Jurnal Mantik, 5 (2) (2021) 1557-1566

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Comparison Analysis Of Naïve Bayes And K-Nearest Neighbor Methods On The Prediction Of Academic Potential In Smk Ti Bali Global Badung

Ni Kadek Winda Patrianingsih[1], Komang Setemen[2], Dewa Gede Hendra Divayana[3]

[1,2,3]Program Studi Ilmu Komputer, Pascasarjana, Universitas Pendidikan Ganesha, Singaraja, Bali

E-mail: windaptr@gmail.com, k.sitemen@undiksha.ac.id, hendra.divayana@undiksha.ac.id

ARTICLE INFO

Article history:
Received: Jun 17, 2022
Revised: Jun 29 , 2022
Accepted: Jul 04, 2022

Keywords:
Data Mining,
Academic Potensial,
Naïve Bayes,
KNN

ABSTRACT

SMK TI Bali Global Badung is a vocational high school that focuses on information technology. SMK TI Bali Global Badung has an academic information system that is still conventional. The potential possessed by students can only be seen in terms of the value obtained, but supporting factors such as place of residence and physical condition can affect the potential of these students. Along with the development of information technology, data mining can provide solutions for schools to find academic potential based on stored data. This data mining system will summarise data from various data and information by analyzing specific patterns or relationships from several data types. Based on the current research results, this study will use the method to predict the academic potential of SMK TI Global Badung students, namely Naive Bayes and K-Nearest Neighbor. Prediction model of academic potential of SMK TI Bali Global Badung using Naive Bayes and KNN methods built using criteria and sub-criteria in the calculation process. In this study, seven criteria were used: character, academic activity, participation in intra-curricular activities, participation in extra-curricular activities, place of residence, and socioeconomic status. Based on the tests, Naive Bayes produces an accuracy value of 43.48%, a precision of 63.64, and a recall of 31.11. K-Nearest Neighbor creates an accuracy value of 57.97%, a precision of 62.5, and a recall of 88.89. In further research, it can also be considered to combine methods to increase the effectiveness of the predictions produced.

1. Introduction

Education has a significant and primary role in the progress of this nation. School is a bridge for students to deepen their knowledge in the world of education, and schools as a system to motivate students and develop their potential of each student. SMK TI Bali Global Badung is a vocational high school focusing on information technology, multimedia, software engineering, and network engineering. The Bali Global Badung IT Vocational School is located on Jalan Tibung Sari, Banjar Kwanji, Dalung, Badung Regency. This school has just turned four years old but already has many achievements in competitions at the provincial and national levels. It is undeniable that the development of information technology is a secondary need for its users.

The assessment and recording system at the SMK TI Bali Global Badung is still conventional. The school experiences problems in conducting further analysis of the potential of its students. Their achievements and activities still assess students' potential at SMK TI Bali Global Badung. At the same time, there is still an academic potential judged by grades, but the distance from home to school and an economic perspective also affect students' potential. Of course, this assessment system is not enough because it is not following the vision and mission of the Bali Global Badung IT Vocational School. One alternative that can be done to help managerial parties improve the quality of the Global Badung IT Vocational School is to take advantage of existing technological developments, especially data mining. This data mining system is

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

expected to provide input for the school in the future to find out the academic potential of its students based on the data that has been stored. Data mining[1] is obtaining or mining the required knowledge from a large amount of data. The process will summarise information or data from extensive raw data. Other fields of science related to data mining, namely Database Systems, Data Warehousing, Statistics, Machine Learning, Information Retrieval, and High-Level Computing. In addition, it is also supported by other sciences such as Neural networks, Pattern Recognition, Spatial Data Analysis, Image Database, and Signal Processing. Data mining is a grouping process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to obtain helpful information stored in large databases. Various data mining methods analyze and predict academic achievement, either using a single algorithm or combining several algorithms to get better results[2].

In this study, the methods used to predict the academic potential of SMK TI Global Badung students are Naive Bayes and K-Nearest Neighbor. The choice of this method is supported by[3], which states that students' understanding will increase with the support of multimedia images to accelerate student understanding. States that it is necessary to add some intermediate developments to the system that can be developed by providing information on the ranking of outstanding students with graphs so that it is easier to read the data[4]. In addition, [5] states that the results of the calculation of the Naïve Bayes data mining classification method show that the accuracy value of the Naïve Bayes method is quite good.

This research was conducted at SMK TI Bali Global Badung. Class X, XI, and XII are the training data used for all majors (Multimedia, Software Engineering, and Computer Network Engineering). The assessment carried out focuses on mastering the concept. The assessment is carried out with a minimum assessment standard and is also based on the assessment of each subject teacher. This assessment system, of course, is not enough because it is not following the mission of the Bali Global Badung IT Vocational School, with this data mining process is expected to produce students who are not only experts in the academic field but also have good morals. One alternative that can be done to help managerial parties improve the quality of the Bali Global Badung IT Vocational School is to take advantage of existing information technology developments, especially data mining system technology. This data mining system is expected to provide a benchmark for teachers to find out the personality of their students based on stored data, assess the teaching process's success, and determine what steps/decisions will be taken next based on the analysis of existing data. The assessment attributes include daily attitude, activeness in the academic field, organization participation, and diligence. Supporting attributes include the distance from home to school, extra-curricular activities, parents' income, and the student's health.

2. Method

2.1 Academic Potential

Academic potential is an ability possessed by someone born and needs to be developed to achieve achievement in the educational field. Education is a place for each individual to develop their potential. Through this, its potential will develop into potential. The role of educators is very influential in the intelligence of each individual[6]. Educators who do not understand each individual's intelligence will have difficulty facilitating the process of developing the potential of each individual[7].

2.2 Data Mining

Data mining is a process used to find a relationship, and so on will be used in selecting a large number of data which will then be stored in a database using various existing methods. In this process, data mining will produce an output in the form of information from many data through the analysis process of the data method that has been collected[8][9].

2.3 Naïve Bayes

Naive Bayes is one of the algorithms that can predict the membership of a class based on Bayes' theorem, which can work like decision trees and neural networks. The process of this algorithm is to perform grouping effectively by optimizing accurate possibilities. Even when the hypothesis is violated, this method is still better than other methods[10]. The Bayes Theorem equation can be seen in Equation 1[10] :

$$_{P}(H|E) = \frac{p(E|H)x\ p(H)}{p(E)} \quad (1)$$

Information:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

p(H|E) = hypothesis H occurs if evidence E occurs
p(E|H) = emergence of evidence E, if hypothesis H occurs
p(H) = hipote hypothesis H regardless of any evidence
P(E) = evidence E regardless of anything

2.4 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is one of the most common data mining methods [11]. This method is included in the lazy learner grouping because it delays the creation process until there is test data that you want to know the functionality of the class label. Then this method can only be run by the algorithm. Working on this algorithm is by grouping based on the similarity of one data to another. In this study, the author only uses the Euclidean distance, then the formula for calculating the distance with Euclidean is as follows[12]can be seen in Equation 2:

$$\sqrt{\sum_{i=1}^{K}(Xi - Yi)}^{2} \quad (2)$$

Information:
Xi value = value of training data
Yi value = value of test data
K value = dimension attribute

The steps in the process of calculating the K-NN algorithm:

a. Determine the value of k.
b. Calculate the square of the Euclidean distance (query instance) of each object to the given training data.
c. Sort the objects into groups that have the smallest Euclidean distance.
d. Collecting class Y labels (Nearest Neighbor classification).
e. Using the Nearest Neighbor algorithm, the calculated query instance value can be predicted.

The measurement of performance indicators used in this study is the Confusion matrix by conducting tests to estimate the right and wrong objects[13].

TABLE 1
Convolution Matrix

	Actual Value	
Prediction Value	TP	TN
	FP	FN

Information:
TP = True Positive
TN= True Negative
FB = False Positive
FN = False Negative

The formula for calculating the confusion matrix is written as below:

a. Precision is useful for measuring the level of accuracy between the information requested by the user and the answer given by the system with the Equation (3):
Pre = TP / TP + FP (3)
b. Recall is useful for measuring the level of success of the system in retrieving an information, in the Equation (4):
Re = TP / TP + FN (4)
c. Accuracy is useful for measuring the performance of a method, with the Equation (5):
Acu =TP + TN/ TP + TN+ FP+ FN (5)

3. Result and Discussion

3.1 Data Analysis

This study requires student data as the primary data source in the study. The data used is active student data in classes X, XI, and XII at the SMK TI Global Badung for the 2018-2020 Academic Year. Data for students majoring in Multimedia (MM), Software Engineering (RPL), and Computer Network Engineering

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

(TKJ). The grouping of data is also adjusted according to the major, and In the calculation process, the data will be divided into 2, namely training data using 2018 and 2019 data and testing data using 2020 batch data. Used in this study revealed 213 students with details as shown in Table 2 below:

TABLE 2
Studen Population

No	Force	Major	Total Student
1	2018	MM	34
		TKJ	18
		RPL	17
2	2019	MM	32
		TKJ	11
		RPL	15
3	2020	MM	40
		TKJ	26
		RPL	20
Total			213

3.2 Analysis of Criteria and Sub-Criteria

Prediction analysis of the academic potential of SMK TI Bali Global Badung using the Naive Bayes method and KNN requires criteria and sub-criteria in the calculation process. The assessment attributes used Character (A1), Academic Activity(A3), Intra-Curricular Participation(A3), and Extra-Curricular Participation (A4). Supporting attributes: Residence(A5) and Socioeconomic Status(A6). The criteria used in the calculation, namely the existing sub-criteria, also need to be defined in the range of weight values. The criteria and sub-criteria and their range of weight values can be seen in Table 3 below:

TABLE 3
Criteria and Sub-Criteroa

Criteria	Sub-Criteria	Value
Character	A (Very Good)	5
	B (Good)	4
	C (Enough)	3
	D (Not Good)	2
	E (Bad)	1
Academic Activity	Active	5
	Currently	3
	Not Active	1
Intra-Curricular Participation	Active	2
	Not Active	1
Extra-Curricular Participation	A (Very Good)	5
	B (Good)	3
	C (Enough)	1
Residence	Near	5
	Medium	3
	Far	1
Sosioeconomic Status	Tall	5
	Medium	3
	Low	1

3.3 Naïve Bayes Calculations

The problems that will be solved here are problems related to the prediction of students' academic potential based on the data that has been collected. Suppose there is a sample of 10 student data from 2 majors in the class of 2018, which is presented in Table 4:

TABLE 4
Studen Data sample

Code	**Attribute**						
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**Potential**
TKJ01	B	Active	Active	B	Medium	Tall	Yes
TKJ02	B	Active	Not Active	B	Medium	Medium	Yes
TKJ03	B	Currently	Not Active	B	Medium	Low	No

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

TKJ04	B	Active	Not Active	B	Medium	Medium	Yes
TKJ05	B	Active	Not Active	B	Medium	Medium	Yes
RPL01	C	Not Active	Not Active	C	Near	Tall	No
RPL02	C	Not Active	Not Active	C	Medium	Low	No
RPL03	B	Not Active	Not Active	C	Far	Medium	No
RPL04	B	Active	Active	B	Far	Tall	Yes
RPL05	B	Active	Active	B	Far	Tall	Yes

The following process determines the sample test data to be calculated using the Naïve Bayes method. The testing stage will require test data to produce decisions related to academic potential, and the training data table can be seen in Table 5 as follows.

TABLE 5
Simple Test Data

Code	Attribute						Potential
	A1	A2	A3	A4	A5	A6	
TKJ06	C	Not Active	Not Active	B	Near	Medium	?

The following are the stages of solving this problem using the Naïve Bayes algorithm based on Equation (1):

1. **Stage 1**, Counting the Number of Classes/Label
   - P(Y=YES) = 6/10
   the number of "Yes" data in the "Potential" column in Table (4) divided by the number of data.
   - P(Y=NO) = 4/10
   the number of "Yes" data in the "Potential" column in Table (4) divided by the number of data.
2. **Stage 2**, Counting the Number of Cases with the Same Class
   - P(Character = A | Y = YES) = 0/6
   - P(Character = A | Y = NO) = 0/4
   - P(Character = B | Y = YES) = 6/6
   - P(Character = B | Y = NO) = 2/4
   - P(Character = C | Y = YES) = 0/6
   - P(Character = C | Y = NO) = 2/4
   - P(Character = D | Y = YES) = 0/6
   - P(Character = D | Y = NO) = 0/4
   - P(Character = E | Y = YES) = 0/6
   - P(Character = E | Y = NO) = 0/4
   - P(Academic Activity = Active | Y = YES) = 6/6
   - P(Academic Activity = Active | Y = NO) = 0/4
   - P(Academic Activity = Currently | Y = YES) = 0/6
   - P(Academic Activity = Currently | Y = NO) = 1/4
   - P(Academic Activity = Not Active | Y = YES) = 0/6
   - P(Academic Activity = Not Active | Y = NO) = 3/4
   - P(Intra-Curricular Partipation = Active| Y = YES) = 3/6
   - P(Intra-Curricular Partipation = Active | Y = NO) = 0/4
   - P(Intra-Curricular Partipation = Not Active | Y = YES) = 3/6
   - P(Intra-Curricular Partipation = Not Active | Y = NO) = 4/4
   - P(Extra-Curricular Partipation = A | Y = YES) = 0/6
   - P(Extra-Curricular Partipation = A | Y = NO) = 0/4
   - P(Extra-Curricular Partipation = B | Y = YES) = 6/6
   - P(Extra-Curricular Partipation = B | Y = NO) = 1/4
   - P(Extra-Curricular Partipation = C | Y = YES) = 0/6
   - P(Extra-Curricular Partipation = C | Y = NO) = 3/4
   - P(Residence = Near | Y = YES) = 0/6
   - P(Residence = Near | Y = NO) = 1/4
   - P(Residence= Medium | Y = YES) = 4/6
   - P(Residence= Medium| Y = NO) = 2/4
   - P(Residence= Far | Y = YES) = 2/6

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

- P(Residence= Far | Y = NO) = 1/4
- P(Sosioeconomic Status = Tall | Y = YES) = 3/6
- P(Sosioeconomic Status = Tall | Y = NO) = 1/4
- P(Sosioeconomic Status = Medium | Y = YES) = 2/6
- P(Sosioeconomic Status = Medium | Y = NO) = 1/4
- P(Sosioeconomic Status = Low | Y = YES) = 1/6
- P(Sosioeconomic Status = Low | Y = NO) = 2/4

3. **Stage 3**, Multiplying All YES and NO Variable Results according to the Problem to be Solved

The next step is to multiply these variables to get the value of each class from the predictions of potential students.

YES

R21 = P(Character = C | Y = YES) * P(Academic Activity = Not Active | Y = YES) * P(Intra-Curricular Partipation= Not Active | Y = YES) * P(Extra-Curricular Partipation = B | Y = YES) * P(Residence = Near | Y = YES) * P(Socioeconomic Status = Medium | Y = YES) * P(Y=YES)
= 0/6 * 0/6 * 3/6 * 6/6 * 0/6 * 2/6 * 6/10
= **0**

NO

R21 = P(Character = C | Y = NO) * P(Academic Activity = Not Active | Y = NO) * P(Intra-Curricular Partipation= Not Active | Y = NO) * P(Extra-Curricular Partipation = B | Y = TIDAK) * P(Residence = Near | Y = No) * P(Socioeconomic Status = Medium | Y = NO) * P(Y=NO)
= 2/4 * 3/4 * 4/4 * 1/4 * 1/4 * 1/4 * 4/10
= **0.00234375**

4. **Stage 4**, compare the results of the YES and NO grades. Because the result (P|NO) is more significant than (P|YES), the conclusion that can be made from this test data is that this student **does NOT** have academic potential.

3.4 K-Nearest Neighbour (KNN) Calculations

The problems that will be solved here are problems related to the prediction of students' academic potential based on the data that has been collected. For example, a sample of 10 student data from 2 majors in the 2018 batch is presented in Table (4). The alternative data in Table (4) is then converted to value based on Table (3) provisions. The conversion of alternative data sub-criteria values can be seen in Table 6 below:

TABLE 6
Conversion Of Almternative Data Sub- Criteria Values

Code	Attribute						Potential
	A1	A2	A3	A4	A5	A6	
TKJ01	4	5	2	3	3	5	Yes
TKJ02	4	5	1	3	3	1	Yes
TKJ03	4	3	1	3	3	1	No
TKJ04	4	5	1	3	3	3	Yes
TKJ05	4	5	1	3	3	3	Yes
RPL01	3	1	1	1	5	5	No
RPL02	3	1	1	1	3	1	No
RPL03	4	1	1	1	1	3	No
RPL04	4	5	2	3	1	1	Yes
RPL05	4	5	2	3	1	1	Yes

The testing stage will require test data to be able to produce decisions related to academic potential. The test data uses the data in Table (5). The value of the test data with the code TKJ06 is then converted according to Table (6), resulting in a value conversion that can be seen in Table 7 below:

TABLE 7
Conversion Test Data Sub-Criteria Value

Code	Attribute					
	A1	A2	A3	A4	A5	A6
TKJ06	3	1	1	3	5	3

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Euclidean distance calculation using Equation (2). Suppose that the 3 nearest neighbors (K=3) are used to determine the class of the new data given. Furthermore, the calculation of the distance between the existing dataset and the new data provided by Euclidean distance is carried out, with the results shown in Table 8 below:

TABLE 8
Euclidean Distance Calculation Results

Student Code	Euclidean Distance	Distance Order
TKJ01	5.09901951	8
TKJ02	5	7
TKJ03	3.60555128	3
TKJ04	4.58257569	4
TKJ05	4.58257569	5
RPL01	2.82842712	1
RPL02	3.46410162	2
RPL03	4.58257569	6
RPL04	6.164414	9
RPL05	6.164414	10

From the calculated distance, it is obtained that the student data RPL01, RPL02, and TKJ03 are the three closest neighbors where the data class of students RPL01, RPL02, and TKJ03 is NO. Because of the three selected neighbors, all neighbors have a label NO, and the new data is estimated to be labeled NO so that the latest data for TKJ06 students are included in the class of students with no academic potential.

3.5 System Implementation Results

Based on the analysis of criteria and calculations that have been carried out, a web-based prediction system for academic potential was built. The following is the resulting implementation interface.

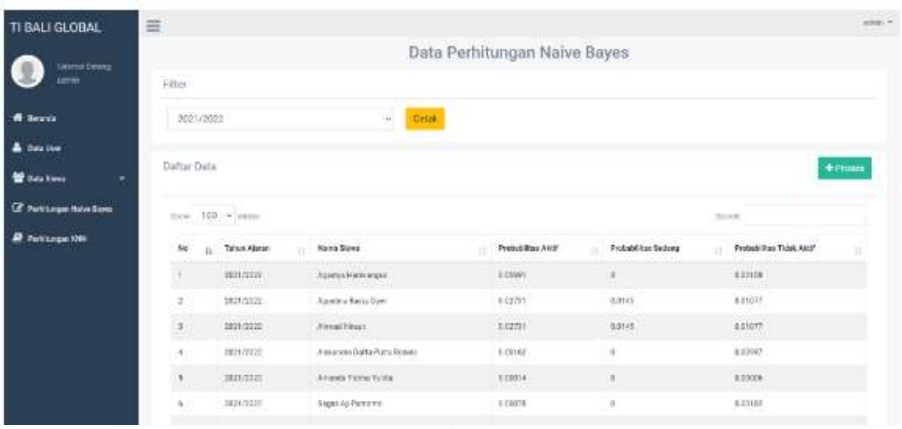

Figure 1. Results of Naïve Bayes Implementation

Figure 1 is a system interface display showing Naïve Bayes calculation data. The implemented system will provide the functionality to enter test data from an excel file. Furthermore, the Naïve Bayes calculation will be carried out to produce the final result in the form of a category of student academic potential.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

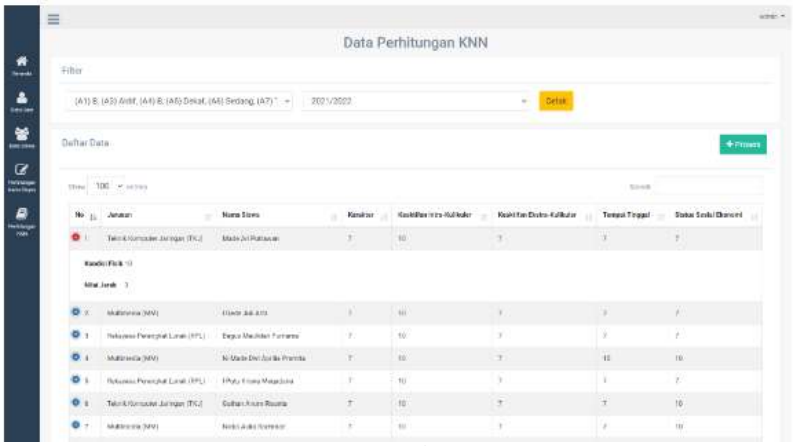

Figure 2. Results of KNN Implementation

Figure 2 is a system interface display that shows the KNN calculation data. The calculation process flow in the system has been adjusted to the manual calculation process described previously. The final results generated from these two methods are then tested and analyzed using a confusion matrix

3.6 Method Test Results

Based on the results of the implementation that has been done, then the testing process is carried out using a confusion matrix. Testing is done by comparing the results given by the system with the results of the assessment provided by the school. The test in this study uses a confusion matrix where the results of each method will be tested how much the accuracy, precision, and recall values are following Equation (3), Equation (4), and Equation (5). The Confusion Matrix table can be seen in Table 9 below:

TABLE 9
Confusion Matrix

Method		
Actual	Predicted	
	Potential	Not Potential
Potential	True Positive (TP)	False Negative (FN)
Not Potential	False Positive (FP)	True Negative (TN)

Information :

a. *True Positive (TP):* the number of students predicted to have academic potential and truly have academic potential.
b. *True Negative (TN):* the number of students predicted to have no academic potential; in fact, these students do not (true) have academic potential.
c. *False Positive (FP):* the number of students who are predicted to have academic potential but do not have academic potential.
d. *False Negative (FN):* the number of students who are predicted to have no academic potential but have academic potential.

Based on the comparison of the results of the Naïve Bayes Classifier method with the actual results of existing students' academic potential, a confusion matrix table is generated, which can be seen in Table 10 below:

TABEL 10
Confusion Matrix Naïve Bayes

Naïve Bayes		
Actual	*Predicted*	
	Potential	Not Potential
Potential	14	31
Not Potential	8	16

a. $Accuracy\ (\%) = \frac{14+16}{69} = 43{,}48\ \%$

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The closeness of the predicted value to the original value was at the level of 43.48%.

b. *Precision* (%) = $\frac{14}{14+8}$ = 63,64 %

The true positive prediction compared to the overall positive predicted outcome was at the level of 63.64%.

c. $Recall\ (\%) = \frac{14}{14+31}$ = 31,11 %

The measure of the accuracy of the desired event is at the level of 31.11%.

Based on comparing the results of the KNN method with the actual results of the existing academic potential of students, a confusion matrix table is generated, which can be seen in Table 11.

TABEL 11
Confusion Matrix Knn

KNN		
Actual	*Predicted*	
	Potential	Not Potential
Potential	40	5
Not Potential	24	0

a. $Accuracy\ (\%) = \frac{40+0}{69}$ = 57,97 %

The closeness of the predicted value to the original value was at the level of 57,97%.

b. *Precision* (%) = $\frac{40}{40+24}$ = 62,5 %

The true positive prediction compared to the overall positive predicted outcome was at the level of 62,5%.

c. $Recall\ (\%) = \frac{40}{40+5}$ = 88,89 %

The measure of the accuracy of the desired event is at the level of 88,89%.

Actually, of the 69 student data tested, 45 student data have academic potential. Based on these data, a comparison is made with the predicted results from the system that has been successfully implemented. Tests are carried out using a confusion matrix, a matrix of predictions that will be compared with the original input data class. The data will be compared with the results of the classification carried out by the system. After testing the Naïve Bayes method, the accuracy value is 43.48%, while KNN produces an accuracy value of 57.97%. The resulting accuracy shows the closeness of the prediction to the actual data of the student's academic potential. So the conclusion that can be obtained from this accuracy test is that KNN produces better accuracy data in predicting students' academic potential. However, it can be seen that the percentage of accuracy made can be increased. Based on the study, a combination of methods can be carried out at the initial stage of determining the weight or value of the criteria used. Lecturer performance assessment was successfully carried out by combining the Profile Matching and TOPSIS methods[14]. In this study, weighting is carried out on alternative data to be processed. In other studies, combinations can be used to determine the primary weight [15].

4. Conclusion

The prediction model for the academic potential of SMK TI Bali Global Badung students using the Naive Bayes and KNN methods was built using criteria and sub-criteria in the calculation process. This study used seven criteria: character, academic activity, participation in intra-curricular activities, extra-curricular activities, place of residence, and socioeconomic status. Existing sub-criteria also need to be defined in the range of weight values. Based on the results of the tests, Naive Bayes produces an accuracy value of 43.48%, a precision of 63.64, and recall of 31.11. K-Nearest Neighbor produces an accuracy value of 57.97%, a precision of 62.5, and a recall of 88.89. The test results show that K-Nearest Neighbor has better-predicted students' academic potential at SMK TI Bali Global Badung. Suggestions are based on the study conducted. A combination of methods can be carried out at the initial stage of determining the weight or value of the criteria used.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

References

[1] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014. doi: 10.1109/ICMIRA.2013.45.

[2] H. Kurniawan, I. R. Munthe, and B. Bangun, "Analysis of Naive Bayes Methods on Scholarship Admissions at Smks Al-Wasliyah 2 Merbau," *J. Mantik*, vol. 6, no. 1, pp. 364–373, 2022.

[3] H. Basri, "Implementasi Algoritma Naive Bayes Untuk Mengetahui Potensi Siswa Dalam Pemahaman Ms. Excel," *Sistemasi*, vol. 8, no. 3, p. 505, 2019, doi: 10.32520/stmsi.v8i3.569.

[4] J. I. Kartika, E. Santoso, and Sutrisno, "Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus: SMP Negeri 3 Mejayan)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 5, pp. 352–360, 2017.

[5] D. Yunita, P. Rosyani, and R. Amalia, "Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes," *J. Inform. Univ. Pamulang*, vol. 3, no. 4, p. 209, 2018, doi: 10.32493/informatika.v3i4.2032.

[6] S. Suryabrata, "Psikologi pendidikan," 2011.

[7] A. Amir, "Pembelajaran Matematika Dengan Menggunakan Kecerdasan Majemuk (Multiple Intelligences)," *Logaritma*, vol. I, no. 01, pp. 1–14, 2013.

[8] N. Nuraeni, "Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC," *J. Tek. Komput. AMIK BSI*, vol. 3, no. 1, pp. 9–15, 2017.

[9] T. S. Kumar, "Data mining based marketing decision support system using hybrid machine learning algorithm," *J. Artif. Intell.*, vol. 2, no. 03, pp. 185–193, 2020.

[10] P. Eko, *Data mining : konsep dan aplikasi menggunakan MATLAB*. CV Andi Offset, 2012.

[11] S. Dewi, "KOMPARASI 5 METODE ALGORITMA KLASIFIKASI DATA MINING PADA PREDIKSI KEBERHASILAN PEMASARAN PRODUK LAYANAN PERBANKAN," *Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 60–66, 2016.

[12] S. Rahmatullah, S. Wahyuni, M. F. Chaining, and F. C. Method, "Penerapan Data Mining Untuk Prediksi Penjualan Produk," no. 2, pp. 75–86, 2020.

[13] A. A. Nababan, M. Khairi, and B. S. Harahap, "Implementation of K-Nearest Neighbors (KNN) Algorithm in Classification of Data Water Quality," *J. Mantik*, vol. 6, no. 1, pp. 30–35, 2022.

[14] I. P. Suarnatha and I. G. Gunadi, "COMBINATION OF THE PROFILE MATCHING AND TOPSIS METHOD IN DECISION SUPPORT SYSTEM OF LECTURER PERFORMANCE ASSESSMENT," *J. Ris. Inform.*, vol. 3, no. 3 SE-Articles, Jun. 2021, doi: 10.34288/jri.v3i3.237.

[15] A. A. S. Widhiyanti, I. M. Candiasa, and K. Y. E. Aryanto, "IMPLEMENTASI AHP-TOPSIS DAN NAÏVE BAYES DALAM SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN BIMBINGAN KONSELING SISWA," *SINTECH (Science Inf. Technol. J.*, vol. 4, no. 2 SE-, pp. 129–138, Oct. 2021, doi: 10.31598/sintechjournal.v4i2.731.