# Comparative Analysis of Random Forest Regressor, Support Vector Regression, and MLP Regressors for Student Graduation Time Prediction

**Yonathan Purbo Santosa[1],**
Department of Informatics, Faculty of Computer Science
Soegijapranata Catholic University, Semarang, Indonesia
[1]yonathansantosa@unika.ac.id

Abstract— Indonesian bachelor degree students are considered graduated in time when the student graduated in 4 year or less with lower graduated times are preferrable. As such, accurate graduation time prediction is critical for institutional resource planning and academic intervention strategies as it impacts the public perceptions of the institutions and its accreditation. Using regression method, this research will test three machine learning model, namely Random Forest Regressor, Support Vector Regression, and Multilayer Perceptron Regressor to predict the total study in years. Compared to the previous approaches, which relies on grade point average (GPA), this approach uses more granular academic study report for each semester and every courses grades. The student data are collected from Department of Informatics, Soegijapranata Catholic University which enrolled between 2017 and 2021. To evaluate the model, each model processed the semester-by-semester student data for every students as the input features and evaluated using several regression scoring method such as R-squared and MSE to picture the dataset generalization. The goal of this research is to lay foundation for further research on early warning systems, especially for department heads to give ability to identify students and to formulate intervention strategies for students with risks of delayed graduation.

Keywords— Graduation Time Prediction, Machine Learning, MLP Regressor, Random Forest, Support Vector Regression

## I. INTRODUCTION

Indonesian undergraduate program has objectives of providing quality education and maintaining the success of study for those who enrolled in the university [1], [2]. Apart from the Grade Point Average (GPA), graduate competencies, and curriculum structure, graduation times are one of the indicators of successful studies. The Indonesian Ministry of Education, Science, and Technology has determined that at least 50% of the students who enrolled in a study program must graduate in-time according to the curriculum [3], which is 4 years or 8 semesters for undergraduate degree. Early interventions by detecting possible late graduation are considered to be important because not only it helps with the department efficient management, it also helps with alleviating financial burden for the student and their families [4]. Using data mining techniques, program coordinators may improve curriculum design while academic advisors may guide students based on the data [5], providing intervention strategy for every student to graduate on time.

Previous research has provided insight into various data mining and machine learning techniques to predict the graduation time for students based on different characteristics of the dataset. One of the approaches is by using the title of the students' thesis with linear regression [6]. This approach shows that the title of thesis is statistically insignificant for the prediction of graduation time [6], meaning thesis title is not suitable for graduation time prediction. The other limitation is that this method only works for students that are in their last moments in their study.

The second approach is using only the academic results such as Grade Point


ISSN 2442-7888 (online) DOI 10.24167/sisforma.v12i2.14394

Averages (GPA) or Grade Point Semester (GPS). One of the research uses C4.5 algorithm to classify graduation time based on Grade Point Semester (GPS) achieving accuracy of 99.64% even without external factors included [7]. This shows that it is possible to use only academic results with no added external features. The next approach uses subjective data such as motivation to graduate on time, research skills, supervisory practices, institutional support, and self-management skills on postgraduate students [8]. The research suggested that research skills followed by institutional support are the main factors of academic success [8]. Compared to undergraduate students, post-graduate students have higher competency standards compared to undergraduate students and might have different access to the facilities provided by the university, so different measurements are needed for undergraduate students.

The last approach, which is the most popular one, is combining academic information and socioeconomics data such as marital status, parental income, etc., as it is believed to add important features to the data. One of the approach is by classification of on-time graduation using GPA for every semester, educational track/specialization, and parent’s income using Support Vector Machine (SVM), Decision Tree, and Logistic Regression, all with or without ensemble model (AdaBoost) [2]. The results show that on average there is an increase of 4.10% in accuracy by using ensemble model with the highest accuracy achieved by using SVM with AdaBoost. This research points out that ensemble models may perform better in predicting students’ graduation time. Rahmadani et al. [9] use external data such as parental income and student part-time jobs besides the GPA to predict the graduation time using linear regression. This approach achieves standard error of 4.07 with apart from the GPA, student part-time jobs affects the most in the delayed graduation, followed by the parental income [9]. Similar approach also implemented with logistic regression with GPA, marital status, student part-time job, and study program interest which achieved MAPE of 8% [10]. Hartati et al. [11] use Naïve Bayes with information gain to classify student graduation delays. The Information gain is used for weighing the features of data showing increases in accuracy, precision, and recall compared to with no weighing [11]. Another approach using decision tree based on particle swarm optimization (PSO) shows that there is an increase in accuracy by 1.05% compared with C4.5 algorithm [12]. PSO works similarly to information gain by selecting important features as the deciding factors for classification. In one of the researches, random forests excel in classifying student graduation time compared to Naive Bayes, KNN, SVM, and Decision Tree with accuracy of 100%, compared to 100%, 97%, 93%, and 99% respectively [13]. While Random Forest and Naive Bayes achieved 100% accuracy, Random Forest achieves higher recall and precision [13]. On high schools demographics, a research on the graduation rate classification was conducted by using the grade of every subjects combined with other achievement and administrative data for each students utilizing random forest [14]. Random forest can shows a high accuracy of 100%, but looking at the data provided, only 8 students are not graduated on-time compared to 152 students which shows highly imbalanced data which can lead to biased prediction [14]. Lastly, Sulehu et al. [15] combined GPA, subjective measures such as stress level, study motivation, and academic confidence, also socioeconomic factors for classification yet only achieve 70% accuracy using Random Forest algorithms. Looking closely at the feature importance, besides GPA study motivation comes the highest while socioeconomic factors are placed among the lowest importance [15], meaning that socio economic factors are probably safe to remove from the dataset.

Although socioeconomic factors provide information for the classifier, these features may not add significant improvements to the classifier models [9], [15]. These findings are reinforced by the success of several machine learning models that able to achieve high performance of classification [7], [8]. Based on previous research, most of the ensemble models can achieve better performance compared to the others [2], [14]. Most of the previous research is also based on classification algorithms instead of regression. In terms of academic interventions, we concluded that it is better to get the general idea on how late students will be, thus regression will be used instead of classification. In light of previous findings, this research aims to find the best performing model for prediction study durations with only using granular academic results such as grades. The contribution of this research is to enabling program coordinators to evaluate their curriculum and create strategies for interventions based on the data that are readily available through the academic information systems.

## II. METHOD

This research objective is to investigate the supervised machine learning capability to predict graduation duration. In order to validate the capabilities, the research method is described in the following passage.

### A. DATA COLLECTION AND DEFINITION

The dataset was collected through the academic information system of the Department of Informatics, Soegijapranata Catholic University (SCU). The first data consists of student academic records which were admitted to the department from 2018 until 2021 and have graduated until October 2025 with the exceptions for transfer and readmission students which excluded in the dataset since this student may have abnormal graduation time. The academic records are collected from the first semester until their last semester and only considering the first time the student takes a course instead of the last. This method is chosen since students may retake the same courses to improve their grades, which will affect their graduation time. For every student, the data consists of the course code, grade, and the semester. The academic record example is shown in Table 1. This data is going to be the independent variable for the regression model.

**Table 1. Academic Reports for Every Student (anonymized)**

| Course Code | Grade | Semester |
|---|---|---|
| IT001 | A | 2021-01 |
| IT002 | AB | 2021-01 |
| … | | |

The second data was the dependent variable which is the graduation time. The data was also collected through the academic information system. For every student, the student ID and their graduation time in years was collected. The data are shown in Table 2. In total there are 213 students collected for the dataset with at least 144 credits for every student spread out among several courses.

**Table 2. Student Graduation Time Data (anonymized)**

| Student Identification Number (ID) | Graduation Time (Years) |
|---|---|
| 21.0001 | 4.6 |
| 21.0002 | 5.6 |
| … | |

### B. DATA PREPROCESSING

Since academic reports are temporal data for every student, we need to transform the data into a vector consisting of all the temporal data present in the academic reports. To transform the data into vector, first we convert the grade from letters into numbers using the lookup table described in Table 3.

**Table 3. Grade to Grade Number Conversion**

| Grade | Grade Numbers |
|---|---|
| A | 4 |
| AB | 3.5 |
| B | 3 |
| BC | 2.5 |
| C | 2 |
| CD | 1.5 |
| D | 1 |
| E | 0 |

After the conversion, the data was flattened into a vector and for every element of the vector is the grade for every course that exists in the curriculum. Any courses that were not taken by the students (some courses are electives) will be replaced with 0. Then the semester when the courses are taken were calculated based on the admission years, for example if a student who was admitted in 2020 who's taking a course with code IT001 in 2021 in odd semester, the said course was taken in their 3rd semester. The semester number then converted into positional encoding taken from Transformer model in order to remove ordering bias (earlier courses will have 1 while the later courses may have larger number) which in regression term, this number will be multiplied by some multipliers [16]. Using positional encoding ensures that every course will still have orders while keeping the values in range of -1 to 1. The formula to calculate the positional encoding is described in equation (1). Once all the courses' positional encoding has been calculated, the positional encoding merged with the vector data as shown in Table 4.

$$\boldsymbol{PE}_{(\boldsymbol{pos})} = \begin{cases} \mathbf{sin}\left(\frac{\boldsymbol{pos}}{\mathbf{100}^{\frac{k}{2}}}\right), \text{ if } \boldsymbol{k} \text{ is even} \\ \mathbf{cos}\left(\frac{\boldsymbol{pos}}{\mathbf{100}^{\frac{k-1}{2}}}\right), \text{ if } \boldsymbol{k} \text{ is odd} \end{cases} \quad \ldots\ldots\ldots(1)$$

**Table 4. Flattened Vector of Course Grade and Positional Encoding**

| ID | IT 001 | IT 001 $PE_1$ | IT 001 $PE_2$ | IT 002 | IT 002 $PE_1$ | IT 002 $PE_2$ | … |
|---|---|---|---|---|---|---|---|
| 01 | 4 | 0.0 | 1.0 | 3.5 | 0.19 | 0.98 | … |
| 02 | 4 | 0.19 | 0.98 | 2.5 | 0.37 | 0.93 | … |
| | | | | … | | | |

Lastly, the GPS information for every semester was added to the data for every student. The GPS on semester which the student has already graduated is filled with 0. The complete feature vector comprises of 194-dimensional data with 16-dimensions data for the GPS, 59-dimensions for the course grades, and 118-dimensions for the positional encoding for every course. The target data consists 1-dimension for the graduation time.

## C. MODEL SELECTION AND ARCHITECTURE

Three machine learning algorithms were selected for their proven effectiveness in handling complex, non-linear regression tasks. The first model is ensemble model, such as Random Forest Regressor (RFR) which performs better compared to several other algorithms [2], [14]. This model employs several decision trees and provides predictions based on the average of each prediction. The second model is Support Vector Regression (SVR) which has regularization techniques to reduce overfitting which is important for student graduation time prediction, especially with small number of data [17]. The third model uses Multi-Layer Regression (MLR) which has shown promising results in predicting the on-time graduation with no signs of overfitting [18].

## D. MODEL TRAINING AND EVALUATION

Each model was trained using random search hyperparameter tuning with maximum samples of 100 combinations of pre-defined hyperparameter with MSE as the evaluator. In order to validate the hyperparameter combination, k-fold cross validation with k = 3 is used. The random search hyperparameter tuning is implemented using Scikit-learn "RandomSearchCV". As such we provide the hyperparameter choice to be used for RFR, SVR, and MLR are described in Table 5, Table 6, and Table 7 respectively.

**Table 5. Hyperparameter Choice for Random Forest Regressor**

| Hyperparameter | Values |
|---|---|
| n_estimators | np.arange(100,1001,100) |
| max_features | log2, sqrt |
| max_depth | np.arange(10,105,10) |
| min_sample_split | [2, 5, 10] |
| min_sample_leaf | [1, 2, 4] |
| bootstrap | True, False |

**Table 6. Hyperparameter Choice for Support Vector Regressor**

| Hyperparameter | Values |
|---|---|
| kernel | linear, poly, rbf, sigmoid |

| | |
|---|---|
| degree | np.arange(1, 6) |
| gamma | scale, auto, np.logspace(-4, 1, 10) |
| coef0 | np.linspace(-1, 1, 10) |
| C | np.logspace(-2, 3, 10) |
| epsilon | np.logspace(-3, 0, 10) |

**Table 7. Hyperparameter Choice for Multilayer Perceptron Regressor**

| **Hyperparameter** | **Values** |
|---|---|
| hidden_layer_sizes | (50, 50); (100, 50, 25); (100, 75, 50, 25) |
| hidden activation | logistic sigmoid, tanh, ReLU |
| solver | lbfgs, sgd, adam |
| alpha | np.logspace(-4, 4, 10) |
| learning_rate | constant, invscaling, adaptive |
| max_iter | 200, 500, 1000 |

The parameter search was conducted with 80% of the dataset as the training data and evaluated on the remaining test data. The evaluation uses Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Squared (R2). This metrics was chosen based on the test conducted by Plevris et al. [19] which shows no indication of poor evaluation on several tests.

## III. RESULTS AND DISCUSSION

This chapter presents the findings following the method described previously. We begin by detailing the descriptive statistics of the training results, followed by a comparative analysis of the performance on test data of the three implemented regression models.

### A. HYPERPARAMETER SEARCH

Firstly, the top three performances based on the hyperparameter search for RFR are shown in Table 8. The mean_test_score contains the MSE sorted from the lowest to highest (left to right). Some of the hyperparameter shows a clear pattern in producing lower training errors in the hyperparameter search. Min_samples_split and min_sample_leaf correlate with the lower results of MSE. Based on the results, using sqrt for max_features and setting bootstrap to be False gives the lower MSE compared with the other. On the other hand, n_estimators and max_depth show no clear pattern, meaning that this hyperparameter dependent on the other hyperparameters. As a result, the lowest training MSE was using the left-most hyperparameter presented in Table 8.

Secondly, hyperparameter search for SVR shows that on average rbf kernel has the lowest MSE of 0.66 followed by linear kernel with MSE of 4.64. The other two kernels are polynomials and sigmoid which on average have MSE of 752.49 and 1041.42 respectively. Polynomial degree parameters have low MSE (less than 0.9) except for degrees with 2 or 5 which have high MSE (3313.6 and 84.95 respectively). Similarly with parameter gamma, only gamma with 0.000359, 0.001292, auto, and scale have average MSE larger than 1 with the largest MSE was 7795.77 for auto. On the other hand, parameter coef0 has the lowest average MSE of 0.62 with the value of -1 with large MSE is observed when using coef0 between -0.7778 and -0.3333 with the largest MSE being 3120.78. On the other hand, epsilon parameters have lower errors with values between 0.01 and 0.02. Lastly, parameter C does not show any pattern of increasing and decreasing MSE. Based on the previous findings, the analysis of hyperparameter for SVR is not conclusive except for the kernel rbf produces the lowest MSE. The top four performances of hyperparameter combinations for SVR are shown in Table 9.

**Table 8. Best Hyperparameter Combination for RFR (rounded until 2 decimals)**

| | **Rank** | | |
|---|---|---|---|
| **Hyperparameters** | **1st** | **2nd** | **3rd** |
| n_estimators | 600 | 400 | 900 |
| min_samples_split | 2 | 2 | 5 |
| min_samples_leaf | 1 | 1 | 1 |
| max_features | sqrt | sqrt | sqrt |
| max_depth | 100 | 80 | 70 |
| bootstrap | FALSE | FALSE | FALSE |
| **mean_test_score** | **0.2440** | **0.2443** | **0.2472** |

**Table 9. Hyperparameter Combination for SVR (rounded until 2 decimals)**

| | **Rank** | | | |
|---|---|---|---|---|
| **Hyper parameters** | **1st** | **2nd** | **3rd** | **4th** |
| kernel | rbf | sigmoid | rbf | linear |
| degree | - | - | - | - |
| gamma | 0.02 | 0.0001 | auto | - |

| | | | | |
|---|---|---|---|---|
| coef0 | - | -0.11 | - | - |
| C | 77.43 | 278.26 | 278.26 | 0.13 |
| epsilon | 0.05 | 0.02 | 0.46 | 0.001 |
| **mean_ test_score** | **0.408** | **0.409** | **0.413** | **0.413** |

Lastly, the results of hyperparameter search results are as follows, first the hyperparameter search MLR show that using three hidden layers performed worse indicated in the increasing of the average MSE compared with the other choices. Second, ReLU activation function resulting in lowest MSE followed by TanH, and lastly logistic. Third, SGD solver algorithm performs worse compared to adam and lbfgs based on the MSE average. Fourth, the lowest average of MSE for the learning rate strategy achieved by adaptive, followed by constant and invscaling. Lastly, lower alpha parameter consistently inducing lower MSE, while 6000 maximum iterations causing the lowest MSE. The combination that resulting in lower MSE described in Table 10 with the first two combinations producing similar results differing only in the maximum iterations, we order the rank based on the number of iterations.

**Table 10. Best Hyperparameter Combination for MLR (rounded until 2 decimals)**

| | **Rank** | | | |
|---|---|---|---|---|
| **Hyper parameters** | **1st** | ***1st** | **3rd** | **4th** |
| hidden_ layer_sizes | (50, 50) | (50, 50) | (100, 50, 25) | (100, 75, 50, 25) |
| activation | tanh | tanh | tanh | tanh |
| solver | lbfgs | lbfgs | lbfgs | lbfgs |
| alpha | 2.783 | 2.783 | 0.359 | 0.046 |
| learning_ rate | const | const | const | invscl |
| max_iter | 4000 | 8000 | 6000 | 10000 |
| **mean_test_ score** | **0.407** | **0.407** | **0.424** | **0.432** |

Based on the hyperparameter search above, we concluded that the results are highly dependent on the combination among the hyperparameter itself for all three algorithms. We use the best performing algorithms instead to test the performance of the regressor models. The training was revalidated again with the best parameter for each model without cross-validation as shown in Table 11. Once the training has concluded, the test data predictions are measured as well as shown in Table 12.

**Table 11. Training Results for All Model**

| **Model** | **MAE** | **MSE** | **RMSE** | **R2** |
|---|---|---|---|---|
| RFR | 0.053 | 0.007 | 0.081 | 0.991 |
| SVR | 0.042 | 0.002 | 0.043 | 0.997 |
| MLR | 0.073 | 0.010 | 0.101 | 0.986 |

**Table 12. Test Results for All Model**

| **Model** | **MAE** | **MSE** | **RMSE** | **R2** |
|---|---|---|---|---|
| RFR | 0.207 | 0.191 | 0.437 | 0.789 |
| SVR | 0.360 | 0.345 | 0.588 | 0.618 |
| MLR | 0.232 | 0.225 | 0.475 | 0.751 |

## B. DISCUSSION

The results presented in the previous subsection confirm that it is feasible to predict a student's final graduation time using only their academic performance data based on courses grades and GPS. We will interpret these findings and explore the implication of these approaches.

Random Forest Regressor (RFR) comes out with the lowest deviation from the test data ($R^2$ = 0.789, MAE = 0.207, MSE = 0.191, RMSE = 0.437) compared to Multilayer Perceptron Regressor (MLR) and Support Vector Regression (SVR) as shown in Table 12. Although the training result shows that SVR has the lowest training error with MAE = 0.042, MSE = 0,002, RMSE = 0.043, $R^2$ = 0.997. Looking at the training results, all models are able to explain the connection between dependent and independent variables with $R^2$ score above 0.98 for all algorithms. In spite of high score for the training data, we observe a decrease of $R^2$ score in the test data 0.618, 0.751, and 0.789 for SVR, MLR, and RFR respectively. SVR excels in the training data yet performs the worst in the test data indicating that SVR is the least suitable in predicting new data. RFR with $R^2$ score of 0.789, which according to Table 13 means a very strong correlations, is the highest for the test data replacing the SVR as the highest performing regressor for training data. Compared with the previous approach which rely on aggregate GPA, Parental Income, and

Student Part-time Job the $R^2$ score only achieved 0.6153 [9], indicating strong correlation. This concluded that using granular approach in making the datasets improves the model performance.

**Table 13. Correlation Between Variable [9]**

| $R^2$ Score | Description |
|---|---|
| 0 | No Correlation |
| >0-0.25 | Very Weak |
| >0.25-0.5 | Quite Strong |
| >0.5-0.75 | Strong |
| >0.75-0.99 | Very Strong |
| 1 | Perfect Correlation |

Looking at the results of RFR closely, it has MSE of 0.191 which corresponds to RMSE as the standard deviation of the prediction which is 0.437 years (5.3 months), almost 1 semester. This finding provides an insight into the fact that improvements could be made

## IV. CONCLUSION

This research sets out to develop and evaluate the best performing regression model for predicting student graduation time (in years), with the primary goal of laying a data-driven foundation for an effective Early Intervention System. The study was designed to test the hypothesis that granular, course-level academic data from a student's first two years of study is a more potent predictor than traditional aggregate metrics like GPA. Based on the results and discussion, the following conclusions are drawn:

1. The hypothesis is strongly validated. Comparative analysis demonstrated that machine learning models can effectively predict final study duration using granular courses grades and Grade Point Semester.
2. The Random Forest Regressor (RFR) was the superior model for this task, achieving the highest performance on the test set with an $R^2$ of 0.789 and the lowest Mean Squared Error (MSE) of 0.191.
3. This level of accuracy, corresponding to a Root Mean Squared Error (RMSE) of approximately 0.437 years (or 5.3 months), is considered practically viable for an institutional early intervention system. It provides a sufficiently precise quantitative tool for heads of study programs to move from reactive to proactive intervention.

While the model shows strong predictive power, it faces limitations in handling extreme outliers, which are often influenced by non-academic factors. Future work should focus on integrating these non-academic variables (such as financial status or campus engagement) and exploring sequential deep learning models (like LSTMs) to capture the temporal dynamics of student performance more explicitly.

## ACKNOWLEDGMENT

The author is grateful to Soegijapranata Catholic University for providing the institutional support and resources that made this research possible. Thanks are extended to the Department of Informatics for their invaluable assistance in facilitating the data collection for this study.

## REFERENCES

[1] M. A. A. Arib, “Survival Analysis of Students Not Graduated on Time Using Cox Proportional Hazard Regression Method and Random Survival Forest Method,” *jsds*, pp. 13–21, May 2023, doi: 10.33369/jsds.v2i1.24312.

[2] A. Desfiandi and B. Soewito, “Student Graduation Time Prediction Using Logistic Regression, Decision Tree, Support Vector Machine, and Adaboost Ensemble Learning,” *IJISCS*, vol. 7, no. 3, p. 195, Oct. 2023, doi: 10.56327/ijiscs.v7i2.1579.

[3] R. Bakri, N. P. Astuti, and A. S. Ahmar, “Machine Learning Algorithms with Parameter Tuning to Predict Students’ Graduation-on-time: A Case Study in Higher Education,” *J. Appl. Sci. Eng. Technol. Educ.*, vol. 4, no. 2, pp. 259–265, Dec. 2022, doi: 10.35877/454RI.asci1581.

[4] H. Andrianof, A. P. Gusman, and O. A. Putra, “Implementasi Algoritma Random Forest untuk Prediksi Kelulusan

Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia," *Jurnal Sains Informatika Terapan*, vol. 4, no. 1, pp. 24–28, Feb. 2025, doi: 10.62357/jsit.v4i2.464.

[5] M. Nachouki and M. Abou Naaj, "Predicting Student Performance to Improve Academic Advising Using the Random Forest Algorithm," *International Journal of Distance Education Technologies (IJDET)*, vol. 20, no. 1, pp. 1–17, 2022, doi: 10.4018/IJDET.296702.

[6] A. Aminuddin, R. Hidayat, G. Sastria, and A. Astried, "Regression Analysis to Predict the Length of Time to Complete a Thesis based on the Title," *SISTEMASI*, vol. 14, no. 2, p. 613, Mar. 2025, doi: 10.32520/stmsi.v14i2.4994.

[7] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journal of Computer and Information Systems Ampera*, vol. 4, no. 2, pp. 55–65, May 2023, doi: 10.51519/journalcisa.v4i2.393.

[8] P. Muthukrishnan, G. K. Sidhu, T. S. Hoon, G. Narayanan, and C. Y. Fook, "Key Factors Influencing Graduation on Time among Postgraduate Students: A PLS-SEM Approach," *Asian Journal of University Education*, vol. 18, no. 1, pp. 51–64, Jan. 2022, Accessed: Nov. 04, 2025. [Online]. Available: https://eric.ed.gov/?id=EJ1336207

[9] N. Rahmadani, E. Kurniawan, N. Nurhasanah, and W. Damanik, "Prediction of on-Time Graduation of Universitas Royal Students Using Multiple Linear Regression Method," *JURTEKSI (jurnal Teknologi dan Sistem Informasi)*, vol. 11, no. 3, pp. 525–532, July 2025, doi: 10.33330/jurteksi.v11i3.3893.

[10] N. Nurmalitasari and E. Purwanto, "Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik," *J.Derivat*, vol. 9, no. 2, pp. 145–152, Dec. 2022, doi: 10.31316/jderivat.v9i2.2639.

[11] S. Hartati, N. A. Ramdhan, and H. A. San, "Prediksi Kelulusan Mahasiswa Dengan Naïve Bayes dan Feature Selection Information Gain," *j.ilm.intech*, vol. 4, no. 02, pp. 223–234, Dec. 2022, doi: 10.46772/intech.v4i02.889.

[12] S. Sudriyanto, R. Rizaldi, and M. A. R. Hariri, "Implementasi Algoritme Decision Tree (C4.5) dengan Optimize Weights (PSO) untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu," *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 2, pp. 252–257, June 2021, doi: 10.32493/informatika.v6i2.9197.

[13] G. Puspa, M. F. R. Amri, and M. P. Nugraha, "Pemanfaatan Random Forest untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa Studi Kasus: Institut Desain dan Bisnis Bali," *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 2, pp. 413–426, June 2025, doi: 10.26798/jiko.v9i2.1886.

[14] A. Fatunnisa and H. Marcos, "Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest," *JAMIKA*, vol. 14, no. 1, pp. 101–111, Apr. 2024, doi: 10.34010/jamika.v14i1.12114.

[15] M. Sulehu, W. Wisda, F. Wanita, and M. Markani, "Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi," *jmp*, vol. 13, no. 2, pp. 2364–2374, Jan. 2025, doi: 10.33395/jmp.v13i2.14472.

[16] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 02, 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.

[17] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, "Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review," *IEEE Access*, vol. 12, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.

[18] N. V. Darenoh, F. A. Bachtiar, and R. S. Perdana, "Prediction of On-time Student

Graduation with Deep Learning Method | Journal of ICT Research and Applications," *J. ICT Res. Appl.*, vol. 18, no. 1, pp. 1–20, June 2024, doi: 10.5614/itbj.ict.res.appl.2023.18.1.1.

[19] V. P. Plevris, G. S. Solorzano, N. B. Bakas, and M. E. A. B. S. Seghier, "Investigation of performance metrics in regression analysis and machine learning-based prediction models," *ECCOMAS Congress 2022 - 8th European Congress on Computational Methods in Applied Sciences and Engineering*, vol. Computational Solid Mechanics, Nov. 2022, doi: 10.23967/eccomas.2022.155.