Accredited SINTA 5
72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

Comparative Study of Naive Bayes Data Mining, Decision Tree C4.5 And Fuzzy Decision Tree (ID3) In Analyzing the Financing of Fitness Center Memberships: Case Study PT Fitindo Sehat Sempurna

Wahyu Joko Saputro[1*], Supriadi Panggabean[2], Faishal Wafiq Zakiy[3]
[1,2,3]Universitas Darunnajah
[1-3]Jl. Ciledug Raya No.01 RT 001 RW003 Kelurahan Ulujami Kecamatan Pesanggrahan
Kota Administrasi Jakarta Selatan Daerah Khusus Jakarta - 12250
E-mail: wahyujs@darunnajah.ac.id[1*]

Submitted Date: 07 Oktober 2024 ***Accepted Date: 18 Oktober 2024***

Abstract - *The business world is always experiencing rapid changes, so it requires companies to be able to respond to these changes quickly and precisely. PT. Fitindo Sehat Sempurna as a company engaged in fitness services wants to do data mining to predict the smooth financing of its members and strive to increase sustainable sales turnover. One of the classification algorithms that is often used and gets a lot of attention from researchers in predicting problematic financing is Naive Bayes, Decision Tree C4.5 and then tries to compare it with the Fuzzy Decision Tree (ID3). Of all the trials of the three methods, the best results were obtained, namely Naive Bayes with the highest calculation result of accuracy of 85,00% and AUC = 0,960 as the best results. This test uses 80 training data and 20 testing data.*
Keywords: Financing, Membership, ID3, Fuzzy, Decision Tree, C4.5, Naive Bayes.

Abstrak - Dunia bisnis selalu mengalami perubahan yang cepat, sehingga menuntut perusahaan untuk dapat merespon perubahan tersebut dengan cepat dan tepat. PT. Fitindo Sehat Sempurna sebagai perusahaan yang bergerak di bidang jasa kebugaran ingin melakukan data mining untuk memprediksi kelancaran pembiayaan para anggotanya dan berupaya untuk meningkatkan omset penjualan yang berkelanjutan. Salah satu algoritma klasifikasi yang sering digunakan dan mendapat banyak perhatian dari para peneliti dalam memprediksi pembiayaan bermasalah adalah Naive Bayes, Decision Tree C4.5 dan kemudian mencoba membandingkannya dengan Fuzzy Decision Tree (ID3). Dari semua uji coba ketiga metode tersebut, didapatkan hasil terbaik yaitu Naive Bayes dengan hasil perhitungan akurasi tertinggi yaitu 85,00% dan AUC = 0,960 sebagai hasil terbaik. Pengujian ini menggunakan 80 data training dan 20 data testing.
Kata kunci: Pembiayaan, Keanggotaan, ID3, Fuzzy, Decision Tree, C4.5, Naive Bayes

1. Introduction

The business world continues to change rapidly, so companies must quickly and appropriately respond to changes. In the era of globalization, the industry demands sophisticated systems to advance, including in fitness services. Fitness centers are growing because they attract different ages who want to exercise comfortably and flexibly time. PT. Fititndo Sehat Sempurna, a fitness service provider, faced problems in membership assessment because the data only focused on one prospective member without considering previous financing experience.

From the above problems can be taken alternatives by utilizing techniques data mining by comparing three algorithms for the identification of member growth and development. Algorithms are measures that are always used in everyday life [1]. Algorithms also serve as supporting tools that ensure the smooth operation of a system. In a variety of contexts, algorithms provide structured instructions and rules, allowing systems to work efficiently and effectively [2]. In this study, the algorithm to be used is the method Naive Bayes, Decision Tree C4.5 and Fuzzy Decision Tree (ID3). With hope after being processed with data mining can help predict the likelihood of stalled member financing.

Naïve Bayes is a classification technique that uses probability and statistical methods to make predictions [3]. This technique is based on Bayes' Theorem, which calculates the probability of an event based on prior knowledge of related conditions [4]. In the context of classification, Naïve Bayes assumes that each feature or attribute of the data used is independent of each other, although in reality this assumption may not always be true [5].

Decision Tree is one of the most powerful and efficient techniques in data mining that it has been widely used by researchers [6]. A decision tree represents a tree-like structure for classified data. Decision trees produce rules for classification [7].

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

Fuzzy decision trees utilize fuzzy set theory to describe the level of connectedness of an attribute, while for the process of forming decision trees using the decision tree induction process [8].

Research on Models Data Mining In determining the feasibility of choosing a place of residence using the method Naïve Bayes, the criteria used include clean water sanitation, location security, prices and house facilities for residence, garbage disposal, transportation, house models, and of course must be flood-free [9]. Method Naive Bayes for determination in the selection of residence, it is used to facilitate consumers in choosing a place to live in accordance with consumer wishes, which can result in a habitable or inappropriate residence [10].

In a study entitled Hypertension Risk Classification Using Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3), Given that hypertension is one of the dangerous diseases, researchers have conducted research related to the classification of hypertension based on existing factors. One of the methods that is often used for classification is Fuzzy Decision Tree [11].

This study will compare the Naive Bayes method with Decission Tree (C4.5), the results of the analysis of the two methods will be compared again with the Fuzzy method to see the ease of use and accuracy of the calculation results of each algorithm to analyze the smooth financing of fitness studio fitness centersat PT Fitindo Sehat Sempurna hopefully this research can help the effectiveness, as well as better system performance in the future.

The development and success of a fitness studio is very dependent on the sales turnover or users of the fitness center. PT. Fitindo Sehat Sempurna wants to maintain customer loyalty and conduct a smooth financing analysis. In this study, Data Mining will be compared using Naive Bayes, Decission Tree C4.5 and Fuzzy Decission Tree (ID3) to predict the smooth financing of fitness center members. From the results of the analysis of the three methods, it will be seen, which method produces the best value.

Based on the identification and scope of the research problem to be carried out, the problems in this study can be summarized in the problem formulation, namely: How to get the best results in predicting the smooth financing of fitness center members with data mining using the Naïve Bayes method, Decision Tree C4.5 and Fuzzy Decision Tree (ID3)? and How to design Data mining so that it can be used to determine the right business strategy?

2. Literature Review

Research conducted by Sinta Amanda Pratiwi et al. related Prediction of Drug Inventory at Pharmacies Using the Decision Tree Algorithm obtained an accuracy of 98.28%, Precision 0.9832, Recall 0.9828, and F1-Score 0.9804. with these results, this research can be continued by implementing drug inventory predictions using Decision Tree into an application [12].

Research conducted by Amelia Ramadhani et al. showed an accuracy of 98.50%, with precision for Drop Out and No Drop Out respectively 93.18% and 100.00%. Recalls of Drop Out and No Drop Out data are 64.44% and 100.00% [13].

Research conducted by Azwanti related to Motorcycle Sales Prediction at PT. Capella Dynamics Nusantara Yellow Face Branch Using Algorithm C4.5, The test results are very good because Rule the resulting is almost the same [14].

Research conducted by Prayoga et al related to Liver Disease Diagnosis Using Method Naïve Bayes, Accuracy testing of 40 test data obtained an accuracy rate of 87.5%. [15].

3. Result and Discussion

The results of research and testing to obtain the model and the results of calculating and comparing the three proposed algorithms, followed by testing the data into the model to get better results.

3.1 Data Training

Data collection techniques are various methods or approaches used to collect information or data required for research or analysis [16].

The amount of data obtained from data collection is as many as 100 records, with attributes of Member Status, Gender, Age, Package, Income, and distance.

3.2 Data Testing

After preparing 20 data test data records, the next step is to perform calculations by multiplying the Current and Non-Current classes/labels.

3.3 Test Result with Naïve Bayes Algorithm

Member data testing is carried out using the RapidMiner 9.7.0 application, the Naive Bayes test flow is as follows;

Accredited SINTA 5
72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

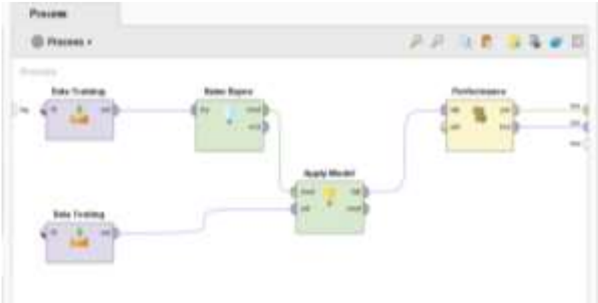

Figure 1. Naive Bayes Test Flow

First, processing begins with inputting training data and testing data that will be used in excel read operators and then connected to the Naive Bayes method in data testing to produce the best performance.

The results of the prediction of CURRENT and NON-CURRENT payments, the data used in the Naive Bayes method is 20 records with 6 attributes as shown in Figure 2.

Figure 2. Naive Bayes Accuracy Results

The test results using the Naïve Bayes method above obtained results with an accuracy rate of 85.00%, which means that the level of data accuracy is good.

Class Precission can be interpreted as a match between a request for information and an answer to the request, so the match between the request and the CURRENT prediction is 81.82% and the match with the CURRENT prediction is 88.89%.

Class Recall is defined as the ratio of the selected relevant items to the total number of relevant items available. So, from the data above, it can be concluded that the relevant level of true CURRENT is 90.00% and true NOT CURRENT is 80.00%. The processing AUC for the Naive Bayes Algorithm is 0.960, as seen in Figure 3.

Figure 3. AUC Naive Bayes Results

AUC (Area Under the Curve) is calculated to measure performance differences. The ROC curve shows accuracy and compares the classification visually with false positives as horizontal lines and true positives as vertical lines. From the data above, analysis using the RapidMiner 9.7.0 application with Naive Bayes measurements obtained AUC results of 0.960 which are included in the excellent category (Excellent Classification).

3.4 Test Results with Algorithm Decision Tree C4.5

Decision Tree C4.5 is one of the algorithms used to form decision trees, the decision tree method of converting facts into decision trees that represent a rule.

Accredited SINTA 5
72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

Data testing was performed using the RapidMiner 9.7.0 application following Figure 4 of the Decision Tree C4.5 test flow.

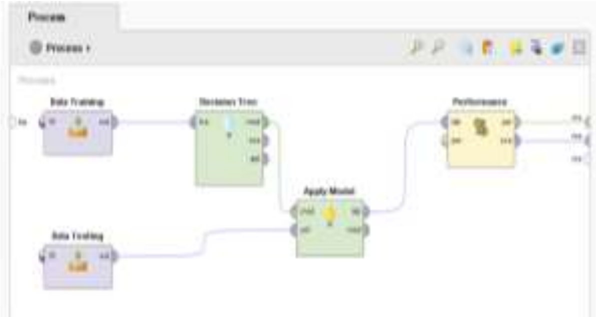

Figure 4. Decision Tree C4.5 Test Flow

First, processing begins with inputting training data and testing data that will be used in excel read operators and then connected to the Decision Tree C4.5 method in data testing to produce the best performance.

The results of the payment prediction are CURRENT and NOT CURRENT, the data used in the Decision Tree C4.5 method is 20 records with 6 attributes as shown in Figure 5.

Figure 5. Decision Tree Accuracy Results C4.5

The test results that have been carried out using the Decision Tree model obtained an accuracy of 80.00% as seen in Figure 4.6.

Class Precission can be interpreted as a match between a request for information and an answer to that request, so the match between the request and the CURRENT prediction is 75.00% and the match with the NOT CURRENT prediction is 87.50%.

Class Recall is defined as the ratio of the selected relevant items to the total number of relevant items available. So, from the data above, it can be concluded that the relevant level of true CURRENT is 90.00% and true NOT CURRENT is 70.00%.

Figure 6. Data Set Decision Tree C4.5

From the results of the analysis above, we can see that there are 11 attribute columns consisting of 6 ordinary attribute columns and 4 attribute special columns with the names Result, Prediction Result, Confidence CURRENT and Confidence NOT CURRENT, this is because these columns determine the results of member data analysis. From the data of 20 members, there are some data that are declared not to match the prediction of Decision Tree C4.5.

Accredited SINTA 5
72/E/KPT/2024
Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

Figure 7. Data Wrong Prediction Decision Tree C4.5

Figure 8. Formation of Decision Tree C4.5

The processing AUC for the C4.5 Decision Tree Algorithm is 0.915, as seen in Figure 9.

Figure 9. AUC Decision Tree C4.5 Results

AUC (Area Under the Curve) is calculated to measure performance differences. The ROC curve shows accuracy and compares the classification visually with false positives as horizontal lines and true positives as vertical lines. From the data above, analysis using the RapidMiner 9.7.0 application with Decision Tree C4.5 measurements obtained AUC results of 0.915 which are included in the excellent category (Excellent Classification).

3.5 Fuzzy Decision Tree (ID3)

This process requires training data taken from the classification data set using the attributes of Member status, Gender, Age, Package, Income and Distance.

Data testing was performed using the RapidMiner 9.7.0 application following Figure 10 of the Fuzzy Decision Tree (ID3) test flow.

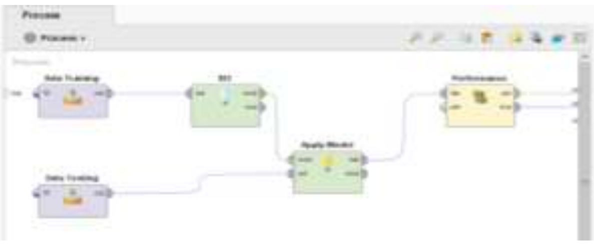

Figure 10. Fuzzy Decision Tree (ID3) Test Flow

Accredited SINTA 5 72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

First, processing begins with inputting training data and testing data that will be used in excel read operators then connected to the Fuzzy Decision Tree (ID3) method in data testing to produce the best performance.

The results of the payment prediction are CURRENT and NOT CURRENT, the data used in the Fuzzy Decision Tree (ID3) method is 20 records with 6 attributes as shown in Figure 11.

Figure 11. Fuzzy Decision Tree (ID3) Accuracy Results

The test results that have been carried out using the Decision Tree model obtained an accuracy of 75.00% as seen in Figure 11.

Class Precission can be interpreted as a match between a request for information and an answer to the request, so the match between the request and the CURRENT prediction is 77.78% and the match with the CURRENT prediction is 72.73%.

Class Recall is defined as the ratio of the selected relevant items to the total number of relevant items available. So, from the data above, it can be concluded that the relevant level of true CURRENT is 70.00% and true NOT CURRENT is 80.00%.

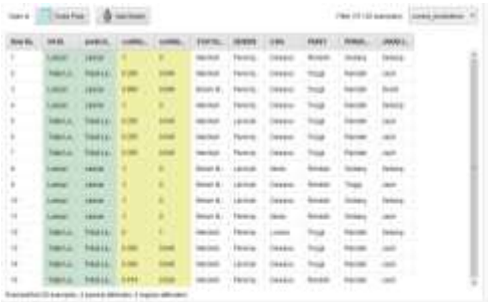

Figure 12. Data Set Fuzzy Decision Tree (ID3)

From the results of the analysis above, we can see that there are 11 attribute columns consisting of 6 ordinary attribute columns and 4 attribute special columns with the names Result, Prediction Result, Confidence CURRENT and Confidence NOT CURRENT, this is because these columns determine the results of member data analysis. From the data of 20 members, there are some data that are declared not to match the predictions of the Fuzzy Decision Tree (ID3).

Figure 13. Data Wrong Prediction Fuzzy Decision Tree (ID3)

Accredited SINTA 5
72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

Figure 14. Tree Formation Rules

The processing AUC for Fuzzy Decision Tree (ID3) is 0.620, as seen in Figure 15.

Figure 15. AUC Fuzzy Decision Tree (ID3) Results

AUC (Area Under the Curve) is calculated to measure performance differences. The ROC curve shows accuracy and compares the classification visually with false positives as horizontal lines and true positives as vertical lines. From the data above, analysis using the RapidMiner 9.7.0 application with Fuzzy Decision Tree (ID3) measurements obtained AUC results of 0.620 which are included in the low category.

3.6 Confusion Matrix Algoritma Naive Bayes

Confusion matrix is a method that uses true positives, false positives, false negatives and true negatives, to calculate the accuracy of data mining concepts as follows:

The test results that have been carried out using the Naive Bayes model, obtained an accuracy of 85.00% as seen in Figure 16 below.

Figure 16. Naive Bayes Accuracy Results

The number of CURRENT predictions whose reality is CURRENT is 9 and the number of CURRENT predictions whose reality is NOT CURRENT is 2, Next The number of CURRENT predictions whose reality is NOT CURRENT is 8 and the number of CURRENT predictions whose reality is CURRENT is 1.

The test results using the Naïve Bayes method above obtained results with an accuracy rate of 85.00%, which means that the level of data accuracy is good.

Class Precission can be interpreted as a match between a request for information and an answer to the request, so the match between the request and the CURRENT prediction is 81.82% and the match with the CURRENT prediction is 88.89%.

Accredited SINTA 5
72/E/KPT/2024
Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

Class Recall is defined as the ratio of the selected relevant items to the total number of relevant items available. So, from the data above, it can be concluded that the relevant level of true CURRENT is 90.00% and true NOT CURRENT is 80.00%.

3.7 Confusion Matrix Decision Tree C4.5

Confusion matrix is a method that uses true positives, false positives, false negatives and true negatives, to calculate the accuracy of data mining concepts as follows:

The results of training data testing that has been carried out using the Decision Tree C4.5 model obtained an accuracy of 80.00% as seen in Figure 17 below.

Figure 17. Decision Tree Accuracy Results C4.5

The number of CURRENT predictions whose reality is CURRENT is 9 and the number of CURRENT predictions whose reality is NOT CURRENT is 3, Next The number of CURRENT predictions whose reality is NOT CURRENT is 7 and the number of CURRENT predictions whose reality is CURRENT is 1.

The test results that have been carried out using the Decision Tree model obtained an accuracy of 80.00% as seen in Figure 4.18.

Class Precission can be interpreted as a match between a request for information and an answer to that request, so the match between the request and the CURRENT prediction is 75.00% and the match with the NOT CURRENT prediction is 87.50%.

Class Recall is defined as the ratio of the selected relevant items to the total number of relevant items available. So, from the data above, it can be concluded that the relevant level of true CURRENT is 90.00% and true NOT CURRENT is 70.00%.

3.8 Confusion Matrix Fuzzy Decision Tree (ID3)

Confusion matrix is a method that uses true positives, false positives, false negatives and true negatives, to calculate the accuracy of data mining concepts as follows:

The test results that have been carried out using the Fuzzy Decision Tree (ID3) model, obtained an accuracy of 75.00% as seen in Figure 18 below.

Figure 18. Fuzzy Decision Tree (ID3) Accuracy Results

The number of CURRENT predictions that are CURRENT reality is 7 and the number of CURRENT predictions that are CURRENT reality is 2, Next The number of CURRENT predictions that reality is NOT CURRENT is 8 and the number of CURRENT predictions that reality is CURRENT is 3.

The test results that have been carried out using the Fuzzy Decision Tree (ID3) model obtained an accuracy of 75.00% as seen in Figure 4.19.

Class Precission can be interpreted as a match between a request for information and an answer to the request, so the match between the request and the CURRENT prediction is 77.78% and the match with the CURRENT prediction is 72.73%.

Class Recall is defined as the ratio of the selected relevant items to the total number of relevant items available. So, from the data above, it can be concluded that the relevant level of true CURRENT is 70.00% and true NOT CURRENT is 80.00%.

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

4. Analysis of Method Comparison Results

Models using Naive Bayes, Decision Tree C4.5 and Fuzzy Decision Tree (ID3) tested for accuracy produced accuracy, precision and recall values from all three models as shown in Table 1.

Table 1. Accuracy, Precision, Recall and ROC values

	Naive Bayes	Decision Tree C4.5	Fuzzy Decision Tree (ID3)
Accuracy	**85,00%**	**80,00%**	**75,00%**
Submitted Date: MMMM dd, yyyy	**Reviewed Date MMMM dd, yyyy**	**Revised Date: MMMM dd, yyyy**	**Accepted Date: MMMM dd, yyyy**
Recall	**80,00%**	**70,00%**	**80,00%**

Based on Table 1 of the performance comparison results of the three algorithms, the test results for Naive Bayes have a higher accuracy value than Decision Tree C4.5 and Fuzzy Decision Tree (ID3) with a Naive Bayes Algorithm accuracy value of 85.00% while the accuracy value of Decision Tree C4.5 is 80.00% and Fuzzy Decision Tree (ID3) has an accuracy of 75.00%.

In terms of prediction true positive FLUENT Naive Bayes is 9 and true positive NOT CURRENT 8, then true positive FLUENT Decision Tree C4.5 is 9 and true positive NOT CURRENT 7, while true positive FLUENT Fuzzy Decision Tree (ID3) is 7 and true positive NOT CURRENT 8, So predictions for CURRENT and NON-CURRENT results, Naïve Bayes Higher accuracy compared to Decision Tree C4.5 and Fuzzy Decision Tree (ID3).

Furthermore, the comparison of AUC (Area Under Curve) values between the three models is Table 2.

Table 2. Method Performance Comparison

	Naive Bayes	Decision Tree C4.5	Fuzzy Decision Tree (ID3)
AUC	0,960	0,915	0,620

Based on Table 2 the Naive Bayes method produces the highest value with an AUC value of 0.960. Included in the Excellent Classification category.

5. Conclusion

From the results of the comparison above, it can be concluded that these three methods namely Naïve Bayes, Decision Tree C4.5 and Fuzzy Decision Tree (ID3) include very good predictions with AUC Naive Bayes = 0.960, Decision Tree C4.5 = 0.915 and Fuzzy Decision Tree (ID3) = 0.620, but the Naive Bayes methodsuperior with an accuracy value of 85.00% compared to the Decision Tree C4.5 method of 80.00% and the Fuzzy Decision Tree (ID3) method of 75.00%.

After conducting this research, the highest accuracy was produced by the Naïve Bayes algorithm because the probability calculation of the Naïve Bayes algorithm was able to predict a greater positive true value compared to Decision Tree C4.5 and Fuzzy Decision Tree (ID3).

Based on the results of research in predictions determine the smooth financing of fitness center members at PT. Fitindo Sehat Sempurna with a comparison of three data mining classification algorithms, it was concluded that Naive Bayes produced 85.00% accuracy and AUC = 0.960, Decision Tree C4.5 produced 80.00% accuracy and AUC = 0.915, while Fuzzy Decision Tree (ID3) produced 75.00% accuracy and AUC = 0.620; This research makes the Naive Bayes algorithm the best model with the highest accuracy of 85.00% and AUC = 0.960 for solving the problem of predicting smooth financing of fitness center members at PT. Fitindo Sehat Sempurna uses 80 data training and 20 data testing to determine data mining design.

Daftar Pustaka

[1] S. Panggabean, Wahyu Joko Saputro, Faishal Wafiq Zakiy, Tutik Lestari, and Ahmad Rifqi, "Darunnajah Vote System Application Design Using PHP Programming Language," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 13, no. 2, pp. 25–38, 2023, doi: 10.35585/inspir.v13i2.61.

[2] S. Panggabean, "Implementasi Algoritma Dijkstra Untuk Menentukan Jalur Terpendek Wilayah Pasar Minggu Dan STMIK Nusamandiri Jakarta," *Swabumi*, vol. 9, no. 1, pp. 78–85, 2021, doi: 10.31294/swabumi.v9i1.9574.

[3] S. Kusumadewi, "Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification," *CommIT (Communication Inf. Technol. J.*, vol. 3, no. 1, p. 6, 2009, doi: 10.21512/commit.v3i1.506.

[4] D. Ananda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 748, 2024, doi: 10.30865/mib.v8i2.7517.

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 7 No. 6, Desember 2024
P-ISSN 2620-8342
E-ISSN 2621-3052

[5] S. Panggabean, W. Gata, and T. A. Setiawan, "Analysis of Twitter Sentiment Towards Madrasahs Using Classification Methods," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 375–389, 2022, doi: 10.37385/jaets.v4i1.1088.

[6] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "A Decision Tree Algorithm Combined with Linear Regression for Data Classification," *2018 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCCEEE 2018*, pp. 1–5, 2018, doi: 10.1109/ICCCEEE.2018.8515759.

[7] A. Kumar and T. R. Singh, "Analysis for biological network properties of Alzheimer's disease associated gene set by enrichment and topological examinations," *Int. J. Bioinform. Res. Appl.*, vol. 13, no. 3, pp. 214–222, 2017, doi: 10.1504/IJBRA.2017.085856.

[8] R. Hasanah, E. W. Hidayat, and N. I. Kurniati, "Implementasi Deteksi Warna Pada Game Finding Color Menggunakan Ekstraksi Fitur Warna dan Fuzzy Decision Tree," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 1, pp. 137–148, 2020, doi: 10.28932/jutisi.v6i1.2388.

[9] Muchamad Bachram Shidiq, W. Gata, S. Kurniawan, D. D. Saputra, and S. Panggabean, "Time Effort Prediction Of Agile Software Development Using Machine Learning Techniques," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 13, no. 2, pp. 39–48, 2023, doi: 10.35585/inspir.v13i2.57.

[10] D. L. Fithri, "Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naive Bayes," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 2, p. 725, 2016, doi: 10.24176/simet.v7i2.787.

[11] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 20–28, 2018.

[12] S. A. Pratiwi, A. Fauzi, S. Arum, P. Lestari, and Y. Cahyana, "KLIK: Kajian Ilmiah Informatika dan Komputer Prediksi Persediaan Obat Pada Apotek Menggunakan Algoritma Decision Tree," *Media Online*, vol. 4, no. 4, pp. 2381–2388, 2024, doi: 10.30865/klik.v4i4.1681.

[13] A. Ramadhani, R. F. Noor, D. Vernanda, and T. Herdiawan, "Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma C4.5 di Politeknik Negeri Subang," *J. Tekno Kompak*, vol. 18, no. 1, p. 101, 2024, doi: 10.33365/jtk.v18i1.3439.

[14] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 13, no. 1, p. 33, 2018, doi: 10.30872/jim.v13i1.629.

[15] N. D. Prayoga, N. Hidayat, and R. K. Dewi, "Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2666–2671, 2018.

[16] M. R. Kusuma, Windu Gata, Sigit Kurniawan, Dedi Dwi Saputra, and Supriadi Panggabean, "Software Defect Prediction For Quality Evaluation Using Learning Techniques Ensemble Stacking," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 13, no. 2, pp. 1–13, 2023, doi: 10.35585/inspir.v13i2.58.