Prediction of Loan Status in Commercial Bank using Machine Learning Classifier

[1]Ch Satyanarayana Vanamala Sudha,[2]Pandit! Mounika, [3]Papanati Divya Sai, [4]Madisetty Navya

[1,2,3,4]Assistant professors, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

Banking Industry always needs a more accurate predictive modeling system for many issues. Predicting credit defaulters is a difficult task for the banking industry. The loan status is one of the quality indicators of the loan. It doesn't show everything immediately, but it is a first step of the loan lending process. The loan status is used for creating a credit scoring model. The credit scoring model is used for accurate analysis of credit data to find defaulters and valid customers. The objective of this paper is to create a credit scoring model for credit data. Various machine learning techniques are used to develop the financial credit scoring model. In this paper, we propose a machine learning classifier based analysis model for credit data. We use the combination of Min-Max normalization and KNearest Neighbor (K-NN) classifier. The objective is implemented using the software package R tool. This proposed model provides the important information with the highest accuracy. It is used to predict the loan status in commercial banks using machine learning classifier.

Keywords: Credit Scoring; K-NN; Loan status; Loan Lending Process;Min-Max Normaliz

1.INTRODUCTION

In commercial loan lending, scoring of borrowers' creditworthiness is one of the most important problems to be addressed in the Banking Industry. Credit risk is defined as the risk that borrowers will fail to meet their loan obligations The Credit scoring system is used to predict the credit risk and to reduce the illegal activities This credit scoring systems are used to make decisions under information about the borrowers In order to make loan decisions, lenders want to minimize the risk of default of each lending decision , and realize the return that compensates for the risk In general, Banking Industry success and failure is based on their credit risk. The credit amount couldn't collect properly, then the bank will be loss. So, bank profit is correlated to their credit risk. Credit risk is a crucial challenge and a complex

task to manage and evaluate Credit scoring tasks can be divided into two groups such as, Commercial loans have always been an important part of the banking industry and lenders are always measured to minimize their credit risk. To solve this credit risk problem is too difficult.

Credit risk evaluation model is used to find the credit risk . To evaluate the existing customer and to classify the new potential customer using credit evaluation models . Data mining is the process to discover useful information from large dataset.It consists of classification, clustering and association rule mining. Classification is a main function of data mining process. There are many classification techniques are available nowadays. Techniques are decision tree, support vector machine, neural

network, k-nearest neighbor and logistic regression, etc. All classification techniques are already used and evaluated for this problem. But, still now couldn't find which technique is suitable for which type of dataset. Therefore, the objective of this paper is to apply a K - NN classifier to develop a credit scoring model for a commercial loan. Furthermore, the aim of this study is to classify loan applications into default customer and nondefault customer group which is used for credit lenders. The results from the study would be very helpful for lenders make loan decisions. The rest of the paper is structured as follows. In section 2 describes the basic concepts of machine learning and normali zation followed by the literature survey on credit

risk, K-NN in section 3. In section 4 discuss on methodology and data to this work followed by result and discussions are made in section 5. In section 6 discuss the conclusion and future work.

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.

Credit risk is defined as the risk that borrowers will fail to pay its loan obligations. In recent years, a large number of banks have developed sophisticated systems and models to help bankers in quantifying, aggregating and managing risk. The outputs of these models also play increasingly important roles in banks' risk management and performance measurement processes. In this study we try to tackle the question of default prediction of short term loans for a Tunisian commercial bank. We use a

database of 924 credit records of Tunisian firms granted by a Tunisian commercial bank from 2003 to 2006. The K-Nearest Neighbor classifier algorithm was conducted and the results indicate that the best information set is relating to accrual and cash-flow and the good classification rate is in order of 88.63 % (for k=3). A curve ROC is plotted to assess the performance of the model. The result shows that the AUC (Area Under Curve) criterion is in order of 87.4% (for the first model), 95% (third model) and 95.6% for the best model with cash flow information.

Babu, Ram, and A. Rama Satish. "Impro ved of K-Nearest Neighbor Techniques in Credit Scoring." Internation al Journal For Developm ent ofComputer Science & Technology I (2013)

Credit scoring has gained more and more attentions both in academic world and the business community today. Many modeling techniques have been developed to tackle the credit scoring tasks. Credit scoring models have been widely used by financial institutions to determine if loan customers belong to either a good applicant group or a bad applicant group. The advantages of using credit scoring models are K-Nearest Neighbor can be described as the benefit from reducing the cost of credit analysis, enabling faster credit decision, insuring credit collections. This model is compare with other models depends upon the each models credit scoring.

Numerous researches have been conducted in the area of credit risk, both using classical models such as Altman Z-score and using machine learning methodology. However, the research using the data from Croatian financial institutions is scarce, especially research

focused on the selection of the demographic and/or behavior variables. In addition, it is important to develop robust models that estimate credit risk as accurately as possible. The goal of this research is to develop a data mining model for prediction of credit risk, using the data from Croatian financial institutions on defaulted clients (demographic and behavior data). Decision tree models are constructed for the prediction of credit risk. Different algorithms for the variable selection are evaluated based on the classification accuracy of the decision trees developed based on the selected variables.

3.SYSTEM ANALYSIS

3.1 Existing system:

In general, Banking Industry success and failure is based on their credit risk. The credit amount couldn't collect properly, then the bank will be loss. So, bank profit is correlated to their credit risk. Credit risk is a crucial challenge and a complex task to manage and evaluate [18]. Credit scoring tasks can be divided into two groups such as, application scoring and behavioral scoring. Application Scoring is to classify the credit applicant into ' good' and ' bad' risk groups. Behavioral scoring task is to classify the existing customers based on their payment history and personal information. Commercial loans have always been an important part of the banking industry and lenders are always measured to minimize their credit risk. To solve this credit risk problem is too difficult. Credit risk evaluation model is used to find the credit risk . To evaluate the existing customer and to classify the new potential customer using credit evaluation models

Disadvantages:
• Less security
• Less storage

3.2 Proposed system:

We propose a machine learning classifier based analysis model for credit data. We use the combination of Min-Max normalization and KNearest Neighbor (K-NN) classifier. The objective is implemented using the software package R tool. This proposed model provides the important information with the highest accuracy. It is used to predict the loan status in commercial banks using machine learning classifier.

Advantages
• More security
• More storage
• More efficient

4. SYSTEM REQUIREMENTS

H/W System Configuration:
• Processor - I3/Intel Processor
• RAM - 4GB (min)
• Hard Disk - 160GB
• Key Board - Standard Windows Keyboard
• Mouse - Two or Three Button Mouse
• Monitor - SVGA

S/W System Configuration:
• Operating System : Windows 10
• Server side Script : Python
• IDE : Google colab or Jupyter Notebook

5.Input and Output Designs

5.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people

keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.

Methods for preparing input validations and steps to follow when error occur.

5.2 OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

5.3 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

4. The output form of an information system should accomplish one or more of the following objectives.

5. Convey information about past activities, current status or projections of the

6. Future.

7. Signal important events, opportunities, problems, or warnings.

8. Trigger an action.
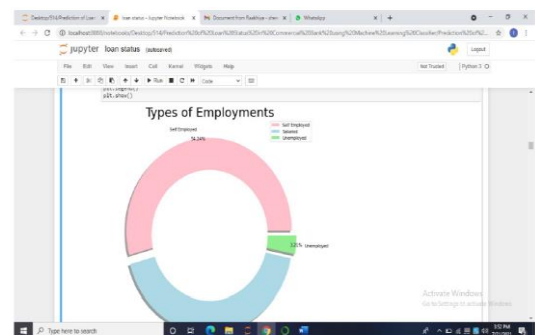
9. Confirm an action.
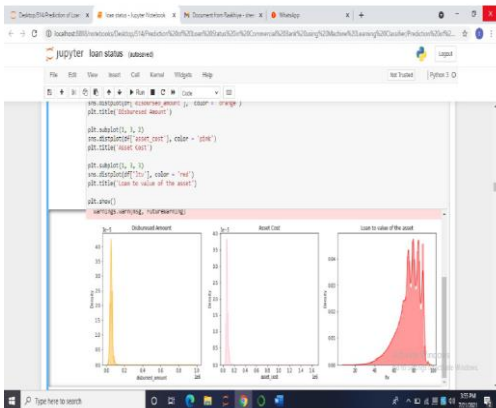

Figure 5.2.1: Types of Employments

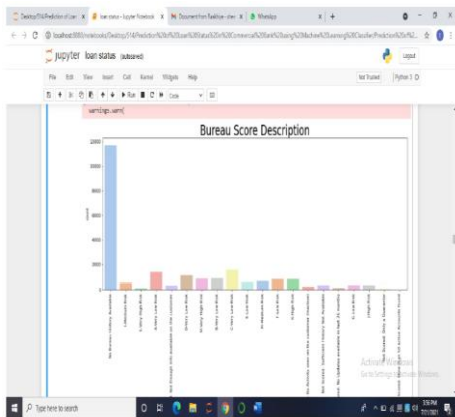Figure52.2: Disbursed amount , Asset cost,
Loan to value of the asset

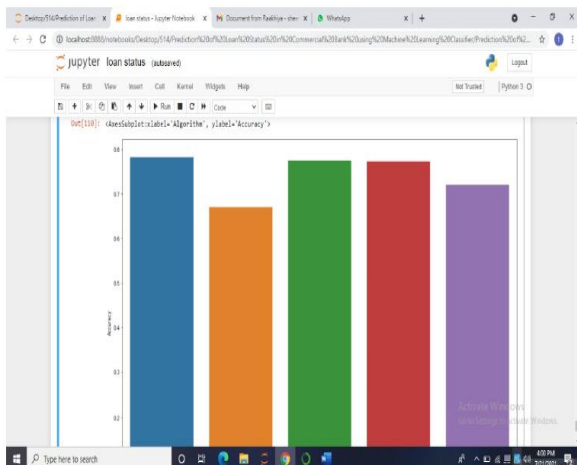Figure 5.4: Bureau Score Description

Figure12.7 Accuracy

6. CONCLUSION

In this project to proposed a loan status model to predict the loan applicant as a valid customer or default customer. The proposed model shows 75.08% accuracy result in classifying credit applicant using python and machine lerning package. The credit lenders can use this model to make a loan decision on loan proposals. Further, the comparison study has been made with different levels of iterations. The iteration level is 30 based k-NN model gives significant accuracy than other levels. This model can be used to avoid the huge loss of commercial bank.

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 Featur e Selection Methods for Credit Risk Assessment ", International Journal of Computer Science, Volume 5, Issue I, No 5, 2017.

[3] Arutjothi,G.,Dr.C.Senth amarai . "Credit Risk Evaluation using Hybrid Feature Selection Method " Software Engineering and Technology 9.2 (201 7): 23-26 .

[4] Attig, Anja, and Petra Perner. "The Problem of Normalization and a Normalized Simil arity Measure by Onlin e Data." Tran . CBR 4.1 (2011) : 3-17.

[5] Babu, Ram, and A. Rama Satish. "Impro ved of K-Nearest Neighbor Techniques in Credit Scoring." Internation al Journal For Developm ent ofComputer Science & Technology I (2013).

[6] Bach, Mirjana Pejic , et al. "Selection of Variables for Credit Risk Data Minin g Models: Preliminary research " MIPRO 2017-40 th Jubil ee International Convention. 2017.

[7] Byanjankar, Ajay, Markku Heikki la, and Jozsef Mezei. "Predicting credit risk in peer-to -peer lending: A neura l network approach." Computational Intelligence , 2015 IEEE Symposium Series on. IEEE, 2015 .

[8] Devi, CR Durga, and R. Manicka Chez ian. "A relative evaluation of the perform ance of ense mble learning in credit scoring." Advances in Computer App lication s (ICACA), IEEE International Conference on. IEEE, 2016 .

[9] G.Arutjothi, Dr.C.Senth amarai , "Effective Ana lysis of Financial Data using Knowl edge Discovery Database", Internation al Journal of Computational Intelligence and Informatics, Vol. 6: No. 2, September 2016

[10] Goel, Dr Himani , and Gurbh ej Singh "Evaluation of Expectation Maximizatio n based Cluster ing Approach for Reusability Prediction of Function based Software Systems." International Journ al 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.