iJRASET

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

www.ijraset.com

Call: ⓒ08813907089 | E-mail ID: ijraset@gmail.com

Credit Risk Analysis Using Naive Bayes in Machine Learning

Prof. S.P. Gade[1], Akshay Nalkol[2], Sainath Ibitwar[3], Durgesh Kirpal[4]

Department of Computer Engineering, PDEA'S College Of Engineering Manjari (BK), Pune, India.

Abstract: A method of data analysis called machine learning automates the creation of analytical models. It is an area of artificial intelligence based on the idea that robots can learn from data, recognise patterns, and form opinions with little help from people. One of the most important and difficult tasks facing banks is the evaluation of credit risk. To prevent financial loss, accuracy is crucial in the classification of credit data. Lending money to people who need it is a major operation in the banking sector. The depositor bank receives the interest earned by the main borrowers in order to repay the principal borrowed from them. Financial risk management is increasingly focusing on credit risk analysis. The ability to estimate credit risk, monitor it, trust the model, and process loans well are essential for decision-making and transparency. In this study, we develop binary classifiers based on machine learning and deep learning models on actual data to forecast the likelihood of loan default. In the aforementioned research, we examine various methods for credit risk analysis that are used to the assessment of credit risk datasets.

Keywords: Decision tree, KNN, K-means clustering, Bayesian classifier, Credit risk, Machine learning, and Naive Bayes classifier.

I. INTRODUCTION

One of the biggest financial issues facing the banking industry is credit risk. However, many lenders have so far been reluctant to take full advantage of the risk digitisation's predictive capacity. This is true even though a recent McKinsey analysis indicated that machine learning may reduce credit losses by up to 10% and that over half of risk managers anticipate a reduction in credit decision times of between 25% and 50%. The banking system assesses the datasets' accuracy before categorising loan applicants into good and negative classifications. The likelihood of candidates who are in good courses returning the money to the bank is very high. As a result of their poor likelihood of returning the money to the bank, applicants in subpar grades frequently default on their loans.

Accuracy is crucial for the classification of credit data since predicting credit risk is a crucial aspect of the banking industry and the biggest difficulty that all institutions must overcome. Banks suffered financial loss. This industry is motivated by the increase in the defaulter's rate in the faulty credit risk data set.

II. LITERATURE SURVEY

Following are some facts that are based on a thought analysis of various authors' works and are revealed in this area of the literature study.

1) In this study, Werasak Kuratach and Pornwatthana Wongchinsri offer a binary classification method that can categorise consumers who ask for loans. As a pre-process, Stepwise Regression (SR), a statistical method, is utilised to choose crucial features for the classifier. The classification model type that has been adopted is known as an Artificial Neural Network (ANN). For the purpose of creating an acceptable classification model, the confusion matrix concept has been employed in conjunction with business rules. According to testing, our method of SR-based Binary classification performs better than using ANN alone (91.30%) with an accuracy rate of 95.65%.

2) In this study, ShishiDahita and N.P. Singh use a hybrid strategy to improve classification accuracy for a more accurate evaluation of consumer loan creditworthiness. Two MLP neural network methods were compared to FS and bagging for improved loan distribution and improved classification accuracy. This paper introduces ANN with a focus on the MultiLayerPreception Architecture. This is followed by a discussion of the ensemble classifier approach with an emphasis on bagging and FS.

3) Bhuvaneswari This study examines a collection of credit portfolios for expensive cars that are distinguished by relevant characteristics. It seeks to evaluate the risk connected to these portfolios and, at the end, gives a prediction model that emphasises key factors and shows how those factors combine to determine whether a client is a defaulter or not. The study uses three alternative decision tree classifiers to integrate machine learning after first using traditional statistical methods.

4) Martey, Peter Bertrand Hassani, Dominique Guegan, and Add We shall concentrate on the algorithms that are used to these decisions in this study. Algorithms are employed in a wide range of contexts and industries. For instance, they are employed by businesses to hire people who fit the suggested profile. Algorithms can streamline a process, make it go more quickly and smoothly, etc. Algorithms, however, are a collection of programmes with particular goals for achieving particular goals. For instance, during the hiring process, it may introduce discrimination or a particular profile, which would "shape" the employees of the company.

III. METHOD

Different types of methods are employed for the evaluation of credit datasets in order to make accurate and trustworthy predictions about credit risk.

A. Decision Tree

A broad predictive modelling tool called decision tree analysis has a wide range of applications. Decision trees are frequently constructed using an algorithmic technique that looks for ways to segment a data collection based on a number of criteria. It is one of the most well-liked and effective methods for supervised learning.

A non-parametric supervised learning technique called decision trees is utilised for both classification and regression applications.
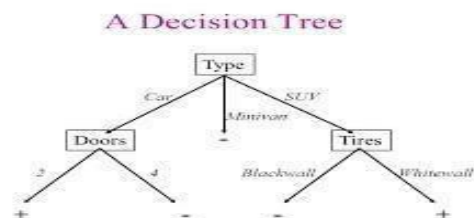

Fig.1- structure of Decision Tree.

B. Naive-Bayes Classifier

The Naive Bayes classifier is a simple and reliable method for categorising data. It is advisable to use the Naive Bayes strategy even if we are working with a data collection that has millions of records with particular features. We achieve good outcomes when we use the Naive Bayes classifier to the analysis of textual data. Natural language processing, for instance.

C. K-Means

When you have unlabeled data, you employ this particular kind of unsupervised learning. This algorithm seeks out groupings of data that match the number of groups indicated by the variable k. Based on the given attributes, the algorithm iteratively assigns each data point to one of K groups. The centroids of K clusters can be employed as labels for training data and low labelled fresh data.

D. Extreme Learning Machine

Extreme learning machines are feedforward neural networks with a single layer or multiple layers of hidden nodes, where the hidden node parameters (rather than just the weights connecting inputs to hidden nodes) need not be tuned. They are used for classification, regression, clustering, sparse approximation, compression, and feature learning.
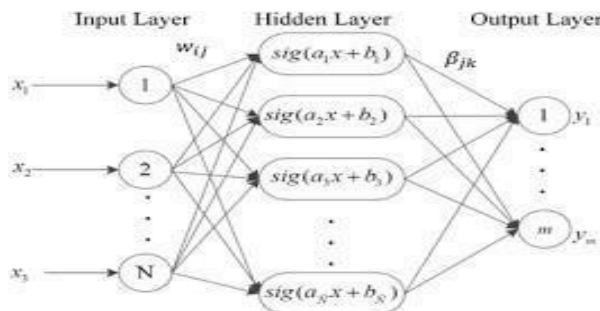

Fig.2- Extreme learning machines model.

These concealed nodes can either be inherited unchanged from their predecessors or randomly allocated and never updated (i.e., random projection but with nonlinear transforms). Typically, hidden node output weights are learned in a single step, which is equivalent to learning a linear model. These models were given the moniker "extreme learning machine" (ELM) by Guang-Bin Huang, the principal inventor.

IV. CONCLUSION

This paper does a survey of the various classifiers employed in the assessment of credit risk. Different types of classifiers, including decision trees, Naive Bayes classifiers, k-means, extreme learning machines, and ensemble classifiers, are explored in this study. In the publication, the dataset that the classifier used is discussed.

Our analysis and comparison of their accuracy using various classifier types revealed that the ELM classifier provides superior accuracy than other classifiers.

REFERENCES

[1] "Towards e-banking: the evolution of business models in financial services," International Journal of Electronic Finance, Vol. 5(4), pp. 333-356, 2011. Merisalo-Ratanen, H., Tinnila, M., and T. Lauraeus with Bask, A.

[2] International Journal of Electronic Finance, "Measuring e-statement quality influence on customer satisfaction and loyalty", Vol. 5(4), pp. 299-315, 2011.

[3] Curran, K., and Orr, J. (2011). "Increasing security by integrating geolocation into electronic financial systems." Journal of Electronic Finance International, Vol. 5(3), pp. 272-285.

[4] "Improved Sparse Least Squares Support Vector Machines", Vol. 48(1-4), pp. 1025-1031, 2002. Cawley, G., and Talbot, N.

[5] "Review and comparison of SVM and ELM based classifiers," Neurocomputing, Vol. 128, pp. 506-516, 2014. Wang, J., Zurada, M.J., and Chorowski, J.

[6] "Credit scoring and rejected instances reassigning by evolutionary computation techniques," Expert Systems with Applications, Chen, M.C., and Huang, S.H. Vol. 24(4), 2003, p. 433–441.

[7] Danenas, P., Garsva, G., and Gudas, "Credit risk evaluation using SVM classifier," International Conferences On Computational Science, 2011, pp. 1699–1709.

[8] Machine Learning, T.G. Dietterich, "Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization", Vol. 40, pp. 139-157, 2000.

[9] Computers and Operations Research, "Neural network credit scoring models," D. West, Vol. 27 (11/12), pp. 1131–1152, 2000. .

[10] Grasva, G. and Danenas, P. (2015). Expert System With Application, "Selection Of Support Vector Machine Based Classifier For Credit Risk,", Vol. 42, pp. 3194-3204.

****

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ◎ (24*7 Support on Whatsapp)