JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Credit Risk Analysis Using Naive Bayes In Machine Learning

Prof. S. P. Gade,

Akshay Nalkol , Sainath Ibitwar, Durgesh Kirpal,

Department of Computer Engineering,

PDEA's College of Engineering, Manjari, Pune, India.

1. Abstract

A key activity within the banking industry is to extend credit to customers. Hence,credit risk analysis is critical for financial risk management. There are various methods used to perform credit risk analysis. Credit risk analysis is becoming an important field in financial risk management. For the evaluation of the credit risk of the customer dataset, numerous credit risk analysis methodologies are employed. The challenging task that requires a thorough examination of the customer credit dataset or the data provided by the customer is the evaluation of the credit risk datasets that results in the decision to grant the loan of the customer or refuse the application of the customer. In this study, we review various credit risk analysis methods that are applied to the evaluation of credit risk datasets. The discussion and comparison of the best method for classifying these datasets serves as the work's conclusion. The machine learning models utilised for this paper include K-Means Clustering, Naive Bayes Classifier, Decision Trees, and Extreme Learning Machine (ELM).

Keywords: Machine Learning, Decision Tree, Naïve Bayes Classifier, K-Means Clustering, Extreme Learning Machine.

2. Introduction

A significant activity of the banking industry is to extend credit to customers. Credit risk analysis evaluates available data and decides how truthful the customer is to the financial institution. This helps in protecting financial institution from fraudulent customer. Sometimes huge loses can be avoided by a little improvement in accuracy of credit evolution of customer. As credit risk evaluation plays a very important role in the financial institutes which lend money to people and it is very critical and most difficult task faced by financial institutes.Credit risk analysis is motivated by increase in number of frauds against financial institutes.

It is important for financial institute to have risk prediction model so that they will be able to identify individuals with higher probability of being frauds. A range of machine learning methods and there application in the areas of credit risk through survey will help financial institute in avoiding frauds. There is lots of data collected by various companies which can be used in the determination of credit risk worthiness of a company or individual through the application of machine learning techniques.

3. Literature Survey

This section of the literature survey eventually reveals some facts based on thought analysis of many authors work as follows:

[1] In this study, WerasakKuratach and PornwatthanaWongchinsri 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. Based on testing, we discovered that our approach, SR-based Binary classification, outperforms utilising ANN alone (91.30%) with an accuracy rate of 95.65%.

[2] ShishiDahita, N.P. Singh adopt a hybrid approach in this research to enhance classification accuracy for a better assessment of consumer loan creditworthiness. Two MLP neural network techniques were benchmarked with FS and bagging for better loan giving and higher 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 analyses a set of credit portfolios for luxurious vehicles that are characterised by pertinent factors. 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. For a variety of purposes and in various sectors, algorithms are used. 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.

4. Methodology

We apply four machine learning classification algorithms. A summary of the methods are as follows:

[1] Naive Bayes Classifier

This algorithm is based on the Bayes Theorem. It computes the probability ofeach potential classification and selects the one with higher probability. The Naive Bayes classifier is a simple and reliable technique for classifying data. It is advised to use the Naive Bayes strategy even if we are working with a data set that has millions of records with certain qualities. When we employ the Naive Bayes classifier for textual data analysis, we get fantastic results. Natural language processing, for instance.

Naive Bayes Classification algorithm is based on Bayes' theorem. Bayes' theorem reads,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\,P(B \mid A)}{P(B)}$$

Where, P(A) is the probability of event A occurring,
P(B) is the probability of event B occurring,
P(A | B) is the probability of A given B,
P(B | A) is the probability of B given A
and P(A ∩ B)) is the probability of both A and B occurring.

In Naïve Bayes, we assume that the features are independent of each other. The classifierproceeds to calculate the probability of each class and selects the class that has the greatest probability.

[2] Extreme Learning Machine

Extreme learning machines are feed-forward neural networks with a single layer or multiple layers of hidden nodes for classification, regression, clustering, sparse approximation, compression, and feature learning, where the hidden node parameters (rather than just the weights connecting inputs to hidden nodes) need not be tuned. These camouflaged nodes can either be inherited unmodified 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 Huan, the principal inventor.

[3] K-Means

K-Means is one of the popular clustering technique or algorithm. K-Means is a unsupervised machine learning technique. It helps to form clusters from the given dataset for further data analytics. Each datapoint belongs to only one cluster.

fig 1: before applying k-means clustering

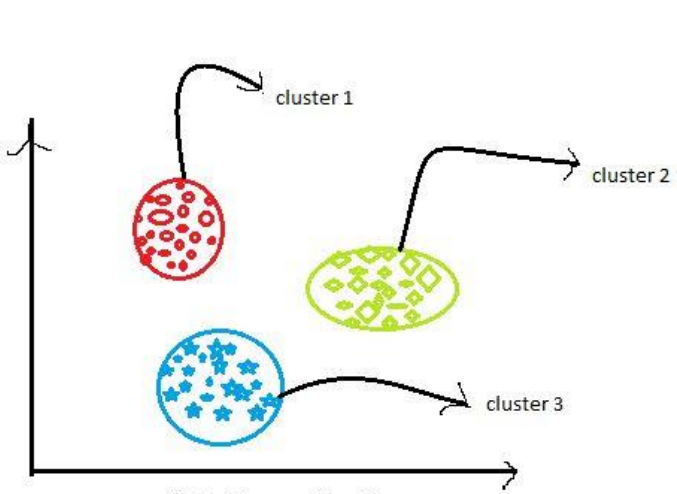
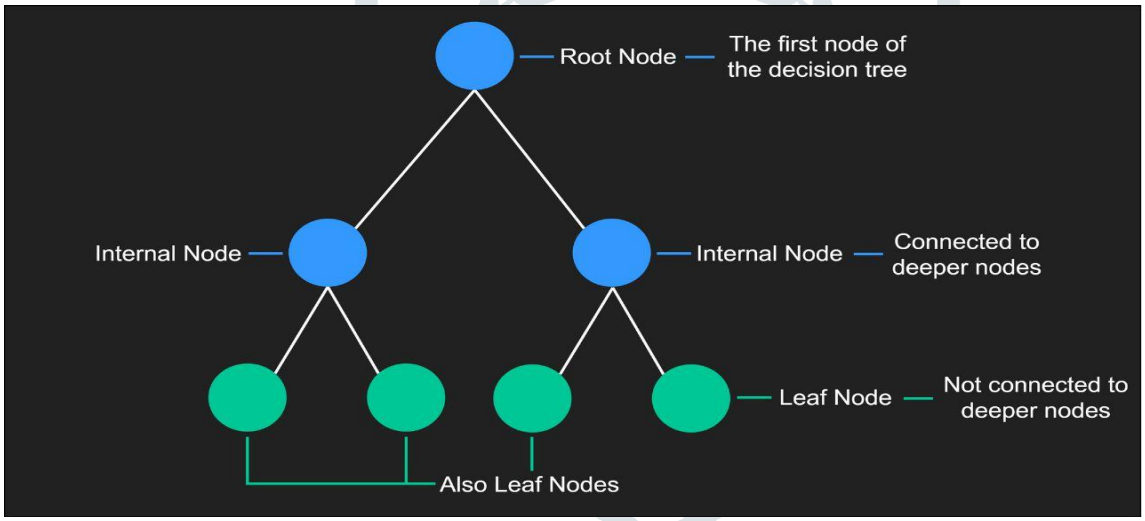


fig 2: After applying K-means clustering

[4] Decision Tree

A decision tree is a hierarchical data structure implementing the divide-and-conquer strategy. It is an efficient nonparametric method, which can be used for both classification and regression. We discuss learning algorithms that build the tree from a given labeled training sample, as well as how the tree can be converted to a set of simple rules that are easy to understand. Another possibility is to learn a rule base directly.



5. Conclusion

It is important to keep in mind that the paper's goal is to survey the various classifiers used in credit risk analysis and choose the best one. Different classifier types are discussed in this study, along with a quick overview of various ensemble classifier types. We examined and compared their accuracy using various classifier types, and we discovered from the comparison table that the ELM classifier provides superior accuracy in comparison to other classifiers.

6. 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. Bask, A., Merisalo-Ratanen, H., Tinnila, M., and Lauraeus, T..

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

[3] Curran, K., and Orr, J. (2011). "Adding geolocation to electronic financial systems for enhanced security." International Journal of Electronic Finance, Vol. 5(3), pp. 272-285.

[4] "Improved sparse least squares support vector machines," Neurocomputing, 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. Chorowski, J., Wang, J., and Zurada, M.J.
[6] Chen, M.C., and Huang, S.H., "Credit scoring and rejected instances reassigning by evolutionary computation techniques," Expert Systems with Applications, 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] T.G. Dietterich, "Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization," Machine Learning, Vol. 40, pp. 139-157, 2000.
[9] D. West, "Neural network credit scoring models," Computers and Operations Research, Vol. 27 (11/12), pp. 1131–1152, 2000.
[10] Grasva, G. and Danenas, P. (2015). "Selection Of Support Vector Machine Based Classifier For Credit Risk," Expert System With Application, Vol. 42, pp. 3194-3204.