Comparison of Individual Algorithms (Decision Tree, Naïve Bayes, and Support Vector Machine) and Ensemble Voting in Predicting Students' On-Time Graduation Based on Course Grades

Sevtian Ferdian[1*], **Miechael**[2], **Arief Wibowo**[3]
[1,2,3] Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
Email: 2311601831@student.budiluhur.ac.id[1], 2311601344@student.budiluhur.ac.id[2*], arief.wibowo@budiluhur.ac.id[3]

Abstract

Education plays an important role in improving the quality of human resources and supporting a country's progress toward becoming a developed nation. Higher education institutions serve as one of the providers of formal education, where the quality of these institutions is measured through accreditation. One of the key indicators influencing accreditation is the outcomes and achievements of the Tri Dharma of higher education, which include the timeliness of student graduation. This study aims to compare models for predicting on-time student graduation using three machine learning algorithms, namely Decision Tree, Naïve Bayes, and Support Vector Machine (SVM), as well as their combination through the Ensemble Voting method. The prediction is based on historical grade data from courses taken during semesters one to four. The research methodology adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM), which consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The dataset used in this study consists of 2,471 records with 11 attributes. Data preprocessing was conducted through data cleaning and class balancing using under sampling techniques. The results indicate that the Ensemble Voting model using the Soft Voting method achieves the best performance, with an accuracy of 91.80%, precision of 91.87%, and recall of 91.80%, outperforming the individual models of Decision Tree, Naïve Bayes, and SVM. The implementation of this model can be utilized to predict students' on-time graduation based on course grade inputs. Therefore, this research can serve as a supporting tool for early detection of potential delays in student graduation.

***Key words**: data mining, decision tree, ensemble model, naïve bayes, support vector machine*

INTRODUCTION

Several indicators determine whether a country is categorized as a developed or developing nation, including per capita income, health levels, and education levels. Developed countries generally have high per capita income, as well as higher levels of health and education. In contrast, developing countries tend to have lower per capita income, lower health standards, and lower levels of education (Yuni, Putra and Hutabarat, 2020). One way to transform a developing country into a developed one is by improving the educational level of its population, which simultaneously enhances the quality of human resources (Suratini, 2017).

Higher education institutions, as providers of formal education, play an important role in improving individuals' knowledge, intelligence, and skills. According to data from the Central Statistics Agency (Badan Pusat Statistik/BPS) published in 2023, only 10.15% of the population are higher education graduates (BPS 2022, 2023). This indicates the importance of higher education institutions in supporting national development through the provision of quality formal education.

The quality of a higher education institution is commonly measured through accreditation. Accreditation is an evaluation process used to assess the quality and feasibility of universities or study programs. Several indicators are considered in the accreditation process, including vision, mission, objectives, and strategies; human resources; financial resources, facilities, and infrastructure; as well as the outcomes and achievements of the Tri Dharma of higher education. Within the outcomes and achievements of the Tri Dharma, assessment criteria include the average student GPA over the last three years, employment level and workplace scale of graduates, the number of publications in academic journals within the last three years, and the percentage of students graduating on time (Sasongko, 2019).

Delayed graduation is an issue that requires attention. In addition to lowering accreditation assessment scores for higher education institutions, delayed graduation can also cause disadvantages for students themselves, particularly in terms of time and financial costs. Therefore, this study focuses on developing a predictive model for on-time student graduation based on historical academic data, specifically course grades obtained from the first to the fourth semester.

Several related studies have previously been conducted. One study entitled "Prediction of Student Graduation at the Faculty of Information Technology UMBy Using the Decision Tree Method with the C4.5 Algorithm" utilized five parameters, namely the total number of completed credits, study program, and semester GPA from semesters five to seven. The study successfully generated rules for predicting graduation timeliness with an accuracy of 82.9% (Nurislamiaty and Rozi, 2021).

Another study entitled "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu" applied the Support Vector Machine (SVM) algorithm to predict on-time graduation. The study achieved the best accuracy of 94.4% on the testing dataset, which consisted of 10% of the total data. The dataset included attributes such as age, gender, GPA, semester GPA, major, and total credits (Rohmawan, 2018).

In 2022, a study entitled "Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma C4.5 Pada STMIK Dharma Wacana" utilized six attributes, namely gender, age, and semester GPA from semesters one to four. The study successfully generated classification rules from the dataset; however, the accuracy value was not reported (Mukti, 2022).

Another study entitled "Student Graduation Prediction Using the Naïve Bayes Method" predicted student graduation levels based on several factors such as marital status, employment status, and cumulative GPA. The research applied the Naïve Bayes algorithm to analyze 379 student records, with 303 records used as training data and 76 records used as testing data (Budiyantara, 2019).

A study conducted by Amri et al. (2023) entitled "Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM" aimed to compare the performance of five machine learning algorithms in predicting student graduation rates. The dataset consisted of 807 student records from the Faculty of Engineering at Hamzanwadi University (2015-2018 cohorts), including variables such as gender, age, semester GPA, cumulative GPA, and graduation status. The research employed the Knowledge Discovery in Databases (KDD) methodology, which includes data selection, preprocessing, transformation, data mining, and evaluation. To address class imbalance, the SMOTEENN technique was applied, and the dataset was divided into 80% training data and 20% testing data. The results showed that the K-Nearest Neighbor (KNN) algorithm achieved the highest accuracy of 96.95%, followed by SVM (93.13%), Naïve Bayes (92.37%), Decision Tree (91.60%), and ANN (90.84%). Most predictions indicated delayed graduation influenced by students' cumulative GPA. Based on these findings, the KNN algorithm was recommended as the best-performing model for predicting graduation rates. Educational institutions are also advised to pay greater attention to students predicted to graduate late by improving their GPA each semester to support timely graduation (Amri, Kusrini and Kusnawi, 2023).

Based on the problems described above and the review of previous studies, most existing research utilizes attributes such as semester GPA, cumulative GPA, or demographic factors. Meanwhile, analysis using course grade data from early semesters is still relatively limited. In addition, the use of combined algorithm approaches, such as ensemble voting, has not been widely explored in the context of graduation prediction. Therefore, this study aims to analyze and compare the performance of three machine learning algorithms—Decision Tree, Naïve Bayes, and Support Vector Machine—both individually and in combination using the ensemble voting method. The dataset used consists of course grade data from semesters one to four, whereas most previous studies rely on semester GPA or cumulative GPA. This study is expected to contribute to the early detection of students who are at risk of delayed graduation and to identify the most effective algorithm

(Decision Tree, Naïve Bayes, SVM, or ensemble voting) for predicting on-time graduation.

Decision Tree (DT) is a classification method based on a tree structure that operates by partitioning a dataset into subsets according to specific attributes. This model constructs decision rules in the form of branches and nodes, which ultimately produce predictions at the leaf nodes. Several algorithms are commonly employed in the construction of Decision Trees, including ID3, C4.5, and CART, which differ primarily in their attribute selection criteria, such as information gain, gain ratio, and the Gini index (Nazifah, Prianto and Id, 2023).

Support Vector Machine (SVM) is a classification algorithm designed to identify the optimal hyperplane that separates data into two or more classes. The core principle of SVM is margin maximization, defined as the distance between the hyperplane and the nearest data points, referred to as support vectors. By maximizing this margin, SVM is capable of producing an optimal class separation (Arifin *et al.*, 2023).

Naïve Bayes is a probabilistic classification method based on Bayes' theorem, which assumes independence among features. Although this assumption is rarely fully satisfied in real-world datasets, Naïve Bayes remains effective and computationally efficient in various applications, particularly in text processing and document classification tasks (Tholib *et al.*, 2023).

Ensemble models refer to approaches that integrate multiple classification algorithms to enhance predictive performance. The fundamental principle of ensemble learning is that combining several models can yield predictions that are more robust, stable, and accurate than those produced by a single model. Among the widely adopted ensemble techniques are Hard Voting and Soft Voting (Indahyanti, Azizah and Setiawan, 2022).

In Hard Voting, the final classification outcome is determined through a majority voting mechanism among the base learners. Each model independently generates a prediction, and the class receiving the highest number of votes is selected as the final decision. This approach is particularly effective when the base models exhibit relatively balanced predictive performance and provide complementary decision patterns, thereby improving the robustness of the overall ensemble model (Rahman *et al.*, 2025).

Soft Voting determines the final class by aggregating the predicted probabilities produced by each base learner and computing their weighted average. Compared to Hard Voting, this approach is more flexible as it incorporates the confidence level of each model in assigning class probabilities. Consequently, Soft Voting often yields improved predictive performance, particularly when the base models provide well-calibrated probability estimates (Rahman *et al.*, 2025).

METHOD

The methodology adopted in this study is the Cross-Industry Standard Process for Data Mining (CRISP-DM), a widely recognized framework that serves as a standard guideline for conducting data mining projects. The CRISP-DM framework consists of six main phases, as shown in Figure 1 (Sulianta, 2024)(Parteek Bhatia, 2019).

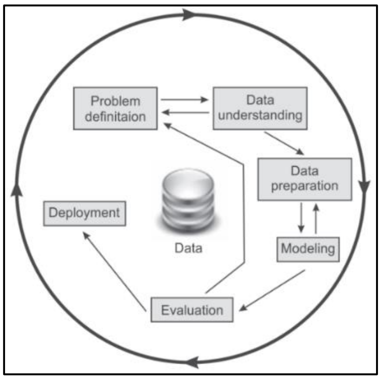


Figure 1. Data Mining Process Based on the CRISP-DM Framework

3.1 Business Understanding

This stage focuses on understanding and identifying the problem domain. Once the problem has been clearly defined, it can be formulated as a research objective to be addressed through the application of data mining techniques.

3.2 Data Understanding

This stage involves acquiring and collecting relevant data from available sources for the data mining process. Once the data have been obtained, a thorough exploration and understanding of the dataset are conducted. This step aims to analyze the characteristics and structure of the data while also assessing whether the dataset is appropriate and sufficient to address the identified research problem.

3.3 Data Preparation

This stage focuses on preparing the dataset through several processes, including data selection, data cleaning, data transformation, and data integration. The data preparation phase is widely recognized as the most time-consuming stage in the data mining workflow, often accounting for up to 90% of the total project time, as it involves ensuring that the dataset is properly structured, consistent, and suitable for subsequent modeling processes.

3.4 Modelling

This stage involves applying appropriate data mining techniques, such as clustering, classification, estimation, and association rule mining, to the prepared dataset. Through the implementation of these analytical methods, the data are processed to extract meaningful patterns and knowledge that can contribute to addressing the identified research problem and providing potential solutions.

3.5 Evaluation

This stage focuses on evaluating the developed model to assess its performance and overall quality. The evaluation results provide insights into whether the model is sufficiently effective in addressing the identified research problem. If the obtained results are not satisfactory, the modeling process may be repeated with further refinements, such as optimizing algorithm parameters, improving data quality, or adjusting the modeling strategy. The evaluation of the model is conducted using several performance metrics, including:

3.5.1 Accuracy

Accuracy is a commonly used evaluation metric that measures the proportion of correctly classified instances relative to the total number of predictions made by the model. It reflects the overall effectiveness of a classification model in correctly identifying both positive and negative instances. Mathematically, accuracy is defined as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

3.5.2 Recall

Recall is an evaluation metric that measures the ability of a classification model to correctly identify all relevant positive instances. It represents the proportion of actual positive cases that are successfully detected by the model, thereby indicating how effectively the model minimizes false negatives. Mathematically, recall is defined as (Parteek Bhatia, 2019):

$$\text{Recall} = \frac{\text{True Positive}}{\text{All Actual Positive}} \quad (2)$$

3.5.3 Precision

Precision is a performance evaluation metric that measures the proportion of correctly predicted positive instances relative to the total number of instances predicted as positive by the model. It reflects the reliability of the model when identifying positive cases, indicating how frequently positive predictions correspond to actual positive outcomes. Mathematically, precision is defined as (Parteek Bhatia, 2019):

$$\text{Precision} = \frac{\text{True Positive}}{\text{All Positive Predicted}} \quad (3)$$

3.6 Deployment

This stage represents the final phase of the data mining process, where the extracted knowledge and developed models are deployed for practical use by organizational stakeholders. The outcomes of the data mining process may be implemented in relatively simple forms, such as analytical reports or decision-support summaries, or in more advanced implementations, including integration into existing operational systems within the organization.

RESULTS AND DISCUSSION

4.1 Business Understanding

On-time graduation is one of the key aspects considered in higher education accreditation. In addition, graduating on time also provides personal benefits for students. Therefore, the data mining process in this study focuses on comparing the performance of three machine learning algorithms individually, as well as their combined implementation through an ensemble approach, in predicting students' on-time graduation.

 ISSN 2656-2855
e-ISSN 2685-5518

4.2 Data Understanding

The dataset was obtained from the Academic Information System database, which stores all student identity and academic records. The collected dataset consists of 2,471 records with 11 initial attributes, as shown in Table 1.

Table 1. Dataset Attributes and Descriptions

No.	Attributes	Descriptions
1	NIM	The unique identification number assigned to each student
2	A	Number of courses in which the student obtained grade A
3	AM	Number of courses in which the student obtained grade A-
4	BP	Number of courses in which the student obtained grade B+
5	B	Number of courses in which the student obtained grade B
6	BM	Number of courses in which the student obtained grade B-
7	CP	Number of courses in which the student obtained grade C+
8	C	Number of courses in which the student obtained grade C
9	TdkLulus	Number of courses that the student did not pass
10	JmlMK	Total number of courses taken during semesters 1-4
11	Kelulusan	Target label indicating whether the student graduated on time or experienced delayed graduation

4.3 Data Preparation

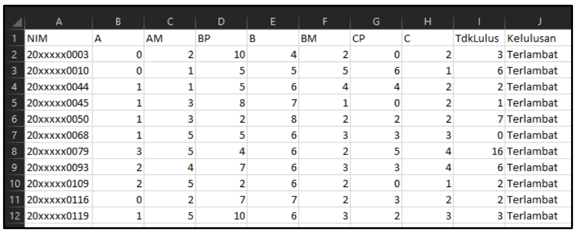
4.3.1 Cleaning

The dataset initially consisted of 11 attributes; however, not all attributes were used in the model development process. An attribute elimination process was performed on two attributes, namely Student ID (NIM) and Total Courses (JmlMK), as they were considered irrelevant to the data mining process.

4.3.2 Sampling

Out of the 2,471 records in the dataset, 305 instances were labeled “Delayed”, while the remaining 2,166 instances were labeled “On Time”, indicating the presence of class imbalance within the dataset. To address class imbalance, a common approach is to apply sampling techniques, such as under sampling, oversampling, or a combination of both (Tan, Steinback and Kumar, 2014).

In this study, the “On Time” class was subjected to under sampling in order to balance the dataset and improve the recall performance of the Decision Tree model. After applying the under sampling technique, the total number of records in the dataset was reduced to 610 instances, resulting in a balanced class distribution with a 1:1 ratio between the two classes.



	A	B	C	D	E	F	G	H	I	J
1	NIM	A	AM	BP	B	BM	CP	C	TdkLulus	Kelulusan
2	20xxxxx0003	0	2	10	4	2	0	2	3	Terlambat
3	20xxxxx0010	0	1	5	5	5	6	1	6	Terlambat
4	20xxxxx0044	1	1	5	6	4	4	2	2	Terlambat
5	20xxxxx0045	1	3	8	7	1	0	2	1	Terlambat
6	20xxxxx0050	1	3	2	8	2	2	2	7	Terlambat
7	20xxxxx0068	1	5	5	6	3	3	3	0	Terlambat
8	20xxxxx0079	3	5	4	6	2	5	4	16	Terlambat
9	20xxxxx0093	2	4	7	6	3	3	4	6	Terlambat
10	20xxxxx0109	2	5	2	6	2	0	1	2	Terlambat
11	20xxxxx0116	0	2	7	7	2	3	2	2	Terlambat
12	20xxxxx0119	1	5	10	6	3	2	3	3	Terlambat

Figure 2. Example of Dataset After the Data Preparation Stage

4.4 Modelling

The modeling process using the Decision Tree method in this study was implemented using the Python programming language and the scikit-learn (sklearn) library. Figure 3 illustrates the program workflow used to build the individual model.

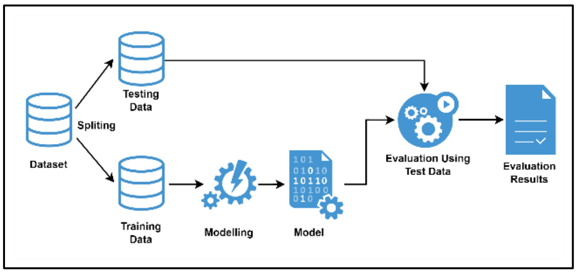


Figure 3. Individual Model Development Process

In the individual model development process, the dataset was divided using an 80:20 split ratio, where 80% of the data was used for the training process, while the remaining 20% was used for testing. The training dataset was then utilized to build predictive models for each algorithm, resulting in separate individual models.

Subsequently, the generated models were evaluated by performing predictions on the prepared testing dataset. The evaluation process produced performance metrics including accuracy, precision, and recall, which were used to assess the effectiveness of each model.

In the ensemble voting modeling process, the dataset was split using an 80:20 ratios, similar to the individual modeling approach. The training dataset was then used to build models for the three algorithms, resulting in three separate individual models.

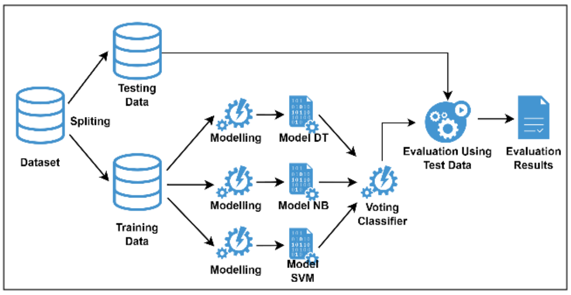


Figure 4. Ensemble Voting Modeling Process

These models were subsequently combined to form an ensemble model using a Voting Classifier. Finally, the ensemble model was evaluated using the testing dataset, producing evaluation metrics to measure the performance of the model.

4.5 Evaluation

To evaluate the performance of the developed models, several evaluation metrics were used, namely accuracy, precision, and recall. The following section presents the evaluation results for each individual model as well as the ensemble model that was developed:

Table 2. Evaluation Results of Each Model

No.	Model	Accuracy	Precision	Recall
1	Decision Tree	86.07%	86.10%	86.07%
2	Support Vector Machine	86.07%	87.06%	86.07%
3	Naïve Bayes	88.52%	89.01%	88.52%
4	Ensemble Model (Voting) - Hard	90.16%	90.40%	90.16%
5	Ensemble Model (Voting) - Soft	91.80%	91.87%	91.80%

Based on the evaluation results obtained for each model, the following section provides an explanation of the performance of each model:

4.5.1 Decision Tree

The Decision Tree model achieved an accuracy of 86.07%, with a precision of 86.10% and a recall of 86.07%. Although these evaluation metrics indicate a reasonably good performance, the Decision Tree model demonstrated the lowest performance compared to the other models evaluated in this study.

4.5.2 Support Vector Machine

The Support Vector Machine (SVM) model achieved an accuracy of 86.07%, which is the same as the Decision Tree model. The model obtained a precision of 87.06%, which is slightly higher than that of the Decision Tree, and a recall of 86.07%, which is identical to the Decision Tree model. Overall, the performance of the Support Vector Machine model is comparable to that of the Decision Tree model; however, it demonstrates a slightly better precision value

4.5.3 Naïve Bayes

The Naïve Bayes model achieved a higher accuracy compared to the two previous models, reaching 88.52%. The model also produced a precision of 89.01% and a recall of 88.52%. Based on these evaluation metrics, the Naïve Bayes model demonstrates better performance compared to the Decision Tree and Support Vector Machine models.

4.5.4 Ensemble Model - Hard Voting

Using the Hard Voting method, the ensemble model achieved an accuracy of 90.16%, with a precision of 90.40% and a recall of 90.16%. These results indicate an improvement in evaluation metrics for the ensemble model compared to the individual models. The improvement is particularly significant when compared with the Decision Tree model.

4.5.5 Ensemble Model - Soft Voting

The ensemble model using the Soft Voting method achieved an accuracy of 91.80%, with a precision of 91.87% and a recall of 91.80%. Based on these evaluation results, the Soft Voting Ensemble Model demonstrates the best performance among all evaluated models, outperforming not only the individual models but also the Hard Voting Ensemble Model.

Figure 5 presents a comparison of the evaluation metrics across all models:

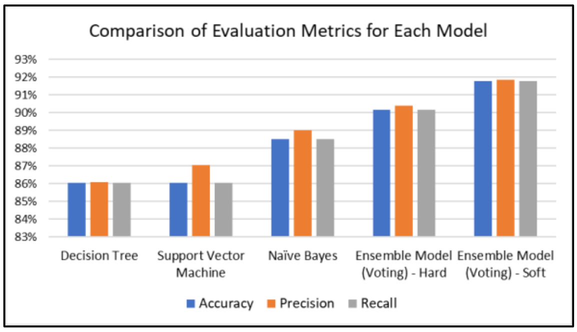


Figure 5. Comparison of Evaluation Metrics for Each Model

Based on the evaluation results described previously, it can be concluded that the ensemble model using the voting approach achieved the best performance compared to the individual algorithms. The results obtained in this study are consistent with the findings of Munandar et al. (2023), which reported that an ensemble model combining C4.5, K-Nearest Neighbor (KNN), and Logistic Regression produced significantly better performance compared to individual models (Munandar, Baihaqi and Nurhopipah, 2023).

4.6 Deployment

After completing the evaluation stage of the models described in the previous section, the deployment stage aims to ensure that the best-performing model can be utilized to provide practical benefits in predicting student graduation outcomes. Based on the evaluation metrics obtained, the Ensemble Model with Soft Voting was selected as the best-performing model according to several performance indicators, including accuracy, precision, and recall. This model achieved the highest values across these evaluation metrics compared to the other models.

Based on these results, the Soft Voting Ensemble Model can be implemented (deployed) either within a newly developed system or integrated into an existing system. Through the implementation of the selected best-performing model, it is expected to contribute to improving the accuracy, efficiency, and effectiveness of predicting student graduation outcomes.

CONCLUSION

This study aims to analyze and compare the performance of three classification algorithms, namely Decision Tree, Naïve Bayes, and Support Vector Machine (SVM), and further integrate them into an ensemble learning model using hard and soft voting techniques to predict students' on-time graduation. Based on the evaluation of the developed models, the results obtained are as follows:

a. Individual Models:
   Decision Tree achieved an accuracy of 86.07%**,** with precision and recall of 86.10% and 86.07%**,** respectively. Support Vector Machine (SVM) showed similar performance to Decision Tree, with an accuracy of 86.07%**,** precision of 87.06%**,** and recall of 86.07%**.** Naïve Bayes demonstrated better performance compared to the other two models, achieving an accuracy of 88.52%**,** precision of 89.01%**,** and recall of 88.52%**.**
b. Ensemble Model:
   The Hard Voting Ensemble model showed a noticeable improvement in predictive performance, achieving an accuracy of 90.16%, precision of 90.40%, and recall of 90.16%. These results indicate that combining multiple classification models can enhance the overall predictive capability of the system.
   The Soft Voting Ensemble produced the best performance among all evaluated models, achieving an accuracy of 91.80%, precision of 91.87%, and recall of 91.80%. These findings suggest that the ensemble approach using soft voting is more effective in improving predictive accuracy and classification performance compared to individual models as well as the hard voting ensemble approach.

Based on the evaluation results, it can be concluded that the soft voting ensemble model demonstrates the best performance in predicting students' on-time graduation. This finding is supported by a notable improvement in the evaluation metrics, including accuracy, precision, and recall, compared to the individual classification models (Decision Tree, Support Vector Machine, and Naïve Bayes). Consequently, the application of the soft voting ensemble approach provides a more reliable and effective solution for identifying students at risk of delayed graduation. Furthermore, this model offers valuable insights that can support data-driven decision-making processes within higher education institutions.

For future research, it is recommended to utilize larger and more diverse datasets obtained from multiple higher education institutions in order to enhance the generalizability of the predictive model. Furthermore, subsequent studies may explore other ensemble learning techniques, such as boosting and bagging, and compare their performance with voting-based ensemble methods to identify the most effective ensemble approach for predicting students' graduation outcomes.

REFERENCES

Amri, Z., Kusrini, K. dan Kusnawi, K. (2023) 'Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM', *Edumatic: Jurnal Pendidikan Informatika*, 7(2), pp. 187–196. Available at: https://doi.org/10.29408/edumatic.v7i2.18620.

Arifin, Z., Rahman, D.F., Rintyarna, B.S. dan Daryanto, D. (2023) 'Penerapan Algoritma Support Vector Machine Berbasis Kernel Radial Basis Function dalam Klasifikasi Sel Kanker', *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, 4(2), pp. 100–106. Available at: https://doi.org/10.37148/bios.v4i2.165.

BPS 2022 (2023) 'Statistik Pendidikan 2023', *Badan Pusat Statistik*, 1101001, p. 790.

Budiyantara (2019) 'Prediksi Mahasiswa Lulus Tepat Waktu Menggunakan Algoritma Decision Tree (C4.5) Pada STMIK Widuri Jakarta', *Infotech, Journal of Technology Information*, 5(2), pp. 1–7.

Indahyanti, U., Azizah, N.L. dan Setiawan, H. (2022) 'Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa', *Jurnal Sains dan Informatika*, 8(2). Available at: https://doi.org/10.34128/jsi.v8i2.459.

Mukti, M.W. (2022) 'Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma C4.5 Pada STMIK Dharma Wacana', *International Research on Big-Data and Computer Technology: I-Robot*, 6(2), pp. 25–29. Available at: https://doi.org/10.53514/ir.v6i2.326.

Munandar, A., Baihaqi, W.M. dan Nurhopipah, A. (2023) 'A Soft Voting Ensemble Classifier to Improve Survival Rate Predictions of Cardiovascular Heart Failure Patients', *ILKOM Jurnal Ilmiah*, 15(2), pp. 344–352.

Nazifah, N., Prianto, C. dan Id, C.A. (2023) 'Analisis Perbandingan Decision Tree Algoritma C4.5 dengan algoritma lainnya: Sistematic Literature Review'. Available at: https://ejurnalunsam.id/index.php/jicom/.

Nurislamiaty, V.A. dan Rozi, A.F. (2021) 'Prediksi Kelulusan Mahasiswa Fakultas Teknologi Informasi Umby Menggunakan Metode Decision Tree Penerapan Algoritma C4. 5', *Journal Of ...*, 5, pp. 1–8.

Parteek Bhatia (2019) *Data Mining Data Warehouseing Principales and Practical Techniques*.

Rahman, A., Zaman, S., Parvej, S., Shill, P.C., Salim, M.S. dan Das, D. (2025) 'Fake News Detection: Exploring the Efficiency of Soft and Hard Voting Ensemble', in *Procedia Computer Science*. Elsevier B.V., pp. 748–757. Available at: https://doi.org/10.1016/j.procs.2025.01.035.

Rohmawan, E.P. (2018) 'Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network', *jurnal ilmiah MATRIK*, p. 3.

Sasongko, D. (2019) 'Akreditasi Perguruan Tinggi Kriteria dan Prosedur 3.0', *Badan Akreditasi Nasional Perguruan Tinggi*, pp. 1–21.

Sulianta, F. (2024) 'Buku Dasar Data Mining from A to Z', (January).

Suratini (2017) 'Pengaruh Pendidikan Dalam Meningkatkan Kualitas Sumber Daya Manusia di Indonesia', *Future: Jurnal Manajemen dan Akuntansi*, 5(1), pp. 68–84.

Tan, P.-N., Steinback, M. dan Kumar, V. (2014) *Introduction to Data Mining*, *Pearson Education Limited*. Available at: https://doi.org/10.1016/b978-155558242-5/50003-6.

Tholib, A., Fadli Hidayat, M.N., Yono, S., Wulanningrum, R. dan Daniati, E. (2023) 'Comparison of C4.5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms', *International Journal of Engineering and Computer Science Applications (IJECSA)*, 2(2), pp. 65–72. Available at: https://doi.org/10.30812/ijecsa.v2i2.3364.

Yuni, R., Putra, P.D. dan Hutabarat, D.L. (2020) 'Sinergi Indonesia Menuju Negara Maju', *Prosiding WEBINAR Fakultas Ekonomi Universitas Negeri Medan*, pp. 35–42.