JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

ANALYSIS NAIVE BAYES TO SELECTION NEW STUDENTS FOR SUPERIOR CLASS STMIK ROYAL

Feby Wulandari Sembiring[1*], Riki Andri Yusda[1], Santoso[1]
[1]Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
email: *febywulandari421@gmail.com

.

Abstract: Superior students represent active and intelligent students who are able to make a direct contribution to the development of the nation. To produce excellent students, one of the efforts that can be made is to foster these superior students from the start by forming a superior class. In forming a superior class, an effective selection process is needed to select students who are truly superior. This research was conducted at STMIK Royal Kisaran where the object of research was prospective new students for the superior class and the related party was the Academic and Student Administration Bureau (BAAK) because the party was directly involved in selecting prospective new students for the superior class. But the problem is, BAAK does not yet have an election process based on a method and there is no information system for selecting prospective new students for superior classes. The purpose of this study is to make predictions related to the selection of prospective new students for superior classes in the future with the Naïve Bayes Algorithm and implemented in an information system.The results of this research are that the Naïve Bayes algorithm produces an accuracy rate of 60% in the good classification category which is measured by the level of accuracy using the confusion matrix, so that the information system produced in this study also has efficient prediction results and is expected to help BAAK.

Keywords: data mining; naïve bayes; superior student class.

Abstrak: Mahasiswa unggulan merupakan representasi dari mahasiswa aktif dan cerdas yang mampu memberikan kontribusi langsung terhadap perkembangan bangsa. Untuk menghasilkan mahasiswa yang benar unggul salah satu upaya yang dapat dilakukan adalah dengan membina mahasiswa unggulan tersebut sejak awal dengan membentuk kelas unggulan. Dalam membentuk kelas unggulan dibutuhkan proses seleksi yang benar efektif untuk memilih mahasiswa yang benar unggul. Penelitian ini dilakukan di STMIK Royal Kisaran dimana objek penelitian merupakan calon mahasiswa baru untuk kelas unggulan dan pihak yang terkait adalah Biro Administrasi Akademik dan Kemahasiswaan (BAAK) karena pihak tersebut terlibat langsung dalam pemilihan calon mahasiswa baru untuk kelas unggulan. Namun permasalahannya, pihak BAAK belum memiliki sebuah proses pemilihan berdasarkan metode dan belum adanya sistem informasi pemilihan calon mahasiswa baru kelas unggulan. Tujuan dari penelitian ini adalah melakukan prediksi terkait pemilihan calon mahasiswa baru kelas unggulan di masa mendatang dengan Algoritma Naïve Bayes dan diimplementasikan dalam sebuah sistem informasi. Hasil penelitian ini algoritma Naïve Bayes menghasilkan tingkat akurasi sebesar 60% dengan kategori *good classification* yang diukur tingkat akurasinya dengan menggunakan *confusion matrix*, sehingga sistem informasi yang dihasilkan dalam penelitian ini juga memiliki hasil prediksi yang efisien dan diharapkan dapat membantu pihak BAAK.

Kata Kunci: data mining; mahasiswa kelas unggulan; naïve bayes.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

College students are students who carry out knowledge in the world of tertiary institutions. In this role, students are required to have a more intelligent mindset because students have a direct contribution to providing their aspirations for national development. Students needed for national development are students who have the ability/skill with very good categorization and are superior compared to others both in terms of academic and non-academic competencies [1].

Smart and active students are a representation of superior students who are nurtured from the start, especially since becoming new students. For this reason, one of the efforts that can be made to foster these superior students is to form a superior class. The superior class is an educational program that focuses on students/students so that they can learn actively with the aim of increasing student competence both cognitively, attitudes and behavior and skills [2].

The aim of forming a superior class is to create a competitive and responsive learning atmosphere and it is hoped that in the future the superior class students who have been prepared from the start will be able to take part in competitions outside the campus both at the national and international levels. As is the case with STMIK Royal Kisaran which is currently actively participating in various activities and competitions involving its students so that a new regulation has emerged, namely the establishment of a superior class at STMIK Royal Kisaran which is launched in 2022.

Students who successfully enter the superior class are assessed based on criteria, namely the value of the entrance screening test (USM), and an English interview which consists of understanding, communicative, right on and pronounciation assessments.

The problem namely there is no specific calculation method in selecting prospective new students for a decent superior class, it is feared that the resulting decisions are not optimal in selecting prospective superior class students who are truly worthy. In addition, there is no information system that is directly integrated with certain methods or algorithms in selecting prospective students for the STMIK Royal Kisaran superior class.

For this reason, a solution is needed to help increase the efficiency and effectiveness of the system in selecting prospective students for superior classes by building an information system for selecting prospective new students for superior classes that is integrated with an algorithm, namely Naïve Bayes. The naïve Bayes algorithm is used for predictions based on existing data classifications from previously formed data patterns [3].

The research with the title "Application Of The Naive Bayes Model To Predict The Potential Of Student Registration In Web-Based Taman Siswa Teluk Betung Vocational School", the Naïve Bayes algorithm was used to predict the potential of students enrolling in SMK Taman Siswa with a resulting accuracy of 86% [4]

Research with the title "Implementation of Orange Data Mining for Classification of Student Graduations Using the K-Nearest Neighbor Model, Decision Tree and Naïve Bayes",

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

recommends Naïve Bayes for graduation classification because the resulting accuracy is higher than other classification algorithms [5].

The study entitled "Application of the Naïve Bayes Method to Determine Employee Performance Assessment at PT.Sinergi Guna Solusindo" uses Naïve Bayes to assess employee performance and the resulting Naïve Bayes accuracy is 94% [6].

The research "Implementation of Naïve Bayes for Student Achievement Prediction Systems" uses Naïve Bayes to select outstanding students and its accuracy is measured by blackbox testing [3].

Research with the title "Implementation of Naïve Bayes Theory in Classification of Prospective New Students STMIK Kharisma Makassar" the accuracy produced by the Naïve Bayes algorithm is 73% [7].

Previous research [8] explained that in his research using criteria in selecting superior class students in the form of grades obtained for one semester consisting of academic scores and spiritual values, the use of the Naïve Bayes algorithm in this study resulted in good grades obtained from the respondents' results. Where the value reaches above 60% is only obtained from the results of the respondents.

Whereas in this study the selection of superior students was obtained from the criteria for the entrance examination scores and English Interview which consisted of communicative, right on, understanding and pronunciation so that the assessment used was more in-depth and detailed with the Naïve Bayes algorithm and the measurement of the accuracy level would be tested using Confusion Matrix.

The purpose of using the Naïve Bayes method to predict the eligibility of prospective new students for superior classes in the future, where after seeing several previous studies the Naïve Bayes algorithm is widely used for a selection process and the level of accuracy produced by Naïve Bayes is also very good with average percentage of 80%.

Therefore, the Naïve Bayes algorithm is suitable to be implemented into the system to classify prospective new students who are eligible to enter superior classes in the future.

METHOD

This research was conducted at STMIK Royal Kisaran, where the research object was prospective new students for the superior class. The party related to the object is the Academic and Student Administration Bureau (BAAK) STMIK Royal where the results of this research are in the form of an information system for selecting prospective new students for superior classes that can help BAAK in predicting prospective superior class students in the future.

The Naïve Bayes method used in this study has a data mining concept. [5]. The used of naïve Bayes algorithm to calculate the probability or likelihood that will occur in the future based on data that has existed in the past [6].

The Naïve Bayes algorithm is used to calculate the probability of a dataset. Naïve Bayes is categorized as Supervised Learning because it has a label/target/class in the dataset which is used as a reference in Bayes calculations. The advantages of the Naïve Bayes algorithm according to [6] include that the method is simple but has

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

high accuracy. The Naïve Bayes process with the concept of data mining in this study uses the Knowledge Discovery in Database (KDD) stage. The following is the Knowledge Discovery in Database (KDD) stage [7].

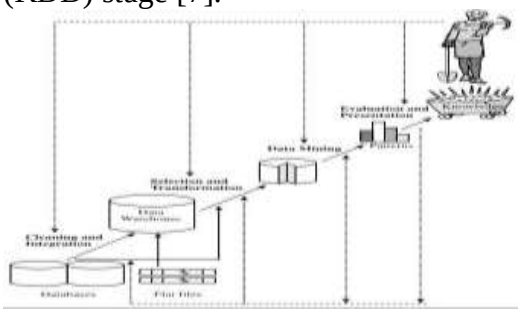

Image 1. The Stage of KDD

Data Cleaning

Data on new students obtained from the Academic and Student Administration Bureau (BAAK) totaled 744 people. However, after going through the data cleaning process, there were 645 people who could be processed. This data is suitable data to be managed by the data mining process.

Data Integration

The integration of the existing data in this study combines the scores of the entrance examination (USM) and the scores of the English interview.

Data Selection

In the data selection process, the data that will be used to calculate Naïve Bayes are the data on the results of the entrance screening test (USM), English language interviews which consist of understanding, right on, communicative, pronunciation and decision values.

Data Transformation

The Naïve Bayes algorithm performs processing on categorical data (letters/words). Because the initial data for prospective new students for superior classes is in the form of numbers (numeric), data transformation is carried out so that Naïve Bayes calculations can be carried out.

Table 1. Transformation of USM

Interval	Transformation
0-50	Low
51 - 79	Medium
80 - 100	High

Table 2.Transformation of understanding

Interval	Transformation
0-9	Not Understand
10 - 15	Enough
16 - 20	Understand
21 - 25	Master

Table 3. Communicative and Right On

Interval	Transformation
0 - 9	Less Good
10 - 15	Enough
16 - 20	Good
21 - 25	Very Good

Table 4.Transformation of Pronunciation

Interval	Transformation
0 - 9	Less Clear
10 - 15	Clear
16 - 20	Enough
21 - 25	Less Clear

Data Mining

In the data mining process, data mining is carried out using the Naïve Bayes method. Classification of prospective new students for superior classes with Naïve Bayes is by looking for probabilities from pre-existing data. The naïve Bayes performance used is dividing the data into training data and testing data then using the resulting

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

probabilities to predict future data with the following formula [8]:

$$P(H|X) = \frac{P(X|H)X\ P(H)}{P(X)} \quad (1)$$

Describe:
X : Unknown class
H : Hypothesis contained in a specific class
P(H|X) : The probability of a hypothesis H under condition X
P(H) : Probability of a hypothesis H (probability prior)
P(X|H) : Probability of X based on condition
P (X) : The probability of X

The distribution of data in this study is training data and testing data in table 5 and table 6

Data is also divided into data testing, namely data used to test data testing and test the Naïve Bayes algorithm. Table 6 is data testing in selecting prospective new students for superior classes

Table 5. Sample of Data Training

no	USM value	understanding	communicative	right on	pronounciation	decisions
1	High	Master	Less Good	Less Good	Less Clear	Very Worthy
2	Low	Master	Very Good	Very Good	Very Clear	Very Worthy
3	High	Master	Very Good	Very Good	Very Clear	Worthy
4	High	Understand	Good	Good	Clear	Unworthy
5	Medium	Not Understand	Less Good	Less Good	Less Clear	Worthy
6	Low	Not Understand	Less Good	Less Good	Less Clear	Unworthy

Table 6. Sample of Data Testing

no	USM Value	understanding	communicative	right on	pronounciation	decisions
1	High	Master	Less Good	Less Good	Less Clear	Very Worthy
2	High	Master	Very Good	Very Good	Clear	Worthy
3	High	Understand	Good	Good	Clear	Unworthy
4	Medium	Understand	Good	Good	Clear	Worthy
5	Medium	Not Understand	Less Good	Less Good	Less Clear	Very Worthy
6	Low	Not Understand	Less Good	Less Good	Less Clear	Unworthy
7	High	Not Understand	Good	Less Good	Very Clear	???

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Evaluation

In this stage to measure the level of accuracy of the predictions generated by the Naïve Bayes algorithm. Accuracy measurement uses a confusion matrix where the resulting value consists of True Positive (TP), namely data that is predicted to be positive and the facts are true positive, True Negative (TN), namely data that is predicted to be negative and the facts are true negative, False Negative (FN) where data is predicted to be positive but the facts are not positive, False Positive (FN) is where the data is predicted to be negative but the facts are not negative [9].

Knowledge Presentations

In this stage the information that has been generated is then presented in the form of an information system. Where the system will display the results of prospective new students for the superior class from the classification produced by the naïve Bayes algorithm so that it is easier for the admin to understand and manage [10].

RESULT AND DISCUSSION

In the process of discussing the results of this study, it describes the performance process of the naïve Bayes algorithm in making predictions from the classification of prospective new students for superior classes. To perform naïve Bayes calculations, the researcher enters 1 sample data which will be processed in data testing to find out the results of the decision. The sample data that will be processed by Naïve Bayes is that there is a new student candidate for the superior class where he has the criteria for the entrance screening exam (USM) score HIGH, but his Understanding ability is in the category of NOT UNDERSTAND, his Communicative is GOOD, his Right On Ability is LESS GOOD, but Pronunciation VERY CLEAR, then based on the probability of the previous data the decision results of the data are?

To find out the results of the decision, a Naïve Bayes calculation is performed with probabilities from previous data as contained in the table 7 to table 12

Table 7. Probability of (USM)

USM	Probability of Decision		
	Very Worthy	Worthy	Unworthy
high	0,387	0,322	0,037
medium	0,612	0,677	0,816
low	0,003	0,003	0,145

Table 8. Probability of Understanding

Understanding	Probability of Decision		
	Very Worthy	Worthy	Unworthy
master	0,580	0,096	0,032
understand	0,290	0,709	0,002
enough	0,032	0,161	0,000
not understand	0,096	0,032	0,998

Table 9. Probability of Communicative

Communicative	Probability of Decision		
	Very Worthy	Worthy	Unworthy
very good	0,580	0,129	0,001
good	0,290	0,645	0,002
enough	0,032	0,193	0,002
less good	0,096	0,032	0,998

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 10. Probability of Right On

Right On	Probability of Decision		
	Very Worthy	Worthy	Unworthy
very good	0,645	0,064	0,032
good	0,225	0,677	0,002
enough	0,032	0,225	0,002
less good	0,096	0,032	0,998

Table 11. Probability of Pronunciation

Pronunciation	Probability of Decision		
	Very Worthy	Worthy	Unworthy
very clear	0,548	0,064	0,001
clear	0,322	0,741	0,000
enough	0,032	0,161	0,000
less clear	0,096	0,032	0,998

Table 12. Probability of Decision

The Decision	Probability of Decision
Very Worthy	0,04806
Worthy	0,04806
Unworthy	0,9039

The next process is to multiply each probability according to the criteria contained in the sample data whose decision results you want to know:

Very Worthy
P(USM = High | Decision = Very Worthy) * P(Understanding = Not Understand | Decision = Very Worthy) * P(Communicative = Good | Decision = Very Worthy) * P(Right On = Less Good | Decision = Very Worthy) * P(Pronounciation = Very Clear | Decision = Very Worthy) * P(Very Worthy)

Worthy
P(USM = High | Decision = Worthy) * P(Understanding = Not Understand | Decision = Worthy) * P(Communicative = Good | Decision = Worthy) * P(Right On = Less Good | Decision = Worthy) * P(Pronounciation = Very Clear | Decision = Worthy) * P(Worthy)

Unworthy
P(USM = High | Decision = Unworthy) * P(Understanding = Not Understand | Decision = Unworthy) * P(Communicative = Good | Decision = Unworthy) * P(Right On = Less Good | Decision = Unworthy) * P(Pronounciation = Very Clear | Decision = Unworthy) * P(Unworthy)

After that, to look at the magnitude of the probability with the Naïve Bayes algorithm :

Very Worthy Probability

In the Very Worthy criteria, a probability of 0.9760 is obtained with an acquisition percentage of 97%.

Worthy Probability

In the Worthy criteria, a probability of 0.2368 is obtained or with a percentage of 2%. This percentage is smaller than the probability percentage resulting from the Very Worthy criterion.

Unworthy Probability

Meanwhile, the Unworthy criteria produces a probability of 0.00035 or 0%, this means that the prospective

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 239 - 248 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2216
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

student does not fit the unworthy criteria.

It can be seen that the prospective students are Very worthy to enter the superior class. Where the greatest probability results are generated from the Very Worthy criterion of 97%.

Table 13. Sample of Superior Student

Name	The Decision	Superior Class
Anjani	Very Worthy	SI-1A
Diana Selvi	Very Worthy	SI-1A
Juli Febri	Very Worthy	SI-1A
Yayang Apriza	Very Worthy	SI-1A
Waldi Sahputra	Very Worthy	SI-1A

The accuracy level of the prediction results is 60% calculated based on the confusion matrix model. The results of the 60% accuracy are categorized as good prediction results.

Tabel 14. Confusion Matrix

		Actual	
Prediction	Worthy	Very Worthy	Not Worthy
Worthy	1	0	1
Very Worthy	1	4	1
Not Worthy	0	0	0

The results of Naïve Bayes calculations in predicting prospective new students for superior classes are then implemented into an information system for selecting prospective new students for superior classes. With the implementation of the Naïve Bayes algorithm in the information system, it is hoped that it will assist the work of the Academic and Student Administration Bureau (BAAK) in selecting prospective new students for superior classes and the resulting information is more relevant and effective due to the implementation of the Naïve Bayes algorithm.

The result of the implementation of the Naïve Bayes algorithm in the information system for selecting prospective new students for superior classes in image 2.

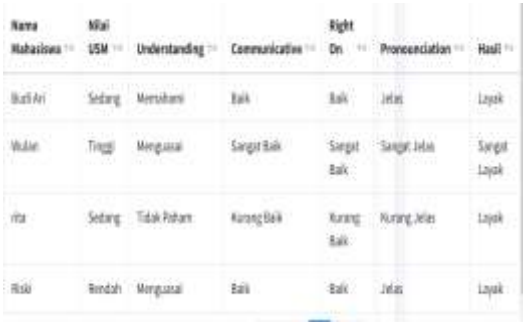

Image 2. Result of Naïve bayes Algorithm in system

CONCLUSION

The information obtained from this research—that the Naïve Bayes algorithm is suitable to be implemented in selecting prospective new students for superior classes. This can be seen from the prediction results generated in predicting a new case where the algorithm produces an accuracy rate of 60% as measured using the confusion matrix. The data sharing process uses a data sharing technique at a ratio of 70:30, namely 70% training data and 30% testing data so that the calculations are more even.

The implementation of the naïve Bayes algorithm can be used easily because it is integrated with the information system for selecting prospective students for superior classes, so that the admin can manage the system properly and the prediction results by the information system for selecting prospective students for superior classes have good accuracy so that the resulting value can help BAAK to select prospective new students for superior classes that are more accurate and effective.

BIBLIOGRAPHY

[1] M. A. Putri and A. Supriyanto, "Pembangunan Karakter Mahasiswa Melalui Keikutsertaan Dalam Organisasi Kemahasiswaan," *J. Manaj. UNM*, pp. 160–166, 2020, [Online]. Available: http://conference.um.ac.id/index.php/apfip/article/view/405

[2] A. R. Z. Wati and S. Trihantoyo, "Strategi Pengelolaan Kelas Unggulan Dalam Meningkatkan Prestasi Belajar Siswa," *J. Din. Manaj. Pendidik.*, vol. 5, no. 1, p. 46, 2020, doi: 10.26740/jdmp.v5n1.p46-57.

[3] S. A. Desi Puspita, "IMPLEMENTASI NAIVE BAYES UNTUK SISTEM PREDIKSI MAHASISWA BERPRESTASI," vol. 8, no. 2, 2022.

[4] S. Eka *et al.*, "Penerapan Model Naive Bayes Untuk Memprediksi Potensi," vol. 1, no. 1, pp. 82–87, 2021.

[5] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.

[6] P. Kinerja, K. Pt, S. Guna, and A. Sudrajat, "Penerapan Metode Naïve Bayes Untuk Menentukan Penilaian Kinerja Karyawan PT.Sinergi Guna Solusindo," vol. 99, no. 99, pp. 1596–1606, 2022.

[7] I. Loelianto, M. S. S. Thayf, and H. Angriani, "Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru Stmik Kharisma Makassar," *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 2, pp. 110–117, 2020, doi: 10.31598/sintechjournal.v3i2.651

[8] S. Nasional *et al.*, "Implementasi Metode Naïve Bayes Untuk Menentukan Kelas Unggulan Pada Siswa SMP," pp. 215–220, 2022.

[9] D. P. Utomo, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," vol. 4, no. April, pp. 437–444, 2020, doi: 10.30865/mib.v4i2.2080.

[10] M. Muhathir and M. H. Santoso, "Analysis Naïve Bayes In Classifying Fruit by Utilizing Hog Feature Extraction," *J. Informatics Telecommun. Eng.*, vol. 4, no. 1, pp. 151–160, 2020, doi: 10.31289/jite.v4i1.3860.

[11] A. N. Syaiful Zuhri Harahap1, "TEKNIK DATA MINING UNTUK PENENTUAN PAKET HEMAT SEMBAKO DAN KEBUTUHAN HARIAN

DENGAN MENGGUNAKAN ALGORITMA FP-GROWTH (STUDI KASUS DI ULFAMART L," vol. 7, no. 3, pp. 111–119, 2019.

[12] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.

[13] N. Sabila *et al.*, "IMPLEMENTATION OF WEB-BASED NAIVE BAYES ALGORITHM FOR DETERMINING DEPARTMENTS AT SMK 10 MUHAMMADIYAH KISARAN IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK PENENTUAN JURUSAN BERBASIS WEB DI SMK 10 MUHAMMADIYAH," vol. 3, no. 6, pp. 1729–1737, 2022.