JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 6, No. 1. December 2023 DOI: https://doi.org/10.34288/jri.v6i1.XXX
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

COMPARISON OF DECISION TREE, NAIVE BAYES AND RANDOM FOREST ALGORITHM TO GET THE BEST PERFORMANCE OF ALGORITHM FOR CUSTOMER CREDIT CLASSIFICATION

Indah Suryani[-1*]

Informatika, Sistem Informasi
Universitas Nusa Mandiri
indah.ihy@nusamandiri.ac.id[-1*],
(*) Corresponding author

Abstract

Credit is a potential income and the most significant business operation risk for a bank. Bad credit has become an ingrained problem in the banking world. Therefore, this research aims to classify customer data profiles who have the opportunity to be able to apply for a loan or not to reduce the risk of bad credit in the future by classifying using three commonly used data mining algorithms, namely the Decision Tree algorithm, Naïve Bayes and Random forest. The research was conducted using an experimental, descriptive method by testing the accuracy of the three methods to get the best performance. Based on the experiments' results, the accuracy performance with the confusion matrix was 73.20% for the Decision Tree algorithm. The accuracy for the Naive Bayes algorithm was 74.4%, and the Random Forest was 77.4%. Meanwhile, performance evaluation is based on the Receiver Operating Characteristics (ROC) curve by looking at the resulting Area Under Curve (AUC) value of 0.717 for the Decision Tree algorithm. At the same time, Naive Bayes produces an AUC value of 0.741, and the largest is Random Forest at 0.796. So it can be concluded that the classification that performed the best was the one that used the Random Forest algorithm. Then, from the validation results using the T-Test of the three methods being compared, the Random Forest produces a significant difference in accuracy compared to the accuracy produced by the Decision Tree, with an alpha value of 0.031.

Keywords: Credit Customer, Classification, Random Forest

Abstrak

Kredit merupakan potensi penghasilan sekaligus resiko operasi bisnis terbesar bagi sebuah bank. Kredit macet telah menjadi permasalahan yang mendarah daging bagi dunia perbankan. Maka dari itu, penelitian ini bertujuan untuk melakukan klasifikasi profil data nasabah yang memiliki peluang untuk dapat mengajukan kredit pinjaman atau tidak, demi mengurangi resiko kredit macet di kemudian hari dengan melakukan klasifikasi menggunakan tiga algoritma data mining yang umum digunakan yaitu algoritma Decision Tree, Naïve Bayes dan Random forest. Penelitian dilakukan menggunakan metode deskriptif eksperimental dengan menguji akurasi dari ketiga metode tersebut untuk mendapatkan performa terbaik dari ketiganya. Berdasarkan hasil eksperimen yang dilakukan, maka didapatkan performa akurasi dengan confusion matrix sebesar 73.20% untuk algoritma Decision Tree, kemudian akurasi untuk algoritma Naive Bayes sebesar 74.4% dan Random Forest sebesar 77.4%. Sedangkan evaluasi performa berdasarkan kurva Receiver Operating Characteristics (ROC) dengan melihat nilai Area Under Curve (AUC) yang dihasilkan adalah sebesar 0.717 bagi algorima Decision Tree, sedangkan Naive Bayes menghasilkan nilai AUC sebesar 0.741 dan terbesar adalah Random Forest sebesar 0.796. Maka dapat disimpulkan bawa performa terbaik dari klasifikasi yang dilakukan adalah yang menggunakan algoritma Random Forest. Kemudian dari hasil validasi menggunakan t-Test dari ketiga metode yang dibandingkan, maka Random Forest menghasilkan perbedaan tingkat akurasi yang signifikan terhadap akurasi yang dihasilkan oleh Decision Tree yaitu dengan nilai alpha sebesar 0.031.

Kata kunci: Kredit Pelanggan; Klasifikasi; Random Forest

INTRODUCTION

According to the Law of the Republic of Indonesia Article 1 Number 10 of 1998, credit is a bill or provision of money that can be expressed as the same, between a bank and another party, by requiring the other party to pay the bill within a predetermined time period. Credit customers are people who use banking services or other financial services (Susilo, 2023). Normally, most of the bank's wealth is obtained from providing credit loans so that a marketing bank must be able to reduce the risk of non-performing credit loans (Religia, Pranoto, & Santosa, 2020).Credit can be the biggest advantage and risk in a banking business. Problematic or bad credit often occurs due to a lack of mature analysis in the credit granting process (Nurjanah, Karaman, Widaningrum, Mustikasari, & Sucipto, 2023).

Therefore, it is necessary to select appropriate customer credit to reduce the risk of loss from providing credit. To support targeted credit offers, data analysis and mathematical calculations are needed so that it will be very efficient if the data used has been processed using an algorithm with fairly accurate output values (Panggabean, 2022). Accurate classification of credit customers is the premise of providing personalized credit services to them (Li, Luo, Tang, & Xie, 2023).

Based on these problems, there are several previous studies which also examined customer credit classification using data from within and outside the country in several financial institutions such as banking, cooperatives and leasing, including research conducted by (Rahmawati, Larasati, & Marsono, 2022),(Yogiek Indra Kurniawan, 2020), (Yusuf & Sestri, 2020),(Darmawan, 2020), (Religia et al., 2020),(Widjiyati, 2021), (Nurdina Rasjid, Nurhikmah Arifin, & Nilam Cahya, 2021) and (Ningsih, Budiman, & Umami, 2022).

Application of machine learning methods to large databases is called data mining (Ethem, 2015). Data mining plays an ever-growing role in both theoretical studies and applications (Vercellis, 2009).

A classification model is a supervised learning method for predicting the value of a target categorical attribute, compared to a regression model that deals with numerical attributes. Given a set of past observations whose target class is known, a classification model is used to generate a set of rules that can predict the target class of future examples (Vercellis, 2009). In classification models, there are several methods that are commonly used, including decision trees, naive Bayes and random forests.

The basic concept of a Decision Tree is to convert data into a decision tree with rules. The selected attributes will produce a partition with more uniform data and can produce a simple decision tree with little repetition. A decision tree consists of a set of rules that aim to divide a number of heterogeneous populations into smaller and more homogeneous ones taking into account the objective variables (Yusuf & Sestri, 2020). Several studies carried out using this decision tree algorithm were carried out by (Bahri & Lubis, 2020).

The Naive Bayes method is a subset of simple probability based on the application of Bayes' theorem with the assumption of strong independence between features (Jasmir, Sika, Mulyadi, & Amelia, 2022). The Naıve Bayes model is tremendously appealing because of its simplicity, elegance, robustness, as well as the speed with which such a model can be constructed, and the speed with which it can be applied to produce a classification.

It is one of the oldest formal classification algorithms, and yet even in its simplest form it is often surprisingly effective (Wu et al., 2008). Due to these advantages, several studies using the Naive Bayes method are research by (Heliyanti Susana, 2022), (Fikrillah, Hudawiguna, & Juliane, 2023), (Putro, Vulandari, & Saptomo, 2020)and (Juwita, Safii, & Efendi Damanik, 2022).

Random Forest is a method that can increase accuracy results in generating attributes for each node which is done randomly (Suci Amaliah, Nusrang, & Aswi, 2022). Likewise, research conducted by (Putri & Wijayanto, 2022), (Buani & Suryani, 2022) in comparing several machine learning algorithms in carrying out classification, the random forest algorithm has produced the best performance.

Based on the literature that has been described, so in this research, customer credit data was classified using the three algorithms mentioned above, namely Decision Tree, Naive Bayes and Random Forest, to then look for the best performance results from these three methods so that it is hoped that they will be able to reduce the risk of bad credit. in the future, which in general will be able to reduce the risk of losses experienced by banks.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v6i1.XXX
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

RESEARCH METHODS

Types of research

This research is quantitative research in the form of experimental research. Experiments were carried out using a commonly used data processing application, namely the rapidminer application version 9.8.

Research Target / Subject

The dataset used in this research consists of 1000 customer credit data records which are divided into two, namely training and testing data. Then 80% of the data is used as training data and 20% of the data is used as testing data.

Data, Instruments, and Data Collection Techniques

This research uses secondary data sourced from the site www.kaggle.com in the form of customer credit data of 1000 records. As stated in Table 1 below. The data used in this research is customer profile data which consists of 20 attributes, which are then processed to obtain a classification of customer profile data which is considered good or bad and then used as consideration in granting customer credit.

Table 1. The Data Attributes of Credit Customer's

No.	Attributes
1	Checking Status
2	Duration
3	Credit_history
4	Purpose
5	Credit Amount
6	Savings Status
7	Employment
8	Installment Commitment
9	Personal Status
10	Other Parties
11	Residence Since
12	Property Magnitude
13	Age
14	Other Payment Plans
15	Housing
16	Existing Credit
17	Job
18	Num Dependents
19	Own Telephone
20	Foreign Worker

Procedure

The modeling carried out in this research was by carrying out the following steps:

1. The first is to prepare a customer credit dataset.
2. So that the dataset is ready for use, the first data pre-processing step used is to replace missing or inappropriate values. Missing values can be replaced by the minimum, maximum or average value of that Attribute. Zero can also be used to replace missing values. Any replenishment value can also be specified as a replacement of missing values. This process is carried out using the replace missing value operator. Then the label and ID attributes are determined using the role set. Once the data is ready to be used,
3. The next stage is to carry out modeling of the three selected algorithms, namely Decision Tree, Naive Bayes and Random Forest. Decision Tree was chosen for the reason that the application of decision tree algorithms in educational data mining has emerged as a powerful (Chen & Lin, 2023). Naive Bayes was chosen for the reason that It has strong model representation, learning ability, and inference ability, while showing high efficiency and high accuracy in learning small data sets stated by (Li et al., 2023). Naive Bayes classifiers are also highly scalable, requiring a number of linear parameters over a variable number (features/predictors) in the learning problem (Jasmir et al., 2022). Meanwhile Random Forest was chosen on the grounds that Random Forest has advantages that the research results shows in study by (Sriyanto and Supriyatna, 2023). that the random forest algorithm can predict diabetes with good performance. The performance evaluation value of the random forest algorithm for predicting diabetes are: accuracy of 99.3% Random Forest produced the best accuracy among the three algorithms tested in research conducted by (Supriyadi *et al.*, 2020). Random Forest with accuracy results of 0.7468 makes this algorithm best used to classify the quality of red wine. In line with research conducted by (Kurniawan *et al.*, 2023) also found random forest can produce very good accuracy. In Rapidminer 9.8, all three modeling can be done simultaneously using the multiply operator with each divides training and testing data using the cross validation operator. Then by adding the performance operator you will get the performance results of the three models.
4. Next, an evaluation is carried out by measuring the accuracy of the confusion matrix and also the AUC value resulting from the ROC curve for each algorithm.

5. The last step is to carry out validation using a t-test to measure the level of significance between the differences in accuracy values of the three algorithms.

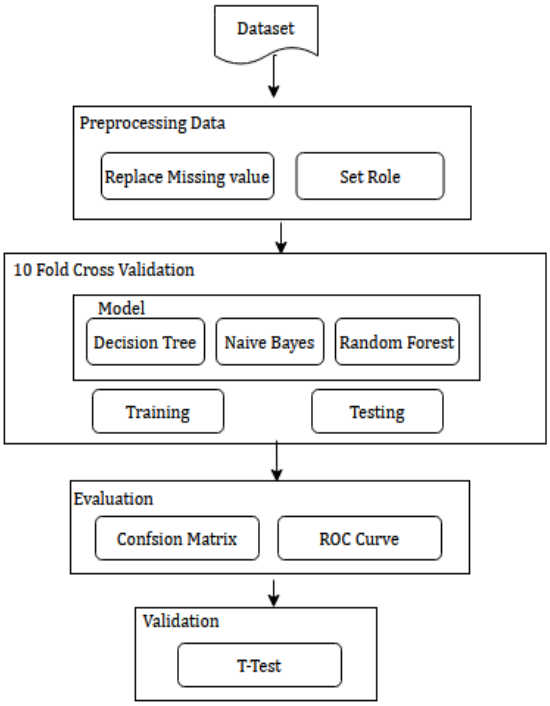

Figure 1. Proposed Method

Figure 1 above shows the flow of steps in the research carried out.

RESULTS AND DISCUSSION

The research stages carried out in classifying customer credit data are as follows:

1. Prepare a dataset for research
2. Carry out data pre-processing
3. Design data modeling using 3 algorithms, namely Decision Tree, Naive Bayes and Random Forest. And then carrying out training and testing on the three models. Training and testing is carried out using the 10 fold cross validation method, which automatically divides the training data into 80% of the total data and the testing data into 20% of the total data.
4. Evaluate the performance produced by each model using a confusion matrix and ROC curve. From evaluation with the confusion matrix, accuracy, precision and recall values will be obtained. Meanwhile, evaluation with the ROC curve is by looking at the resulting value in the Area Under the Curve (AUC).

In the first experiment, classification was carried out using a Decision Tree algorithm which obtained results as shown in Table 2. as follows.

Table 2. Decision Tree Performance

	True Good	True Bad
Pred Good	577	145
Pred Bad	123	155

Based on Table 2, it can be seen that the accuracy produced by the Decision Tree is 73.2% which is obtained using the following formula:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$
$$Accuracy = 73.2\%$$

The resulting precision value is 79.9% which is obtained using the following calculation formula.

$$Precision = \frac{\text{TP}}{\text{TP + FP}} \times 100\%$$
$$Precision = 79.9\%$$

And the recall value obtained is 82.4% from the following calculation:

$$Recall = \frac{\text{TP}}{\text{TP + FN}} \times 100\%$$
$$Recall = 82.4\%$$

Figure 2 shows the AUC resulting from the decision tree experiment.The value on the red line shows the TPR (True Positive Rate) value while the blue line shows the FPR (False Positive Rate) value. From the area where the two intersect, if you add up the areas, you can get the Area Under Curve which is 0.717 as shown in Figure 2 below.

Figure 2. Decision Tree AUC

Table 3 shows the performance produced by Naive Bayes.

Table 3. Naive Bayes Performance

	True Good	True Bad
Pred Good	582	138
Pred Bad	118	162

Based on the values in Table 3 above, the accuracy, precision, and recall values are obtained using the following formula.
Accuracy = 74.4%
Precission = 80.8%
Recall = 83.1%

As seen in Figure 3 below, it also shows the ROC curve resulting from experiments using Naive Bayes with the resulting AUC area of 0.741.

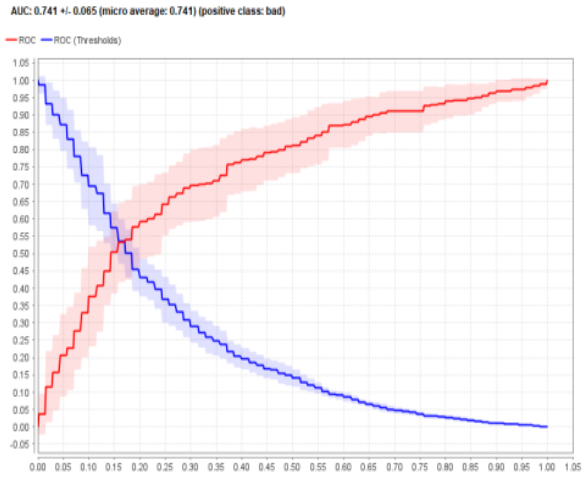

Figure 3. Naive Bayes AUC

The accuracy results from the third model with Random Forest produce accuracy, precision, and recall values, as seen in Figure 3 below. This accuracy describes how accurate the model is in classifying correctly. Meanwhile, precision describes the accuracy between the requested data and the prediction results provided by the model. Recall or sensitivity describes the success of the model in recovering information.

Table 4. Random Forest Performance

	True Good	True Bad
Pred Good	673	199
Pred Bad	27	101

Furthermore, the accuracy, precision, and recall values resulting from the values in Table 4 are as follows.
$Accuracy = 77.4\%$
$Precision = 96.1\%$
$Recall = 77.1\%$

The AUC value resulting from the ROC curve of the random forest algorithm is 0.796, as seen in Figure 4.

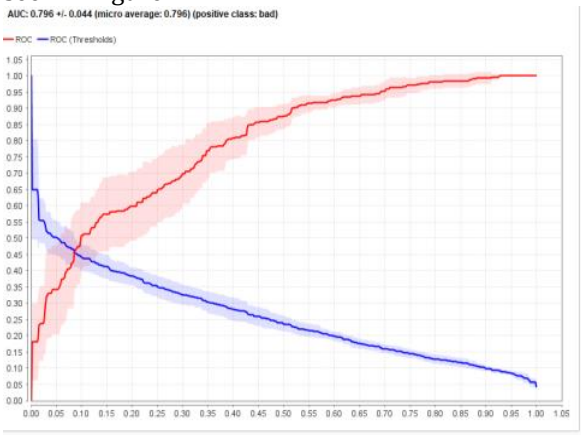

Figure 4. Random Forest AUC

A summary of the results of the three models can be obtained based on the test results shown in Table 5 below.

Table 5. Model Testing Result

	DT	NB	RF
Accuracy	73.2%	74.4%	77.4%
Precision	79.9%	80.8%	96.1%
Recall	82.4%	83.1%	77.1%

In Table 5, it can be seen that the results of testing the Random Forest algorithm produce superior accuracy and precision. In contrast, the Naive Bayes model produces the highest value for recall.

5. The final step is to test the validity of the three models. Based on the literature review of previous research, in research on comparative analysis of the accuracy of machine learning methods for classification, in general, there are still many who have not yet reached the stage of validating the accuracy results of the several models being compared and have only reached the evaluation of performance using the confusion matrix. In research conducted by (Susilo, 2023) under the title "Performance Comparison of K-Nearest Neighbors and Naive Bayes For Classification of Customer Behavior in Bank Credit Payments," Then other research by (Yasir & Suraji, 2023) with the title "Comparison Of Naïve Bayes, Decision Tree, Random Forest Classification Methods On Sentiment Analysis Of Increasing The Cost Of Hajj 2023 On Youtube Social Media" and research by (Putri & Wijayanto, 2022) with the title "Comparative Analysis of Data Mining

Classification Algorithms in Phishing Website Classification" and research "Research on Credit Customer Management Based on Customer Classification and Classification Preference" by (Li et al., 2023). Therefore, at the final stage of this research, a validity test was carried out to evaluate the accuracy performance of the three models studied. The validation method chosen in this research is the t-test method. The t-test is carried out by alternately comparing the accuracy values between two models to test the difference between the two. The results, as shown in Table 6, were obtained from the tests carried out.

Table 6. t-Test Result

Compared Algorithm	Alpha
Decision Tree Vs. Naive Bayes	0.453
Naive Bayes Vs. Random Forest	0.128
Random Forest Vs. Decision Tree	0.031

Based on the alpha value produced in Table 6, it can be seen that the difference between Naive Bayes and Decision Tree is 0.453. Then, the comparison of Random Forest to Naive Bayes is 0.128. The alpha value from comparing the Random Forest to the Decision Tree is 0.031. So, it can be concluded that there is a significant difference in the accuracy comparison from comparing the Random Forest method to the decision tree, which produces an alpha value under 0.05.

CONCLUSIONS AND SUGGESTIONS

Conclusion

Based on the experiments, this research uses three machine learning methods, Decision Tree, Naive Bayes, and Random Forest, to classify customer credit. Of the three algorithms, it is known that the Random Forest method shows the highest accuracy and AUC performance with values of 77.4% and 0.796. Meanwhile, the accuracy and AUC of the Decision Tree are 73.2% and 0.717. Then, the accuracy and AUC of Naive Bayes are 74.4% and 0.741. Then, after validation was carried out using the t-test, an alpha value of 0.031 was obtained by comparing the Random Forest performance with Decision Trees. This means there is a significant difference in the performance of random forests versus decision trees. So, it can be said that in this research, Random Forest produced the most superior performance in classifying customer credit data.

Suggestion

The results of the research were that the random forest algorithm produced the best accuracy performance of the three algorithms tested. In other words, in this research, the Random Forest algorithm was superior to Decision Tree and Naive Bayes. However, the resulting accuracy value is still below 80%. So, in future research, we can try to optimize or select features so that it is hoped that we can further improve the accuracy performance of the Random Forest.

REFERENCES

Bahri, S., & Lubis, A. (2020). Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League. *Jurnal Sintaksis*, *2*(04), 63–70. Retrieved from https://www.ojs.yayasanalmaksum.ac.id/index.php/Sintaksis/article/view/47

Buani, D. C. P., & Suryani, I. (2022). Implementation of Machine Learning Algorithms for Early Detection of Cervical Cancer Based on Behavioral Determinants. *Jurnal Riset Informatika*, *5*(1), 445–450. https://doi.org/10.34288/jri.v5i1.453

Chen, S., & Lin, X. (2023). *Application of Decision Tree Algorithm in Educational Data Mining*. *6*, 120–127. https://doi.org/10.23977/curtm.2023.060818

Darmawan, T. (2020). Credit Classification Using CRISP-DM Method On Bank ABC Customers. *International Journal of Emerging Trends in Engineering Research*, *8*(6), 2375–2380. https://doi.org/10.30534/ijeter/2020/288620

Ethem, A. (2015). Introduction to Machine Learning Second Edition Adaptive Computation and Machine Learning. In *Massachusetts Institute of Technology*. Retrieved from https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf

Fikrillah, H. N. F., Hudawiguna, S., & Juliane, C. (2023). Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *10*(1), 683–695. Retrieved from https://jurnal.mdp.ac.id/index.php/jatisi/article/view/3624

Heliyanti Susana. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *4*(1), 1–8.

https://doi.org/10.52005/jursistekni.v4i1.96

Jasmir, J., Sika, X., Mulyadi, M., & Amelia, R. (2022). Klasifikasi Kelayakan Pemberian Kredit Pada Calon Debitur Menggunakan Naïve Bayes. *JURIKOM (Jurnal Riset Komputer)*, *9*(6), 1833. https://doi.org/10.30865/jurikom.v9i6.5131

Juwita, Safii, M., & Efendi Damanik, B. (2022). Algoritma Naïve Bayes Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar. *Journal of Machine Learning and Artificial Intelligence*, *1*(4), 337–346. https://doi.org/10.55123/jomlai.v1i4.1674

Li, C., Luo, X., Tang, S., & Xie, M. (2023). Research on Credit Customer Management Based on Customer Classification and Classification Preference. *Journal of Global Humanities and Social Sciences*, *4*(6), 282–287. https://doi.org/10.61360/bonighss232015310604

Ningsih, W., Budiman, B., & Umami, I. (2022). Implementasi Algoritma Naïve Bayes Untuk Menentukan Calon Penerima Beasiswa Di SMK YPM 14 Sumobito Jombang. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *4*(2), 446–454. https://doi.org/10.47233/jteksis.v4i2.570

Nurdina Rasjid, Nurhikmah Arifin, & Nilam Cahya. (2021). Klasifikasi Nasabah Bank Layak Kredit Menggunakan Metode Naive Bayes. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, *1*(1), 01–10. https://doi.org/10.55606/juisik.v2i2.187

Nurjanah, I., Karaman, J., Widaningrum, I., Mustikasari, D., & Sucipto, S. (2023). Penggunaan Algoritma Naïve Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa. *Explorer*, *3*(2), 77–87. https://doi.org/10.47065/explorer.v3i2.766

Panggabean, I. M. (2022). *Analisis Prediksi Kelayakan Nasabah Kredit Menggunakan Algoritma Random Forest Menggunakan PEGA dan WEKA*. *5*, 78–90.

Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, *11*(1), 59–66. https://doi.org/10.34010/komputika.v11i1.4350

Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *8*(2). https://doi.org/10.30646/tikomsin.v8i2.500

Rahmawati, P., Larasati, A., & Marsono, M. (2022). Pengembangan Model Persetujuan Kredit Nasabah Bank Dengan Algoritma Klasifikasi Naïve Bayes, Decision Tree, Dan Artificial Neural Network. *J@ti Undip: Jurnal Teknik Industri*, *17*(1), 1–12. https://doi.org/10.14710/jati.1.1.1-12

Religia, Y., Pranoto, G. T., & Santosa, E. D. (2020). South German Credit Data Classification Using Random Forest Algorithm to Predict Bank Credit Receipts. *JISA(Jurnal Informatika Dan Sains)*, *3*(2), 62–66. https://doi.org/10.31326/jisa.v3i2.837

Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(3), 121–127. https://doi.org/10.35580/variansiunm31

Susilo, A. (2023). Perbandingan Kinerja K-Nearest Neighbors dan Naive Bayes Untuk Klasifikasi Perilaku Nasabah Pada Pembayaran Kredit Bank. *Jurnal Sains Dan Teknologi (JSIT)*, *3*(3), 364–379. https://doi.org/10.47233/jsit.v3i3.1264

Vercellis, C. (2009). Business Intelligence: Data Mining and Optimization for Decision Making. In *Business Intelligence: Data Mining and Optimization for Decision Making*. https://doi.org/10.1002/9780470753866

Widjiyati, N. (2021). *Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval Implementation of Random Forest Algorithm in The Classification of Credit Approval Dataset*. *1*(1), 1–7. https://doi.org/10.25008/janitra.v1i1.118

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., … Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, Vol. 14, pp. 1–37. https://doi.org/10.1007/s10115-007-0114-2

Yasir, M., & Suraji, R. (2023). Perbandingan Metode Klasifikasi Naive Bayes, Decision, Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 pada Media Sosial Youtube. *Jurnal Cahaya Mandalika (JCM)*, *3*(2), 180–192. https://doi.org/10.36312/jcm.v3i2.1520

Yogiek Indra Kurniawan, T. I. B. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan. *Jurnal Ilmiah Matrik Universitas Bina Darma*, *22*(1), 73–82. https://doi.org/https://doi.org/10.33557/jurnalmatrik.v22i1.843

Yusuf, D., & Sestri, E. (2020). Metode Decision Tree Dalam Klasifikasi Kredit Pada Nasabah PT Bank Perkreditan Rakyat (Studi Kasus : PT BPR Lubuk Raya Mandiri). *Jurnal Sistem Informasi (JUSIN)*, *1*(1), 21–28. https://doi.org/10.32546/jusin.v1i1.855