

Analysis Forecasting Students Using Random Forest and Linear Regression Algorithms

Muhammad Nurdin^{1)*}, Fauziah²⁾,

^{1,2)}Nasional University, Indonesia

¹⁾ muhhammadnurdin2022@student.unas.ac.id, ²⁾ fauziah@civitas.unas.ac.id

Submitted : July 17, 2024 | **Accepted** : Aug 21, 2024 | **Published** : Oct 7, 2024

Abstract: Planning the number of new student enrollments (PMB) is a crucial aspect for universities as it has a significant impact on human resource planning and budget allocation. However, accurately predicting the number of new students poses various challenges due to external factors such as economic conditions, government policies, and competition among educational institutions. This study aims to analyze the forecasting of new student enrollments at Universitas Nasional using the Random Forest (RF) and Linear Regression (LR) algorithms. The data used in this research includes historical data on student application form purchases from 2017 to 2023, with a focus on form purchases from 2017 to 2023. During this period, the number of form purchases totaled 37,078, with the following breakdown per year: 5,178 in 2017, 6,733 in 2018, 5,978 in 2019, 5,042 in 2020, 5,184 in 2021, 5,360 in 2022, and 3,603 in 2023. Two forecasting methods, RF and LR, were applied to evaluate and contrast the effectiveness of each predictive method. The analysis results show that the RF model outperformed the LR model. This is demonstrated by the reduced Mean Absolute Error (MAE) in RF (202.614) compared to LR (503.233), signifying a lower average absolute error in the RF model. Additionally, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were also lower in Random Forest (69,764.003 and 264.178) compared to LR (268,276.644 and 518.023), suggesting that RF is more effective in handling outliers and has a lower average squared error. The lower Mean Absolute Percentage Error (MAPE) in RF (4.109%) compared to LR (10.647%) indicates a smaller relative error. Furthermore, the lower Median Absolute Deviation (MAD) in RF (187.781) also demonstrates more consistent predictions compared to LR (467.999). Overall, the RF model provides more accurate and consistent predictions, leading to the conclusion that RF is the superior model for forecasting the number of new students compared to LR

Keywords: Student Admission Forecasting, Random Forest Algorithm, Linear Regression, Historical Admission Data, Prediction Accuracy

INTRODUCTION

Nasional University is one of the leading higher education institutions in Indonesia, and it was established in 1949. Since its founding, it has experienced significant growth and has become an important choice for prospective students seeking high-quality and reputable higher education. In the real of higher education, forecasting new student admissions (PMB) is a crucial aspect for universities to accurately plan their human resource and budgetary needs. Forecasting has consistently been crucial in decision-making and planning, helping to navigate the uncertainties of the future, which can be both fascinating and difficult. This drives individuals and organizations to focus on reducing risks while increasing potential benefits. (Petropoulos et al., 2022). The accuracy in predicting the number of prospective students affects the university's ability to manage class capacity, allocate faculty, and provide adequate supporting facilities. However, making accurate predictions about new student admissions is a challenge due to various external elements like economic conditions, government policies, and competition among universities. In an effort to enhance the precision of new student admission forecasts, this study employs two popular machine learning algorithms: RF and LR. Random forest is a type of ensemble learning method classified under homogeneous base learners, where all the base learners are decision trees, as implied by the name, giving the random forest a simpler structure compared to similar methods (Savargiv et al., 2021). LR is a mathematical approach utilized to evaluate and measure the relationship between the variables under study (Maulud & Abdulazeez, 2020). These algorithms were chosen for their capability to process historical admission data to

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

identify significant patterns and trends. The goal of this study is to assess the effectiveness of these two algorithms in predicting the number of new students who will be admitted to Nasional University. By utilizing historical admission data from the past several years, this research will compare the predictive accuracy of both methods. RF, known for being an ensemble method that combines the results from multiple decision trees, is expected to provide more stable and realistic predictions. Meanwhile, LR, which attempts to fit the best straight line through the data, will offer a different perspective in this forecasting analysis. The findings of this study are anticipated to offer valuable insights into more effective methods for predicting new student admissions and to provide a basis for the university to improve its decision-making in planning and resource management.

LITERATURE REVIEW

In Indonesia, the count the number of higher education graduates is determined by their completion status. However, many undergraduate students have reached the maximum study duration without having their completion status determined. period without having their completion status confirmed. This issue complicates the calculation of the actual number of graduates, which is crucial for higher education evaluation and policy-making. To address this, It is essential to predict the completion status of these students, whose outcomes are currently unknown. (Nirmala et al., 2022). The linear regression graph closely surrounds this line, demonstrating the highest accuracy. Conversely, the decision tree model shows the lowest accuracy, as it is more dispersed with numerous outliers, indicating that this model yields less precise results (Narnaware, 2023).

This study predicts the final grades for mathematics and Portuguese language courses using datasets with 369 and 649 records, respectively. The results show that both the SVM and RF algorithms can reach up to 93% accuracy in binary classification for these datasets. In regression tasks, the RF algorithm yields the lowest RMSE of 1.13 (Alamri et al., 2020). The study introduced and compared the RF and SVM algorithms. The findings showed that six features were sufficient to accurately predict the yield, with an R^2 greater than 0.85 and an RMSE below 5.7%, while only three features were required to predict the compositions. Additionally, the study derived significant insights from the models, revealing that pyrolysis conditions had a greater impact on the yield (55%), CO₂ (73%), and H₂ (81%) than biomass characteristics, while the reverse was true for CO (12%) and CH₄ (38%) (Tang et al., 2021). Models such as SVR, RF, CART, and beta regression were employed on the training data to predict success rates for the test data.

The beta regression and RF models achieved the highest R^2 values, 0.5909 and 0.6638, respectively. Among the machine learning algorithms, the RF model had the lowest MSE (0.0038), making it the most effective for predicting success rates. While the beta regression model performed well in predicting success rates for Çanakkale province, the RF model demonstrated greater accuracy for Ankara. The key variables identified in the beta regression model, such as GDP, divorce rate, and book reading rate, were also significant in the RF model (Koc & Akin, 2022). One of the major global issues today is the high rate of road traffic accidents (RTAs), particularly in low- and middle-income countries. In Rwanda, RTAs are a leading cause of death. Accurate prediction of RTAs is crucial for effective transportation management and the development of Intelligent Transportation Systems. This study used two modeling techniques, RF and SVM, to forecast short-term road accidents, with data sourced from the police for model evaluation. The models' performance was compared using accuracy measures such as lower error indices (MAE, MSE, RMSE) and a higher coefficient of determination (R^2). The RF model outperformed the SVM model, achieving a higher R^2 value of 0.91 compared to SVM's 0.86. Machine learning methods like these show promise for predicting road accidents, helping to enhance safety and formulate regulations to prevent future incidents. An effective prediction method would assist policymakers and healthcare providers in improving strategies for accident management (Gatera et al., 2023).

The findings underscore the outstanding predictive performance of the Random Forest model across all chosen fintech companies. The low MSE values for Niyogin Fintech Ltd., One 97 Communications Ltd. (Paytm), and PB Fintech Ltd. (Policybazaar) indicate that the model's predictions closely match the actual stock prices, highlighting its effectiveness in capturing stock price trends. Additionally, the RMSE values of 0.07736, 0.3875, and 0.3855 for these companies further demonstrate the model's precision, indicating minimal average deviation from actual prices. Generally, lower RMSE values are preferred in forecasting models, emphasizing the model's reliability in generating accurate predictions (Meher et al., 2024).

The effectiveness of the hybrid model is assessed using five statistical metrics: coefficient of determination (R^2), RMSE, MAE, RAE, and TIC. The results demonstrate that the hybrid model delivers high predictive accuracy, with the Arithmetic Optimization Algorithm significantly enhancing the RF performance. Key statistical results include $R^2 = 0.977201$, RMSE = 0.1179, MAE = 0.0573, RAE = 0.0930, and TIC = 0.0187. Additionally, the Ant Lion Optimizer shows excellent convergence, achieving a TIC value of 0.014986 after 101 iterations (Xu et al., 2024).

The study also presents a hybrid approach that combines the RF algorithm combined with a genetic algorithm for multi-class classification. This combination is intended to enhance accuracy and reliability compared to current methods. A case study demonstrates that the suggested method predicts student performance with an average

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

accuracy of 93.11%, marking an improvement of at least 2.25% over baseline techniques. The effectiveness of this approach is further supported by analysis using various evaluation metrics, including Accuracy, Precision, Recall, and F-Measure (Chen & Liu, 2024).

After experimental validation, the RF model has shown exceptional performance in detecting financial fraud, providing strong evidence for its use in this area. Consequently, the financial detection model, which utilizes artificial intelligence technology along with the random forest algorithm, is of significant theoretical and practical importance, effectively reducing corporate financial risks (Du & Zhai, 2024).

METHOD

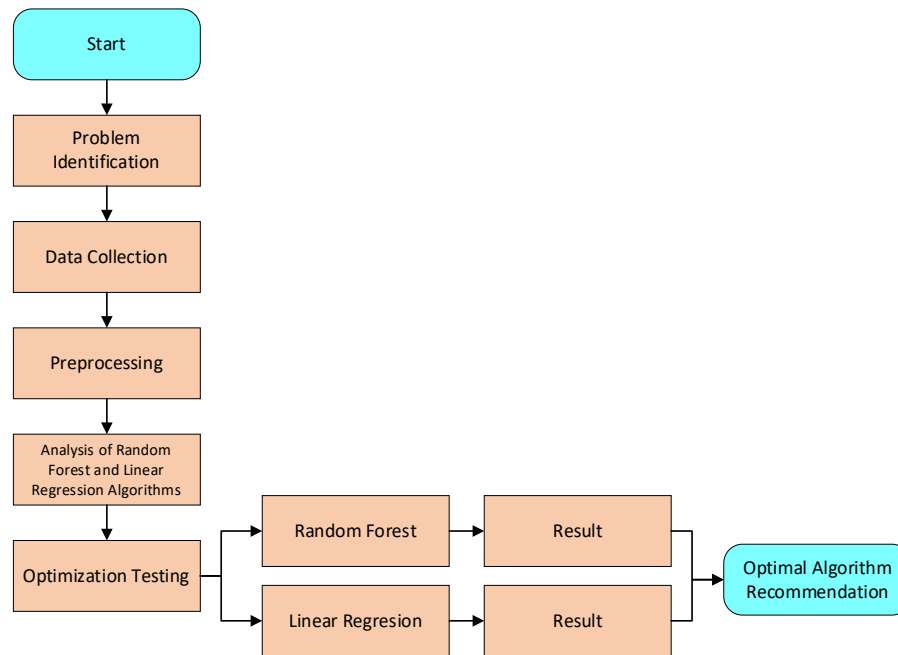


Figure 1. Purpose Research Method

The stages of Research Development include several key steps. First, Problem Identification: Define the problem to be solved, which is predicting new student admissions, and establish the objectives and scope of the analysis. Second, Data Collection: Gather 10 years of PMB (New Student Admissions) data from reliable sources such as university websites or student admission information systems, ensuring the data includes information such as the number of applicants, the number of students admitted, registration fees, trends in program interest, and accreditation information over the past 10 years. Third, Data Preprocessing: Remove missing values and outliers from the data, apply transformations such as normalization or standardization to ensure all variables are on a consistent scale, and split the data into training and testing sets for model training and evaluation. Fourth, Algorithm Analysis Using Two Prediction Models: Initialize and train the RF and LR models using the preprocessed data, and evaluate each model's performance using appropriate metrics. Train both models with the training set and optimal parameters. Fifth, Model Evaluation: Assess the performance of both models using accuracy metrics such as MAE, MSE/RMSE, MAPE, and MAD, and compare their performance. Finally, Result Analysis: Analyze the evaluation results to determine which model performs better in predicting admissions, visualize the prediction outcomes for a clearer view, discuss the implications of the research findings, and provide recommendations for applying the prediction model in practice. This research aims to make a significant contribution to higher education institutions by improving the accuracy of PMB admission predictions, thereby helping universities make more informed and effective strategic decisions regarding new student admissions.

DATA SET

The extensive dataset utilized in this study comprises form purchase transactions from 2017 to 2023, totaling 37,078 data entries. Parameters used: Price, representing the average prevailing price; Accreditation, describing the accreditation level; and Favorite, indicating the popularity or important level of a program.

*name of corresponding author



Years	2015	2016	2017	2018	2019	2020	2021	2022	2023
Buyers	3271	4082	5028	6771	6040	4792	5286	5547	3603
Price	304205,14	308441,94	309018,50	309414,86	313473,92	312739,98	317508,51	316017,67	322901,39
Accreditation	4	4	4	5	6	7	8	9	10
Favorite	8	10	10	12	12	14	14	16	16

Figure 2. The dataset used

Data Processing

In this phase, data processing involves managing missing values by eliminating empty or nonexistent entries. Data normalization follows, which standardizes values within every record to account for variations in range across different datasets. Finally, the data is divided into training and testing sets. The metric used compares the last 5 years of actual data with 5 years of predictions, providing a more precise comparison. Additionally, this approach includes adding the prediction results of each year to the dataset and retraining the model with the updated dataset for predicting each subsequent year.

Modeling

Forecasts are necessary for strategic planners only when the future is uncertain . For example, forecasting the sunrise is unnecessary because it is a certain event. Similarly, when outcomes can be controlled, such as predicting the temperature in your home, forecasting is not needed. However, many decisions involve uncertainty, and in these cases, formal forecasting methods (referred to simply as forecasting) can be highly beneficial (Armstrong & Franke, 2001). MAE, MSE/RMSE, MAPE, and MAD, which measure the average absolute difference, squared difference, percentage difference, and median absolute difference between predicted and actual values, respectively.

MAE : $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ (1)

MSE : $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (2)

RMSE : $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ (3)

MAD : $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ (4)

MAPE : $=(1/N) * \sum[(|Prediction- Real| / |Real|) * 100\%]$ (5)

RESULT

The experimental modeling in this study was carried out on a laptop featuring an 11th Gen Intel(R) Core(TM) i3-1115G4 processor running at 3.00GHz, 8 GB of RAM, and a 500 GB HDD, using Windows 11 Home, version 23H2. The LR, RF, and Prophet algorithms were implemented for data processing using Python, with each model undergoing three tests using different training-to-testing data ratios: 70:30, 50:50, and 30:70. The dataset, consisting of buyer data from 2015 to 2023, was split using a validation method such as the 80:20 Holdout Method, where 80% of the data was used for training and 20% for testing. The model's performance was evaluated based on prediction results and accuracy metrics, including MAE, MSE, RMSE, MAD, and MAPE.

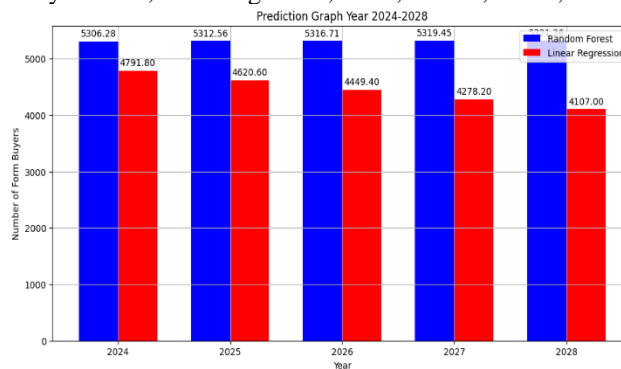


Figure 3. Prediction Graph

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Analysis of the prediction results for new student admissions using the RF and LR algorithms for the next 5 years based on the given data:

```
Random Forest predictions for the next 5 years: [5306.28,
5312.5648000000008, 5316.7127680000002, 5319.4504268800008,
5321.2572817408145]
Linear Regression Prediction for the next 5 years: [4791.799999999988,
4620.599999999977, 4449.399999999965, 4278.199999999953,
4106.999999999942]
Matrix Random Forestt: (508.2122226278376, 514729.77701226796,
717.4467067401368, 8.221103426106708, 274.71276800000214)
Linear Regression Metrics: (1210.000000000035, 1670489.920000076,
1292.4743401708508, 20.82817061934828, 1253.0000000000582)
The dataset used for Random Forest predictions is annually:
[5178, 6733, 5978, 5042, 5184, 5360]
[5360.0, 5978.0, 5184.0, 5042.0, 5306.28]
[5360.0, 5978.0, 5184.0, 5042.0, 5312.5648000000008]
[5360.0, 5978.0, 5184.0, 5042.0, 5316.7127680000002]
[5360.0, 5978.0, 5184.0, 5042.0, 5319.4504268800008]
[5360.0, 5978.0, 5184.0, 5042.0, 5321.2572817408145]
The dataset used for Linear Regression prediction is for each year.:
[5178, 6733, 5978, 5042, 5184, 5360]
[5360.0, 5978.0, 5184.0, 5042.0, 4791.799999999988]
[5360.0, 5978.0, 5184.0, 5042.0, 4620.599999999977]
[5360.0, 5978.0, 5184.0, 5042.0, 4449.399999999965]
[5360.0, 5978.0, 5184.0, 5042.0, 4278.199999999953]
[5360.0, 5978.0, 5184.0, 5042.0, 4106.999999999942]
```

Figure 4. Overall results of RF and LR

RF for the next 5 years:

[5306.28, 5312.5648000000008, 5316.7127680000002, 5319.4504268800008, 5321.2572817408145]

LR predictions for the next 5 years:

[4791.799999999988, 4620.599999999977, 4449.399999999965, 4278.199999999953, 4106.999999999942]

RF Predictions

Years 2024 : 5306.28
Years 2025 : 5312.56
Years 2026 : 5316.71
Years 2027 : 5319.45
Years 2028 : 5321.26

LR Predictions

Years 2024 : 4791.80
Years 2025 : 4620.60
Years 2026 : 4449.40
Years 2027 : 4278.20
Years 2028 : 4107.00

Based on the predictions, the RF algorithm provides a higher and more stable forecast for new student admissions compared to LR. RF predicts an increase in the number of new students from 2024 to 2028, starting at 5306.28 in 2024 and rising slightly each year. In contrast, LR shows a decline in new student admissions each year, starting at 4791.80 in 2024 and continuing to decrease to 4107.00 by 2028. This difference indicates that RF forecasts a more stable growth, while LR predicts a consistent decrease in new student admissions.

ACCURACY ANALYSIS

In this analysis, the metrics and dataset used for predictions in both RF and LR models.

Metrik RF :

(508.2122226278376, 514729.77701226796, 717.4467067401368, 8.221103426106708, 274.71276800000214)

Metrik LR :

(1210.000000000035, 1670489.920000076, 1292.4743401708508, 20.82817061934828, 1253.0000000000582)

ANALYSIS OF METRIC RESULTS

In the graph below, the RF model outperforms the LR model in terms of accuracy and prediction consistency. The RF metrics—MAE of 508.21, MSE of 514,729.78, RMSE of 717.45, MAPE of 8.22%, and MAD of 274.71—indicate lower prediction errors and better stability. In contrast, the LR model shows higher error metrics, with an MAE of 1,210.00, MSE of 1,670,489.92, RMSE of 1,292.47, MAPE of 20.83%, and MAD of 1,253.00, reflecting greater prediction errors and lower consistency. Overall, RF is the more reliable model for forecasting.

*name of corresponding author



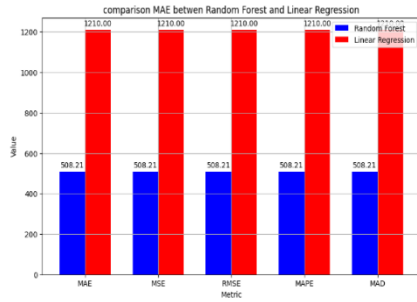


Figure 5. Comparison MAE between RF and

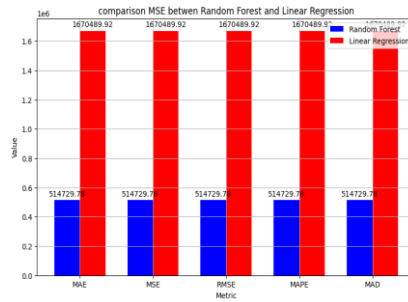


Figure 6. Comparison MSE between RF and

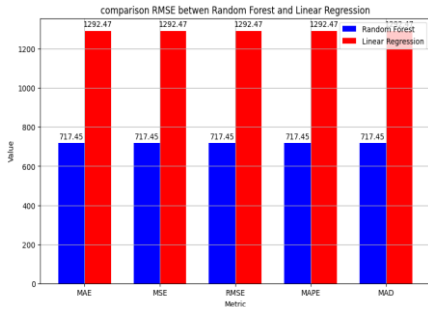


Figure 7. Comparison RMSE between RF

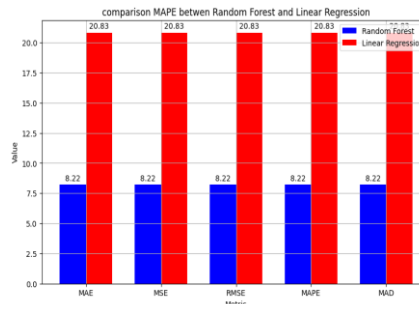


Figure 8. Comparison MAPE between RF

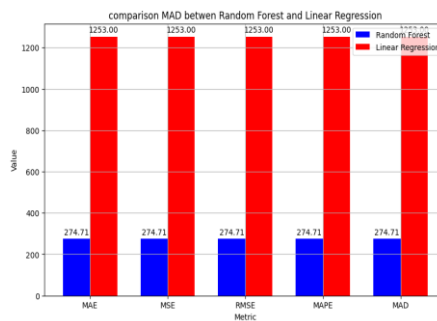


Figure 9. Comparison MAD between RF and RL

TESTING RATIO

80% training data, 20% validation data training, 7 items out of a total of 9 data points (9 years dataset) Using data indices: Model Performance Analysis

```
# Creating a DataFrame
data = pd.DataFrame({'Year': years, 'Students': students})
# Preparing data for training
X = data['Year'].values.reshape(-1, 1)
y = data['Students'].values
# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Model Random Forest
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
# Model Linear Regression
lr_model = LinearRegression()
lr_model.fit(X_train, y_train)
```

Figure 10. Training data and testing data

ANALYSIS OF THE DATASET USED

RF Dataset:

- 5178, 6733, 5978, 5042, 5184, 5360 for 2024.
- 5360.0, 5978.0, 5184.0, 5042.0, 5306.28 for 2025.
- 5360.0, 5978.0, 5184.0, 5042.0, 5312.564800000008 for 2026.
- 5360.0, 5978.0, 5184.0, 5042.0, 5316.712768000002 for 2027.
- 5360.0, 5978.0, 5184.0, 5042.0, 5319.450426880008 for 2028.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

6. 5360.0, 5978.0, 5184.0, 5042.0, 5321.2572817408145 2029 and onwards.

LR Dataset:

1. 5178, 6733, 5978, 5042, 5184, 5360 for 2024.
2. 5360.0, 5978.0, 5184.0, 5042.0, 4791.799999999988 for 2025.
3. 5360.0, 5978.0, 5184.0, 5042.0, 4620.599999999977 for 2026.
4. 5360.0, 5978.0, 5184.0, 5042.0, 4449.399999999965 for 2027.
5. 5360.0, 5978.0, 5184.0, 5042.0, 4278.199999999953 for 2028.
6. 5360.0, 5978.0, 5184.0, 5042.0, 4106.999999999942 for 2029 and onwards.

The datasets used for predicting new student admissions with RF and LR models reveal distinct trends. For RF, the dataset shows a consistent set of values for the years 2024 through 2028, with incremental increases in predicted admissions from 5306.28 in 2024 to 5321.26 in 2029 and beyond, indicating a stable and slightly growing trend. In contrast, the LR dataset indicates a declining trend, with predictions starting at 4791.80 in 2024 and decreasing to 4107.00 by 2029. This contrast highlights RF capacity to forecast a gradual increase, while LR projects a steady reduction in new student admissions.

ANALYSIS OF ACCURACY METRIC COMPARISON

The performance metrics for the RF and LR models reveal significant differences in their predictive accuracy. For RF, the metrics are as follows: MAE of 508.21, MSE of 514,729.78, RMSE of 717.45, MAPE of 8.22%, and MAD of 274.71. In contrast, the LR metrics show a MAE of 1210.00, MSE of 1,670,489.92, RMSE of 1292.47, MAPE of 20.83%, and MAD of 1253.00. These results indicate that RF generally provides more accurate and reliable predictions with lower error values compared to LR, which has higher errors and less precision in its forecasts.

```
# Function to calculate accuracy value
def calculate_metrics(y_true, y_pred):
    mae = mean_absolute_error(y_true, y_pred)
    mse = mean_squared_error(y_true, y_pred)
    rmse = np.sqrt(mse)
    mape = np.mean(np.abs((y_true - y_pred) / y_true)) * 100
    mad = np.median(np.abs(y_true - y_pred))
    return mae, mse, rmse, mape, mad

# New student admissions predictions for the next 5 years
future_years = np.arange(2024, 2029)
rf_future_predictions = []
lr_future_predictions = []
rf_datasets = [students.tolist()]
lr_datasets = [students.tolist()]

# Predictions using additional data each year
for year in future_years:
    # Random Forest
    rf_pred = rf_model.predict(np.array([[year]])) [0]
    rf_future_predictions.append(rf_pred)
    new_X_train = np.append(X_train, [[year]], axis=0)
    new_y_train = np.append(y_train, rf_pred)
    rf_model.fit(new_X_train, new_y_train)
    rf_datasets.append(new_y_train.tolist())
    # Linear Regression
    lr_pred = lr_model.predict(np.array([[year]])) [0]
    lr_future_predictions.append(lr_pred)
    new_X_train_lr = np.append(X_train, [[year]], axis=0)
    new_y_train_lr = np.append(y_train, lr_pred)
    lr_model.fit(new_X_train_lr, new_y_train_lr)
    lr_datasets.append(new_y_train_lr.tolist())
```

Figure 11. Code snippet for predictions

```
# Calculate metrics for future predictions by comparing the past years with the next year's
# predictions.
rf_metrics = calculate_metrics(y[-5:], np.array(rf_future_predictions))
lr_metrics = calculate_metrics(y[-5:], np.array(lr_future_predictions))
# Showing results
print("Random Forest predictions for the next 5 years:", rf_future_predictions)
print("Linear Regression Prediction for the next 5 years:", lr_future_predictions)
print("Matrix Random Forestt:", rf_metrics)
print("Linear Regression Metrics:", lr_metrics)
print("The dataset used for Random Forest predictions is annually:")
for dataset in rf_datasets:
    print(dataset)
```

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

```
print("The dataset used for Linear Regression prediction is for each year.:")  
for dataset in lr_datasets:  
    print(dataset)
```

Figure 12. Code Analysis of accuracy metrics of MAE, MSE, RMSE, MAD, and MAPE

CONCLUSION

The RF predictions for the next five years show a stable and slightly increasing trend, with projections starting at 5,306.28 forms in 2024 and rising to 5,321.26 by 2028. In contrast, LR predicts a consistent decline, starting from 4,791.80 in 2024 and dropping to 4,106.99 by 2028. In terms of performance metrics, RF demonstrates better results compared to LR, with a MAE of 508.21, significantly lower than 1,210.00, an MSE of 514,729.78 compared to 1,670,489.92, and an RMSE of 717.45 versus 1,292.47. The MAPE for RF is 8.22%, lower than the 20.83% for LR, and the MAD for RF is 274.71, showing a more consistent error distribution compared to 1,253.00 for LR. The dataset for Random Forest each year utilizes updated historical data with the latest predictions, while LR shows a consistent declining trend in the dataset. Overall, the Random Forest model provides more accurate and stable predictions and is more effective in handling errors and outliers, making it more recommended for forecasting student numbers.

REFERENCES

- Alamri, L. H., Almuslim, R. S., Alotibi, M. S., Alkadi, D. K., Ullah Khan, I., & Aslam, N. (2020). Predicting Student Academic Performance using Support Vector Machine and Random Forest. *ACM International Conference Proceeding Series, Part F16898*(July), 100–107. <https://doi.org/10.1145/3446590.3446607>
- Armstrong, J., & Franke, G. (2001). Principles of forecasting. In *Journal of Marketing Research*. <http://elibrary.ru/item.asp?id=4643952%5Cnpapers2://publication/uuid/2EF07596-D9F9-463E-8773-080AE5DD8ADC>
- Chen, M., & Liu, Z. (2024). Heliyon Predicting performance of students by optimizing tree components of random forest using genetic algorithm. *Heliyon*, *10*(12), e32570. <https://doi.org/10.1016/j.heliyon.2024.e32570>
- Du, Q., & Zhai, J. (2024). *Measurement : Sensors Application of artificial intelligence Sensors based on random forest algorithm in financial recognition models*. *33*(June).
- Gatera, A., Kuradusenge, M., Bajpai, G., Mikeka, C., & Shrivastava, S. (2023). Comparison of random forest and support vector machine regression models for forecasting road accidents. *Scientific African*, *21*, e01739. <https://doi.org/10.1016/j.sciaf.2023.e01739>
- KOC, T., & AKIN, P. (2022). Estimation of High School Entrance Examination Success Rates Using Machine Learning and Beta Regression Models. *Journal of Intelligent Systems: Theory and Applications*, *5*(1), 9–15. <https://doi.org/10.38016/jista.922663>
- Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(2), 140–147. <https://doi.org/10.38094/jastt1457>
- Meher, B. K., Singh, M., Birau, R., & Anand, A. (2024). Forecasting stock prices of fintech companies of India using random forest with high-frequency data. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1), 100180. <https://doi.org/10.1016/j.joitmc.2023.100180>
- Narnaware, M. (2023). College Admission Prediction using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *11*(11), 520–523. <https://doi.org/10.22214/ijraset.2023.56539>
- Nirmala, I., Wijayanto, H., & Notodiputro, K. A. (2022). Prediction of Undergraduate Student's Study Completion Status Using MissForest Imputation in Random Forest and XGBoost Models. *ComTech: Computer, Mathematics and Engineering Applications*, *13*(1), 53–62. <https://doi.org/10.21512/comtech.v13i1.7388>
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., ... Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Savargiv, M., Masoumi, B., & Keyvanpour, M. R. (2021). A new random forest algorithm based on learning automata. *Computational Intelligence and Neuroscience*, *2021*. <https://doi.org/10.1155/2021/5572781>
- Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wang, S., Chen, H., & Raza Naqvi, S. (2021). Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: Effects of pyrolysis conditions and biomass characteristics. *Bioresource Technology*, *339*(July), 125581. <https://doi.org/10.1016/j.biortech.2021.125581>
- Xu, W., Tu, J., Xu, N., & Liu, Z. (2024). Predicting daily heating energy consumption in residential buildings through integration of random forest model and meta-heuristic algorithms. *Energy*, *301*(May), 131726. <https://doi.org/10.1016/j.energy.2024.131726>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.