International Journal of Multidisciplinary Engineering in Current Research

Volume 6, Issue 6, June 2021, http://ijmec.com/

Prediction of Loan Status in Commercial Bank using Machine Learning Classifier

T Rajesh singh
professor
Cse Department
SRM University Chennai - SRM Institute of Science and Technology

ABSTRACT

To address a range of difficulties in the banking business, the creation of a more accurate predictive modelling system is required. It is impossible to forecast who would fail on a loan, which makes it an onerous job for the banking industry. When it comes to loans, one of the quality factors is how they are currently doing. Despite the fact that it does not show all of the information right away, it is the first stage in the loan application process and must be completed. The loan status is taken into account while developing a credit rating model. Credit scoring algorithms are utilised for trustworthy credit data analysis, since they allow for the identification of defaulters and legitimate customers in the credit data. One of the objectives of this project is to create a credit scoring model that is based on credit information. For the purpose of developing the financial credit score model, a number of machine learning algorithms are used in tandem with one another. This study includes the development of a credit data analysis approach that is based on machine learning classifiers as part of the overall project. KNearest Neighbor (K-NN) classifier is used in combination with Min-Max normalisation to provide the best results. The goal is realised via the usage of the R programming language as well as other resources. In terms of the most important facts, the proposed model provides the most accurate information conceivable. Machine learning classifiers are used by commercial banks to forecast the status of loan applications.Keywords: Credit Scoring; K-NN; Loan status; Loan Lending Process;Min-Max Normaliz

1. INTRODUCTION

If you're looking for a commercial loan, one of the most important concerns that the banking industry must address is the grading of borrowers' creditworthiness, which is one of the most pressing challenges that the banking industry must address. When it comes to commercial loan lending, one of the most important concerns that the banking industry must address is the grading of borrowers' creditworthiness. As a result of borrowing money, the probability that borrowers may not be able to meet their financial commitments is referred to as credit risk in finance. Credit scoring systems are used to anticipate credit risk and prevent criminal behaviour, with the goal of lowering crime rates in the United States as a result. In order to make decisions about borrowers based on the information they supply about themselves and their financial circumstances, credit rating systems are utilised. For each loan decision they make, lenders strive diligently to reduce the likelihood of default in order to maximise the return that pays for the risk they take. Risk of credit default is what decides whether or not the banking sector as a whole will be successful or unsuccessful in its endeavours. It will suffer a financial loss if the bank is unable to recover the whole amount of the credit card debt. Consequently, a bank's profit is tied to the risk of a loan arrangement defaulting. Additionally, credit risk management and assessment are both time-consuming and complicated operations to do, making them a substantial source of worry. Separating credit-scoring tasks into two categories is a viable option. Throughout the history of the banking business, commercial loans have been a significant component of the company's operations, and lenders are subjected to ongoing scrutiny in order to reduce their credit risk exposure. Just as importantly, the credit risk issue is just too complex to be handled in a reasonable period of time, and there is no acceptable amount of time available.

International Journal of Multidisciplinary Engineering in Current Research
Volume 6, Issue 6, June 2021, http://ijmec.com/

The use of a credit risk assessment model aids in the estimation of the amount of credit risk exposure. With the aid of credit assessment models, it is possible to analyse a current customer and categorise a new prospective consumer. Among those who work in the subject of data mining, data mining is the term used to describe the practise of extracting meaningful information from enormous databases. In this technique, the categorization of data, the grouping of data, and the mining of association rules among the data are all performed. The classification function is a vital component of the data mining process, and it should not be taken for granted in any manner. To make your choice from among the several categorization systems available at the time, look no further than the list below. Techniques like as decision trees, support vector machines, neural networks, k-nearest neighbour, and logistic regression, to mention a few examples, are examples of what is possible to implement. In order to accomplish this study, we have tested and assessed every categorization technique that has been made available to us. To be completely honest, I'm still not sure which strategy would be the most appropriate for any particular dataset. Therefore, the goal of this project is to establish a credit rating model for a commercial loan by merging the findings of a K-NN classifier with a regression model in order to create a more accurate credit rating model. Another goal is to divide loan applications into two groups: default customers and nondefault customers. This classification will be used by financial institutions to determine loan eligibility and payback conditions for future loans. It is predicted that the findings of the research would be very beneficial to lenders when it comes to loan selection and selection criteria. The following is how the remaining portions of the paper are organised: A review of the literature on credit risk and K-NN is provided in the next section, which is followed by a description of the basic ideas of machine learning and normalisation in the following section. It is in Section 5 of this report that you will find a comprehensive description of how the strategy and data used in this project were developed, as well as the conclusions and recommendations made. It is important to include in Section 6 your discussion of the conclusion, as well as your goals for the future.

2. LITERATURE SURVEY:

Abdelmoula, Aida Krichene . "Bank credit risk analysis with knearest-neighbor classifier: Case of Tunisian banks." Accounting and Management Information Systems 14.1 (2015): 79.

It is possible for borrowers to fail to meet their financial obligations in line with the terms of the agreement, which is referred to as credit risk. The development of sophisticated tools and models to assist bankers in the analysis, aggregation, and risk management processes has been more common in recent years. Among the tools and models available are the following: Because of this, the results of these models are becoming more significant in the risk management and performance assessment operations that financial institutions carry out on a daily basis. Using machine learning approaches, we want to discover a solution to the issue of predicting short-term loan default for a Tunisian commercial bank, which is now being investigated. From 2003 to 2006, we made use of information from a database of 924 credit records of Tunisian businesses that had been supplied to us by a Tunisian commercial bank. Moreover, as shown by the results of the K-Nearest Neighbor classifier technique, which achieves a respectable classification performance of around 88.63 percent (for k=3), the best information set is associated with accrual and cash-flow accounting principles. Using a receiver operating characteristic curve (ROC), it is important to evaluate the performance of the model under consideration (ROC curve). According to the results of the research, the

International Journal of Multidisciplinary Engineering in Current Research
Volume 6, Issue 6, June 2021, http://ijmec.com/

AUC (Area Under the Curve) criteria for the first model was determined to be 87.4 percent, 95 percent for the third model, and 95.6 percent for the best model utilising cash flow data, with the first model scoring the highest.

Among those present were A. Rama Satish, Babu, and other members of the group It has been discovered something new: Improvements have been made in credit scoring techniques that make use of K-Nearest Neighbor techniques. This is the first issue of Computer Science and Technology Development, a new journal that has just been established. The government of the United States has issued a directive on the advancement of computer science and technology (2013)

Credit scoring has lately garnered a great deal of interest in the domains of academia and business, and this is especially true in the realm of academics. A variety of alternative modelling approaches have been created in order to cope with the difficulties associated with credit rating. In order to evaluate whether a loan application is a good or a terrible match for a financial institution, financial organisations have long used credit scoring models. There are a number of benefits to using credit scoring models, including the potential to save money on credit investigation, the ability to make credit decisions more rapidly, and the assurance that credit will be collected. This model is compared to the other models on the basis of the credit score assigned to each of the models in question.

There have been a large number of credit risk studies conducted, utilising both conventional models such as theAltman Z-score and machine learning techniques, to determine the chance of default. Research that integrates data from Croatian financial institutions, especially research that focuses on the selection of demographic and/or behavioural characteristics, is severely lacking. Furthermore, the development of dependable models that can effectively predict credit risk is crucial to the achievement of success in this field. As part of this research project, researchers want to construct a credit risk prediction model that will be based on defaulting customer data from Croatian financial institutions (demographic and behaviour data). In the financial sector, the use of decision tree models is utilised to forecast credit risk in order to minimise losses. It is necessary to investigate various variable selection approaches in order to assess the classification accuracy of decision trees created with the supplied variables.

3. SYSTEM ANALYSIS

3.1 Existing system:

The risk of credit default determines whether or not the banking industry will be successful or unsuccessful as a whole. If the bank does not collect the whole amount of the credit, it will incur a loss in terms of finances. Because of this, bank profits are inversely proportionate to the likelihood of loan defaults occurring. It takes a long time to monitor and analyse credit risk, and it may be quite stressful [18]. Credit scoring may be separated into two categories: application scoring and behaviour scoring. Application scoring is the first of these categories. The first of these categories is the grading of applications. Credit application scoring is used to categorise credit applicants into "good" and "bad" risk categories in order to determine their creditworthiness. Existing customers are categorised based on their payment history and personal information, which is determined through behavioural scoring. Throughout the banking industry's history, commercial loans have been an important component of the company, and lenders

International Journal of Multidisciplinary Engineering in Current Research
Volume 6, Issue 6, June 2021, http://ijmec.com/

are constantly scrutinised in order to limit their credit risk exposure. When it comes to addressing the credit risk problem, there is just too much trouble. The usage of a credit risk assessment model helps to estimate the level of credit risk. For both current customers and new potential consumers, credit evaluation models are used in order to analyse and categorise them.

Disadvantages:

• Less security

• Less storage

3.2 Proposed system:

An in-depth description is provided of the technique to credit data analysis that relies on machine learning classifiers. The Min-Max normalisation approach is used in conjunction with the KNearest Neighbor (K-NN) classifier to provide the best results. In order to do this, the R programming language and its libraries are used in conjunction with one another. When it comes to the most important data, the proposed model provides the most accurate results possible. An automated machine learning classifier is used by commercial banks in order to forecast the status of loan applications.

Advantages

• More security
• More storage
• More efficient

4. Input and Output Designs

4.1 INPUT DESIGN

The input design is the way through which information systems and their users are connected to one another. Develop the data preparation requirements and processes, as well as the actions necessary to convert transaction data into a format that can be processed by the business application, in order to ensure that the data is ready for processing. To do this, examine the computer to see if it is capable of reading data from a written or printed document, or ask users to manually enter the data into the system, as appropriate. In order to reduce the amount of data entry necessary, control errors and delays, and eliminate needless processes and simplify the process, this approach is being used. The usage of the input provides security and convenience, while ensuring the confidentiality of the information. Information to be provided as input should include the following items: • What kind of information should be offered as a consequence of the procedure and how should it be presented?

For which data sets should a certain coding or organisation system be used. A chat will be conducted in order to help operational employees with submitting their feedback to the system.During this part, you will learn about several input validation procedures, as well as what to do in the event that an error occurs.

OBJECTIVES

Feedback on your effort is much appreciated. In computer-based systems, designing is the process of converting a user-oriented description of an input into a computer-based system. A well-planned computerised system is critical for minimising data input errors and leading management in the proper path in order to acquire reliable data from the computerised system. A poorly designed computerised system might result in data that is untrustworthy.

International Journal of Multidisciplinary Engineering in Current Research
Volume 6, Issue 6, June 2021, http://ijmec.com/

The construction of user-friendly data entry panels that are capable of processing vast volumes of information ishow this is achieved in part 2. Aim for simplicity and error-free data input when it comes to data entry. The data input panel has been created in such a way that you may make all of the necessary data changes with relative simplicity. It also provides you with the opportunity to browse through your documents.

The correctness of the data will be checked once it has been input. 3. It is possible to enter information into computers via the use of displays. This prevents the user from being taken off guard since appropriate notifications are sent in a timely manner when they are necessary. It is for this reason that the purpose of input design is to provide an input layout that is simple to grasp.

4.2 DESIGN OF THE OUTPUT FORMULA

In addition to meeting the demands of the end user, a high-quality output displays the data in an understandable way.

straight-forward and plain presentation of facts It is the outputs of a system that are responsible for transferring the results of its processing from one system to another and from one user to another. In order for the information to be moved for immediate usage as well as for hard copy printing, this option must be taken throughout the output design phase. As far as the end user is concerned, it is the most important and direct source of information accessible. In order to improve the interaction between the system and the user, it is necessary to build efficient and intelligent output designs.

Ensure that computer output is created in a logical and well-thought-out manner; the required output must be generated while ensuring that each output component is designed in a way that users will find the system simple to use, and that it is effective. As part of the process of analysing and generating computer-generated output, it is critical to establish the precise nature of the output that will be needed in order to meet the criteria being reviewed.

2. Select a method for disseminating the information.

5. If one does not already exist, a paper or report or other document containing the information collected from the system will be generated.
6. Fourteenth, the output form of an information system should accomplish one or more of the goals indicated below.
7. Obtain and disseminate information regarding the company's prior activities as well as its current state and future prospects
8. Within the next several months or years.
9. Make critical notifications about forthcoming events, opportunities, difficulties, or warning signals that must not be ignored.
10. Set off a series of events that will continue indefinitely.
11. Create a plan for how you will proceed going forward.

International Journal of Multidisciplinary Engineering in Current Research
Volume 6, Issue 6, June 2021, http://ijmec.com/

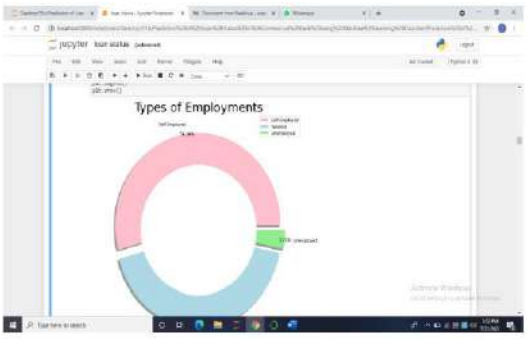

Figure 4.2.1: Types of Employments

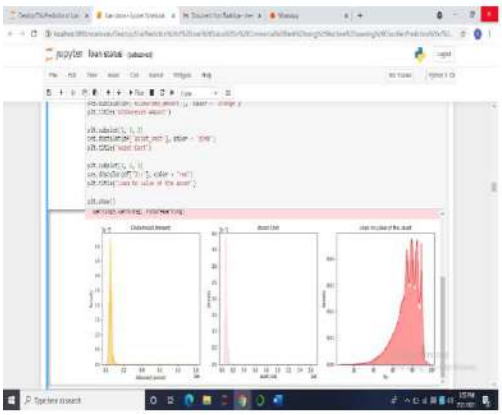

Figure4.2.2: Disbursed amount , Asset cost, Loan to value of the asset

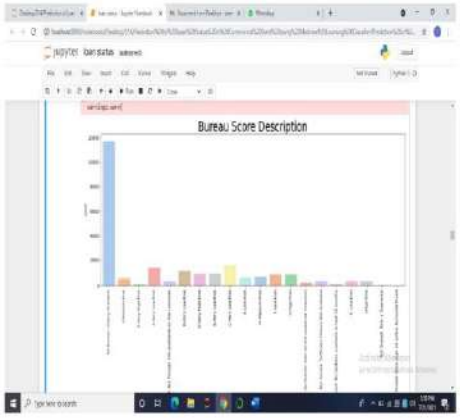

Figure 4.4: Bureau Score Description

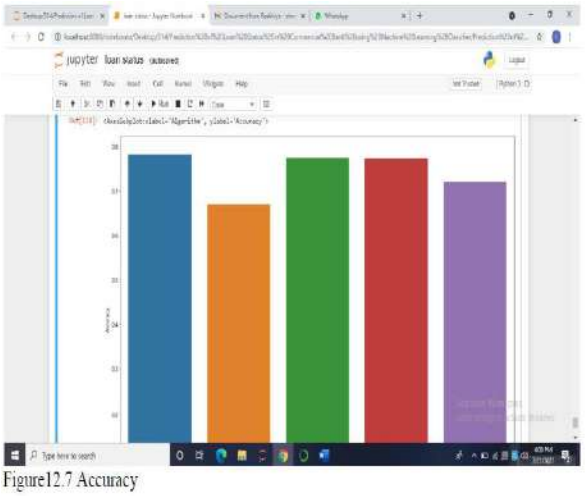

Figure12.7 Accuracy

5. CONCLUSION

Specifically, we are working on developing a loan status model that will be able to determine whether or not a loan application has been submitted legitimately. When it comes to categorising credit applications, the suggested model, which is based on Python and the machine learning package, has a classification accuracy rate of 75.08 percent, according to the researchers. When faced with a large number of loan offers, credit lenders could choose for this method of making lending judgments. In addition, a number of repetition levels at varying degrees of difficulty were used in the comparative investigation. This is a significant improvement over the accuracy of the previous iteration levels of the k-NN model at its 30th iteration level. Using this strategy, it is possible that a commercial bank will avoid suffering a big loss.

REFERENCES:

[I] Abdelmoula, Aida Krichene . "Bank credit risk analysis with knearest-neighbor classifier: Case of Tunisian banks."

Accounting and Management Information Systems 14.1 (2015): 79.
[2] Arutjothi, G., Dr. C. Senthamarai. "Comparison of Feature Selection Methods for Credit Risk Assessment ", International Journal of Computer Science, Volume 5, Issue I, No 5, 2017.
[3] Arutjothi, G., Dr.C.Senthamarai. "Credit Risk Evaluation using Hybrid Feature Selection Method " Software Engineering and Technology 9.2 (2017): 23-26.
[4] Attig, Anja, and Petra Perner. "The Problem of Normalization and a Normalized Similarity Measure by Online Data." Tran. CBR 4.1 (2011): 3-17.
[5] Babu, Ram, and A. Rama Satish. "Improved of K-Nearest Neighbor Techniques in Credit Scoring." International Journal For Development of Computer Science & Technology I (2013).
[6] Bach, Mirjana Pejic, et al. "Selection of Variables for Credit Risk Data Mining Models: Preliminary research " MIPRO 2017-40th Jubilee International Convention. 2017.
[7] Byanjankar, Ajay, Markku Heikkila, and Jozsef Mezei. "Predicting credit risk in peer-to-peer lending: A neural network approach." Computational Intelligence, 2015 IEEE Symposium Series on. IEEE, 2015.
[8] Devi, CR Durga, and R. Manicka Chezian. "A relative evaluation of the performance of ensemble learning in credit scoring." Advances in Computer Applications (ICACA), IEEE International Conference on. IEEE, 2016.
[9] G.Arutjothi, Dr.C.Senthamarai, "Effective Analysis of Financial Data using Knowledge Discovery Database", International Journal of Computational Intelligence and Informatics, Vol. 6: No. 2, September 2016
[10] Goel, Dr Himani, and Gurbhej Singh "Evaluation of Expectation Maximization based Clustering Approach for Reusability Prediction of Function based Software Systems." International Journal of Computer Applications (0975-8887) Volume (2010) process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.