Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

A COMPARATIVE ANALYSIS OF THE NAIVE BAYES AND C4.5 ALGORITHMS IN DETERMINING ASSISTANT BRANCH MANAGERS AT PT. BANK XYZ

Absony[1], Yulistia[2]
[1,2] Universitas MDP, South Sumatera, Indonesia
[1] mailto:absony_2426311006@mhs.mpd.ac.id, [2],mailto:Yulistia@mdp.ac.id

ABSTRACT
Determining the right leader is a crucial factor in organizational success, including in the banking sector. This study aims to compare the performance of two popular classification algorithms, namely Naïve Bayes and C4.5, in the selection process of Branch Sub-Leaders at PT. Bank XYZ. Using a data mining approach, the research analyzes historical employee data encompassing personal attributes, competencies, and strategic priorities. The evaluation was conducted using a confusion matrix and ROC curve to measure accuracy, precision, recall, and F1-score for each algorithm. The experimental results show that C4.5 delivers superior performance, achieving an accuracy of 0.985 and an AUC of 1.000 in the binary scenario, while Naïve Bayes only reached an accuracy of 0.296 and an AUC of 0.8365. This study confirms that C4.5 is recommended as the primary model to support decision-making by providing the most suitable classification method for objective and transparent leadership placement. Furthermore, it contributes to sustainable managerial strategies through high accuracy and strong interpretability..

Keywords : Data Mining, Naive Bayes, C4.5, Assistant Branch Manager, Confusion Matrix, ROC

INTRODUCTION

A company needs a succession planning strategy, with the goal of forming a solid talent pool through a selection process and the development of programs aimed at positions such as Assistant Branch Manager, which require high potential and capabilities on an ongoing basis, as well as the ability to continue the good performance that has been achieved. the selection process for Assistant Branch Leaders must therefore be carried out accurately, taking into account various factors such as exposure or suitability of experience or expertise gained, suitability of competencies, and fulfillment of expected capabilities so as to improve company performance and ensure organizational sustainability.

To that end, it is necessary to conduct research related to the readiness of determining Assistant Branch Leaders by analyzing several data mining algorithms. Data mining is a process of discovering new patterns from large data sets. This process involves a combination of various fields of science such as artificial intelligence, machine learning, statistics, and database systems with the main objective of obtaining meaningful and easy-to-understand knowledge from complex data (Amna, et al., 2023).

According to several studies and techniques in analysis using the Naive Bayes algorithm and the C4.5 algorithm, both are very effective in finding patterns in data management. The C4.5 algorithm is superior in terms of accuracy and classification capabilities compared to Naive Bayes for marketing personnel placement. However, Naive Bayes has an advantage in terms of classification processing speed (Royadi, 2018). The Naive Bayes algorithm is superior in terms of accuracy and classification capabilities compared to the C4.5 algorithm for determining employee promotions at PT. South Pacific Viscose. Naive Bayes also has an advantage in terms of classification speed (Muhyidin, et al., 2021).

The research scope includes analysis and comparison of Naive Bayes and C4.5 algorithms in the classification process for Assistant Branch Manager placements at PT. Bank XYZ. This research focuses on the use of historical employee data covering personal attributes, competencies, and strategic priorities. The evaluation was conducted using a confusion matrix and ROC curve to measure the accuracy, precision, recall, and F1-score of each algorithm. This study does not cover the development of new algorithms, but rather compares the effectiveness of two existing algorithms in the context of banking organizations.

LITERATURE REVIEW

Naive Bayes and C4.5 algorithms have been widely used in various studies to process data and generate accurate

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

predictions. For example, research by (Nasrullah, et al., 2024) related to the importance of data mining as the basis for applying classification algorithms, Wulandari, et al. (2024) compared the performance of the two algorithms in predicting chronic kidney disease, and Nurhadi (2017) tested the effectiveness of Naive Bayes and C4.5 in medical cases. Additionally, Rudianto et al. (2022) emphasized the superiority of C4.5 in terms of accuracy and interpretability, while Naive Bayes is known for its simplicity and computational speed. Both algorithms have their own advantages depending on the type and characteristics of the data used.

Data Mining

Data mining is the process of extracting previously unknown but useful information from available data so that it can be used to support decision making. In the context of business, government, and research, data mining is often used for customer behavior analysis, fraud detection, trend prediction, and data classification (Amna, et al., 2023). The techniques used in data mining are diverse, including classification, regression, clustering, and association (Rahayu, et al., 2024). One of the most commonly used techniques is classification, which is the process of mapping data into specific categories based on the attributes it possesses or contains (Hapsari, et al., 2024).

Naive Bayes Classification Algorithm The assumption of attribute independence in Naive Bayes is rarely perfectly fulfilled, but this method remains effective, especially with clean and uncorrelated data. the Bayes theorem that forms the basis of the calculation can be seen in Equation 2:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)}$$

in which:
- P(C|X) : Probability of class C given data X.
- P(X|C) : Probability of data X appearing if the class is C.
- P(C) : Initial probability (prior) of class C.
- P(X) : Probability of data X (constant for all classes).

In classification, choose the class with the largest P(C|X) value.
as shown in the following figure

C4.5 Classification Algorithm

The C4.5 algorithm is a decision tree-based classification method developed by Quinlan as an extension of the ID3 algorithm.

$$Entropy(S) = -\sum_{i=1}^{n} \rho_i Log_2 \rho_i$$

where:
ρ_i = proportion of data in class i.
n = number of classes.

Classification Model Performance Evaluation

an important stage in the data mining process to assess how well the model predicts data classes accurately. One of the most commonly used evaluation methods is the confusion matrix, which is a table that shows the number of correct and incorrect predictions made by the classification model compared to the actual data (S, et al., 2023).

Figure 1 explains that the ROC curve provides a visual representation of the model's ability to distinguish between positive and negative classes. The closer the ROC curve is to the upper left corner of the graph, the better the model's performance. To measure performance quantitatively, AUC (Area Under Curve) is used, which is the area under the ROC curve. The AUC value ranges from 0 to 1, where a value close to 1 indicates an excellent model, while a value close to 0.5 indicates a model that is no better than random guessing (S, et al., 2023).

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

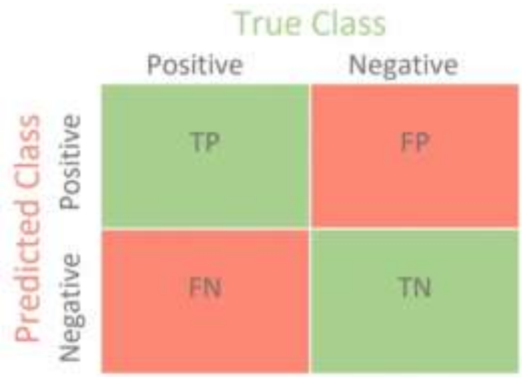


Fig 1. Bagan ROC (Receiver Operating Characteristic)

METHOD

The Research was conducted through several systematic stages designed to achieve the research objectives, namely comparing the performance of the Naive Bayes and C4.5 algorithms in classifying Assistant Branch Manager placements. The stages of research conducted are as follows:

Methodological Framework Figure 2 shows the stages of the research flow that are used as guidelines in this study. in Figure 2

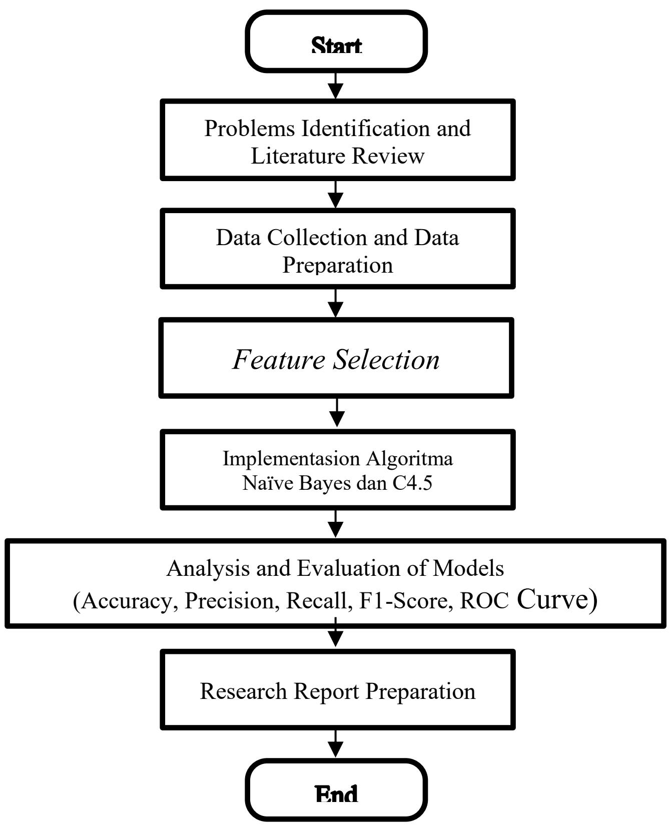


Fig. 2 Framework of Research Flow

Data Collection and Preparation

This stage is an important step in the research process because data quality greatly determines the accuracy and success of the classification model to be built.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

a. Data Sources
The data used in this study was obtained from internal sources at PT. Bank XYZ in 2025, with a total of 1392 records and 39 labels. In accordance with research ethics, personal identities (PII) were anonymized and only analytical attributes were used for modeling.
b. Data Preparation (Data Preprocessing)
c. Attribute Selection (Feature Selection)The Chi-Square (Chi2) method was applied to the TALENT_CLASS_PERIODE_2025 label to identify the most influential attributes. The selection results show that 2024 YEAR-END EVALUATION, TOTAL EVALUATION FOR THE LAST 3 YEARS, 2024 TALENT_CLASS, TOTAL TALENT CLASS FOR 3 YEARS, GENDER, and LENGTH OF SERVICE (YEARS) have the highest scores, making them core feature candidates in the modeling stage. Data Division (Train-Test Split)In the model testing stage, the dataset of 1,392 records that had undergone cleaning and transformation and the removal of empty target rows in the 2025 Talent Class Period became 1,055 valid records, which were then divided into two parts, namely the training data (training set) and the test data (testing set). This division was carried out with a ratio of 80% (844 records) for training and 20% (211 records) for testing.

Analysis and Interpretation of Results

This study is categorized as quantitative, using a comparative experimental method. This approach was chosen because the study aims to test and compare the performance of two classification algorithms, namely Naive Bayes and C4.5, in determining the suitability of Assistant Branch Leaders based on historical employee data.

Quantitative characteristics are reflected in the use of measurable numerical and categorical data, as well as statistics-based analysis to assess model performance. The comparative experimental method was used because this study not only built a classification model, but also tested both algorithms using the same procedure, thereby enabling an objective comparison of performance.

The selection of this approach was based on several considerations, namely:

a. The research variables are measurable, so they can be analyzed using statistical methods and classification algorithms.
b. The need for objectivity and reproducibility, where the research results must be retested using the same procedure to obtain consistent results.
c. Ease of interpreting results, particularly in the C4.5 algorithm, which produces decision trees and rules that can be understood by decision makers.

RESULT

Research results that were obtained through data processing, algorithm modeling, and performance evaluation to answer the research objectives that had been formulated, namely analyzing and comparing the performance of the Naive Bayes and C4.5 algorithms in classifying Assistant Branch Manager placements, identifying the advantages and disadvantages of each method, and providing recommendations on the best algorithm to support objective and efficient decision making.

Research data was sourced from PT. Bank XYZ's internal historical data, which has been anonymized to maintain confidentiality. Data processing was carried out to ensure the quality of the dataset before it was used in algorithm modeling. The dataset used was sourced from PT. Bank XYZ's historical employee data, consisting of 1,392 records and 39 attributes. The following steps were taken to improve the quality of the dataset and analytical readiness prior to modeling:

1. Data Validation and Cleaning
Of the 1,392 initial records, duplication checks, missing value handling, and deletion of rows without target labels were performed. Empty values in numerical attributes were filled using median imputation, while rows without Talent Class Period 2025 were deleted.
2. Transformation Stage
Data transformation is performed to ensure that all attributes are in numerical format to meet the requirements of the Naïve Bayes and C4.5 algorithms.
3. Normalization Stage
Normalization is performed to standardize the scale of numerical attributes into a range of 0–1 using the MinMaxScaler method. Normalization has a significant effect on data quality and the feature selection process
4. Feature Selection Stage
The feature selection process applies the Chi-Square method to identify from 39 attributes, which have previously undergone validation, data cleaning, transformation, and normalization, those that have a significant influence

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

on the target variable, namely Talent Class Period 2025. The implementation of this method was carried out using the scikit-learn library in the Python programming language, which enables efficient calculation of Chi-Square scores in figure 3

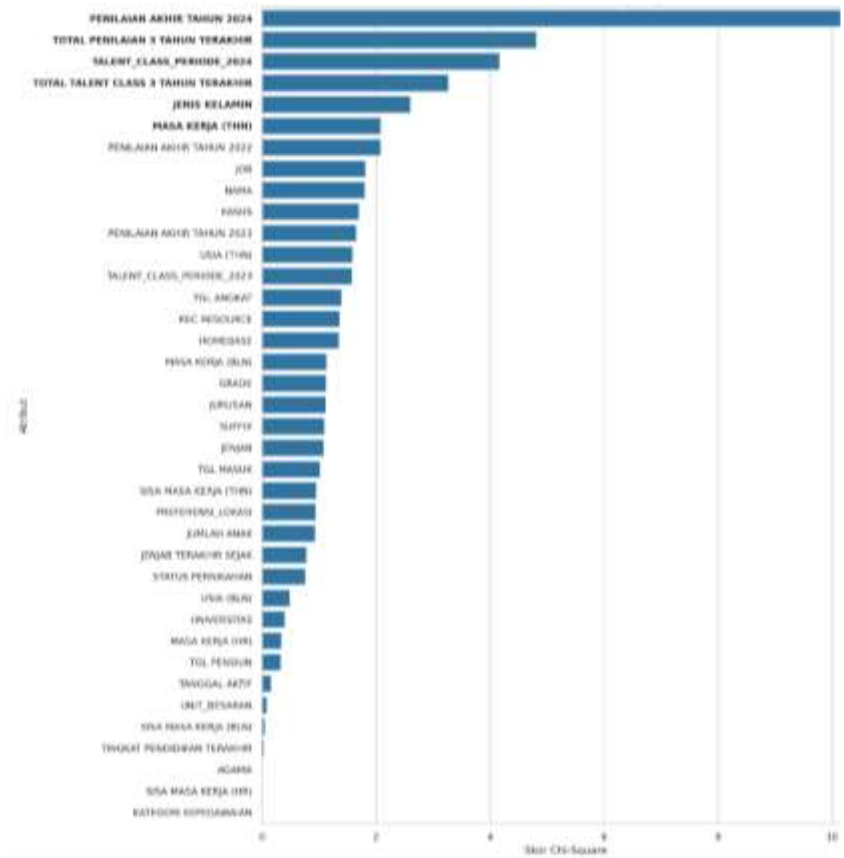


Fig. 3 Attribute Results Based on Chi-Square Scores

This difference can be explained in Figure 4, where the characteristics of the C4.5 algorithm are shown. C4.5 is capable of effectively handling categorical and numerical attributes and building decision tree-based rules that correspond to complex patterns in the data. In contrast, Naive Bayes assumes independence between features, which is rarely the case in human resource data, resulting in a drastic decrease in accuracy.

In the context of selecting branch managers, accuracy and precision are crucial because classification errors can impact leadership quality and organizational performance. Models with high precision, such as C4.5, are more reliable in ensuring that candidates predicted as “Very Good” truly have superior competencies. This is important to avoid the risk of placing unsuitable individuals in strategic positions. Although C4.5's Recall is not yet perfect at 0.500 (50%), this model is still better at identifying potential candidates than Naive Bayes. Thus, C4.5 can be used as a decision-making tool that supports the principles of meritocracy and data-driven management.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

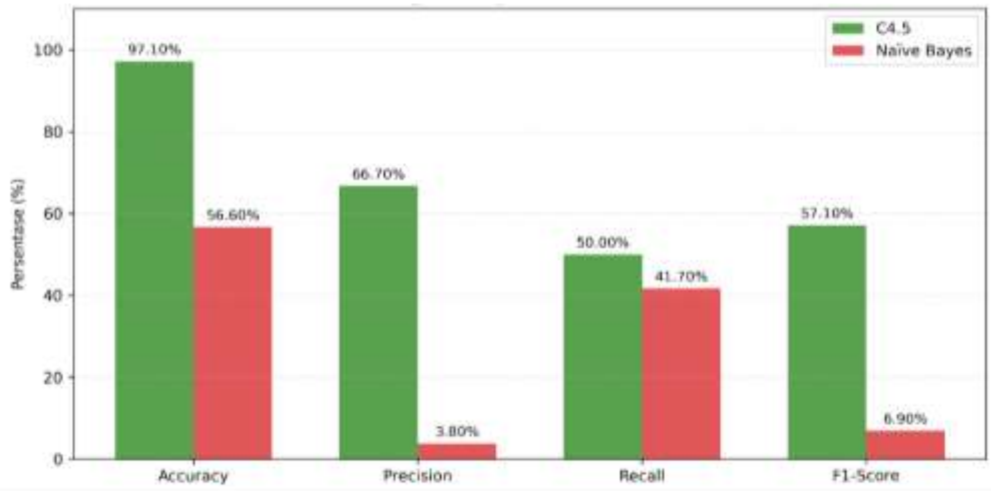


Fig. 4 Comparison of Final Performance Results for Binary Scenarios

Comparative Analysis Steps and Model Recommendations

Comparative analysis is conducted through four main steps to ensure that the selected model is not only superior in terms of metrics, but also relevant to business needs, namely:

1. The Talent Class dataset used in this study comes from historical employee data at PT. Bank XYZ, which has undergone data preprocessing to ensure quality and consistency. The first stage was feature selection using the Chi-Square method, which produced six attributes that were most influential on the 2025 Talent Class classification, namely the 2024 Year-End Assessment, Total Assessment for the Last 3 Years, 2024 Talent Class, Total Talent Class for the Last 3 Years, Gender, and Length of Service. Next, data division (train-test split) was performed with a ratio of 80:20, where 844 records were used for training and 211 records for testing. The final stage was modeling, where both algorithms, Naïve Bayes and C4.5, were applied to the processed dataset. Naïve Bayes builds a probabilistic model based on the assumption of independence between attributes, while C4.5 forms a decision tree based on the gain ratio value to determine the most informative attributes. This process ensures that both algorithms are tested using the same procedure, so that the comparison results can answer the problem formulation related to the effectiveness of each method in supporting the determination of Assistant Branch Leaders.
2. The test results show a significant difference between the performance of the C4.5 and Naïve Bayes algorithms in two evaluation scenarios. In the multi-class scenario (Very Good, Good, Fair, Poor), C4.5 recorded an Accuracy of 0.9621 with Precision and Recall of 0.5669 and 0.5649, respectively, while Naïve Bayes only achieved an Accuracy of 0.2133 and an F1-Score of 0.1325. These values indicate that although C4.5 excels overall, the complexity of separating four categories causes a decrease in precision and recall, especially in minority classes. This shows that decision tree models are more adaptive to attribute variations but still face challenges in classifying classes with unbalanced data distribution. Conversely, in the binary scenario (Good vs. Very Good), C4.5 showed near-perfect performance with an accuracy of 0.985, precision of 0.985, recall of 0.750, and an F1-score of 0.800, while Naïve Bayes only recorded an accuracy of 0.296 and a precision of 0.034. The high precision value in C4.5 indicates that almost all recommended candidates are truly from the superior category, so the risk of mispromotion is very small. The binary scenario is considered more relevant for promotion needs because the main focus of the selection process is to distinguish superior candidates (Very Good) from good candidates, not to separate all categories. With only two classes, the model can form rules that are sharper and more suited to business needs, thereby supporting the promotion process.
3. The effectiveness of classification models on imbalanced data, such as in the case of branch leader selection, cannot be adequately measured using ROC–AUC alone. Although ROC provides a general overview of discriminatory ability, this metric tends to be optimistic when class distributions are highly imbalanced. Therefore, Precision–Recall Curve (PR–AUC) is recommended as an advanced evaluation approach. PR–AUC is more sensitive to changes in precision and recall in the positive class (Very Good), thus providing a more accurate picture of the model's ability to identify priority candidates. In addition to PR–AUC, positive class F1 is an important metric because it emphasizes the balance between precision and recall for the most relevant category in the promotion process. In this context, high precision ensures that the recommended candidates are truly superior, while high recall ensures that as many

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

potential candidates as possible are identified. The combination of the two through the positive class F1 supports fairer and more effective decision making.

4. The Practical Implications of these results support more objective and transparent promotion decisions through the use of data-based classification algorithms. With high accuracy and interpretability of C.45, companies can reduce the risk of misplacing assistant branch leaders, thereby improving business, service, and operational performance. In addition, the application of this model accelerates the candidate selection process and can form the basis for the development of a Decision Support System integrated with HR applications to support sustainable succession planning.

DISCUSSIONS

The results In Figure 5, ROC reinforces this finding. The red curve (C4.5) is almost flush with the upper axis with an AUC = 0.985, indicating a very high discrimination ability between the "Good" and "Very Good" classes. In contrast, the blue curve (Naive Bayes) approaches the random diagonal line with an AUC of 0.495, meaning its performance is almost equivalent to random guessing. This difference confirms that C4.5 is more effective at learning complex patterns in human resource data than Naive Bayes, which is hampered by the assumption of feature independence.

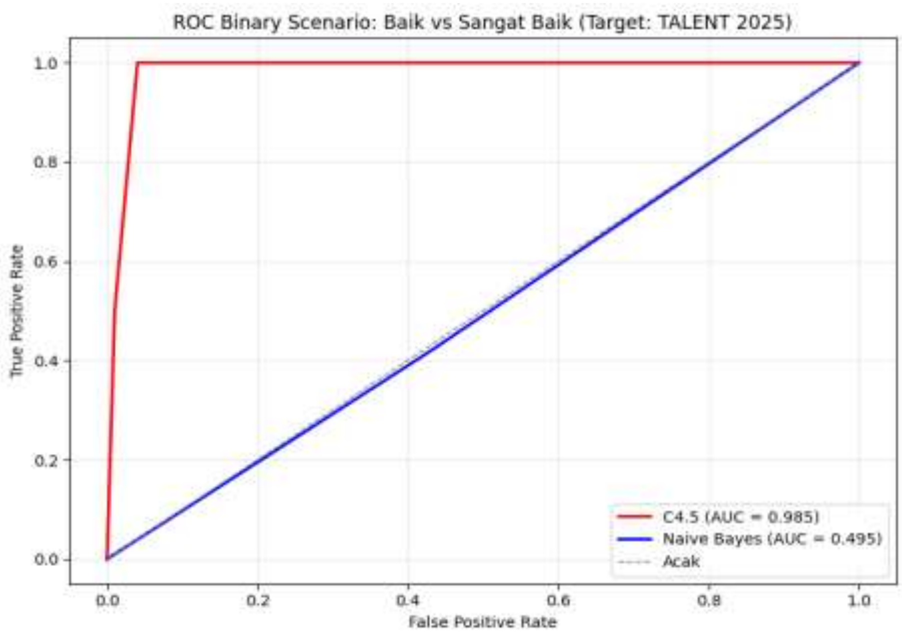
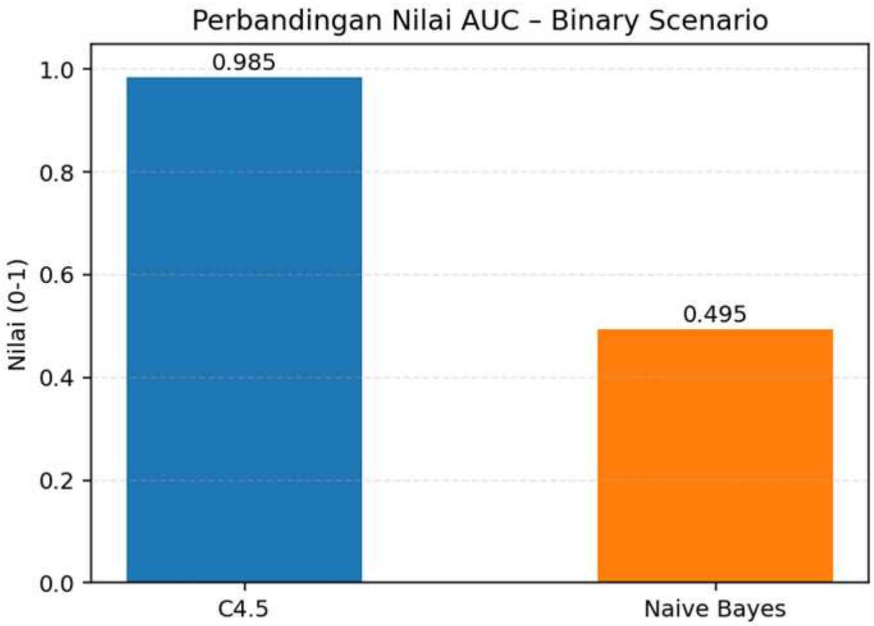


Fig. 5 ROC and AUC Results for Binary Scenario (Good vs. Very Good)

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16



` Fig.6 AUC Comparison Results – Binary Scenario

In the binary scenario in Figure that distinguishes between the Very Good and Good categories, the near-perfect AUC value in C4.5 indicates that this model is consistently able to give higher probability scores to candidates in the "Very Good" category compared to "Good" across the entire threshold range, while Naïve Bayes fails to demonstrate this ranking quality. This difference is consistent with the performance pattern seen in the binary AUC graph, confirming that C4.5 is more suitable for use in the promotion selection process, which requires accurate identification of candidates with superior potential.In the context of selecting branch leaders, the accuracy and discrimination ability of the model are very important because this position requires high competence, integrity, and leadership potential. Errors in identifying "Very Good" candidates can have an impact on organizational performance and business strategy success.

CONCLUSION

Based According to the results of the research and analysis conducted, the following conclusions can be drawn:

The C4.5 algorithm shows significantly better performance than Naïve Bayes in Talent Class classification to support the selection of branch leaders. In the binary scenario (Good vs. Very Good), C4.5 achieved Accuracy = 0.985, Precision = 0.985, Recall = 0.750, F1-Score = 0.800, and AUC Curve = 1.000, which indicates near-optimal discrimination capabilities. In contrast, Naïve Bayes only recorded Accuracy = 0.296, Precision = 0.034, Recall = 0.625, F1-Score = 0.065, and AUC Curve = 0.8365.

Based on the test results, the C4.5 algorithm is recommended as the main model because it has high accuracy and interpretability, which facilitates translation into transparent policies.

The very high AUC curve value in the C4.5 algorithm indicates the potential for overfitting and dependence on historical features. This condition can reduce the ability to generalize to new data

This research plays a direct role in supporting the process of determining Assistant Branch Leaders through an objective data-based approach. By utilizing data-based classification algorithms, companies can reduce the risk of errors in determining assistant branch leaders, improve the efficiency of the selection process, and strengthen succession planning strategies to ensure that superior candidates are consistently identified.

REFERENCES

Alfarobi, I. (2017). KOMPARASI ALGORITMA C4.5, NAIVE BAYES, DAN RANDOM FOREST UNTUK KLASIFIKASI DATA KELULUSAN MAHASISWA.

Amna, S, W., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wadhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). Data Mining (Vol. 2, Issue January 2013). https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.7938

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

Penerima Beasiswa. 8(1), 2088–4591. https://ejournal.upm.ac.id/index.php/energy/article/view/111

Astuti, M. (2024). Perbandingan Metode Random Forest Dan Naive Bayes Pada Klasifikasi Perilaku Mahasiswa Di LMS Spada Indonesia. 15(1), 37–48. https://repository.unhas.ac.id/id/eprint/41519/2/H062201003_tesis_27-08-2024 bab I-II%28FILEminimizer%29.pdf

Astuti, R. D. (2019). Analisis Perbandingan Algoritma K-Means Dan K-Medoids Untuk Menerapkan Segmentasi Pelanggan. https://repository.nusamandiri.ac.id/repo/files/250264/download/FILE TESIS.pdf

Baskoro. (2022). Klasifikasi Penyakit Jantung Menggunakan Particle Swarm Optimization Dan Algoritma K-Nearest Neighbors. http://repo.darmajaya.ac.id/10193/1/Tesis Full.pdf

Daewana, F. (2025). Perbandingan Algoritma Naïve Bayes Dan C4 . 5 Dalam Prediksi Keputusan Karyawan Untuk Meninggalkan Perusahaan. 5, 1712–1726.

Dharmawan, W. S. (2018). Feature Selection Berbasis Abc-Svm Dan Pso-Svm Dalam Masalah Klasifikasi. Repository.Bsi.Ac.Id. https://repository.bsi.ac.id/index.php/unduh/item/340225/Thesis_Weiskhy-Steven-Dharmawan.pdf

Fitriani, E. (2018). Komparasi Algoritma Klasifikasi Data Mining Bantuan Program Keluarga Harapan. https://repository.bsi.ac.id/repo/files/346147/download/Full-Tesis.pdf

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. Jurnal Teknologi Informasi Dan Ilmu Komputer, 11(1), 215–224. https://doi.org/10.25126/jtiik.20241118074

Hapsari, R. K., Sumarni, T., Hidayatulloh, T., Farida, Herianto, Elfaladonna, F., Aini, N., Nurfitria, Kurniawan, H., Kraugusteeliana, Yusuf, L., & Widiatama, Y. (2024). Data Mining (Vol. 11, Issue 1).

Herteno, R., Budiman, I., Kartini, D., & Mazdadi, M. I. (2025). Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4 . 5. 01.

Iriadi, N., & Nuraeni, N. (2016). Kajian Penerapan Metode Klasifikasi Data Kelayakan Kredit Pada Bank. II(1), 132–137.

Jayanto, I., & Benisius. (2024). Analisis Perbandingan Algoritma Decision Tree untuk Prediksi Karyawan dengan Potensi Atrisi di PT. XYZ. 22(1), 49–59. https://doi.org/10.61805/fahma.v22i1.112

Mega, M., & Jasmir. (2023). Prediksi Masa Studi Mahasiswa Unama Jambi Menggunakan Metode Algoritma C4.5. 8(1), 140–151. https://doi.org/10.33998/jurnalmsi.2023.8.1.770

Mendrofa, H. K. K., Waruwu, M. H., Mendrofa, Y., & Bate'e, M. M. (2024). Analisis Strategi Suksesi Pimpinan Yang Efektif Di PT Bank Negara Indonesia (Persero) Tbk Cabang

Alfarobi, I. (2017). KOMPARASI ALGORITMA C4.5, NAIVE BAYES, DAN RANDOM FOREST UNTUK KLASIFIKASI DATA KELULUSAN MAHASISWA.

Amna, S, W., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wadhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). Data Mining (Vol. 2, Issue January 2013). https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. 8(1), 2088–4591. https://ejournal.upm.ac.id/index.php/energy/article/view/111

Astuti, M. (2024). Perbandingan Metode Random Forest Dan Naive Bayes Pada Klasifikasi Perilaku Mahasiswa Di LMS Spada Indonesia. 15(1), 37–48. https://repository.unhas.ac.id/id/eprint/41519/2/H062201003_tesis_27-08-2024 bab I-II%28FILEminimizer%29.pdf

Astuti, R. D. (2019). Analisis Perbandingan Algoritma K-Means Dan K-Medoids Untuk Menerapkan Segmentasi Pelanggan. https://repository.nusamandiri.ac.id/repo/files/250264/download/FILE TESIS.pdf

Baskoro. (2022). Klasifikasi Penyakit Jantung Menggunakan Particle Swarm Optimization Dan Algoritma K-Nearest Neighbors. http://repo.darmajaya.ac.id/10193/1/Tesis Full.pdf

Fitriani, E. (2018). Komparasi Algoritma Klasifikasi Data Mining Bantuan Program Keluarga Harapan. https://repository.bsi.ac.id/repo/files/346147/download/Full-Tesis.pdf

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. Jurnal Teknologi Informasi Dan Ilmu Komputer, 11(1), 215–224. https://doi.org/10.25126/jtiik.20241118074

Hapsari, R. K., Sumarni, T., Hidayatulloh, T., Farida, Herianto, Elfaladonna, F., Aini, N., Nurfitria, Kurniawan, H., Kraugusteeliana, Yusuf, L., & Widiatama, Y. (2024). Data Mining (Vol. 11, Issue 1).

Herteno, R., Budiman, I., Kartini, D., & Mazdadi, M. I. (2025). Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4 . 5. 01.

Iriadi, N., & Nuraeni, N. (2016). Kajian Penerapan Metode Klasifikasi Data Kelayakan Kredit Pada Bank. II(1), 132–137.

* Corresponding author

Submitted : 2026-01-31
Accepted : 2026-02-04
Published : 2026-04-16

Jayanto, I., & Benisius. (2024). Analisis Perbandingan Algoritma Decision Tree untuk Prediksi Karyawan dengan Potensi Atrisi di PT. XYZ. 22(1), 49–59. https://doi.org/10.61805/fahma.v22i1.112

Mega, M., & Jasmir. (2023). Prediksi Masa Studi Mahasiswa Unama Jambi Menggunakan Metode Algoritma C4.5. 8(1), 140–151. https://doi.org/10.33998/jurnalmsi.2023.8.1.770

Muhyidin, Y., Alfiansyah, K., & Hutapea, E. P. (2021). Perbandingan Performa Algoritma Naive Bayes Dan C45 (Studi Kasus pemilihan promosi karyawan PT. South Pacific Viscose). 45.

Nasrullah, M. F., Saedudin, R. R., & Hamami, F. (2024). Perbandingan Akurasi Algoritma C4.5 Dan K-Nearest Neighbors Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Indonesia. 9(2), 628–638. https://doi.org/10.29100/jipi.v9i2.4655

Nurhadi, A. A. (2017). Perbandingan Algoritma C4.5 Dan Naive Bayes Pada Kasus Hepatitis Dan Jantung. 356797.

Pauziah, U. (2017). Analisis Penentuan Karyawan Terbaik Menggunakan Metode Algoritma Naive Bayes (Studi Kasus PT. XYZ). 1(1), 94–102.

Pradytya, A. D. (2018). Kajian Data Mining Untuk Memprediksi Kelulusan Mahasiswa Dengan Metode Klasifikasi (Studi Kasus : STMIK LIKMI Bandung).

Pramitasari, A. E., & Nataliani, Y. (2021). Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja Dengan Algoritma K-Means Dan Fuzzy C-Means. 8(3), 1119–1132. https://doi.org/10.35957/jatisi.v8i3.957

Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M. (2024). Buku Ajar Data Mining (Vol. 1, Issue January 2024).

Rakhmawati, H., & Sindyka, A. M. (2025). PERBANDINGAN ALGORITMA NAIVE BAYES DAN C4 . 5 DALAM DIAGNOSIS PENYAKIT PARU-PARU. 15(3), 424–432.

Ridwan, A., & Khoiriyah, A. T. (2020). PENERAPAN TEKNIK BAGGING PADA ALGORITMA NAIVE BAYES DAN ALGORITMA C4.5 UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS. 1, 41–48.

Rohman, A., & Rochcham, M. (2019). Komporasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. 5(1).

Royadi. (2018). Perbandingan Metode Data Mining Naïve Bayes Dan Algoritma C4 . 5 Untuk Klasifikasi Penempatan Tenaga Marketing Perbandingan Metode Data Mining Naïve Bayes.

S, W., Rismayani, Sihotang, J. I., Aisa, S., Gunawan, H., Tamsir, N., Masturoh, S., Radiyah, U., Gustiana, Z., Harlina, S., & Muslihi, M. T. (2023). Data Warehouse Dan Data Mining.

Santoso, L., & Priyadi. (2024). Comparative Study of Feature Engineering Techniques for Predictive Data Analytics. Journal of Technology Informatics and Engineering, 3(2), 417–435. https://doi.org/10.51903/jtie.v3i2.225

Saptarini, N. G. A. P. H. (2016). Penentuan Talenta Karyawan Berdasarkan Menggunakan Konsep Data Mining. 2(1), 34. https://doi.org/10.32511/jiflash.v2i1.22

Septiani, D. (2017). Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. 13(1), 76–84. http://archive.ics.uci.edu/ml/.

Setiawan, A. B. (2018). Kajian Komparasi Penerapan Algoritma C4.5 Dan Naïve Bayes Sebagai Penunjang Keputusan Pinjaman Uang (Studi Kasus Di Koperasi Karyawan PT. Karyamitra Budisentosa Pandaan Pasuruan).

Stephanie, C. (2019). Penerapan Metode C4.5, KNN, SVM, RF Dalam Klasifikasi Talenta Karyawan Untuk Memperoleh Tingkat Akurasi Tertinggi.

Sugara, B., Adidarma, D., & Budilaksono, S. (2019). Perbandingan Akurasi Algoritma C4.5 dan Naïve Bayes untuk Deteksi Dini Gangguan Autisme pada Anak. 3(1), 119–128.

Sugiyarto, I. (2019). Perbandingan Kinerja Algoritma Data Mining Prediksi Persetujuan Kartu Kredit. Faktor Exacta, 12(3), 180. https://doi.org/10.30998/faktorexacta.v12i3.4310

Supriyadi, A. (2023). Perbandingan Algoritma Naive Bayes dan Decision Tree(C4.5) dalam Klasifikasi Dosen Berprestasi. 7(1), 39–49. https://doi.org/10.29407/gj.v7i1.19797

Suyadi, Setyanto, A., & Fattah, H. Al. (2017). Analisis Perbandingan Algoritma Decision Tree (C4 . 5) Dan K-Naive Bayes Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas. 2(1), 59–68.

Walim. (2018). Analisis perbandingan algoritma naive bayes , random forest dan c.45 dalam klasifikasi kelayakan masyarakat untuk mendapatkan bantuan dana desa.

Windarti, M., & Suradi, A. (2019). Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa. 12(1), 14. https://doi.org/10.35671/telematika.v12i1.778

Wulandari, V., Sari, W. J., & Alfian, Z. (2024). Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik. 4(April), 710–718.

Yang, S., & Berdine, G. (2017). The receiver operating characteristic (ROC) curve. 5(19), 34. https://doi.org/10.12746/swrccc.v5i19.391

* Corresponding author