Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 6, Issue 1, January 2026; Page, 527-535
Email: pantherajurnal@gmail.com

PREDICTION MODEL GRADUATION STUDENT WITH NAIVE BAYES ALGORITHM

Lolyta Damora Simbolon[1]*, Rani Farida Sinaga[2], Lena Rosdiana Pangaribuan[3], & Sofy Kristiani Manalu[4]
[1,2,3,&4]Mathematics Education Study Program, Faculty of Teacher Training and Education, HKBP Nommensen University, Sutomo Street Number 4A, Medan, North Sumatera 20235, Indonesia
**Email: lolyta.simbolon@uhn.ac.id*
Submit: 10-12-2025; Revised: 29-01-2026; Accepted: 30-01-2026; Published: 31-01-2026

ABSTRACT: One of indicator important in evaluation quality something college tall is level graduation students. Graduation rate appropriate time students become a reflection from the quality of the learning process which ultimately also influences accreditation institutions or study program. The purpose of this research is to analyze how to increase student graduation rates and reduce dropout rates through more accurate and effective data-based policy recommendations. There are Lots factor affecting graduation student so that need determined factors significant influence level graduation said. With do analysis to data and predictions graduation students, institutions expected can give appropriate intervention for increase level graduates and identify at-risk students experience late graduation. In research this using the Naïve Bayes method for predict graduation student based on various factors. The data used in study this is the data obtained from students of the Mathematics Education Study Program, Faculty of Teacher Training and Education, HKBP Nommensen University. The results of study this show accuracy 80%, precision 88.24%, and recall 88.24%.

Keywords: Graduation, Model, Naïve Bayes, Prediction, Student.

How to Cite: Simbolon, L. D., Sinaga, R. F., Pangaribuan, L. R., & Manalu, S. K. (2026). Prediction Model Graduation Student with Naive Bayes Algorithm. *Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan, 6*(1), 527-535. https://doi.org/10.36312/panthera.v6i1.876

INTRODUCTION

College tall own obligation for produce competent graduates. This can assessed from level graduation his students (Khasanah et al., 2022). Graduation appropriate time is one of the indicator important in evaluate quality college tall. High graduation rate no only reflect effectiveness of the learning process but also influences reputation and accreditation college tall the (Basuki, 2023). In the context accreditation, accuracy time graduation students become one of the criteria evaluation main, as stated by the National Accreditation Board for Higher Education (BAN-PT) and various Accreditation Institutions Independent (LAM) (Legowo & Indiarto, 2017).

One of the reject measuring quality college tall seen from many students who can finish studies in accordance with specified time or the term students graduate on time time (Gunawan et al., 2021). The more lots students who graduate on time time so the more good performance too college tall said, so that level graduation student appropriate time becomes one criteria evaluation accreditation for something college tall or study program (Azahari et al., 2020). Ideal study duration for a student level bachelor for finish his education is 4 years, will but still there is students who do not finish studies appropriate time (Perez & Perez, 2021). This condition indicates the need for further evaluation of academic, administrative, and student support systems to improve on-time graduation rates.

Many factors influence success education students, good academic or non-academic (Issah et al., 2023). Understand factors that influence it the can give evaluation about accuracy time study which is success in management education (Rolansa et al., 2020). Some factors that influence graduation student late, such as marital status student, student status (working/ not working), level understanding student to material lectures that can seen from the student's GPA (Anwarudin et al., 2022). This factor is the source data main thing that can utilized for managed, so that existing data patterns can understood (Jananto et al., 2021). Retrieval decision no enough only depend on action repair after problem appears, but is also necessary action preventive with do data analysis and classification or prediction for know data patterns and transform them into useful information (Pangestu, 2023). By do data analysis and creating predictive models, institutions education can identify patterns and contributing factors to graduation students, so that can strategize for raise level graduation (Sembiring & Tambunan, 2021).

For form a prediction model graduation students in research this will used approach machine learning with Naïve Bayes method. Powerful machine learning own potential for help institutions in taking decision policy (Putra & Harahap, 2024) and the Naïve Bayes method are used for predict graduation student based on various factor with using historical data students (Wijiyanto et al., 2024). The results of study this expected can give contribution in increase level graduation student as well as reduce number separated studies through recommendation policy more data-based accurate and effective (Hussen & Saikhu, 2024). Based on the description above, then formulation problem from study this is how do prediction graduation student with an optimal prediction model. For solve problem the so done analysis to factors that influence graduation students and then the prediction model graduation student formed with the Naïve Bayes method (Mehta, 2023).

METHOD

Approach quantitative in research this referring to to Naïve Bayes method classification used for analyze data for produce output in the form of classification (Rachmawati & Miasary, 2024). While data analysis was carried out to object research and variables research (Allaam, 2021). The data used in study this is secondary data namely student data obtained from the database of the Mathematics Education Study Program at HKBP Nommensen University (Mulyana, 2006). The student data in question is data on students who graduated in 2024 and 2025.

Variables Study

Study this involving a number of variables that then grouped into variables dependent and variable independent. Variable dependent in study this is graduation students (right) time and late, where the indicators seen from the length of study students (Viet et al., 2021). As for the variables independent in study this is student data in the form of type gender, employment status, IPS 1, IPS 2, IPS 3, IPS 4, IPS 5, and IPS 6, length of service end (Mehdi & Nachouki, 2023).

Procedure Study

Study this done in a number of the stages that begin from data collecting, preprocessing, splitting data, modeling and evaluation.

Data Collecting

Data collecting is stage first for collect the required initial data in study this. Good data is the data obtained in a way objective and has verified (Awaludin et al., 2022).

Preprocessing

At the stage this done selection influential attributes big to predictions. Empty data and missing attributes influential to calculation will deleted (Marzuqi et al., 2021).

Splitting Data

This process done for share the data that has been through stage preprocessing to in train data and test data.

Modeling (Naive Bayes)

At the stage this analysis on the data is carried out with creating a Naïve Bayes method in do prediction graduation students. At this stage this, done manual calculations with use formula Naive Bayes algorithm (Rawal & Lal, 2023).

Evaluation

Applications used in calculation for study this is RapidMiner. Evaluation process done with compare calculation with manual applications and calculations. At this stage this will done calculation accuracy, precision, and recall use confusion matrix. Confusion matrix provides classification performance assessment based on object with correct or wrong and produce mark accuracy, recall, precision. Study this expected produce an optimal prediction model with mark high accuracy. Accuracy value show level model accuracy in classify data with correct (Kartianom et al., 2022).

RESULTS AND DISCUSSION

Data Collection

Data used in study this is the data of students of the Mathematics Education Study Program, Faculty of Teacher Training and Education, HKBP Nommensen University, who have passed the exam green in 2 year period last. Data used a total of 68 student data. The following is displayed sample data before preprocessing (Bakri et al., 2022).

Table 1. Research Data Sample Table.

Name	NPM	Gender	Employment status during student study period	Marital status during student study period	Grade Point Average (GPA) Semester 1	Grade Point Average (GPA) Semester 2	Grade Point Average (GPA) Semester 3	Grade Point Average (GPA) Semester 4	Grade Point Average (GPA) Semester 5	Grade Point Average (GPA) Semester 6	Length of final assignment (thesis)	Student graduation status
Hernando Silitonga	20150056	Man	Doesn't work	Not married yet	3.97	4.00	3.86	4.00	3.92	3.77	> 6 months	On time (<= 4 years / 8 semesters)

Preprocessing

Data cleaning was performed by removing variables or columns that were irrelevant to prediction and to avoid privacy concerns. Sample data after variable cleaning is shown in Table 2 (Zheng & Li, 2024).

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 6, Issue 1, January 2026; Page, 527-535
Email: pantherajurnal@gmail.com

Table 2. Research Data Samples After the Data Cleaning Process.

JK	Employment status during student study period	Grade Point Average (GPA) Semester 1	Grade Point Average (GPA) Semester 2	Grade Point Average (GPA) Semester 3	Grade Point Average (GPA) Semester 4	Grade Point Average (GPA) Semester 5	Grade Point Average (GPA) Semester 6	Length of final assignment period (thesis/month)	Student graduation status
Man	Doesn't work	3.97	4.00	3.86	4.00	3.92	3.77	> 6	On time (<= 4 years / 8 semesters)

Data Transformation

The next step is to transform or label the data in the categorical variables into numeric data types (Nuraeni et al., 2021). This is necessary for the Naïve Bayes algorithm to process the data effectively. Table 3 shows the results of the data transformation or labeling for the categorical variables.

Table 3. Results of the Data Transformation Process.

Gender	Employment status during student study period	Grade Point Average (GPA) Semester 1	Grade Point Average (GPA) Semester 2	Grade Point Average (GPA) Semester 3	Grade Point Average (GPA) Semester 4	Grade Point Average (GPA) Semester 5	Grade Point Average (GPA) Semester 6	Length of final assignment (thesis)	Student graduation status
0	0	3.97	4	3.86	4	3.92	3.77	0	On time

Splitting Data

Before starting modeling with Naïve Bayes algorithm, the data must be divided into two parts: training data and testing data. Of the total dataset, 30% is for testing data and 70% is for training data, respectively (Dikriani & Karim, 2023).

Table 4. Comparison of Training Data and Testing Data.

Information	Training Data	Testing Data	Total
Proportion	70%	30%	100%
Amount	48	20	68

Modeling with Naïve Bayes

At this stage, the model design is formed and implemented through training and data testing with RapidMiner 9.10.008. The prediction model is obtained based on the probability values of each attribute from the data training process. The model evaluation process is conducted to measure the accuracy and reliability of the resulting Naïve Bayes method. This evaluation uses testing data that are separate from the training data, allowing the model's performance to be assessed objectively through metrics such as accuracy, precision, recall, and the confusion matrix. The evaluation results serve as a basis for determining the model's classification capability and ensuring that the developed model can provide consistent and reliable predictions on new data.

The evaluation results are used as a basis for further analysis of the strengths and weaknesses of the developed Naïve Bayes method. By examining the evaluation metrics and misclassification patterns shown in the confusion matrix, it is possible to identify the most influential attributes as well as potential biases or prediction errors. This analysis enables model refinement through the selection of more relevant attributes, parameter adjustments, or the inclusion of additional training data, with the aim of improving overall performance and ensuring that the model can be applied effectively and reliably in real-world scenarios.

Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan
E-ISSN 2808-246X; P-ISSN 2808-3636
Volume 6, Issue 1, January 2026; Page, 527-535
Email: pantherajurnal@gmail.com

Attribute	Parameter	Tepat waktu	Terlambat
Jenis Kelamin	mean	0.775	0.500
Jenis Kelamin	standard deviation	0.423	0.535
Status pekerjaan saat masa studi mahasiswa	mean	0.200	0
Status pekerjaan saat masa studi mahasiswa	standard deviation	0.405	0.001
Indeks Prestasi (IP) Semester 1	mean	3.757	3.449
Indeks Prestasi (IP) Semester 1	standard deviation	0.158	0.261
Indeks Prestasi (IP) Semester 2	mean	3.610	3.519
Indeks Prestasi (IP) Semester 2	standard deviation	0.216	0.214
Indeks Prestasi (IP) Semester 3	mean	3.780	3.692
Indeks Prestasi (IP) Semester 3	standard deviation	0.161	0.200
Indeks Prestasi (IP) Semester 4	mean	3.729	3.547
Indeks Prestasi (IP) Semester 4	standard deviation	0.182	0.105
Indeks Prestasi (IP) Semester 5	mean	3.672	3.608
Indeks Prestasi (IP) Semester 5	standard deviation	0.265	0.150
Indeks Prestasi (IP) Semester 6	mean	3.721	3.550
Indeks Prestasi (IP) Semester 6	standard deviation	0.148	0.085
Lama masa tugas akhir (skripsi)	mean	0.725	0.125
Lama masa tugas akhir (skripsi)	standard deviation	0.452	0.354

Figure 1. Attribute Probability Values from Training with RapidMiner.

The results of training the data with RapidMiner also display the statistical distribution of each class attribute used. After the model design is configured, the next step is data testing. The prediction results from the data testing are shown in Table 6 below (Nakhipova et al., 2024).

Row No.	Status kelul...	prediction(S...	confidence(...	confidence(...	Jenis Kelamin	Status peke...	Indeks Pres...	Indeks Pres...	Indeks Pres...	Indeks Pres...	Indeks Pres...	Indeks Pres...	Lama masa ...
1	Tepat waktu	Tepat waktu	1.000	0.000	1	0	3.850	3.810	3.730	4	3.350	3.640	1
2	Terlambat	Tepat waktu	0.546	0.454	0	0	3.810	3.870	3.810	3.800	3.710	3.770	0
3	Tepat waktu	Tepat waktu	0.788	0.212	0	0	3.820	3.170	3.930	3.800	3.880	3.650	1
4	Tepat waktu	Tepat waktu	1	0	1	1	3.600	3.700	3.200	3.300	3.400	3.500	1
5	Tepat waktu	Tepat waktu	1.000	0.000	1	0	3.850	3.420	3.610	3.800	2.820	3.480	1
6	Tepat waktu	Tepat waktu	1	0	1	1	3.920	3.730	4	3.790	3.840	3.800	1
7	Tepat waktu	Terlambat	0.489	0.511	1	0	3.820	3.770	3.750	3.780	3.850	3.600	1
8	Tepat waktu	Tepat waktu	1.000	0.000	1	0	3.880	3.730	3.830	4	3.540	3.800	1
9	Tepat waktu	Tepat waktu	1	0	0	1	3.920	3.750	4	3.710	3.750	3.850	1
10	Tepat waktu	Tepat waktu	0.999	0.001	1	0	3.670	3.750	3.800	3.750	3.970	3.920	0
11	Tepat waktu	Tepat waktu	0.992	0.008	1	0	3.750	3.710	3.820	3.790	3.370	3.750	1
12	Terlambat	Terlambat	0.129	0.871	1	0	3.850	3.670	3.550	3.820	4	3.890	0
13	Terlambat	Tepat waktu	0.978	0.022	1	0	3.870	3.820	3.440	3.630	4	3.050	0
14	Tepat waktu	Tepat waktu	1	0	1	1	3.790	3.650	3.670	3.670	3.880	3.650	1
15	Tepat waktu	Tepat waktu	0.945	0.055	1	0	3.770	3.750	3.840	3.700	3.830	3.750	1
16	Tepat waktu	Terlambat	0.000	1.000	1	0	3.570	3.310	3.830	3.430	3.260	3.530	0
17	Tepat waktu	Tepat waktu	1	0	0	1	3.760	3.640	3.790	3.930	4	3.750	1
18	Tepat waktu	Tepat waktu	0.949	0.051	1	0	3.780	3.600	4	3.760	3.670	3.720	1
19	Tepat waktu	Tepat waktu	0.999	0.001	1	0	3.820	3.700	4	3.780	3.750	3.810	1
20	Tepat waktu	Tepat waktu	0.891	0.109	1	0	3.460	3.100	3.350	3.200	2.900	3.280	1

Figure 2. Prediction Results from Data Testing.

Based on Figure 2 above, it appears that the prediction model successfully classified graduation rates. It was found that 17 students were in the "on-time" class and 3 students were in the "late" class.

Evaluation

The following is the calculation of the evaluation results with:

accuracy: 80.00%

	true Tepat waktu	true Terlambat	class precision
pred. Tepat waktu	15	2	88.24%
pred. Terlambat	2	1	33.33%
class recall	88.24%	33.33%	

Figure 3. Confusion Matrix Results.

1) Accuracy

The confusion matrix results show a performance vector with 80% accuracy, which means this model correctly classifies 80% of the test data (Akanbi, 2023).

2) Confusion Matrix

True Positive (TP) is 15 records that were predicted and correct as students who graduated "on-time", while False Negative (FN) is 2 records that were predicted and correct as students who graduated "late". Furthermore, 2 True Negative (TN) records were predicted as students who graduated "late" but graduated "on-time", and 1 False Positive (FP) record was predicted as students who graduated "late" but graduated "on-time".

3) Precision/Recall

Based on Figure 3, the precision/recall for the "on-time" class shows 88%, which means the results are accurate, while for the "late" class the results are 33%.

CONCLUSION

The research on predicting the graduation rate of Mathematics Education Study Program students used 68 data sets, with details of 48 training data and 20 testing data. The attributes used include: gender, employment status, social studies 1, social studies 2, social studies 3, social studies 4, social studies 5, and social studies 6, length of final assignment period and graduation status. The research results obtained were accuracy = 80%, precision = 88.24%, and recall = 88.24%.

SUGGESTION

This research shows that the strategy of this research is expected to contribute to increasing student graduation rates and reducing dropout rates through more accurate and effective data-based policy recommendations.

ACKNOWLEDGEMENT

We would like to express our gratitude to the Head of the Mathematics Education Study Program at HKBP Nommensen University and the students who have supported and been involved in completing this research.

REFERENCE

Akanbi, O. B. (2023). Application of Naive Bayes to Students' Performance Classification. *Asian Journal of Probability and Statistics, 25*(1), 35-47. https://doi.org/10.9734/ajpas/2023/v25i1536

Allaam, M. R. R. (2021). Klasifikasi Genus Tanaman Anggrek Menggunakan Metode *Convolutional Neural Network* (CNN). In *e-Proceeding of Engineering* (pp. 3147-3179). Bandung, Indonesia: Program Studi Sarjana Informatika, Fakultas Informatika, Universitas Telkom.

Anwarudin, A., Andriyani, W., DP, B. P., & Kristomo, D. (2022). The Prediction on the Students' Graduation Timeliness Using Naive Bayes Classification and K-Nearest Neighbor. *Journal of Intelligent Software Systems*, *1*(1), 75-88. https://dx.doi.org/10.26798/jiss.v1i1.597

Awaludin, M., Yasin, V., & Wahyuningsih, M. (2022). Optimization of Naïve

Bayes Algorithm Parameters for Student Graduation Prediction at Universitas Dirgantara Marsekal Suryadarma. *Journal of Information System, Informatic and Computing*, *6*(1), 91-106. https://doi.org/10.52362/jisicom.v6i1.785

Azahari, A., Yulindawati, Y., Rosita, D., & Mallala, S. (2020). Komparasi Data *Mining Naive Bayes* dan *Neural Network* Memprediksi Masa Studi Mahasiswa S1. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *7*(3), 443-452. https://doi.org/10.25126/jtiik.2020732093

Bakri, R., Astuti, N. P., & Ahmar, A. S. (2022). Machine Learning Algorithms with Parameter Tuning to Predict Students' Graduation-on-Time: A Case Study in Higher Education. *Journal of Applied Science, Engineering, Technology, and Education, 4*(2), 259-265. https://doi.org/10.35877/454RI.asci1581

Basuki, M. (2023). Kontribusi Mahasiswa dalam Akreditasi Program Studi. *Nusantara: Jurnal Pengabdian kepada Masyarakat*, *3*(2), 48-54. https://doi.org/10.55606/nusantara.v3i2.1038

Dikriani, D., & Karim, A. T. I. (2023). Comparison of C4.5 and Naive Bayes Algorithm Methods in Prediction of Student Graduation on Time (Case Study: Information Systems Study Program). *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, *3*(1), 40-44. https://doi.org/10.20895/dinda.v3i1.782

Gunawan, G., Hanes, H., & Catherine, C. (2021). C4.5, K-Nearest Neighbor, Naïve Bayes, and Random Forest Algorithms Comparison to Predict Students' on Time Graduation. *Indonesian Journal of Artificial Intelligence and Data Mining*, *4*(2), 62-71. http://dx.doi.org/10.24014/ijaidm.v4i2.10833

Hussen, A. A. M. E., & Saikhu, A. (2024). Modeling of Student Graduation Prediction Using the Naive Bayes Classifier Algorithm. In *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)* (pp. 1-8). Surabaya, Indonesia: Institut Teknologi Sepuluh Nopember.

Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A Systematic Review of the Literature on Machine Learning Application of Determining the Attributes Influencing Academic Performance. *Decision Analytics Journal*, *7*, 100204. https://doi.org/10.1016/j.dajour.2023.100204

Jananto, A., Sulastri, S., Wahyudi, E. N., & Sunardi, S. (2021). Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, *10*(1), 71-78. https://doi.org/10.32736/sisfokom.v10i1.991

Kartianom, K., Arpandi, A., Kassymova, G. K., & Ndayizeye, O. (2022). Prediction Model of Teacher Candidate Student Graduation Status: Decision Tree C4. 5, Naive Bayes, and k-NN. *Ekspose: Jurnal Penelitian Hukum dan Pendidikan*, *21*(2), 1419-1427. https://doi.org/10.30863/ekspose.v21i2.3407

Khasanah, N., Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). Prediksi Kelulusan Mahasiswa dengan Metode *Naïve Bayes*. *Technologia Jurnal Ilmiah*, *13*(3) 207-214. https://doi.org/10.31602/tji.v13i3.7312

Legowo, M. B., & Indiarto, B. (2017). Model Sistem Penjaminan Mutu Berbasis Integrasi Standar Akreditasi BAN-PT dan ISO 9001:2008. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, *1*(2), 90-98. https://doi.org/10.29207/resti.v1i2.51

Marzuqi, A., Laksitowening, K. A., & Asror, I. (2021). Temporal Prediction on Students' Graduation using Naïve Bayes and K-Nearest Neighbor Algorithm. *Jurnal Media Informatika Budidarma*, *5*(2), 682-686. https://doi.org/10.30865/mib.v5i2.2919

Mehdi, R., & Nachouki, M. (2023). A Neuro-Fuzzy Model for Predicting and Analyzing Student Graduation Performance in Computing Programs. *Education and Information Technologies*, *28*(3), 2455-2484. https://doi.org/10.1007/s10639-022-11205-2

Mehta, S. (2023). Playing Smart with Numbers: Predicting Student Graduation Using the Magic of Naive Bayes. *International Transactions on Artificial Intelligence*, *2*(1), 60-75. https://doi.org/10.33050/italic.v2i1.405

Mulyana, D. (2006). *Metodologi Penelitian Kualitatif: Paradigma Baru Ilmu Komunikasi dan Ilmu Sosial Lainnya*. Bandung Barat: PT. Remaja Rosdakarya.

Nakhipova, V., Kerimbekov, Y., Umarova, Z., Suleimenova, L., Botayeva, S., Ibashova, A., & Zhumatayev, N. (2024). Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction. *International Journal of Information and Education Technology*, *14*(1), 92-98. https://doi.org/10.18178/ijiet.2024.14.1.2028

Nuraeni, F., Agustin, Y. H., Rahayu, S., Kurniadi, D., Septiana, Y., & Lestari, S. M. (2021). Student Study Timeline Prediction Model Using Naïve Bayes Based Forward Selection Feature. In *2021 International Conference on ICT for Smart Society (ICISS)* (pp. 1-5). Bandung, Indonesia: Institut Teknologi Bandung.

Pangestu, B. H. (2023). Data *Mining* Menggunakan Algoritma *Naïve Bayes Classifier* untuk Evaluasi Kinerja Karyawan. *Jurnal Riset Matematika*, *3*(2), 177-184. https://doi.org/10.29313/jrm.v3i2.2837

Perez, J. G., & Perez, E. S. (2021). Predicting Student Program Completion Using Naïve Bayes Classification Algorithm. *International Journal of Modern Education and Computer Science*, *12*(3), 57-67. https://doi.org/10.5815/ijmecs.2021.03.05

Putra, M., & Harahap, E. (2024). *Machine Learning* pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Random Forest*. *Jurnal Riset Matematika (JRM)*, *4*(2), 127-136. https://doi.org/10.29313/jrm.v4i2.5102

Rachmawati, A. K., & Miasary, S. D. (2024). Performance Analysis of C5. 0 and Naïve Bayes Classification Algorithm for Pattern Recognition of Student Graduates. In *AIP Conference Proceedings* (pp. 1-7). Cambridge, United States of America: Harvard University.

Rawal, A., & Lal, B. (2023). Predictive Model for Admission Uncertainty in High Education Using Naïve Bayes Classifier. *Journal of Indian Business Research*, *15*(2), 262-277. https://doi.org/10.1108/JIBR-08-2022-0209

Rolansa, F., Yunita, Y., & Suheri, S. (2020). Sistem Prediksi dan Evaluasi Prestasi

Akademik Mahasiswa di Program Studi Teknik Informatika Menggunakan Data *Mining*. *Jurnal Pendidikan Informatika dan Sains*, *9*(1), 75-85. https://doi.org/10.31571/saintek.v9i1.1696

Sembiring, M. T., & Tambunan, R. H. (2021). Analysis of Graduation Prediction on Time Based on Student Academic Performance Using the Naïve Bayes Algorithm with Data Mining Implementation (Case study: Department of Industrial Engineering USU). *IOP Conference Series: Materials Science and Engineering*, *1122*(1), 12069. https://doi.org/10.1088/1757-899X/1122/1/012069

Viet, T. N., Le Minh, H., Hieu, L. C., & Anh, T. H. (2021). The Naïve Bayes Algorithm for Learning Data Analytics. *Indian Journal of Computer Science and Engineering*, *12*(4), 1038-1043. https://doi.org/10.21817/indjcse/2021/v12i4/211204191

Wijiyanto, W., Pradana, A., & Sopingi, S. (2024). Perbandingan Data untuk Memprediksi Ketepatan Studi Berdasarkan Atribut Keluarga Menggunakan *Machine Learning*. *JIKA (Jurnal of Informatics) Universitas Muhammadiyah Tangerang*, *8*(2), 221-228. https://doi.org/10.31000/jika.v8i2.10752

Zheng, X., & Li, C. (2024). Predicting Students' Academic Performance through Machine Learning Classifiers: A Study Employing the Naive Bayes Classifier (NBC). *International Journal of Advanced Computer Science and Applications*, *15*(1), 994-1008. https://doi.org/10.14569/ijacsa.2024.0150199