**MATICS** : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
(*Journal of Computer Science and Information Technology)* *Volume* 15, No. 2(2023), pp 96-101
**Received** : November 8th 2023; **Accepted** : November 19th 2023 ; **Avalaible Online** : November 28th 2023

# Comparison of Different Classification Techniques to Predict Student Graduation

Aan Fuad. S, Ririen Kusumawati, M. Imamudin

***Abstract*— Every year, the number of students accepted at the Maulana Malik Ibrahim State Islamic University of Malang continues to increase. However, not all students can be graduated on time according to the specified study period, resulting in a buildup of students who have not been graduated according to their graduation period. One of the aspects that is evaluated in the Study Program accreditation process is the student graduation rate. Apart from that, for each semester, Study Programs are also required to report educational data to DIKTI, and student graduation is one of the factors considered in the report. There is an imbalance between the number of students graduating each year and the number of new students accepted. To overcome this problem, it is necessary to predict student graduation to determine whether they will graduate on time. In science and data analysis, predictions are often used to make estimations based on existing data and information. Classification models in predicting student graduation include the Nave Bayes method, Support Vector Machine SVM, and Random Forest, as well as the level of accuracy of these three methods. From the results of experiments and model evaluations carried out, with data from 458 Informatics Engineering Study Program students, consisting of 366 training data and 92 testing data, it was found that the SVM model had the highest accuracy, reaching around 87%, and Random Forest also had good accuracy, around 82%. At the same time, the Naïve Bayes model has lower accuracy, around 76%.**

***Index Terms*— predictions, Classification models, Naive Bayes, SVM, Random Forest, model evaluation, accuracy.**

## I. INTRODUCTION

The number of students accepted at the Maulana Malik Ibrahim Malang State Islamic University continues to increase yearly. Still, not all students can graduate on time according to the specified study period, resulting in a buildup of students who have not graduated according to their graduation period. [1].

One aspect that is evaluated in the Study Program accreditation process is the student graduation rate. In addition, for each semester, Study Programs are also required to report education data to DIKTI, and student graduation is one of the factors considered in the report. At the Faculty of Science and Technology, UIN Maulana Malik Ibrahim Malang, the Informatics Engineering Study Program has the largest number of students, so the student graduation rate time is very important to increase the assessment and accreditation of the Study Program.[1]

There is an imbalance between the number of students graduating each year and the number of new students accepted. To overcome this problem, it is necessary to predict student graduation to determine whether they will graduate on time. This prediction system requires available information to evaluate a student's chances of graduating on time. Suppose students can find out their graduation status early. In that case, academic authorities can implement appropriate policies to minimize the number of students who do not graduate on time and complete their study period efficiently. [2].

In science and data analysis, predictions are often used to make predictions based on existing data and information. However, we must keep the principles of ethics, integrity, and responsibility in mind when using predictive methods or conducting data analysis. Awareness of our actions and paying attention to their impacts are values that can be applied in the context of predictions and everyday life.

Prediction systematically estimates something most likely to happen based on previous and current available information so that prediction errors (the difference between something that happens and the predicted result) can be minimized [3]. This research aims to carry out a data classification model for students from the Informatics Engineering Study Program, Faculty of Science and Technology, UIN Maulana Malik Ibrahim Malang for 2014 - 2019 as training data and testing data in predicting student graduation using the *Naïve Bayes method, Support Vector Machine (SVM), and Random Forest*, and what is the level of accuracy of the three ways.

## II. LITERATURE REVIEW

Several previous studies examined the comparison of classification methods with data mining, including::

Research conducted by Agus, Yustina & Wawan Application of the K-Nearest Neighbors Algorithm to predict student graduation at STMIK Sinar Nusantara

Manuscript received March 22, 2007. This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University

Aan Fuad, S is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia (email aan.ti@uin-malang.ac..id)

Ririen Kusumawati., with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia (email ririen.kusumawati@ti.uin-malang.ac.id)

M. Imamudin is is the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University, Malang, Indonesia (email imamudin@ti.uin-malang.ac.id)

Surakarta using the K-Nearest Neighbors (k-NN) algorithm. The research results show that the k-NN algorithm can be used to predict student graduation with a good level of accuracy, namely around 81.11%[4].

Research conducted by Rohmawan in 2018 predicted student graduation on time using the Decision Tree and Artificial Neural Network methods. Aims to predict student graduation on time using the Decision Tree and Artificial Neural Network (ANN) methods. The research results show that the Decision Tree method can predict student graduation on time with an accuracy of 79.76%, while the ANN has an accuracy of 84.86%. In addition, factors that influence students' graduation on time, such as GPA, number of credits taken, and classes taken, have also been identified in this research. [5].

Research conducted by Thaniket in 2019 predicted student graduation on time using the Support Vector Machine (SVM) algorithm. The research results show that the SVM method produces quite high prediction accuracy, namely around 85.3%[2].

Research conducted by Hozairi in 2021 Implementation of orange data mining for student graduation classification using the K-Nearest Neighbor, Decision Tree, and Naïve Bayes models. Aims to compare three different classification methods in predicting student graduation. The methods used are K-Nearest Neighbor (KNN), Decision Tree, and Naïve Bayes, which are implemented using Orange Data Mining software. The data used is student data from the Informatics Engineering study program at Madura Islamic University, class of 2016, with 35 test data used. The data attributes in this research are NIM, Student Name, Gender, Student Age, temporary GPA for semesters 1-8, and graduation status. Data was analyzed using the Orange Data Mining application using three algorithms, namely K-NN, Decision Tree, and Naïve Bayes. The data testing process applies K-Fold Cross Validation (K=5), while the evaluation models used are Confusion Matrix and ROC. The comparison results of the three algorithms are: K-NN has an accuracy level of 77%, precision of 76%, Decision Tree has an accuracy level of 74%, precision of 84%, and Naïve Bayes has an accuracy level of 89%, precision 88%. It can be concluded that from the 35 trial data used in this research, the best results for predicting student graduation rates were using the Naïve Bayes algorithm.e high prediction accuracy, namely around 85.3% [6].

Research conducted by Sabathos Mananta and Arther Sandag in 2021 aims to predict student graduation in choosing a master's program using the K-Nearest Neighbor (K-NN) algorithm. This research uses data from students who have graduated and taken a master's program at a university in Indonesia. The K-NN method is used to classify students based on factors influencing graduation, such as average GPA, age, gender, and length of study. The research results show that the K-NN algorithm can predict student graduation with fairly high accuracy, with an accuracy value of 96.25%, precision of 98.08%, and recall of 70.00% for the independent test, while for test results that use cross-validation has an accuracy of 91.88%, precision of 81.29%, and recall of 61.15%. This research provides important information for universities in selecting students who have the potential to succeed in Master's programs and providing appropriate guidance for students who are at risk of not graduating. [7].

Performance of two algorithm methods, namely C4.5 and Naïve Bayes. The research results show that both algorithm methods can be used to predict student graduation with a fairly high level of accuracy. However, of the two ways, the C4.5 algorithm performs better in predicting student graduation with an accuracy rate of 82.7% compared to Naïve Bayes, which has an accuracy rate of 79.8%. Therefore, this research suggests using the C4.5 algorithm in predicting student graduation. [3].

Research conducted by Oon, Darmawan, Lim, and Susanti in 2022 studied the student graduation classification process using the Random Forest method. The classification variables formed are based on several categories, namely on time with a high GPA, on time with a medium GPA, on time with a low GPA, not on time with a high GPA, not on time with a medium GPA, and not on time with a low GPA. In testing the Random Forest method using the Python programming language, an accuracy rate of 0.98 or 98% was found. With an accuracy rate of 98%, the Random Forest method for classifying student graduation is considered very accurate. [8].

## III. RESEARCH METHOD

This case study is aresearch that uses modeling to project student graduation according to schedule by applying Data Mining Classification techniques with the Naïve Bayes method (Support Vector Machine), SVM, and Random Forest (RF).

The design of our proposed research system is

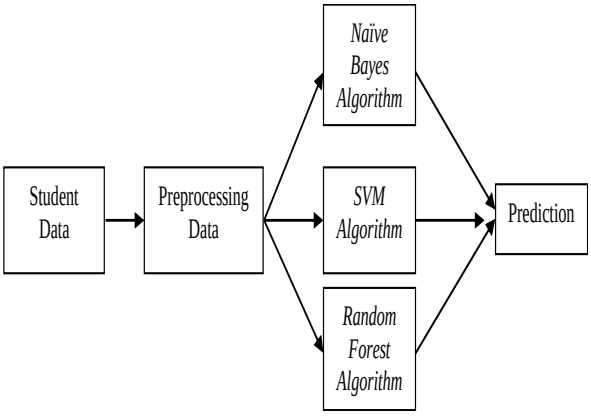


Figure 1. System Development Design

described in Figure 1 with student data and then goes through a data preprocessing process. After the preprocessing stage, the data can be processed with a predetermined algorithm. The following parameters, type features, and student data that have been preprocessed can be seen at.

| Parameter | Type |
|---|---|
| NIM | Numeric |
| Nama Mahasiswa | Text |
| Jenis Kelamin | Categorical (laki-laki, perempuan) |
| Asal Sekolah | Categorical (MAN, MAS, SMEA, SMKN, SMKS, SMUN, SMUS) |
| Propinsi Sekolah | Categorical |
| Pernah Mondok di Pesantren | Categorical (ya, tidak) |
| Jalur PMB | Categorical (Beasiswa luar negeri, Mandiri asing, Mandiri Prestasi, Mandiri Tertulis, Mandiri Tertulis Bidikmisi, Santri Berprestasi, SBMPTN, SBMPTN Bidikmisi, SNMPTN, SNMPTN Bidikmisi, SPAN-PTKIN, SPAN-PTKIN Bidikmisi, UM-PTKIN) |
| IP Semester 1 | Numeric |
| IP Semester 2 | Numeric |
| IP Semester 3 | Numeric |
| IP Semester 4 | Numeric |
| IP Semester 5 | Numeric |
| IPK Semester 1-5 | Numeric |
| Status | Categorical (Tepat Waktu, Tidak Tepat) |

Figure 2. Parameters and Fitur Type

### *A. Naïve Bayes*

The first method that will be used is the Naïve Bayes method. This algorithm produces a model in the form of class probability that can predict whether a student will graduate on time based on the appearance of features in the data. The advantage of this algorithm is that it is fast in data processing and effective on large datasets [9]. The stages of the Naïve Bayes method in this research can be seen in Figure 3 below:

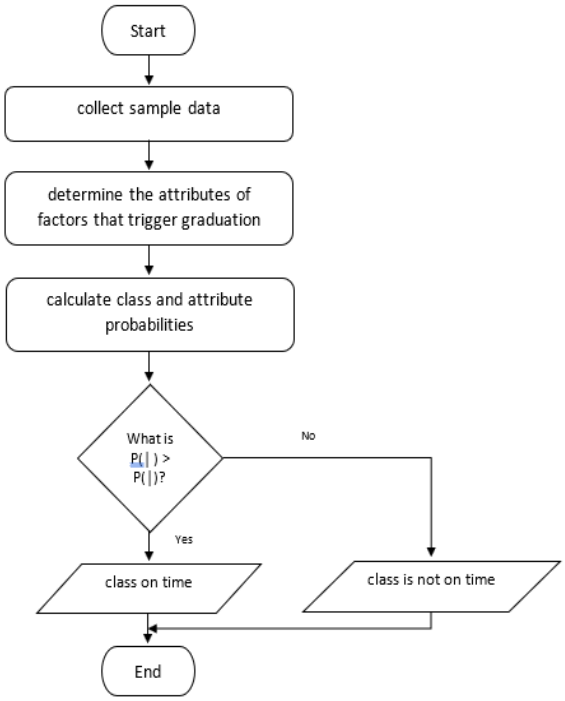


Figure 3. Naïve Bayes Flowchart

### *B. SUPPORT VECTOR MACHINE (SVM)*

Support Vector Machine (SVM) is a relatively new technique used to make predictions, both in the case of classification and regression, and is considered one of the best techniques currently [10]. This technique was first introduced in 1995 and is included in the Supervised Learning class, the same as Artificial Neural Network (ANN). However, SVM has more accurate performance in predicting the type of new data, so it has become a technique that is widely used in the fields of Data Mining and Machine Learning. The following is the flow of the Random Forest algorithm as in Figure 4.

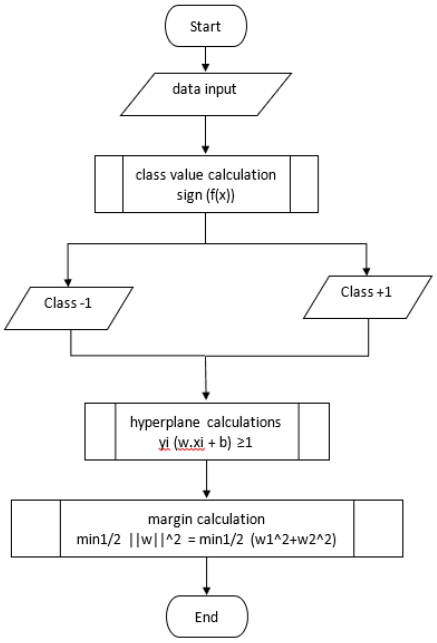


Figure 4. SVM Flowchart

In this technique, data is entered into the system. Next, the class value will be searched using an equation formula, where the class value can be class +1 (On Time) or class -1 (Not Exact). Once the class is known, this technique will look for two data from different types with the closest distance (support vector). From these two data, this technique will look for the dividing line (hyperplane) using the equation formula. Finally, this technique will find the maximum distance from the hyperplane with a support vector (margin) using an equation to find the best hyperplane from the two data.

SVM is used to separate two classes of data by looking for a hyperplane that has the largest distance to data samples from both types. The hyperplane functions to maximize the margin between the two types and minimize classification errors. A linear equation defines this hyperplane:

$$\boldsymbol{w.x + b = 0}$$

Where w is the weight vector, and b is the bias. To classify a new sample x, we can use the signum function of the hyperplane equation:

$$\boldsymbol{f(x) = sign(wtx + b)}$$

If f(x) > 0, then sample x is classified as a positive class, and if f(x) < 0, then sample x is classified as a negative class. The SVM formulation can be optimized by maximizing the margin between the two types and minimizing the norm of the weight vector with the following notation:

$$\boldsymbol{1/2\ ||w||\hat{}2}$$

### *C. RANDOM FOREST (RF)*

The Random Forest algorithm is one of the algorithms included in supervised learning. This algorithm can solve classification or regression problems. Apart from being simple, this algorithm can

be used to solve classification or regression problems [11].

The Random Forest algorithm is included in ensemble learning. This ensemble learning works with several methods that work together to carry out its performance. Because working together, the version of this model is better than the model that works alone. Because this model works with several methods, the final result of the values will be combined. For the regression case, the combination of values uses the mean or average value; for the classification case, the highest occurrence of the mode value is used [12].

The Random Forest method is a development of the CART method, namely by applying the bootstrap aggregating (bagging) and random feature selection methods. In Random Forest, many trees are grown to form a forest; then, analysis is carried out on this collection of trees [13]. The following is the path of the Random Forest algorithm as shown in Figure 5.

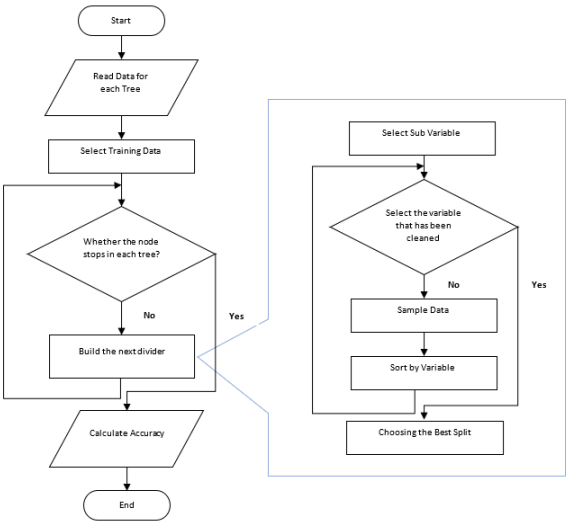


Figure 5. Random Forest Flowchart

The steps in the Random Forest algorithm are as follows:

1. Enter data into each tree (tree)
2. Choosing training data, for example, taking 80% as training data and the remaining 20% as testing data.
3. Evaluate whether each knot stops in each tree or not.
   a) If the node does not stop, then build the next separator by selecting sub-variables and then selecting variables that have been cleaned.
   b) If the knot has stopped, then choose the best split.
   c) If the node has not stopped, select the sample data, sort it based on variables, and repeat the process to get the best split.
   d) Repeat the above process until the knot stops at each tree. e) If the knot has stopped, then calculate the accuracy value.
   e) If the knot has stopped, then calculate the accuracy value.

## IV. RESULT

From the results of the analysis of the count of 3 methods, namely Naïve Bayes, Support Vector Machine, and Random Forest, testing will be carried out using Python programming with a dataset of 458 data on Informatics Engineering Study Program Students of UIN Maulana Malik Ibrahim Malang. The dataset is divided into training data, which amounts to 80% of the total data, and will be used as testing data, totaling 20%. The following is the process of testing data on the classification system using Python programming, as shown in Figure 6.

| | NIM | Nama Mahasiswa | Jenis Kelamin | Asal Sekolah | Propinsi Sekolah | Pernah Mondok di Pesanten | Jalur PMB | IP1 | IP2 | IP3 | IP4 | IP5 | IPK1-5 | Status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 14650001 | ASIFATUL MU`AWANAH | Perempuan | MAN | JAWA TIMUR | Tidak | SPAN-PTKIN | 3.57 | 3.46 | 3.50 | 3.75 | 3.48 | 3.552 | TEPAT WAKTU |
| 1 | 14650002 | ACHMAD FATHULLAH | Laki-laki | SMUN | JAWA TIMUR | Tidak | SNMPTN | 3.71 | 3.43 | 3.73 | 3.60 | 3.65 | 3.624 | TIDAK TEPAT |
| 2 | 14650005 | FARHATUL AMMAH | Perempuan | SMUS | JAWA TENGAH | Ya | SNMPTN | 3.71 | 3.60 | 3.43 | 3.55 | 3.10 | 3.478 | TIDAK TEPAT |
| 3 | 14650006 | FATHURRAHMAN | Laki-laki | SMUN | SULAWESI TENGGARA | Tidak | SNMPTN | 3.33 | 3.42 | 3.00 | 3.27 | 3.08 | 3.220 | TIDAK TEPAT |
| 4 | 14650007 | JANITA PUTRI MEIYANA ERMANSYAH | Perempuan | SMUN | NUSA TENGGARA BARAT | Tidak | SNMPTN | 3.38 | 3.00 | 3.29 | 3.50 | 3.04 | 3.242 | TIDAK TEPAT |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 453 | 19650135 | BISYRI SYAMSURI | Laki-laki | MAN | JAWA TIMUR | Tidak | MANDIRI TERTULIS | 3.62 | 3.54 | 3.83 | 3.58 | 3.70 | 3.654 | TEPAT WAKTU |
| 454 | 19650138 | MIFTAH FURQAANUL HAQ | Laki-laki | SMUN | JAWA TIMUR | Tidak | MANDIRI TERTULIS | 3.33 | 3.70 | 3.60 | 3.23 | 3.42 | 3.456 | TEPAT WAKTU |
| 455 | 19650142 | ACHMAD RIDHA | Laki-laki | SMUS | SULAWESI SELATAN | Tidak | MANDIRI TERTULIS | 3.48 | 3.61 | 3.59 | 2.46 | 3.28 | 3.284 | TEPAT WAKTU |
| 456 | 19650153 | YASSIN MOHAMED ZAKARIA ABDELFATAH ALI | Laki-laki | MAN | LUAR NEGERI | Tidak | MANDIRI ASING | 3.79 | 3.54 | 3.48 | 2.58 | 3.60 | 3.398 | TEPAT WAKTU |
| 457 | 19650156 | NABIL RAHMAD ILLAHI | Laki-laki | SMKN | JAWA TIMUR | Tidak | MANDIRI TERTULIS | 3.14 | 3.41 | 3.67 | 3.25 | 3.22 | 3.338 | TEPAT WAKTU |

458 rows × 14 columns

Figure 6. Student Data

The dataset in the picture above contains data attributes that will be processed in the classification process with three methods; before the classification process is carried out, a column drop process will be carried out that is not needed. Columns that will be processed for classification contain gender, school origin, school province, ever boarding at a pesantren, PMB path, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IP Semester 5, GPA 1 to 5, with labeling status on time or not on time. The data to be tested is shown in Figure 7.

| | Jenis Kelamin | Asal Sekolah | Propinsi Sekolah | Pernah Mondok di Pesanten | Jalur PMB | IP1 | IP2 | IP3 | IP4 | IP5 | IPK1-5 | Status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Perempuan | MAN | JAWA TIMUR | Tidak | SPAN-PTKIN | 3.57 | 3.46 | 3.50 | 3.75 | 3.48 | 3.552 | TEPAT WAKTU |
| 1 | Laki-laki | SMUN | JAWA TIMUR | Tidak | SNMPTN | 3.71 | 3.43 | 3.73 | 3.60 | 3.65 | 3.624 | TIDAK TEPAT |
| 2 | Perempuan | SMUS | JAWA TENGAH | Ya | SNMPTN | 3.71 | 3.60 | 3.43 | 3.55 | 3.10 | 3.478 | TIDAK TEPAT |
| 3 | Laki-laki | SMUN | SULAWESI TENGGARA | Tidak | SNMPTN | 3.33 | 3.42 | 3.00 | 3.27 | 3.08 | 3.220 | TIDAK TEPAT |
| 4 | Perempuan | SMUN | NUSA TENGGARA BARAT | Tidak | SNMPTN | 3.38 | 3.00 | 3.29 | 3.50 | 3.04 | 3.242 | TIDAK TEPAT |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 453 | Laki-laki | MAN | JAWA TIMUR | Tidak | MANDIRI TERTULIS | 3.62 | 3.54 | 3.83 | 3.58 | 3.70 | 3.654 | TEPAT WAKTU |
| 454 | Laki-laki | SMUN | JAWA TIMUR | Tidak | MANDIRI TERTULIS | 3.33 | 3.70 | 3.60 | 3.23 | 3.42 | 3.456 | TEPAT WAKTU |
| 455 | Laki-laki | SMUS | SULAWESI SELATAN | Tidak | MANDIRI TERTULIS | 3.48 | 3.61 | 3.59 | 2.46 | 3.28 | 3.284 | TEPAT WAKTU |
| 456 | Laki-laki | MAN | LUAR NEGERI | Tidak | MANDIRI ASING | 3.79 | 3.54 | 3.48 | 2.58 | 3.60 | 3.398 | TEPAT WAKTU |
| 457 | Laki-laki | SMKN | JAWA TIMUR | Tidak | MANDIRI TERTULIS | 3.14 | 3.41 | 3.67 | 3.25 | 3.22 | 3.338 | TEPAT WAKTU |

458 rows × 12 columns

Figure 7. Dataset to be tested

The results of the three machine learning models with 366 training data and 92 testing data will be seen and compared from the accuracy value. Accuracy is a metric that measures how well the model does in making correct predictions. The higher the accuracy value, the better the model performance. Furthermore, it can compare the accuracy of the three models to determine which model is best at generalizing data that has never been seen during training. The following are the results of the classification model performance of 3 methods, namely Naïve Bayes, SVM, and Random Forest, by looking at the results of Accuracy (Accuracy), Precision (Precision), Recall (Recall or Sensitivity), and F1-Score. The results of

the Confusion Matrix accuracy value of the classification model of the three methods Naïve Bayes, Support Vector Machine (SVM), and Random Forest can be seen in the table and graph below.

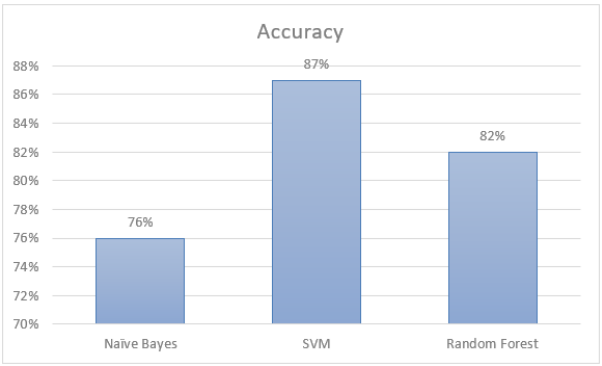


Figure 8. Accuracy result graph

The results of the comparison of the three methods are described in Table 1. The results of the comparison show that the SVM method has higher accuracy than the others.

Table 1. Confusion Matrix Results

| | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| *Naïve Bayes* | 76% | 94% | 73% | 82% |
| SVM | 87% | 88% | 96% | 92% |
| *Random Forest* | 82% | 85% | 91% | 88% |

## V. CONCLUTION

This research has implemented three Data Mining-based classification algorithms, Naïve Bayes, SVM, and Random Forest, to predict student graduation. After conducting experiments and model evaluation, the following are the conclusions:

1. Model Accuracy

The SVM model has the highest accuracy, reaching about 87%. The Random Forest model also has good accuracy, about 82%. The Naïve Bayes model has a lower accuracy, which is about 76%.

2. Relative Performance

In this study, the SVM Model is superior in predicting student graduation, followed by the Random Forest Model and the Naïve Bayes Model.

3. Model Choice

The choice of the most suitable model depends on the goal and the characteristics of the data. If the highest accuracy is the priority, SVM may be the best choice. However, if a balance between accuracy and model interpretation is desired, Random Forest may be a good choice.

4. Additional Evaluation

In addition to accuracy, it is necessary to perform additional evaluations to understand the model's performance in more detail. This can help in assessing specific aspects of the model performance, especially in the discussion of graduating students

## REFERENCES

[1] Mashlahah, *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*. 2013.

[2] R. Thaniket, Kusrini, and E. T. Luthf, "Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. FATEKSA J. Teknol. dan rekayasa*, vol. 13, no. 2, pp. 69–83, 2019.

[3] A. Rahmayanti, L. Rusdiana, and S. Suratno, "Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa," *Walisongo J. Inf. Technol.*, vol. 4, no. 1, pp. 11–22, 2022, doi: 10.21580/wjit.2022.4.1.9654.

[4] P. Agus, Yu. R. W. Utami, and W. L. YS, "Penerapan Algoritma K-Nearest Neighbors Untuk Prediksi Kelulusan Mahasiswa Pada STMIK Sinar Nusantara Surakarta," *TIKomSiN*, vol. Vol 5,No 1, pp. 27–31, 2017.

[5] E. P. Rohmawan, "281628-Prediksi-Kelulusan-Mahasiswa-Tepat-Waktu-42Eb4C1B," *J. Ilm. MATRIK*, p. 3, 2018.

[6] 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.

[7] A. Sabathos Mananta and G. Arther Sandag, "Prediksi Kelulusan Mahasiswa Dalam Memilih Program Magister Menggunakan Algoritma K-NN," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 10, no. 2, pp. 90–96, 2021, doi: 10.30591/smartcomp.v10i2.2488.

[8] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, Dec. 2022, doi: 10.33372/stn.v8i2.885.

[9] M. R. Qisthiano, T. B. Kurniawan, E. S. Negara, and M. Akbar, "Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 987, 2021, doi: 10.30865/mib.v5i3.3030.

[10] O. Bangun, H. Mawengkang, and S. Efendi, "Metode Algoritma Support Vector Machine (SVM) Linier Dalam Memprediksi Kelulusan Mahasiswa," *J. Media Inform. Budidarma*, vol.

6, no. 4, p. 2006, 2022, doi: 10.30865/mib.v6i4.4572.

[11] M. L. Mu'tashim and A. Zaidiah, "Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)," *Pros. Semin. Nas. Mhs. Bid. Ilmu Komput. dan Apl.*, vol. 4, no. 1, pp. 155–166, 2023.

[12] B. Yusuf, M. Qalbi, B. Basrul, I. Dwitawati, M. Malahayati, and M. Ellyadi, "Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh," *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 4, no. 1, p. 50, 2020, doi: 10.22373/cj.v4i1.7247.

[13] G. S. Suwardika and I. K. P. Suniantara, "Analisis Random Forest Pada Klasifikasi Cart Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 13, no. 3, pp. 177–184, 2019, doi: 10.30598/barekengvol13iss3pp177-184ar910.