Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2024. Vol. 3. No. 2; e-ISSN : 2808-4519

Performance Analysis of Ensemble Learning and Feature Selection Methods in Loan Approval Prediction at Banks

Muhammad Iqbal[1*], Rizka Dahlia[2], Muhammad Ifan Rifani Ihsan[3], Lisnawanty[4], Rabiatus Sa'adah[5]

[1,4,5] *Informatics Engineering Study Program, Bina Sarana Informatika University*
[2] *Information Systems Study Program, Bina Sarana Informatika University*
[3] *Data Science Study Program, Nusa Mandiri University*
**Corresponding author E-mail: iqbal.mdq@bsi.ac.id*

Abstract

Applying for a loan at a bank has a series of relevant assessments based on data and credit scores in determining a borrower's eligibility to receive a loan from the bank. Machine learning is the basis for evaluating whether an individual is worthy of obtaining a loan, in order to reduce the potential risks faced by banks. This research aims to obtain the best accuracy value from the Loan Approval Prediction dataset which is sourced from the open dataset provider website, namely Kaggle. This Loan Approval Prediction dataset has 14 features with 4,269 data. The results of dataset analysis carried out on 4,269 data showed that the amount of data that could be studied was 4,173 data (2,599 data were approved and 1,574 data were rejected). The results of the feature importance evaluation on 14 features show that loan amount is the most important feature compared to other features, while bank asset value is the feature that has the lowest influence. Research on the Loan Approval Prediction dataset was also carried out by testing several Decision Tree ensemble models, including Extreme Gradient Boosting or XGBoost, Light Gradient Boosting Machine (Light GBM), Gradient Boosting, Random Forest, Adaptive Boosting (Adaboost) and Extra Trees. The comparison results show that the XGBoost (Extreme Gradient Boosting) model is the best model, with Accuracy 0.9974, AUC 0.9998, Recall 0.9963, Prec 0.9969, F1 0.9966.

Keywords: *Ensemble learning, Feature selection, Prediction, Loan Approval*

1. Introduction

Bank loans in Indonesia have become an important pillar in supporting various aspects of people's economic lives. Loans make it easy for people to get large amounts of money but payments are made periodically according to the agreement so that it is not burdensome for the community [1], [2]. In addition to facilitating individual needs, bank loans are also a catalyst for business growth and investment. The role of bank loans in Indonesia as a driver of economic growth is inseparable from the terms of application which serve as a basis for banks to measure the eligibility of borrowers [3], [4].

In the context of applying for a bank loan, there is a set of criteria that a potential borrower must meet. These criteria go beyond an evaluation of an individual's credit history to include an in-depth analysis of their financial stability and capacity to manage loan repayments. All of these steps are geared towards minimizing the risks associated with defaulting on loan repayments, which could potentially disrupt the financial stability of the bank [5]. In addition, as a precaution against potential non-performing loans, these criteria also take into account the ability of potential borrowers to make consistent payments. This evaluation involves an in-depth assessment of the individual's income, financial obligations, and overall financial stability [6].

By ensuring that all these criteria are met, banks can reduce the risks associated with loans that cannot be honored, and ensure safety and stability in their financial operations. This is essentially to protect the interests of the various parties involved, including the borrower and the bank itself [7].

However, when loan application criteria are not met, this can have serious repercussions for all parties involved. These risks are particularly relevant in the context of lending in Indonesia's financial sector. They include the potential for increased risk of non-performing loans, financial losses for borrowers and banks, and the possibility of affecting market confidence in the banks concerned [8].

Therefore, relevant assessments based on data and credit scores are essential in determining the eligibility of a borrower to receive a loan from a bank. Modern technologies such as machine learning provide valuable insights for banks in evaluating whether an individual is eligible or not, aiming to reduce the potential risks faced by banks [9].

There was a study conducted by Bhargav and Shashireka that compared Random Forest and Decision Trees algorithms in predicting loan approval. The results show Random Forest is more accurate (79.4490% precision, 21.0310% loss) than Decision Tree (67.2860% prediction, 32.7140% loss) [10].

In addition, Bhargav and Malathi also conducted a study comparing Random Forest with Logistic Regression in predicting non-performing loans. Random Forest has a slightly lower loss (19.1080%) than Logistic Regression (19.0970%), although the accuracy is slightly lower (80.8920% vs. 81.2030%). The two are not significantly different according to statistical tests with a 95% confidence level. Random Forest seems to be more effective in predicting non-performing loans than Logistic Regression [11].

Another study conducted by Bahrgav and Pharvathy compared the effectiveness of loan prediction between Random Forest and Naive Bayes. Although Random Forest had slightly higher accuracy (80.8920%) and lower loss (19.1080%) than Naive Bayes (80.8520% accuracy, 19.1480% loss), the difference was not statistically significant [12].
The same method was also used in a study conducted by Sravani and Mahaveerakannan who compared the loan prediction capabilities between Random Forest and Support Vector Machine algorithms. The results show that Random Forest accuracy (85.30%) is superior to SVM (75.10%) in predicting loans [13].
Another comparison was also done by Shandu, Sharma and Jassi who used several algorithms such as Random Forest, Logistic Regression, Naive Bayes, Decision Tree, and SVM to train the model. The comparison results show that Random Forest provides the highest accuracy, around 99%, compared to other algorithms [14].
In addition, Dasari, Rishitha and Gandhi conducted research on loan acceptance status using, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. Initially, when the four algorithms were run separately, they gave an accuracy result of 82%. By feeding the results from the bagged classifiers to the voting classifier, the accuracy increased to 94% [15].

2. Literature Review

Related to the case of bank loan prediction, previous research has been conducted, namely by P. Bhargav and P. Rama Parvathy L with the title "Comparing Random Forest with the Naive Bayes Algorithm with Improved Accuracy: An Effective Machine Learning Method for Loan Prediction" [12] with the problem of determining the approach in machine learning for loan prediction by comparing the Random Forest algorithm with Naive Bayes. To achieve accuracy, the Novel Random Forest Classifier is used. The calculation is done using G-power of 80%. The results of the Random Forest algorithm accuracy are 80.8920% and a loss of 19.1080%, and the results of the Naive Bayes algorithm are 80.8520% and a loss of 19.1480%, respectively. with the conclusion that loan prediction significantly seems better in the novel Random Forest Classifiers which have a significantly stronger accuracy value than Naive Bayes.
Furthermore, P. Bhargav and K. Sashirekha with the title "A Machine Learning Method for Predicting Loan Approval by Comparing the Random Forest and Decision Tree Algorithms" [10] which discusses the prediction of loan approval by comparing the Random Forest algorithm with Decision Trees. The calculation is done using G-power of 80%. The results of the random forest algorithm have a precision of 79.4490% and a loss of 21.0310%, and the results of the traditional decision tree algorithm are 67.2860% and a loss of 32.7140% respectively. With the conclusion that Random Forests are more accurate in predicting loan acceptance than Decision Trees.

2.1. Basic Theory

Data mining is a term to describe the discovery of knowledge in databases, through the use of mathematics, statistical techniques, machine learning, or artificial intelligence to identify and extract various information that can be utilized, as well as relevant knowledge from various large databases [16]. Machine learning is a branch of computer science that examines how a machine can solve problems without explicit programming [17].
Feature selection is one of the classification preprocessing stages. Feature selection is done by selecting relevant features that affect the classification results. Feature selection is used to reduce data dimension and irrelevant features. Feature selection is used to improve the effectiveness and performance efficiency of classification algorithms [18].
Prediction is an activity that can be carried out to estimate what might happen in the future, through the utilization of various old data based on certain indicators [16]. Python is one of the new programming languages today, in this programming language we are simpler and shorter in making a program, every program we make will definitely and will definitely need input and output results [19].
Extreme Gradient Boosting or XGBoost is an advanced Gradient Boosting method which is an ensemble method of models used in decision trees developed to get faster running time even in processing large data [20].
Light Gradient Boosting Machine (LightGBM) is a fast, distributed and high-performance gradient boosting method based on decision tree. LightGBM is an ensemble method that aggregates the predictions from multiple decision trees (by adding each tree) [21].
Gradient boosting is an algorithm that uses the ensemble technique of decision trees, it is able to solve data classification and prediction problems. Gradient boosting solves the problem by adjusting the weak learning with the negative gradient of the loss function and boosting the trees with parameters representing the split variables fitted to each terminal node of the tree [22].
Random Forest (RF) is an algorithm that uses a recursive binary splitting method to reach the final nodes in a tree structure based on classification and regression trees [23].
Adaptive Boosting (Adaboost) is one of several variants of the boosting algorithm. Adaboost is an ensemble learning that is often used in boosting algorithms [24].

3. Research Methodology

3.1. Research Phases

Research should have structured steps or stages so that the desired results can be achieved by the initial objectives of the research. The following are the stages carried out by the author in doing this research. which is described using a Flowchart Diagram:

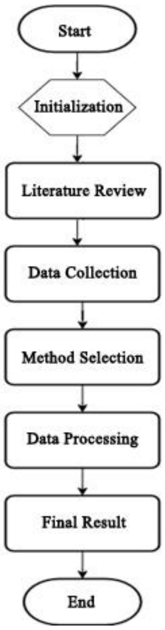

Figure 1: Research flow

3.2. Data Collection

At the data collection stage, the author carries out the process of retrieving secondary datasets from the open dataset provider website, Kaggle. In this case, the dataset taken is the Loan Approval Prediction dataset published by contributor Ajay M in 2018.

3.3. Data Analysis

The data processed in this study amounted to 604 data. It has a total of 14 attributes, including: loan id, gender, married, dependents, educations, self-employed, applicant income, co applicant income, loan amount, loan amount term, credit history, property area and loan status. With the target of predicting the status of loan applications (Loan Status).

4. Results and Discussion

4.1. Analysis of Datasets

At the stage of analyzing the dataset, the author collects a number of information that can be used in research on predictions or possibilities that bank loans can be approved. Based on the dataset used as a reference in this study, important features in determining credit acceptance decisions are described in table 1.

Table 1: Bank Loan Dataset Features

Feature	Description
Loan id	Borrower ID
No of dependents	Number of Dependents
education	Education
Self employed	Employment Status
Income annum	Income per year
Loan amount	Loan Amount
Loan term	Loan Duration
Cibil score	Sibil Score (Based on Credit History)
Residential assets value	Residential Asset Value
Commercial assets value	Commercial Asset Value
luxury_assets_value	Luxury Asset Value
Bank asset value	Bank Asset Value
Loan status	Loan Status

In table 1. the total number of features on the bank loan approval dataset is 13 fields where the Loan Id field is defined as an identity label that is not included in the calculation to determine whether the credit criteria are accepted or not, Loan status is used as the target variable. So that the initialization of input features based on the results of this research analysis is 11 features and produces one target value.
After conducting feature analysis on the dataset under study, a visualization based on 604 data obtained comparison between approved and rejected data is depicted in Figure 2.

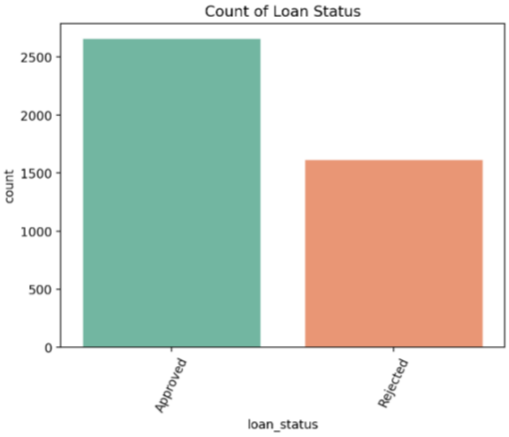

Figure 2: Visualization of Loan Status

Visualization in Figure 2 of a total of 4,269 data shows the amount of data approved/approved is 2656 and the amount rejected/rejected is 1613.
To optimize the accuracy value of the model test results, what is done at the next stage is to perform outlier removal, this stage is a process used in data analysis to identify and eliminate values that are far different from most data in the dataset. The results of Outlier Removal are shown in Figure 3 and Figure 4.

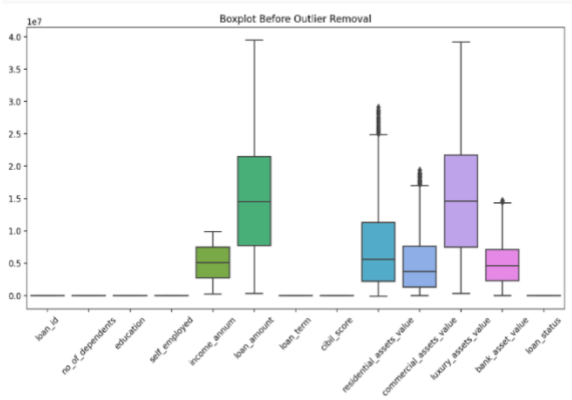

Figure 3: Before Removal of Outliers

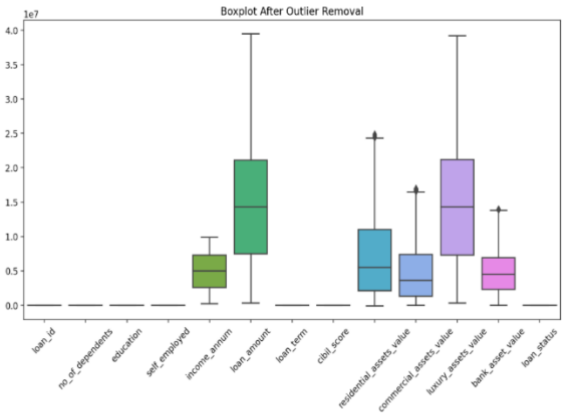

Figure 4: After Removal of Outliers

If you look at Figure 3, the black dots above or outliers are visible and long above the curve, after removing outliers the dots above the curve almost disappear or do not exist. After removing outliers or missing in the dataset, the number of data becomes 4,564 datasets.

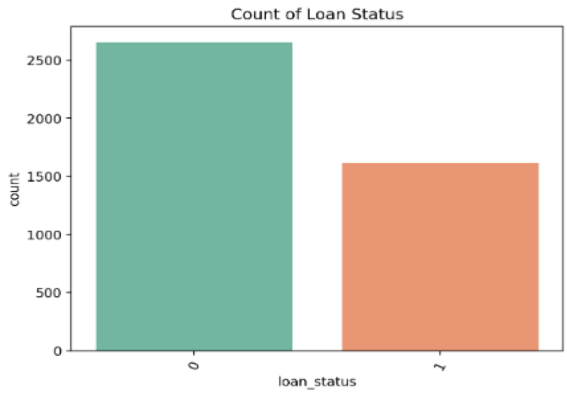

Figure 5: Visualization of Loan Status After Outlier Removal

In Figure 5 after removing outliers the amount of data studied becomes 4,173 where 2,599 is approved/approved data and 1,574 data is rejected/rejected.

4.2. Testing Methods

To get the best accuracy value in modeling the research dataset a number of models were tested to get the best accuracy. The models tested in this study are ensemble models of decision tree classifications including Extreme Gradient Boosting or XGBoost, Light Gradient Boosting Machine (Light GBM), Gradient boosting, Random Forest, Adaptive Boosting (Adaboost) and Extra Trees. The model testing results are shown in table 2.

Table 2: Model Testing Results

Model		Accuracy	AUC	Recall	Prec.	F1
xgboost	Extreme Gradient Boosting	0.9829	0.9973	0.9722	0.9824	0.9772
lightgbm	Light Gradient Boosting Machine	0.9829	0.9977	0.9714	0.9832	0.9772
gbc	Gradient Boosting Classifier	0.9796	0.9971	0.9659	0.9799	0.9728
rf	Random Forest Classifier	0.9766	0.9957	0.9619	0.9759	0.9688
ada	Ada Boost Classifier	0.9670	0.9935	0.9595	0.9536	0.9564
et	Extra Trees Classifier	0.9587	0.9911	0.9476	0.9432	0.9453

Table 2 explains the results of testing the XGBoost (Extreme Gradient Boosting) and LightGBM (Light Gradient Boosting Machine) model models have almost the same and excellent performance, with an Accuracy value of 0.9829 and AUC of more than 0.99 compared to other models tested in this study, but of course these two models have advantages and disadvantages in evaluating other models. The XGBoost model is better in the recall and timing process while LightGBM is superior in its precission and AUC values. Therefore, this study proposes that based on the results of model comparison using machine learning tools, the XGBoost (Extreme Gradient Boosting) model is the best model in determining predictions based on the results of testing the dataset.

Table 3: Best Model Testing Results

Model	Accuracy	AUC	Recall	Prec.	F1
XGBoost	0.9974	0.9998	0.9963	0.9969	0.9966

4.3. Evaluation

After getting the best model for this research, the next step is to study the hidden patterns and extract some features that are considered important and influential in predicting credit approval decisions.

4.3.1. AUC/ROC

ROC AUC (Receiver Operating Characteristic - Area Under Curve) is used in statistics and machine learning to evaluate the performance of classification models. Here are the results in figure 6.

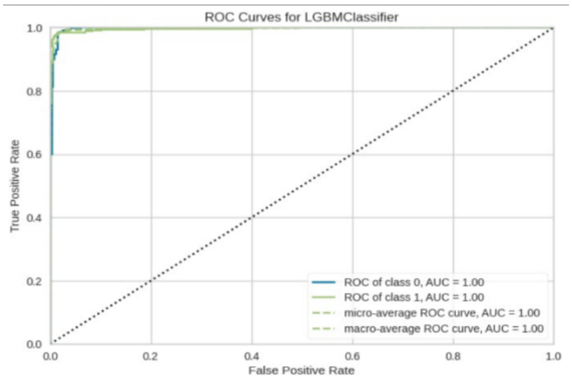

Figure 6: Evaluation of AUC/ROC

In Figure 6, the AUC/ROC value is almost close to 1, indicating a very good classification as shown in the bright green line.

4.3.2. Confusion Matrix

In classifications that produce binary Yes and No decision values, the confusion matrix is used to measure how much the model can help in evaluating the model more holistically, not only based on accuracy, but also other important aspects such as precision and recall. The confusion matrix is shown in Figure 7.

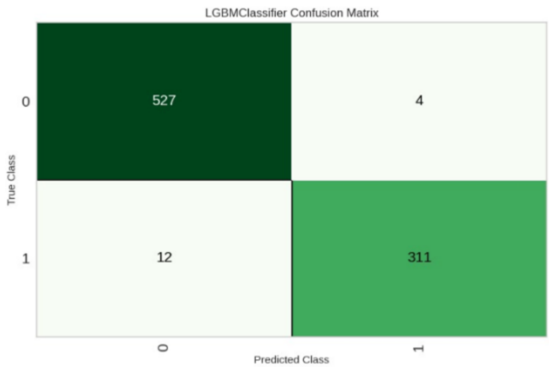

Figure 7: Evaluation of Confusion Matrix

In Figure 7, it is explained that the dark green color shows the evaluation of datasets that are predicted to be wrong, the number of data predicted to be wrong but the result is correct is 4 data, the number of datasets predicted to be correct but the result is wrong is 12 datasets and 311 datasets are declared correct and the results are also correct shown in the light green box.

4.3.3. Feature Importance

Feature Importance is an approach in machine learning that aims to identify and measure how significant each feature (variable) in a dataset is to the performance of a predictive model. This concept is very useful, especially in complex models, to understand what factors are most influential in making predictions. Figure 8 below shows the evaluation results based on the most important features.

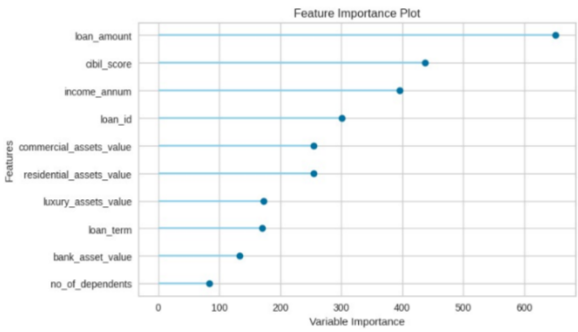

Figure 8: Results of Feature Importance Evaluation

Based on the results in figure 8, here are some research results:

1. Loan amount is the most important feature with a much higher importance value than other features. This shows that the loan amount has a very significant influence in the model prediction.
2. Cibil score emerged as the second most important feature, indicating that the credit score of the applicant also has a significant influence on the model's decision.
3. Income annum and commercial assets value are also considered important, but at a lower value than the top two features.

4. Features such as bank asset value and no of dependents seem to have the least influence in this model, corresponding to their lower relative importance values.

4.3.4. Evaluation of Prediction Model

Prediction model evaluation is used to measure how much accuracy score is generated by the model based on input features and produces a target or prediction. Based on the prediction model evaluation results shown in table 3.

Table 3: Prediction Model Results

no of dependents	Education	self employed	income annum	loan amount	loan term	cibil score	residential assets value	commercial assets value	luxury assets value	bank asset value	loan status	prediction label	prediction score
2	1	0	9600000	29900000	12	778	2400000	17600000	22700000	8000000	0	0	0.9999
0	0	1	4100000	12200000	8	417	2700000	2200000	8800000	3300000	1	1	0.9999
3	1	0	9100000	29700000	20	506	7100000	4500000	33300000	12800000	1	1	0.9994
3	1	0	8200000	30700000	8	467	18200000	3300000	23300000	7900000	1	1	0.9998
5	0	1	9800000	24200000	20	382	12400000	8200000	29400000	5000000	1	1	0.9999

In Table 3, the input value based on 11 features produces one target variable loan status where the predicted value of the result is the same as the actual value by displaying the prediction of the average score value above 99%.

5. Conclusion

This research using Loan Approval Prediction dataset has 14 features with 4,269 data. The results of dataset analysis conducted on 4,269 data show that the amount of data that can be studied is 4,173 data (2,599 approved/approve data and 1,574 rejected/reject data). This study tested several Decision Tree ensemble models, including Extreme Gradient Boosting or XGBoost, Light Gradient Boosting Machine (Light GBM), Gradient Boosting, Random Forest, Adaptive Boosting (Adaboost) and Extra Trees. The test results show that the XGBoost (Extreme Gradient Boosting) and LightGBM (Light Gradient Boosting Machine) models. The model comparison results show that the XGBoost (Extreme Gradient Boosting) Model is the best model, with an accuracy value of 0.9974, AUC 0.9998, Recall 0.9963, Prec 0.9969, F1 0.9966. The results of the evaluation of feature importance on 14 features show that loan amount is the most important feature compared to other features, while bank asset value is the feature that has the lowest influence. And in the prediction model produces one target variable loan status where the predicted value of the results is the same as the actual value by displaying the prediction of the average score value above 99%.
Suggestions for future research include testing sampling and optimization techniques that have not been applied in this study. This could be done by exploring novelty in dataset size, or through a banking case study, particularly in the context of feature selection to predict bank loan approval.

Acknowledgement

The author would like to thank LPPM Bina Sarana Informatika University for providing moral support to the author so that he can complete this research.

References

[1] D. A. Saputri and R. K. Dewi, "PENGARUH PENDAPATAN TERHADAP PEMBIAYAAN GADAI (RAHN) PADA PEGADAIAN SYARIAH WAY HALIM 2016-2018," *J. BISNIS Akunt. UNSURYA*, 2020, doi: 10.35968/jbau.v5i2.433.
[2] P. Rahmawati, A. Larasati, and M. Marsono, "PENGEMBANGAN MODEL PERSETUJUAN KREDIT NASABAH BANK DENGAN ALGORITMA KLASIFIKASI NAÏVE BAYES, DECISION TREE, DAN ARTIFICIAL NEURAL NETWORK," *J@ti Undip J. Tek. Ind.*, 2022, doi: 10.14710/jati.1.1.1-12.
[3] H. Huzain, A. D. Saputri, S. R. Rasyid, and M. M. Tsani, "Studi Kelayakan Bisnis Aspek finansial," *Uin Alauddin Makassar*, 2021.
[4] S. Mutmainah and R. P. Putra, "PROSEDUR PENDAFTARAN PINJAMAN KREDIT USAHA RAKYAT (KUR) PADA PT. BANK RAKYAT INDONSIA (Persero) Tbk. UNT TANJUNGSARI CABANG PAMANUKAN," *World Financ. Adm. J.*, 2022, doi: 10.37950/wfaj.v4i1.1341.
[5] Y. Zhou, "Loan Default Prediction Based on Machine Learning Methods," in *Proceedings of the 3rd International Conference on Big Data Economy and Information Management, BDEIM 2022, December 2-3, 2022, Zhengzhou, China*, EAI, 2023. doi: 10.4108/eai.2-12-2022.2328740.
[6] Mia Muchia Desda and Mai Yuliza, "Pengaruh Risiko Kredit Terhadap Likuiditas Melalui Perputaran Piutang Pada PT. BPR Swadaya Anak Nagari," *WACANA Ekon. (Jurnal Ekon. Bisnis dan Akuntansi)*, 2021, doi: 10.22225/we.20.2.2021.161-169.
[7] E. Octavia, "ANALISIS PROSES PEMBERIAN KREDIT UNTUK MENGURANGI RESIKO KREDIT MACET DI PT. BANK ARTHA GRAHA INTERNASIONAL TBK BANDUNG.," *J. Akunt. Bisnis dan Ekon.*, 2021, doi: 10.33197/jabe.vol6.iss2.2020.622.
[8] M. Murtala, "ANALISIS RISIKO KREDIT USAHA PADA NASABAH PT. BANK BRI (STUDI KASUS BRI UNIT T. NYAK ARIEF)," *J. Ekon. dan Pembang.*, 2020, doi: 10.22373/jep.v11i1.72.
[9] P. Ziemba, J. Becker, A. Becker, A. Radomska-Zalas, M. Pawluk, and D. Wierzba, "Credit decision support based on real set of cash loans using integrated machine learning algorithms," *Electron.*, 2021, doi: 10.3390/electronics10172099.
[10] P. Bhargav and K. Sashirekha, "A Machine Learning Method for Predicting Loan Approval by Comparing the Random Forest and Decision Tree Algorithms .," vol. 10, pp. 1803–1813, 2023, doi: https://doi.org/10.17762/sfs.v10i1S.414.
[11] P. Bhargav and K. Malathi, "Using Machine Learning , the Random Forest Algorithm and Logistic Regression to Predict Default Loan Approval .," vol. 10, pp. 1814–1824, 2023, doi: https://doi.org/10.17762/sfs.v10i1S.415.
[12] P. Bhargav and P. Rama Parvathy, "Comparing Random Forest with the Naive Bayes Algorithm with Improved Accuracy: An Effective Machine Learning Method for Loan Prediction," *J. Surv. Fish. Sci.*, vol. 10, no. 1S, pp. 2018–2029, 2023, [Online]. Available:

http://sifisheriessciences.com/journal/index.php/journal/article/view/436

[13] K. Sravani and R. Mahaveerakannan, “Using Random Forest as a Novel Approach to Loan Prediction and Comparing Accuracy to the Support Vector Machine Algorithm,” vol. 10, pp. 1174–1181, 2023.

[14] F. Chollet, “Loan Approval Prediction Using Machine Learning,” *Mach. Learn.*, vol. 45, no. 13, pp. 40–48, 2017, [Online]. Available: https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA

[15] Y. Dasari, K. Rishitha, and O. Gandhi, “Prediction of Bank Loan Status Using Machine Learning Algorithms,” *Int. J. Comput. Digit. Syst.*, vol. 14, no. 1, pp. 139–146, 2023, doi: 10.12785/ijcds/140113.

[16] N. M. Aji, V. Atina, and N. A. Sudibyo, “View of PEMODELAN PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES DI UNIBA.pdf.” pp. 148–158, 2023.

[17] M. Ula, A. F. Ulva, and Mauliza, “View of IMPLEMENTASI MACHINE LEARNING DENGAN MODEL CASE BASED REASONING DALAM MENDAGNOSA GIZI BURUK PADA ANAK.pdf.” pp. 333–339, 2021.

[18] D. Cahya Putri Buani, “Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung,” *EVOLUSI J. Sains dan Manaj.*, vol. 9, no. 2, pp. 43–48, 2021, doi: 10.31294/evolusi.v9i2.11141.

[19] A. N. Syahrudin and T. Kurniawan, “Input dan Output pada Bahasa Pemrograman Python,” *J. Dasar Pemrograman Python STMIK*, no. June 2018, pp. 1–7, 2018.

[20] A. C. Nugraha and M. I. Irawan, “Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost),” *J. Sains dan Seni ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.107032.

[21] F. I. Kurniadi and P. D. Larasati, “Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 67–72, 2022, doi: 10.47970/siskom-kb.v6i1.328.

[22] E. Ismanto and M. Novalia, “dan Gradient Boosting untuk Klasifikasi Komoditas Performance Comparison Between C4.5 Algorithm, Random Forests, and Gradient Boosting for Commodity Classification,” *Agustus*, vol. 20, no. 3, pp. 400–410, 2021.

[23] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, “Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, 2021, doi: 10.29207/resti.v5i1.2813.

[24] Qadrini L, Sepperwali A, and Aina A, “Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial,” *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1966, 2021.