JEPIN
(Jurnal Edukasi dan Penelitian Informatika)
ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. 9
No. 3
Desember
2023

Predicting Student Academic Performance Based on Psychological Test using Machine Learning

Mario E. S. Simaremare[#1], San A. Limbong[#2], Estomihi R. Sirait[#3], Cristina S. Hasibuan[#4]

[#]Department of Information System, Institut Teknologi Del, Indonesia
Jl. Sisingamangaraja, Tobasa, Sumatera Utara

[1]mario@del.ac.id
[2]sanantoniolimbong@gmail.com
[3]estomihisirait@gmail.com
[4]cristina.hasibuan28@gmail.com

Abstract— **It is essential to consider the psychological aspect of selecting new students to determine the success of prospective students. The psychological aspect is measured by a psychological test that shows the level of prospective students' abilities in social, emotional, personality, and potential to live at university. This paper proposes an approach to predicting student performance based on their psychological test scores using the Decision Tree and Random Forest algorithms. The dataset used in this study was taken from the student academic record at Institut Teknologi Del, which includes years of psychological test scores and the Grade Point Average (GPA) from studying at the Institute. More specifically, the dataset used includes the 2019, 2020, and 2021 class years. However, there are gaps in the dataset used, including missing values and psychological test attributes such as TIU, TIU Category, Work Achievement, Work Tempo, Accuracy, and Consistency, which are unavailable in other datasets. This is shown in the correlation heatmap, which shows the level of correlation for each attribute, which is still classified as a very weak correlation. Therefore, we came up with two approaches. The first approach is to use as many records as possible (Analysis on records), and the opposite of the second is to take advantage of more features (Analysis based on features). The two approaches are compared to determine which performs better for the classification model. Our results show that studies that emphasize the use of records produce slightly better performance than analyses that emphasize features. In more detail, the random forest algorithm produces the best performance compared to the decision tree algorithm in each Analysis, the RMSE value is 0.4552, and the MAE value is 0.3514. Moreover, none of the psychological test attributes strongly correlate to GPA and hence do not guarantee student performance.**

Keywords— **Decision Tree, Random Forest, Machine Learning, Psychological Test, RMSE, MAE**

I. Introduction

Student learning success is inseparable from the influence of various factors, such as the learning environment or other factors (both internal and external). This is supported by [1], which concluded that one factor that significantly influences student success is motivation and whether the student has learning talent. Another illustration is that student motivation and satisfaction positively correlate with student learning outcomes.

A psychological test is a test used to measure individual differences and individual reactions on different occasions. Psychological tests are used to get candidates according to the abilities expected to achieve organizational needs [2]. The application of psychological tests is essential to determine the suitability or eligibility of the individual for the organization or institution.

The Institut Teknologi Del (IT Del) student admission process includes academic tests, after which a psychological test measures prospective students' ability level in the social, emotional, personality, and potential fields. Psychological tests are provided at each entrance after the academic tests are carried out. The psychological test measurement is intended to see whether the candidate can adapt to the campus lifestyle.

Based on this, admission is decided by examining the psychological and academic aspects. The psychological test conducted at IT Del has several measurement categories such as General Intelligence Test (TIU), Emotional Stability, Work Achievement, Work Tempo, Accuracy, Consistency, Endurance, and Intellectual Quotient (IQ) or Work Attitude and Intelligence. Furthermore, each aspect of the psychological test is measured according to the applicable rating scale. Psychological test aspects are calculated using a letter scale (grade) with two formats. The first format includes Very Poor (KS), Poor (K), Somewhat Poor (AK), Fair/Average, Somewhat Good (AB), Good (B), and Excellent (BS). In contrast, the second format includes Poor (K), Moderately Poor (S-), Moderate (S), Sufficiently Poor (C-), Sufficient/Adequate (C), Sufficiently Good (C+), Moderately Good (B) and Good (B).

The head of the Study Program decides the candidate's eligibility when the candidate applies. In the Information Systems study program, the requirements for prospective students who are considered eligible are measured through the General Intelligence Test (TIU) with a range of scores greater than or equal to 10 (TIU $\geqslant$ 10), IQ greater than or equal to 105 (IQ $\geqslant$ 105). For the results of each aspect with a Somewhat Poor (AK) value, there can be no more than 3, and there are no aspects with a Poor (K). This assessment has the potential to result in human error, as well as subjective decisions. In other words, the assessment can override the application of the prerequisite scale that has been determined by considering other aspects.

Based on the problems above, we need a machine learning model that can provide predictions for prospective students based on their academic achievements when they enter IT Del. The model work by comparing aspects of the psychological test assessment and comparing the Grade Point Average (GPA) of the previous students while participating in active lectures at IT Del by applies two analytic approaches, which are analysis based on records and analysis based on features, to address the variances in the database structures of student psychological test results from 2019, 2020, and 2021. These discrepancies include differences in attribute forms and the presence of incomplete data. Implementing both analytical methods aims to achieve a more precise and reliable prediction outcomes. The GPA has been studied through previous research [2], which predicted student GPA based on first-semester results and used computer science course data, followed by grades from six courses, one laboratory result, and GPA in the graduation year. This research focus on developing a machine learning model using a decision tree and random forest algorithm. The machine learning algorithms help overcome the limitations of decision trees by optimizing the tree structure, reducing overfitting, and capturing complex patterns in the data. Thus, machine learning models using Decision Tree and random forest help predict the right prospective new Del Technology Institute students according to their academic achievements.

II. LITERATURE REVIEW

A. Psychological Test

Psychological tests have various data collection techniques, such as tests, interviews, case studies, behavioural observations, and other procedures [2]. Based on research [3], the implementation of psychological data collection tests that are commonly used, such as paper and pencil tests, objective and essay tests, standard and non-standard tests, individual and group tests, verbal or nonverbal tests, personality tests, interest tests, aptitude tests, achievement test, intelligence test, and vocational test. Reliability is dependability, durability [4], stability [5], consistency, predictability, and accuracy. If it meets the reliability criteria, then the assessment results from the test can be interpreted as reliable.

B. Decision Tree

The supervised Learning algorithm can be used for regression or classification. In other words, the decision tree can be used for numerical and categorical data. The decision tree algorithm works like a tree, where class labels are leaves and features (or conditions) are branches [6]. Decision trees are used to deal effectively with large non-linear data sets. The decision tree divides the dataset into smaller subsets, forming decisions in stages [7]. Several algorithms used to build a Decision Tree, such as CART, C4.5, SPRINT, ID3 or algorithm SLIQ [8]. Decision trees are used to deal effectively with large non-linear data sets. Besides that, decision tree algorithms are easy to understand, interpret and visualize [9].

C. Random Forest

Random Forest algorithms can make predictions or classifications by combining decisions from the primary model sequence [7]. Simply put, the Random Forest algorithm is an algorithm that consists of many decision trees [9]. Random Forest is capable of working on smaller data sets as well as with large data sets or data sets containing many predictors. The accuracy level of Random Forest is obtained by allowing the development of trees to grow very deep and combining each tree to avoid overfitting [8]. The previous study [10] showed that Random Forest is the suitable algorithm for predicting management whether the student should be retained or not and provides short-term accuracy to predict which students are most suited to be maintained.

D. The Relationship between Psychologycal Tests and Machine Learning

There are similarities between psychological tests and Machine Learning, as stated in previous research [2], which says that psychological testing is a measurement method that provides an objective and standard way to predict future behaviour based on a sample of current behaviour. Both psychological tests and machine learning have in common predicting future data based on a given sample of data. Psychometry is the development of measurement instruments and assessments of individual psychology through reliable and valid measurements [11]. In other words, the psychological test uses psychometric principles to ensure that the tests used are accurate and reliable. So, further research is needed to examine the use of psychological tests in machine learning models. Machine Learning models for cognitive theory are rarely used in analyzing psychological experiments and developing psychometric tests [12].

E. Evaluation Metrics

Based on research [13], the evaluation metrics used to measure forecasting errors and assess predictive models in the regression model are MAE, MSE, RMSE, and MAPE. In what follows, we elaborate more on research-based error evaluation metrics used in this research.

1) RMSE (Root Mean Squared Error): RMSE measures the error gap analysis between actual and estimated values by taking the roots of MSE [14]

$$RMSE = \sum_{i=1}^{n} (Xi - X'i)^2 / n \quad (1)$$

2) MSE (Mean Squared Error): MSE is the difference - the average square of the difference between the predicted and actual values [14]. The greater the MSE value, the worse the model performance will be, and vice versa.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Xi^2 - X'i^2) \quad (2)$$

3) *MAE (Mean Absolute Error):* MAE measures how well the regression model predicts the actual target value. MAE is calculated by taking the average absolute error (model prediction minus the actual value). The greater the MAE value, the worse the model's performance in predicting the target value, and vice versa. MAE gives less weight to outliers [13].

$$MAE = 1/n\sum_{i=1}^{n}|Xi - X'i| \quad (3)$$

4) MAPE (Mean Absolute Percentage Error): MAPE measures the average absolute percentage error between predicted and actual values. The lower the MAPE value, the smaller the prediction error in the model.

$$MAPE = 1/n\sum_{i=1}^{n}|Xi - X'i/Xi| * 100 \quad (4)$$

Further description of the evaluation in research display the evaluation results. However, the evaluation using MAE is more emphasized. This is as described in research [13], suggesting that the MAE algorithms are more appropriate for determining the accuracy of predictions. In line with that study, research [10] indicates that the MAE evaluation metrics are better used to compare performance between different regression models.

F. Correlations Coefficient Analysis

Correlation coefficient analysis is a statistical technique used to measure the strength and direction of the relationship between two variables. It provides a numerical value that indicates the extent to which the variables are linearly related. The following is an interpretation of the range of correlation coefficient analysis.

TABLE I
CORRELATIONS COEFFICIENT ANALYSIS [15]

Positive	Negative	Interpretation
+1.00	-1.00	Perfect
+0.80 to +0.99	-0.80 to -0.99	Very Strong
+0.60 to +0.79	-0.60 to -0.79	Strong
+0.40 to +0.59	-0.40 to -0.59	Moderate
+0.20 to +0.39	-0.20 to -0.39	Weak
+0.01 to +0.19	-0.01 to -0.19	Very Weak

Based on TABLE I, Positive correlation refers to the relationship between two variables where a difference follows a change in the value of one variable in the same direction as the other variable. Conversely, a negative correlation occurs when a change follows a change in the value of one variable.

III. METHODOLOGY

A. Dataset

The dataset was taken from the student academic records at IT Del from intake 2019, 2020, and 2021. The dataset consists of psychological test results and the GPA of active students majoring in Informatics and Information Systems from 2019/2020 to 2022/2023 academic years. The following is the description of the dataset and the features/variables in it.

TABLE II
COMPARISON OF ATTRIBUTE DATASETS

Attribute	2019	2020	2021
Seq	Available	Available	Available
GPA1	Available	Available	Available
GPA2	Available	Available	Available
GPA3	Available	Available	Available
GPA4	Available	Available	None
GPA5	Available	None	Available
GPA6	Available	None	Available
GPA	None	None	Available
Batches	None	None	Available
TIU	None	Available	Available
TIU Category	None	Available	Available
Emotional Stability	Available	Available	Available
Work Achievement	None	Available	Available
Work Tempo	None	Available	Available
accuracy	None	Available	Available
Consistency	None	Available	Available
endurance	None	Available	Available
IQ	Available	Available	Available
IQ Category	None	Available	None
Intelligence	Available	Limited	None
work attitude	Available	Limited	None
IQ. 1	None	Limited	None
Emotional Stability.1	None	Limited	None

From TABLE II, it can be analyzed that there are three categories related to the availability of attribute columns in the dataset: Available, Limited, and None. This research not use the Limited category because it only contains 14 out of 50 records in the 2020 batch dataset. It has a different format from other records, which does not allow a combination of column attributes.

Based on the analysis results, it divided into two analyses, the first analysis is *based on records* focuses on each attribute available in each dataset, namely the 'Seq,' 'GPA1', 'GPA2', 'GPA3', 'Emotional Stability,' and 'IQ.' This approach is aimed to use as many records as possible by using agreed attributes. In comparison, the second analysis is based on attributes where using as many attributes as possible is the primary objectives. Later we may decide which approach produces the better results.

For comparison, the second analysis consists of all available attributes in each dataset for 2020 and 2021. In this case, the available attributes are TIU, TIU Category, Emotional Stability, Work Achievement, Work Tempo, Accuracy, Consistency, Endurance, GPA1, GPA2, and GPA3. This is done to enrich the features' availability in this study.

The purpose of each attribute used is that TIU is designed to assess an individual's numerical, verbal, and figural reasoning abilities. The Emotional Stability component evaluates the stability of a person's emotions, pinpointing specific areas where the candidate may exhibit instability. The Work Achievement metric measures an individual's skill level, accomplishments, or knowledge in a designated area. The Accuracy parameter measures the proficiency of an individual in managing detail-oriented tasks. Consistency assesses the capability to devise solutions irrespective of the situational context.

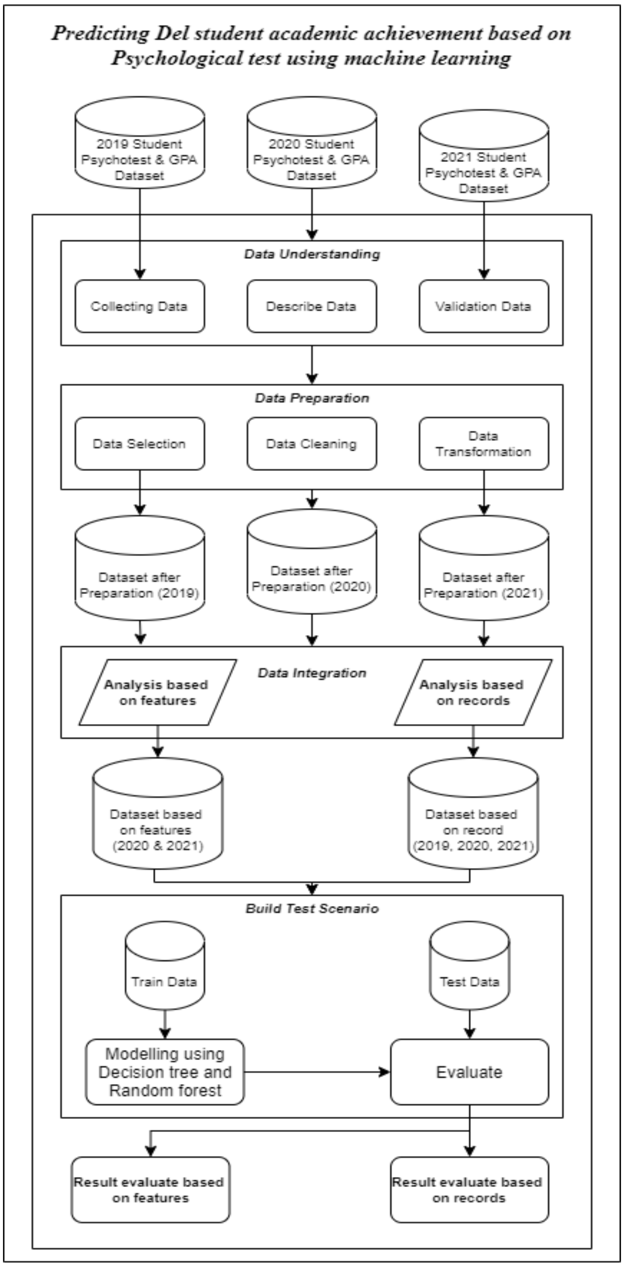

Figure. 1 Architectural design models

B. Architectural Models

Based on the architectural design models in Figure. 1, the research began with the data understanding to understand the state or context of the data. Later, the validated data was fed into the data preparation stage. Data preparation consists of several steps: data selection, data cleaning, and data transformation. In the next stage, data integration was carried out based on the analysis, namely, analysis based on record and analysis based on features that produced their own datasets. Furthermore, the two datasets were carried into the build test scenario stage, the modelling stage. In this stage, we split each dataset with 80:20 ratio. As much as 80% of the dataset were used for training data (train data) and the remaining 20% for test data (test data).

C. Data Correlation

The following section aims to analyze the relationship of each attribute used based on the analysis based on records and analysis based on features. Heatmap was used to help us analyzed the correlation (Figure. 2).

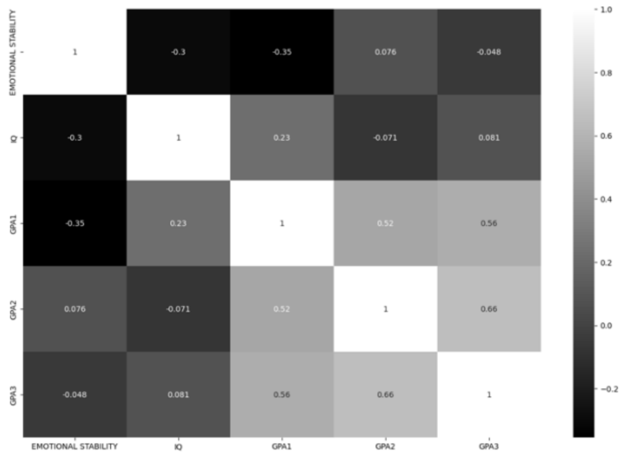

Figure. 2 Heatmap correlation based on attributes

Based on the heatmap, analysis based on records showed a relatively very weak correlations (both positive and negative) between the input features and the target features. The heatmap goes darker when the correlation is low, otherwise it goes lighter. On the heatmap, the emotional stability, IQ, GPA1, GPA2, and GPA3 attributes are displayed in lighter, which shows that there is a strong correlation between each of these attributes. GPA2 is used as a target feature to support the objectives of this study by using GPA as a reference, followed by psychological test scores. GPA1 is not used as a feature selection because IT Del has a special event to the students carry out the adaptation process well through lectures. Whereas GPA3 was chosen for data, this means that data for this study has yet to be available, compared to GPA2, which can be used as a research comparison for the 2022 batch dataset if available in the future.

In addition, if you use the GPA1 and GPA3 attributes as input features, this conflict with research objectives which measure the performance of prospective students based on a combination of psychological tests by comparing GPA results taken by students during their active lecture period or in other words when using GPA1, GPA2, and GPA3 for student performance, there is no need to make predictions involving psychological test attributes. Based on these results, there are two input features, namely 'EMOTIONAL STABILITY' and 'IQ,' and a target feature, GPA2. As a note, the attributes described in these results used in future datasets based on records.

All available attributes displayed after combining the 2019, 2020, and 2021 datasets (datasets based on records).

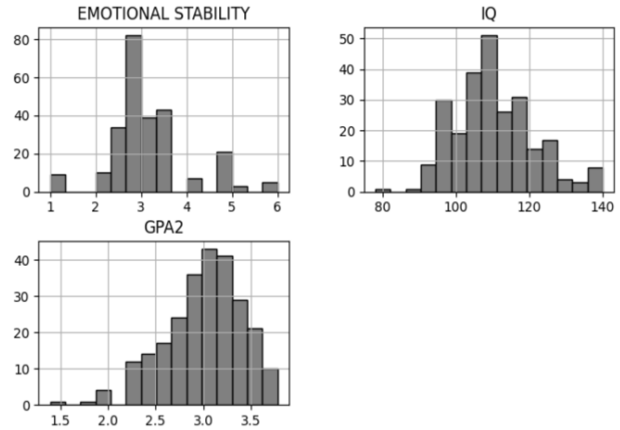

Figure. 3 Available attribute of datasets based on records

As shown in Figure. 3, the applied feature selection have stages similar to those done in the analysis based on records of deleted columns on the GPA1 and GPA3 attributes.

The heatmap analysis based on feature correlation showed that each psychological test input feature from 'TIU' to 'IQ' showed a very weak correlation with the target feature, namely 'GPA2', which is displayed with a heatmap colour that tends to be darker.

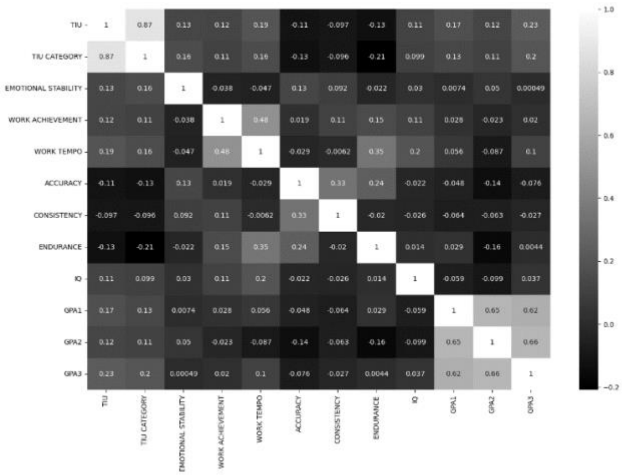

Figure. 4 Heatmap correlation based on feature

The correlation heatmap in Figure. 4 indicates a very weak correlation between the input feature and the target feature, namely a correlation of 0.12 between GPA2 and TIU, a correlation of 0.11 between GPA2 and TIU category, a correlation of 0.05 between GPA2 and Emotional Stability, and negative correlations of -0.023 with work achievement, -0.14 with accuracy, -0.063 with consistency, -0.16 with endurance, and -0.099 with IQ. All available attributes displayed after combining the 2020 and 2021 datasets (datasets based on features).

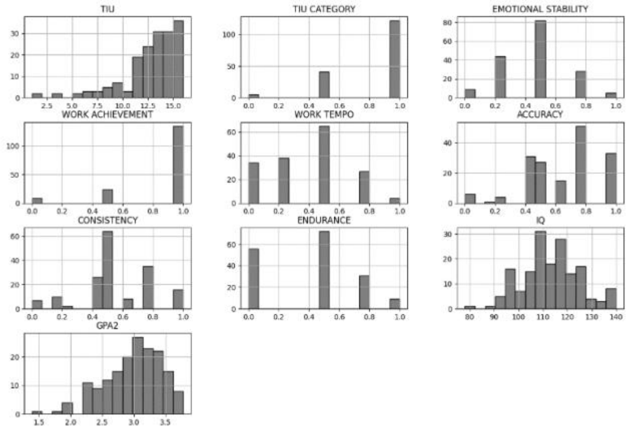

Figure. 5 Available attribute of datasets based on features

So based on Figure. 5, there nine input features, namely TIU, TIU Category, Emotional Stability, Work Achievement, Work Tempo, Accuracy, Consistency, Endurance, IQ, and GPA2 as target features. These attributes describe the attributes available in the dataset based on features. As an explanation, TIU is General Intelligence Test, and TIU Category is labelled in letter form for TIU attributes.

IV. RESULT

The following showed a bar chart comparing the evaluation results of the first Analysis or analysis based on records using 253 data with the results of the second Analysis or Analysis based on features using 168 data used to predict the results of GPA2.

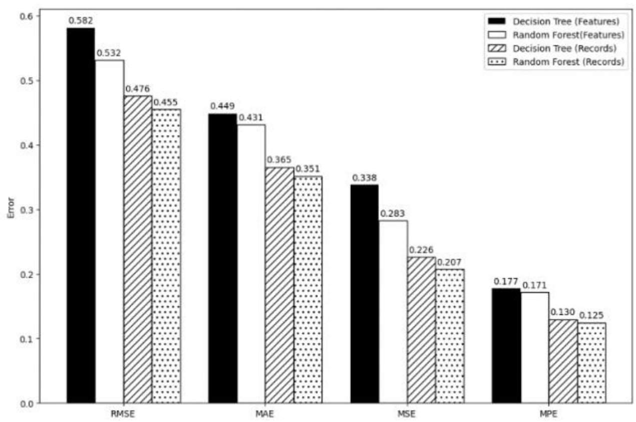

Figure. 6 Comparison of model error evaluation

For special values generated from the evaluation results in Figure. 6 displayed as follows.

TABLE III
EVALUATION RESULTS

Metrics	Decision Tree based on Features	Random Forest based on Features	Decision Tree based on Records	Random Forest based on Records
RMSE	0.5817	0.5317	0.4757	0.4552
MAE	0.4488	0.4312	0.3654	0.3514
MSE	0.3383	0.2827	0.2263	0.2072
MPE	0.1775	0.1714	0.1296	0.1245

Based on TABLE III, the model that produces better performance determined by comparing the two error evaluations, namely MAE and RMSE. The analysis results show that the Random Forest algorithm produces low error evaluations or the model performs better using more features. Model training and testing from among the four analyses were performed. In this context, there is a comparison between the use of features and records in forming Decision Tree and Random Forest models. Models based on Random Forest tend to provide better performance (with lower metric values) compared to models that only use Decision Trees. Besides that, using records as a basis for building models also tends to provide better results than using features. Thus, developing the best model and producing the lowest error by the research title Predicting Del student academic achievement based on Psychological test using Machine Learning is the Random Forest algorithm based on features or the Random Forest algorithm with complex records.

V. CONCLUSION

Based on the research analysis that has been done, namely Analysis based on features and Analysis based on records, a comparison is made between the evaluation error of the decision tree algorithm. The best performance is an analysis based on records using the Random Forest algorithm. The Analysis based on records showed a very weak correlation between the input feature and the target feature, specifically a correlation of 0.076 between emotional Stability and GPA 2, indicating a very weak positive correlation. Additionally, a negative correlation of -0.071 between IQ and GPA 2 indicates a very weak negative correlation. This is shown in the correlation heatmap, which shows the level of correlation for each attribute, which is still classified as a very weak correlation.

Meanwhile, the evaluation results showed low and good errors, with RMSE of 0.4552, MAE of 0.3514, MSE of 0.2072, and MPE of 0.1245. Even though the development of machine learning models has been successfully carried out in research, certain psychological test aspects do not guarantee each student's performance. Therefore, it is essential to recognize that assessing student achievement should not only focus on psychological or academic tests but also consider aspects such as motivation, personality, interests, social skills, and creativity. A holistic and comprehensive approach to evaluating student achievement can provide a completer and more accurate picture of student's abilities and potential.

ACKNOWLEDGMENT

This publication is supported by Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM), Institut Teknologi Del (009.16/ITDel/LPPM/Penelitian/III/2023).

REFERENCES

[1] H. Altabrawee, O. A. J. Ali, and S. Q. Ajmi, “Predicting Students’ Performance Using Machine Learning Techniques,” *J. Univ. BABYLON Pure Appl. Sci.*, vol. 27, no. 1, pp. 194–205, 2019, doi: 10.29196/jubpas.v27i1.2108.

[2] I. Baron, H. Agustina, and Melania, “Journal of Management and Marketing Review The Role of Psychological Testing As an Effort to Improve Employee Competency,” *J. Manag. Mark. Rev.*, vol. 5, no. 1, pp. 1–15, 2020, [Online]. Available: https://doi.org/10.35609/jmmr.2020.5.1.

[3] E. Tanuar, Y. Heryadi, Lukas, B. S. Abbas, and F. L. Gaol, “Using Machine Learning Techniques to Earlier Predict Student’s Performance,” *1st 2018 Indones. Assoc. Pattern Recognit. Int. Conf. Ina. 2018 - Proc.*, pp. 85–89, 2019, doi: 10.1109/INAPR.2018.8626856.

[4] M. L. Brewer *et al.*, “Resilience in higher education students: a scoping review,” *High. Educ. Res. Dev.*, vol. 38, no. 6, pp. 1105–1120, 2019, doi: 10.1080/07294360.2019.1626810.

[5] R. S. Gregorio, “Emotional stability as a key competence of managers,” *Int. J. Humanit. Cult. Stud.*, vol. 3, 2018.

[6] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[7] M. S. Acharya, A. Armaan, and A. S. Antony, “A comparison of regression models for prediction of graduate admissions,” *ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc.*, pp. 1–5, 2019, doi: 10.1109/ICCIDS.2019.8862140.

[8] L. Falát and T. Piscová, “Predicting GPA of University Students with Supervised Regression Machine Learning Models,” *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178403.

[9] H. Dabiri, V. Farhangi, M. J. Moradi, M. Zadehmohamad, and M. Karakouzian, “Applications of Decision Tree and Random Forest as Tree-Based Machine Learning Techniques for Analyzing the Ultimate Strain of Spliced and Non-Spliced Reinforcement Bars,” *Appl. Sci.*, vol. 12, no. 10, pp. 1–13, 2022, doi: 10.3390/app12104851.

[10] L. He, S. Diego, R. A. Levine, and S. Diego, “Random Forest as a Predictive Analytics Alternative to Regression in Institutional Research,” vol. 23, no. 1, 2018.

[11] D. Graziotin, P. Lenberg, R. Feldt, and S. Wagner, “Psychometrics in Behavioral Software Engineering: A Methodological Introduction with Guidelines,” *ACM Trans. Softw. Eng. Methodol.*, vol. 31, no. 1, pp. 1–36, 2022, doi: 10.1145/3469888.

[12] G. Orrù, M. Monaro, C. Conversano, A. Gemignani, and G. Sartori, “Machine Learning in Psychometrics and Psychological Research,” vol. 10, no. January, pp. 1–10, 2020, doi: 10.3389/fpsyg.2019.02970.

[13] R. Kumar, P. Kumar, and Y. Kumar, “Time Series Data Prediction using IoT and Machine Learning Technique,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 373–381, 2020, doi: 10.1016/j.procs.2020.03.240.

[14] A. Taufiqurrahman, A. G. Putrada, and F. Dawani, “Decision Tree Regression with AdaBoost Ensemble Learning for Water Temperature Forecasting in Aquaponic Ecosystem,” *6th Int. Conf. Interact. Digit. Media, ICIDM 2020*, no. Icidm, 2020, doi: 10.1109/ICIDM51048.2020.9339669.

[15] M. G. Uddin and M. Uddin, “E-Government Development & Digital Economy: Relationship,” *Am. Econ. Soc. Rev.*, vol. 6, no. 1, pp. 39–54, 2020, doi: 10.46281/aesr.v6i1.580.