuracy. The process of analyzing data from different perspectives and extracting useful knowledge from it. Tithe core of knowledge discovery process. Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree-based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.