Prediction of Informatics Engineering Student Graduation using Naïve Bayes Method

Antonius Suban Koten[1], Anggela M Bouk[2], Fatchulloh Reza Ar Rozi[3], Imam Auwal Salisu[4*]
[1,2,3] Department of Informatics Engineering, University of Widya Gama Malang, Jl. Borobudur No. 35 Malang, Indonesia
[4] Department of Economic Management, Centre for Management Development, Nigeria

Article Info

Article history:

Received January 05, 2026
Revised February 20, 2026
Accepted February 25, 2026

Keywords:

Naive bayes, Graudation prediction

ABSTRACT

Student graduation is one of the indicators of the success of the educational process in higher education. This study aims to predict the graduation of students in the Informatics Engineering study program using the Naive Bayes method, by considering the Final Semester Exam (UAS), Mid-Semester Exam (UTS), assignments, and attendance as the main variables. The Naive Bayes method was chosen because of its simplicity in handling multivariable data and its ability to produce accurate classification models.

Corresponding Author:

Imam Auwal Salisu
Department of Economic Management
Centre for Management Development
Nigeria
Email: auwalimam14@gmail.com

1. INTRODUCTION

Higher education plays a very important role in forming quality human resources who are ready to compete in the professional world. In this era of globalization, the quality of education is one of the main factors that determines the success of individuals in the future. One of the main indicators used to measure the success of education is the timely graduation of students. However, to achieve timely graduation, students need special attention in terms of education management and academic evaluation. Universities face major challenges in ensuring that students not only graduate, but also graduate on time and with good quality. Therefore, predicting student graduation is an important thing to pay attention to. Predicting student graduation can help universities take more effective and targeted actions to improve the quality of education and help students who are at risk of experiencing delayed graduation. One approach that can be used to predict student graduation is by using Data Mining techniques, which allow processing large and complex academic data to produce useful patterns or information. In this context, data such as Final Semester Exam (UAS) scores, Mid-Semester Exam (UTS) scores, assignment scores, and student attendance rates can be used as variables that affect their graduation.

The Naive Bayes method is one of the classification methods in Data Mining that can be used to predict student graduation. Naive Bayes is a probabilistic algorithm that works based on the assumption that each attribute (variable) in the data is independent of each other. Although this assumption rarely occurs in reality, the Naive Bayes method has proven effective in producing accurate predictions with relatively fast computing time. Therefore, Naive Bayes is a good choice for processing diverse academic data and producing a classification model that can predict student graduation quite well. Several previous studies have used various methods to predict student graduation. For example, a study by Pratama, Wihandika, and Ratnawati (2018) used the Support Vector Machine (SVM) algorithm to predict the timeliness of student graduation with an accuracy of 80.55%. Another study by Putri and Waspada (2018) used the C4.5 algorithm to predict student graduation and produced an accuracy of 62.44%, using variables such as gender, region of origin, GPA, and TOEFL score as factors influencing graduation. However, the results of this study indicate the potential to improve prediction accuracy, especially by including more specific and relevant data, such as final exam scores, mid-term exam scores, assignments, and attendance.

The Naive Bayes method offers a simple and efficient approach to handling such data. In this study, Naive Bayes will be used to predict the graduation of Informatics Engineering students using academic variables including UAS, UTS, assignments, and attendance scores. By using this method, it is expected that a system can be built that can help lecturers and universities monitor students' academic progress and provide earlier intervention if needed. The purpose of this study is to build a prediction model for Informatics Engineering students' graduation using the Naive Bayes method, which can provide more accurate predictions regarding the likelihood of students graduating based on their academic data. With this system, lecturers and university administrators can more easily identify students who are at risk of experiencing delayed graduation and take appropriate action to help them. This study also aims to provide contribution in improving the efficiency of higher education management and improving the quality of learning in the Informatics Engineering study program.

2. METHOD

1.1 Data collection

In this study, a document analysis approach was used to collect the necessary data. The data collected was in the form of a list of grades that had been taken by students, which included the Final Semester Exam (UAS) scores, Mid-Semester Exam (UTS) scores, assignment scores, and student attendance levels. This approach allows the study to gain a deep understanding of the context and complexity of the factors that influence student graduation. The documents analyzed came from the academic information system.

1.2 Naive Bayes Algorithm

Naive Bayes Classifier is known as a simple Bayesian classifier and has become an important and successful probabilistic model in practice. Despite its strong independence assumption, Naive Bayes Classifier has proven effective in text classification, medical diagnosis and computer performance management.

$$P(X|C) = \frac{P(X|C).P(C)}{P(X)}$$

Information :

- P(C|X) is the outcome we want to predict (pass or fail).
- P(X|C) is the probability of data X (UAS, UTS, assignments, and attendance scores) occurring in class C.
- P(C) is the prior probability for class C (e.g., the proportion of students who graduate).

3. RESULTS AND DISCUSSION

3.1. Dataset

The following table shows the results of student evaluation based on mid-term exam, final exam, assignments, attendance, and Pass and Fail grades.

Nim	UTS	UAS	Task	Presence	Information
231552018154813	90	88	90	100%	Passed
231552018154616	95	85	90	100%	Passed
231552018154690	89	80	90	100%	Passed
231552018154777	95	83	90	100%	Passed
231552018154664	80	85	70	80%	Not pass
231552018154699	75	79	60	80%	Not pass

3.2. Implementation of calculations in Naive Bayes

For example, we want to calculate the probability that a student with the following data is declared "Passed".

1. Prior probability P(C)

Calculate the proportion of students who passed (C=passed) and failed (C=failed) from the dataset. Total students are 6, students who passed 4, students who failed 2.
So :

$$P(Lulus) \quad = \frac{4}{6} = 0{,}67$$

$$P(Tidak\ lulus) \quad = \frac{2}{6} = 0{,}33$$

2. Probability Likelihood P(X)|C)
   For example, we use feature x = namely UTS 90, UAS 88, Assignments 90, attendance 100%. After that we calculate each feature in the passing class.

$$P(UTS = 90|Lulus) \quad = \frac{1}{4} = 0{,}25$$

$$P(UAS = 88|\ Lulus) \quad = \frac{1}{4} = 0{,}25$$

$$P(Tugas = 90|\ Lulus) \quad = \frac{4}{4} = 1$$

$$P(Kehadiran = 100\%|\ Lulus) \quad = \frac{4}{4} = 1$$

Combined:
P(X|Pass) = 0.25 x 0.25 x 1 x 1 = 0.0625

- For example, we use feature x = namely UTS 80, UAS 85, Assignments 70, attendance 80%. After that we calculate each feature in the failed class.

$$P(UTS = 80|Tidak\ Lulus) \quad = \frac{1}{2} = 0{,}5$$

$$P(UAS = 85|Tidak\ Lulus) \quad = \frac{1}{2} = 0{,}5$$

$$P(Tugas = 70|Tidak\ Lulus) \quad = \frac{1}{2} = 0{,}5$$

$$P(Kehadiran = 80\%|Tidak\ Lulus) \quad = \frac{2}{2} = 1$$

P(X|Fail) = 0.5 x 0.5 x 1 x 1 = 0.25

3. P(X) is the total probability for data X, which is used for normalization.

So :
P(X) = P(X|Pass) . P(Pass) + P(X|Fail) . P(Fail)
= (0.0625 . 0.67) + (0.25 . 0.33)
= 0.0418 + 0.0825
= 0.1243

After obtaining P(X), we can calculate the posterior probability using Bayes' formula.

$$P(X|C) = \frac{P(X|C).P(C)}{P(X)}$$

- Probability of Passing P(Pass|x)

$$P(X|C) = \frac{P(X|C).P(C)}{P(X)} = \frac{0{,}0625\ .\ 0{,}67}{0{,}1243} = 0{,}3368$$

- Probability of Failing

$$P(X|C) = \frac{P(X|C).P(C)}{P(X)}$$
$$= \frac{0{,}25\ .\ 0{,}33}{0{,}1243}$$
$$= 0.6637$$

3.3 Confusion Matrix

	Passed (Actual)	**Failed (Actual)**
Pass (prediction)	4 (TP)	0 (FP)
Failed (prediction)	0 (FN)	2 (TN)

Information :

- True Positive (TP): 4 students who were predicted to pass and actually passed.
- True Negative (TN): 2 students who were predicted to fail actually failed.
- False Positive (FP): 0 students who were predicted to Pass but actually Failed.
- False Negative (FN): 0 students who were predicted to Fail but actually Passed.

3.4 Matrix evaluation

- Accuracy

$$\text{Accuracy}= \frac{TP+TN}{TP+TN+FP+FN} = \frac{4+2}{4+2+0+0}$$
$$= 0.1\ (\ 100\%)$$

- Precision

$$\text{Precision}= \frac{TP}{TP+FP} = \frac{4}{4+0} = 0.1\ (\ 100\%)$$

- Recall

$$\text{Recall}= \frac{TP}{TP+FN} = \frac{4}{4+0}$$
$$= 0.1\ (\ 100\%)$$

- F1-Score

$$\text{F1-Score}= 2 \text{ x } \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}$$
$$= 2 \text{ x } \frac{0{,}1 \text{ x } 0{,}1}{0{,}1+0{,}1}$$
$$= 0.1\ (\ 100\%)$$

3.5 Results Analysis

The evaluation results of the Naive Bayes model show very good performance in predicting student graduation based on the dataset used. Based on the confusion matrix, the model successfully predicted all students in the dataset correctly, resulting in a True Positive (TP) evaluation value of 4 students who were predicted to graduate actually graduated based on actual data. This shows that the model has good ability in identifying students who will graduate. True Negative (TN) of 2 students who were predicted not to graduate actually did not graduate. This shows that the model can identify students who are at risk of failing with perfect accuracy. False Positive (FP) there are no students who are predicted to pass but actually do not pass, indicating that the model does not make mistakes in predicting graduation. False Negative (FN) there are no students who are predicted to fail but actually pass, indicating that the model does not fail to identify students who meet the graduation criteria. With the confusion matrix Accuracy: 100%, indicating that the model predicts all data correctly without error. Precision: 100%, indicating that all students who are predicted to pass actually pass. Recall: 100%, indicating that the model successfully captures all students who actually pass. F1-Score: 100%, indicating a perfect balance between precision and recall.

4. CONCLUSION

This study shows that the Naive Bayes method is effective in predicting the graduation of Informatics Engineering students based on the scores of the Final Semester Exam (UAS), Mid-Semester Exam (UTS), assignments, and attendance levels. By using this approach, students who are at risk of experiencing delays in graduation can be identified more accurately. The evaluation results show that the model built is able to achieve high accuracy, and makes a positive contribution to the management of higher education. This study also emphasizes the importance of utilizing academic data in improving the quality of education and supporting earlier intervention efforts for students who need help. Therefore, the application of this method can be a useful tool for lecturers and universities in monitoring students' academic progress and taking appropriate actions to improve their learning outcomes.

REFERENCES

[1] P. & P. Nugraha, "Application of Educational Data Mining in Predicting Student Performance in Class: Literature Study,"*Infos Journal-Information …*, vol. 2, no. 1, pp. 40–45, 2019, [Online]. Available: https://ojs.amikom.ac.id/index.php/infosjournal/article/view/2427

[2] Andri and Marlindawati, "Grouping of Student Learning Interests Using Data Mining Techniques with the Clustering Method,"*J. Ilm. MATRIX*, vol. 17, no. 1, pp. 67–76, 2015, [Online]. Available: https://media.neliti.com/media/publications/224929-pengelompokan-minat-belajar-mahasiswa-me-c66149b1.pdf

[3] Y. Asriningtias and R. Mardhiyah, "Data Mining Application to Display Student Graduation Rate Information,"*J. Inform.*, vol. 8, no. 1, pp. 837–848, 2014.

[4] DR Yusian, "Analysis of Data Mining Application in Student Admissions at Lhokseumawe State Polytechnic Using the K-Means Algorithm,"*J. Informatics Comput. …*, vol. 4, no. 2, pp. 208–216, 2018, [Online]. Available: http://jurnal.uui.ac.id/index.php/jics/article/view/749

[5] MP Agustin, AA Supianto, and W. Purnomo, "Data Mining Application to Predict Students with Potential to Drop Out Using K-Nearest Neighbor (K-NN) Algorithm,"*J. Development of Information Technology and Computer Science.*, vol. 3, no. 7, pp. 6311–6319, 2019.

[6] H. Susanto and S. Sudiyatno, "Data mining to predict student achievement based on socio-economic, motivation, discipline and past achievement,"*J. Educator. Vocational*, vol. 4, no. 2, pp. 222–231, 2014, doi: 10.21831/jpv.v4i2.2547.