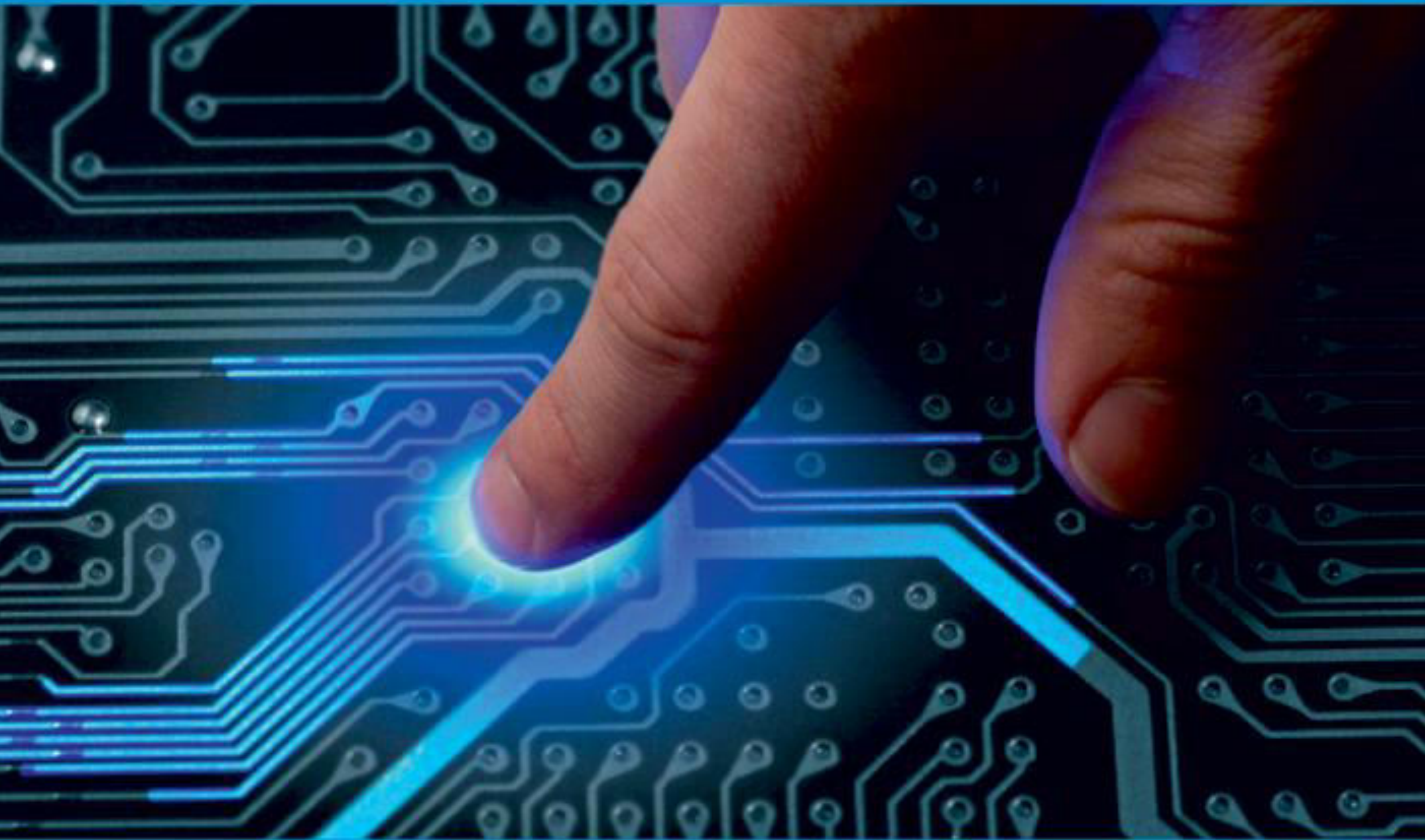
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 7, July 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107075 |

A Comparative Study of Deep Learning and Machine Learning Techniques in Credit Score Classification

Rahul Shukla[1], Rupali Sawant[2], Renuka Pawar[3]

Final Year Student, Department of Information Technology, Sardar Patel Institute of Technology, Mumbai, India[1]

Assistant Professor, Department of Information Technology, Sardar Patel Institute of Technology, Mumbai, India[2]

Assistant Professor, Department of Information Technology, Sardar Patel Institute of Technology, Mumbai, India[3]

ABSTRACT: The field of credit score classification has experienced notable progress through the introduction of deep learning (DL) and machine learning (ML) techniques, empowering financial institutions to make well-informed decisions about assessing creditworthiness. However, existing research often focuses on just a few classifiers pertaining to either ML or DL techniques, lacking a comprehensive comparative analysis between the two. This gap calls for a thorough study that evaluates and compares a wide range of ML classifiers and DL models in the context of credit scoring. Our work aims to address this limitation by presenting an extensive comparative analysis between different ML and DL approaches. We provide novel insights into the strengths and weaknesses of each model, enabling financial institutions to select the most suitable approach for their specific needs. Through conducting extensive experiments on a credit records dataset, we evaluated the accuracy, precision, recall, and F1 score of various ML classifiers, such as logistic regression, decision trees, and random forests. Additionally, we delved into the capabilities of DL models, which included multi-layer perceptron (MLP), convolutional neural networks (CNN), recurrent neural networks (RNN), and hybrid models. Our findings revealed that Random Forest achieved the highest test accuracy of 90.27, while MLP and CNN closely followed with the second-highest accuracies at 87.08 and 87.16, respectively. These results also demonstrated the potential of both MLP and CNN in credit scoring assessment. MLP's strength lies in its capacity to handle non-linear relationships between features, providing a viable alternative to decision tree-based models. On the other hand, CNN excels in capturing spatial patterns and dependencies among features, presenting a distinct advantage in credit score classification. Overall, our study presents a broad-spectrum overview of the analysis, encompassing each model's performance and effectiveness in credit score classification. The findings empower financial institutions to leverage the benefits of DL and ML techniques, optimizing their decision-making processes and enhancing risk management strategies. By selecting the most suitable credit score classification model based on the insights gained from this comparative analysis, institutions can make informed choices and effectively evaluate creditworthiness, leading to improved risk assessment and lending decisions.

KEYWORDS: Credit Score, Machine Learning, Deep Learning, Random Forest, Multi-Layer Perceptron, CNN

I. INTRODUCTION

Credit score classification plays a crucial role in the financial industry, aiding lenders and institutions in evaluating the creditworthiness of individuals and businesses. Traditionally, ML techniques have been employed to build credit scoring models, which have provided valuable insights into credit risk assessment. However, with the advent of DL techniques, there has been a surge of interest in exploring their potential to revolutionize credit score classification. DL algorithms, characterized by their ability to automatically learn intricate patterns and representations from data, offer promising advantages over traditional ML approaches. They can potentially capture complex relationships in credit data, leading to more accurate credit risk assessments and enhanced decision-making processes for lenders.

In this study, we aim to conduct a comprehensive comparative analysis to evaluate the performance and effectiveness of DL and ML techniques in credit score classification. We utilize a diverse dataset of credit records to empirically assess the accuracy, precision, recall, and F1-score of various ML classifiers, such as logistic regression, decision trees, and random forests. Additionally, we delve into the capabilities of DL models, including multi-layer perceptron (MLP), convolutional neural networks (CNN), recurrent neural networks (RNN), and hybrid models.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107075 |

Through extensive experimentation, we present a broad-spectrum overview of each model's performance in credit score classification tasks. Our objective is to provide valuable insights into the strengths and limitations of each model, thereby empowering financial institutions to make informed decisions while selecting credit score classification models that align with their specific needs. We also categorize the models according to the suitable needs of financial institutions.

The remainder of this paper is organized as follows: Section II provides an overview of the related work concerning ML and DL techniques in credit scoring systems. Section III describes the methodology of our work. Section IV discusses the results and evaluation of the proposed system with Section V concluding the paper.

II. LITERATURE REVIEW

Several studies have explored the application of machine learning (ML) and deep learning (DL) techniques in credit scoring. These studies aim to evaluate the performance and effectiveness of ML and DL models in credit score classification tasks [1]—[8]. The paper [1] proposed by Shrawan Kumar Trivedi, utilized ML classifiers in the form of Bayesian networks, Naïve Bayes classifier, SVM, C5.0, and RF of which RF achieved the highest accuracy of 93% in comparison to the other models implemented. The paper also highlighted the effectiveness of various feature selection methods, such as Chi-Square, Gain Ratio, and Information Gain, with Chi-Square showing the most promising results when used in conjunction with the RF classifier. While the study offered an extensive investigation into various ML classifiers, it overlooked the exploration of DL techniques' potential in credit scoring. In the work [2], presented by Hongyi Qian et al., the authors propose a novel SR-1D-CNN architecture to address the challenge of mining patterns between features in data that lacks spatial local correlation structure characteristics similar to images or texts. The proposed method in the study involves mapping each original feature to another feature space and then reshaping them into a multichannel form using a soft reordering mechanism. This transformed data is then input into a 1D CNN, enabling better feature extraction and modeling. Through conducting an extensive comparative analysis among various DL models, the study demonstrated that the SR-1D-CNN outperformed other deep learning models such as DeepFM, DCN-V2, and TabNet in terms of both performance and computational efficiency. While this work is extensively comprehensive in terms of the ML and DL models it implements, it lacks the inclusion of the application of these models for financial organizations. [3] Bart H. L. Overes and Michel van der Wel modeled sovereign credit ratings by undertaking a performance evaluation and comparison of different ML and DL classifiers in the form of MLP, SVM, Naïve Bayes, Ordered Logit (OL) and Classification, and Regression Trees (CART), of which MLP reported the highest random cross-validated accuracy of 68%. While this work comes close to achieving a part of the scope of our work, it falls short in terms of the poor performance of the implemented models. The work [4] proposed by Y. Song et al. presented a novel rating-specific and multi-objective ensemble classification method to address the imbalanced credit risk assessment task. The method demonstrated a better trade-off between default identification ability and overall prediction performance while considering loan companies' risk preferences. Although the study extensively explored the proposed ensemble technique and conducted performance evaluations and comparisons involving both ML and DL models, similar to [2], it did not include the practical application of these models for financial organizations. [5] Yadong Wang et al. proposed a novel deep reinforcement learning model, termed the deep Q-network with confusion-matrix-based dynamic reward function (DQN-CMDRF), for customer credit scoring. Through comprehensive experiments on five customer credit scoring datasets, the DQN-CMDRF model outperformed eight traditional classification models, demonstrating its effectiveness in improving customer credit scoring performance. While the study showcased promising results in customer credit scoring, it is important to note that our project's focus might not directly involve dynamic credit scoring or deep reinforcement learning. However, the concept of dynamic reward functions and adaptive decision-making processes remains valuable in various financial domains. This idea may be a scope for future research for our concerned study. Therefore, while not directly applicable to our project, the study's findings on dynamic modeling and adaptive decision-making hold significance for the broader field of finance and data-driven decision-making. In their proposed work [6], Peng Du and Dong Shu aimed to effectively manage the financial market and comprehensively assess personal credit to reduce the risk for financial enterprises. The authors presented an integrated deep-learning model that included RNN and BRNN models which utilized bionic optimization algorithms to overcome the limitations of shallow models and to optimize path analysis. The proposed system outperformed single deep learning models and improved the accuracy of the financial credit risk management system by 2.3%. Despite these achievements, the authors acknowledged some shortcomings, such as the need for data pre-processing and optimization of all parameters for model performance, which served as fundamental in the application of our hybrid model comprising RNN and CNN in its implementation. The work [7] presented by Maher Ala'raj et al. addressed the crucial need for effective credit card client scoring using machine learning. Their study introduced two models, MP-LSTM and PE-LSTM, which predicted the probability of missed payments and total monthly purchase

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107075 |

amounts for credit card customers, respectively. Through comprehensive experimentation and comparison with traditional classification models, the authors demonstrated the superiority of their LSTM-based models in improving consumer credit scoring. The study formulated the base on which we considered and understood the importance of deep learning models in credit score classification. We built upon the same, by introducing DL models other than the LSTM ones as proposed in their work. [8] Vincenzo Moscato et al. conducted a benchmarking study on credit risk scoring models for peer-to-peer (P2P) lending platforms. The study aimed to predict loan repayment and manage the challenges posed by high dimensionality and imbalanced data. Leveraging various machine learning classifiers in the form of LR, MLP, and RF, and sampling techniques namely undersampling, and oversampling, the authors evaluated their proposed approach on a real social lending platform dataset. By comparing various ML classifiers and DL models, the research showcased the advantages and potential of incorporating DL models alongside traditional ML classifiers. This comparison allowed for a better understanding of how DL models can address the challenges of high dimensionality and class imbalance in credit risk prediction. The study also proved beneficial in proving insight into the typical features of a lending dataset. In summary, the literature review highlights various studies exploring the application of machine learning and deep learning techniques in credit scoring, with insights into their performance, advantages, and limitations. Our research aims to build upon these findings and allow for tailored solutions that address the challenges of credit risk assessment in various financial domains.

III. METHODOLOGY

This section of our work describes the proposed approach. It consists of different stages as seen in Fig. 1.

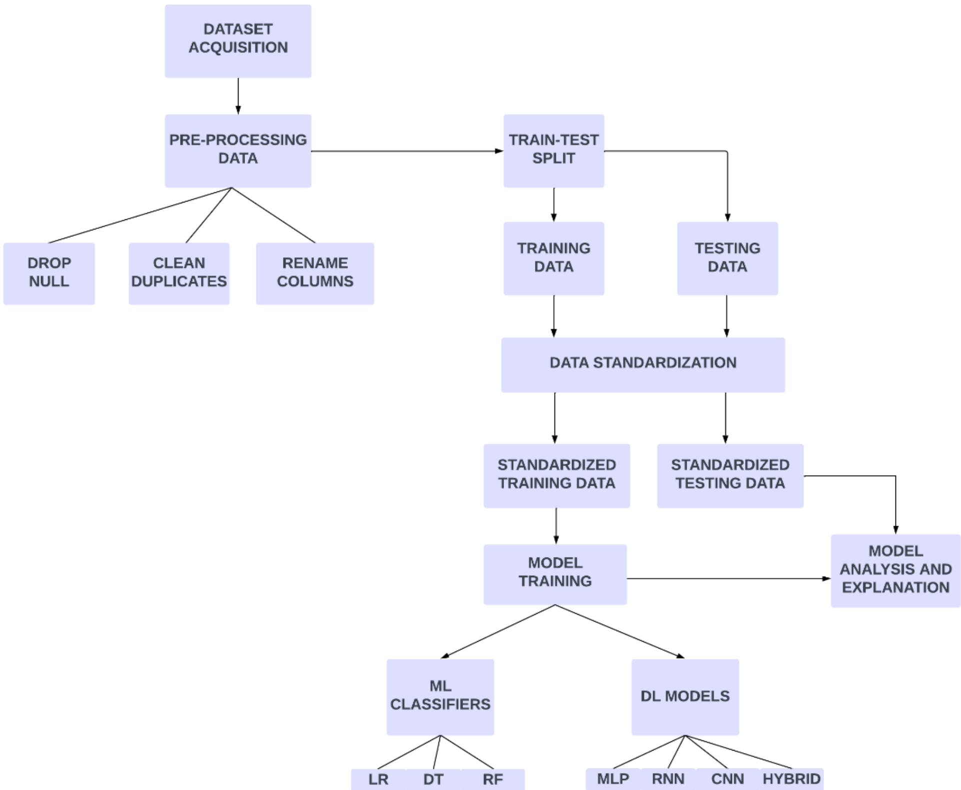


Fig. 1: Overview of the proposed methodology

We start off by first acquiring our Credit Score Dataset [9] comprising 13 columns and 5960 entries. The dataset as available, has several anomalies in the form of null values, duplicate entries, inconsistent columns, and column names. We remove these inconsistencies under the data pre-processing step which involves data cleaning, data transformation, and finally data reduction. As a result of obtaining a cleaned dataset, we split it into two parts—one for training and the other for testing, via sklearn's train_test_split() function. Approximately about 20% of the dataset is reserved for testing while the rest 80% is utilised for training. Next, the training and testing data are standardized to ensure uniformity and comparability between different features. Under data standardization, the data for every feature is reduced to a common

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107075 |

scale with a mean of 0 and a standard deviation of 1. The standardized training data is then used for model training, which involves training both ML classifiers and DL models. The ML classifiers we choose for this study are Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) whereas the DL models include MLP, CNN, RNN, and a CNN + RNN hybrid model. We test all models against a common set of performance metrics involving accuracy, precision, recall, and F1 score to gather insights into the strength and weaknesses of every model. Finally, we summarize the results under the Model Evaluation and Analysis phase.

IV. RESULTS AND DISCUSSIONS

This section describes our findings from the model analysis and evaluation phase. However, before directly jumping onto the model results, we first summarize our reduced and cleaned dataset(s) in Table 1 below. It is evident from Table 1 that the number of rows does not change as instead of dropping the null entries we simply replaced them with the column mean for that particular column. The proportion of the duplicate entries dropped and the number of null values populated causes the number of rows between the two datasets to be unchanged.

TABLE I
COMPARATIVE ANALYSIS BETWEEN ORIGINAL AND REDUCED DATASET

PARAMETER	ORIGINAL DATASET	REDUCED DATASET
Number of Columns	13	11 (col="JOB" && col="REASON" dropped)
Number of Rows	5960	5960
Number of Duplicate Entries	200	0
Max Number of Null Values for any column	1267 (col="DEBTINC")	0 (for all col)
Any Renamed columns	0	1 (col="BAD" renamed to col="TARGET")

Table 2 below compares various credit scoring models implemented, with their relative advantages or drawbacks mentioned under the "Remarks" section. The analysis of models is done in a way to provide financial institutions insight into which model can best suit their needs.

TABLE II
COMPARATIVE ANALYSIS OF IMPLEMENTED ML CLASSIFIERS AND DL MODELS

Model	Accuracy	F1-Score (Class 0)	Precision (Class 0)	Recall (Class 0)	F1-Score (Class 1)	Precision (Class 1)	Recall (Class 1)	Remarks
LR	0.8356	0.9	0.85	0.97	0.43	0.7	0.31	Relatively low precision and recall for defaulters (Class 1). May not be the best choice if accurately identifying defaulters is crucial.
DT	0.87	0.92	0.9	0.94	0.65	0.7	0.6	Shows potential for identifying

								defaulters but still has scope for improvement.
RF	0.9027	0.94	0.92	0.97	0.73	0.83	0.64	Performs well in identifying defaulters with a good balance between precision and recall.
MLP	0.8708	0.92	0.89	0.95	0.63	0.74	0.55	Alternative to Decision Tree, with the advantage of handling non-linear relationships between features.
CNN	0.8716	0.92	0.88	0.97	0.6	0.8	0.47	Shows potential for credit score classification, especially in capturing spatial patterns or dependencies among features.
RNN	0.8347	0.91	0.83	0.99	0.34	0.84	0.21	May not be the best choice for identifying defaulters, struggles with recall.
Hybrid	0.8431	0.91	0.86	0.96	0.49	0.7	0.38	Combines strengths of CNN and RNN, but does not show significant improvement over other models.

Overall, we can draw the result that for companies:

A. Prioritizing Accurate Identification of Defaulters may consider:

A.A.1 Random Forest (RF): With the highest accuracy (90.27%) and good precision and recall for defaulters (Class 1), RF is well-suited for companies that emphasize accurate identification of defaulters while maintaining a balanced trade-off between precision and recall.

A.A.2 Convolutional Neural Network (CNN): CNN also shows promise in accurately identifying defaulters (Class 1) with an accuracy of 87.16%. Its ability to capture spatial patterns and dependencies among features makes it an attractive choice for companies seeking accurate credit score classification, especially in cases where the relationships among features are complex.

B. Focusing on Balanced Performance may prioritize:

A.B.1 Decision Tree (DT): DT demonstrates balanced performance with an accuracy of 87.00%. It shows potential for identifying defaulters (Class 1) with room for improvement, making it suitable for companies that prioritize balanced performance in credit risk assessment.

A.B.2 Multi-Layer Perceptron (MLP): MLP offers an alternative to Decision Tree, with an accuracy of 87.08%. Its capability to handle non-linear relationships between features and decent performance in identifying defaulters (Class 1) make it a viable choice for companies seeking a more sophisticated model.

A.B.3 Hybrid (Combination of CNN and RNN): The hybrid model achieves an accuracy of 84.31% with balanced precision and recall for both classes. It combines the strengths of CNN and RNN, providing a reasonable option for companies seeking a balanced approach to credit score classification.

C. With Sequential Data or Temporal Dependencies may prioritize:

A.C.1 Recurrent Neural Network (RNN): Although RNN achieves an accuracy of 83.47%, it struggles with recall for defaulters (Class 1). Companies with sequential credit data or temporal dependencies may still consider RNN as it can capture temporal patterns and dependencies in the data.

V. CONCLUSION

In conclusion, this paper provides valuable insights into the application of different ML and DL approaches in credit scoring for financial institutions. By conducting an extensive comparative analysis, this study successfully addresses the previous gap in comprehensive classifier evaluation. The evaluation of accuracy, precision, recall, and F1 score offers a clear understanding of each model's strengths and weaknesses, empowering financial institutions to make informed choices in selecting the most suitable models according to their specific requirements. The classifiers with the highest potential, as presented in our findings, were Random Forest which achieved the highest accuracy of 90.27%, closely followed by MLP and CNN both at an accuracy of roughly around 87%. While RF performed well in terms of identifying defaulters with a good balance between precision and recall, MLP and CNN provided an alternative to DT by handling linear and non-linear patterns, and capturing spatial patterns between features, respectively. The work also summarized the cases wherein DT, RNN, and hybrid models may be adopted by financial organizations. Overall, it can be concluded that this work facilitates data-driven decision-making for financial institutions by providing valuable insights into the performance and suitability of various ML and DL classifiers in credit scoring.

REFERENCES

1. S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," *Technology in Society*, vol. 63, 2020, article no. 101413, ISSN: 0160-791X, doi: 10.1016/j.techsoc.2020.101413.
2. H. Qian, P. Ma, S. Gao, and Y. Song, "Soft reordering one-dimensional convolutional neural network for credit scoring," *Knowledge-Based Systems*, vol. 266, p. 110414, 2023. doi: 10.1016/j.knosys.2023.110414.
3. Overes, B.H.L., Michel van der Wel, "Modelling Sovereign Credit Ratings: Evaluating the Accuracy and Driving Factors using Machine Learning Techniques," *Computational Economics,* vol. 61, pp. 1273–1303 (2023). doi: 10.1007/s10614-022-10245-7
4. Y. Song, Y. Wang, X. Ye, R. Zaretzki, and C. Liu, "Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme," *Information Sciences*, vol. 629, pp. 599-617, 2023. doi: 10.1016/j.ins.2023.02.014.
5. Yadong Wang, Yanlin Jia, Yuhang Tian, and Jin Xiao, "Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring," *Expert Systems with Applications*, vol. 200, pp. 117013, 2022, ISSN 0957-4174, doi: 10.1016/j.eswa.2022.117013.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107075 |

6. Du Peng, and Hong Shu. "Exploration of Financial Market Credit Scoring and Risk Management and Prediction Using Deep Learning and Bionic Algorithm." *JGIM* vol.30, no.9 2022: pp.1-29. doi: 10.4018/JGIM.293286
7. Maher Ala'raj, Maysam F. Abbod, Munir Majdalawieh, Luay Jum'a, "A deep learning model for behavioural credit scoring in banks," *Neural Computing and Applications*, vol. 34, no. 8, pp. 5839-5866, Apr. 2022. doi: 10.1007/s00521-021-06695-z.
8. V. Moscato, A. Picariello, and G. Sperlí, "A benchmark of machine learning approaches for credit score prediction," *Expert Systems with Applications*, vol. 165, pp. 113986, 2021. doi: 10.1016/j.eswa.2020.113986.
9. *Home and Equity (HMEQ) Dataset.* Available: https://www.kaggle.com/datasets/ajay1735/hmeq-data

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING