Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13860
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation of the C4.5 and Naive Bayes Algorithms to Predict Student Graduation

Lianah[1)*], Syaiful Zuhri Harahap[2)], Irmayanti[3)]
[1,2,3)] Universitas Labuhanbatu, Indonesia
[1)] lialianah60@gmail.com, [2)] syaifulzuhriharahap@gmail.com, [3)] irmayantiritonga2@gmail.com

Submitted : Jul 12, 2024 | **Accepted** : Jul 25, 2024 | **Published** : Jul 27, 2024

Abstract: This research aims to determine student graduation using two data mining methods, namely the Naive Bayes Classifier and the C4.5 Algorithm. Research stages include data analysis, data pre-processing, model design in data mining, classification results, method evaluation, and evaluation results. This research uses student data consisting of training data and testing data to evaluate the performance of the two methods in predicting student graduation based on attributes such as attendance scores, behavior scores, Final Semester Examination (UAS) scores, and report card scores. The classification results show significant differences between the two methods. The Naive Bayes Classifier produces predictions that 37 students pass and 17 students do not pass, while the C4.5 Algorithm predicts that 30 students pass and 24 students do not pass. This difference in results indicates that there are differences in the approaches of the two methods to student graduation data, with the Naive Bayes Classifier tending to provide more positive predictions than the C4.5 Algorithm. Evaluation of the performance of the method shows that the Naive Bayes Classifier has an accuracy rate of 100%, which is a perfect result, while the C4.5 Algorithm has an accuracy rate of 89%. This significant difference in evaluation results confirms that the Naive Bayes Classifier is superior in classifying student graduation compared to the C4.5 Algorithm in the context of this research. These findings can help in making decisions regarding student graduation evaluations in the future.

Keywords: C4.5 algorithm; Classification; Confusion Matrix; Data Mining; Naïve Bayes Classifier Method; Tree Viewer

INTRODUCTION

Graduation is an important moment in one's educational journey, marking academic achievement after years of study and hard work (Almufqi & Voutama, 2023). This moment is not only the culmination of an individual's struggle in a particular field of study, but also marks the beginning of a new chapter in life, whether it is continuing to higher education or entering the world of work. For many people, graduation is a symbol of their ability to overcome challenges and achieve set goals, as well as an opportunity to celebrate success with friends, family and mentors who have provided support throughout the learning process. More than just a ceremonial ceremony, graduation also brings new responsibilities and expectations. Graduates are expected to be able to apply the knowledge and skills acquired during their studies to contribute positively to society and their chosen profession. In this context, graduation becomes the starting point for pursuing the desired career, developing further skills, and facing real challenges in the outside world. Additionally, this moment also often triggers reflection on the past and planning for the future, encouraging graduates to continue learning and developing throughout their lives.

Graduation of students at upper levels such as high school or vocational school is a very important and valuable moment in their educational journey (Sibuea & Safta, 2017). It marks the end of the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

secondary education phase and is a stepping stone to a more mature stage of life, whether that is progressing to higher education or directly entering the world of work. Graduation at this level indicates that students have successfully completed a complex and challenging curriculum, and have developed the skills and knowledge required for the next stage. It is a time to celebrate accomplishments, reflect on struggles, and appreciate the support of family, teachers, and friends. For high school students, graduation opens the door to college, where they can pursue a more specific and in-depth field of study suited to their interests and talents. Meanwhile, for vocational school students, graduation often means being ready to enter the world of work with the practical and technical skills they have learned. The importance of graduation also lies in the transition from adolescence to adulthood, where students begin to take greater responsibility for their choices and future. With a graduation certificate in hand, they have important capital to build a productive and successful career and life in the future.

Graduating students and girls is not always an easy and smooth process. Often, there are gaps in graduation standards that result in students who rarely attend class or show poor behavior to still be able to graduate (Nurjana, Perdana Windarto, Qurniawan, & Tunas Bangsa, 2022). This can occur due to a variety of factors, including pressure to increase graduation rates, a lack of resources to provide additional support for struggling students, and educational policies that may place more emphasis on the end result than the learning process itself. In some cases, students who are academically ineligible or demonstrate inadequate behavior can still earn graduation, while diligent and well-behaved students may face unfair challenges in achieving their goals. On the other hand, there are also students who attend class diligently, behave well, and work hard throughout the year, but still face difficulties in graduating due to various reasons such as personal problems, lack of support at home, or unfavorable socio-economic conditions. This situation reflects existing gaps in the education system, where external and non-academic factors can influence students' academic outcomes. It is therefore important for education systems to address these disparities with a more equitable and inclusive approach, providing the support needed for all students to reach their full potential and ensuring that graduation truly reflects true academic achievement and personal development.

Graduating students have certain categories or attributes that are used to determine whether they can be declared passed or not passed (E. A. Saputra & Nataliani, 2021). These attributes include attendance scores, behavior scores, Final Semester Examination (UAS) scores, and report card scores. Attendance scores reflect students' consistency and responsibility in attending lessons, while behavior scores show students' attitudes and ethics while at school. The UAS score provides an overview of the student's understanding of the material that has been taught throughout the semester, and the report card score includes an overall assessment of the student's academic performance during one school year (Susanto & Sudiyatno, 2014). By evaluating these four attributes, schools can ensure that graduation is based on comprehensive academic achievement and good behavior. The existence of these four attributes makes the process of determining graduation more transparent and objective. Students who have high attendance scores, good behavior, satisfactory UAS scores, and good report cards are likely to pass, while students who lack one or more of these attributes may need to improve their performance before they can graduate. By using these various attributes, schools can provide a fairer and more comprehensive assessment of students' abilities and attitudes, ensuring that only students who truly meet the established educational standards can be declared graduates. This also encourages students to not only focus on academic achievement but also on aspects of good attendance and behavior at school.

The author will conduct research on graduating students using four main attributes: attendance scores, behavior scores, Final Semester Examination (UAS) scores, and report card scores. With these values, this research aims to provide clear results regarding students who passed and did not pass. The samples used in this research were students of class XII (twelve) at SMA Negeri 1 Silangkitang. The data collected will be analyzed to determine the most significant factors influencing student graduation, so as to provide deeper insight into important aspects of the educational process at the school. This research will be carried out using data mining techniques with two main methods: Naive Bayes Classifier and the C4.5 algorithm. The Naive Bayes Classifier method will be used to classify data based on the probability of each attribute, while the C4.5 algorithm will build a decision tree to identify patterns and relationships between attributes. As was done in research (Amalia, 2020) the Naïve Bayes method can be used to classify student graduation. Research conducted in (Lizsara, Oyama, & Wardani,

*name of corresponding author

2020) shows that the application of the Naïve Bayes method to determine graduation can be done in data mining. By using these two methods, the author hopes to obtain accurate and reliable results regarding student graduation, as well as provide useful recommendations for schools in improving the assessment and learning process. It is hoped that the results of this research will help in better understanding the factors that influence graduation and create more effective strategies to support students' academic success. In research (Punkastyo, Septian, & Syaripudin, 2024) the application of the Naïve Bayes method to data mining to determine student graduation can be done with an accuracy of 98%.

LITERATURE REVIEW

Data mining is the process of analyzing large amounts of data to discover hidden patterns, relationships and trends in it (Abas et al., 2023) (A. D. S. Saputra, Hindarto, & Haryono, 2023). This process uses statistical techniques, mathematics, and machine learning algorithms to identify useful information that can assist in decision making (Bustomi, Nugraha, Juliane, & Rahayu, 2023). Data mining is often used in various fields such as business, health, social sciences, and education to optimize performance, predict behavior, and gain deeper insights from available data (Aji & Devi, 2023). With the ability to process and analyze data effectively, data mining enables organizations and researchers to transform raw data into valuable knowledge and strategic actions. In research conducted (S. A. Hasibuan, Sihombing, & Nasution, 2023) data mining can be used to classify data. Just like in research (Mawaddah, Dar, & Yanris, 2023) that data mining can be used for data classification. The Naive Bayes Classifier method is a machine learning algorithm used for classification, based on Bayes' Theorem with the assumption of independence between features

The Naive Bayes Classifier method is a machine learning algorithm used for classification, based on Bayes' Theorem with the assumption of independence between features (Anam, Rahmiati, Paradila, Mardainis, & Machdalena, 2023) (Rahman & Fauzi Abdulloh, 2023). This method is very effective and efficient, especially for large datasets, because it is simple but powerful in handling text classification problems such as spam filters, sentiment analysis, and pattern recognition (Alam, Alana, & Juliane, 2023) (Apriyani, Maskuri, Ratsanjani, Pramudhita, & Rawansyah, 2023). Naive Bayes works by calculating the probability of each possible class given certain attributes and then selecting the class with the highest probability (Madjid, Ratnawati, & Rahayudi, 2023). Although the independence assumption is rarely strictly met in practice, this method often provides good performance and accurate results, making it a popular choice in a variety of data mining and machine learning applications. In research conducted by (F. F. Hasibuan, Dar, & Yanris, 2023) the naïve Bayes method can be used to classify levels of satisfaction carried out in data mining and obtain accuracy results of 97%. In research (Siregar, Irmayani, & Sari, 2023) the naïve Bayes method is able to classify people who deserve social assistance which is able to provide 100% accuracy results.

The C4.5 algorithm is a machine learning algorithm used to build decision trees, which are useful for classification purposes (Maizura, Sihombing, & Dar, 2023). This algorithm works by dividing the dataset based on the attributes that provide the highest information about the target class, using the concepts of entropy and information gain to determine the best division. At each step of tree formation, C4.5 selects the attributes that are most effective in separating the data into different classes, until all the data is well classified or no attributes remain. The advantage of C4.5 is its ability to handle missing data, work with continuous and discrete attributes, and produce trees that are easy to understand and interpret, making it one of the most influential and widely used algorithms in data analysis and machine learning.

METHOD

The Naive Bayes Classifier method and the C4.5 algorithm are two effective approaches used to classify graduating students and female students at SMA Negeri 1 Silangkitang. Naive Bayes Classifier uses probability principles and the assumption of independence between attributes to predict whether a student will graduate or not. By analyzing attendance scores, behavior scores, Final Semester Examination (UAS) scores, and report card scores, this method calculates the probability of each passing category and determines the results based on the highest probability. The advantage of Naive Bayes lies

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

in its simplicity and computational speed, so it can handle large datasets efficiently and produce accurate predictions even though the assumption of independence between attributes is often not fully met in practice.

Meanwhile, the C4.5 algorithm builds a decision tree based on the attributes that provide the most information about graduation classification. By calculating entropy and information gain, C4.5 selects the most effective attributes for separating data into different classes. Each branch of the decision tree represents a specific attribute value, and each leaf node represents the final decision about student graduation. C4.5 has the ability to handle missing data, work with continuous and discrete attributes, and produce models that are easy to interpret. By using these two methods comparatively, this research can provide in-depth insight into the factors that influence student graduation and provide a reliable predictive tool to help improve the assessment and decision-making process in these schools.

With these two methods, the Naive Bayes Classifier and the C4.5 Algorithm, the classification of student graduation at SMA Negeri 1 Silangkitang can be done more accurately and reliably. The Naive Bayes Classifier provides probability-based predictions by utilizing attendance, behavior, UAS scores, and report card grades to calculate the probability of graduation, while the C4.5 Algorithm builds a clear and easy to understand decision tree based on important attributes that influence graduation outcomes. This combination enables comprehensive analysis of student data, minimizes misclassification, and provides more precise results to determine which students are eligible to graduate. Thus, these two methods provide effective tools to support the assessment process in schools and increase fairness and accuracy in determining student graduation. But not only that, there are stages that the author will carry out to be able to carry out this research, the stages are as follows.

The C4.5 Algorithm formula

$$Entropy\ (S) = \sum -pi\ n\ i{=}1\ \log 2(pi)$$

Information:
S = Set of cases
n = Number of S partitions
pi = Proportion of Si to S

Naive Bayes Method Formula

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

Information:

X = Sample data that has an unknown class (label).
C = Hypothesis that X is a data class (label).
P(C) = Probability of hypothesis C.
P(X) = Probability of observed sample data (probability C).
P(X|C) = Probability based on the conditions in the hypothesis.

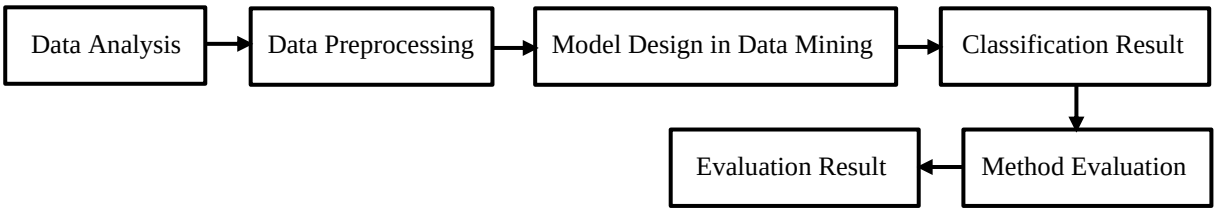

Fig 1. Classification Stages in Data Mining

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

An explanation of each stage is as follows.

Data analysis : The stages used to collect the data set that will be used in this research.
Data Preprocessing : The stages used to clean and select data that is suitable for use are arranged according to the needs of this research.
Model Design in Data Mining : The stages used to design the model that will be used for data classification.
Classification Results : The stage contains the results of the classification that has been carried out from the design model that has been created previously.
Method Evaluation : The stages used to evaluate the method are by designing an evaluation model.
Evaluation result : The stages contain evaluation results such as accuracy, precision and recall in this research.

For the C4.5 algorithm workflow and the Naive Bayes method, it works using a classification model. So the flow of this method goes through the stages of data collection which will be used for data analysis, then data preprocessing which is used to clean data that is not suitable for use. After that, create a classification model that is designed to be used as a classification tool. If so, the classification results will come out. Then create an evaluation model that is used to determine the results of the method's capabilities or what is often called accuracy.

RESULT

Data Analysis

In this stage, the author conducted research on the classification of graduating students at SMA Negeri 1 Silangkitang using the Naive Bayes Classifier method and the C4.5 algorithm. This research uses two sets of data, namely training data consisting of 30 student and female students' data, and testing data consisting of 54 student and female students' data. By using these two methods, it is hoped that it can provide accurate and reliable classification results to support the student graduation assessment process. Data was obtained by conducting field observations, namely at SMA Negeri 1 Silangkitang.

Data Training

Training data is training data that is used to help the data classification process in data mining. Training data will later help the data classification process.

Table 1. Data Training

Name	Gender	Presence	Behavior	UAS	Report card	Category
Abilqis Laela Rahmadana	Man	Baik	Good	Tall	Tall	Passed
Adam Mulyana Syaban	Man	Good	Good	Tall	Tall	Passed
Affrilia Ilham Islami	Man	Good	Good	Tall	Not Enough	Not pass
Ajeng Viviana Sari	Woman	Enough	Enough	Not Enough	Tall	Not pass
Alhairi	Man	Baik	Good	Currently	Tall	Passed
Amanda Rasya Putri	Woman	Enough	Good	Not Enough	Not Enough	Not pass
Amril Maiza	Man	Good	Good	Tall	Currently	Passed
Andika Fajar Ramadhan	Man	Good	Good	Tall	Tall	Passed
Ankobi Lanang Dmingge	Man	Good	Good	Tall	Tall	Passed

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13860
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Faisal Fadilah	Man	Enough	Enough	Not Enough	Not Enough	Not pass
Faneezha Syifa Azzahra	Woman	Enough	Enough	Not Enough	Not Enough	Not pass
Farah Aini Nurazizah	Woman	Good	Enough	Not Enough	Not Enough	Not pass
Hilda Elvina	Woman	Good	Good	Currently	Tall	Passed
Huda Nur Alif Aryansyah	Woman	Enough	Enough	Currently	Not Enough	Not pass
Ines Ayu Puji Lestari	Woman	Good	Good	Tall	Tall	Passed
Ivan Amjad Al Baihaqi	Man	Enough	Enough	Not Enough	Not Enough	Not pass
Khairunisa Fitriani	Man	Good	Good	Not Enough	Tall	Passed
Nayla Dwi Septiani	Woman	Baik	Enough	Not Enough	Not Enough	Not pass
Nesa Nur Aryda	Woman	Good	Good	Tall	Tall	Passed
Nita Aulia	Woman	Good	Good	Tall	Currently	Passed
Pandu Zildjian Oktafianto	Man	Enough	Enough	Not Enough	Currently	Not pass
Putri Puspaningtyas	Woman	Enough	Enough	Tall	Not Enough	Not pass
Putry Oktavia Romadona	Woman	Enough	Good	Not Enough	Not Enough	Not pass
Ravi Azmi Maskur	Man	Good	Good	Tall	Tall	Passed
Salsabila Assyifa	Woman	Enough	Enough	Not Enough	Not Enough	Not pass
Sanju Kurniyansyah	Man	Good	Good	Not Enough	Not Enough	Not pass
Sazkya Dwi Nurhaliza	Woman	Good	Good	Tall	Tall	Passed
Tika Yulianingsih	Woman	Enough	Enough	Not Enough	Not Enough	Not pass
Tony Aliyana	Man	Enough	Good	Tall	Currently	Passed
Wulan Rahmadani	Woman	Enough	Good	Currently	Tall	Passed

In the table above is the training data used for 30 students and female students. The data above was obtained by observing the Silangkitang 1 Public High School. The data above will later become training data which will help the classification process.

Data Testing

Testing data is data that will be classified in data mining using the Naïve Bayes Classifier method and the C4.5 algorithm. Testing data is also often referred to as sample data that will be processed and obtained results for students who pass or fail.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

Table 2. Data Testing

Name	Gender	Presence	Behavior	UAS	Report card
Adelia Arsita	Woman	Good	Good	Tall	Currently
Alan Mulyadi	Man	Enough	Enough	Tall	Not Enough
Alvira Agustiana	Woman	Good	Good	Tall	Not Enough
Amanda	Woman	Enough	Good	Not Enough	Not Enough
Aprilia Wahyuningsih	Woman	Enough	Enough	Not Enough	Not Enough
Arif Hidayat	Man	Enough	Enough	Not Enough	Currently
Azra Aurel Mareza	Woman	Good	Enough	Tall	Tall
Bachtiar Ghoniyya	Man	Enough	Baik	Not Enough	Not Enough
Berliana Syifaul Ummah	Woman	Good	Enough	Not Enough	Not Enough
Bredly Freesq Zullio	Man	Good	Good	Tall	Tall
Camelia Aulia Rahma	Woman	Good	Baik	Tall	Tall
Desi Revana	Woman	Good	Good	Not Enough	Tall
Deva Nurmala	Woman	Good	Good	Currently	Tall
Devi Anggreini Putri	Woman	Good	Enough	Tall	Tall
Dimas Aprianto	Man	Good	Baik	Not Enough	Tall
Diska Selfiana	Woman	Enough	Good	Tall	Tall
Dwi Raditya Siregar	Woman	Good	Baik	Not Enough	Tall
Elsa Zahra	Woman	Good	Good	Tall	Tall
Evi Nurfiana	Woman	Good	Baik	Tall	Tall
Gita Afrianti	Woman	Enough	Enough	Currently	Not Enough
Haliza Putri Nurul Ulfah	Woman	Enough	Baik	Not Enough	Not Enough
Handoko Wibowo	Man	Good	Good	Tall	Tall
Iqbal Pratama Putra	Man	Good	Good	Not Enough	Tall
Jerri Hikmanah Sitompul	Man	Enough	Enough	Not Enough	Tall
Laudy Santi Anjani	Woman	Good	Enough	Not Enough	Not Enough
Maulidia Nurul Khairani	Woman	Good	Good	Tall	Tall
Mhd Rizky Aidil Ritonga	Man	Good	Good	Tall	Not Enough
Milda Suci Wahyuni	Woman	Enough	Enough	Not Enough	Not Enough
Muhammad Abdul Rochim	Man	Enough	Good	Tall	Tall

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

Muhammad Akbar	Man	Good	Good	Not Enough	Tall
Muhammad Arief	Man	Good	Good	Tall	Tall
Muhammad Azkia	Man	Good	Good	Not Enough	Tall
Muhammad Gajali	Man	Good	Good	Tall	Currently
Muhammad Iqbal Alfaruq	Man	Good	Good	Tall	Not Enough
Muhammad Misbahul Khair	Man	Good	Good	Tall	Not Enough
Muhammad Nur Ikhsan	Man	Good	Good	Tall	Tall
Muhammad Rapi'i	Man	Good	Baik	Not Enough	Tall
Muhammad Rizki	Man	Good	Good	Tall	Tall
Muhammad Syahid Fadillah	Man	Good	Enough	Tall	Tall
Muhammad Zairu Irfan	Man	Good	Good	Not Enough	Not Enough
Nabila Syadewi	Woman	Enough	Enough	Not Enough	Currently
Nadya Bella Ramadany	Woman	Baik	Good	Not Enough	Not Enough
Nurmala Cantika Dewi	Woman	Good	Good	Currently	Tall
Pikri Rahmadan	Man	Good	Good	Currently	Currently
Puput Hartini	Woman	Good	Enough	Not Enough	Not Enough
Rizqi Purnama	Man	Good	Baik	Tall	Tall
Rojana	Woman	Baik	Good	Tall	Tall
Roxana Stefani Ritonga	Woman	Enough	Enough	Not Enough	Not Enough
Siti Ramadani	Woman	Good	Good	Tall	Tall
Syuaaip Batu Bara	Man	Enough	Enough	Not Enough	Currently
Tiara	Woman	Good	Good	Tall	Tall
Venny Ade Mulia	Woman	Baik	Good	Currently	Tall
Wayang Dharmayasa	Man	Good	Good	Tall	Tall
Yasmine Athaullah	Woman	Enough	Good	Tall	Tall

The table above is research sample data that has obtained 54 research sample data. With sample data, research can be carried out and later the method used can be applied.

Data Preprocessing

This stage is a process carried out to clean data that is not suitable for use or select data that is suitable for use, then at this stage the data will also be arranged based on the format that has been determined in this research.

Model Design in Data Mining

At this stage, a design model is used to classify data using the Naïve Bayes Classifier method and the C4.5 algorithm. With this model, it will be easy to classify data in data mining.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

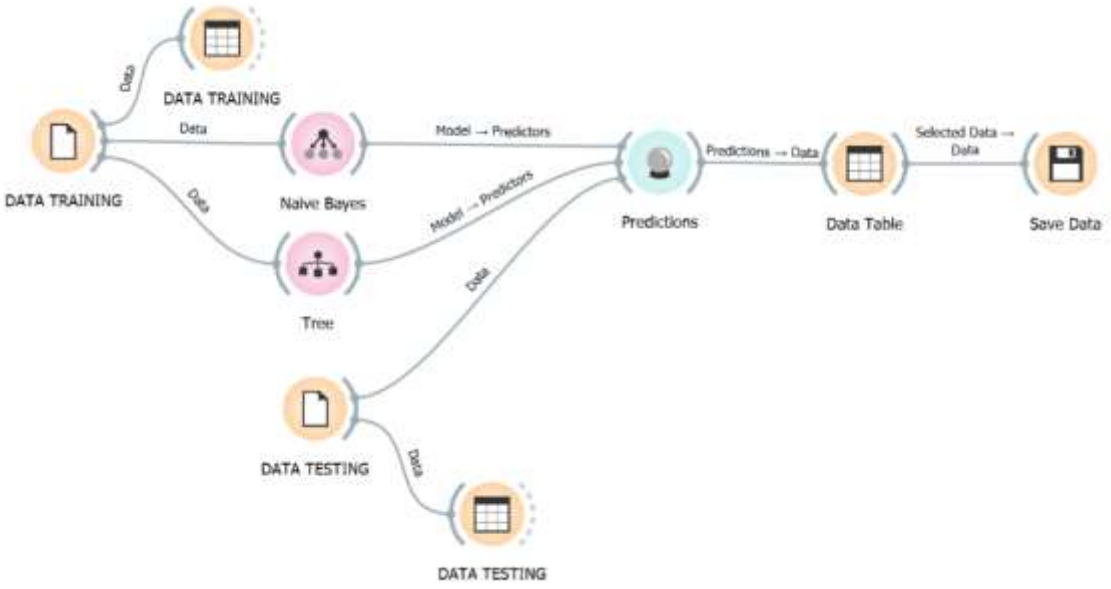

Fig 2. Planning Model on Data Mining

In the image above is the model used to classify data. The model above was designed using the orange application. This design will later provide classification results for the research sample data used. The method used can be seen in the widget in the red box. This widget is a method that will be used to classify data in data mining. The methods are the Naïve Bayes Classifier method and the C4.5 Algorithm.

Classification Results

At this stage, the results of the classification have been carried out using a design model in data mining with the design carried out in the orange application. The classification results can be seen in the image below.

Table 3. Classification Results

Name	Gender	Presence	Behavior	UAS	Report card	Naïve Bayes	Algoritma C4.5
Adelia Arsita	Woman	Good	Good	Tall	Currently	Passed	Passed
Alan Mulyadi	Man	Enough	Enough	Tall	Not Enough	Not pass	Not pass
Alvira Agustiana	Woman	Good	Good	Tall	Not Enough	Passed	Not pass
Amanda	Woman	Enough	Good	Not Enough	Not Enough	Not pass	Not pass
Aprilia Wahyuningsih	Woman	Enough	Enough	Not Enough	Not Enough	Not pass	Not pass
Arif Hidayat	Man	Enough	Enough	Not Enough	Currently	Not pass	Not pass
Azra Aurel Mareza	Woman	Good	Enough	Tall	Tall	Passed	Not pass
Bachtiar Ghoniyya	Man	Enough	Baik	Not Enough	Not Enough	Not pass	Not pass
Berliana Syifaul Ummah	Woman	Good	Enough	Not Enough	Not Enough	Not pass	Not pass
Bredly Freesq Zullio	Man	Good	Good	Tall	Tall	Passed	Passed
Camelia Aulia Rahma	Woman	Good	Baik	Tall	Tall	Passed	Passed

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

Desi Revana	Woman	Good	Good	Not Enough	Tall	Passed	Passed
Deva Nurmala	Woman	Good	Good	Currently	Tall	Passed	Passed
Devi Anggreini Putri	Woman	Good	Enough	Tall	Tall	Passed	Not pass
Dimas Aprianto	Man	Good	Baik	Not Enough	Tall	Passed	Passed
Diska Selfiana	Woman	Enough	Good	Tall	Tall	Passed	Passed
Dwi Raditya Siregar	Woman	Good	Baik	Not Enough	Tall	Passed	Passed
Elsa Zahra	Woman	Good	Good	Tall	Tall	Passed	Passed
Evi Nurfiana	Woman	Good	Baik	Tall	Tall	Passed	Passed
Gita Afrianti	Woman	Enough	Enough	Currently	Not Enough	Not pass	Not pass
Haliza Putri Nurul Ulfah	Woman	Enough	Baik	Not Enough	Not Enough	Not pass	Not pass
Handoko Wibowo	Man	Good	Good	Tall	Tall	Passed	Passed
Iqbal Pratama Putra	Man	Good	Good	Not Enough	Tall	Passed	Passed
Jerri Hikmanah Sitompul	Man	Enough	Enough	Not Enough	Tall	Not pass	Not pass
Laudy Santi Anjani	Woman	Good	Enough	Not Enough	Not Enough	Not pass	Not pass
Maulidia Nurul Khairani	Woman	Good	Good	Tall	Tall	Passed	Passed
Mhd Rizky Aidil Ritonga	Man	Good	Good	Tall	Not Enough	Passed	Not pass
Milda Suci Wahyuni	Woman	Enough	Enough	Not Enough	Not Enough	Not pass	Not pass
Muhammad Abdul Rochim	Man	Enough	Good	Tall	Tall	Passed	Passed
Muhammad Akbar	Man	Good	Good	Not Enough	Tall	Passed	Passed
Muhammad Arief	Man	Good	Good	Tall	Tall	Passed	Passed
Muhammad Azkia	Man	Good	Good	Not Enough	Tall	Passed	Passed
Muhammad Gajali	Man	Good	Good	Tall	Currently	Passed	Passed
Muhammad Iqbal Alfaruq	Man	Good	Good	Tall	Not Enough	Passed	Not pass
Muhammad Misbahul Khair	Man	Good	Good	Tall	Not Enough	Passed	Not pass
Muhammad Nur Ikhsan	Man	Good	Good	Tall	Tall	Passed	Passed
Muhammad Rapi'i	Man	Good	Baik	Not Enough	Tall	Passed	Passed
Muhammad Rizki	Man	Good	Good	Tall	Tall	Passed	Passed
Muhammad Syahid Fadillah	Man	Good	Enough	Tall	Tall	Passed	Not pass
Muhammad Zairu Irfan	Man	Good	Good	Not Enough	Not Enough	Not pass	Not pass
Nabila Syadewi	Woman	Enough	Enough	Not Enough	Currently	Not pass	Not pass

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

Nadya Bella Ramadany	Woman	Baik	Good	Not Enough	Not Enough	Not pass	Not pass
Nurmala Cantika Dewi	Woman	Good	Good	Currently	Tall	Passed	Passed
Pikri Rahmadan	Man	Good	Good	Currently	Currently	Passed	Passed
Puput Hartini	Woman	Good	Enough	Not Enough	Not Enough	Not pass	Not pass
Rizqi Purnama	Man	Good	Baik	Tall	Tall	Passed	Passed
Rojana	Woman	Baik	Good	Tall	Tall	Passed	Passed
Roxana Stefani Ritonga	Woman	Enough	Enough	Not Enough	Not Enough	Not pass	Not pass
Siti Ramadani	Woman	Good	Good	Tall	Tall	Passed	Passed
Syuaaip Batu Bara	Man	Enough	Enough	Not Enough	Currently	Not pass	Not pass
Tiara	Woman	Good	Good	Tall	Tall	Passed	Passed
Venny Ade Mulia	Woman	Baik	Good	Currently	Tall	Passed	Passed
Wayang Dharmayasa	Man	Good	Good	Tall	Tall	Passed	Passed
Yasmine Athaullah	Woman	Enough	Good	Tall	Tall	Passed	Passed

The classification results were obtained using the naïve Bayes classifier method and the c4.5 algorithm in data mining. However, the classification results obtained from the 2 methods used have different results. For classification results using the Naïve Bayes Classifier method, the results obtained were 37 data of students and female students who had graduated from school and the remaining data were 17 data of students and female students who had not graduated from school, while the results obtained by the C4.5 Algorithm were 30 data of students and female students who had graduated from school and the remaining data was 24 data on students and girls who did not graduate from school. From the two classification results, the methods used have quite different results.

Method Evaluation

At the method evaluation stage, the aim is to obtain results from the ability or performance of the method in carrying out classification. So to be able to find out the reasons for the differences in classification results from the two methods used, namely by carrying out an evaluation to determine how capable each method is of classifying data. So at the method evaluation stage the author also uses a design model that will be used to carry out the evaluation.

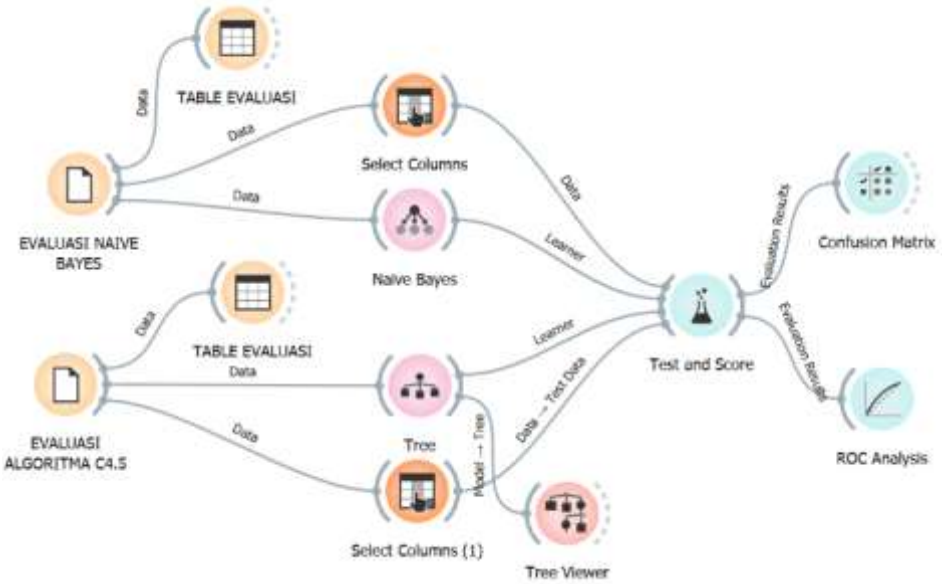

Fig 3. Method Evaluation Model

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13860
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The design model above is the design model that will be used to evaluate the method used in this research. With this model, it will be easy for the author to obtain method evaluation results. The files used in this evaluation are the same, but only there are attributes that are used in the Naïve Bayes method, but are not used in the C4.5 algorithm and vice versa.

Evaluation Result

For the evaluation results that will be carried out the author uses 2 widgets that can support and obtain evaluation results of the method used, for the first widget the author uses the test and score widget where the widget will provide evaluation results in the form of method accuracy and for the other widget, namely the Confusion Matrix which is an evaluation widget that also provides method accuracy results. There is 1 more widget, but it is not used to provide evaluation results, but rather provides graphic results from the two methods used.

Test and Score

The evaluation results from the test and score widget provide results that can be used to explain how suitable and good the method's ability to classify data is. The results will be presented in table form and later the results of the test and score widget will be explained.

Table 4. Result of Test and Score

Model	AUC	CA	F1	Precision	Recall
Naïve Bayes Classifier	1.000	1.000	1.000	1.000	1.000
Algoritma C4.5	0.893	0.926	0.925	0.926	0.926

In the table above are the results of method evaluation that has been carried out using a design model that has been created previously. From the evaluation results in the table above, the Naïve Bayes Classifier method obtained perfect results, namely 100%, while the evaluation results obtained by the C4.5 Algorithm were 89%. The evaluation results obtained have quite large differences. This explains that the naïve Bayes classifier method is superior and more suitable for classification in this research than the c4.5 algorithm. But this is only the result of 1 widget, there is still 1 more widget, namely the confusion matrix which will provide the evaluation results.

Confusion Matrix

Confusion Matrix Results of the Naïve Bayes Classifier Method

Table 5
Confusion Matrix Results

		Predicted		
		Interest	Not Interested	∑
Actual	Interest	37	0	37
	Not Interested	0	17	17
	∑	37	17	54

The Confusion Matrix results can be seen in the table above. For the results, there are True Positive (TP) results which are 37 data, for True Negative (TN) results which are 9 data, for False Positive (FP) results which are 0 and for False Negative (FN) results which are 0 data. For these results, it is not possible to directly measure the accuracy value, the data above must be calculated first using the formula in the confusion matrix, which is as follows.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

$$\boldsymbol{Accuracy} = \frac{37+17}{37+17+0+0} + \mathbf{100\%}$$ Then the Accuracy value = 100%

$$\boldsymbol{Presisi} = \frac{37}{37+0} + \mathbf{100\%}$$ Then the Precision value = 100%

$$\boldsymbol{Recall} = \frac{37}{37+0} + \mathbf{100\%}$$ Then the Recall value = 100%

Confusion Matrix Algorithm C4.5 Results

Table 6
Confusion Matrix Results

		Predicted		
		Interest	Not Interested	∑
Actual	Interest	36	1	148
	Not Interested	3	14	2
	∑	147	3	150

The Confusion Matrix results can be seen in the table above. For the results, there are True Positive (TP) results which are 36 data, for True Negative (TN) results which are 14 data, for False Positive (FP) results which are 1 and for False Negative (FN) results which are 3 data. For these results, it is not possible to directly measure the accuracy value, the data above must be calculated first using the formula in the confusion matrix, which is as follows.

$$\boldsymbol{Accuracy} = \frac{36+14}{36+14+1+3} + \mathbf{100\%}$$ Then the Accuracy value = 92%

$$\boldsymbol{Presisi} = \frac{36}{36+1} + \mathbf{100\%}$$ Then the Precision value = 97%

$$\boldsymbol{Recall} = \frac{36}{36+3} + \mathbf{100\%}$$ Then the Recall value = 92%

Tree Viewer Algorithm C4.5

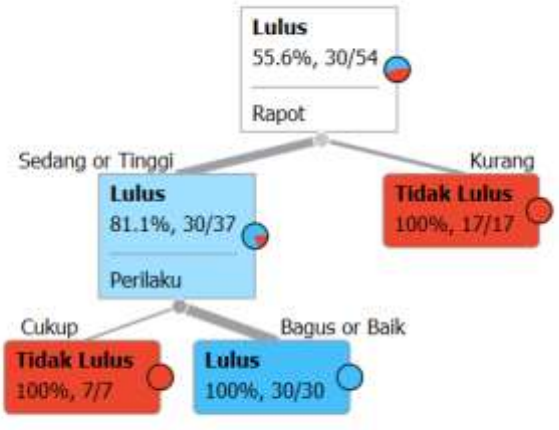

Fig 4. Results of Tree Viewer Algorithm C4.5

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13860 p-ISSN : 2541-044X

In the image above are the results of the c4.5 algorithm tree viewer which is the result of evaluating the c4.5 algorithm decision tree. There are percentage results for each attribute given from the tree viewer widget.

ROC Analysis

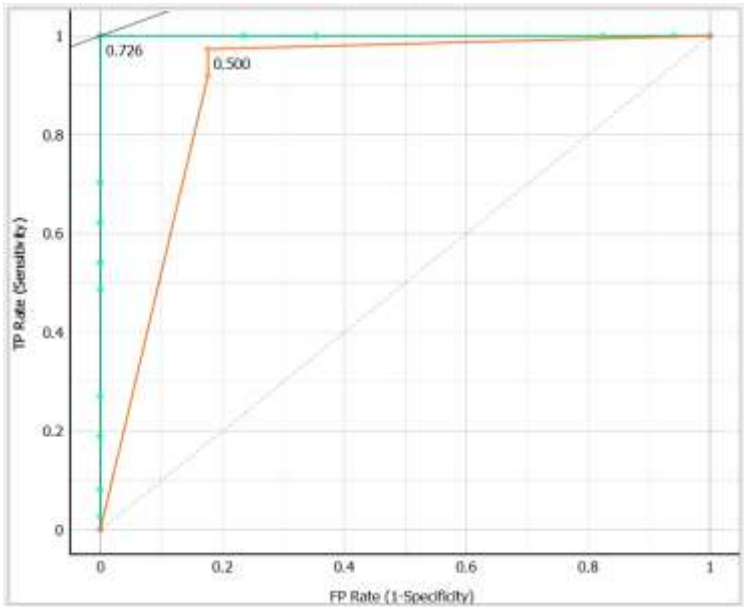

Fig 5. ROC Analysis of Students and Girls Who Graduated from School

From the picture above is the graphic result of students and girls graduating from school. The graph in green is the graph result from the naïve Bayes classifier method and the graph in orange is the graph result from the c4.5 algorithm. The results obtained state that the graphic results of the naïve Bayes classifier method are better than the results of the c4.5 algorithm. This is because the results obtained by the naïve Bayes classifier method are higher and almost reach 100%, while the c4.5 algorithm is only around 50%. The difference obtained from the two methods is quite large. These results are the results of students who graduated from school.

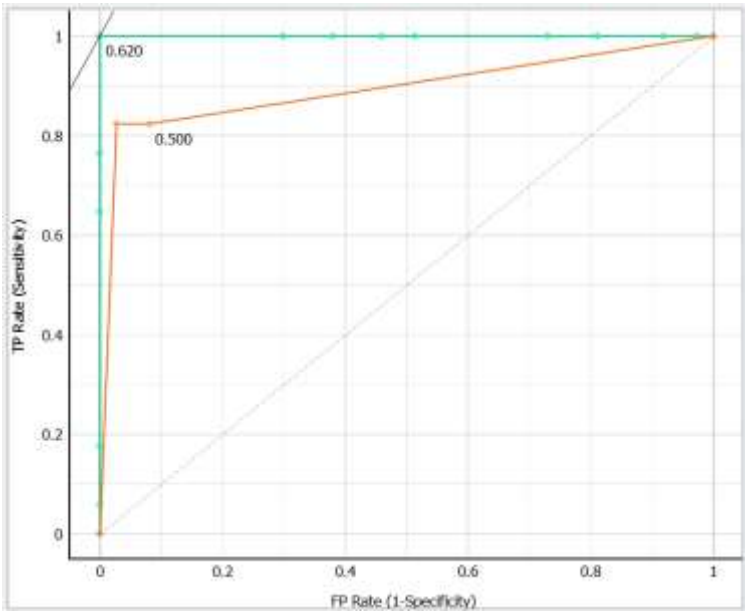

Fig 6. ROC Analysis of Students and Girls Who Have Not Graduated from School

*name of corresponding author

The graphic results for students and girls who did not graduate from school also had quite different results and the results obtained still show that the naïve Bayes classifier method is superior to the c4.5 algorithm in this study.

DISCUSSIONS

Research conducted to determine student graduation at SMA Negeri 1 Silangkitang used two classification methods, namely the Naive Bayes Classifier and the C4.5 Algorithm. This research aims to evaluate the performance of the two methods in predicting student graduation based on attributes such as attendance scores, behavior scores, Final Semester Examination (UAS) scores, and report card scores. The data used consists of two sets, namely training data which contains 30 student data and female students, and testing data which contains 54 student data. The results of this study show that the two methods produce different classification results. The Naive Bayes Classifier method produces predictions that 37 data of students and female students graduated from school, while 17 data of other students and female students did not graduate. On the other hand, the C4.5 Algorithm predicts that 30 data of students and female students will graduate from school, while 24 data of students and female students will not graduate. The difference in results between these two methods shows a significant difference in their ability to classify student graduation data.

In terms of performance evaluation, the Naive Bayes Classifier method shows excellent results with an accuracy rate of 100%, which means all predictions made by this method are correct. In contrast, the C4.5 Algorithm has an accuracy rate of 89%, which, although still quite high, indicates that there are some inaccurate predictions. The quite large difference in the results of this evaluation indicates that the Naive Bayes Classifier method is superior in classifying student graduation in this research compared to the C4.5 Algorithm. Comparison of evaluation results shows that the Naive Bayes Classifier is more effective and more precise in predicting student graduation at SMA Negeri 1 Silangkitang. With an accuracy level of 100%, this method is proven to be more reliable than the C4.5 Algorithm which has an accuracy of 89%. Therefore, in the context of this research, the Naive Bayes Classifier can be considered a superior method for student graduation classification.

CONCLUSION

This research at SMA Negeri 1 Silangkitang aimed to determine student graduation using two classification methods: Naive Bayes Classifier and the C4.5 Algorithm. Utilizing training data of 30 students and testing data of 54 students, the study evaluated their performance in predicting graduation based on attributes such as attendance, behavior, UAS scores, and report card scores. Naive Bayes predicted 37 students would graduate and 17 would not, while C4.5 predicted 30 would graduate and 24 would not, highlighting significant differences in their classification approaches. Naive Bayes tends to provide more positive graduation predictions, whereas C4.5 is more conservative. These findings suggest that schools should carefully consider which method aligns better with their educational evaluation needs.

REFERENCES

Abas, M. I., Ibrahim, I., Syahrial, S., Lamusu, R., Baderan, U. S., & Kango, R. (2023). Analysis of Covid-19 Growth Trends Through Data Mining Approach As Decision Support. *Sinkron*, *8*(1), 101–108. https://doi.org/10.33395/sinkron.v8i1.11861

Aji, G. W., & Devi, P. A. R. (2023). Data Mining Implementation For Product Transaction Patterns Using Apriori Method. *Sinkron*, *8*(1), 421–432. https://doi.org/10.33395/sinkron.v8i1.12071

Alam, A., Alana, D. A. F., & Juliane, C. (2023). Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes. *Sinkron*, *8*(4), 2641–2650. https://doi.org/10.33395/sinkron.v8i4.12998

Almufqi, F. M., & Voutama, A. (2023). Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa. *Jurnal Teknika*, *15*(1), 61–66. https://doi.org/10.30736/jt.v15i1.929

*name of corresponding author

Amalia, R. (2020). Penerapan Data Mining untuk Memprediksi Hasil KePassedan Siswa Menggunakan Metode Naïve Bayes. *Juisi*, *06*(01), 33–42.

Anam, M. K., Rahmiati, R., Paradila, D., Mardainis, M., & Machdalena, M. (2023). Application of Naïve Bayes Algorithm for Non-Cash Food Assistance Recipients in Kampar Regency. *Sinkron*, *8*(1), 433–441. https://doi.org/10.33395/sinkron.v8i1.12032

Apriyani, M. E., Maskuri, R. A., Ratsanjani, M. H., Pramudhita, A. N., & Rawansyah, R. (2023). Digital Forensic Investigates Sexual Harassment on Telegram using Naïve Bayes. *Sinkron*, *8*(3), 1409–1417. https://doi.org/10.33395/sinkron.v8i3.12514

Bustomi, Y., Nugraha, A., Juliane, C., & Rahayu, S. (2023). Data Mining Selection of Prospective Government Employees with Employment Agreements using Naive Bayes Classifier. *Sinkron*, *8*(1), 1–8. https://doi.org/10.33395/sinkron.v8i1.11968

Hasibuan, F. F., Dar, M. H., & Yanris, G. J. (2023). Implementation of the Naïve Bayes Method to determine the Level of Consumer Satisfaction. *SinkrOn*, *8*(2), 1000–1011. https://doi.org/10.33395/sinkron.v8i2.12349

Hasibuan, S. A., Sihombing, V., & Nasution, F. A. (2023). Analysis of Community Satisfaction Levels using the Neural Network Method in Data Mining. *Sinkron*, *8*(3), 1724–1735. https://doi.org/10.33395/sinkron.v8i3.12634

Lizsara, P. A., Oyama, S., & Wardani, S. (2020). Implementasi Data Mining Menggunakan Metode Naïve Bayes Untuk Memprediksi Ketepatan Waktu Tingkat KePassedan Mahasiswa (Study Kasus: Program Studi Informatika Universitas PGRI Yogyakarta). *Seri Prosiding Seminar Nasional Dinamika Informatika*, *4*(1), 34–37. Retrieved from http://prosiding.senadi.upy.ac.id/index.php/senadi/article/view/121

Madjid, F. M., Ratnawati, D. E., & Rahayudi, B. (2023). Sentiment Analysis on App Reviews Using Support Vector Machine and Naïve Bayes Classification. *Jurnal Dan Penelitian Teknik Informatika*, *8*(1), 556–562. Retrieved from https://doi.org/10.33395/sinkron.v8i1.12161

Maizura, S., Sihombing, V., & Dar, M. H. (2023). Analysis of the Decision Tree Method for Determining Interest in Prospective Student College. *SinkrOn*, *8*(2), 956–979. https://doi.org/10.33395/sinkron.v8i2.12258

Mawaddah, A., Dar, M. H., & Yanris, G. J. (2023). Analysis of the SVM Method to Determine the Level of Online Shopping Satisfaction in the Community. *SinkrOn*, *8*(2), 838–855. https://doi.org/10.33395/sinkron.v8i2.12261

Nurjana, A., Perdana Windarto, A., Qurniawan, H., & Tunas Bangsa, S. (2022). Implementasi Data Mining Dalam Memprediksi Prestasi Siswa Dengan Algoritma C4.5. *SmartEDU*, *1*(4), 171–180. Retrieved from https://ejournal.abivasi.id/index.php/SmartEDU

Punkastyo, D. A., Septian, F., & Syaripudin, A. (2024). Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi KePassedan Siswa. *Journal of System and Computer Engineering (JSCE)*, *5*(1), 24–35. https://doi.org/10.61628/jsce.v5i1.1073

Rahman, R., & Fauzi Abdulloh, F. (2023). Performance of Various Naïve Bayes Using GridSearch Approach In Phishing Email Dataset. *Sinkron*, *8*(4), 2336–2344. https://doi.org/10.33395/sinkron.v8i4.12958

*name of corresponding author

Saputra, A. D. S., Hindarto, D., & Haryono, H. (2023). Supervised Learning from Data Mining on Process Data Loggers on Micro-Controllers. *Sinkron*, *8*(1), 157–165. https://doi.org/10.33395/sinkron.v8i1.11942

Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, *3*(3), 424–439. https://doi.org/10.51519/journalisi.v3i3.164

Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring. *Jurteksi*, *4*(1), 85–92. https://doi.org/10.33330/jurteksi.v4i1.28

Siregar, A. P., Irmayani, D., & Sari, M. N. (2023). Analysis of the Naïve Bayes Method for Determining Social Assistance Eligibility Public. *SinkrOn*, *8*(2), 805–817. https://doi.org/10.33395/sinkron.v8i2.12259

Susanto, H., & Sudiyatno, S. (2014). Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Pendidikan Vokasi*, *4*(2), 222–231. https://doi.org/10.21831/jpv.v4i2.2547

*name of corresponding author