Application of Linear Regression and Random Forest Algorithms for Predicting the Human Development Index in Aceh

Mulyati[1], **Nur Aynun Siregar**[2], **Khairunnisa**[3]

[123]Computer Science Study Program, Faculty of Science, Technology and Health Sciences, Bina Bangsa Getsempena University, Banda Aceh, Aceh, 23111, Indonesia

Abstract

The Human Development Index (HDI) is an important indicator for assessing the welfare and quality of life of the population in a region. Differences in HDI indices between districts/cities reflect real variations in development, requiring accurate data-based analysis to predict HDI values. This study aims to compare the performance of Linear Regression and Random Forest algorithms in modeling and predicting HDI values based on the indicators of Expected Schooling Years (HLS), Average Schooling Years (RRLS), Life Expectancy (UHH), and Adjusted Per Capita Income (PP). The dataset used consists of district/city HDI data in Aceh from 2020 to 2025. Based on the model evaluation results, Linear Regression shows excellent performance with an R^2 value of 0.99 and a low prediction error rate as seen from the RMSE value of 0.32 and MAE of 0.23. Meanwhile, Random Forest also performed very well with an R^2 value of 0.98, but the RMSE value of 0.55 and MAE of 0.43 were higher than those of Linear Regression. Overall, both models were able to explain the variation in HDI data very well, but Linear Regression provided more accurate prediction results, making it more suitable for use in the context of HDI prediction in this research..

Keywords: *Human Development Index (HDI); K-fold Cross Validation; Linear Regression; Prediction; Random Forest*

1. Introduction

One of the benchmarks for a country's success in economic growth is the quality of its human resources (HR) [1]. The quality of human resources can be measured based on the Human Development Index (HDI) [2]. HDI measures three main dimensions, namely health, education, and a decent standard of living (www.bps.go.id). HDI is an important parameter in evaluating development results and formulating national policies. The National Human Development Index (HDI) in 2025 increased by 0.88 points compared to the previous year, this shows that human development policies are being implemented in a focused, consistent, and sustainable manner [3].

In Aceh province, the Human Development Index (HDI) has increased annually. During the 2020–2025 period, the HDI grew by an average of 0.79 percent per year. In 2025, it reached 76.23, above the national average of 75.90. However, compared to other regions in Indonesia, Aceh still ranks 13th nationally [4]. The inequality of the Human Development Index between districts/cities in Aceh reflects differences in human development outcomes. This is caused by a lack of coordination of development across regions in Aceh [5]. The differences in HDI indices between districts/cities reflect real variations in development, requiring accurate data-based analysis to predict future HDI developments.

One method that can be applied to predict HDI growth is by utilizing data-based predictive analysis techniques. Two frequently used approaches are linear regression and random forest. Linear regression is a basic statistical technique used to model the linear relationship between the dependent variable (HDI) and the independent variables (factors influencing the HDI) [6]. Although simple, linear regression has limitations when dealing with non-linear relationships and data containing outliers [7]. Random forest is an ensemble-based machine learning technique that can handle more complex data, including non-linear interactions between variables, and produces more precise output that is resistant to overfitting [8]. Research by Liaw and Wiener (2002) [9] shows that random forests perform better in predicting socio-economic outcomes than conventional methods, especially when the data used contains many variables and

*Corresponding author. *E-mail address:* mulyamulyati3@gmail.com
Received: 20 December 2025, Accepted: 30 January 2026 and available online 31 January 2026
DOI: https://doi.org/10.33751/komputasi.v19i2.5260

noise. A comparison between linear regression and random forest can provide valuable insights into the most effective method for predicting future HDI values. Several previous studies have shown the importance of selecting the right model for HDI analysis and prediction. Fitriyah Z, et al. (2021) [10] examined the factors that significantly influence the Human Development Index (HDI) in West Java and Banten using linear regression analysis. Wahyudi et al. (2023) [11] conducted a study using linear regression analysis to determine the factors that statistically significantly influence the HDI level in East Kalimantan. Riza N, et al. (2024) [12] predicted the Human Development Index (HDI) in West Java Province using linear regression. Puspa F G, et al. (2025)[13] also analyzed the factors influencing the HDI using linear regression in West Papua Province. Haliza et al. (2025) [14] predicted the HDI in West Sumatra using linear regression. Meanwhile, Arisandi, A & Syarifuddin, S (2023)[15] also applied the Random Forest algorithm to predict the Human Development Index (HDI) in Eastern Indonesia. Therefore, this study aims to compare the performance of linear regression and random forest in predicting the HDI of districts/cities in Aceh Province for the period 2020-2025 and to determine the method that is most suitable for the characteristics of the data and multi-year trends. This study also divides the training data and test data using a time-based split approach, which is the separation of data based on time, so that the model learns from historical patterns before predicting the next period.

2. Methods

The research stages include dataset collection, data preprocessing, data analysis using linear regression and random forest, and model evaluation. Figure 1 shows the research stages carried out.

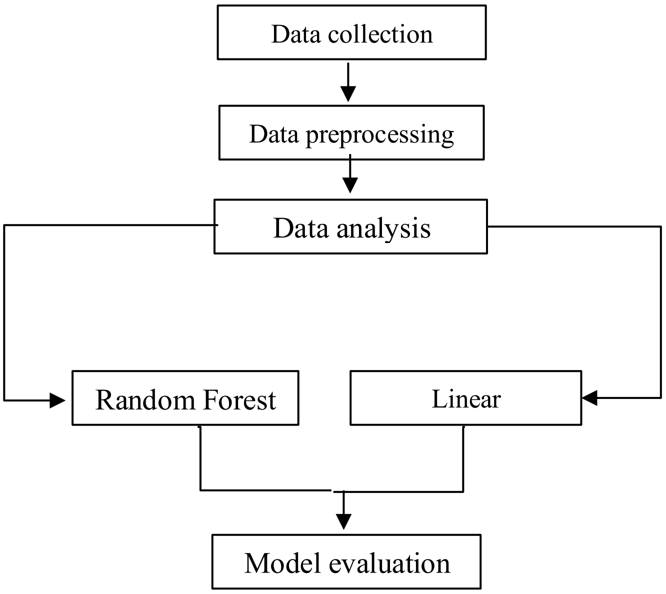


Figure 1. Research Stages

2.1. Data Collection

The data used in this study is secondary data, namely the Human Development Index (HDI) for Aceh Province, obtained from the BPS website www.bps.go.id. The data used ranges from 2020 to 2025, totaling 144 items. The variables used are Life Expectancy (UHH), Expected Years of Schooling (HLS), Average Years of Schooling (RRLS), and Adjusted Per Capita Expenditure (PP).

2.2. Data preprocessing

The data was collected and then preprocessed by checking for missing values and outliers, followed by dividing the data into training and test data based on temporal order to maintain the characteristics of the time series. Next, data normalization was performed using the Z-score method to equalize the scale between variables. The mean and standard deviation were calculated based on the training data, then used to transform the training and test data to prevent data leakage in the modeling process.

2.3. Data Analysis

The pre-processed data was then analyzed, namely by implementing it into linear regression and Random Forest. Regression analysis is a statistical analysis method that studies dependent variables that are influenced by one or more independent variables [16] with the aim of measuring the average value of the dependent variable based on the known value of the independent variable. If it involves one independent variable, it is called simple linear regression, while if it involves more than one independent variable, it is called multiple linear regression [17]. This study used multiple linear regression because it involves several independent variables. The general form of multiple linear regression is as seen in Equation (1) [18]:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \tag{1}$$

with y_i : the value of the dependent variable at the *i-th* observation, $\beta_0, \beta_1 \dots, \beta_k$: Regression coefficient parameters, x_{ij} : the value of the independent variable j in the i-th observation, and e_i : random error.

Random forest is a method of machine learning in the form of a collection of decision trees that are used in data classification and prediction, where this algorithm is included in the ensemble learning technique, namely an approach that combines prediction results from various models to obtain a better level of accuracy and stability [19]. The modeling stage uses a pipeline in Python (scikit-learn 1.3.1) to combine preprocessing and modeling. Numeric features are scaled with StandardScaler, then predicted using Linear Regression (default) and Random Forest with parameters n_estimators=200, max_features='sqrt', random_state=42). This pipeline maintains consistency in the training and prediction processes and prevents data leakage [20]. The entire data analysis process was run using Python version 3.9.21 with support from the Pandas library version 2.1.4 and NumPy version 1.26.3.4.

2.4. Model evaluation

Model validation was performed using a time-based approach, using HDI data from 2020–2024 as training data to predict HDI in 2025. This method prevents data leakage and ensures that the model learns historical trends before predicting the future. The evaluation method for each model is using the Correlation Coefficient ($\boldsymbol{R^2}$), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [21]. $\boldsymbol{R^2}$ is a measure that assesses how much variation in the target variable can be explained by the model [22]. This value is in the range of 0 to 1, where the value closer to 1 indicates a good quality model. MAE serves to assess the average absolute difference between the actual and predicted percentage values. A lower MAE value indicates that the model has a better ability to predict values close to the actual [23]. The formula for calculating MAE can be seen in equation (2)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2}$$

Meanwhile, RMSE is a measure of the average prediction error by giving greater weight to significant errors. A smaller RMSE value indicates a more accurate model in capturing data patterns [24]. The formula for calculating the RMSE value can be seen in equation (3).

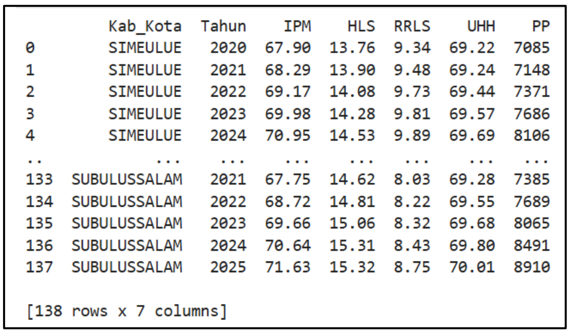
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3}$$

With yi: Actual value, $\hat{y}_i$: Predicted value, and n is the number of data

3. Result and Discussion

3.1. Data Collection

The 138 data collected through the website www.bps.go.id were then used to determine the dependent and independent variables. The dependent variable was the Human Development Index (HDI)(IPM), while the independent variables consisted of Expected Years of Schooling (HLS), Average Years of Schooling (RRLS), Life Expectancy (UHH), and Per Capita Expenditure (PP), with numeric data types for each variable. Figure 2 shows an example of the dataset used**.**

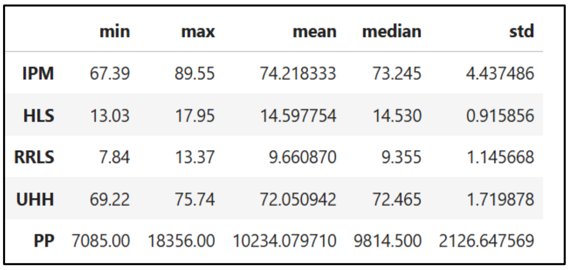


```
          Kab_Kota  Tahun    IPM    HLS  RRLS    UHH    PP
0        SIMEULUE   2020  67.90  13.76  9.34  69.22  7085
1        SIMEULUE   2021  68.29  13.90  9.48  69.24  7148
2        SIMEULUE   2022  69.17  14.08  9.73  69.44  7371
3        SIMEULUE   2023  69.98  14.28  9.81  69.57  7686
4        SIMEULUE   2024  70.95  14.53  9.89  69.69  8106
..            ...    ...    ...    ...   ...    ...   ...
133  SUBULUSSALAM   2021  67.75  14.62  8.03  69.28  7385
134  SUBULUSSALAM   2022  68.72  14.81  8.22  69.55  7689
135  SUBULUSSALAM   2023  69.66  15.06  8.32  69.68  8065
136  SUBULUSSALAM   2024  70.64  15.31  8.43  69.80  8491
137  SUBULUSSALAM   2025  71.63  15.32  8.75  70.01  8910

[138 rows x 7 columns]
```

Figure 2. Dataset used

The dataset was subjected to descriptive analysis to provide an initial overview of the characteristics of the data that will be used in the research. Figure 3 shows that the HDI variable has an average value of 74.21, with a range of values between 67.39 and 89.55. The Per Capita Expenditure (PP) variable exhibits the greatest variation compared to the other variables, as indicated by its relatively high standard deviation. This indicates differences in economic conditions between regions.



	min	max	mean	median	std
IPM	67.39	89.55	74.218333	73.245	4.437486
HLS	13.03	17.95	14.597754	14.530	0.915856
RRLS	7.84	13.37	9.660870	9.355	1.145668
UHH	69.22	75.74	72.050942	72.465	1.719878
PP	7085.00	18356.00	10234.079710	9814.500	2126.647569

Figure 3. Descriptive analysis results

A correlation analysis was conducted between the dependent and independent variables used in this study. The correlation results can be seen in Figure 4. The correlation heatmap visualization indicates that most independent variables have a positive relationship with the Human Development Index (HDI) (IPM). Per capita expenditure (PP) and average years of schooling (RRLS) variables show a relatively strong relationship compared to the other variables. This indicates that economic and educational aspects play a significant role in improving the HDI

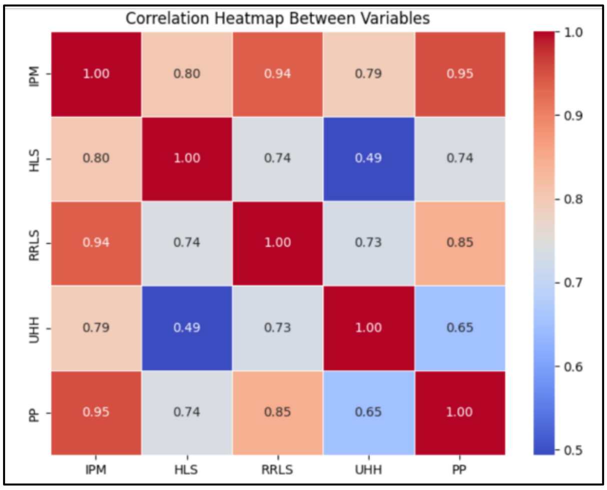
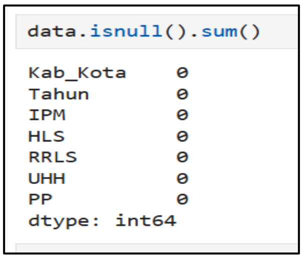


Figure 4. Heatmap Visualization Results between Variables

3.2. Data Preprocessing

Before using the data in predictions, a missing value check was performed using the Python Jupiter notebook. The results of this check showed zero values for each variable, indicating no missing or empty data, as seen in Figure 5.



```
data.isnull().sum()

Kab_Kota      0
Tahun         0
IPM           0
HLS           0
RRLS          0
UHH           0
PP            0
dtype: int64
```

Figure 5. Missing value check results

Next, outliers were checked. This check was performed to determine whether the data for each variable met reasonable limits and whether extreme values were found, so that all data could be used in the next stage. Based on the boxplot visualization results in Figure 5, the Adjusted Per Capita Expenditure (PP) variable showed the greatest data variation and several outliers above the whisker limit. This indicates inequality in expenditure levels between regions. Meanwhile, the IPM, HLS, RRLS, and UHH variables had relatively homogeneous data distributions and did not show extreme outliers. Therefore, the data were retained in subsequent analyses.

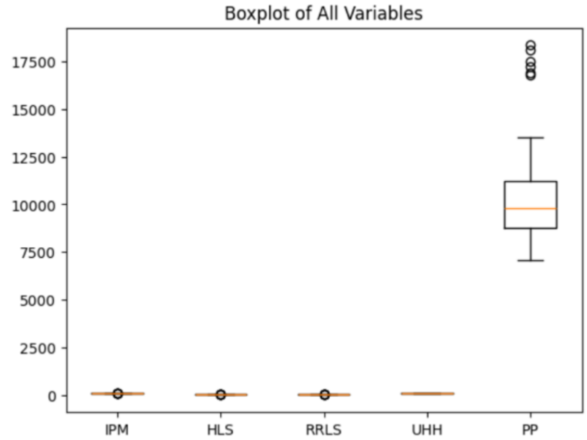
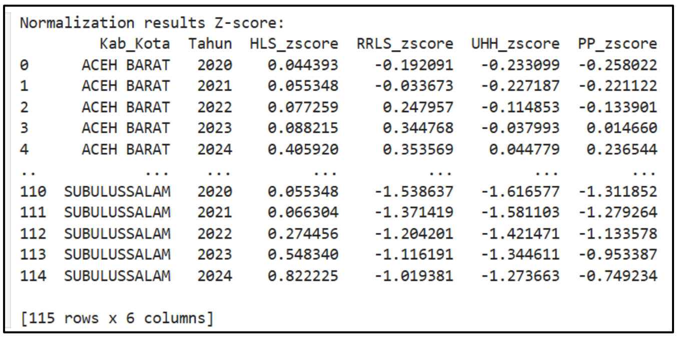


Figure 5. Boxplot visualization results

The dataset was divided into training and testing sets based on time, then the training data was normalized using z-scores so that all variables had an equal contribution. The same transformation was applied to the test data. Figure 6 shows the results of data normalization.



```
Normalization results Z-score:
          Kab_Kota  Tahun  HLS_zscore  RRLS_zscore  UHH_zscore  PP_zscore
0       ACEH BARAT   2020    0.044393    -0.192091   -0.233099  -0.258022
1       ACEH BARAT   2021    0.055348    -0.033673   -0.227187  -0.221122
2       ACEH BARAT   2022    0.077259     0.247957   -0.114853  -0.133901
3       ACEH BARAT   2023    0.088215     0.344768   -0.037993   0.014660
4       ACEH BARAT   2024    0.405920     0.353569    0.044779   0.236544
..             ...    ...         ...          ...         ...        ...
110   SUBULUSSALAM   2020    0.055348    -1.538637   -1.616577  -1.311852
111   SUBULUSSALAM   2021    0.066304    -1.371419   -1.581103  -1.279264
112   SUBULUSSALAM   2022    0.274456    -1.204201   -1.421471  -1.133578
113   SUBULUSSALAM   2023    0.548340    -1.116191   -1.344611  -0.953387
114   SUBULUSSALAM   2024    0.822225    -1.019381   -1.273663  -0.749234

[115 rows x 6 columns]
```

Figure 6. Results of data normalization with z-score

3.3. Data Analysis

3.3.1. Prediction with Linear Regression

The linear regression algorithm is used to model the relationship between HDI, which is the dependent variable, with the variables Expected Schooling Length (HLS), Average Schooling Length (RRLS), Life Expectancy (UHH), and Per Capita Expenditure (PP) as independent variables. This regression model was trained using district/city HDI data in Aceh for 2020–2024 as training data, then used to predict HDI values in 2025 as test data. This is the separation of data based on time, so that the model learns from historical patterns before predicting the next period. The prediction results compared with the actual data can be seen from the scatter plot visualization in Figure 7. Based on this figure, the HDI prediction points are close to the actual HDI values and follow the regression line. This shows that the linear regression model can predict the HDI well. The difference between the actual and predicted values is relatively small, and no significant deviations were found. The calculation of means and standard deviations on the prediction data obtained a prediction mean of 75.9 with a relatively symmetrical distribution around the regression line. The standard deviation of 4.41 indicates that the prediction results are quite consistent and do not deviate far from the actual values.

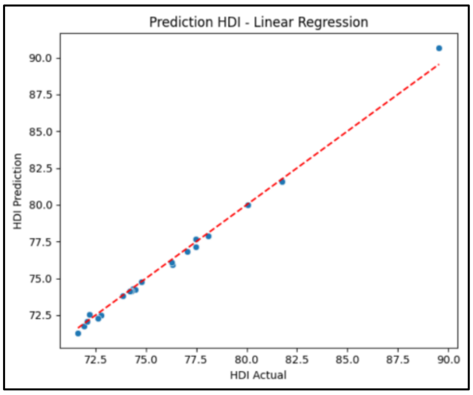


Figure 7. Scatter plot of predicted and actual values of linear regression

3.3.2. Prediction with Random Forest

Predictions with random forest were made using the same training and test data as Linear Regression. The parameters used were the number of decision trees, n_estimators=200, with the number of features considered being max_features='sqrt', and the model was set with random_state=42. The prediction results compared with the actual values in random forest can be seen in Figure 8. In the low to medium HDI range (around 72–78), the points appear relatively close to each other and consistent. However, at higher HDI values (above 80), there is a deviation from the diagonal line, indicating that the model tends to deviate slightly at extreme values. In addition, the average prediction was calculated to be 75.8 and the standard deviation of the prediction was 3.858, indicating that the Random Forest prediction results were more concentrated and stable around the average value. However, the standard deviation of the error of 0.524 indicates that the prediction error is more variable, so the consistency of this model is lower than that of linear regression.

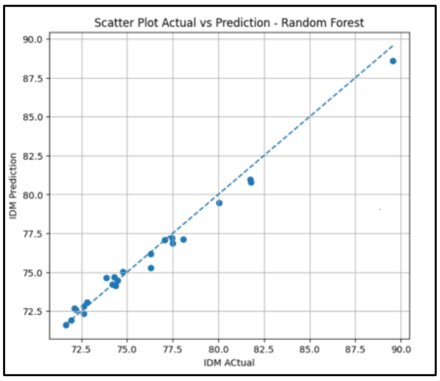
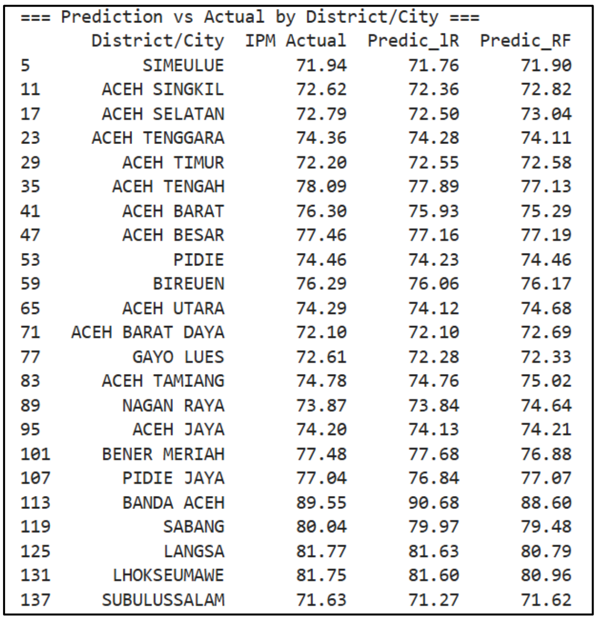


Figure 8. Scatter plot of predicted and actual values of Random Forest

3.4. Model Evaluation

Model validation was performed using training data from districts/cities for the years 2020-2024, while testing data was from 2025. The prediction results for each district/city can be seen in the following figure 9.



```
=== Prediction vs Actual by District/City ===
      District/City  IPM Actual  Predic_lR  Predic_RF
5          SIMEULUE       71.94      71.76      71.90
11     ACEH SINGKIL       72.62      72.36      72.82
17     ACEH SELATAN       72.79      72.50      73.04
23    ACEH TENGGARA       74.36      74.28      74.11
29       ACEH TIMUR       72.20      72.55      72.58
35      ACEH TENGAH       78.09      77.89      77.13
41       ACEH BARAT       76.30      75.93      75.29
47       ACEH BESAR       77.46      77.16      77.19
53            PIDIE       74.46      74.23      74.46
59          BIREUEN       76.29      76.06      76.17
65       ACEH UTARA       74.29      74.12      74.68
71  ACEH BARAT DAYA       72.10      72.10      72.69
77        GAYO LUES       72.61      72.28      72.33
83     ACEH TAMIANG       74.78      74.76      75.02
89      NAGAN RAYA        73.87      73.84      74.64
95       ACEH JAYA        74.20      74.13      74.21
101    BENER MERIAH       77.48      77.68      76.88
107      PIDIE JAYA       77.04      76.84      77.07
113     BANDA ACEH        89.55      90.68      88.60
119          SABANG       80.04      79.97      79.48
125          LANGSA       81.77      81.63      80.79
131     LHOKSEUMAWE       81.75      81.60      80.96
137    SUBULUSSALAM       71.63      71.27      71.62
```

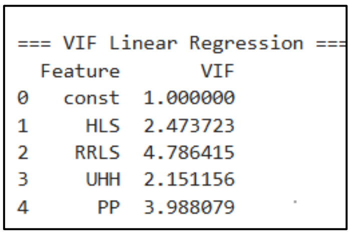
Figure 9. Prediction results and actual results in linear regression and random forest

Next, the prediction results were evaluated for each model, as shown in Table 1. Linear Regression shows slightly better performance with an R^2 value of 0.99, RMSE of 0.32, and MAE of 0.23, indicating a very low prediction error rate. Meanwhile, Random Forest also has excellent performance with an R^2 of 0.98, but its RMSE (0.55) and MAE (0.43) values are higher than those of Linear Regression. Overall, both models were able to explain the variation in HDI data very well, but Linear Regression provided more accurate prediction results, making it more suitable for use in the context of HDI prediction in this research.

Table 1. Evaluation results of each fold in linear regression and random forest

Model	R^2	RMSE	MAE
Linier Regression	0,99	0,32	0,23
Random Forest	0,98	0,55	0,43

In addition, multicollinearity analysis was also performed on the Regression Model. Figure 10 shows that the Variance Inflation Factor (VIF) value for all independent variables is below the critical threshold of <10, so it can be concluded that the regression coefficients can be interpreted reliably and the model is not too sensitive to data fluctuations [25].



```
=== VIF Linear Regression ===
  Feature       VIF
0   const  1.000000
1     HLS  2.473723
2    RRLS  4.786415
3     UHH  2.151156
4      PP  3.988079
```

Figure 10. Variance Inflation Factor (VIF)

Residual plot were also performed on the linear regression model to show the distribution of residual values against the model's predicted values. Figure 11 shows that the linear regression model is quite good because most of the prediction errors are around zero and do not form a specific pattern. However, the model still tends to predict values slightly lower than the actual data, and there is one extreme data point that deviates significantly. Overall, the linear regression model is quite good, but its performance may decline in certain observations.

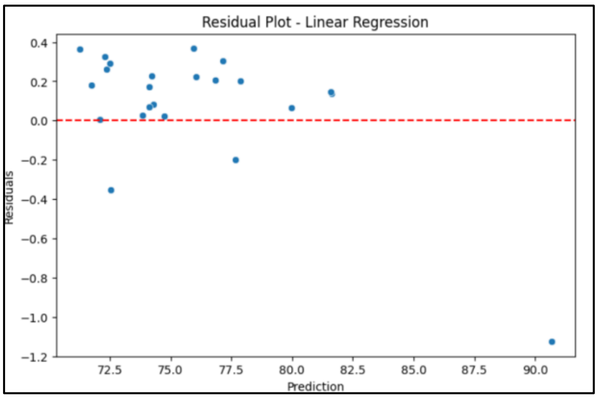


Figure 11. Residual Plot in Linear Regression

In addition, visualization of important features is also necessary to determine the extent of each independent variable's influence on the dependent variable. Based on the Permutation Feature Importance results in the Random Forest model (Figure 12), RRLS is the feature that has the most influence on model performance, followed by PP as the second most important factor. UHH makes a moderate contribution, while HLS has the least influence on model predictions. These analysis results indicate that the most effective improvement in IPM is achieved through education level (RRLS) and economic welfare and purchasing power of the community (PP).

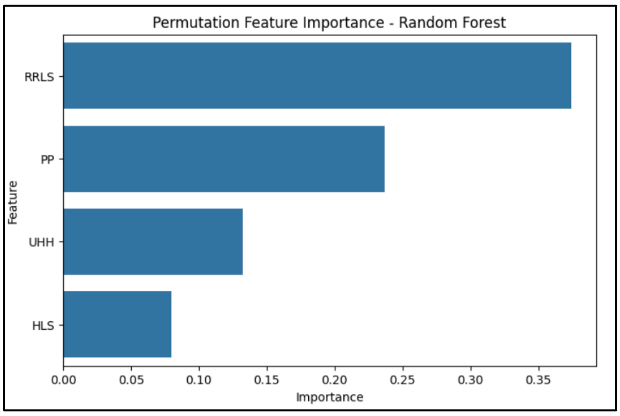


Figure 12. Permutation Feature Importance results in the Random Forest model

4. Conclusion

Based on the model evaluation results, Linear Regression showed excellent performance with an R^2 value of 0.99 and a low prediction error rate as seen from the RMSE value of 0.32 and MAE of 0.23. Meanwhile, Random Forest also performed very well with an R^2 value of 0.98, but the RMSE value of 0.55 and MAE of 0.43 were higher than those of Linear Regression. In linear regression, multicollinearity testing was also performed using the Variance Inflation Factor (VIF) for all variables used, which was < 10. This indicates that there was no multicollinearity between independent variables in the linear regression model. In addition, a residual plot test was also performed, where most of the residual values were scattered around the zero line, indicating that the linear regression model was able to predict the HDI quite well without a systematic error pattern. In the random forest, a Permutation Feature Importance test was also conducted, which showed that RRLS was the feature that most influenced model performance, followed by PP as the second most important factor. This indicates that the most effective increase in HDI is achieved based on the level of education (RRLS) and population income (PP),

Further research is recommended using a longer data period, considering that the data currently used only covers 2020-2025, so that HDI predictions can be made more accurately and capture the dynamics and differences in characteristics in each district/city. In addition, other time series predictive methods such as Gradient Boosting, XGBoost, ARIMA, or SARIMA should be used to compare the prediction performance with linear regression and Random Forest.

References

[1] A. Hamdan, A. Sarea, R. Khamis, and M. Anasweh, "Heliyon A causality analysis of the link between higher education and economic development : empirical evidence," Heliyon, vol. 6, no. 6, p. e04046, 2020, doi: 10.1016/j.heliyon.2020.e04046.

[2] L. Mukaromah, Z. Hanifatuzzahra, A. Nasrullah, T. M. Latifah, F. Ekonomi, and U. Lampung, "PENGARUH INDEKS PEMBANGUNAN MANUSIA (IPM), TINGKAT UPAH MINIMUM , DAN TINGKAT PENGANGGURAN TERHADAP PERTUMBUHAN EKONOMI INDONESIA TAHUN 2022," vol. 13, no. 02, pp. 228–245, 2023, doi: 10.37478/als.v13i2.2874.

[3] Kemenko PMK, "N," https://www.kemenkopmk.go.id/ipm-naik-bukti-nyata-sinergi-koordinasi-pembangunan-manusia.

[4] Badan Pusat Statistik (BPS), "Selama 2020–2025, IPM Provinsi Aceh rata-rata meningkat sebesar 0,79 persen per tahun," https://aceh.bps.go.id/id/pressrelease/2025/11/05/1166/selama-2020-2025--ipm-provinsi-aceh-rata-rata-meningkat-sebesar-0-79-persen-per-tahun.html.

[5] Y. Perwira, "Tantangan Pembangunan sumber daya manusia di aceh," J. Transform. Adm., vol. 07, no. 02, pp. 1369–1384, 2017.

[6] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, Second Edi. USA: Springer.

[7] A. Zulkarnain, S. W. Rizki, and H. Perdana, "Analisis regresi robust estimasi-mm dalam mengatasi pencilan pada regresi linear berganda," vol. 09, no. 1, pp. 123–128, 2020.

[8] L. E. O. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: DOI: 10.1023/A:1010950718922.

[9] A. Liaw and M. Wiener, "Classification and Regression by randomForest," R news, vol. 2, no. 3, pp. 18–22, 2002.

[10] Z. Fitriyah, S. Irsalina, E. Widodo, and U. I. Indonesia, "Analisis faktor yang berpengaruh terhadap ipm menggunakan regresi linear berganda," vol. 2, no. 3, pp. 282–291, 2021.

[11] H. Wahyudi, F. Naufal, F. W. Pongoh, and R. Amelia, "Analisis Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kalimantan Timur," vol. 2, no. 1, pp. 121–135, 2023.

[12] N. Riza, F. A. Maresti, S. S. Azzahra, and S. P. P. Ningsih, "IMPLEMENTASI REGRESI LINEAR BERGANDA PREDIKSI FAKTOR-FAKTOR INDEKS PEMBANGUNAN MANUSIA DI PROVINSI JAWA BARAT," Masal. J. Pendidik. dan Sains, vol. 5, no. 1, pp. 69–86, 2025.

[13] F. G. Puspa, E. R. Matulessy, and L. O. Muhlis, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI INDEKS PEMBANGUNAN MANUSIA DI PROVINSI PAPUA BARAT

MENGGUNAKAN REGRESI LINIER BERGANDA," J. Rev. Pendidik. dan Pengajaran, vol. 8, no. 1, pp. 422–429, 2025.

[14] P. Y. Haliza, R. Rafiza, and J. Simanullang, "Penerapan Regresi Linier Berganda Dalam Memprediksi IPM Berdasarkan Faktor Ekonomi Dan Sosial Di Sumatera Barat," vol. 2, no. June, pp. 544–554, 2025.

[15] A. Arisandi and S. Syarifuddin, "Memprediksikan Indeks Pembangunan Manusia di Wilayah Indonesia Bagian Timur Menggunakan Random Forest Classification," vol. 5, no. 1, pp. 1–6, 2023, doi: 10.31605/jomta.v5i1.2402.

[16] M. P. Prof. Dr. A. Muri Yusuf, Metode Penelitian Kuantitatif, Kualitatif, &Penelitian Gabungan, 4th ed. kencana, 2017.

[17] E. D. Kartiningrum, H. B. Notobroto, N. E. Kumarijati, and E. Yuswatiningsih, Aplikasi Regresi dan Korelasi Dalam Analisis Data Hasil Penelitian, Dr. Rifaat. Mojokerto: STikes Majapahit Mojokerto, 2022.

[18] D. C. MONTGOMERY, E. A. PECK, and G. G. VINING, INTRODUCTION TO LINEAR REGRESSION ANALYSIS, 5th ed. A JOHN WILEY & SONS, INC., PUBLICATION, 2012.

[19] C. A. Bahri and K. D. Tania, "Perbandingan Kinerja LSTM , Random Forest , dan SVR Berbasis Knowledge Discovery untuk Prediksi Harga Beras Sumatera Selatan," vol. 12, no. 5, pp. 721–732, 2025, doi: 10.30865/jurikom.v12i5.9140.

[20] R. Hidayat et al., "IMPLEMENTASI ALGORITMA RANDOM FOREST REGRESSION UNTUK MEMPREDIKSI PENJUALAN PRODUK DI SUPERMARKET," vol. 10, no. 1, pp. 101–109, 2025.

[21] E. C. Wibowo and A. D. Cahyono, "Analisis Perbandingan Algoritma Regresi Linier dengan Neural Network untuk Prediksi Harga Saham," Sist. J. Sist. Inf., vol. 14, no. 4, pp. 1879–1896, 2025.

[22] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, Applied Linear Statistical Models Fifth Edition.

[23] T. Rohana, H. Y. Novita, and E. Nurlaelasari, "KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI KAPASITAS PRODUKSI POTENSIAL AIR BERSIH DI INDONESIA," J. Teknol. Terpadu, vol. 11, no. 1, pp. 36–43, 2025.

[24] J. Halif, D. Wahiddin, I. Sanjaya, and S. Faisal, "Model Regresi Linear Berganda untuk Prediksi Tingkat Pengangguran di Provinsi Jawa Barat," pp. 324–335, 2025, doi: 10.33364/algoritma/v.22-1.2312.

[25] A. Rodiyah, D. P. Kusuma, N. B. Ramadani, R. M. Ibrahim, R. A. Huda, and A. Rahajeng, "Pengaruh Rata Rata Lama Sekolah , Garis Kemiskinan , dan Usia Harapan Hidup terhadap IPM di Provinsi Jawa Tengah Anisatu Rodiyah dikembangkan oleh United Nations Development Programme (UNDP) dan di Indonesia dihitung oleh Badan Pusat Statistik (BPS).," J. Ilm. Ekon. Dan Manaj., vol. 3, no. 5, pp. 345–357, 2025.