JOURNAL ON COMMUNICATIONS

ISSN:1000-436X

Scopus

REGISTERED

www.jocs.review

EVALUATION OF DATA MINING MODELS FOR STUDENT GRADUATION PREDICTION USING NAÏVE BAYES, KNN, AND C4.5 ALGORITHMS

Yuri Dimas Arta Ginanjar [1] , Abdul Syukur [2] and Moch Arief Soeleman [3]

[1] *Universitas Dian Nuswantoro*
[2] *Universitas Dian Nuswantoro*
[3] *Universitas Dian Nuswantoro*

Abstract: *This research was conducted to help colleges or universities monitor or predict student graduation. Admission selection is carried out to attract high-quality students. Student quality can be measured by the length of their stay at the university. Graduated on time with the title of Outstanding Student. The number of degrees has a significant impact on the university's accreditation ranking. Many factors, such as study program, admission route, gender, and Semester Achievement Index, have a significant impact on when a student can graduate in their studies. These variables are processed using the KNN (K-Nearest Neighbors) algorithm, Naïve Bayes, C4.5. Data preprocessing uses student data such as study program, gender, admission route, and semester Achievement Index from semesters 1 to 6. Based on graduation data, 452 students graduated on time and 48 students graduated late. All data is taken from the graduation database from 2018 to 2022.*

Keywords: **Data mining, Naïve Bayes Algorithm, K-Nearest Neighbor (KNN), C4.5 Algorithm, Student graduation prediction**

1. INTRODUCTION

Higher education is required to not only organize the learning process but also to map and predict student academic success more accurately and measurably. Student graduation predictions are crucial as a basis for decision-making in academic policy development, development planning, and early detection of students at risk of not graduating on time. Through a systematic approach, educational institutions can improve the quality of academic services and the efficiency of resource management.
The development of information technology has encouraged the use of academic data stored in information systems as a source of new knowledge through the application of data mining techniques. Data mining allows the extraction of hidden patterns from historical student data, such as grades, attendance, and academic history, so that it can be used to build more objective graduation prediction models. In education, the use of data mining has developed into a specialized branch known as educational data mining, which focuses on analyzing student data to support academic decision-making.

Various classification algorithms in data mining, such as Naive Bayes, K-Nearest Neighbor (KNN), and C4.5, are widely used for prediction tasks because they are able to group data into specific classes based on patterns learned from training data. Each algorithm has different characteristics, advantages, and disadvantages, both in terms of accuracy, data requirements, and computational complexity. Therefore, a comprehensive model evaluation is needed to compare the performance of these three algorithms in the context of student graduation prediction, so that the most appropriate and optimal algorithm can be determined for use in higher education environments.

1.1.Formulation of the problem

1 How accurate are the naive Bayes, KNN, and C4.5 methods in predicting student graduation in real time?

2 How intense is the prediction of on-time graduation using the Naïve Bayes, KNN, and C4.5 methods that occur in the student environment?

1.2.Objective

1. Can choose the right method to predict student graduation accurately by using the naive Bayes, KNN, and C4.5 methods.
2. To determine the accuracy level of the naive Bayes, KNN, and C4.5 methods in predicting student graduation.

2. LITERATURE REVIEW

2.1. Understanding data mining

Data mining is the process of extracting hidden information or patterns from large data sets using statistical techniques, machine learning, and artificial intelligence. In education, the development of data mining has given rise to a specialized branch called educational data mining, which focuses on utilizing student academic data to support decision-making, such as predicting graduation rates, detecting dropout risks, and mapping learning performance.

2.2. Data mining in higher education

Various studies have shown that data mining can be used to predict student graduation or on-time study based on historical data such as GPA, number of credits, attendance, and demographic attributes. Classification models built from this data help universities identify students at risk of late graduation, allowing for early academic intervention. Furthermore, the application of data mining also contributes to improving service quality and study program accreditation through more evidence-based data management.

2.3. Naive Bayes Algorithm

K-Nearest Neighbor is a proximity-based classification algorithm, where new data is classified based on the majority class of its k nearest neighbors in the feature space. KNN is non-parametric, making it easy to implement, but it is sensitive to the choice of k value, data scale, and number of features. In the context of student graduation prediction, KNN is used to group students into categories of graduating on time or not, by exploiting similarities in academic patterns with previous students.

2.4. K-Nearest Neighbor (KNN) Algorithm

K-Nearest Neighbor is a proximity-based classification algorithm, where new data is classified based on the majority class of its k nearest neighbors in the feature space. KNN is non-parametric, making it easy to implement, but it is sensitive to the choice of k value, data scale, and number of features. In the context of student graduation prediction, KNN is used to group students into categories of graduating on time or not, by exploiting similarities in academic patterns with previous students.

2.5. C4.5 Algorithm (Decision Tree)

C4.5 is a development of the ID3 decision tree algorithm that constructs decision trees based on selecting the best attributes using the gain ratio criterion. The resulting decision tree is easy to understand because it represents if-then rules that explain the classification rationale. Several studies on student graduation prediction have shown that C4.5 often produces competitive accuracy, and in some cases, superior accuracy, compared to Naïve Bayes and other classification methods.

2.6. Previous research on graduation prediction

Previous research has largely discussed the application of the Naïve Bayes, KNN, and C4.5 algorithms to predict student graduation or on-time study using variables such as GPA, number of credits, attendance, and personal data. Some studies only compare two algorithms, such as Naïve Bayes and C4.5 or decision trees and Random Forest, and report differences in accuracy and other evaluation metrics such as precision, recall, and F1-score. Research that simultaneously examines all three algorithms (Naïve Bayes, KNN, and C4.5) within a single evaluation framework is still relatively limited, thus opening up opportunities to contribute to providing a comprehensive overview of the performance of each algorithm in predicting student graduation.

3. RESEARCH METHODS

3.1. Research method design

This study uses a quantitative method with a data mining approach to build and evaluate a student graduation prediction model using the Naïve Bayes algorithm, K-Nearest Neighbor (KNN), and C4.5. The data used are secondary data in the form of academic data of students who have been declared graduated from three study programs, namely Nautical, Engineering, and Maritime Transportation and Port Management, with a total of 500 data. The class label used is Graduated Status with two categories, namely On Time and Late.

3.2. Data sources and objects

The research data comes from an academic database containing information on students who have graduated from three study programs. The data for each study program is 184 Nautical graduates, 180 Engineering graduates, and 136 Maritime Transportation and Port Management graduates. Each data record represents a single student with several academic and administrative attributes that will be used as predictors for graduation status.

3.3. Data preprocessing stage

The preprocessing stage is carried out to ensure data quality before it is used in modeling. The steps taken include:

Data validation, namely checking the completeness, consistency and suitability of data formats, as well as handling empty or invalid data.

Data integration and transformation, namely combining data from several sources if any, data type alignment, categorical attribute encoding (label encoding/one-hot), and normalization or standardization of numeric attributes if necessary.

Data reduction, namely the selection of relevant attributes and the removal of redundant or less influential attributes on predicting graduation status, so that the model becomes more efficient and reduces the risk of overfitting.

3.4. Application of Naïve Bayes, KNN, and C4.5 algorithms

After preprocessing, the data is divided into training and test data using a specific splitting scheme (e.g., 70% training and 30% test data or k-fold cross-validation). At this stage, three classification algorithms are applied:

Naïve Bayes is used to build a probabilistic model based on the attribute distribution over the On-Time and Off-Time classes.

K-Nearest Neighbor is used by specifying a certain k value and a distance metric (e.g. Euclidean) to classify test data based on its proximity to the training data.

C4.5 is used to build a decision tree based on selecting the best attributes with gain ratio criteria, thus producing a model in the form of classification rules .

3.5. Evaluation and analysis of results

Model evaluation was conducted by comparing the performance of the three algorithms on the test data using classification evaluation metrics, such as accuracy, precision, recall, and F1-score. Additionally, a confusion matrix was used to observe the distribution of correct and incorrect predictions in each of the On-Time and Late classes. The evaluation results were then analyzed to determine the algorithm with the best performance in predicting graduation status and to interpret the patterns formed by the models, particularly in the C4.5 decision tree.

4. RESEARCH RESULTS AND DISCUSSION

4.1. Validity, Integration , transformation and data reduction

The initial data processing process in this study includes three main stages: data validation, data integration and transformation , and data reduction. These three stages are interrelated and produce a final data set ready for use as modeling input with the class label Pass Status (On Time/Late). This stage is crucial to ensure that the data used is clean, consistent, and representative. In the data validation stage, 500 student data were checked for attribute completeness, format consistency, and the presence of duplicate data. Data with missing or illogical critical attributes (e.g., values out of range) were corrected or eliminated according to data quality criteria. After validation, 500 records of data suitable for processing remained, which then served as the basis for the integration and transformation process. The data integration and transformation stage was carried out by combining data from several variable categories, namely admission pathway, study program, and gender, into a single structured fact table. At this stage, categorical values (e.g., Polbit, Regular, Mandiri; Nautika, Teknika, Tata Laksana; Male, Female) are also encoded into numerical form so that they can be processed by the classification algorithm, as well as the alignment of the Pass Status class labels into two categories: On Time and Late.

The data reduction stage is not carried out by reducing the number of records (since all 500 records are still used), but by simplifying and selecting the attributes most relevant to graduation prediction. Variables such as admission pathway, study program, and gender are retained as features because they show distribution variation between the On-Time and Late-Time categories, while redundant or uninformative attributes are removed. This approach aligns with feature selection practices in data mining research for student graduation prediction, which aims to improve the efficiency and accuracy of classification models.

The following is a summary of data distribution after the validation, integration, transformation and data reduction processes.

Table 1: Data Validity

Validity Based on				
Entry Path	Polbit	Appropriate	133	500 student data
		Late	10	
	Regular	Appropriate	123	
		Late	15	
	Independent	Appropriate	196	
		Late	23	
Study program	Nautical	Appropriate	166	500 student data
		Late	19	
	Technique	Appropriate	168	
		Late	11	
	Naval and port administration	Appropriate	118	
		Late	18	
Gender	Man	Appropriate	385	500 student data
		Late	41	
	Women	Appropriate	67	
		Late	7	

Source: 2023 Observation Results

4.2. Naive Bayes

The Naïve Bayes test yielded prediction results from 500 test data sets, yielding an accuracy rate of 90.80%. The precision class for the correct prediction was 95.93% and the accuracy class for the late prediction was 51.72%. The recall class for the correct prediction was 93.81%, while the recall class for the correct prediction was 62.50%.

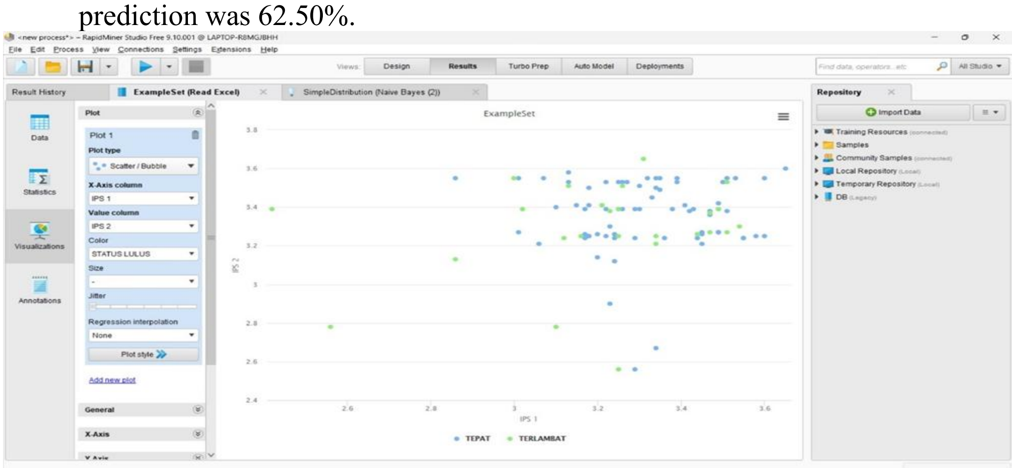


Figure 1. Visualization of Naive Bayes Testing

Table 2: Naïve Bay

Naive Bayes			
	true EXACTLY	true LATE	class precision
pred. RIGHT	424	18	95.93%
pred. LATE	28	30	51.72%
class recall	93.81%	62.50%	
accuracy	90.80%		

Source: 2023 Observation Results

4.2.1. AUC (optimistic)

An optimistic approach to performance evaluation is known as optimistic AUC. This is a method for achieving results consistent with established criteria. The Cross Validation method resulted in an AUC (optimistic) result of 0.809.

4.2.2. AUC neutral

Neutral AUC is a calculation of the measurement of the difference in performance of the method used to produce a comparison of values that match the criteria and values that do not match the specified criteria, the Cross Validation method obtained an AUC value of 0.809.

4.2.3. AUC (pessimistic)

AUC (pessimistic) is a measure of how well a method performs compared to other methods used to produce results that do not meet the specified criteria, the Cross Validation method obtained an AUC (pessimistic) result of 0.585.

4.3. Decision Tree C4.5

The C4.5 Decision Tree model builds a decision tree based on selecting the best attributes using the gain ratio. The C4.5 test results show the following classification performance:

1. The accuracy of the C4.5 algorithm in predicting student graduation is 89.60%.
2. Precision C4.5 indicates the proportion of correct On or Off predictions out of all positive predictions. The precision value is 40.91%.
3. Recall C4.5 shows the proportion of actual data that was correctly predicted. The recall value was 97.12% for true EXACT and 18.75% for true LATE.

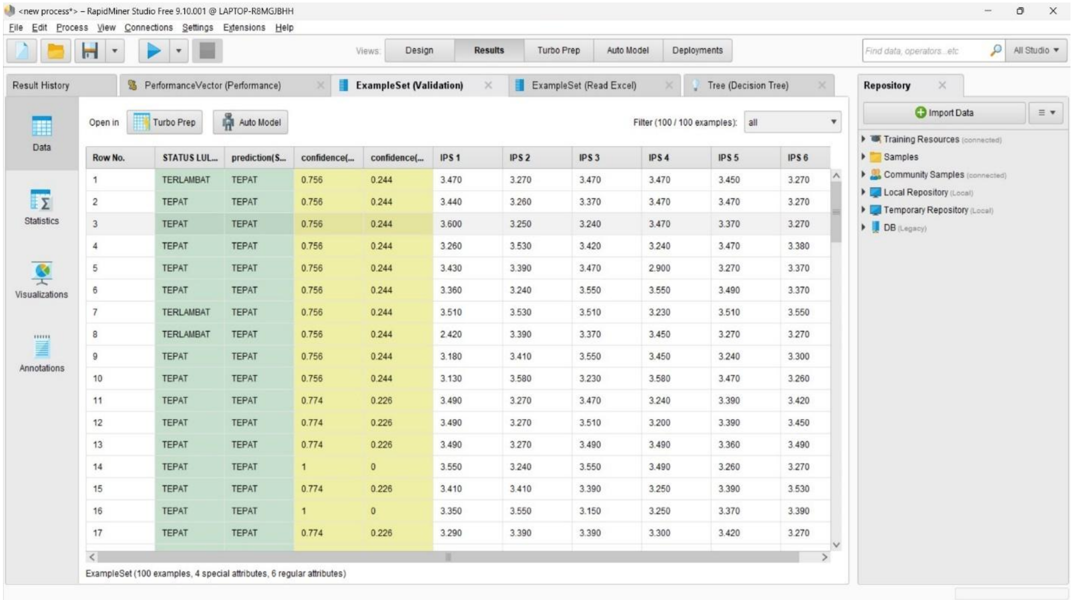


Figure 2 Confidence Value from Decision Tree C4.5 Testing

Table 3. Decision Tree C4.5 Test Results

Naive Bayes			
	true EXACTLY	true LATE	class precision
pred. RIGHT	439	39	91.84%
pred. LATE	13	9	40.91%
class recall	97.12%	18.75%	
accuracy	89.60%		

Source: 2023 Observation Results

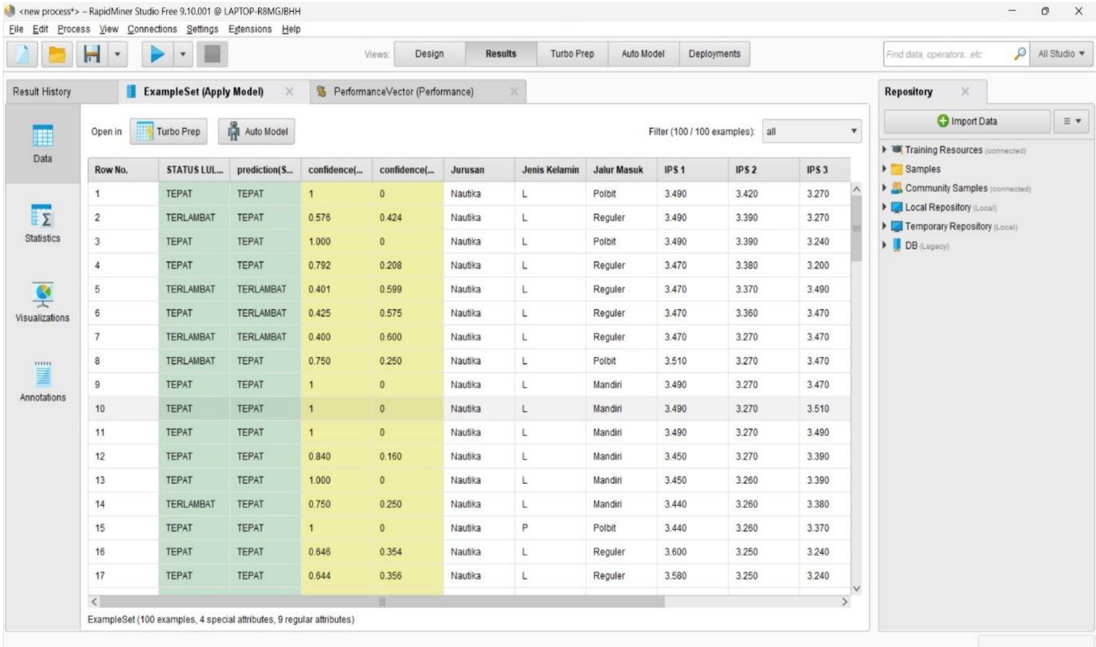
4.3.1. AUC (optimistic)
Testing using the Decision Tree C4.5 method, so that the results obtained from the AUC (optimistic) were 0.839.

4.3.2. AUC neutral
Decision Tree testing obtained an AUC of 0.577, this result was obtained from the results of Decision Tree testing using the cross validation method.

4.3.3. AUC (pessimistic)
The AUC (pessimistic) results generated from the Decision Tree test obtained a value of 0.317.

4.4. K-Nearest Neighbor

The KNN model classifies data based on the proximity of its k nearest neighbors. Test results on the test data demonstrate its performance. K-Nearest Neighbor is a versatile algorithm that offers simplicity and flexibility in solving classification and regression problems. Its use cases span a wide range of domains, making it an essential tool in the field of learning.



Row No.	STATUS LUL...	prediction(S...	confidence(...	confidence(...	Jurusan	Jenis Kelamin	Jalur Masuk	IPS 1	IPS 2	IPS 3
1	TEPAT	TEPAT	1	0	Nautika	L	Polbit	3.490	3.420	3.270
2	TERLAMBAT	TEPAT	0.576	0.424	Nautika	L	Reguler	3.490	3.390	3.270
3	TEPAT	TEPAT	1.000	0	Nautika	L	Polbit	3.490	3.390	3.240
4	TEPAT	TEPAT	0.792	0.208	Nautika	L	Reguler	3.470	3.380	3.200
5	TERLAMBAT	TERLAMBAT	0.401	0.599	Nautika	L	Reguler	3.470	3.370	3.490
6	TEPAT	TERLAMBAT	0.425	0.575	Nautika	L	Reguler	3.470	3.360	3.470
7	TERLAMBAT	TERLAMBAT	0.400	0.600	Nautika	L	Reguler	3.470	3.270	3.470
8	TERLAMBAT	TEPAT	0.750	0.250	Nautika	L	Polbit	3.510	3.270	3.470
9	TEPAT	TEPAT	1	0	Nautika	L	Mandiri	3.490	3.270	3.470
10	TEPAT	TEPAT	1	0	Nautika	L	Mandiri	3.490	3.270	3.510
11	TEPAT	TEPAT	1	0	Nautika	L	Mandiri	3.490	3.270	3.490
12	TEPAT	TEPAT	0.840	0.160	Nautika	L	Mandiri	3.450	3.270	3.390
13	TEPAT	TEPAT	1.000	0	Nautika	L	Mandiri	3.450	3.260	3.390
14	TERLAMBAT	TEPAT	0.750	0.250	Nautika	L	Mandiri	3.440	3.260	3.380
15	TEPAT	TEPAT	1	0	Nautika	P	Polbit	3.440	3.260	3.370
16	TEPAT	TEPAT	0.646	0.354	Nautika	L	Reguler	3.600	3.250	3.240
17	TEPAT	TEPAT	0.644	0.356	Nautika	L	Reguler	3.580	3.250	3.240

ExampleSet (100 examples, 4 special attributes, 9 regular attributes)

Figure 3 KNN Test Table

Table 4 Results of Data Testing Using KNN

Naive Bayes			
	true EXACTLY	true LATE	class precision
pred. RIGHT	451	35	92.80%
pred. LATE	1	13	92.86%
class recall	99.78%	29.08%	
accuracy	92.80%		

Source: 2023 Observation Results

4.4.1. AUC (optimistic)
testing using the K-Nearest Neighbor method, so that the results obtained from the AUC (optimistic) were 0.912.

4.4.2. AUC neutral
From the K-Nearest Neighbor algorithm test, the AUC result was 0.749, this result was obtained from the K-Nearest Neighbor test results using the cross validation method.

4.4.3. AUC (pessimistic)
The K-Nearest Neighbor test using the Cross Validation method obtained a value of 0.585.

4.5. Algorithm Comparison

From the data obtained after going through several stages of sample data testing, varying accuracy values were obtained for the various algorithms used for testing. These results provide a comparison of the various algorithms tested.

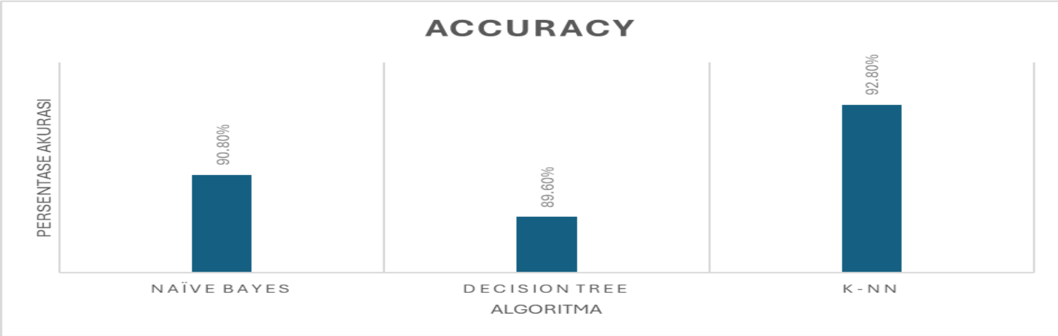


Figure 4. Comparison Results of Accuracy of Naive Bayes, Decision Tree, and KNN Algorithms

Figure 4.17 shows the comparison of the accuracy of the Naïve Bayes, Decision Tree, and K-NN algorithms. The highest accuracy was obtained from the K-NN algorithm with a percentage of 92.80%, followed by the Naïve Bayes algorithm with a percentage of 90.80%, and finally the Decision Tree algorithm with a percentage of 89.60%.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

The results of the study show that all three algorithms—KNN, Naive Bayes, and C4.5 or Decision Tree—can be easily used to classify student graduation data. The K-Nearest Neighbor algorithm has the highest accuracy of 92.80%, the Decision Tree or C4.5 algorithm achieved an accuracy of 89.60%, and the Naïve Bayes algorithm achieved a score of 90.80%. In this study, the K-NN algorithm has a higher level of accuracy when compared to the Naïve Bayes and Decision Tree or C4.5 algorithms.

Among the methods used in this study were KNN, Naive Bayes, and C4.5, as well as Decision Tree, which was validated using the cross-validation method, resulting in an accuracy of 92.80% for the K-Nearest Neighbor algorithm. At this time, the K-Nearest Neighbor algorithm appears to be superior compared to the C4.5 algorithm or Decision Tree and Naïve Bayes. The highest Review result was 92.80 percent, while the highest accuracy value was 92.80% achieved by the K-Nearest Neighbor algorithm. Also, Decision Tree had an AUC of 0.978, which is a good value. As a result, KNN achieved a high AUC of 0.258, which indicates that the model performed poorly or poorly when analyzing data from students. There are a number of different classification algorithms, including KNN, Naive Bayes, C4.5, and Decision Tree, which can be used in conjunction with previous research findings.

5.2. Recommendations

There are also suggestions for writing papers that are similar to or use the same topic as this paper to make it longer.

1. In developing data mining applications, students are influenced by various factors, including parents' occupation, parents' income level, and other supporting factors that can be used as a reference in predicting graduation.
2. For long-term research, it is recommended to choose a different algorithm or increase the volume of relevant data using a similar algorithm.

REFERENCE S

[1] *CD Scott and RE Smalley, "Diagnostic Ultrasound: Principles and Instruments" , Journal of Nanosci. Nanotechnology. , vol. 3, no. 2,* ***(2003)*** *, pp. 75 -80.*

[2] *C. Zhang, P. Zhong, M. Liu, Q. Song, Z. Liang, and X. Wang, "Hybrid Metric K-Nearest Neighbor Algorithm and Applications," Mathematics Probl Eng, vol. 2022, 2022, doi: 10.1155/2022/8212546.*

[3] *F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," International Journal of Information Technology (Singapore), vol. 13, no. 4, pp. 1503–1511, Aug. 2021, doi: 10.1007/s41870-020-00430-y.*

[4] *A. Ella Hassanien Siddhartha Bhattacharyya Satyajit Chakrabati Abhishek Bhattacharya Soumi Dutta Editors, "Advances in Intelligent Systems and Computing 1300 Emerging Technologies in Data Mining and Information Security." [On line]. Available: http://www.springer.com/series/11156*

[5] *P. Kaur, M. Singh, and GS Josan, "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector," in Procedia Computer Science, Elsevier, 2015, pp. 500–508. doi: 10.1016/j.procs.2015.07.372.*

[6] *L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analysis of Student Graduation Prediction on Time Using Naive Bayes Data Mining Method: Systematic Review," Faktor Exacta, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.*

[7] *SS Alfere and AY Maghari, "Prediction of Student's Performance Using Modified KNN Classifiers," 2018. [Online]. Available: https://ssrn.com/abstract=3704733*

[8] *D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System on Students Using C4.5 Algorithm," MATRIK: Journal of Management, Informatics Engineering and Computer Engineering, vol. 19, no. 2, pp. 358–365, May 2020, doi: 10.30812/matrik.v19i2.685.*

[9] *B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, March. 2021, doi: 10.38094/jastt20165.*

[10] *DD Hariyanti, "Combination of Naive Bayes and K-Medoid Methods in Predicting Student Majors in Senior High School".*

[11] *EF Wati and B. Rudianto, "Bina Sarana Informatika University 1 Informatics Engineering, Nusa Mandiri University, 2 Jl. Kramat Raya No. 98, Senen, Central Jakarta 10450 1 Jl," Engineering and Informatics, 2022.*

[12] *AA Priadi, D. Wahdiana, and N. Mutia, "Quality-Based Implementation Model for the Applied Masters in Transportation Postgraduate Program at the Semarang Maritime Polytechnic," Maritime Dynamics, vol. 3, no. 1, pp. 1–19, May 2022, doi: 10.46484/db.v3i1.297.*

[13] *A. Rohman and A. Rufiyanto, IMPLEMENTATION OF DATA MINING WITH THE DECISION TREE C4.5 ALGORITHM FOR STUDENT GRADUATION PREDICTION AT PANDANARAN UNIVERSITY.*

[14] *A. Ashari Muin, "Naive Bayes Method for Graduation Prediction (Case Study: Data of New College Students)," Scientific Journal of Computer Science, vol. 2, no. 1, 2016, [Online]. Available: http://ejournal.fikom-unasman.ac.id*

[15] *A. Rohman, "K-NEAREST NEIGHBOR (K-NN) ALGORITHM MODEL FOR STUDENT GRADUATION PREDICTION."*

[16] *SNH Bukhari, J. Webber, and A. Mehbodniya, "Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates," Sci Rep, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-11731-6.*

[17] *M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," Decision Analytics Journal, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.*

[18] *Z. Lyu et al., "Back-Propagation Neural Network Optimized by K-Fold Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam," Materials, vol. 15, no. 4, Feb. 2022, doi: 10.3390/ma15041477.*

[19] *ZA Varzaneh, A. Orooji, L. Erfannia, and M. Shanbehzadeh, "A new COVID-19 intubation prediction strategy using an intelligent feature selection and K-NN method," Inform Med Unlocked, vol. 28, Jan. 2022, doi: 10.1016/j.imu.2021.100825.*

[20] *Z. Lyu et al., "Back-Propagation Neural Network Optimized by K-Fold Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam," Materials, vol. 15, no. 4, Feb. 2022, doi: 10.3390/ma15041477.*

[21] *D. Ramos, P. Faria, A. Morais, and Z. Vale, "Using decision trees to select forecasting algorithms in distinct electricity consumption context of an office building," Energy Reports, vol. 8, pp. 417–422, Jun. 2022, doi: 10.1016/j.egyr.2022.01.046. 105-Article Text-142-1-10-20190125".*

[22] *R. Puspita, S. Putri, I. Waspada, and UD Semarang, "Implementation of the C4.5 Algorithm in the Application for Predicting the Graduation of Informatics Study Program Students," 2018.*

[23] *A. Ella Hassanien Siddhartha Bhattacharyya Satyajit Chakrabati Abhishek Bhattacharya Soumi Dutta Editors, "Advances in Intelligent Systems and Computing 1300 Emerging Technologies in Data Mining and Information Security." [On line]. Available: http://www.springer.com/series/11156*

[24] *A. Husejinović, "Credit card fraud detection using naive Bayesian and C4.5 decision tree classifiers," vol. 8, no. 1, pp. 1–5, 2020.*

[25] *DD Hariyanti, "Combination of Naive Bayes and K-Medoid Methods in Predicting Student Majors in Senior High School".*

[26] *L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analysis of Student Graduation Prediction on Time Using Naive Bayes Data Mining Method: Systematic Review," Faktor Exacta, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.*

[27] *"Comparison of Naive Bayes and K-NN Method on Tuition Fee Payment Overdue Prediction."*

[28] *A. Ashari Muin, "Naive Bayes Method for Graduation Prediction (Case Study: Data of New College Students)," Scientific Journal of Computer Science, vol. 2, no. 1, 2016, [Online]. Available: http://ejournal.fikom-unasman.ac.id*

[29] *SS Alfere and AY Maghari, "Prediction of Student's Performance Using Modified KNN Classifiers," 2018. [Online]. Available: https://ssrn.com/abstract=3704733*

[30] *A. Rohman, "K-NEAREST NEIGHBOR (K-NN) ALGORITHM MODEL FOR STUDENT GRADUATION PREDICTION."*

[31] *B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, March. 2021, doi: 10.38094/jastt20165.*