Journal of Electrical Engineering and Computers (JEECOM)
Vol. 7, No. 1 (2025), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Classification Of Direction Using Naive Bayes Classifier Method (Case Study Of Hidayatul Islam Leces Vocational School)

Dwi Yanto[1], Heri Susanto[2], Ninanesia[3], Kiki Zulkifli[4]
[12]Information System, Taruna Informatics and Computer Management Academy, Probolinggo, Indonesia
[34]Accounting Information System, Taruna Informatics and Computer Management Academy, Probolinggo, Indonesia

Article Info

Article history:

Received April, 24 2025
Revision April 28 2025
Published April 29, 2025

Keywords:

Classification Student Choice Of Majors
Naive Bayes Classifier
Vocational School Hidayatul Islam

ABSTRACT

Abstract— Determining student choice of majors is an important process in the world of education that can affect students' future. In this research, we conducted a study on determining student choice of majors using the Naive Bayes Classifier algorithm at Vocational School Hidayatul Islam. The purpose of this study was to test the accuracy of the Naive Bayes Classifier algorithm in predicting student choice of majors and to provide recommendations that can support decision making in determining student majors. This study uses historical data of Vocational School Hidayatul Islam students which includes various attributes such as academic grades, Mathematics, Science, Language, Science, and Average report card grades. The data was processed and trained on the Naive Bayes Classifier algorithm using supervised learning methods. Furthermore, the algorithm was tested using separate test data. The results showed that the Naive Bayes Classifier algorithm provided an accuracy of 97.50% in determining student choice of majors at Vocational School Hidayatul Islam. This shows a very good ability to predict student choice of majors based on existing attributes. With high accuracy, this algorithm can be an effective tool in helping the student choice of major decision-making process. However, it should be noted that the results of this study need to be considered in the specific context of Vocational School Hidayatul Islam and the characteristics of its students. Factors such as students' interests and talents, parents' views, and job market needs should also be important considerations in determining students choice of majors. Therefore, the Naive Bayes Classifier algorithm should be used as one component in a broader decision-making process, which involves consideration of these various factors.

Corresponding Author:
Dwi Yanto
Taruna Informatics and Computer Management Academy, Jl. Raya Leces No.A3 Leces, Probolinggo 67202, Indonesia
Email: dwiyanto.atp@gmail.com

1. INTRODUCTION

Education is the main foundation in shaping the quality of individuals and society. Therefore, the education process needs to be continuously improved, both from cognitive and administrative aspects. One of the important stages in secondary education is majoring, which is the process of placing students in expertise programs according to their interests, talents, and academic abilities. The right majoring will encourage students to develop more and be comfortable in the learning process, while the wrong majoring can hinder students' potential.[3]

SMK Hidayatul Islam Leces, which was established in 2013 under the auspices of the Hidayatul Islam Education and Islamic Boarding School Foundation, has two departments: Computer and Network Engineering (TKJ), and Office Automation and Management (OTKP). The department selection process at this school is still done manually using interest questionnaires and student report card scores from previous schools. This process is considered less than optimal because it is still less objective and quite time-consuming.

To improve efficiency and accuracy in the majoring process, this study proposes the application of the Naive Bayes Classifier method. This method is a probability-based classification technique that can predict majors that match students' academic characteristics, such as Indonesian, Social Studies, Science, and Mathematics scores. With this approach, the majoring process can be carried out more quickly, precisely, and based on data.[7]

This research is expected to contribute to the development of a more effective new student majoring system, as well as assist schools in making more accurate decisions for the placement of expertise programs at SMK Hidayatul Islam Leces.

2. LITERATURE STUDY

2.1 Data Mining

Data Mining is a series of processes to extract added value from a data set in the form of knowledge that has not been known manually. Data Mining is the automatic analysis of large or complex data with the aim of finding important patterns or trends that are usually not realized. Data Mining is the analysis of reviewing data sets to find unexpected relationships and summarize data in a different way than before, which can be understood and is useful for data owners. Data Mining is a field of several scientific fields that combines techniques from machine learning, pattern recognition, statistics, databases, and visualization to handle the problem of retrieving information from large databases.[11]

2.2 Classification

A technique that looks at the behavior and attributes of a defined group. This technique can provide classification to new data by manipulating existing data that has been classified and using the results to provide a number of rules. One easy and popular example is the Decision tree, which is one of the most popular classification methods because it is easy to interpret. Decision tree is a prediction model using a tree structure or hierarchical structure.[1]

2.3 Naive Bayes

The classification method using probability and statistical methods proposed by British scientist Thomas Bayes, namely predicting future opportunities based on previous experiences, is known as Bayes' Theorem. The main characteristic of this Naive Bayes Classifier is a very strong (naive) assumption of the independence of each condition / event.[15]

2.4 Rapid Miner

RapidMiner is a solution for analyzing data mining, text mining and predictive analysis. RapidMiner uses various descriptive and predictive techniques to provide insights to users so they can make the best decisions. RapidMiner has approximately 500 data mining operators, including operators for input, output, data preprocessing and visualization. RapidMiner is a stand-alone software for data analysis and as a data mining engine that can be integrated into its own products. RapidMiner is written using the Java language so it can work on all operating systems.[13]

3. METHOD

In research, references to theories and findings from previous research are very important to provide valid supporting data. One source of supporting data that needs to be considered is previous research that is relevant to the current research topic. In this case, the previous research that is focused on is related to the problem of determining student majors. Therefore, the researcher conducted a study of several research results that include:

Table 1. Relevant Research

No.	Title	Method	Results
1	Implementation of C4.5 Algorithm for Senior High School Students' Majors (Case Study: SD Negeri 9 Air Kumbang)	C4.5	The majoring in SMA 6 Surakarta is done when registering new students, before the students are accepted as class X students. The large number of registrants with a manual system results in errors and long data processing. So this study aims to develop a system to classify majors for students accurately, effectively and efficiently to determine science, social studies or language majors. The

ISSN:2715-6427

			C4.5 algorithm is used to find rule patterns based on supporting variables in the form of average Junior High School (SMP) report card scores, science, social studies and Indonesian language test scores. And the resulting accuracy rate is 97.42%.
2	Application of Naïve Bayes Classifier Method for Student Majors at Al-Falah Islamic High School, Jakarta	Naïve Bayes Classifier	Madrasah Aliyah Al-Falah has three departments, namely Religious Sciences (IIK), Mathematics Natural Sciences (MIA), and Social Sciences (IIS). The majoring mechanism at MA Al-Falah is carried out by conducting four types of tests, namely mathematics, Indonesian, English, and religion on students. The results of these tests are used as the basis for determining students' majors. The school only relies on data processing and sorting with Microsoft Excel. In a study conducted by Ahmad Zainul Mafakhir and Achmad, the Naïve Bayes Classifier method was applied to classify test scores and produce recommendations for student majors. The student majoring classification system that was developed can help the student majoring process more easily, quickly, and accurately. Based on the tests that have been carried out, the student majoring system can provide recommendations for student majors with an accuracy value of 33.34% (Mafakhir & Achmad, 2020)
3	Classification of Vocational High School Major Selection Using Gradient Boosting Classifier	Gradient Boosting Classifier	In this study, a classification model is proposed to predict school majors with students' abilities and interests. This study compares five classifiers on a dataset of major selection with the RIASEC model which maps student interests into realistic, investigative, artistic, social, active, and conventional. The next process in this study is to perform hyperparameter tuning using GridsearchCV to obtain the most influential parameters of the selected classification algorithm.

Of course, this research at SMK Yadika 12 Depok aims to predict major choices. The algorithms used in this study are Multinomial Naïve Bayes, Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Gradient Boosting Classifier, Decision Tree Classifier, K Neighbors Classifier, and Logistic Regression. Where in this study shows the results that the Gradient Boosting Classifier with Hyperparameter Tuning using GridSearchCV obtains an accuracy of 72% and class recall reaches 76%.

This research framework is a method for collecting data regularly so that it can avoid the validity of data information, in this research framework it can design efficiently and in detail in collecting data. In collecting data, namely by using qualitative methods, namely by conducting direct observations and interviews so that it can sweeten the data needed in the study. The following is a description of the research framework

The framework for this research is as follows :

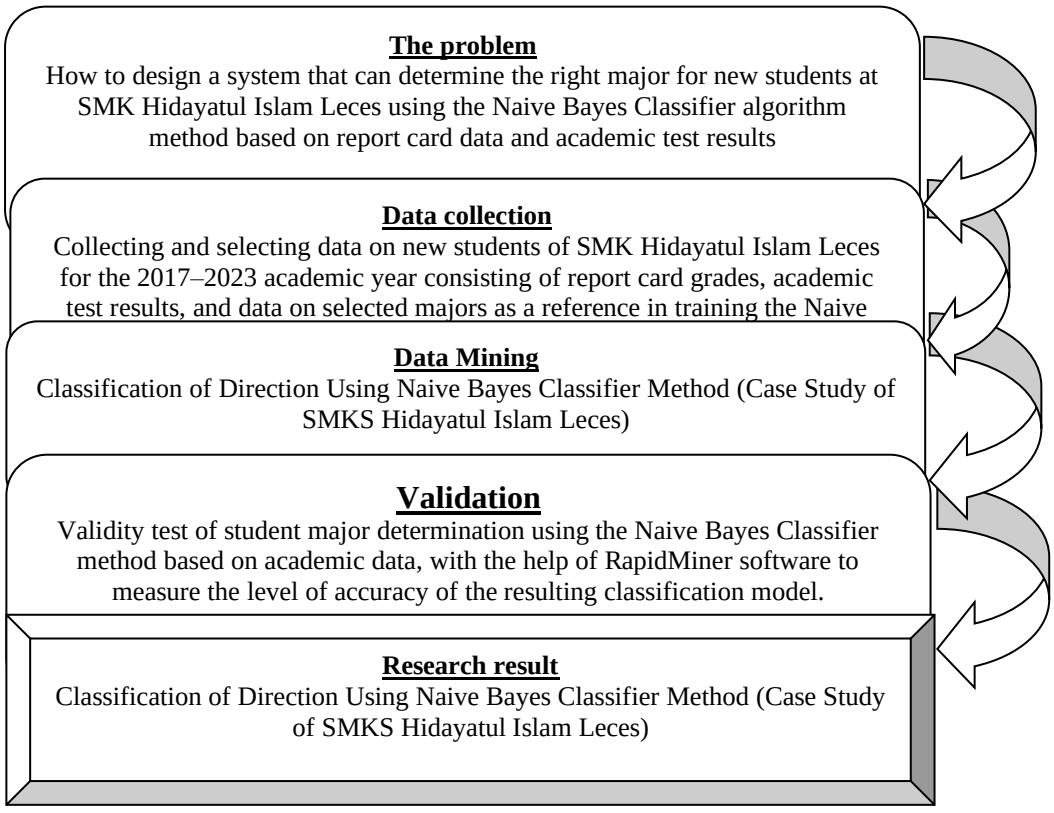

Figure 1. Research Framework

Channel The stages in this research are as follows :

ISSN:2715-6427

1. Selection
   The initial dataset was taken from the report card data of junior high school/Islamic junior high school students and the academic test results of prospective new students of SMKS Hidayatul Islam in the previous academic year. The data was collected and selected by the New Student Admissions (PPDB) committee. The variables used in this study include report card scores for the subjects of Indonesian, Mathematics, Science, Social Studies, and the results of the academic test for entering SMK.
2. Preprocessing
   This stage is carried out to prepare the dataset to be suitable for use in the classification process. Some of the preprocessing processes carried out include filling in empty data (if any), normalizing values, coding class categories (TKJ and OTKP), and checking data consistency. With this preprocessing, the data becomes cleaner and ready to be entered into the Naïve Bayes model.
3. Naïve Bayes
   At this stage, the classification process is carried out using the Naïve Bayes algorithm. This algorithm will process the values of student attributes (such as report card grades and academic test results) to estimate the most suitable majors for each student. Naïve Bayes works based on the principle of probability and the relationship between variables that are assumed to be independent. The classification results will show the recommended majors for each student, whether Computer and Network Engineering (TKJ) or Office Management Automation (OTKP).

4. RESULTS AND DISCUSSION

1. Selection Stage
   As explained previously, the dataset in this study was taken from report card data and academic test results of prospective new students of SMKS Hidayatul Islam Leces for the 2017–2023 academic year. Data was obtained from PPDB archives and student academic documents.
   a. Data Collection
   b. After the data is collected, a cleaning process is carried out to remove irrelevant or incomplete data. Not all variables are used in this study, only variables that influence the determination of the chosen major, namely academic values. From the available data, 395 data are used as the population, and 25 data are used as samples for analysis and classification testing using the Naïve Bayes Classifier method.:

Tabel. 2 Dataset

NO	NAME	MTK	SCIENCE	LANGUAGE	IPS	AVERAGE	MAJOR
1	Abdullah	86	94	73	72	70	TKJ
2	Ahmad Dae Robi	90	87	75	75	79	TKJ
3	Aminatul Rohmah	92	80	76	75	90	TKJ
4	Aminatus Sahro	95	82	78	88	89	TKJ
5	Amri Yahya	80	82	70	77	70	TKJ
6	Aprilia Valentina	83	92	78	80	82	TKJ
7	The Story of Usmawati	71	77	86	94	77	OTKP
8	Ayu Fitria	72	78	90	87	89	OTKP
9	Eka Yuniar Sri Utami	70	77	92	80	76	OTKP
10	Elijah Ramadani	71	78	80	82	88	OTKP
11	The Age of Fazira	69	77	83	92	77	OTKP
12	English: Febri Nur Komariyah	78	81	90	88	87	TKJ
13	Fendi Febrianto	94	89	70	72	72	TKJ
14	The Wisdom of Nurhasanah	70	78	78	81	70	OTKP
15	Khofifa	85	88	92	88	89	TKJ

16	Laili Nuroh	68	77	94	89	79	OTKP
17	Lailul Fijriah	69	78	85	88	90	OTKP
18	Mohd Jamal	93	84	75	72	74	TKJ
19	Moh. Zaenal Abidin	79	91	86	81	84	TKJ
20	Muhammad Andika	88	93	89	85	87	TKJ
21	Museum	81	95	78	75	76	TKJ
22	Princess Wahyuni Hasanah	67	77	93	84	70	OTKP
23	Ririn Andriyana	68	78	79	91	82	OTKP
24	Saiful Islam	95	78	94	92	91	TKJ
25	Sherly Angela Wibowo	84	93	70	68	66	TKJ

2. Preprocessing

This stage is used to prepare the dataset to have better quality and be effective before being modeled. The data that has been collected from the report card scores and academic test results of students is first checked for completeness, then the cleaning process is carried out on incomplete data or data containing input errors.

Some of the preprocessing steps carried out in this study include:

Empty data check: Identifies and removes entries that do not have complete values.

Normalization of values: Adjusting the scale of values to be uniform in the same range, to facilitate the classification process.

Major label encoding: Converting majors into categorical formats such as “TKJ” and “OTKP” to facilitate processing in the classification system.

Data grouping: Grouping students based on academic attributes to make it easier to analyze according to relevant major categories.

With this preprocessing process, the dataset becomes more structured and can represent the actual conditions in the major selection process at Vocational School Hidayatul Islam.

3. Naive Bayes Classifier Implementation and Calculation Results

a. Analysis and Design Results

The results of the analysis and design in this section of the thesis explain how the analysis is carried out to solve the existing problem formulation by designing it according to your wishes as a form of solution in solving the problem. The tool used to build the model is rapidminer version 10.1. The dataset is uploaded to the rapidminer application in the form of Excel format which is then determined the type of data for each attribute or output class label shown in Figure 2.

ISSN:2715-6427

	MTK *polynominal*	IPA *polynominal*	BAHASA *polynominal*	IPS *polynominal*	RATA-RATA *polynominal*	JURUSAN *binominal label*
1	B	A	C	C	C	TKJ
2	A	B	C	C	C	TKJ
3	A	B	C	C	A	TKJ
4	B	B	C	C	C	TKJ
5	B	A	C	B	B	TKJ
6	C	B	A	B	B	TKJ
7	A	B	C	C	C	TKJ
8	B	B	A	B	B	TKJ
9	A	B	C	C	C	TKJ
10	C	A	B	B	B	TKJ
11	B	A	B	B	B	TKJ
12	B	A	C	C	C	TKJ

Figure 2. Snippet of the uploaded dataset in rapidminer

The caption for Figure 3 is as follows:
All attributes related to grades are of the “polynominal” type, starting from grades for Mathematics, Science, Language, Social Studies and the average, while majors are of the “binominal” type.

Name	Type	Missing	Statistics		Filter (6 / 6 attributes): Search for Attributes
Label **JURUSAN**	Binominal	0	Negative TKJ	Positive OTKP	Values TKJ (223), OTKP (172)
MTK	Nominal	0	Least D (17)	Most B (172)	Values B (172), C (125), ...[2 more]
IPA	Nominal	0	Least D (7)	Most B (160)	Values B (160), C (134), ...[2 more]
BAHASA	Nominal	0	Least D (13)	Most B (170)	Values B (170), C (148), ...[2 more]
IPS	Nominal	0	Least D (3)	Most C (205)	Values C (205), B (114), ...[2 more]
RATA-RATA	Nominal	0	Least D (7)	Most B (178)	Values B (178), C (176), ...[2 more]

Figure 3. Statistics of rapidminer output dataset

b. Naive Bayes Classifier Method

In the Naïve Bayes Classifier method, it has been explained from the explanation and advantages of this method in the previous chapter. This study believes that the Naïve Bayes Classifier method has high accuracy because it has the same purpose as the previous study. Figure 4 shows the configuration of the naive bayes algorithm model with the rapidminer 10.1 application.

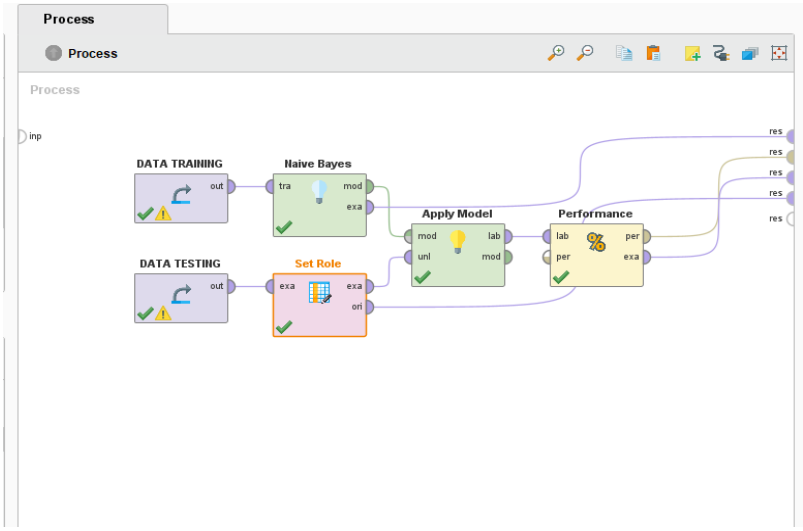

Figure 4. Algorithm model building configuration *naive bayes*

Figure 5. shows the resulting naive Bayes model in the form of a description of the class distribution on the major attribute.

SimpleDistribution

```
Distribution model for label attribute JURUSAN

Class TKJ (0.565)
5 distributions

Class OTKP (0.435)
5 distributions
```

Figure 5. Description of distribution model *naive bayes*

c. Implementation

Processand the results of the classification model using rapid miner are intended so that the model performance is general in classifying data. In this study, K-fold cross validation was applied with a K-value = 10, which is a standard number in commonly used validation techniques and is available in the rapidminer application.

d. Classification Model Testing Configuration

With rapidminer 10.1 software, figure 4.5 shows the configuration of K-fold cross validation testing with the naive bayes algorithm. Where in the cross validation operator there is a configuration that refers to the algorithm model used, in this case it is naive bayes shown in figure 6.

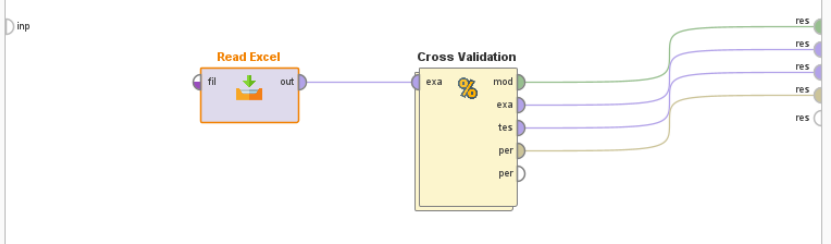

Figure 6. Operator configuration*Cross Validation*

Aftermakemodel as shown in figure 7, then the cross validation operator is clicked twice to be able to carry out the next configuration as shown in figure 7.

ISSN:2715-6427

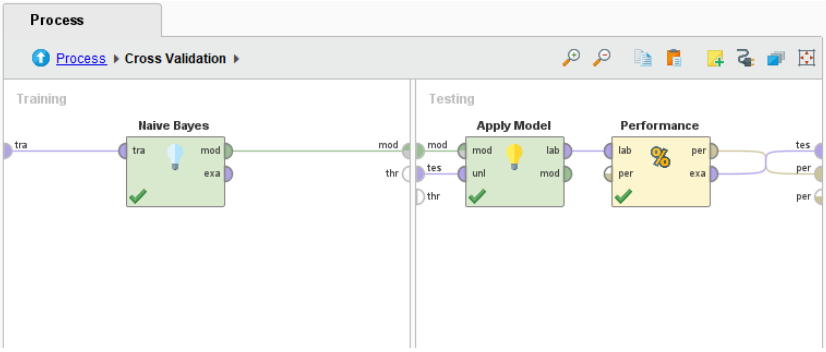

Figure 7. Model testing configuration *naive bayes*

e. Classification Model Testing Results

Figure 8. Show performanceVector screenshot showing the performance values of the naive bayes model.

PerformanceVector

```
PerformanceVector:
accuracy: 97.50% +/- 4.03% (micro average: 97.50%)
ConfusionMatrix:
True:   TKJ     OTKP
TKJ:    59      2
OTKP:   1       58
precision: 98.57% +/- 4.52% (micro average: 98.31%) (positive class: OTKP)
ConfusionMatrix:
True:   TKJ     OTKP
TKJ:    59      2
OTKP:   1       58
recall: 96.67% +/- 7.03% (micro average: 96.67%) (positive class: OTKP)
ConfusionMatrix:
True:   TKJ     OTKP
TKJ:    59      2
OTKP:   1       58
AUC (optimistic): 0.986 +/- 0.044 (micro average: 0.986) (positive class: OTKP)
AUC: 0.986 +/- 0.044 (micro average: 0.986) (positive class: OTKP)
AUC (pessimistic): 0.986 +/- 0.044 (micro average: 0.986) (positive class: OTKP)
```

Figure 8. Model performance *naive bayes*

The results of the accuracy of naive bayes are shown in Figure 9, from the student choice of major dataset with the rapidminer 10.1 tool application using the Naive Bayes Classifier method. What is obtained is the result of the Naive Bayes Classifier method using the K = 10 cross validation method with an accuracy of 97.50%.

accuracy: 97.50% +/- 4.03% (micro average: 97.50%)

	true TKJ	true OTKP	class precision
pred. TKJ	59	2	96.72%
pred. OTKP	1	58	98.31%
class recall	98.33%	96.67%	

Figure 9. Calculation results *naive bayes cross validation*

f. Evaluation and Validation of Methods

At the evaluation stage, the results are obtained by obtaining Accuracy, then evaluating the model and measuring accuracy with a confusion matrix that focuses on the class in general.

Table 3.*Confusion Matrix*

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	TrueNegative (TN)

So from the formula in table 4. will produce the calculation value in table 4 :

Table 4 *Confusion Matrix and Naive Bayes method*

	true true	true OTKP	class precision
TKJ Pred.	59	2	96.72%
Pred. OTKP	1	58	98.31%
class recall	98.33%	96.67%	

5. CONCLUSION

Based on the results of the research conducted from the initial stage to testing, the results of the average calculation of naive Bayes validation, it has shown very good accuracy, reaching 97.50%.. This accuracy shows that the Naive Bayes Classifier algorithm is able to determine choice of majors with a high level of success at Vocational School Hidayatul Islam. The accuracy of 97.50% shows that the Naive Bayes Classifier algorithm has the ability to predict students' choice of majors with a high level of success. In the context of Vocational School Hidayatul Islam, this means that the naive bayes algorithm can be an effective model in helping the process of determining students' choice of majors.

ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to all parties who have supported the preparation of this journal. Our special thanks go to:

1. Leadership for your direction and inspiration.
2. Colleagues for their constructive input and criticism.
3. Family and friends for their moral support and encouragement.
4. Institutions that have provided financial assistance and facilities.

We also thank all parties who cannot be mentioned one by one, but have contributed directly or indirectly. Hopefully this good cooperation can continue in the future.

Best regards,

ISSN:2715-6427

REFERENCES

[1] Mochammad Haldi Widianto. 2019. Naive Bayes Algorithm. BINUS University. <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/.>.

[2] Rina Kurniasari and Azizah Fatmawati. 2019. "Implementation of C4.5 Algorithm for Senior High School Students' Majors". Scientific Journal of Computer and Informatics (KOMPUTA) ISSN: 2089-9033, eISSN: 2715-7849, Vol. 8, No. 1, March 2019.

[3] Ahmad Zainul Mafakhir and Achmad. 2020. "Application of Naive Bayes Classifier Method for Student Majors at Madrasah Aliyah Al-Falah Jakarta". Fountain of Informatics Journal ISSN: 2541-4313, Vol. 5, No. 1, May 2020.

[4] Hadi Priyono, Retno Sari and Tati Mardiana. 2022. "Classification of Vocational High School Major Selection Using Gradient Boosting Classifier". JOURNAL OF INFORMATICS ISSN: 2355-6579, E-ISSN: 2528-2247, Vol. 9 No. 2 October 2022

[5] Dr. Rahmat Hidayat, MA and Dr. Abdillah, S.Ag, M.Pd. 2019. Educational Science "Concept, Theory and Its Application". Indonesian Education Development Care Institute (LPPPI) ISBN: 978-623-90653-8-6

[6] Naparin, H. (2016). Classification of High School Students' Interests Using the Naive Bayes Method. Systemic: Information System and Informatics Journal, 2(1), 25–32

[7] Edy Budiman. 2019. LEARNING THE BASICS OF ALGORITHMS & PROGRAMMING. Samarinda, ISBN: 978-602-14706-5-7.

[8] Dennis, Donny, Lia, & I Wayan. (2019). Learning Data Mining with RapidMiner. Jakarta: Open Content Model.

[9] Purnamasari, Detty. Henharta, Jonathan. 2019. Get Easy Using Weka. East Jakarta. Dapur Buku.

[10] Alfa Saleh. 2018. "Implementation of Naïve Bayes Classification Method in Predicting Household Electricity Usage". Citec Journal, ISSN: 2354-5771, Vol. 2, No. 3, May 2018

[11] Wijaya, N., Endah, M., Feliati, M., Studi, P., Program, I., & Korespondensi, P. (2020). Application of the C.45 decision tree algorithm for classification of occupancy status data for post-Merapi eruption rehabilitation houses. 424–430.

[12] Saleh, Alfa, & Nasari, F. (2018). The Use of Unsupervised Discretization Techniques in the Naive Bayes Method in Determining the Majors of Madrasah Aliyah Students. Journal of Information Technology and Computer Science, 5(3), 353

[13] Dennis, Donny, Lia, & I Wayan. (2019). Learning Data Mining with RapidMiner. Jakarta: Open Content Model

[14] Amalia, H. 2018. "Comparison of SVM and NN Data Mining Methods for Chronic Kidney Disease Classification". Pilar Nusa Mandiri Journal, 14(1), 1–6. Retrieved from https://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/80/67.

[15] Baker, R. C. (1989). Nonlinear unstable systems. International Journal of Control, 23(4), 123–145. Bisri, M. H. (2015). Implementation of Naïve Bayes Algorithm to Predict Student Majors at SMA Kesatrian 1 Semarang. Informatics Journal, 1–7