

# Analysis on Credit Risk Assessment for a Multi-Purpose Cooperative Using Neural Network Algorithm

Jeffrey F. Papa<sup>1</sup>, Reagan B. Ricafort<sup>2</sup>

<sup>1</sup>Cavite State University, Philippines

<sup>2</sup>AMA University, Philippines

DOI: <https://doi.org/10.51244/IJRSI.2025.120700030>

Received: 03 July 2025; Accepted: 07 July 2025; Published: 30 July 2025

## ABSTRACT

Machine learning has become a useful tool in improving financial decision-making, especially in predicting credit risk. For multipurpose cooperatives in the Philippines, accurately identifying members who are likely to repay or default on loans is important to maintain financial stability and fairness in lending. This study aimed to compare the performance of four neural network algorithms in credit risk assessment using real-world cooperative data from 2019 to 2025. The models were evaluated based on accuracy, precision, recall, F1 score, and ROC AUC. Results showed that ANN performed the best overall, with an accuracy of 86%, a precision of 70%, a recall of 60%, an F1 score of 65%, and a high ROC AUC of 90%. RNN also showed good results, while CNN, though high in precision, had low recall. Based on the findings, ANN and RNN are recommended for cooperatives as reliable tools to support loan decision-making, helping reduce financial risks while promoting responsible and inclusive lending.

**Keywords:** Credit Risk, Neural Network Algorithm, ANN, CNN, RNN, MLP, Deep Learning

## INTRODUCTION

The broad use of data obtained from various sources, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to guide choices and actions toward the right stakeholders is known as business analytics (Davenport & Harris, 2007); Soltanpoor & Sellis, 2016)[1][2]. According to the study (Lepenioti et al., 2020) [3], to generate commercial value, business analytics aims to empower businesses to make decisions more quickly, effectively, and intelligently. Credit risk assessment is a crucial component of financial decision-making, especially within banking and lending institutions, where the accuracy of loan approvals and interest rate determinations significantly impacts profitability and market stability. According to Galindo, J., & Tamayo, P. (2000)[4], a more effective use of resources could result from a precise assessment of risk and its application in business or international financial risk models. When estimating credit risk, most deep learning models perform better than traditional machine learning and statistical algorithms, while ensemble approaches yield higher accuracy than single models (Shi, S. et.al. 2022)[5].

Models for credit risk assessment are crucial to this process because they provide a systematic approach to data analysis and default prediction. The capabilities of Deep Learning models, including multi-layer perceptrons, convolutional neural networks, recurrent neural networks, and hybrid models, offer insightful information on the use of various ML and DL models in credit scoring for financial institutions. In the study of Shukla et.al. (2023) [6], Random Forest achieved the best test accuracy of 90.27%, followed closely by MLP and CNN with accuracies of 87.08% and 87.16%. Their study helps as a reference in comparative analysis for models used in credit risk.

In the study conducted by Wang (2022) [7], A modification to an algorithm is combined with a backpropagation neural network to improve commercial banks' credit risk assessment models. While the algorithm optimizes important network parameters to boost performance, the neural network is the main modeling tool. According to experimental results, the combined model obtains an accuracy of over 65% and an acceptability rate of over

85% for evaluation results. Comparing these numbers to more conventional credit scoring techniques, which usually have an accuracy rate of about 50%, shows a notable improvement.

The Bangko Sentral ng Pilipinas article in 2021 [8] examines how machine learning might be used in central banking, specifically to improve procedures like data validation, forecasting, and monitoring. Now casting regional inflation, which enhances macroeconomic models, and identifying anomalous data to enhance data validation procedures are two important applications. This is consistent with machine learning's increasing significance in the financial industry, particularly in credit management. Neural network algorithms in credit risk assessment can increase the precision and effectiveness of forecasting financial risks and managing credit portfolios, just like the BSP uses machine learning to improve decision-making procedures and operational efficiencies.

The primary goal of this study is to compare the performance of various neural network algorithms in credit risk prediction on a Multipurpose Cooperative. Specifically, the following objectives were pursued: (1) to compare the performance of various neural network algorithms using Artificial Neural Network (ANN), (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) in credit risk assessment for a multi-purpose cooperative; (2) analyzing its performance using metrics such as accuracy, precision, recall, F1 score, and ROC AUC in real-world cooperative credit data.

## METHODOLOGY

The research procedure for this study follows a systematic approach from data collection, data preprocessing, model development, training, and metric evaluation. The key steps are as shown in figure 1.



Fig. 1. Research Procedure

### Data Collection

The study begins with data collection and preprocessing being conducted. Members' record and historical loan data was requested from the Cavite College of Fisheries Multi-Purpose Cooperative (CCF-MPC). The dataset contains 1,720 records with various members' attributes such as Credit Score, Capital Share, Age, Tenure, Balance, Monthly Salary, and the target variable Risk\_Category as shown in Figure 2.

```

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1720 entries, 0 to 1719
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  ---                -
0   RowNumber             1720 non-null   int64
1   CustomerId            1720 non-null   int64
2   MembershipStatus      1720 non-null   int64
3   CreditScore           1720 non-null   int64
4   Education             1720 non-null   object
5   CapitalShare          1720 non-null   int64
6   Gender               1720 non-null   object
7   Age                  1720 non-null   int64
8   Tenure               1720 non-null   int64
9   Balance              1720 non-null   float64
10  MaritalStatus         1720 non-null   int64
11  HasCrCard             1720 non-null   int64
12  MonthlySalary         1720 non-null   float64
13  Risk_Category         1720 non-null   int64
dtypes: float64(2), int64(10), object(2)
memory usage: 188.3+ KB
  
```

Fig. 2. Dataset info from CCF-MPC

## Data Preprocessing

The next process is data preprocessing, which involves cleaning and preparing the data. First, missing data was not a concern in our dataset since all entries were complete. Second, the columns RowNumber and CustomerId were removed because they do not contribute meaningful information to the prediction power. Third, the LabelEncoder function was used to columns Gender and Education to convert them into a numerical format suitable for a neural network. Fourth, the dataset was then split into features (X) and the target variable (y), which is the Risk\_Category indicating whether a customer is classified as low or high risk. The fifth step in the preprocessing phase is normalization. The StandardScaler function was used for normalization to ensure that features contribute equally to the model and to improve convergence during training, transforming the input data into a standardized range. The last step in preprocessing is to address the class imbalance using the Synthetic Minority Oversampling Technique or SMOTE.

## Modeling and Training

To assess the accuracy and robustness of the model, the dataset is divided into training (80%) and testing (20%) sets during the model-building and training phase. There are four different kinds of neural network models used in this study: ANN, MLP, CNN, and RNN. The preprocessed dataset is used to train each model, and hyperparameters are adjusted to maximize efficiency and performance.

*ANN:* The ANN model was built using Keras' Sequential API. To generate probabilities appropriate for binary classification, it had an input layer, two hidden layers with Rectified Linear Unit activation functions, and an output layer with a sigmoid activation function. Binary cross-entropy loss function and Adam optimizer were used for binary classification problems to create the model [9]. Samples were used to train the model across 50 epochs. To track performance and identify overfitting during training, a subset of the training data was used for validation.

*MLP:* The MLP model was constructed using TensorFlow's Keras API. ReLU activation function was used to host non-linearity, it has two hidden layers with 64 and 32 neurons, and an input layer that matched the number of features. The sigmoid activation function was used in the output layer. The Adam optimizer, a well-liked and successful neural network optimizer that uses binary cross-entropy as the loss function, was used to build the model. In order to monitor performance on unseen data during each epoch, the model was fitted on the training data for 50 epochs with a batch size of 32 and a validation split of 10% from the training set.

*CNN:* CNN model architecture began with an explicit Input layer to define the shape of the input data. The first computational layer was a Conv1D layer with 64 filters and a kernel size of 3, applying convolutional operations across the feature sequence to detect patterns. This was followed by a MaxPooling1D layer to reduce dimensionality and focus on the most prominent features, and a Dropout layer to mitigate overfitting by randomly disabling a fraction of neurons during training (Zhang et al., 2015)[10]. The output from these layers was then flattened and passed through a fully connected Dense layer for further abstraction. Another Dropout layer was added before the final Dense layer with a sigmoid activation function, which outputs a probability value indicating the likelihood of each sample belonging to the positive class. The CNN model was compiled using the Adam optimizer and binary cross-entropy as the loss function, optimized for binary classification problems. It was validated on the test set after being trained for 30 epochs with a batch size of 32 using the training data.

*RNN:* The data was reshaped to fit the input requirements of an RNN, which expects a 3D input format specifically. In this case, each data point was treated as a sequence with a single timestep, and the features represented the attributes of each sample. The dataset was split into training and testing sets to evaluate the model's generalization performance. The RNN model was also built using Keras Sequential API. Binary cross-entropy loss and the Adam optimizer were used to compile the model. Using a validation split to monitor the performance during training, it was then trained over 20 epochs with a batch size of 32.

## Evaluation

The model's performance is measured after the model has been trained. The models are evaluated using several

performance indicators, such as the Area Under the ROC Curve (AUC-ROC), F1-score, recall, accuracy, and precision.

**Accuracy:** An essential indicator for assessing a classification model's performance is accuracy, which gives a brief overview of the model's performance in terms of making accurate predictions. It is determined by dividing the total number of input samples by the number of accurate predictions [12].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

**Precision:** A model's success is gauged by its precision, which indicates the proportion of positive predictions that the model is truly right. It is especially useful in high-risk domains where false positives need to be minimized (Chicco & Jurman, 2020) [13].

$$\text{Precision} = \frac{TP}{TP + FP}$$

**Recall:** The proportion of accurately predicted positive instances to all actual positive instances is known as recall. It gauges how well every pertinent positive example is captured by the model. It is crucial when the cost of missing a positive instance (false negative) is high (Saito & Rehmsmeier, 2015) [14].

$$\text{Recall} = \frac{TP}{TP + FN}$$

**F1-Score:** The F1-Score is the harmonic mean of precision and recall. [0,1] is its range. The precision (number of instances properly classified) and robustness (number of instances missed) of our classifier are typically shown by this parameter. Great accuracy is achieved with lower recall and higher precision, but many occurrences are missed. Performance will improve with a higher F1 score.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

**AUC-ROC:** The capacity of a model to differentiate between classes at all classification thresholds is assessed using the Area Under the ROC Curve (AUC-ROC). Better model performance in separating positives from negatives is indicated by higher AUC (Fawcett, 2006) [15].

$$AUC = \int_0^1 TPR(FPR) dFPR$$

## RESULTS AND DISCUSSION

The findings and analysis of the study on credit risk assessment using neural network algorithms for the Cavite College of Fisheries Multi-Purpose Cooperative (CCF-MPC) are presented in this part. The study's dataset is made up of demographic and financial information gathered from CCF-MPC records between 2019-2025. To ascertain how well four neural network models forecast credit risk, a comparison analysis was carried out. Each model was evaluated using performance indicators such as accuracy, precision, recall, F1 score, and the AUC-ROC. To better demonstrate each algorithm's classification performance and predictive power, visual diagrams like ROC curves and the confusion matrix are presented. The results provide information on which neural network design is most suited to improving risk management in cooperative credit systems.

Table I Comparative Analysis Of Implemented Neural Network Models

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
ANN	0.86	0.70	0.60	0.65	0.90
MLP	0.85	0.68	0.60	0.64	0.89
CNN	0.84	0.75	0.42	0.54	0.84
RNN	0.85	0.69	0.59	0.64	0.90

Table 1 shows the result of the comparison of four different neural network models used to evaluate credit risk. ANN had the highest accuracy (0.86) and F1 score (0.65) among all models, it made the most correct predictions and had a good balance between precision and recall. ANN also had a strong ROC AUC score of 0.90, showing that it can distinguish between good and risky customers well. MLP and RNN performed similarly, with slightly lower accuracy (0.85) and F1 scores (0.64). Their ROC AUC scores were also high at 0.90 and 0.89, indicating reliable classification performance. CNN had the lowest recall (0.42) and F1 score (0.54), even though it had the highest precision (0.75). This means CNN was good at correctly identifying low-risk customers but missed many of the high-risk ones, making it less effective for this task. In summary, ANN performed the best overall for credit risk assessment in this study, followed closely by MLP and RNN. CNN, while precise, was not as reliable due to its low recall.

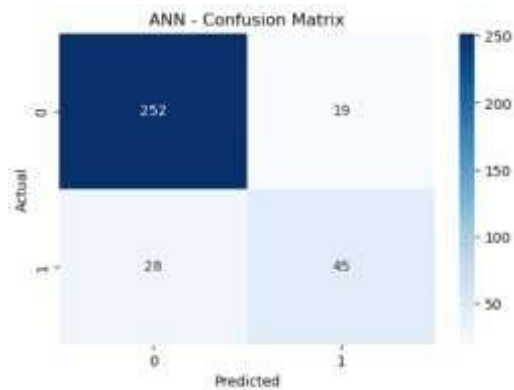


Fig. 3 Confusion Matrix for ANN Model

Figure 3 shows the confusion matrix for the ANN model. It shows that it accurately identified 252 low-risk and 45 high-risk customers, while misclassifying 19 low-risk as high-risk and 28 high-risk as low-risk. This indicates that the model performs well in recognizing safe borrowers and moderately well in detecting risky ones. The relatively low number of false positives suggests that it avoids rejecting too many good clients, while the moderate number of false negatives shows that some risky borrowers are still missed.



Fig. 4 Confusion Matrix for MLP Model

Figure 4 shows the confusion matrix for the MLP (Multilayer Perceptron) model. It shows that it correctly classified 247 low-risk and 50 high-risk customers. It made 24 false positive errors by labeling low-risk customers as high risk and 23 false negative errors by missing some high-risk customers and classifying them as low risk. These results indicate a fairly balanced performance, with good identification of both risk classes. The number of false positives and false negatives is low and nearly equal, suggesting the MLP model maintains a solid trade-off between avoiding wrongly rejected good borrowers and successfully identifying risky ones.

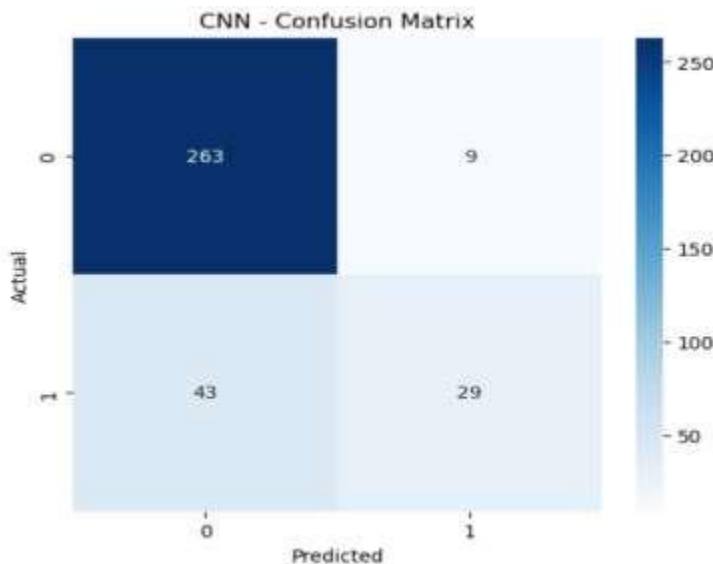


Fig. 5 Confusion Matrix for CNN Model

Figure 5 shows the confusion matrix for the CNN (Convolutional Neural Network) model. It shows that it accurately predicted 263 low-risk and 29 high-risk customers. However, it misclassified 43 high-risk customers as low risk (false negatives) and 9 low-risk customers as high risk (false positives). This indicates that while the CNN model is very good at identifying low-risk clients (high precision), it struggles to detect a significant portion of high-risk borrowers, as shown by the high number of false negatives. This weakness in recall limits the model's usefulness in credit risk prediction, as many risky borrowers may go undetected.

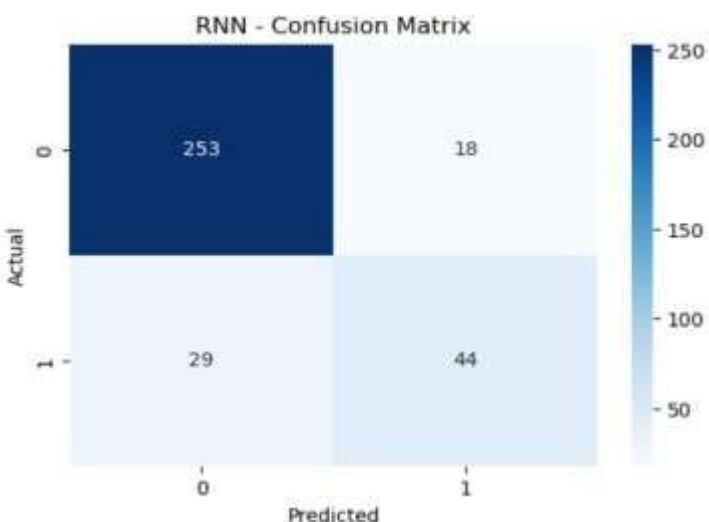


Fig. 6 Confusion Matrix for RNN Model

Figure 6 shows the confusion matrix for the RNN (Recurrent Neural Network) model. It shows that it correctly identified 253 low-risk and 44 high-risk customers. It made 18 false positive predictions by incorrectly classifying low-risk clients as high risk, and 29 false negatives by missing high-risk clients and labeling them as low risk. These results indicate a well-balanced performance, with a low number of misclassifications in both directions.



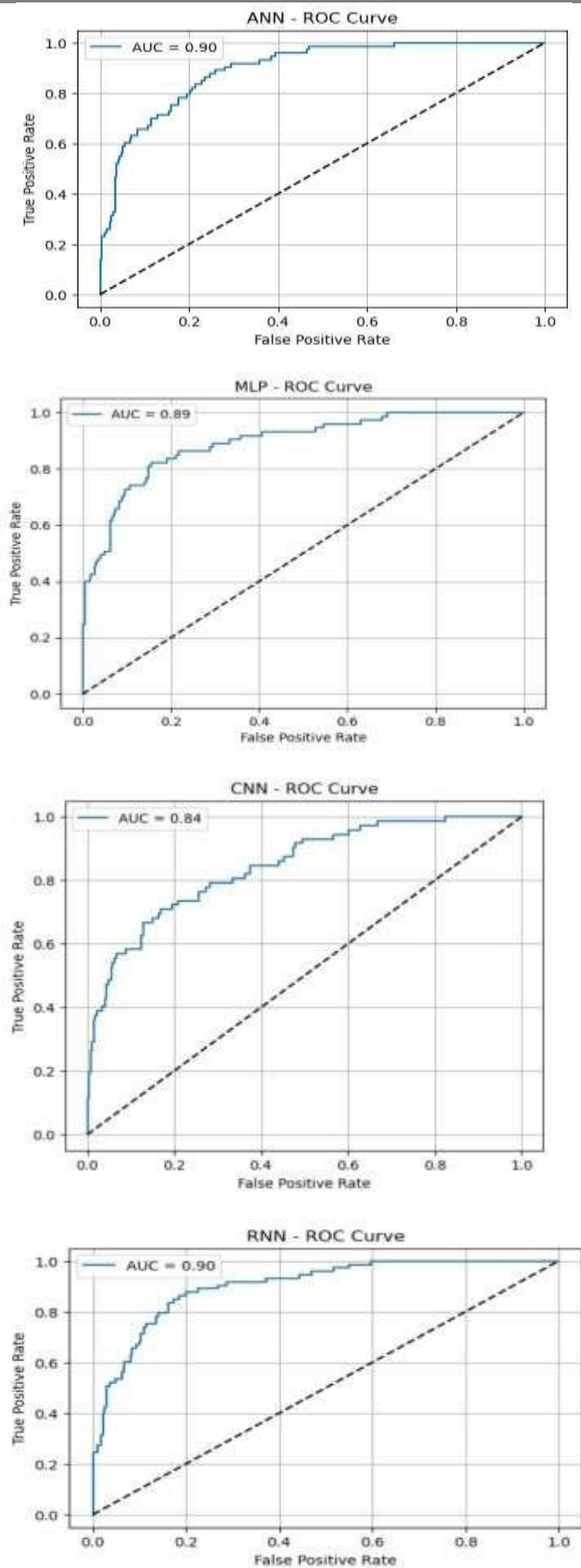


Fig. 7 ROC Curve of the 4 Neural Network Models.

Figure 7 shows the ROC curves for the four neural network models. It demonstrates their ability to distinguish between high-risk and low-risk borrowers. Both ANN and RNN achieved the highest AUC score of 0.90, indicating excellent classification performance and a strong ability to correctly rank risky clients higher than safe ones. With an AUC of 0.89, the MLP model came second, demonstrating great reliability as well. The CNN model's AUC of 0.84 indicated that it performed worse than the other models in terms of differentiating across risk categories. According to the ROC curve diagram, ANN and RNN were better at predicting credit risk.

## CONCLUSION

This study positively compared the performance of four neural network algorithms for credit risk prediction in a multipurpose cooperative using real-world data. Based on the results, the ANN model showed the best overall performance, with the highest accuracy of 86%, a good balance between precision 70% and recall of 60%, and the highest F1 score of 65%, along with a strong AUC-ROC of 90%. The RNN closely matches the ANN in most metrics. The MLP followed closely, while the CNN had the highest precision of 75% but the lowest recall of 42%, making it less effective in detecting high-risk borrowers. In conclusion, ANN and RNN are the most likely fit models for credit risk assessment in a cooperative setting, providing reliable and balanced predictions that can help reduce lending risks. These findings can guide cooperatives in adopting AI-based tools for smarter financial decision-making.

This study can significantly benefit cooperatives in the Philippines. Deploying these models comes with practical suggestions. Periodic training requirements must be considered, and it will require access to updated data and basic technical expertise. The model also needs monitoring after deployment to ensure continued accuracy. For cooperatives with limited technical setup, it is advisable to provide training to IT personnel to manage updates, and to include explainability modules to help users understand the prediction outcomes to enhance trust and transparency among members.

## ACKNOWLEDGEMENT

The author would like to thank God for the success of this research and, second, their family for the love and support. To the Cavite College of Fisheries Multipurpose Cooperative (CCF-MPC) for providing the data that is required for this study. To Dr. Reagan B. Ricafort and Dr. Maksuda Sulatana of AMA University for sharing their expertise and unending support. Finally, Mr. Papa would like to thank the Faculty and Staff Development Program of Cavite State University for their support and financial aid.

## REFERENCES

1. Davenport and Harris, 2007. T.H. Davenport, J.G. Harris. *Competing on analytics: The new science of winning*. Harvard Business Press (2007). <https://cs.brown.edu/courses/cs295-11/competing.pdf>.
2. Soltanpoor, R., Sellis, T. (2016). *Predictive Analytics for Big Data*. In: Cheema, M., Zhang, W., Chang, L. (eds) *Databases Theory and Applications. ADC 2016. Lecture Notes in Computer Science()*, vol 9877. Springer, Cham. [https://doi.org/10.1007/978-3-319-46922-5\\_19](https://doi.org/10.1007/978-3-319-46922-5_19).
3. Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020, February 1). *Predictive analytics: Literature review and research challenges*. *International Journal of Information Management*. Elsevier Ltd. <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>.
4. Galindo, J., & Tamayo, P. (2000). *Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications*. *Computational Economics*, 15(1–2), 107–143. <https://doi.org/10.1023/a:1008699112516>.
5. Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022, September 1). *Machine learning-driven credit risk: a systemic review*. *Neural Computing and Applications*. Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00521-022-07472-2>.
6. Shukla, Rahul & Sawant, Rupali & Pawar, Renuka. (2023). *A Comparative Study of Deep Learning and Machine Learning Techniques in Credit Score Classification*. *International Journal of Innovative Research in Computer and Communication Engineering*. 11. 10.15680/IJIRCCCE.2023.1107075.
7. Wang, Xiaogang, *Analysis of Bank Credit Risk Evaluation Model Based on BP Neural Network*,



- Computational Intelligence and Neuroscience, 2022, 2724842, 11 pages, 2022.  
<https://doi.org/10.1155/2022/2724842>
8. Amodia, R. A., Gabriel, V. M. C., & Mapa, C. R. (2021, December). Thinking AI-head: Exploring machine learning applications in central banks. Economic Newsletter, No. 21-03. <https://www.bsp.gov.ph/Sites/researchsite/Publications/BSP-Economic-Newsletter/EN21-03.pdf>
  9. Brownlee, J. (2019, October 23). Loss and loss functions for training deep learning neural networks. Machine Learning Mastery. <https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/>
  10. Zhang, Y., & Wallace, B.C. (2015). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. ArXiv, abs/1510.03820.
  11. Heaton, Jeffrey. (2017). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618. Genetic Programming and Evolvable Machines. 19. 10.1007/s10710-017-9314-z.
  12. Brownlee, J. (2020). Machine learning performance metrics. Machine Learning Mastery. <https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/>
  13. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
  14. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
  15. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>