E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

ITAF Kupang New Student Admission Prediction Using The Random Forest Method

Mohamad Iqbal Ulumando[1*], **Orry Adrianus Mokola**[2]
[1,2]Informatics Study Program, Institut Teknologi Alberth Foenay Kupang, Indonesia
[1*]iqbal77ulumando@gmail.com, [2]orrym5992@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 25-05-2026
Accepted: 06-06-2026
Published: 11-06-2026

Keywords:
New Student Prediction; Random Forest; Data Mining; Student Admission; ITAF Kupang.

ABSTRACT

New student admission is a crucial aspect of higher education academic planning. The Alberth Foenay Institute of Technology (ITAF) Kupang requires a data-driven approach to predict the number of new students in each study program to support more accurate decision-making. This study aims to predict the number of new student admissions at ITAF Kupang in the 2026/2027 academic year using the Random Forest method. The data used comes from historical data on new student admissions over the past five years (2021–2025) in three study programs: Informatics, Environmental Engineering, and Mechanical Engineering. The year and study program variables are used as input variables, while the number of new students is used as the output variable. The research stages include data pre-processing, transformation and encoding of categorical variables, Random Forest modeling, and model evaluation using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The model evaluation results show an MAE value of 9.11 and an RMSE of 10.58, indicating that the model has quite good predictive performance. The prediction results show that the number of new students in the 2026/2027 academic year is estimated to be 41 students for the Informatics Study Program, 24 students for Environmental Engineering, and 16 students for Mechanical Engineering. This research is expected to be a supporting basis for planning new student admissions at ITAF Kupang.

INTRODUCTION

New student admission is an important indicator of the sustainability and development of higher education (Tuadingo et al., 2025). The number of new students accepted each year directly impacts academic planning, human resource availability, facility and infrastructure management, and the sustainability of study programs (Wijaya et al., 2025). Inaccuracy in estimating the number of new students can cause problems, both in the form of excess and insufficient learning capacity (Miftahussa'adiah et al., 2025). Therefore, a predictive approach is needed that can provide a more accurate and data-based estimate of the number of new students (Fathir et al., 2024).

The Alberth Foenay Institute of Technology (ITAF) Kupang, as a growing higher education institution, faces the dynamics of new student admissions that fluctuate annually. Based on new student admission data for three study programs—Informatics, Environmental Engineering, and Mechanical Engineering—over the past five years (2021/2022 to 2025/2026 academic years, significant changes in student numbers are evident. In the Informatics Study Program, the number of new students experienced a sharp increase in the 2023/2024 and 2024/2025 academic years, before declining in the 2025/2026 academic year. Meanwhile, the Environmental Engineering and Mechanical Engineering Study Programs exhibited a fluctuating pattern, with a tendency to increase in certain periods and decrease in others.

In the last three years, the number of new students registered at ITAF Kupang has reached 263 students, with details in the 2023/2024 academic year of 78 students, the 2024/2025 academic year of 127 students, and the 2025/2026 academic year of 58 students. These data indicate that new student admissions are not stable and tend to change from year to year. This condition requires the institution to have analytical tools that are able to predict the number of new student admissions in a more scientific and measurable manner (Rizaldi & Aliyyah, 2024).

To date, the new student admissions planning process is generally still carried out based on simple evaluations of historical data or manual estimates (Kuswanto & Hakim, 2025). This approach has limitations because it is not yet fully capable of capturing non-linear and complex data patterns (Arsanti et al., 2025). Along with the development of information technology, the application of machine learning methods has become one solution that can be used to improve prediction accuracy. One method that is widely used in numerical data prediction is Random Forest (Efendi et al., 2024), namely an ensemble learning algorithm that combines a number of decision trees to produce a more stable and accurate prediction model (Azis et al., 2024). The Random Forest method has the advantage of handling limited historical data and is able to study non-linear patterns of relationships between variables (Ratna Sari & Alfin, 2025). In this study, new student admission data for the last five years, namely from 2021 to 2025, is used as input and output

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

variables in the prediction process. The model built is then used to predict the level of new student admissions in the 2026/2027 academic year for each study program at ITAF Kupang. With this approach, predictions are made scientifically based on real data and formal models (Raditia Vindua et al., 2025), so that the prediction results are expected to support more accurate and evidence-based decision making (Hidayat et al., 2025).

The better the quality of the data and models used, the higher the level of accuracy of the predictions produced (Puspitasari et al., 2025). Researchers also previously conducted research with the same object using the Random Forest method with the title Classification of Dropout Risk of ITAF Kupang Students Using Random Forest as an Early Warning System (Ulumando, 2026a).

LITERATURE REVIEW

New Student Admission Prediction

Forecasting is the process of estimating future values or conditions based on available historical data. In higher education, student admissions forecasting is a crucial aspect of academic planning, resource management, promotional strategy development, and determining student enrollment capacity. Information on the number of students projected to enroll in the next enrollment period can help universities make more informed, data-driven decisions (Kurniasih, 2024).

New student admissions are influenced by various factors, such as the number of applicants in previous years, the community's economic conditions, competition between universities, the effectiveness of promotions, the study programs offered, and government education policies. Therefore, data analysis methods capable of identifying patterns and relationships between variables are necessary to produce accurate predictions (Susilo et al., 2024).

Advances in information technology and artificial intelligence have encouraged the use of machine learning techniques in various prediction problems, including predicting the number of new students. Machine learning methods have the ability to learn patterns from historical data and generate models that can be used to predict future conditions with greater accuracy than conventional methods (Sulehu et al., 2025).

Data Mining

Data mining is the process of discovering patterns, relationships, trends, or valuable information from large data sets using statistical, mathematical, and artificial intelligence techniques. According to the Knowledge Discovery in Databases (KDD) process, data mining is one of the main stages aimed at extracting useful knowledge from available data. The stages of data mining generally include data selection, data cleaning, data transformation, modeling, and evaluation of results. Data mining techniques can be used for various purposes, such as classification, clustering, association, and prediction (Raditia Vindua et al., 2025). In this study, data mining was used to predict the number of new student admissions based on historical data held by the Flores Institute of Technology and Administration (ITAF) Kupang. By utilizing past data, the system can build a predictive model that provides an overview of the number of new students in the following period (Putra & Harahap, 2024).

Machine Learning

Machine learning is a branch of artificial intelligence that allows computers to learn from data without having to be explicitly programmed. Machine learning algorithms work by recognizing patterns in training data and using these patterns to generate predictions or decisions about new data. In general, machine learning is divided into three main categories: supervised learning, unsupervised learning, and reinforcement learning (Kuswanto & Hakim, 2025). In supervised learning, models are built using labeled or target data, allowing them to be used for classification and prediction. In the context of this research, the Random Forest method falls into the supervised learning category and is used to build predictive models based on historical student admissions data. Machine learning's ability to process complex data makes it a widely used approach in predictive research across various fields (Sitanggang et al., 2026).

Random Forest Method

Random Forest is a machine learning algorithm developed by Leo Breiman in 2001. This algorithm belongs to the ensemble learning group which works by building a number of decision trees and combining the results from all the trees to produce more accurate and stable predictions (Puspitasari et al., 2025). The basic principle of Random Forest is to generate multiple decision trees using different data samples using bootstrap sampling. For each tree, attributes are randomly selected, resulting in model variation that reduces the risk of overfitting (Ulumando, 2026b). The final results are obtained through a voting mechanism for classification or an average of the predicted results for regression. The advantages of Random Forest include high accuracy, the ability to handle large amounts of data, robustness to noise and missing values, and the ability to measure the importance of each variable (feature importance). Furthermore, Random Forest is relatively easy to implement and can perform well on a variety of prediction problems (Efendi et al., 2024).

PERFECT JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

METHOD

This study uses a quantitative approach with machine learning methods to predict the number of new student admissions at the Alberth Foenay Institute of Technology (ITAF) Kupang in the 2026/2027 academic year. The Random Forest method used utilizes historical data on new student admissions over the past five years as the basis for developing a predictive model.

Data Types and Sources

The data used in this study is secondary data in the form of new student admissions data at ITAF Kupang in three study programs: Informatics, Environmental Engineering, and Mechanical Engineering. The data was taken from the institution's academic archives and covers a five-year period, namely the 2021/2022 to 2025/2026 academic years. This data is used as input and output variables in the predictive modeling process (Harkamsyah Andrianof et al., 2025).

- **Research Variables**
  The variables used in this study consist of:
  – Input variable (X): the number of new students in each study program based on data from the last five years (2021/2022 to 2025/2026 academic years).
  – Output variable (Y): the predicted number of new students in the 2026/2027 academic year for each study program.
- **Research Stages**
  The research stages are as shown in Figure 1.

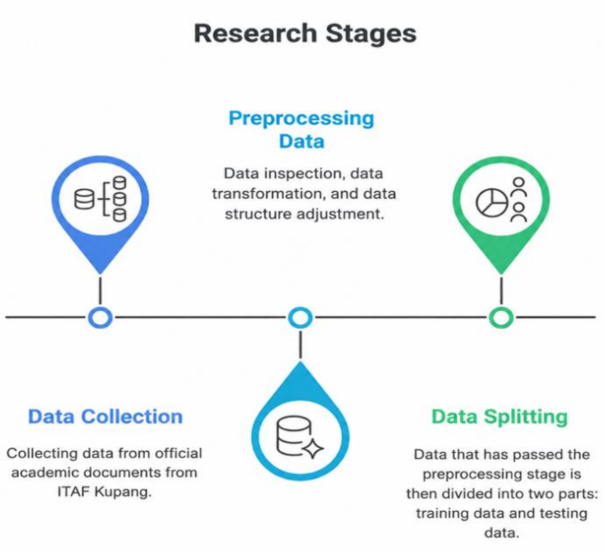

Figure 1. Research Stages

– Data collection
  New student admissions data was collected from official ITAF Kupang academic documents. The data was tabulated by study program and academic year to facilitate analysis and modeling.
– Data Preprocessing
  The pre-processing stage is carried out to ensure data quality before it is used in modeling. This process includes:
  1) Data inspection to ensure there are no missing or inconsistent data.
  2) Data transformation into a numeric format suitable for the Random Forest algorithm.
  3) Data structure adjustment so that each data record can be used as input for the prediction model.
– Data Sharing
  The pre-processed data was then divided into two parts: training data and testing data. The data was divided into 80% training data and 20% testing data. The training data was used to build the Random Forest model, while the testing data was used to measure the model's predictive performance.

PERFECT JOURNAL OF SMART ALGORITHMS
E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

- **Random Forest Modeling**

Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to obtain a final prediction value. Each decision tree is constructed from a randomly selected subset of the training data (bootstrap sampling), using a different subset of features at each node split. This process aims to improve model accuracy and reduce the risk of overfitting.

The main parameters used in Random Forest modeling include the number of decision trees, the maximum depth, and the minimum amount of data per node. Parameters are determined to obtain a model with optimal predictive performance.

- **Model Evaluation**

The performance evaluation of the Random Forest model was carried out using several prediction error measurement metrics, namely:

a) Mean Absolute Error (MAE), to measure the average absolute difference between actual and predicted values.
b) Root Mean Square Error (RMSE), to measure the magnitude of prediction error by imposing a larger penalty on larger errors.

The evaluation results are used to assess the level of accuracy and reliability of the Random Forest model in predicting the number of new student admissions.

- **New Student Admission Prediction**

The developed and evaluated Random Forest model was then used to predict the number of new student admissions for the 2026/2027 academic year for each study program at ITAF Kupang. These predictions are expected to inform more accurate, data-driven academic planning and decision-making.

RESULT

Data Sources and Forms

The data used in this study comes from new student admission data at the Alberth Foenay Institute of Technology (ITAF) Kupang for the last 5 years (2021–2025) in three study programs shown in Table 1 below.

Table 1. Student Data of the Alberth Foenay Institute of Technology (ITAF) Kupang

Study Program	2021	2022	2023	2024	2025
Informatics	5	3	40	78	35
Environmental Engineering	12	10	20	30	11
Mechanical Engineering	0	3	18	19	12

The data above is then converted into numerical form and prepared for Random Forest modeling with: Year and study program variables as input and the number of students variable as output in the form of a graph shown in Figure 2 below:

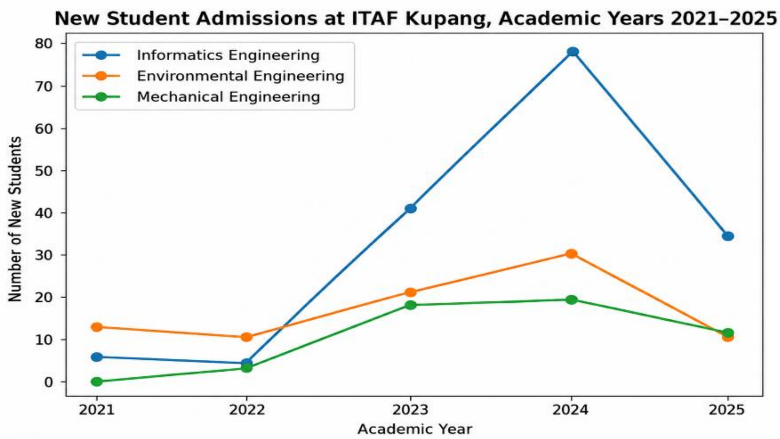

Figure 2. Graph of New Student Admissions at ITAF Kupang 2021-2025

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

Data pre-processing

Data pre-processing is the initial stage aimed at ensuring the data used in modeling is of good quality and meets the requirements of the Random Forest algorithm. In this study, new student admissions data was obtained in tabular form by study program and academic year. The pre-processing stages included:

1. Data checking to ensure there are no missing values and that the data is consistent.
2. Data transformation: converting tabular data into a numeric format so it can be processed by a machine learning algorithm.
3. Categorical variable encoding: converting the study program variable into a numeric format using label encoding techniques.
4. Final dataset compilation: each row of data represents a single observation, with input variables being the year and study program, and output variables being the number of new students.

The result of this pre-processing stage is a numeric dataset ready for use in training and testing the Random Forest model. The following are the steps in data pre-processing using Python:

Data check

The data checking stage was conducted to ensure that the data used in the study was complete, consistent, and suitable for further processing using the Random Forest algorithm. This checking aimed to identify missing values, typos, and data type conformity for each variable. The following is the data checking process using the Python programming language, and the results are displayed in Table 2 below.

```
# Import library
import pandas as pd

# Membuat dataset
df = pd.DataFrame({
    "Program_Studi": [
        "Informatika","Informatika","Informatika","Informatika","Informatika",
                  "Teknik  Lingkungan","Teknik  Lingkungan","Teknik  Lingkungan","Teknik
Lingkungan","Teknik Lingkungan",
        "Teknik Mesin","Teknik Mesin","Teknik Mesin","Teknik Mesin","Teknik Mesin"
    ],
    "Tahun": [
        2021,2022,2023,2024,2025,
        2021,2022,2023,2024,2025,
        2021,2022,2023,2024,2025
    ],
    "Jumlah_Mahasiswa": [
        5,3,40,78,35,
        12,10,20,30,11,
        0,3,18,19,12
    ]
})
# Menampilkan data awal
df
```

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

Table 2. Data Examination Results

Index	Study Program	Year	Number of Students
0	Informatics	2021	5
1	Informatics	2022	3
2	Informatics	2023	40
3	Informatics	2024	78
4	Informatics	2025	35
5	Environmental Engineering	2021	12
6	Environmental Engineering	2022	10
7	Environmental Engineering	2023	20
8	Environmental Engineering	2024	30
9	Environmental Engineering	2025	11
10	Mechanical Engineering	2021	0
11	Mechanical Engineering	2022	3
12	Mechanical Engineering	2023	18
13	Mechanical Engineering	2024	19
14	Mechanical Engineering	2025	12

Based on the results of the data examination, it can be concluded that the new student admissions data used in this study is in a complete and consistent condition so that it can be continued to the next pre-processing stage.

Data transformation

The data transformation stage is carried out to convert raw data into a form suitable for processing by the Random Forest algorithm. In this study, data transformation focused on converting categorical variables, namely study programs, into numerical form through an encoding process. This transformation is necessary because machine learning algorithms cannot process data in text format. The following is the data transformation process using the Python programming language, and the results are displayed in Table 3 below.

```
# Encoding program studi menjadi numerik
df["Kode_Prodi"] = df["Program_Studi"].map({
    "Informatika": 0,
    "Teknik Lingkungan": 1,
    "Teknik Mesin": 2
})
# Menampilkan hasil transformasi
df
```

Table 3. Data Transformation Results

Index	Study Program	Year	Number of Students	Program Code
0	Informatics	2021	5	0
1	Informatics	2022	3	0
2	Informatics	2023	40	0
3	Informatics	2024	78	0
4	Informatics	2025	35	0
5	Environmental Engineering	2021	12	1
6	Environmental Engineering	2022	10	1
7	Environmental Engineering	2023	20	1
8	Environmental Engineering	2024	30	1
9	Environmental Engineering	2025	11	1
10	Mechanical Engineering	2021	0	2
11	Mechanical Engineering	2022	3	2
12	Mechanical Engineering	2023	18	2
13	Mechanical Engineering	2024	19	2
14	Mechanical Engineering	2025	12	2

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

Data transformation was performed by converting the study program variables into numerical form using label encoding techniques. This process ensured the dataset could be processed by the Random Forest algorithm. The transformation results showed that all data was now in numerical format and ready for use in the modeling phase.

Encoding categorical variables

The categorical variable encoding stage aims to convert non-numerical data into numeric form so that it can be processed by a machine learning algorithm. In this study, the categorical variable is the study program. The encoding process is necessary because the Random Forest algorithm can only work with numeric data. The following is the categorical variable encoding process using the Python programming language, and the results of the categorical variable encoding are displayed in Table 4 below.

```
# Encoding variabel kategorikal (Program Studi)
df["Kode_Prodi"] = df["Program_Studi"].map({
    "Informatika": 0,
    "Teknik Lingkungan": 1,
    "Teknik Mesin": 2
})

# Menampilkan hasil encoding
df
```

Table 4. Results of Encoding Categorical Variables

Index	Study Program	Year	Number of Students	Program Code
0	Informatics	2021	5	0
1	Informatics	2022	3	0
2	Informatics	2023	40	0
3	Informatics	2024	78	0
4	Informatics	2025	35	0
5	Environmental Engineering	2021	12	1
6	Environmental Engineering	2022	10	1
7	Environmental Engineering	2023	20	1
8	Environmental Engineering	2024	30	1
9	Environmental Engineering	2025	11	1
10	Mechanical Engineering	2021	0	2
11	Mechanical Engineering	2022	3	2
12	Mechanical Engineering	2023	18	2
13	Mechanical Engineering	2024	19	2
14	Mechanical Engineering	2025	12	2

Categorical variables were encoded using the label encoding method to convert study program variables into numeric form. This process generates a numeric code representing each study program and ensures that all variables in the dataset can be processed by the Random Forest algorithm.

Final dataset compilation

The final dataset preparation stage aims to create a data structure ready for use in the Random Forest modeling process. At this stage, the data that has undergone the inspection and encoding process is reorganized by separating the input variables (X) and output variables (y) according to the requirements of the machine learning algorithm. The following is the process of compiling the final dataset using the Python programming language.

```
# Menyusun dataset akhir
X = df[["Tahun", "Kode_Prodi"]]   # Variabel input
y = df["Jumlah_Mahasiswa"]        # Variabel output

# Menampilkan dataset akhir
       X, y
```

PERFECT JOURNAL OF SMART ALGORITHMS
E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

```
(    Year  Code Program
 0   2021            0
 1   2022            0
 2   2023            0
 3   2024            0
 4   2025            0
 5   2021            1
 6   2022            1
 7   2023            1
 8   2024            1
 9   2025            1
 10  2021            2
 11  2022            2
 12  2023            2
 13  2024            2
 14  2025            2,
 0      5
 1      3
 2     40
 3     78
 4     35
 5     12
 6     10
 7     20
 8     30
 9     11
 10     0
 11     3
 12    18
 13    19
 14    12
 Name: Number of Students, dtype: int64)
```

The final dataset was compiled by separating the input variables, academic year and study program code, from the output variable, the number of new students. This dataset served as the basis for training and testing a Random Forest model to predict the number of new student admissions for the following academic year.

Random Forest Modeling

The modeling phase aims to build a Random Forest model capable of studying the relationship between academic year and study program and the number of new students. This model is used to predict the number of new student admissions for the 2026/2027 academic year based on historical data from the previous five years (2021–2025). Random Forest was chosen because it can handle non-linear data, reduce the risk of overfitting, and provide stable prediction results despite a relatively limited data set. Due to the relatively small data set, the data was divided into 80% training data and 20% testing data. This division aims to test the model's ability to predict on previously unseen data. The following is the source code for the training data using the Python programming language.

```
from sklearn.model_selection import train_test_split

# Membagi data
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

X_train.shape, X_test.shape
((12, 2), (3, 2))
```

The Random Forest model is built using multiple decision trees working collectively. Each tree produces a prediction, and the final result is obtained by averaging the predictions of all the trees. Below is the source code for the test data using the Python programming language.

PERFECT JOURNAL OF SMART ALGORITHMS
E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

```
from sklearn.ensemble import RandomForestRegressor

# Membuat model Random Forest
rf_model = RandomForestRegressor(
    n_estimators=100,
    random_state=42
)

# Melatih model
rf_model.fit(X_train, y_train)
```

The modeling was performed using the Random Forest algorithm, utilizing historical student admissions data as training data. The dataset was divided into training and test data in an 80:20 ratio. The model was built using 100 decision trees and trained to learn the relationship between academic year, study program, and the number of new students.

Model Evaluation

Model evaluation was conducted to measure the accuracy and prediction error of the Random Forest model. This stage aimed to determine the model's ability to predict the number of new students based on historical data used as training data. In this study, the evaluation was conducted using several error metrics commonly used in regression, namely Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results of the model evaluation using the Python programming language are shown in Figure 3.

```
from sklearn.metrics import mean_absolute_error, mean_squared_error
import numpy as np

# Menghitung Mean Absolute Error (MAE)
mae = mean_absolute_error(y_test, y_pred)

# Menghitung Root Mean Squared Error (RMSE)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print("Mean Absolute Error (MAE):", mae)
print("Root Mean Squared Error (RMSE):", rmse)

Mean Absolute Error (MAE): 9.11
Root Mean Squared Error (RMSE): 10.58435480634822
```

Figure 3. Results of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)

The Random Forest model evaluation results showed a Mean Absolute Error (MAE) of 9.11. This value indicates that, on average, the absolute difference between the number of new students predicted by the model and the actual data is approximately 9 students. In other words, the model is able to predict the number of new student admissions with a relatively small error rate for each study program and academic year.

Furthermore, the Root Mean Squared Error (RMSE) value was 10.58. The RMSE value, which is slightly larger than the MAE, indicates some prediction errors that are larger, but still within reasonable limits. This indicates that the model does not produce significant outliers and has a fairly good level of stability.

Based on these evaluation results, it can be concluded that the Random Forest model is able to learn new student admission patterns based on historical data from the past five years quite well. Although the data used is relatively limited, the MAE and RMSE values obtained indicate that the model has adequate performance for predicting new student admissions for the 2026/2027 academic year.

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

Predicted Results Study Program

Based on the Random Forest model, which was built and evaluated using historical data on new student admissions at the Alberth Foenay Institute of Technology, Kupang, over the past five years (2021–2025), a prediction process for the number of new student admissions for the 2026/2027 academic year was performed. This prediction was conducted separately for each study program, using the academic year and study program variables as model inputs.

The prediction process aims to provide an initial overview of the potential number of new students admitted to each study program in the upcoming academic year. The prediction results are estimates based on historical data and patterns learned by the model. Therefore, they can be used as considerations in academic planning, resource management, and strategic decision-making at the Alberth Foenay Institute of Technology, Kupang.

The following are the prediction results for new student admissions at the Alberth Foenay Institute of Technology, Kupang, for the 2026/2027 academic year, displayed using the Python programming language. Table 5 shows the prediction results.

```
import pandas as pd


# Data prediksi untuk tahun 2026
data_prediksi = pd.DataFrame({
    "Tahun": [2026, 2026, 2026],
    "Kode_Prodi": [0, 1, 2]  # 0=Informatika, 1=Teknik Lingkungan, 2=Teknik Mesin
})


data_prediksi
# Prediksi jumlah mahasiswa baru 2026/2027
hasil_prediksi = rf_model.predict(data_prediksi)


hasil_prediksi
# Membuat tabel hasil prediksi
hasil_prediksi_df = pd.DataFrame({
    "Program Studi": ["Informatika", "Teknik Lingkungan", "Teknik Mesin"],
    "Prediksi Mahasiswa Baru 2026/2027": hasil_prediksi.round().astype(int)
})


hasil_prediksi_df
```

Table 5. Predicted Results of New Student Admissions at ITAF Kupang for the 2026/2027 Academic Year

index	Study Program	New Student Predictions 2026/2027
0	Informatics	41
1	Environmental Engineering	24
2	Mechanical Engineering	16

Based on modeling and prediction results using the Random Forest algorithm, estimates were obtained for the number of new student admissions at the Alberth Foenay Kupang Institute of Technology for the 2026/2027 academic year for each study program. The predictions indicate that the Informatics Study Program is expected to admit 41 new students, the Environmental Engineering Study Program 24 new students, and the Mechanical Engineering Study Program 16 new students. These predictions indicate that the Informatics Study Program remains the study program with the highest new student admission rate compared to other study programs. This aligns with historical trends in previous years, where the Informatics Study Program demonstrated relatively higher growth and applicant interest. Meanwhile, the Environmental Engineering and Mechanical Engineering Study Programs showed more moderate, but stable, admission rates based on historical data patterns studied by the model.

Overall, these predictions provide a quantitative overview of the potential for new student admissions in the 2026/2027 academic year. These predictions can be used as a basis for academic planning, resource allocation, and the development of promotion and admission strategies at the Alberth Foenay Institute of Technology, Kupang. However,

PERFECT JOURNAL OF SMART ALGORITHMS
E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

these predictions are still dependent on historical data patterns, so external factors outside the model, such as institutional policies, economic conditions, and prospective student interests, also need to be considered in decision-making.

Graph of Predicted Results for New Students at ITAF Kupang for the 2026/2027 Academic Year

To facilitate understanding of the predicted number of new student admissions for the 2026/2027 academic year, the results of the Random Forest modeling are presented in graphical form. The graphical presentation aims to provide a visual representation of the comparison of the predicted number of new students for each study program at the Alberth Foenay Institute of Technology, Kupang.

The graphs displayed show the distribution of predicted new student admissions for each study program, facilitating analysis of differences and trends in the number of new students across study programs. Through this visualization, the institution can more easily identify study programs with the highest admission potential and those requiring greater attention in promotional strategies and academic planning. Figure 4 shows the graph of the predicted results of new students at ITAF Kupang for the 2026/2027 academic year.

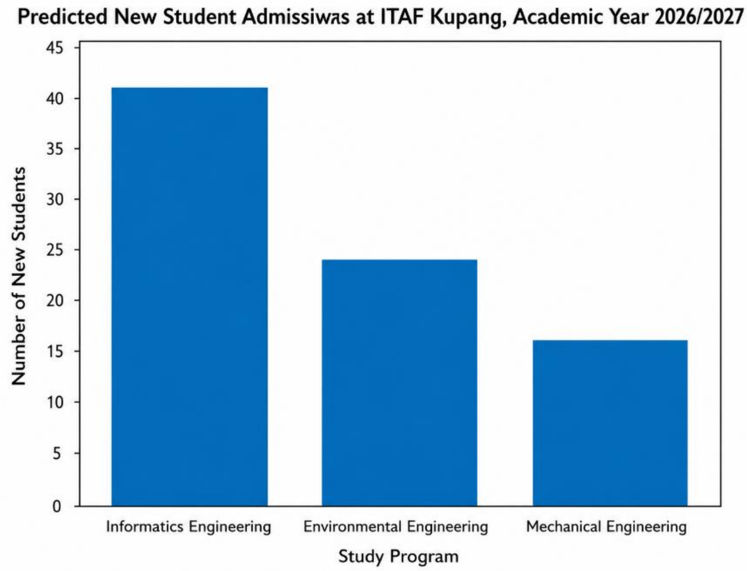

Figure 4. Graph of Predicted Results for New Students at ITAF Kupang for the 2026/2027 Academic Year

Overall, this prediction graph shows differences in new student acceptance rates across study programs at the Alberth Foenay Institute of Technology, Kupang. The graphic visualization helps clarify the predictions obtained from the Random Forest model and can be used as a basis for strategic decision-making related to academic planning, resource management, and new student admission strategies for the 2026/2027 academic year.

DISCUSSION

The results of this study indicate that the Random Forest algorithm can be used to predict the number of new student admissions at the Alberth Foenay Institute of Technology (ITAF) Kupang based on historical student admissions data for the past five years (2021–2025). The model evaluation values of MAE 9.11 and RMSE 10.58 indicate that the model has a relatively low prediction error rate and is within an acceptable range for academic planning purposes. The RMSE value, which is not significantly greater than the MAE, indicates that the model does not produce many extreme prediction errors, thus the prediction performance can be considered quite stable.

This finding aligns with the characteristics of Random Forest as an ensemble learning method that combines multiple decision trees to increase accuracy and reduce the risk of overfitting. In this research dataset, the new student admission pattern is fluctuating and not entirely linear, particularly in the Informatics Study Program, which experienced a sharp increase in 2023–2024 before declining in 2025. Random Forest's ability to learn nonlinear relationships allows the model to maintain logical predictions even though the data shows changes in patterns between years. This strengthens the argument that ensemble methods are more appropriate than simple approaches that assume a stable linear trend.

Predictions for the 2026/2027 academic year show that the Informatics Study Program is expected to admit 41 new students, Environmental Engineering 24 students, and Mechanical Engineering 16 students. The dominant

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

predictions for the Informatics Study Program are consistent with historical trends showing higher growth rates and prospective student interest than the other two study programs. Substantively, these results can be interpreted as indicating that the information technology field still has strong market appeal, both due to the need for a digital workforce and the perception of better career prospects. In other words, the model not only captures past numerical patterns but also reflects the relatively sustainable trend in demand for education in Informatics.

On the other hand, the prediction results for Environmental Engineering and Mechanical Engineering show a more moderate growth pattern. The predictions of 24 and 16 new students do not mean that these two study programs are less promising, but rather indicate that the historical pattern of admissions in these two study programs tends to be stable and does not experience spikes as large as those for Informatics. In the context of institutional planning, this information is important because it suggests that program promotion and development strategies can be differentiated across study programs. Programs with high growth potential require greater academic capacity, while programs with moderate growth may require strengthened promotion or program differentiation to increase their appeal to prospective students.

The results also indicate that the use of simple variables—academic year and study program—is sufficient to produce a model with adequate performance. This finding indicates that these two variables contain significant information about the variation in the number of new students at ITAF Kupang. However, the model's accuracy could potentially be improved by adding other variables more closely related to enrollment behavior, such as the number of applicants, the selection pass rate, promotion intensity, prospective students' school of origin, regional economic conditions, and the level of competition between universities. The absence of these variables makes the current model more representative of aggregate historical patterns than the mechanisms causing changes in the number of new students. From a practical perspective, the prediction results provide a quantitative basis for planning new student admissions. The estimated number of students per study program can be used to determine lecturer needs, class capacity, class schedules, laboratory facilities, and operational budgeting. The use of a machine learning approach also helps reduce reliance on subjective manual estimates. Thus, the research findings have implementable value as a decision support system in academic management at universities.

However, interpretation of the results requires consideration of the study's limitations. The dataset used only covers five years of historical data with a total of 15 observations, thus limiting the model's ability to capture long-term patterns. Furthermore, changes in government policy, economic conditions, institutional reputation, promotional strategies, and prospective student preferences can significantly alter future admission patterns without being reflected in the historical data used. Therefore, the prediction results should be treated as estimates based on historical patterns, not as exact figures. Further research with a longer data period and richer variables is likely to improve the model's generalizability and accuracy.

Overall, this study demonstrates that Random Forest is a viable approach for predicting new student admissions at ITAF Kupang. The combination of good evaluation performance, the ability to handle nonlinear patterns, and prediction results consistent with historical trends support the use of this method as a tool for academic planning and data-driven strategic decision-making in higher education.

CONCLUSION

Based on the results of research that has been conducted regarding the prediction of new student admissions at the Alberth Foenay Institute of Technology (ITAF) Kupang using the Random Forest method, several conclusions can be drawn as follows.

1. The Random Forest method has been proven effective in modeling and predicting new student admissions based on historical admissions data for the past five years (2021–2025) across three study programs: Informatics, Environmental Engineering, and Mechanical Engineering. This model is capable of capturing nonlinear patterns in the data despite the relatively limited dataset size.
2. The model evaluation results show a Mean Absolute Error (MAE) of 9.11 and a Root Mean Squared Error (RMSE) of 10.58. These values indicate that the average model prediction error is within acceptable limits for small-scale data, thus demonstrating a fairly good level of accuracy in predicting the number of new students.
3. The prediction results for new student admissions for the 2026/2027 academic year indicate that the Informatics Study Program is predicted to accept the highest number of new students, at 41 students, followed by the Environmental Engineering Study Program with 24 students, and the Mechanical Engineering Study Program with 16 students. The differences in predicted numbers between study programs reflect the differences in prospective students' interests in each field of study.
4. The graphical visualization of the prediction results provides a clear and informative picture of the distribution of new students in each study program. This graph can be used as a tool in strategic decision-making related to academic planning and resource management at ITAF Kupang.

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

Overall, this study shows that a data-driven prediction approach using Random Forest can provide objective and evidence-based support in planning new student admissions in higher education environments.

REFERENCES

Arsanti, Y., Minsaris, L. O. A., & Arifin, W. A. (2025). Perbandingan Model Prediksi Suhu Permukaan Laut Menggunakan Smoothing dan Long Short-Term Memory. *Jurnal Algoritma*, *22*(1), 1026–1038. https://doi.org/10.33364/algoritma/v.22-1.2113

Azis, H., Alisma, A., Purnawansyah, P., & Nirmala, N. (2024). Analisis Kinerja Algoritma Pembelajaran Mesin Ensembel Pada Dataset Multi Kelas Citra Jaffe. *Ilmiah NERO*, *9*(2), 107–118. https://doi.org/10.21107/nero.v9i2.27872

Efendi, M. S., Sarwido, & Zyen, A. K. (2024). Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *RESOLUSI : Rekayasa Teknik Informatika Dan Informasi*, *5*(1), 12–20. https://doi.org/10.30865/resolusi.v5i1.2149

Fathir, Sutriawan, & Zumhur Alamin. (2024). Pengembangan Sistem Penerimaan Mahasiswa Baru Menggunakan Arsitektur Golang Framework. *Scientific : Journal of Computers Sciences and Informatics*, *1*(1), 1–8. https://doi.org/10.34304/scientific.v1i1.226

Harkamsyah Andrianof, Aggy Pramana Gusman, & Okta Andrica Putra. (2025). Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia. *Jurnal Sains Informatika Terapan (JSIT)*, *4*(1), 24–28.

Hidayat, R., Tri Saputra, H., Husnah, M., Nabila, N., Hidayatullah, M. B., Naufal Nazhmi, M., Azra, J., & Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produk di Supermarket. *Simkom (Sistem Informasi Dan Sistem Komputer)*, *10*(1), 101–109. https://doi.org/10.51717/simkom.v10i1.703

Kurniasih, R. (2024). Prediksi Jumlah Mahasiswa Baru Dengan Menggunakan Regresi Linier Sederhana. *Jurnal Matematika*, *24*(1), 84–89. https://journals.unisba.ac.id/index.php/matematika%0ADiterima:

Kuswanto, J., & Hakim, L. (2025). Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa. *Decode: Jurnal Pendidikan Teknologi Informasi*, *5*(1), 262–270. https://doi.org/10.51454/decode.v5i1.1103

Miftahussa'adiah, M., Adi, W. C., & Hadi, A. (2025). Analisis Swot Untuk Penerimaan Mahasiswa Baru (PMB) PTKIN 2024. *Research and Development Journal of Education*, *11*(2), 863–875. https://doi.org/10.30998/rdje.v11i2.11180

Puspitasari, D., Nada, N. Q., & Jaka Harjanta, A. T. (2025). Penerapan Algoritma Random Forest Untuk Prediksi Biaya Kontruksi Berbasis Web. *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, *8*(2), 314–326. https://doi.org/10.47324/ilkominfo.v8i2.385

Putra, M., & Harahap, E. (2024). Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika (JRM)*, *4*(2), 127–136. https://doi.org/https://doi.org/10.29313/jrm.v4i2.5102

Raditia Vindua, Salsa Sayida Bilqis, Hamdan Qo'du Ilal Hakim, & Rido Ramadhan. (2025). Prediksi Kemacetan Lalu Lintas Menggunakan Algoritma Random Forest. *TIN: Terapan Informatika Nusantara*, *6*(7), 1071–1077. https://doi.org/10.47065/tin.v6i7.8806

Ratna Sari, N., & Alfin, A. (2025). Optimalisasi Rencana Produksi untuk Mengurangi Overstock dan Stockout di Divisi Production Planning and Inventory Control (PPIC) Menggunakan Random Forest. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, *5*(2), 149–165. https://doi.org/10.55382/jurnalpustakaai.v5i2.1040

Rizaldi, M. A., & Aliyyah, R. R. (2024). Strategi Pemasaran Penerimaan Mahasiswa Baru. *Karimah Tauhid*, *3*(2), 2288–2311. https://doi.org/10.30997/karimahtauhid.v3i2.12079

Sitanggang, A. A., Lase, M. Y., Sipayung, S. P., Informatika, T., Katolik, U., & Thomas, S. (2026). Prediksi Kelulusan Mahasiswa Menggunakan Logistic Regression dan Random Forest Berdasarkan Data Akademik. *Jurnal Pendidikan Tambusai*, *10*(1), 3503–3513.

Sulehu, M., Wisda, W., Wanita, F., & Markani, M. (2025). Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi. *Jurnal Minfo Polgan*, *13*(2), 2364–2374. https://doi.org/10.33395/jmp.v13i2.14472

Susilo, B., Ramdhan, N. A., & Bachri, O. S. (2024). Application of the K-Nearest Neighbor Algorithm for Predicting Digital Product Sales. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(4), 1466–1476. https://doi.org/https://doi.org/10.57152/malcom.v4i4.1517

Tuadingo, R. D., Rachmat, E., & Gobel, L. Van. (2025). Efektivitas Program Penerimaan Mahasiswa Baru (PMB)

E-ISSN: E-ISSN: 3064-0377
Volume 3, **Number** 2, July 2026
https://doi.org/10.62671/perfect.v3i2.284

Dalam Meningkatkan Minat Masyarakat Untuk Melanjutkan Studi Di Universitas Bina Taruna Gorontalo. *Journal of Innovative and Creativity*, *5*(2), 23930–23938.

Ulumando, M. I. (2026a). Klasifikasi Resiko DropOut Mahasiswa ITAF Kupang Menggunakan Random Forest Sebagai Sistem Peringatan Dini. *Simkom (Sistem Informasi Dan Sistem Komputer)*, *11*(1), 116–130. https://doi.org/https://doi.org/10.51717/simkom.v11i1.1255

Ulumando, M. I. (2026b). Prediksi Harga Motor Bekas Di Kota Kupang Menggunakan Metode Random Forest. *SULIWA: Jurnal Multidisiplin Teknik, Sains, Pendidikan Dan Teknologi*, *3*(2), 83–93. https://doi.org/https://doi.org/10.62671/suliwa.v3i2.258

Wijaya, H. C., Tungadi, D., Wangarry, V., Santoso, R. C., & Sanapang, G. M. (2025). Faktor-Faktor Penentu Keputusan Mahasiswa Dalam Memilih Perguruan Tinggi : Studi Empiris Pada Kampus X Di Makassar. *Jurnal Online Manajemen ELPEI (JOMEL)*, *5*(1), 1262–1270.