Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation of the K-Nearest Neighbor (kNN) Method to Determine Outstanding Student Classes

Nanda Fahrezi Munazhif[1)*], Gomal Juni Yanris[2)], Mila Nirmala Sari Hasibuan[3)]
[1,2,3)]Universitas Labuhanbatu
[1)]munazhifalfatih@gmail.com, [2)]gomaljuniyanris@gmail.com, [3)]milanirmalasari7@gmail.com

Submitted : Feb 28, 2023 | **Accepted** : Mar 22, 2023 | **Published** : Apr 1, 2023

Abstract: Education being one factor supporting students / I to be able to increase their knowledge. Each student has their own potential that they have obtained in the world of education. Therefore, every school has created an education program that functions to increase the potential of high achieving students. The program is a flagship class program. What is meant by a superior class program is a process of selecting and classifying students to be placed in the classroom superior (grade student / I achievement). Therefore, this study aims to implement classification on student data using the K-Nearest Neighbor (kNN) algorithm. K-Nearest Neighbor (kNN) is a method used to classify data based on training data (data set). The data that the writer will use is student data of 60 student data. In this classification using the kNN method aims to classify data on students who are eligible to enter the superior class (class of outstanding students). The first step is the process of determining data requirements. Then cleaning or pre-processing and the next is to design a widget model of the kNN method on the orange application to carry out the data classification process. The test results using 60 student data using the KNN method and using the Confusion Matrix obtained an Accuracy value of 91.6%, then a Precision value of 89.2% and a Recall value of 92.5%. The conclusion is that this study succeeded in obtaining a method that the best and also get the best results for Classification of superior student classes.
Keywords: Classification, Confusion Matrix, Data Mining, Nearest Neighbor, ROC analysis, Superior Class,

INTRODUCTION

The MTs Swasta Islamiyah Kotapinang was established on August 12, 1947. The establishment of the Madrasah Islamiyah was at a time when the Indonesian people were fighting against the Dutch, with bullets, mortars, cannons and sweat and blood drenching the earth of pervade fighting against enemies with unequal weapons. This Madrasah is one of the junior high school level Islamic schools in Kotapinang District. This Madrasah has many students who are interested in attending this Madrasah. In it there is a lot of knowledge about religion, especially Islam and also a lot of general knowledge, such as mathematics, Indonesian and many others.

For the process of selecting and determining students who are eligible to enter and be accepted at this Madrasah school, they must meet the conditions set by the Madrasah. Not only that, to determine student classrooms, the school also selects superior students and will be placed in the first room and vice versa. To determine superior and non-excellent classes, they still use a manual system, namely calculating data or grades which are indicators of student excellence, still using a manual system. So that the calculation

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12227 p-ISSN : 2541-044X

method used must be replaced with an appropriate and accurate method, according to the desired needs. This method can also speed up data calculations.

The superior class is a number of students who have outstanding achievements and will later be grouped in a special class. This is done to develop intelligence, skills, abilities, talents, interests, attitudes and behavior of students/I so that they can have high, good and superior achievement indicators in accordance with their respective potentials. Pre-eminent classes are carried out in order to improve the quality and quantity of students towards the intelligent generation. There are many who have a positive impact in carrying out superior class activities or programs. The superior class program is also implemented to provide encouragement to students to be more active in learning and improving their skills, both formal and non-formal.

Data mining is a process of using statistics, mathematics, and artificial intelligence to extract and identify useful information from data. (Baharuddin, Azis, & Hasanuddin, 2019) The data mining will later be stored in a repository using reasoning pattern technology, statistical techniques and mathematics. (Wijaya & Girsang, 2015) Mining is carried out according to the needs of a writer. Data mining will later be stored using statistical and mathematical techniques. Data mining is a process of mining data with a large enough amount. So, to do data mining, we need to use the methods or algorithms that exist in data mining. In data mining later we will carry out a classification using existing methods in data mining. Classification is a learning model used to manage data mining. This classification will later determine the class of each desired category, meaning that in data mining the classification functions to determine the category or class of each data.(Saputra, Widiyaningtyas, & Wibawa, 2018) Data mining can later also turn data into useful information and knowledge. (Hussain, Dahan, Ba-Alwib, & Ribata, 2018) This means that we will analyze the data that we classify to determine each class from each category. (Kumar, Chatterjee, & Díaz, 2020) Classification of superior classes is very important in the world of education, because it can provide encouragement to students to study more actively.(Ibrahim, Zulkifli, Sabri, Shari, & Noordin, 2019) This classification can also later determine and see the quality of each student.(Zaki, Zulkifley, Mohd Stofa, Kamari, & Mohamed, 2020)

The K-Nearest Neighbor (KNN) method is an appropriate way to classify data based on data that is close to the desired object.(Lubis, Lubis, & Al-Khowarizmi, 2020) While classification is the process of finding a function or shape of something to serve as a differentiator between the concept and class.(Zheng & Wu, 2006) Therefore we can use the K-Nearest Neighbor algorithm which can be a good classification method. Classification is the main technique in carrying out this analysis.(George, Shaikat, Ferdawoos Hossain, Parvez, & Uddin, 2019) The K-Nearest Neighbor method later aims to classify data with superior and non-superior conditions.(Isnain, Supriyanto, & Kharisma, 2021) Classification using the K-method Nearest Neighbor will later be carried out on a technology basis, so in its implementation later it will use the orange application.(Jadhav, Udupi, & Patil, 2019)

METHOD

The K-Nearest Neighbor (KNN) algorithm is a good classification method with close proximity to data mining based on an object and its terms.(Liantoni, 2016) K-Nearest Neighbor can be a good classification method.(Kurniawan & Barokah, 2020) This method is carried out by testing and comparing data, between training data and testing data.(Liantoni, 2016) Using the K-Nearest Neighbor method can help us solve a problem and have high accuracy.(Farhad Khorshid & Mohsin Abdulazeez, 2021) The K-Nearest Neighbor method has several advantages, this method can conduct training quickly, how to understand it is simple and also easy to learn. The K-Nearest Neighbor method can withstand data that has noise and this method is also effective if used for training large enough data.(Rosso, 2019) For the general form of the formula for the K-Nearest Neighbor method we will use the Euclidean distance formula, namely as follows:

$$\boldsymbol{Dxy} = \sqrt{\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{y}_i)}^2$$

Information:
D : Proximity Distance

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12227 p-ISSN : 2541-044X

x : data training
y : data testing
n : the number of individual attributes between 1 s/d n
f : function similitary attribute I between x and y
I : individual attributes between 1s/d n

Confusion Matrix is a tool that can measure and calculate the level of classification of a data to determine the accuracy, precision, recall that exist in data mining.

TABLE 1
CONFUSION MATRIX TESTING

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

Where table 1 contains:
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 a wrong value.
4. FP (False Positive), namely the amount of data that is positive but which has a false value.

The precision value can be obtained from the number of positive category data that is classified correctly divided by the total data classified as positive, precision can be obtained using the formula:

$$\boldsymbol{Acuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FN + FP}}$$

The precision value can be obtained from the number of positive category data that is classified correctly divided by the total data classified as positive, precision can be obtained using the formula:

$$\boldsymbol{Presisi} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}}$$

Recall is used to indicate the number of offenders from positive category data that are correctly classified. To find the Recall value using the formula:

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$$

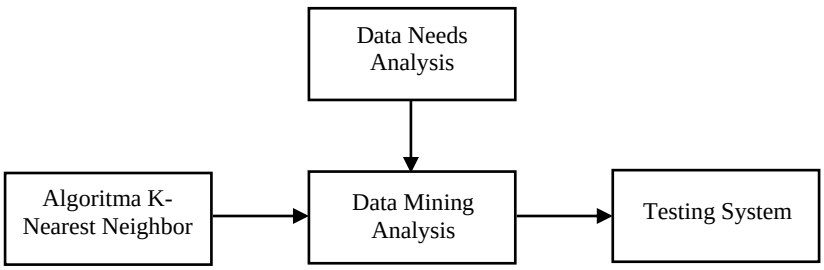

***Figure 1**. Data Mining Process Design*

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12227 p-ISSN : 2541-044X

In Figure 1 is a process design that will be carried out in the data mining process. Analysis of data requirements is data that will later be used to carry out an analysis or classification in data mining. The K-Nearest Neighbor algorithm is a data classification method that is close to the desired object. Data mining analysis is a process that will be carried out to determine patterns or rules that exist in the K-Nearest Neighbor method. System testing is the final process of data mining, namely the process of testing the system to see the errors that exist in the system and see the results of the system.

RESULT

3.1. Data Analysis

In the picture below is the data of MTs Swasta Islamiyah Kotapinang students. This data then becomes the criterion or determinant for students who qualify for the superior class. The data includes the student's name, gender, ethnicity, knowledge, skills, and attitudes. In the knowledge column is the value of student knowledge, the value of knowledge means the knowledge that students have such as counting, reading and answering questions given by the school. So, if the student can answer the question correctly, he will get a high score. Therefore, knowledge is one of the criteria for students to enter the superior class. Skills are the value of student practice. In Madrasah education there is a lot of knowledge about religion. Not only that, religious lessons will not only discuss theory but will be practiced directly. Then the value of skills is the value of student practice, such as prayer, fasting, etc. Finally, attitude is student behavior starting from being polite to the teacher, speaking fluently and listening when the teacher explains. This is the criterion that determines a student's eligibility for the next grade. Those are the three criteria that determine the eligibility of students to enter the superior class. These criteria are very important in the world of education.

Name	Gender	NIS	Knowledge	Skills	Attitude	Category
Ahmad Arrivat Nasution	Man	229166	93	97	96	Superior
Ahmad Rifai Ginting	Man	229167	92	98	97	Superior
Ahmad Saud Maritho Rambe	Man	229328	86	84	82	Not Superior
Ahmad Vidy Damira	Man	161758	80	84	82	Not Superior
Akbar Dalimunthe	Man	229331	87	84	86	Not Superior
Alika Salsabila Harahap	Woman	229168	93	94	92	Superior
Alisah Johan	Woman	229169	92	96	97	Superior
Anggi Febriansyah Daulay	Man	229170	93	97	92	Superior
Annisa Ayuni	Woman	229368	89	84	86	Not Superior
Aulia Ritonga	Woman	229332	85	82	84	Not Superior
Azlila Alnatun Karmila Siregar	Woman	229333	86	84	87	Not Superior
Deni Ajriyansyah	Man	229335	87	84	86	Not Superior
Dinggol Negara Ritonga	Man	229336	81	88	86	Not Superior

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

El Guaje Perdana Siregar	Man	229338	84	82	86	Not Superior
Fikri Nabhan Hasibuan	Man	229173	92	91	95	Superior
Geisha Rahayu Alsin	Woman	229174	92	97	93	Superior
Harisa Azra Harahap	Woman	229339	85	83	84	Not Superior
Hazmico Yudayana Ritonga	Man	229340	85	87	88	Not Superior
Hendryanta Kusuma	Man	229341	86	89	83	Not Superior
Ibnu Hafiz Siregar	Man	229176	92	97	95	Superior
Indah Dwi Jannika	Woman	229177	92	94	97	Superior
Indah Sari Hasibuan	Woman	229342	85	84	81	Not Superior
Iskandar Zulkarenain Pulungan	Man	229368	85	84	87	Not Superior
Jihan Khairani	Woman	229179	93	92	96	Superior
Kanaya Syafira Rahmad Siregar	Woman	229181	93	98	92	Superior
M. Maulana Pazri Harahap	Man	161832	80	84	82	Not Superior
Muhammad Basit Siregar	Man	229346	84	82	88	Not Superior
Muhammad Farhan Dalimunthe	Man	229183	92	91	97	Superior
Muhammad Rahmad Hamdi Nasution	Man	229184	93	97	94	Superior
Muhammad Rava Hasibuan	Man	229347	80	82	84	Not Superior
Muhammad Rofiq Matondang	Man	182344	85	84	88	Not Superior
Muhammad Soleh	Man	229349	83	85	82	Not Superior
Nabila Shafa Hasibuan	Woman	229186	92	97	94	Superior
Nazwa Nabila Siregar	Woman	229187	92	93	95	Superior
Pazri	Man	229350	85	87	84	Not Superior
Putra Siregar	Man	229189	92	93	92	Superior
Putri Rahayu	Woman	161852	86	82	84	Not Superior
Qoni Atun Ni'mah	Woman	161853	93	92	95	Superior

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Queensha Dian Hafizhah	Woman	229191	97	98	96	Superior
Rahayu	Woman	229352	85	82	87	Not Superior
Raja Fatir Abdillah Tanjung	Man	229960	92	94	98	Superior
Ray Aldyansyah Harahap	Man	229193	93	96	92	Superior
Rayhan Ardian Syah	Man	229354	85	87	83	Not Superior
Reny Lestari Harahap	Woman	161862	86	82	89	Not Superior
Rifat Al Faridji Hasibuan	Man	229195	94	92	96	Superior
Rimadani	Woman	161868	92	94	96	Superior
Riza Alfarizy Ritonga	Man	229357	86	84	82	Not Superior
Roni	Man	229197	92	94	96	Superior
Salia Dahliati	Woman	229198	92	96	93	Superior
Salsa Safitri	Woman	161873	83	87	82	Not Superior
Salsabila Indah Sari	Woman	102645	85	82	89	Not Superior
Salwa As-Shifa Hasni Lubis	Woman	229360	82	87	84	Not Superior
Siti Patimah Harahap	Woman	229361	85	87	88	Not Superior
Sultan Arif Sahlan Siregar	Man	229199	93	97	94	Superior
Syifa Nursyafri	Woman	161884	92	94	98	Superior
Umairoh Putri Dalimunthe	Woman	229362	86	84	89	Not Superior
Willzlah Abe Ready Pratama	Man	229364	85	87	82	Not Superior
Wulan Dari	Woman	229365	85	87	89	Not Superior
Zahra Mafaza	Woman	229203	94	96	98	Not Superior
Zian Raditiya	Man	229366	82	87	83	Not Superior

Figure 2. *Student data*

In Figure 2, the data table above is student data obtained from MTs Islamiyah Kotapinang. The data was taken from grade 1 junior high school level of 60 students. The picture above consists of name, gender, ethnicity, knowledge value, skill value and attitude value.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

TABEL 2
STUDENT DATA ATTRIBUTES

No	Attribute	Text	Description
1	Nama	Category	Student's full name
2	Gender	Text	Student gender
3	Nis	numeric	Student identification number
4	Knowledge	numeric	Value of student knowledge
5	Skills	numeric	Assess students' skills
6	Attitude	Text	Assess students' attitudes

In the attribute table. The research attribute is data that has been obtained from MTs Swasta Islamiyah Kotapinang. The attribute data above has been given the type of each attribute and has also been given a description of each existing attribute.

3.2. Data Training

The training data is the data that we will use as a research sample, the data is obtained from MTs Swasta Islamiyah Kotapinang in the form of a .xlsx file which will be an indicator for determining student eligibility to become or enter as a superior class student.

Name	Gender	NIS	Knowledge	Skills	Attitude	Category
Ahmad Arrivat Nasution	Man	229166	93	97	96	Superior
Ahmad Rifai Ginting	Man	229167	92	98	97	Superior
Ahmad Saud Maritho Rambe	Man	229328	86	84	82	Not Superior
Ahmad Vidy Damira	Man	161758	80	84	82	Not Superior
Akbar Dalimunthe	Man	229331	87	84	86	Not Superior
Alika Salsabila Harahap	Woman	229168	93	94	92	Superior
Alisah Johan	Woman	229169	92	96	97	Superior
Anggi Febriansyah Daulay	Man	229170	93	97	92	Superior
Annisa Ayuni	Woman	229368	89	84	86	Not Superior
Aulia Ritonga	Woman	229332	85	82	84	Not Superior
Azlila Alnatun Karmila Siregar	Woman	229333	86	84	87	Not Superior
Deni Ajriyansyah	Man	229335	87	84	86	Not Superior
Dinggol Negara Ritonga	Man	229336	81	88	86	Not Superior
El Guaje Perdana Siregar	Man	229338	84	82	86	Not Superior
Fikri Nabhan Hasibuan	Man	229173	92	91	95	Superior
Geisha Rahayu Alsin	Woman	229174	92	97	93	Superior
Harisa Azra Harahap	Woman	229339	85	83	84	Not Superior
Hazmico Yudayana Ritonga	Man	229340	85	87	88	Not Superior

*name of corresponding author

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

Hendryanta Kusuma	Man	229341	86	89	83	Not Superior
Ibnu Hafiz Siregar	Man	229176	92	97	95	Superior
Indah Dwi Jannika	Woman	229177	92	94	97	Superior
Indah Sari Hasibuan	Woman	229342	85	84	81	Not Superior
Iskandar Zulkarenain Pulungan	Man	229368	85	84	87	Not Superior
Jihan Khairani	Woman	229179	93	92	96	Superior
Kanaya Syafira Rahmad Siregar	Woman	229181	93	98	92	Superior
M. Maulana Pazri Harahap	Man	161832	80	84	82	Not Superior
Muhammad Basit Siregar	Man	229346	84	82	88	Not Superior
Muhammad Farhan Dalimunthe	Man	229183	92	91	97	Superior
Muhammad Rahmad Hamdi Nasution	Man	229184	93	97	94	Superior
Muhammad Rava Hasibuan	Man	229347	80	82	84	Not Superior
Muhammad Rofiq Matondang	Man	182344	85	84	88	Not Superior
Muhammad Soleh	Man	229349	83	85	82	Not Superior
Nabila Shafa Hasibuan	Woman	229186	92	97	94	Superior
Nazwa Nabila Siregar	Woman	229187	92	93	95	Superior
Pazri	Man	229350	85	87	84	Not Superior
Putra Siregar	Man	229189	92	93	92	Superior
Putri Rahayu	Woman	161852	86	82	84	Not Superior
Qoni Atun Ni'mah	Woman	161853	93	92	95	Superior
Queensha Dian Hafizhah	Woman	229191	97	98	96	Superior
Rahayu	Woman	229352	85	82	87	Not Superior
Raja Fatir Abdillah Tanjung	Man	229960	92	94	98	Superior
Ray Aldyansyah Harahap	Man	229193	93	96	92	Superior
Rayhan Ardian Syah	Man	229354	85	87	83	Not Superior
Reny Lestari Harahap	Woman	161862	86	82	89	Not Superior
Rifat Al Faridji Hasibuan	Man	229195	94	92	96	Superior
Rimadani	Woman	161868	92	94	96	Superior
Riza Alfarizy Ritonga	Man	229357	86	84	82	Not Superior
Roni	Man	229197	92	94	96	Superior
Salia Dahliati	Woman	229198	92	96	93	Superior
Salsa Safitri	Woman	161873	83	87	82	Not Superior

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Salsabila Indah Sari	Woman	102645	85	82	89	Not Superior
Salwa As-Shifa Hasni Lubis	Woman	229360	82	87	84	Not Superior
Siti Patimah Harahap	Woman	229361	85	87	88	Not Superior
Sultan Arif Sahlan Siregar	Man	229199	93	97	94	Superior
Syifa Nursyafri	Woman	161884	92	94	98	Superior
Umairoh Putri Dalimunthe	Woman	229362	86	84	89	Not Superior
Willzlah Abe Ready Pratama	Man	229364	85	87	82	Not Superior
Wulan Dari	Woman	229365	85	87	89	Not Superior
Zahra Mafaza	Woman	229203	94	96	98	Not Superior
Zian Raditiya	Man	229366	82	87	83	Not Superior

Figure 3. *Data Training*

Figure 3 contains data on prospective students for the superior class that have been obtained from MTs Swasta Islamiyah Kotapinang which contains all the data needed in the selection process for prospective students who are eligible to enter the superior class.

TABEL 3
STUDENT DATA COLUMNS

No	Attribute	Type	Role	Values
1	Nama	Text	Meta	
2	Gender	Category	Feature	Man, Woman
3	Nis	numeric	Meta	
4	Knowledge	numeric	Feature	
5	Skills	numeric	Feature	
6	Attitude	Text	Feature	
7	Category	Categorical	Target	Superior, Not Superior

In the data column table which is the training data that we will use to determine the data classification of the superior class of MTs Islamiyah Kotapinang students. by using the K-Nearest Neighbor (kNN) method, namely by changing the role for the category from feature to target so that we can find results from existing information or data.

3.3. Data Selection Process (Preprocessing)

The data selection process is the process of selecting and determining the data needed before entering the data mining process using the K-Nearest Neighbor (kNN) method, then the preprocessing This data will be selected first. So that data determined and selected to enter the data mining process are name, gender, family name, knowledge, skills and attitudes. These data will be used and included in the data mining process using the K-Nearest Neighbor (kNN) method.

3.4. Data Mining Process

The data mining process is carried out using a classification model using the orange application using the K-Nearest Neighbor method which will determine the eligibility of students to enter the superior class of MTs Private Islamiyah Kotapinang.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

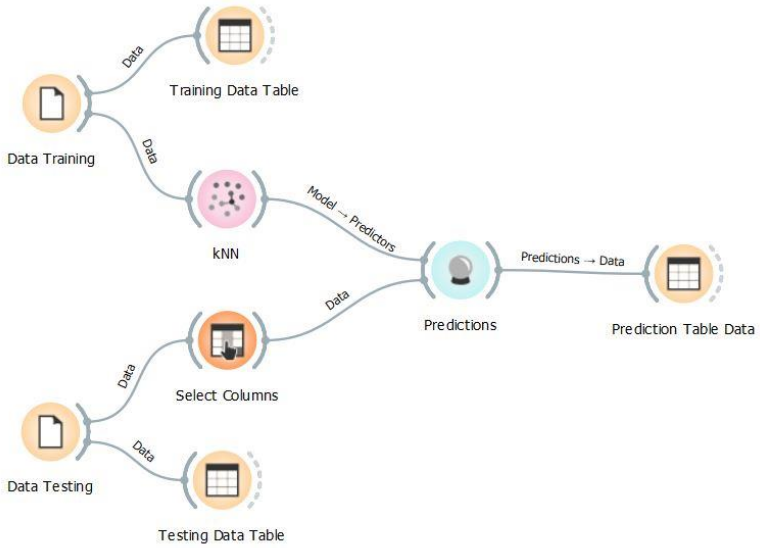

Figure 4. *Data Mining Process*

In Figure 4, the data mining process is a widget design using the orange application. This data mining was carried out using the K-Nearest Neighbor method with student data that had been obtained from MTs Islamiyah Kotapinang. The data is classified by using the K-Nearest Neighbor method to get predetermined target data.

3.5. Classification Model Testing Process

In the process of testing the K-Nearest Neighbor method that has been done previously, training data and test data are needed to test the results of data classification using the K-Nearest Neighbor method as shown below.

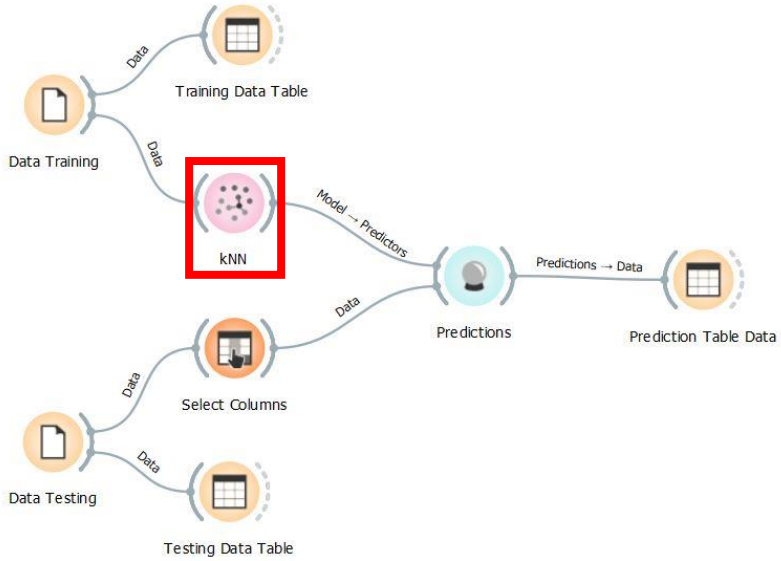

Figure 5. *Widget design model classification of scholarship acceptance status dataset*

In Figure 5. The classification model of the widget dataset category of superior class students is a pattern design that has been added to the data classification process. The widget section with a red box

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12227 p-ISSN : 2541-044X

is the K-Nearest Neighbor method which is used to test the data that has been given to determine or classify student data that falls into the superior and non-excellent categories.

3.6. Classification Model Predictions Process

In this process, it is a process for predicting the results of the data mining classification model by using predictions to determine the results of the eligibility of students who fall into the category of superior or non-excellent classes. The prediction results obtained from 60 student data at Islamic Private MTs Kotapinang using the K-Nearest Neighbor (KNN) method, namely the results of the Classification of students/I who are eligible to enter the superior class as many as 27 students/I or equal to 45% and for students who not eligible to enter the superior class as many as 33 students / i or equal to 55%. This means that the K-Nearest Neighbor (kNN) method is very suitable to be used to carry out a data classification.

3.7. Classification Model Evaluation Results

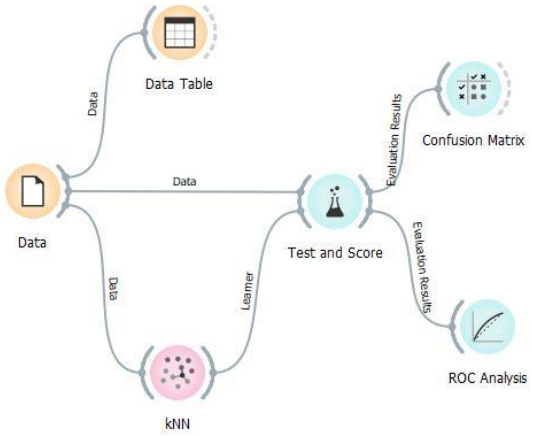

Figure 6. *Classification Evaluation Widget*

Figure 6 contains a classification evaluation consisting of patterns for determining tests and scores. Confusion Matrix and ROC Analysis were carried out using the K-Nearest Neighbor (kNN) method. To get the Confusion Matrix and ROC Analysis, namely by using training data and test data that have been combined into test data with 1 attribute as a target. 2 attributes as text, namely name and nis, 1 attribute as a category, namely gender and finally 3 attributes as numeric, namely the value of knowledge, skills and attitudes.

TABEL 4
RESULT OF TEST AND SCORE

Model	AUC	CA	F1	Precision	Recall
kNN	0.922	0.917	0.917	0.917	0.917

After conducting an evaluation with the Classification model using the K-Nearest Neighbor (kNN) method, the results obtained from the tests and scores were carried out using the K-Nearest Neighbor (kNN) method and applied using the orange application, the results obtained from the AUC were 0.922, the result of CA is 0.917, the result of F1 is 0.917, the result of Precision is 0.917 and the result of Recall is 0.917. For the process in test and score we choose Cross validation with a number of folds of 2.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

3.8. Evaluation Result with Confusion Matrix

Confusion Matrix is a measuring tool for prediction methods by calculating the correctness of data that has been classified using the K-Nearest Neighbor method.

TABEL 5
RESULT OF CONFUSION MATRIX

		Predicted		
		Superior	Not Superior	∑
Actual	Superior	25	3	28
	Not Superior	2	30	32
	∑	27	33	60

Figure 12. The True Positive (TP) result is 30. True Negative (TN) is 25, False Positive (FP) is 2 and False Negative (FN) is 3. Then the values for accuracy, precision and recall are as follows:

$$\boldsymbol{Accuracy} = \frac{30+25}{30+25+3+2} + \mathbf{100\%}$$ Then the Accuracy value = 91%

$$\boldsymbol{Presisi} = \frac{30}{30+2} + \mathbf{100\%}$$ Then the Presisi value = 93%

$$\boldsymbol{Recall} = \frac{30}{30+3} + \mathbf{100\%}$$ Then the Recall value = 90%

3.9. Evaluation Result with ROC Curve

The Roc Curve is obtained from the true signal (sensitivity) and (1 specificity) over the entire cut off point range to obtain the ROC curve visualized from the Confusion Matrix. The results of the ROC graph can be seen in Figures 12 and 13.

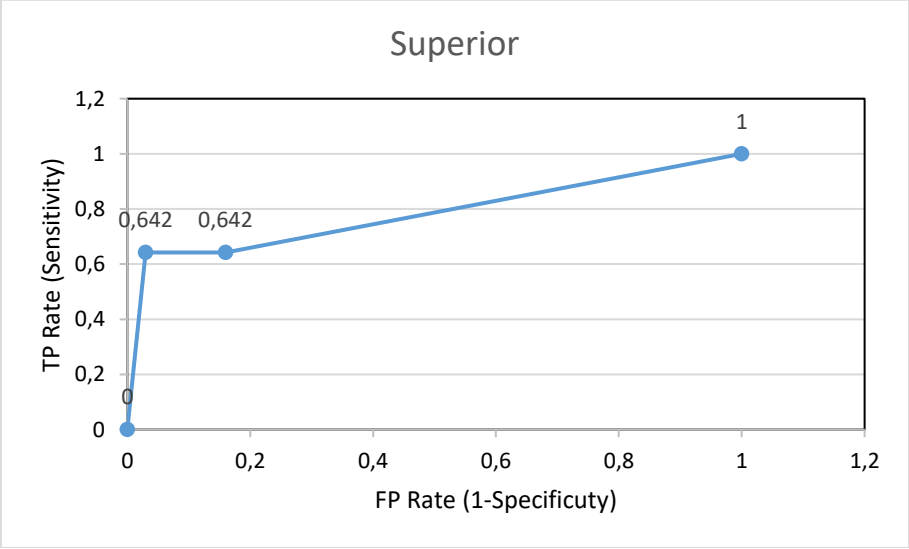

Figure7. *ROC analysis with the target of students entering the superior class*

Figure 7 states the result ROC analysis the eligibility of students to enter the superior class category of Islamic Private MTs Kotapinang, superior class with the K-Nearest Neighbor method of 0.642.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

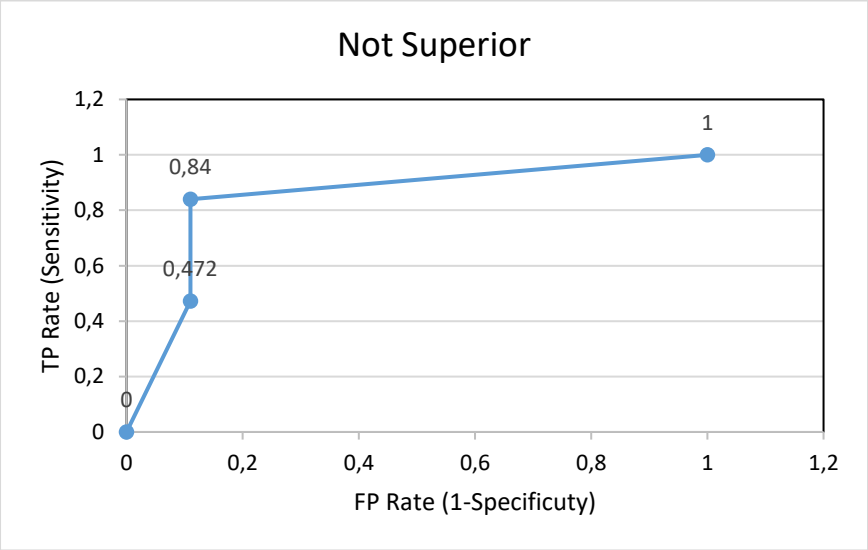

Figure 4. *ROC analysis with the target of students who enter the class are not superior*

Figure 8 states the result ROC Analysis the feasibility of students entering the superior class category of MTs Islamiyah Kotapinang, the class is not superior with the K-Nearest Neighbor method of 0.472.

DISCUSSIONS

Based on the number of datasets, namely as many as 60 data used as test data and which became research sample data, then carried out several tests with the composition of the training data and testing data with the number of neighbors being 3 using the K-Nearest Neighbor (kNN) method with the test and score widget using number of folds is 2, then the accuracy for Classification is 92.2% with a total of 27 data for training data and 60 data for testing data. But when compared with the accuracy obtained with the confusion matrix widget with the number of folds on the test and score of 2, an accuracy of 91.6% is obtained.

The results obtained using the K-Nearest Neighbor (kNN) method with test and score or added with the confusion matrix widget, but the results obtained can still be said to be high, because it is capable of producing representations with an accuracy of more than 90%.

CONCLUSION

In this classification, it can be concluded that from the K-Nearest Neighbor (kNN) method with test and score results obtained an average accuracy of 92.2%, an average precision of 91.7%, and an average recall of 91. .7%. Compared with the same method, namely K-Nearest Neighbor (kNN), but with the addition of a confusion matrix widget. Then the results obtained in the confusion matrix are for an average accuracy of 91.2%, for an average precision of 89.2%, and for an average recall of 92.5%. This means that using the K-Nearest Neighbor (kNN) method is very efficient and is the best method when used to carry out a Classification. So in determining the superior class of students/I, the K-Nearest Neighbor (KNN) method is very good to use. The superior class of students / I becomes one of the supporters to increase the potential of students / I to be able to compete to achieve good performance.Implementation of the K-Nearest Neighbor (kNN) Method to Determine Outstanding Student Classes.

REFERENCES

Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, *11*(3), 269–274. https://doi.org/10.33096/ilkom.v11i3.489.269-274

Farhad Khorshid, S., & Mohsin Abdulazeez, A. (2021). Breast Cancer Diagnosis Based on K-Nearest Neighbors: a Review. *Journal Of Archaeology Of Egypt/Egyptology*, *18*(4), 1927–1951.

*name of corresponding author

George, F. P., Shaikat, I. M., Ferdawoos Hossain, P. S., Parvez, M. Z., & Uddin, J. (2019). Recognition of emotional states using EEG signals based on time-frequency analysis and SVM classifier. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*(2), 1012. https://doi.org/10.11591/ijece.v9i2.pp1012-1020

Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447–459. https://doi.org/10.11591/ijeecs.v9.i2.pp447-459

Ibrahim, S., Zulkifli, N. A., Sabri, N., Shari, A. A., & Noordin, M. R. M. (2019). Rice grain classification using multi-class support vector machine (SVM). *IAES International Journal of Artificial Intelligence*, *8*(3), 215–220. https://doi.org/10.11591/ijai.v8.i3.pp215-220

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121. https://doi.org/10.22146/ijccs.65176

Jadhav, S. B., Udupi, V. R., & Patil, S. B. (2019). Convolutional neural networks for leaf image-based plant disease classification. *IAES International Journal of Artificial Intelligence*, *8*(4), 328–341. https://doi.org/10.11591/ijai.v8.i4.pp328-341

Kumar, A., Chatterjee, J. M., & Díaz, V. G. (2020). A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing. *International Journal of Electrical and Computer Engineering*, *10*(1), 486–493. https://doi.org/10.11591/ijece.v10i1.pp486-493

Kurniawan, Y. I., & Barokah, T. I. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor. *Jurnal Ilmiah Matrik*, *22*(1), 73–82. https://doi.org/10.33557/jurnalmatrik.v22i1.843

Liantoni, F. (2016). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *Jurnal ULTIMATICS*, *7*(2), 98–104. https://doi.org/10.31937/ti.v7i2.356

Lubis, A. R., Lubis, M., & Al-Khowarizmi. (2020). Optimization of distance formula in k-nearest neighbor method. *Bulletin of Electrical Engineering and Informatics*, *9*(1), 326–338. https://doi.org/10.11591/eei.v9i1.1464

Rosso, G. A. (2019). Milton. *William Blake in Context*, (September), 184–191. https://doi.org/10.1017/9781316534946.021

Saputra, M. F. A., Widiyaningtyas, T., & Wibawa, A. P. (2018). Illiteracy classification using K means-naïve bayes algorithm. *International Journal on Informatics Visualization*, *2*(3), 153–158. https://doi.org/10.30630/joiv.2.3.129

Wijaya, A., & Girsang, A. S. (2015). Use of Data Mining for Prediction of Customer Loyalty. *CommIT (Communication and Information Technology) Journal*, *10*(1), 41. https://doi.org/10.21512/commit.v10i1.1660

Zaki, S. Z. M., Zulkifley, M. A., Mohd Stofa, M., Kamari, N. A. M., & Mohamed, N. A. (2020). Classification of tomato leaf diseases using mobilenet v2. *IAES International Journal of Artificial Intelligence*, *9*(2), 290–296. https://doi.org/10.11591/ijai.v9.i2.pp290-296

Zheng, W. C., & Wu, X. X. (2006). Investigations of the spin Hamiltonian parameters for the trigonal Co 2+ center in ZnS0.001Se0.999 mixed crystal. *Optical Materials*, *28*(4), 370–373. https://doi.org/10.1016/j.optmat.2004.12.020

*name of corresponding author