International Journal of Computer, Network Security and Information System (IJCONSIST)
Vol. 7 No. 1, September 2025

Explainable Artificial Intelligence (xAI) for Reliable Financial Decision-Making in Credit Scoring System

Nabila Wafiqotul Azizah
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya, Indonesia
nabilaazizah1328@gmail.com

Muhammad Reza Pahlawan
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya, Indonesia
muhammad_reza.si@upnjatim.ac.id

Imroatul Ajizah
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya, Indonesia
imroatul_ajizah.si@upnjatim.ac.id

Mohammad Al Hafidz
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya, Indonesia
hafidz.si@upnjatim.ac.id

***Abstract*— Finance plays a vital role as one of the key elements necessary for sustaining life. Since financial stability is closely linked to overall well-being, many individuals resort to borrowing from financial institutions. As a result, the increasing number of loan applications has led to a rise in financial burdens and fund congestion within these institutions. To mitigate such risks, credit scoring has become an essential predictive approach widely adopted in financial institutions to evaluate customer creditworthiness. Through credit scoring, institutions can determine whether a customer is eligible to receive a loan. This study employs an open-source dataset obtained from Kaggle and follows the CRISP-DM methodology, which consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The research implements a classification approach by comparing two algorithms—Random Forest Regression and XGBoost. The results show that the Random Forest Regression model performs better, achieving the highest accuracy, recall, and precision, with an AUC value of 0.796 and a Coefficient of Variation (CV) of 0.712.**

Keywords - Credit scoring, XGBoost, Random Forest, Accuracy, CRISP-DM

I. INTRODUCTION

According to the Indonesian Financial Services Authority (OJK), there is an increase in the number of non-performing loans (NPL) in the range of February 2020 to February 2021. NPLs are borrowers who cannot return loans according to a predetermined schedule, so banks can experience financial disruptions. Overdue NPL can increase the product cost, and the interest rate of the credit product remains high [1]. This is not surprising because loan activity plays an important role in the banking sector. To minimize this risk, the implementation of credit risk assessment through credit scoring was important to evaluate customer eligibility. Hopefully, a bank or any other lending agency will not give credit to the wrong customer.

Credit scoring serves as a method for estimating the likelihood that potential customers may fail to repay their loans. Predicting credit risk plays a vital role in the financial industry, as it directly impacts decisions related to loan approvals, credit card issuance, and other credit-related evaluations [2], [3]. Credit scoring system on each agency must have an optimal ability to analyze customers who are eligible or not to be given a loan.

The agency can utilize a machine learning algorithm to gain valuable information about the probability of a customer to default [4]. The development of artificial intelligence through a machine learning model is extensively employed in a credit scoring system. This model was believed to offer a more robust and data-driven solution [5]. Besides that, one of the main challenges in credit risk prediction lies in the imbalance of datasets, where default cases are far fewer than non-default cases [6]. Based on previous research, machine learning algorithm with ensemble approach can be utilized to handling imbalanced credit scoring [7].

Machine learning already played an important role in propelling the lending industry sector, such as default prediction insight [8]. There are several algorithms, especially ensemble learning, that can be employed for credit scoring. Random Forest (RF), for instance, is defined as a group of unpruned decision tree models. After a large number of trees have been generated, it votes for the most popular class [9]. Besides of its ability to generate a prediction, it can be used to rank and select relevant features [10].

Extreme Gradient Boosting (XGBoost) is an advanced implementation of gradient tree to efficiently address various tasks such as regression, classification, and ranking [8]. This model already shown an excellent performance in many domains [11], [12]. XGBoost with proper parameter can result as the best overall accuracy and have particularly good AUC score [12]. Its parallel processing and decision tree optimization makes it became well-suited for credit scoring problem [13].

This study focuses on comparing the performance of two machine learning algorithms, Random Forest and XGBoost, in predicting customer loan eligibility. The primary objective is to evaluate how effectively each model can classify potential borrowers as either eligible or ineligible for loan approval. Based on our literature review, these both algorithms are widely recognized for their robustness in handling complex, high-dimensional datasets; however, their performance may vary depending on factors such as feature importance, parameter tuning, and data imbalance.

The remainder of this article is as follows: Section 2, we provide the research methodology for developing a machine learning model. In Section 3, we present the analysis results from the experiments. In the last section, we present the conclusion and potential improvement towards the research.

II. RESEARCH METHODOLOGY

This research uses a series of stages focused on five main phases to determine optimal results. The first step in this research is data acquisition, followed by preprocessing to remove outliers. After the data is cleaned, the data is processed into a model using random forest regression and xgboost. After obtaining the curve from the modeling process, the next stage is evaluation, where the data will be calculated using a confusion matrix. After this stage is completed properly, the final step is implementation using Power BI. This research stage is visualized in the form of a flowchart, which can be seen in Fig. 1.

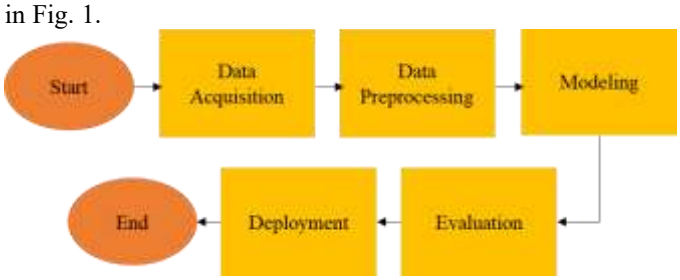

Fig. 1. Research Methodology

A. Data Acquisition

The goal of this stage is to obtain a relevant, representative, and high-quality dataset for the credit assessment analysis process. Several activities are carried out at this stage, including identifying data sources, collecting data, checking for completeness, and managing data storage and security. In this study, we used open-source secondary data obtained from Kaggle, originating from Home Credit Indonesia. This dataset is a collection of other datasets, as it contains eight supporting dataset types, each with its own function.

B. Data Preprocessing

This stage ensures the data is clean, consistent, and ready to be used for credit assessment analysis modeling. Several activities are carried out at this stage, including data cleaning, feature coding, feature scaling, feature engineering, feature selection, and feature splitting. At this stage, data cleaning is carried out, and data that does not have its own function is discarded. This stage is also used to remove outliers and redundant data.

C. Modeling

At this stage, the data will be processed using two supporting algorithms: Random Forest Regression and XGBoost, to compare their performance in predicting credit risk. The use of these two algorithms aims to determine the effectiveness of the model.

Random forest regression is one of the types of algorithms used to classify large amounts of data. Random forest works by creating several decision trees and then combining them with the aim of getting more stable and accurate predictions. In line with that, in this research random forest is implemented to predict the creditworthiness of loan applicants. In the prediction process, random forest uses various formulas as its constituent. The following is the formula of random forest regression.

Based on Fig. 2, information can be obtained stating that the random forest regression algorithm consists of different decision trees, each with the same nodes, but with different data that produce different leaves. Correspondingly, the decisions from several trees are combined to find the average of all the decision trees.

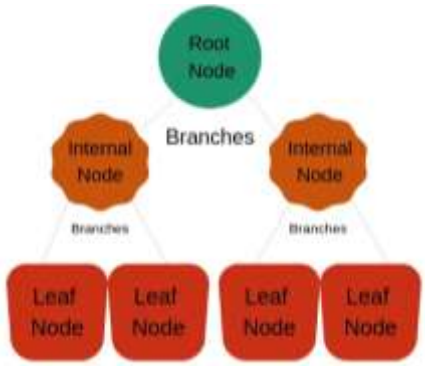

Fig. 2. Decision Tree Structure in Random Forest

XGBoost is one of the effective algorithms in data analysis using large datasets. It is very popular due to its speed and efficiency in generating highly accurate predictions. XGBoost uses ensemble learning technique, which is based on boosting technique. Based on Fig. 3, information can be obtained stating that ai and ri are regularization parameters and residuals calculated with the i-th tree. And hi is the function trained to predict the residuals.

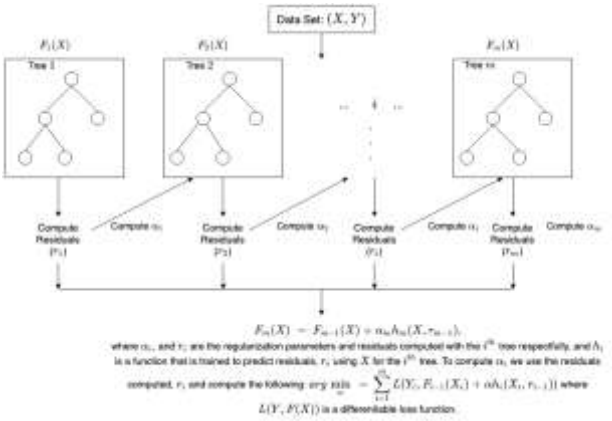

Fig. 3. XGBoost Architecture

D. Evaluation

This stage aims to evaluate and compare the performance of the two models based on relevant evaluation metrics. This stage is the stage used to interpret a modeling that has been done before. This modeling is done with Confusion Matrix and ROC Curve. Confusion matrix is one of the methods used to measure the performance of classification models that use data mining. This method is a table containing formulas to calculate accuracy, recall, precision, and error. Based on Fig. 4, information is obtained stating that accuracy is calculated based on the formula (TP+TN)/(TP+TN+FN+FP), recall is obtained from the formula (TP)/(TP+FN), while precision is obtained from the formula (TP)/(TP+FP)

		Actual	
		Positive	Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Fig. 4. Confussion Matrix

III. RESULTS AND DISCUSSION

The dataset in this study is divided into two, consisting of training data and testing data. The training data used has a percentage of 80% of the total data, while the testing data used has a percentage of 20% of the total data. In line with that, this research compares Random Forest and XGBoost.

A. Random Forest

To evaluate the diagnostic performance of the Random Forest Regression algorithm, two evaluation models were used, namely the Confusion Matrix and the ROC Curve**.** These methods were applied to assess the model's classification accuracy and to measure its sensitivity and specificity across each prediction category.

In this research, the confusion matrix is applied to two algorithms at once, one of which is random forest regression, so that the results are obtained in Fig. 5. The information is obtained stating that accuracy with the formula (52865 + 1048) / (52865 + 1047 + 3918 + 3673) so that the result is 0.451. Precision with the formula TN / TP + FP. So that the result is 0.946. Correspondingly, 0.946 belongs to class 0 and 0.054 belongs to class 1. While 0.083 belongs to class 0, and 0.1716 belongs to class 1. Then recall with the TN / TP + FN formula. So that the result is 0.797. Correspondingly, 0.797 belongs to class 0 and 0.52, belongs to class 1. While 0.0203 belongs to class 0, and 0.48 belongs to class 1.

Fig. 5. Confussion Matrix from Random Forest Model

Furthermore, an evaluation was conducted using ROC_AUC to assess the diagnostic performance of the Random Forest Regression algorithm. ROC_AUC is implemented into two algorithms at once, one of which is random forest regression. Based on Fig. 6, it can be obtained information that the graph of ROC AUC fluctuates, the highest accuracy is at max_depth with a value of 0.674. Meanwhile, the lowest value at max_depth is located at 0.669.

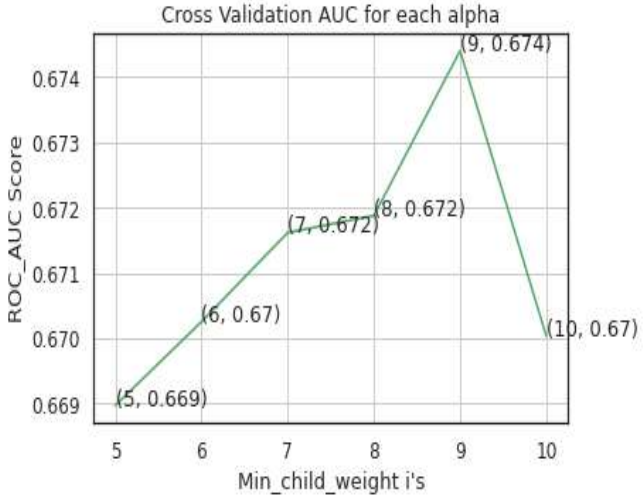

Fig. 6. ROC-AUC from Random Forest

B. XGBoost

In this research, the confusion matrix is applied to two algorithms at once, one of which is XGBoost, so that the results are obtained as Fig. 7 below. Based on the figure, the information is obtained stating that accuracy with the formula (53275 + 849) / 61503 so that the result is 0.88. Precision with the TN/TP + FP formula. So that the result is 0.946.

Correspondingly, 0.946 belongs to class 0 and 0.054 belongs to class 1. While 0.083 belongs to class 0, and 0.1716 belongs to class 1. While recall with the TN / TP + FN formula. So that the result is 0.797. Correspondingly, 0.797 belongs to class 0 and 0.52, belongs to class 1. While 0.0203 belongs to class 0, and 0.48 belongs to class 1.

Fig. 7. Confussion Matrix from XGBoost Model

Furthermore, an evaluation was conducted using ROC_AUC to assess the diagnostic performance of the Random Forest Regression algorithm. ROC_AUC is implemented into two algorithms at once, one of which is xgboost. Based on Fig. 8, it can be obtained information that the graph of ROC AUC fluctuates, the highest accuracy at max_depth with a value of 0.678. Meanwhile, the lowest value at max_depth is located at 0.656.

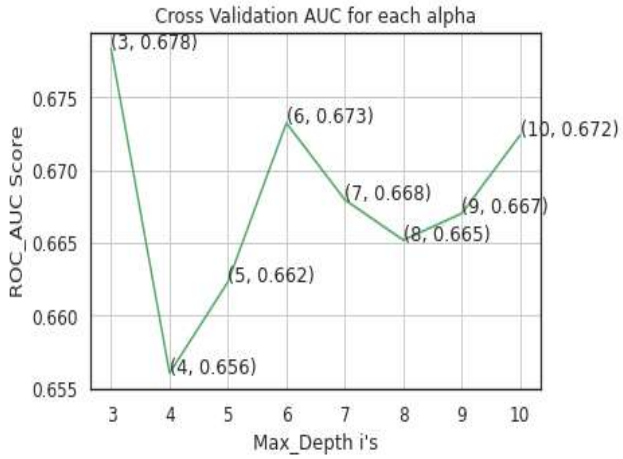

Fig. 8. ROC-AUC from XGBoost

C. The comparison of the two algorithms

The dataset in this study is divided into two, consisting of training data and testing data. The training data used has a percentage of 80% of the total data, while the testing data used has a percentage of 20% of the total data. In line with that, this research compares Random Forest Regression and XGBoost. Based on table I, information can be obtained stating that XGBoost has better performance than Random Forest Regression. This can be seen from the accuracy value which reaches 88%, the precision value obtained is 94.2%, the recall value is 94%, and 75% for ROC.

TABLE 1. Model Performance Comparison

Model	Performance			
	Accuracy	**Precision**	**Recall**	**ROC**
Random Forest	45,1%	94,6%	79,7%	79,6%
XGBoost	88%	94,2%	94%	75%

1. Random forest regression

In this study, this algorithm performed less well than the model using the XGBoost algorithm. The random forest model produced an accuracy of 45.1%, precision of 94.6%, and recall of 79.7%. Based on Fig. 9, information can be obtained stating that random forest gets a greater AUC value than XGboost, this modeling gets an AUC value of 79.6%.

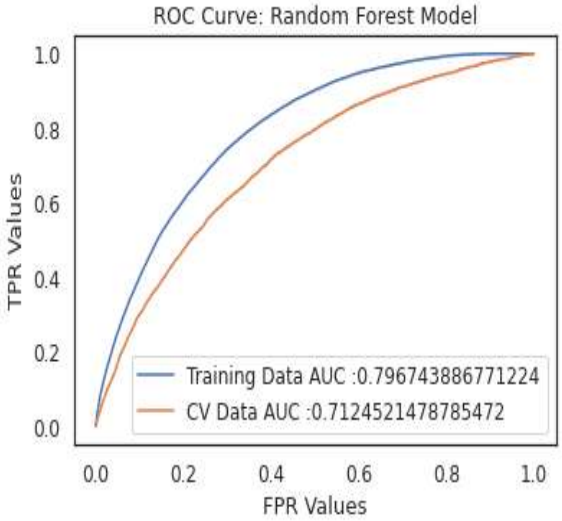

Fig. 9. ROC Curve of Random Forest

.

2. XGBoost

In this research, the XGBoost algorithm produces very good performance compared to the performance of Random Forest Regression. This can be seen from the acquisition of accuracy of 88%, precision of 94.2%, and recall of 94%. Based on Fig. 10, information can be obtained stating that XGBoost produces a lower AUC value than Random Forest. Given, this modeling only produces an AUC value of 75.2%.

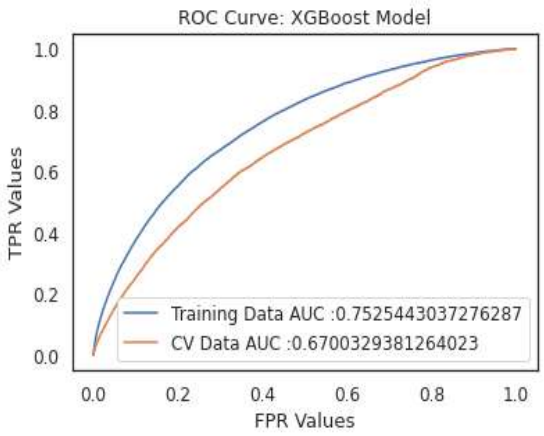

Fig. 10. ROC-Curve of XGBoost

D. Model Deployment

This Deployment stage will explain the process after applying the Machine Learning model or algorithm to the production environment. In the context of this research, we use Power BI after going through the steps of data acquisition, preprocessing, and modeling with Random Forest Regression and XGBoost algorithms.

The deployment step aims to generate loan repayment probability estimates. This estimate is obtained from data on customers who are able and unable to pay, used to predict the customer's repayment ability, and distinguish whether the customer deserves a loan. Based on Fig. 11, information can be obtained that states some useful recommendations for Home Credit Indonesia. Here are the recommendations:

- Provide longer installment period options for clients with income below 300,000
- Limit/reduce credit allocation to clients with External Source 1 values below 0.3
- Increase credit allocation to clients with more than 5 years of service

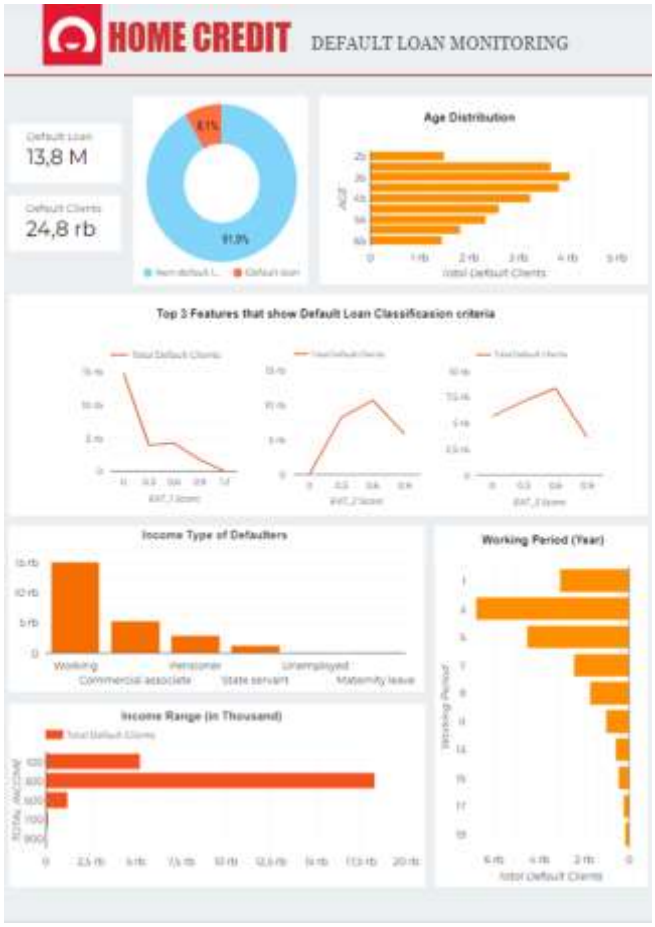

Fig. 11. Important Feature Analysis

IV. CONCLUSION

Research using a dataset from the Home Credit Default company using the XGBoost and Random Forest methods resulted in a prediction with 88% accuracy in the XGBoost method, while Random Forest resulted in 45.1% accuracy.

Based on this, the suitable algorithm to be implemented in the future is the XGBoost algorithm. Given that this algorithm has a high accuracy value compared to other algorithms. Thus, predictions related to customer eligibility in borrowing to the company are more accurate. Further development is needed in the research method to determine the feasibility of customers in lending to banks, so an algorithm that has a high accuracy value is needed so that the resulting predictions are more optimal

REFERENCES

[1] L. Liu, Q. Meng, K. Lei, and Y. Teng, "Self-control and knowledge sharing on Chinese farmers' credit default: Part-time salary as a moderator," *Systems and Soft Computing*, vol. 6, p. 200167, Dec. 2024, doi: 10.1016/j.sasc.2024.200167.

[2] Y. Li and W. Chen, "A Comparative Performance Assessment of Ensemble Learning for Credit Scoring,"

Mathematics, vol. 8, no. 10, p. 1756, Oct. 2020, doi: 10.3390/math8101756.

[3] W. Bao, N. Lianju, and K. Yue, "Integration of unsupervised and supervised machine learning algorithms for credit risk assessment," *Expert Syst Appl*, vol. 128, pp. 301–315, Aug. 2019, doi: 10.1016/j.eswa.2019.02.033.

[4] Z. Wang, Y. Tian, S. Li, and J. Xiao, "A secure cross-silo collaborative method for imbalanced credit scoring," *Eur J Oper Res*, vol. 326, no. 2, pp. 357–373, Oct. 2025, doi: 10.1016/j.ejor.2025.04.020.

[5] J. A. Darwish, "Optimization and prediction of corporate credit rating through advanced feature selection based on AI and deep learning," *Alexandria Engineering Journal*, vol. 127, pp. 586–594, Aug. 2025, doi: 10.1016/j.aej.2025.05.043.

[6] I. Aruleba and Y. Sun, "Enhanced credit risk prediction using deep learning and SMOTE-ENN resampling," *Machine Learning with Applications*, vol. 21, p. 100692, Sep. 2025, doi: 10.1016/j.mlwa.2025.100692.

[7] M. Xia, Z. Wang, X. Lan, W. Liu, and J. Wu, "Confidence-driven under-sampling decision forest for imbalanced credit scoring," *Eng Appl Artif Intell*, vol. 147, p. 110278, May 2025, doi: 10.1016/j.engappai.2025.110278.

[8] Sugiarto *et al.*, "Optimizing The XGBoost Model with Grid Search Hyperparameter Tuning for Maximum Temperature Forecasting," *Journal of Applied Data Sciences*, vol. 6, no. 4, pp. 2517–2529, Dec. 2025, doi: 10.47738/jads.v6i4.885.

[9] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst Appl*, vol. 39, no. 3, pp. 3446–3453, Feb. 2012, doi: 10.1016/j.eswa.2011.09.033.

[10] J.-Y. Lee and P.-Y. Chen, "Optimization of Heat Recovery Networks for Energy Savings in Industrial Processes," *Processes*, vol. 11, no. 2, p. 321, Jan. 2023, doi: 10.3390/pr11020321.

[11] L. Munkhdalai, T. Munkhdalai, O.-E. Namsrai, J. Y. Lee, and K. H. Ryu, "An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments," *Sustainability*, vol. 11, no. 3, p. 699, Jan. 2019, doi: 10.3390/su11030699.

[12] W. Yotsawat, P. Wattuya, and A. Srivihok, "Improved credit scoring model using XGBoost with Bayesian hyper-parameter optimization," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, p. 5477, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5477-5487.

[13] M. Alblooshi, H. Alhajeri, M. Almatrooshi, and M. Alaraj, "Unlocking Transparency in Credit Scoring: Leveraging XGBoost with XAI for Informed Business Decision-Making," in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/ACDSA59508.2024.10467573.