International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

# Unveiling the Future Machine Learning Predicts Credit Card Scores

**Shaik Arshad[1], Mrs. K. Kavitha[2]**

MCA Student[1], Assistant Professor[2]

Department of Computer Applications, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India

ARTICLEINFO

**Article History:**

Accepted: 01 Aug 2023

Published: 20 Aug 2023

**Publication Issue**

Volume 9, Issue 4

July-August-2023

**Page Number**

357-364

ABSTRACT

The review of credit issuance decisions has undergone significant enhancements by incorporating manual judgement and statistical analysis into the decision-making processes. As financial institution databases grow in size, this integration has remarkably improved the reliability and efficiency of credit issuance decisions. Machine learning algorithms, especially Artificial Neural Network (ANN), have played a pivotal role in assisting with credit approval decisions. However, the varying algorithms and parameter selections among prediction models have led to differences in prediction performance. This study aims to improve model construction in the credit scoring process and analyze the forecast effectiveness of prevalent models. By setting a predetermined performance objective, numerous regression models and classifiers, including Decision Trees and Random Forest, were evaluated for their prediction accuracy. Through rigorous experimentation, ANN emerged as the top-performing model, exhibiting the highest performance score in terms of balanced accuracy. The findings of this research contribute to refining credit approval decision-making and offer valuable insights for financial institutions seeking to adopt robust machine learning models for credit scoring, ultimately enhancing the overall credit assessment process.

Keywords: Credit card, Prediction, Artificial Neural Network (ANN), Decision tree, Random Forest.

## I. INTRODUCTION

In recent years, the financial industry has witnessed a significant surge in credit card applications, necessitating more efficient and accurate credit approval processes. To meet these challenges, prediction-driven data analysis has emerged as a pivotal tool in facilitating the credit card approval process for banks. Leveraging the power of machine learning algorithms, credit scoring has demonstrated

remarkable capabilities in assessing the likelihood of customers' compliance with credit policies and making informed credit card approval decisions. Among the various models used in credit scoring, parametric models have played a fundamental role. Linear Discriminant Analysis, a prominent parametric model, has faced criticism due to its limitations in handling categorical data and potential inequalities in covariance matrices for different classes. To address these shortcomings, Logistic Analysis, another parametric model, has gained popularity by employing the cumulative distribution function of the logistic distribution. This approach has improved prediction accuracy and become widely adopted in credit scoring applications. In addition to parametric models, nonparametric models have garnered attention in credit scoring. Artificial Neural Network (ANN), Decision tree, and Random Forest are among the typical nonparametric models employed to predict creditworthiness. As researchers delve into data mining approaches, the future trend of credit scoring is expected to encompass even more complex and sophisticated methods. Recent studies by Sadatrasoul et al. have emphasized the growing preference for hybrid data mining methods in credit scoring. The integration of these methods has demonstrated higher prediction accuracy and improved time efficiency. For instance, Harris successfully integrated the Clustered method proposed by GU and Han into the credit scoring model, leading to substantial reductions in computational time compared to ANN. Moreover, the integration of neural networks has shown significant enhancements in credit score and bankruptcy prediction accuracy compared to using single neural networks. In this context, this paper aims to explore the advancements in credit scoring through prediction-driven data analysis, focusing on the evolution of parametric and nonparametric models, as well as the increasing trend of deploying hybrid data mining approaches. By understanding and leveraging these cutting-edge methodologies, financial institutions can achieve more reliable and efficient credit card approval processes, ultimately benefiting both the banks and their valued customers.

## II. RELATED WORKS

**1.C. Support-Vector Networks Machine Learning, vol:** Support-vector networks (SVNs) have emerged as a promising learning machine for two-group classification problems. By nonlinearly mapping input vectors to a high-dimensional feature space and constructing a linear decision surface within it, SVNs exhibit exceptional generalization capability. Initially limited to separable training data, this paper extends the concept to accommodate non-separable cases. Demonstrating the effectiveness of SVNs with polynomial input transformations further emphasizes their potential. To validate their performance, we conduct a comparative analysis between SVNs and various classical learning algorithms in the context of Optical Character Recognition, thus highlighting the significance of SVNs in real-world applications. This study aims to shed light on the superiority of SVNs as a versatile and powerful classification tool.

**Credit scoring in banks and financial institutions via data mining techniques:** In the realm of credit scoring, the application of data mining techniques has garnered significant attention and research during the period from 2000 to 2012. However, despite the abundance of studies in this field, there appears to be a gap in the literature regarding data mining applications specifically in credit scoring. In light of this, this paper adopts a novel research approach, conducting a comprehensive investigation through academic and systematic literature review, encompassing all relevant journals within the ScienceDirect online database. The articles are meticulously categorized based on the nature of credit scoring (enterprise, individual, and SME), as well as the types of data mining techniques employed (single classifier, Hybrid methods, and Ensembles). Variable selection methods are also thoroughly examined due to their crucial significance in addressing credit scoring challenges. The results of

this review indicate a predominant focus on individual credit scoring, with limited research in the domain of enterprise and SME credit scoring. Notably, ensemble methods, support vector machines, and neural network techniques emerge as the most favored approaches in recent studies. Among the hybrid methods, the combinations of "classification and classification" and "clustering and classification" have been most commonly employed. This paper's analysis aims to provide valuable insights and direction for future research endeavors, culminating in a series of practical recommendations for further investigations in the field.

**Credit scoring using the clustered support vector machine:** The present study explores credit scoring practices and proposes a novel approach, the clustered support vector machine (CSVM), for credit scorecard development. Traditional nonlinear SVM methods for classification suffer from computational inefficiencies with large credit scoring datasets. This research aims to address this issue by introducing the CSVM algorithm, which overcomes the limitations of conventional nonlinear SVM techniques. By comparing the CSVM with other nonlinear SVM methods, this study demonstrates that the CSVM achieves comparable classification performance while being more computationally efficient. The findings highlight the potential of CSVM as a promising tool for credit scoring and risk assessment in the financial industry.

**Using neural network ensembles for bankruptcy prediction and credit scoring:** Bankruptcy prediction and credit scoring are crucial areas extensively studied in accounting and finance literature. The use of artificial intelligence and machine learning techniques, particularly the multilayer perceptron (MLP) network with back-propagation learning, has become prevalent in solving financial decision-making problems. While combining multiple classifiers has shown promise in recent studies, their effectiveness in bankruptcy prediction and credit scoring remains unclear. In this paper, we explore the performance of a single classifier as the baseline and compare it with multiple classifiers, including diversified ones using neural networks on three datasets. Our findings shed light on making optimal financial decisions based on these classifier architectures.

**Ensemble Neural Networks for Bankruptcy Prediction and Credit Scoring: A Comparative Study:** The introduction of this comparative study on "Ensemble Neural Networks for Bankruptcy Prediction and Credit Scoring" aims to investigate the efficacy of ensemble techniques in enhancing the accuracy of bankruptcy prediction and credit scoring models. As financial institutions face increased risks in lending and investment decisions, accurate and reliable predictive models are crucial. This research focuses on employing ensemble neural networks, which combine the strengths of multiple models, to overcome the limitations of individual classifiers. By comparing the performance of ensemble-based approaches against traditional single-model methods, this study seeks to provide valuable insights into the potential benefits of ensemble techniques in improving the precision and robustness of bankruptcy prediction and credit scoring systems.

## III.Methodology

**Proposed system:**

The proposed system for credit card score prediction addresses the challenge of imbalanced data by utilizing advanced machine learning techniques. It incorporates Decision Trees, Artificial Neural Networks (ANN), and Random Forest algorithms to predict credit card scores based on annual income and salary information. By leveraging these models, the system aims to achieve more accurate and balanced predictions, enabling lenders and financial institutions to make informed decisions while mitigating risks associated with credit scoring. Through the integration of these methods, the proposed system offers a reliable and efficient solution for credit card score prediction in a highly imbalanced data environment.

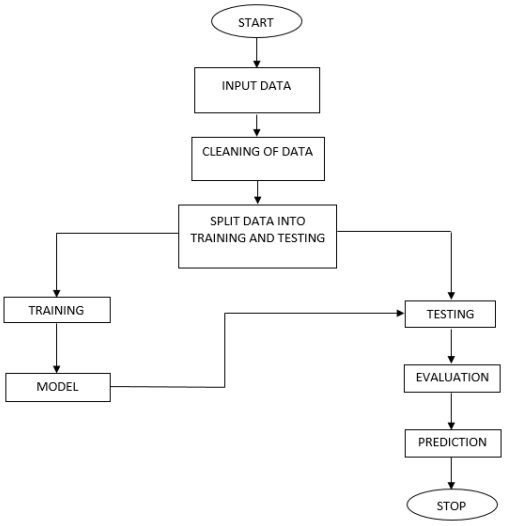



**Figure 1: Block diagram**

## IV. Implementation

The algorithms listed below were used to complete the project.

### 1. Random Forest:

Random Forest is a powerful and popular machine learning algorithm used for both classification and regression tasks. It operates by constructing multiple decision trees during the training phase and combines heir outputs to make predictions. The algorithm works as follows: Given a dataset, it randomly selects subsets of the data and features to build individual decision trees. Each tree is trained on these subsets independently, and during the prediction phase, the majority vote or the average of the predictions from all trees is taken as the final output. This approach helps to reduce overfitting and improve the model's generalization ability. Random Forest offers several advantages, such as handling large datasets, dealing with missing values, and assessing feature importance. It is widely used in various domains, including finance, healthcare, and marketing, due to its robustness and ability to handle high-dimensional data., Random Forest is a versatile and effective machine learning algorithm that provides accurate and reliable predictions, making it a valuable tool for data scientists and researchers in diverse fields.

### 2. Decision tree:

A decision tree is a popular machine learning algorithm used for both classification and regression tasks. It resembles a flowchart-like structure, where each internal node represents a feature or attribute, each branch corresponds to a decision based on that feature, and each leaf node represents a final decision or outcome. The tree is constructed by recursively partitioning the data into subsets based on the best features that maximize information gain or Gini impurity.

Decision trees are highly interpretable and useful for handling both categorical and numerical data. They excel in handling complex datasets, as they can capture non-linear relationships and interactions between variables. However, they are prone to overfitting when the tree becomes too deep. To address this, techniques like pruning and setting depth limits are commonly used. Moreover, decision trees have been widely applied in various fields, such as finance, healthcare, and marketing, due to their ability to provide insights into the decision-making process. Their simplicity, effectiveness, and ease of implementation make them a valuable tool for understanding and solving real-world problems with high accuracy.

### 3. Artificial Neural Networks:

Artificial Neural Networks (ANNs) are a subset of machine learning algorithms inspired by the biological neural networks in the human brain. They are widely used for various tasks, including pattern recognition, classification, regression, and optimization problems.

An ANN is composed of interconnected nodes, known as neurons, organized into layers: an input layer, one or more hidden layers, and an output layer. Each neuron takes inputs, applies a weight to them, performs

a non-linear activation function, and produces an output. During training, the network adjusts the weights based on the data, learning from examples to make predictions on new data.

The process of training ANNs is often referred to as "deep learning," and networks with many hidden layers are called deep neural networks. These deep architectures have revolutionized AI, leading to significant breakthroughs in image recognition, natural language processing, and game playing.

Though powerful, ANNs require substantial computational resources and large labeled datasets for training. Ongoing research aims to improve their efficiency and expand their applicability to advance various fields like healthcare, robotics, and autonomous systems.

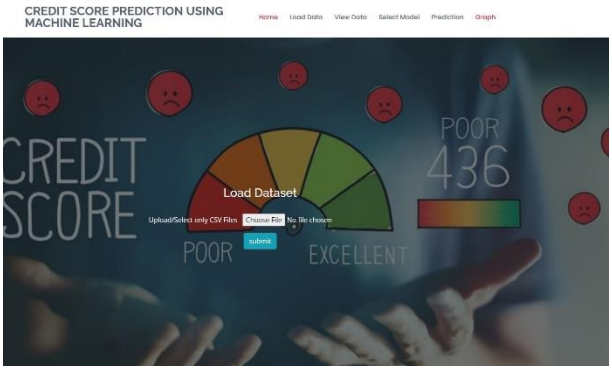
## V. Results and Discussion

The following screenshots are depicted the flow and working process of project.

**Home Page:** Here user view the home page for A Credit score prediction appellation.


**Load Page:**

User will Load The Data set

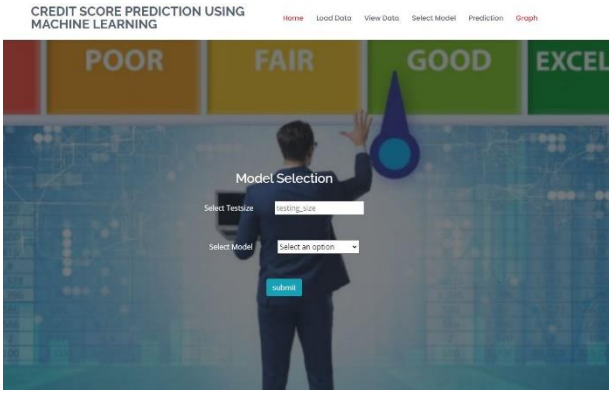



**View Page:**

User View The Data

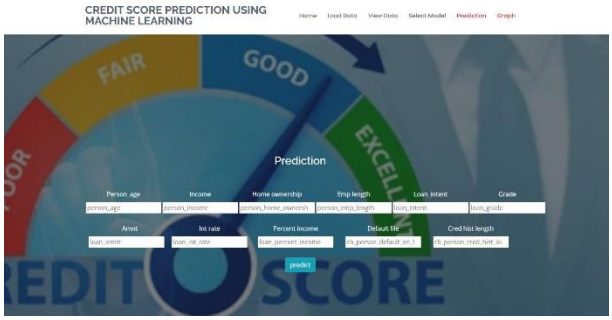

**Model:**

User will View the accuracy on every algorithm.



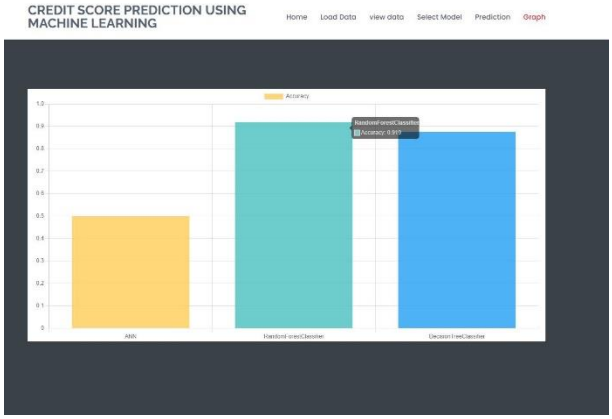

**Prediction page:**

User will give a proper input and view the result




**Graph:**

User view the accuracy graph




## VI.Conclusion

In this research, we employed three distinct machine learning techniques, namely the Decision Tree algorithm, the Random Forest algorithm, and the Artificial Neural Network (ANN), to forecast credit card scores and determine loan statuses as either non-default or default. After rigorous evaluation, we identified the best-performing techniques. By analyzing variables such as salary, income, and tax rate, our models successfully predicted credit card scores, enabling us to ascertain the likelihood of loan default. These findings hold significant implications for credit risk assessment and can aid financial institutions in making well-informed lending decisions, ultimately promoting responsible lending practices and mitigating potential credit losses.

.

## VII.REFERENCES

[1] Q. Gu and J. Han, "Clustered support vector machines", Artifi- cial Intelligence and Statistics, pp. 307-315, 2013

[2] R. Y. Goh and L. S. Lee, "Credit scoring: A review on support vector machines and metaheuristic approaches", Advances in Operations Research, vol. 2019, pp. 1-30, 2019.

[3] S. M. Sadatrasoul, M. R. Gholamian, M. Siami and Z. Hajimo-Hammadi, "Credit scoring in banks and financial institutions via data mining techniques: a literature review", Journal of AI and Data Mining, vol. 1, no. 2, pp. 119-129, 2013.

[4] W. E. Henley and D. J. Hand, "A K-nearest-neighbour classifier for assessing consumer credit risk", The Statistician, vol. 45, no. 1, pp. 77, 1996.

[5] C. Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[6] J. C. Wiginton, "A note on the comparison of logit and discriminant models of consumer credit behavior", The Journal of Financial and Quantitative Analysis, vol. 15, no. 3, pp. 757, 1980.

[7] A. K. Reichert, C. Cho and G. M. Wagner, "An examination of the conceptual issues involved in developing credit-scoring models", Journal of Business & Economic Statistics, vol. 1, no. 2, pp. 101, 1983..

[8] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees", Neural Computation, vol. 9, no. 7, pp. 1545-1588, 1997.

[9] T. Harris, "Credit scoring using the clustered support vector machine", Expert Systems with Applications, vol. 42, no. 2, pp. 741-750, 2015.

[10] C. TSAI and J WU, "Using neural network ensembles for bankruptcy prediction and credit

scoring", Expert Systems with Applications, vol. 34, no. 4, pp. 2639-2649, 2008.

**Cite this article as :**

Shaik Arshad, Mrs. K. Kavitha, "", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp. 385-391, July-August-2023.