Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Analysis of Student Excellence Classes in Data Mining Using the KNN Method

Arvida Ritonga[1)*], Masrizal[2)], Irmayanti[3)]
[1)2)3)]Universitas Labuhanbatu, Indonesia
[1)]arvidaritonga02@gmail.com, [2)]masrizal120405@gmail.com, [3)]irmayantiritonga2@gmail.com

Submitted :April 5, 2024 | **Accepted** : April 20, 2024 | **Published** : April 22, 2024

Abstract: Excellent classes are programs designed to maximize the academic and non-academic potential of students and girls, with the aim of improving their overall achievement. This program aims to provide more intensive learning and a curriculum tailored to students' needs and abilities, so that they can develop their talents and competencies optimally. In order to evaluate the effectiveness of the superior class program and to identify students who are most suitable for the program, this research was conducted using the K-Nearest Neighbors (KNN) method in data mining. The research process includes several critical stages, namely determining relevant data, designing a machine learning model, testing the model to ensure its effectiveness, and evaluating the model to assess the accuracy and reliability of the results. This research used sample data consisting of 92 male and female students, where the results of the analysis showed that 42 of them met the criteria to enter the superior class, while 50 other students did not. These criteria are determined based on various factors, including academic achievement, participation in extracurricular activities, and other individual characteristics assessed through the KNN method. The accuracy results obtained from the model evaluation show excellent performance, confirming that the approach used is effective in classifying students based on their potential to excel in superior class programs. The conclusion of this research shows that the use of the KNN method in data mining can accurately identify students who will benefit most from superior class programs. Thus, this approach offers a valuable tool for educational institutions to optimize student potential and raise overall standards of achievement.

Keywords: Classification; Confusion Matrix; Data Mining; K-Nearest Neighbor (kNN); Superior;

INTRODUCTION

Student excellence classes are special education programs designed to accommodate and develop the maximum potential of students who demonstrate extraordinary achievements in various fields, both academic and non-academic. These programs typically offer a more challenging curriculum and broader learning opportunities compared to regular classes, including extracurricular activities specifically designed to enhance students' skills and talents. The main goal of superior classes is to equip students with the knowledge, skills, and experience necessary to achieve success at a higher level, both in further education and in the world of work. Through innovative educational approaches and a supportive learning environment, superior classes strive to encourage students to reach their full potential. The superior class program has a noble goal of developing the potential of outstanding students and students. However, in its implementation, errors often arise in the selection process, which results in inappropriate student placement.

Mistakes in the selection and placement of students in superior class programs give rise to various other problems, such as injustice and gaps in motivation among students. This situation can affect the dynamics of learning at school, where students who should have the opportunity to develop in a more challenging environment feel disappointed and underappreciated. Meanwhile, students who do not fully meet the criteria feel pressured to meet the high expectations set. This psychological impact can lead to mental stress for students, reduced self-confidence, and perhaps even reduce their enthusiasm for learning. Therefore, it is very important for educational institutions to revise and improve their selection systems, ensuring that the process of determining students for superior classes is carried out fairly and transparently, based on clear and objective criteria. Based on the case study described previously regarding errors in determining students who are worthy of entering the superior class, the author feels compelled to conduct research which aims to analyze and determine which students are worthy of entering the superior class more accurately. This research will utilize the K-Nearest Neighbor (KNN) method,

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

which is one of the techniques in Data Mining. The KNN method was chosen because of its good ability to classify data based on the closeness of the characteristics of the data to samples whose class is known. Thus, it is hoped that this research can produce an objective and accurate model to determine which students have the right to enter superior classes.

This research will be carried out at Madrasah Tsanawiyah (MTs) Guppi Tanjung Harapan, which is one of the educational institutions that has superior class programs and faces challenges in the student selection process. MTs Guppi Tanjung Harapan was chosen as the research location because it is representative and has sufficient data for analysis. The research will involve all class VII students as samples, considering that they are the group that will be observed and analyzed to determine their eligibility to enter the superior class based on predetermined criteria. The data that will be used includes academic grades, non-academic achievements, and other relevant characteristics, which will then be processed using the KNN algorithm to predict their eligibility to enter superior classes. It is hoped that the use of the KNN method in this research will not only increase objectivity and accuracy in determining students for superior classes, but also provide new insights to schools regarding the potential for applying Data Mining techniques in the educational decision making process. It is hoped that the results of this research can be used as a reference for MTs Guppi Tanjung Harapan in improving and redesigning their superior class selection system, so that in the future the determination of students who are worthy of entering superior classes can be done more fairly, objectively and transparently.

The data that will be used in this research includes three main aspects that determine the suitability of students to enter superior classes, namely attitude values, knowledge values and skills values. Attitude values describe student behavior and ethics in the learning process, reflecting non-academic aspects that are important for character development. Knowledge scores measure students' mastery of subject matter, showing how far they understand basic concepts in various fields of study. Meanwhile, skills scores assess students' practical abilities, such as the application of knowledge in real situations, creativity, and problem-solving abilities. Plus 1 more attribute, namely the attendance score, this is a form of attitude or behavior of students who comply with school rules. So, when a student is not present, they can take permission or do other things according to the environment and existing school rules. These four types of data will be used as independent variables in the K-Nearest Neighbor (KNN) model to analyze and determine which students are most worthy of entering the superior class based on their overall performance.

LITERATURE REVIEW

The K-Nearest Neighbor (KNN) method is a simple but very effective machine learning algorithm, used in classification and regression (Zai, Sirait, Nainggolan, Sihombing, & Banjarnahor, 2023) (Istiadi, Rahman, & Wisnu, 2023). This algorithm works by predicting new data labels based on similarities with already known data (Adjani, Fauzia, & Juliane, 2023) (Arifuddin, Pinastawa, Anugraha, & Pradana, 2023) (Violita, Yanris, & Hasibuan, 2023). In principle, KNN identifies the 'k' nearest neighbors of the data points to be classified, based on a certain distance metric, such as Euclidean distance. The labels of these 'k' nearest neighbors are then used to determine the most likely label for the unknown data point. Uniquely, KNN does not require an explicit learning phase, which makes it fall into the 'lazy learning' algorithm category. This means that the algorithm does not create a general model from the training data, but 'decides' based on a subset of the data that is relevant at the time a prediction is needed, making it very flexible in dealing with data changes. In research conducted by (Munazhif, Yanris, Nirmala, & Hasibuan, 2023) the knn method can be used to conduct research to determine students' superior classes. In research conducted by (Irmayani & Sinaga, 2023) the knn method can be used to classify data to determine the level of community satisfaction and also in research by (Pratama, Yanris, Nirmala, & Hasibuan, 2023) that the knn method can be used to conduct research on satisfaction levels as well. The knn method was also used in research conducted by (Triani, Dar, & Yanris, 2023) that the knn method can be used to determine the level of community satisfaction and in research conducted by (Violita et al., 2023) also used the knn method for level classification customer satisfaction. The average accuracy obtained from this research is also very good, because the average accuracy results obtained are more than 90%.

The main advantages of KNN are its ease of implementation and effectiveness in various situations, especially on data that is intrinsically divided into small groups (Pratama et al., 2023). However, this algorithm also has some advantages, such as sensitivity to feature scale and high data dimensionality, which can cause performance degradation due to the “curse of dimensionality”. In addition, choosing the optimal value of 'k' is crucial, because a value that is too small can make the model too sensitive to noise, while a value of 'k' that is too large can blur the boundaries between classes. Nonetheless, with proper parameter selection and careful data pre-processing, KNN can be a very powerful tool in data analysis and machine learning.

Data mining is an analytical process that aims to explore and analyze large amounts of data to discover significant patterns, trends and relationships that may be hidden in the dataset (Aji & Devi, 2023). This technique combines statistical methods, machine learning, and data analysis to transform raw data into useful information that can support decision making (Saputra, Hindarto, & Santoso, 2023). In practice, data mining is used in various fields such as marketing, medical research, computer science, and finance, to help organizations identify new opportunities, increase operational efficiency, and minimize risks (Sinaga, Marpaung, Tarigan, & Tania, 2023).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

By analyzing datasets from various sources, data mining allows companies to gain deep insights into customer behavior, market trends, and consumption patterns, so they can create more informed and targeted business strategies.

METHOD

The K-Nearest Neighbor (KNN) method has been proven to be a very effective tool in data mining analysis, especially in identifying superior classes of students (Hermawan & Prianggono, 2023). By relying on the similarity principle, KNN allows educators and researchers to classify students into classes based on their characteristics and achievements that are similar to previously defined 'superior' students (Triani et al., 2023). This algorithm measures the distance between each student in the dataset to other students, selecting the closest 'k' students as a reference to determine whether a student is included in the superior category (A. W. Sari, Hermanto, & Defriani, 2023). The advantage of using KNN in this context is its ability to adapt classification to different types of data and class structures, providing more dynamic and adaptive insights into students' academic potential. In this way, KNN supports more informed and personalized educational decision making, assisting educational institutions in designing more effective learning programs to improve student achievement. In implementing this research, there are implementation stages that will be carried out. The implementation stages are as follows.

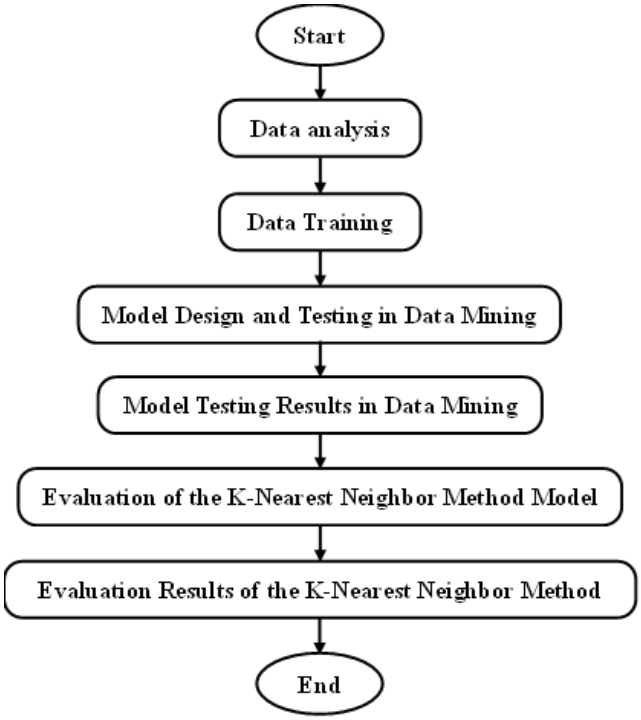

Fig 1. Research Framework on Data Mining

In table 1 above, the initial stage is data analysis, which is the stage carried out to determine the data that will be used. As was done in research(M. Sari, Yanris, & Hasibuan, 2023), data analysis is very necessary in research, namely the aim is to obtain data that suits the needs of the research being carried out. Then training data is data that is used to assist the data processing process in data mining using the knn method. This stage is a very important stage and is also explained in the research conducted (Violita et al., 2023) that training data is used to help the data classification process using methods that are appropriate to the classification model, if the training data does not exist then the research cannot This is done because there is no target for the classification to be carried out. Next is the stage of designing and testing the model that has been created in data mining, which is the process of creating a system and testing the system. Next, the test results are the results obtained from system testing using data that has been obtained previously. Furthermore, evaluation is the process of designing an evaluation model that is used to see the effectiveness and accuracy of the method used. Lastly, the evaluation results are the results obtained from the evaluation process. Furthermore, the design, testing and evaluation stages are also very important because these stages are the results of research. It was also carried out in research conducted by (F. F. Hasibuan, Dar, & Yanris, 2023) that these stages were the core of the research and became the results of the research.

Confusion Matrix

Confusion matrix is an important evaluation tool in data mining to measure the performance of classification models. This matrix is a table that visualizes model performance by comparing the predictions produced by the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

model to the known actual values. The confusion matrix consists of four main elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These elements help in calculating important metrics such as accuracy, precision, sensitivity, and specificity, which provide deep insights into how well the model can identify different classes in the dataset. The Confusion Matrix table can be seen below.

Table 1
Confusion Matrix

		Prediction Class	
	Class	**True**	**False**
Attribute Class	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

From the table above for the explanation below.

1. TP (True Positive), namely the amount of positive data that has a true value.
2. TN (True Negative), namely the amount of negative data that has a true value.
3. FN (False Negative), namely the amount of negative data but which has the wrong value.
4. FP (False Positive), namely the amount of data that is positive but has the wrong value.

$$\boldsymbol{Acuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FN+FP}} \times \mathbf{100\%}$$ (Indah, Sari, & Dar, 2023)

$$\boldsymbol{Presisi} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \times \mathbf{100\%}$$ (S. A. Hasibuan, Sihombing, & Nasution, 2023)

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \times \mathbf{100\%}$$ (Maizura, Sihombing, & Dar, 2023)

RESULT

Data analysis

In analyzing this data, the author will use several key attributes to gain a deeper understanding of student achievement. These attributes include the full names of students and students, which will make it easier to identify and document personal data. Apart from that, other attributes used are Knowledge value, attitude value and skill value. Knowledge scores measure students' understanding of the material being taught, attitude scores assess how students behave and interact in the school environment, and skills scores evaluate students' practical abilities in applying what they have learned. By combining these three aspects of assessment, the author aims to provide a comprehensive picture of student competence in various dimensions, which does not only focus on academics but also includes aspects of attitudes and skills. The data that has been obtained is as follows.

Table 2. Research Sample Data

Full Name	Presence	Attitude	Knowledge	Skills
Abul Miqdad Almadani Marbun	93	93	97	96
Adam Syahputra Hasibuan	92	92	98	97
Adzwar Al Buqori	83	83	85	82
Afril Anzas Mara	86	86	84	82
Ahmad Fahrezi	80	80	84	82
Ahmad Soleh Siahaan	87	87	84	86
Aldi Syahputra	93	93	94	92
Alvino	92	92	97	94
Anggi Khadijah Munthe	92	92	96	97
Anisa Wahyuni	93	93	97	92
Annisa Syafirah	89	89	84	86
Aprilia Lestari	85	85	82	84
Ardiansyah Pitra Simanjuntak	92	92	93	95
Aria Syahfitri	86	86	84	87
Aulia Saskia	85	85	87	84
Azka Maharani	87	87	84	86

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Bella Shintia	92	92	93	92
Dea Agiz	81	81	88	86
Delpin Arapah Hasibuan	86	86	82	84
Desnita Sari	84	84	82	86
Dicky Afandi	92	92	91	95
Dika Alviandi	92	92	97	93
Dwi Pebriyanti	85	85	83	84
Ejiy Adriyansah	85	85	87	88
Erlina Ritonga	86	86	89	83
Fajri Ananda	92	92	97	95
Farida Hannum Rangkuti	93	93	92	95
Fatur Rahman Munthe	97	97	98	96
Fikri Ardiansyah	92	92	94	97
Fitri Samsari Dewi	85	85	82	87
Habib Akbar Siahaan	85	85	84	81
Hafiz Ilman Hakim	92	92	94	98
Imam Maulana	85	85	84	87
Intan Mutiara	93	93	96	92
Iqbal Maulana	93	93	92	96
Irma Fadillah	93	93	98	92
Jesika Dewi Putri	80	80	84	82
Juwita Pratiwi	84	84	82	88
Kartika Maharani	92	92	91	97
Khaimi Firdiansyah Siregar	93	93	97	94
Latisyah Putri	80	80	82	84
Linda Risky Mawaddah	85	85	84	88
Listia Ningsih	85	85	87	83
Lufhi Darmawansah Pulungan	86	86	62	89
Mita Afrilia	83	83	85	82
Muhammad Arifin	94	94	92	96
Muhammad Ariya Fadillah	92	92	97	94
Muhammad Daffa Malikul Ka`Ab	92	92	93	95
Muhammad Fahri	92	92	94	96
Muhammad Ibnu Sabilillah	85	85	87	84
Muhammad Ihsan Aldiansyah	92	92	93	92
Muhammad Rendi Irawan	86	86	82	84
Muhammad Rizki	86	86	84	82
Muhammad Rizky Ritonga	93	93	97	96
Mutia Putri Wulandari	93	93	92	95
Mutiara Cahyani	97	97	98	96
Nah Indah Harahap	92	92	98	97
Naila Putri	85	85	82	87
Nanda Ardiansyah	92	92	94	98
Nazril Ilham	93	93	96	92
Nazwa Rahmadini Gultom	85	85	87	83
Nur Azizah Br Siregar	86	86	62	89
Nur Ica Septina	86	86	84	82
Nurul Patmasari	94	94	92	96
Panji Irwansyah	80	80	84	82
Putri Sintia Bella	87	87	84	86
Putri Zahara Azni	92	92	94	96
Raja Kesya Alvaro Pakpahan	93	93	94	92
Ramadani	86	86	84	82
Reza Wijaya	92	92	94	96
Rezza Putra Jambak	92	92	96	97
Rinda Mei Citra Pane	92	92	96	93
Riska Fitria	83	83	87	82

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Saripan	85	85	82	89
Saskia Rizkiyani	82	82	87	84
Sazia Parapat	85	85	87	88
Selly Maulani	93	93	97	92
Silmi Kaffa	89	89	84	86
Sindi Afrilia	93	93	97	94
Sintia Aura Ritonga	92	92	94	98
Soleh	86	86	84	89
Suci Ermina Ritonga	85	85	82	84
Suci Rahmadani	85	85	87	82
Syahbila	86	86	84	87
Syifa Nur Kayla	87	87	84	86
Taufik Maulana	81	81	88	86
Tiara Rambe	85	85	87	89
Tirani Akhirunisa Rambe	94	94	96	98
Ummi Umayyah	82	82	87	83
Wardatul Jannah Nasution	85	85	84	88
Yoga Dimas Pratama	84	84	82	86
Zakaryah Nasution	92	92	91	95

In table 2 above is sample data taken from the MTS Guppi Tanjung Harapan School. The class taken is class 7. This is because in class 7 students will take an exam to determine whether they are in the superior class or not in the superior class. This means that the table above is the data obtained which will later be used to be classified in data mining using the KNN method.

Data Training

Training data is a collection of data used in machine learning to train models so they can recognize patterns or make predictions. This data set includes diverse inputs along with expected outputs or labels, allowing machine learning algorithms to learn from these examples. In the learning process, the model will use this training data to adjust parameters and correct errors, so that when faced with new data, the model can make accurate predictions or classifications. The use of representative and high-quality training data is the main key in developing effective and reliable models. The training data that will be used is as follows.

Table 3. Data Training

Full Name	Presence	Attitude	Knowledge	Skills	Category
Alisha Putri Maharani	89	93	97	96	Superior
Bagas Rizki Pratama	87	92	88	85	Not Superior
Cahaya Intan Permata	92	92	94	96	Superior
Daffa Rayhan Alfarez	97	98	97	82	Superior
Elina Siti Rahmawati	92	97	89	97	Superior
Farrel Aditya Nugraha	95	96	84	97	Superior
Ghina Ayu Lestari	91	87	87	87	Not Superior
Hafizh Akbar Maulana	98	86	86	84	Not Superior
Indah Puspita Sari	95	84	89	82	Not Superior
Jovan Alif Ramadhan	82	89	85	81	Not Superior

In table 3 above is a training data table that will be used to assist the data classification process. So the data above will later be used to become a classification target. The data used as training data is 10 data which is data from students who have already taken the exam and obtained a score indicating that they are or are not in the top class.

Planning and Model Testing on Data Mining

The design and testing of data mining models that will be carried out by the author is a comprehensive process, starting with designing a model that will be applied to training data to identify patterns and make predictions. This process involves the use of the K-Nearest Neighbor (KNN) method, a machine learning technique that functions by classifying new data based on similar characteristics to data in known categories. After the model is designed, the next step is to test the model using testing data to evaluate the performance and accuracy of the model in

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

conditions that have never been seen before. To facilitate this design and testing process, the author will use the Orange application, a visual and user-friendly data analytics software, supporting the process of data exploration, modeling and model evaluation in an intuitive and efficient way.

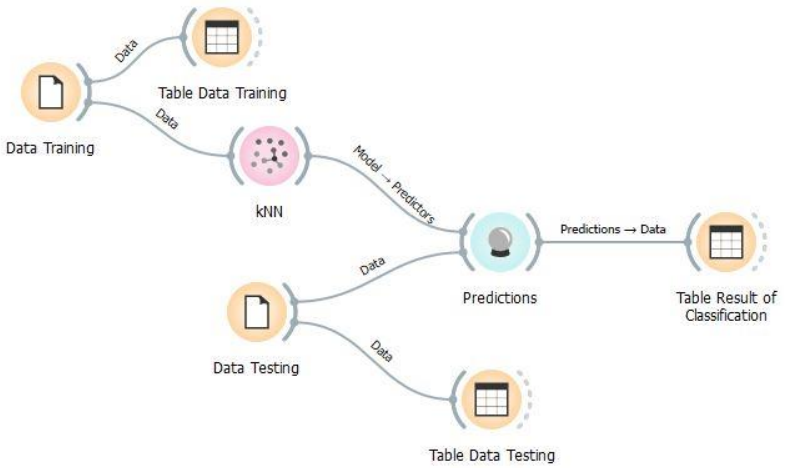

Fig 2. Data Testing Model Design in Data Mining

In figure 2 above is a model designed and designed in the orange application. Later the design model above will be used to carry out data testing or what can be called data classification. So the sample data that has been obtained will be classified using the design model in the image above. The method can be seen in the model above knn which is the method used to classify data in data mining.

Model Testing Results in Data Mining

At this stage, it is the result of predicting the superior class of students using the K-Nearest Neighbor (kNN) method in data mining.

Table 4. Model Testing Results Data

Full Name	Presence	Attitude	Knowledge	Skills	Category
Abul Miqdad Almadani Marbun	91	93	97	96	Superior
Adam Syahputra Hasibuan	90	92	98	97	Superior
Adzwar Al Buqori	90	83	85	82	Not Superior
Afril Anzas Mara	91	86	84	82	Not Superior
Ahmad Fahrezi	95	80	84	82	Not Superior
Ahmad Soleh Siahaan	96	87	84	86	Not Superior
Aldi Syahputra	90	93	94	92	Superior
Alvino	92	92	97	94	Superior
Anggi Khadijah Munthe	94	92	96	97	Superior
Anisa Wahyuni	90	93	97	92	Superior
Annisa Syafirah	88	89	84	86	Not Superior
Aprilia Lestari	89	85	82	84	Not Superior
Ardiansyah Pitra Simanjuntak	90	92	93	95	Superior
Aria Syahfitri	89	86	84	87	Not Superior
Aulia Saskia	86	85	87	84	Not Superior
Azka Maharani	88	87	84	86	Not Superior
Bella Shintia	97	92	93	92	Superior
Dea Agiz	98	81	88	86	Not Superior
Delpin Arapah Hasibuan	99	86	82	84	Not Superior
Desnita Sari	98	84	82	86	Not Superior
Dicky Afandi	87	92	91	95	Superior
Dika Alviandi	89	92	97	93	Superior
Dwi Pebriyanti	87	85	83	84	Not Superior
Ejiy Adriyansah	95	85	87	88	Not Superior
Erlina Ritonga	90	86	89	83	Not Superior
Fajri Ananda	93	92	97	95	Superior
Farida Hannum Rangkuti	88	93	92	95	Superior

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fatur Rahman Munthe	96	97	98	96	Superior
Fikri Ardiansyah	90	92	94	97	Superior
Fitri Samsari Dewi	88	85	82	87	Not Superior
Habib Akbar Siahaan	89	85	84	81	Not Superior
Hafiz Ilman Hakim	90	92	94	98	Superior
Imam Maulana	87	85	84	87	Not Superior
Intan Mutiara	95	93	96	92	Superior
Iqbal Maulana	89	93	92	96	Superior
Irma Fadillah	94	93	98	92	Superior
Jesika Dewi Putri	97	80	84	82	Not Superior
Juwita Pratiwi	98	84	82	88	Not Superior
Kartika Maharani	89	92	91	97	Superior
Khaimi Firdiansyah Siregar	87	93	97	94	Superior
Latisyah Putri	86	80	82	84	Not Superior
Linda Risky Mawaddah	95	85	84	88	Not Superior
Listia Ningsih	98	85	87	83	Not Superior
Lufhi Darmawansah Pulungan	97	86	62	89	Not Superior
Mita Afrilia	90	83	85	82	Not Superior
Muhammad Arifin	94	94	92	96	Superior
Muhammad Ariya Fadillah	95	92	97	94	Superior
Muhammad Daffa Malikul Ka`Ab	88	92	93	95	Superior
Muhammad Fahri	89	92	94	96	Superior
Muhammad Ibnu Sabilillah	90	85	87	84	Not Superior
Muhammad Ihsan Aldiansyah	97	92	93	92	Superior
Muhammad Rendi Irawan	87	86	82	84	Not Superior
Muhammad Rizki	88	86	84	82	Not Superior
Muhammad Rizky Ritonga	87	93	97	96	Superior
Mutia Putri Wulandari	87	93	92	95	Superior
Mutiara Cahyani	96	97	98	96	Superior
Nah Indah Harahap	94	92	98	97	Superior
Naila Putri	91	85	82	87	Not Superior
Nanda Ardiansyah	90	92	94	98	Superior
Nazril Ilham	95	93	96	92	Superior
Nazwa Rahmadini Gultom	99	85	87	83	Not Superior
Nur Azizah Br Siregar	97	86	62	89	Not Superior
Nur Ica Septina	90	86	84	82	Not Superior
Nurul Patmasari	98	94	92	96	Superior
Panji Irwansyah	95	80	84	82	Not Superior
Putri Sintia Bella	96	87	84	86	Not Superior
Putri Zahara Azni	89	92	94	96	Superior
Raja Kesya Alvaro Pakpahan	88	93	94	92	Superior
Ramadani	86	86	84	82	Not Superior
Reza Wijaya	85	92	94	96	Superior
Rezza Putra Jambak	96	92	96	97	Superior
Rinda Mei Citra Pane	89	92	96	93	Superior
Riska Fitria	91	83	87	82	Not Superior
Saripan	93	85	82	89	Not Superior
Saskia Rizkiyani	96	82	87	84	Not Superior
Sazia Parapat	98	85	87	88	Not Superior
Selly Maulani	96	93	97	92	Superior
Silmi Kaffa	88	89	84	86	Not Superior
Sindi Afrilia	87	93	97	94	Superior
Sintia Aura Ritonga	94	92	94	98	Superior
Soleh	99	86	84	89	Not Superior
Suci Ermina Ritonga	90	85	82	84	Not Superior
Suci Rahmadani	89	85	87	82	Not Superior
Syahbila	87	86	84	87	Not Superior

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Syifa Nur Kayla	88	87	84	86	Not Superior
Taufik Maulana	99	81	88	86	Not Superior
Tiara Rambe	95	85	87	89	Not Superior
Tirani Akhirunisa Rambe	96	94	96	98	Superior
Ummi Umayyah	99	82	87	83	Not Superior
Wardatul Jannah Nasution	88	85	84	88	Not Superior
Yoga Dimas Pratama	89	84	82	86	Not Superior
Zakaryah Nasution	90	92	91	95	Superior

In table 4 above are the results obtained from the classification results in data mining using the knn method. Of the 92 sample data used, there were 42 data of students and female students who entered the superior class and as many as 50 data of students and female students who did not enter the superior class.

Evaluation of the K-Nearest Neighbor Method Model

At the evaluation stage, researchers test how effective the method used is in producing accurate predictions or classifications. This stage focuses on measuring the accuracy, precision and recall of the model that has been developed. Accuracy measures how often the model makes correct predictions across cases. Precision assesses how accurate the model is in identifying positive cases, while recall measures the model's ability to find all relevant positive cases in the dataset. This evaluation is important to understand the model's performance in real-world conditions and to identify areas that need improvement. Thus, this evaluation stage is key in ensuring that the data mining method used is reliable and effective in applying the Knowledge or insights obtained from the analyzed data.

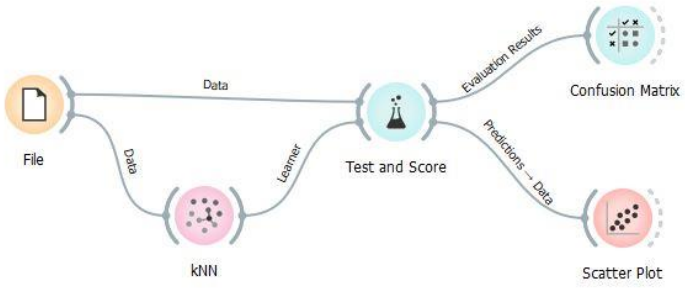

Fig 3. Method Evaluation Model in Data Mining

Figure 3 shows a model designed to evaluate the knn method. The model above was also designed using the orange application. With this model, you can carry out evaluations well. The aim of designing this evaluation model is to see and determine the effectiveness of the method used. This aims to see the accuracy of the method used. This accuracy can later provide results that can explain how good and suitable the method used in this research is.

Evaluation Results of the K-Nearest Neighbor Method

In the results of this evaluation, the author will present details regarding the accuracy, precision and recall obtained from applying the K-Nearest Neighbors (KNN) method. The author uses two widgets in the Orange application to carry out this evaluation, namely the 'Test and Score' widget and the 'Confusion Matrix' widget. The 'Test and Score' widget provides a comprehensive evaluation of model performance, including accuracy, precision and recall, in an intuitive and easy to understand manner. Meanwhile, the 'Confusion Matrix' widget helps in visualizing the model performance in more detail, allowing the authors to understand how the model predicts different classes and where prediction errors occur. The combination of these two widgets provides a clear picture of the effectiveness of the KNN method in data classification, allowing the authors to conduct an in-depth analysis of the strengths and limitations of the developed model.

Result of Test and Score

Table 5. Result of Test and Score

Model	AUC	CA	F1	Precision	Recall
K-Nearest neighbor	1.000	1.000	1.000	1.000	1.000

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13627
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In table 5 are the results of the Test and Score which are the results of the evaluation of the method used. The accuracy results obtained from the test and score widget obtained very good results. This is because the accuracy results are perfect.

Result of Confusion Matrix

Table 6
Result of Confusion Matrix

		Predicted		
		Superior	Not Superior	∑
Actual	Superior	42	0	42
	Not Superior	0	50	50
	∑	42	50	92

In table 6 are the results of the evaluation using the confusion matrix widget, the table above has provided the results, but the results above cannot yet be stated that the accuracy obtained from the knn method is good. Calculations need to be carried out for accuracy, precision and recall values. However, previously from the table above, the True Positive value was obtained, namely 42 data, for the True Negative value, 50 data were obtained, for the False Positive value, the value was 0 data and for the False Negative value, the value was 0 data. To calculate the accuracy, precision and recall values for the confusion matrix widget, they are as follows.

$$\boldsymbol{Accuracy} = \frac{42+50}{42+50+0+0} + \mathbf{100\%}$$ Then the Accuracy value = 100%

$$\boldsymbol{Presisi} = \frac{42}{42+0} + \mathbf{100\%}$$ Then the Precision value = 100%

$$\boldsymbol{Recall} = \frac{42}{42+0} + \mathbf{100\%}$$ Then the Recall value = 100%

From the results of the calculations carried out, the accuracy, precision and recall results obtained from the confusion matrix widget are 100%. It can be stated that the knn method can be used well to carry out data classification and predictions, because the accuracy obtained is perfect.

Scatter Plot Result

Scatter Plot results are used to visualize the relationship between two variables and to identify patterns or trends that may exist in a data set. Scatter Plots map each data point in two-dimensional space with one variable on the x-axis and another variable on the y-axis, allowing researchers to see how the data correlates, whether positively, negatively, or not at all.

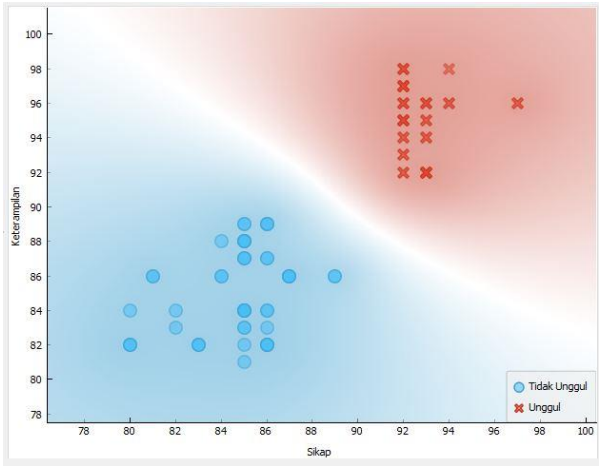

Fig 4. Scatter Plot Results

Figure 4 shows a scatter plot of skill and attitude attributes that fall into the superior and non-superior categories. But it can be seen in the picture above that the results that are not superior are more numerous than the results that are superior. So in the image above there are 2 colors that explain the number of method evaluation

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

results. So it can be seen that between students who are in superior or non-superior classes, the comparison is not too far apart. But there are more students and students who do not excel.

DISCUSSIONS

Research on the analysis of superior class students at MTS Guppi Tanjung Harapan using the K-Nearest Neighbors (KNN) method has provided impressive results. This method, known for its superior ability to classify data, has proven to be very effective in identifying students who have the potential to enter superior classes. KNN, with its distance-based approach to classifying data, allows the system to accurately group students based on important variables such as attitude, skill and knowledge scores and also attendance scores. In this study, kNN was implemented to classify students based on their academic performance, using a dataset that includes various features such as academic grades, attendance, and participation in extracurricular activities. The kNN model is configured with optimal parameters determined through a cross-validation process to obtain the most accurate predictions.

The results of model evaluation using the "Test and Score" widget in data mining software show that prediction accuracy reaches 100%. This shows that the kNN model is able to classify students into superior classes perfectly, without errors. Furthermore, use of the “Confusion Matrix” widget confirmed similar results, with no misclassification between superior and non-distinguished students, also indicating 100% accuracy. Comparison of the two evaluation matrices, namely the results of the "Test and Score" and the "Confusion Matrix", gives a ratio of 1:1, indicating suitability and consistency in the excellent results of the two evaluation methods. This consistency confirms the reliability of the kNN method in educational applications, especially in identifying and predicting students' superior classes based on available data. This research underlines the potential of data mining in the education sector to assist in decision making and improve the quality of education. In research conducted by (Irmayani & Sinaga, 2023) the accuracy obtained using the knn method also obtained very good accuracy results, namely 100%. This means that the KNN method is very good for classifying data.

CONCLUSION

The results of the analysis of the superior classes of students and female students at MTS Guppi Tanjung Harapan showed very satisfying results. Of the 92 sample data analyzed using the K-Nearest Neighbors (KNN) method, it was found that 42 students were successfully classified into superior classes, while 50 other students were not. The success of this classification confirms the effectiveness of the KNN method in analyzing and identifying students' academic potential. Furthermore, the consistency of good accuracy results from two different evaluation widgets, namely 'Test and Score' and 'Confusion Matrix', strengthens the reliability of the analysis carried out. These results not only provide valuable insights for school decision makers to design more targeted educational programs, but also demonstrate the potential of using data mining in education to optimize student learning and development.

REFERENCES

Adjani, K., Fauzia, F. A., & Juliane, C. (2023). Comparison of K-N Earest Neighbor and Naïve Bayes Algorithms for Prediction of Aptikom Membership Activity Extension in 2023. *SinkrOn*, *8*(2), 700–707. https://doi.org/10.33395/sinkron.v8i2.12081

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

Arifuddin, N. A., Pinastawa, I. W. R., Anugraha, N., & Pradana, M. G. (2023). Classification of Stroke Opportunities with Neural Network and K-Nearest Neighbor Approaches. *SinkrOn*, *8*(2), 688–693. https://doi.org/10.33395/sinkron.v8i2.12228

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

Hermawan, F., & Prianggono, J. (2023). Crime of theft prediction using Machine Learning K-Nearest Neighbour Algorithm at Polresta Bandar Lampung. *Sinkron*, *8*(3), 1515–1527. https://doi.org/10.33395/sinkron.v8i3.12422

Indah, I. C., Sari, M. N., & Dar, M. H. (2023). Application of the K-Means Clustering Agorithm to Group Train Passengers in Labuhanbatu. *SinkrOn*, *8*(2), 825–837. https://doi.org/10.33395/sinkron.v8i2.12260

Irmayani, D., & Sinaga, F. A. (2023). *Analysis of the Level of Public Satisfaction on the TikTok Application as an E-Commerce*. *7*(4), 2579–2591.

Istiadi, I., Rahman, A. Y., & Wisnu, A. D. R. (2023). Identification of Tempe Fermentation Maturity Using Principal Component Analysis and K-Nearest Neighbor. *Sinkron*, *8*(1), 286–294. https://doi.org/10.33395/sinkron.v8i1.12006

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

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

Munazhif, N. F., Yanris, G. J., Nirmala, M., & Hasibuan, S. (2023). *Implementation of the K-Nearest Neighbor (kNN) Method to Determine Outstanding Student Classes*. *7*(2), 719–732.

Pratama, H. A., Yanris, G. J., Nirmala, M., & Hasibuan, S. (2023). *Implementation of Data Mining for Data Classification of Visitor Satisfaction Levels*. *8*(3), 1832–1851.

Saputra, A. D., Hindarto, D., & Santoso, H. (2023). Disease Classification on Rice Leaves using DenseNet121, DenseNet169, DenseNet201. *Sinkron*, *8*(1), 48–55. https://doi.org/10.33395/sinkron.v8i1.11906

Sari, A. W., Hermanto, T. I., & Defriani, M. (2023). Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine. *Sinkron*, *8*(3), 1366–1378. https://doi.org/10.33395/sinkron.v8i3.12447

Sari, M., Yanris, G. J., & Hasibuan, M. N. S. (2023). Analysis of the Neural Network Method to Determine Interest in Buying Pertamax Fuel. *SinkrOn*, *8*(2), 1031–1039. https://doi.org/10.33395/sinkron.v8i2.12292

Sinaga, B., Marpaung, M., Tarigan, I. R. B., & Tania, K. (2023). Implementation of Stock Goods Data Mining Using the Apriori Algorithm. *Sinkron*, *8*(3), 1280–1292. https://doi.org/10.33395/sinkron.v8i3.12852

Triani, D. J., Dar, M. H., & Yanris, G. J. (2023). Analysis of Public Purchase Interest in Yamaha Motorcycles Using the K-Nearest Neighbor Method. *Sinkron*, *8*(3), 1238–1254. https://doi.org/10.33395/sinkron.v8i3.12433

Violita, P., Yanris, G. J., & Hasibuan, M. N. S. (2023). Analysis of Visitor Satisfaction Levels Using the K-Nearest Neighbor Method. *SinkrOn*, *8*(2), 898–914. https://doi.org/10.33395/sinkron.v8i2.12257

Zai, F., Sirait, J., Nainggolan, D. W., Sihombing, N. G. D., & Banjarnahor, J. (2023). Comparison Analysis of C4.5 Algorithm and KNN Algorithm for Predicting Data of Non-Active Students at Prima Indonesia University. *Sinkron*, *8*(4), 2027–2035. https://doi.org/10.33395/sinkron.v8i4.12879

*name of corresponding author