Al-Zaytoonah Journal of Business, Vol. 01 issue 02, july 2025

Enhancing Credit Risk Prediction Using Deep Learning Techniques

Qeethara Al-Shayea[1] , Hanadi Saad [2], Nisrein Alarameen

[1,2] Faculty of Business, Al-Zaytoonah University of Jordan, Amman, Jordan
drqeethara@zuj.edu.jo

Abstract. Accurate credit risk prediction is crucial for financial institutions seeking to minimize loan defaults as well as make more informed decisions. The study proposes the utilization of a deep learning-based approach for loan applicants to be classified as creditworthy or non-creditworthy by deploying a multilayer perceptron (MLP) neural network. Demographic characteristics and some credit-related characteristics are utilized as input in the proposed model. To address the class imbalance problem common in credit data, SMOTEENN (Synthetic Minority Oversampling Technique–Edited Nearest Neighbours) is implemented. The quality of models is tested utilizing the major classification metrics: precision, recall, F1-score, and confusion matrix assessment. The results validate that the proposed MLP architecture effectively identifies patterns in consumer credit behavior and offers excellent predictive power, even when confronted with unbalanced data. This paper supports the potential of deep neural networks to be promising tools for enhancing credit risk evaluation in modern banking systems.

Keywords: Artificial Intelligence. Credit Risk · Prediction · Deep Neural Network

1 Introduction

In today's fast-paced and highly competitive global financial environment, banking institutions operate within a volatile market influenced by technological advancement, demand diversification, and international competition [1]. In such a climate, the ability of financial institutions to effectively manage risk—particularly credit risk—has become essential to sustaining performance and maintaining financial stability. Credit risk, defined as the likelihood of a borrower's failure to meet financial obligations, is recognized as one of the most significant risks faced by commercial banks [2].

The importance of credit risk lies in its direct impact on a bank's financial health and profitability. As loans represent the largest portion of bank assets, exposure to credit risk has wide-ranging implications for capital adequacy, revenue generation, and overall institutional resilience. Poor credit risk management practices, including inadequate borrower assessment and weak portfolio oversight, can lead to defaults, financial losses, and reputational damage [3].

The ongoing challenge of managing credit risk has driven the need for more advanced and accurate predictive models. Traditional statistical methods have limitations in handling the complexity and nonlinearity of financial data. In contrast, artificial neural networks (ANNs), particularly deep learning models like multilayer perceptrons (MLPs), offer a promising solution. These models are capable of learning intricate patterns in historical data, enabling better classification of borrowers into creditworthy and non-creditworthy categories [4].

This study aims to develop and evaluate a deep learning-based credit risk prediction model using a multilayer perceptron neural network. The model is designed to classify consumer loan applicants based on selected financial and behavioral parameters. In doing so, the study not only contributes to the growing body of research on AI-driven financial analytics but also provides practical insights for improving credit decision-making in banking institutions.

By exploring the intersection of machine learning and financial risk management, this research provides a framework for commercial banks to enhance their credit evaluation processes, mitigate default risk, and support more informed lending strategies. Ultimately, it underscores the potential of deep learning tools to transform the predictive capabilities of the banking sector.

The rest of the work has been convened as follows: Section 2 presents the literature review. In Section 3, the data collection and methodology are described. Results and discussion are illustrated in Section 4. Finally, Section 5 concludes the paper.

2 Literature Review

In the nowadays, many researchers have proposed deep learning algorithms for predicting credit score based on specific data.

Credit risk is known to be the hugely important assortment level in which it happens in obligation instruments because of the differences in obligated people and the quality of counter parties' credit. This credit chance is a vital risk for the banks. The credit chance is the peril of default on a commitment which will develop from a borrower ignoring to form required installments. For the time being, diverse machine learning (ML) and data mining (DM) algorithms have been utilized to progress different perspectives of credit scoring expectations [5]. Machine learning for credit scoring can offer assistance in analyzing a tremendous run of information to offer more precise forecasts. Machine learning models are taken into consideration as critical tools for building predictive models. Past demographic and financial data of indebted individuals is vital to constructing a computerized credit score forecast model based on a machine learning classifier [6]. Goh and Lee [7] investigated the utilization of support vector machines (SVMs) for credit scoring and illustrates their viability in dealing with classification errands. Dastile et al. [8] displayed a brief literature study of machine learning models on credit scoring applications. They also found that in common, an ensemble of classifiers performs way better than single classifiers. Even though deep learning models have not been connected broadly in credit scoring literature, they appear promising outcomes. Li [9] executed the XGBoost algorithm to distinguish the bad clients who don't pay cashback from the good clients. As a comparison, the results show that the XGBoost algorithm works much way better than logistic regression. Pincovsky et al. [10] presented a literature review of the published studies about credit scoring. Tran et al. [11] proposed a hybrid model to combine the deep learning network and genetic algorithms which are extracted rules to build a robust credit paradigm. Ariza-Garzón et al. [12] evaluated the well-known logistic regression demonstration and a few machine learning algorithms for giving scoring in peer-to-peer loaning. Their results illustrate that's possible to have machine learning algorithms accurate and straightforward. Pławiak et al. [13] presented a novel strategy, named Deep Genetic Hierarchical Network of Learners (DGHNL). The proposed strategy comprises diverse sorts of learners, counting support vector machines, K-nearest neighbors, probabilistic neural networks, and fuzzy systems.

3 Methodology

3.1 Data Collection and Preprocessing

For this study, the available "Default of Credit Card Clients" dataset of the UCI Machine Learning Repository is utilized. The dataset contains 30,000 credit card customers in Taiwan, including demographic, financial, and behavioral information accumulated from April 2005 to September 2005. The task is to predict whether a client will default on the payment in the following month.

The database has 25 variables, which are demographic characteristics such as age, gender, educational attainment, and marital status. Credit-related characteristics such as credit limit, previous bill values, payment history, and repayment status during the last six months. A payment default target variable (1 = default, 0 = no default).

A notable imbalance was observed in the target variable, with the majority of records belonging to non-defaulting customers. To address this, the SMOTEENN technique (a combination of synthetic minority oversampling and noise reduction) was used. This approach balances the dataset while preserving data integrity, ensuring more robust and fair model training.

3.2 Deep Learning Models

A multilayer perceptron (MLP) neural network is used to determine whether candidates are creditworthy or non-creditworthy based on relevant financial and behavioral features.

Several major metrics are utilized to quantify the performance of DL models. These metrics include the confusion matrix, accuracy, precision, recall, and F1 score. Precision-recall and Receiver Operating Characteristics (ROC) curves are also used to measure the performance of ML models.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{3}$$

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

4 Results and Discussion

A general evaluation of the performance of the developed deep learning model using ten-fold cross-validation is submitted in this section. The evaluation is conducted by the performance metrics computed over all the folds for the sake of generalizability and robustness. These metrics are accuracy, recall, precision, and F1-score. Furthermore, visualizations and descriptive statistics have been provided for an enhanced comprehension of the classification capability of the model.

The performance of the model in each of the 10 folds in cross-validation is shown in Table 1. It gives us the ability to observe the consistency and generalization ability of the model across different subsets of the data. The reported accuracy scores display the model demonstrates strong and consistent predictive performance across different data splits.

Table 1. The proposed model accuracy.

accuracy: 0.8478
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8419
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8562
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8474
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8394
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8510
86/86 [==============================] - 0s 2ms/step
accuracy: 0.8459
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8375
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8317
86/86 [==============================] - 0s 1ms/step
accuracy: 0.8492
86/86 [==============================] - 0s 1ms/step

Figure 1 shows the confusion matrices for all 10 folds within a 2-row figure. One confusion matrix per fold is shown. Each confusion matrix shows the classification results as True Negatives, False Positives, False Negatives, and True Positives. This presentation allows for a comparison of the performance of the model across different validation splits.

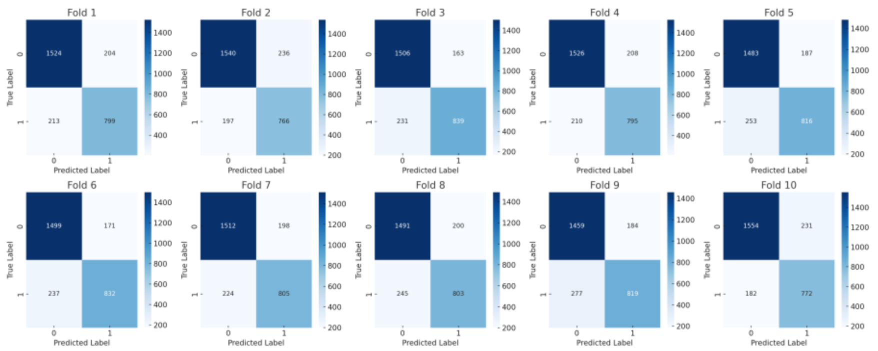

Fig. 1. Confusion matrices for all ten folds.

Table 2 summarizes the recall, precision, and F1-score values calculated for all folds in cross-validation. Recall represents the proportion of correctly identified positive samples, precision indicates the proportion of true positives out of all positive predictions, and the F1 score provides a harmonic mean of precision and recall. It facilitates an insight into the performance of the classification model over different sets of validation. Figure 2 shows the performance of the model, with precision generally higher than recall, and the F1 score serving as a consistent composite indicator. The observed variations emphasize the importance of evaluating models across multiple metrics and validation sets.

Table 2. Evaluation matrix per fold.

Fold	Recall	Precision	F1-Score
1	0.8774	0.8819	0.8797
2	0.8866	0.8671	0.8767
3	0.8670	0.9023	0.8843
4	0.8790	0.8800	0.8795
5	0.8543	0.8880	0.8708
6	0.8635	0.8976	0.8802
7	0.8710	0.8842	0.8775
8	0.8589	0.8817	0.8701
9	0.8404	0.8880	0.8636
10	0.8952	0.8706	0.8827

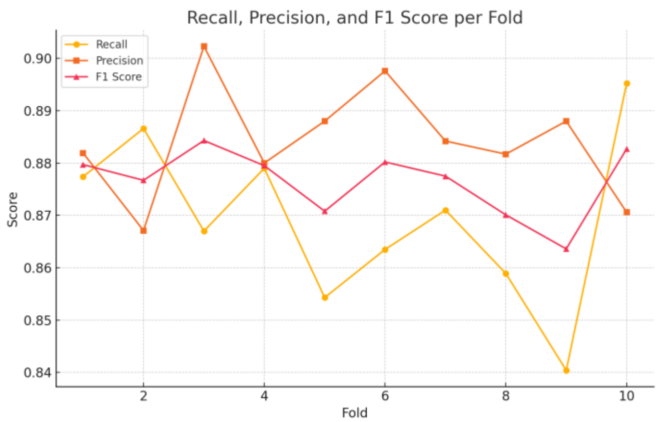

Fig. 2. Precision, Recall, and F1 Scores across Folds.

To further analyze performance, a Precision-Recall Curve (PRC) was plotted and the area under the curve (AUC-PRC) calculated. Figure 3 illustrate the Precision, Recall curve. The model had an AUC-PRC of 0.9161, indicating a good trade-off between precision and recall and good separability of positive and negative classes.

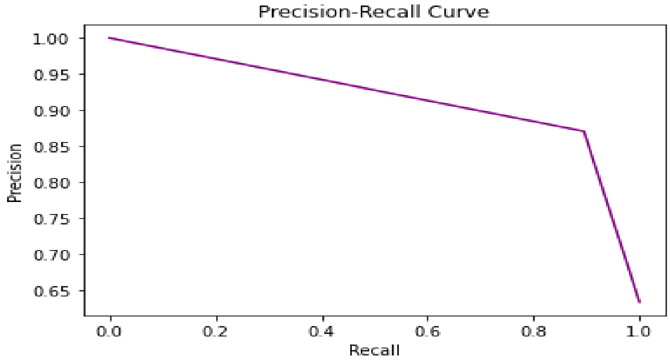

Fig. 3. Precision, Recall curve.

5 Conclusion

This study has explored the use of deep learning, specifically a multilayer perceptron (MLP) neural network, for credit risk prediction in the banking sector. By classifying loan applicants as either creditworthy or non-creditworthy based on financial and behavioral attributes, the proposed model offers a data-driven solution to enhance credit decision-making processes. The results indicate that the MLP model is able to learn complex, nonlinear relationships within the data and performs well in terms of predictive power across a range of evaluation metrics, such as precision, recall, F1-score, and confusion matrices. The promising issue in this work is addressed in the class imbalance problem by applying effective data balancing techniques, the SMOTEENN method. This preprocessing strategy significantly improved the model’s sensitivity to high-risk cases, thus growing its practical utility in real-world banking applications. Overall, the results show that the deep learning models, when properly trained and preprocessed, can outperform traditional methods in credit risk prediction tasks.

6 References

1. S. T. Cavusgil, and G. Knight (2015). The born global firm: An entrepreneurial and capabilities perspective on early and rapid internationalization. Journal of International Business Studies, 46(1), 3–16.
2. G. N. Al-Eitan, and T. O. Bani-Khalid (2019). The effect of credit risk on the performance of Jordanian commercial banks. International Journal of Economics and Financial Issues, 9(3), 246–252.
3. M. Z. Rehman, M. S. Saeed, and M. Zahid (2019). Assessing the impact of credit risk on banks' performance in Pakistan. Asian Economic and Financial Review, 9(1), 46–61.
4. C. L. Huang, M. C. Chen, and C. J. Wang (2006). Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications, 33(4), 847–856.
5. L. C. Thomas, D. B. Edelman, and J. N. Crook (2017). Credit Scoring and Its Applications. SIAM.
6. S. K. Trivedi (2020). A study on credit scoring modeling with different feature selection and machine learning approaches, Technology in Society, vol. 63, 2020,101413, https://doi.org/10.1016/j.techsoc.101413.
7. R. Y. Goh, L. S. Lee (1019). "Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches", Advances in Operations Research, vol. 2019, pages 30, https://doi.org/10.1155/2019/1974794
8. X. Dastile, T. Celik, and M. Potsane (2020). Statistical and machine learning models in credit scoring: A systematic literature survey, Applied Soft Computing, vol. 91, pp. 9809-9826.
9. Y. Li (2019). "Credit Risk Prediction Based on Machine Learning Methods," 2019 14th International Conference on Computer Science & Education (ICCSE), Toronto, ON, Canada, pp. 1011-1013, doi: 10.1109/ICCSE.2019.8845444.
10. M. Pincovsky, A. Falcão, W. N. Nunes, A. Paula Furtado and R. C. L. V. Cunha (2021). Machine Learning applied to credit analysis: a Systematic Literature Review, 2021 16th Iberian Conference

on Information Systems and Technologies (CISTI), Chaves, Portugal, pp. 1-5, doi: 10.23919/CISTI52073.2021.9476350.
11. K. Tran, T. Duong and Q. Ho (2016). "Credit scoring model: A combination of genetic programming and deep learning," 2016 Future Technologies Conference (FTC), San Francisco, CA, USA, pp. 145-149, 2016, doi: 10.1109/FTC.2016.7821603.
12. M. J. Ariza-Garzón, J. Arroyo, A. Caparrini and M. -J. Segovia-Vargas (2020). "Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending", IEEE Access, vol. 8, pp. 64873-64890, doi: 10.1109/ACCESS.2020.2984412.
13. P. Pławiak, M. Abdar, J. Pławiak, V. Makarenkov, U. R. Acharya (2020). DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring, Information Sciences, vol. 516, pp. 401-418, https://doi.org/10.1016/j.ins.2019.12.045.