NUANSA INFORMATIKA
Volume 19 Number 2 July 2025

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

Predicting the Happiness Index Based on the HDI Indicator in Indonesia Using the Ensemble Learning Approach

Rofi Nafiis Zain[1], Iwan Setiawan[2], Virdiandry Putratama[3], Syafrial Fachri Pane*[4]
[1, 2, 4]Applied Informatics Engineering, Universitas Logistik dan Bisnis Internasional, Indonesia
[3]Applied Management Informatics, Universitas Logistik dan Bisnis Internasional, Indonesia
E-mail: [1]rofinafiisr@ulbi.ac.id, [2]iwan.setiawan@ulbi.ac.id, [3]virdiandry@ulbi.ac.id,
*[4]**syafrial.fachri@ulbi.ac.id**

Abstract

Machine Learning is widely utilized to analyse complex data across various research domains. In this study, we employed an ensemble learning approach comprising Random Forest Regression (RF), XGBoost Regression (XGB), Decision Tree Regression (DT), Pearson correlation analysis, and Shapley Additive Explanations (SHAP) to examine the relationship between the Human Development Index (HDI) and happiness indicators in Indonesia. The study had two main objectives. First, to investigate the correlation between HDI and happiness. The Pearson correlation analysis yielded a value of 0.88, indicating a very strong positive relationship. This was further supported by the results of Permutation Importance (PI) and SHAP analysis, which identified HDI as one of the most influential variables in predicting happiness scores. Second, we developed a predictive model using a stacking ensemble approach, integrating RF, XGB, and DT regressors. The resulting model demonstrated high predictive performance, with an R-squared value of 97.68%, a Mean Absolute Error (MAE) of 0.002900, a Mean Squared Error (MSE) of 0.000021, and a Root Mean Squared Error (RMSE) of 0.004604.

Keywords*—Prediction, Happiness, HDI, Stacking Ensemble Learning, Indonesia*

Submission: June 04, 2025 *Accepted: June 14, 2025* *Published: July 10, 2025*
DOI Article: https://doi.org/10.25134/ilkom.v19i2.410

1. INTRODUCTION

Happiness is widely recognized as a key indicator of a nation's overall well-being, particularly in evaluating social and economic progress. The Happiness Index offers a quantitative measure of individuals' life satisfaction and overall happiness, providing valuable insights for policymakers across various domains [1]. In Indonesia, the pursuit of improved quality of life and happiness is explicitly identified as a national priority, reflecting the government's recognition of its significance [2]. Nevertheless, accurately measuring and predicting happiness requires a comprehensive, data-driven approach to support the development of effective and targeted policies [3].

The Human Development Index (HDI) plays a crucial role in evaluating societal well-being, as it encompasses key dimensions such as life expectancy, education levels, and employment rates—all of which substantially influence individual and collective happiness [4]. A high HDI is generally associated with greater well-being, as it reflects the quality of a nation's healthcare and education systems—fundamental pillars of social development [5]. However, the relationship between HDI and happiness is not always straightforward, necessitating more nuanced analytical approaches to fully understand its complexity. While techniques such as Pearson correlation analysis can identify linear associations between HDI components and happiness [6], they are limited in scope. To gain deeper insights, advanced methods like Shapley Additive Explanations (SHAP) offer a more comprehensive view by quantifying the contribution of each HDI indicator in predicting happiness. These approaches enhance model interpretability and help identify the most influential factors affecting well-being [7].

Moreover, the advent of the Industrial Revolution 4.0 has heightened the significance of machine learning in analyzing complex datasets, including those related to happiness [8]. Machine learning excels at identifying intricate patterns within data, surpassing the capabilities of traditional statistical methods, and enabling the development of advanced predictive models that incorporate multiple HDI indicators. However, relying on a single model may not yield optimal predictive accuracy. Therefore, ensemble learning techniques—which combine the strengths of multiple models—are increasingly adopted to enhance forecasting

p-ISSN :1858-3911 , e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

performance [2], [7]. In this study, we implemented three core models: Random Forest Regression, XGBoost Regression, and Decision Tree Regression, each contributing to improved predictions of happiness levels in Indonesia [9].

Previous studies have leveraged machine learning to predict happiness indices with promising results. For example, research by Jannani et al. identified GNI per capita, HDI, and life expectancy as key factors strongly associated with happiness. Among the models tested, Random Forest yielded the highest accuracy at 93.66%, with a Root Mean Square Error (RMSE) of 0.05748, highlighting GNI per capita as the most influential variable [10]. Similarly, Çiftci and Yıldız demonstrated that social media addiction negatively impacts happiness and life satisfaction, with life satisfaction serving as a mediating variable. Their study employed Elastic Net Regression, achieving 84.5% accuracy and underscoring the critical role of these psychosocial factors [11]. Additionally, research by Zhiqiang Jiang applied machine learning to predict regional economic growth and industrial structure, with the K-Nearest Neighbors (KNN) algorithm attaining an accuracy of 79.46% [12].

This study makes two key contributions. First, it assesses the significance of HDI-related variables in predicting happiness using both Pearson correlation and SHAP analysis. Second, it develops a predictive model for happiness based on HDI indicators using an ensemble learning approach.

2. RELATED WORKS

Machine learning models have been extensively applied to predict the Happiness Index in various studies. Jannani et al. demonstrated that the Random Forest model achieved the highest accuracy in predicting happiness based on GNI per capita, HDI, and life expectancy, with a precision of 93.66% and a Root Mean Square Error (RMSE) of 0.05748. Their findings identified GNI per capita as the most influential factor affecting happiness predictions [10]. In another study, Çiftci and Yıldız explored the negative impact of social media addiction on happiness and life satisfaction, with life satisfaction acting as a mediating variable. Using the Elastic Net Regression model, their study reached an accuracy of 84.5%, emphasizing the importance of social determinants in understanding happiness [11]. Furthermore, Zhiqiang Jiang applied machine learning techniques to forecast regional economic growth and industrial structure. The K-Nearest Neighbors (KNN) algorithm achieved an accuracy of 79.46%, underscoring the significant role of economic variables in influencing overall well-being [12].

This study employs an ensemble learning approach to predict happiness based on Human Development Index (HDI) data by integrating Random Forest Regression (RF), XGBoost Regression (XGB), and Decision Tree Regression (DT) models. The ensemble method, combined with Permutation Importance (PI) and Shapley Additive Explanations (SHAP) analysis, outperforms individual models, demonstrating the advantages of aggregating multiple algorithms to enhance predictive accuracy. Moreover, the use of PI and SHAP not only strengthens model performance but also offers interpretable insights into the relative contributions of HDI indicators to happiness outcomes. This research contributes to the advancement of happiness prediction by presenting a robust and transparent ensemble learning framework.

3. RESEARCH DESIGN

3.1. Data Preparation

The first phase of this study involves data collection and pre-processing. The dataset was obtained from the Indonesian Central Statistics Agency and includes HDI-related indicators such as Life Expectancy, Gender Development Index, Expected Years of Schooling, Female and Male Labour Force Participation Rates, Mean Years of Schooling, and the overall Human Development Index. The data spans the period from 2018 to 2020. Following data acquisition, the pre-processing stage involves cleaning and preparing the dataset for modeling [13]. This includes data cleansing and distribution analysis [14] to identify and address missing values and to verify the distribution of each variable. Additionally, data types are checked to ensure compatibility with the analysis requirements. After these steps, all variables are standardized using the Standard Scaler method [15], which ensures that the features are on a comparable scale, thereby improving the reliability and performance of machine learning algorithms [16]. These procedures enhance data quality, improve analytical precision, and contribute to more robust and trustworthy regression model outcomes.

3.2. Descriptive and Statistical Analysis

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

The second phase involves descriptive and statistical analysis, including Exploratory Data Analysis (EDA) and correlation testing. EDA is used to visualize data distributions and identify trends, providing insights into variables that may influence the predicted outcomes. Subsequently, correlation analysis is conducted to assess linear relationships between variables, which helps determine the relative weight of each factor in predicting the Happiness Index and identify the most influential HDI indicators. Based on prior literature [10], this study employs Pearson correlation analysis, a widely used method for evaluating linear associations between continuous variables.

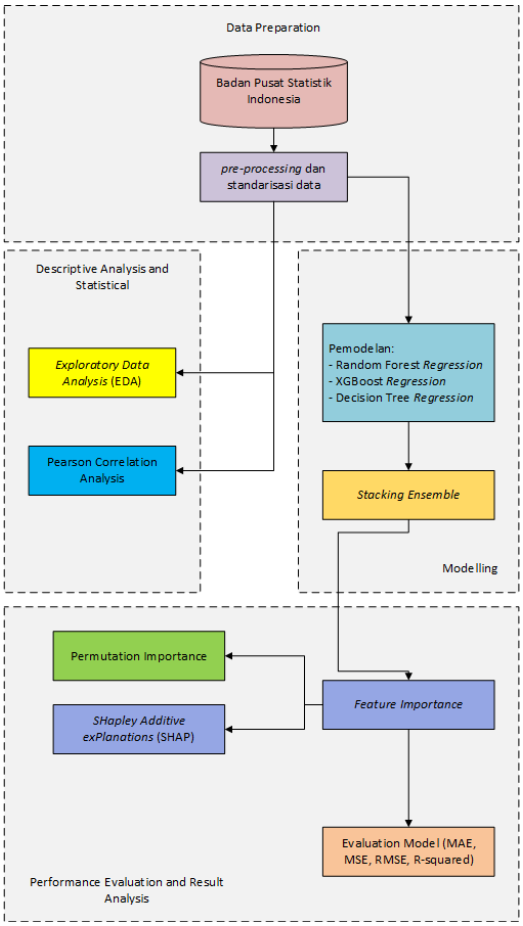

Figure 1 The proposed research methodology

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \qquad (1)$$

Equation (1) presents the Pearson correlation formula, which measures the strength and direction of the linear relationship between two variables. A correlation coefficient of r=1r = 1r=1 indicates a perfect positive correlation, where both variables increase or decrease together. In contrast, r=−1r = -1r=−1 signifies a perfect negative correlation, meaning one variable increases as the other decreases. A value of r=0r = 0r=0 suggests no linear relationship exists between the two variables.

3.3. Modelling

The modeling stage is the core component of this study and consists of two main parts: individual model training and Stacking Ensemble implementation. A range of machine learning algorithms is employed to develop a regression model for predicting happiness scores, with each model trained on standardized data and evaluated for accuracy. Specifically, this study utilizes three regression algorithms: Random Forest Regression, XGBoost Regression, and Decision Tree Regression.

To further improve predictive performance, a Stacking Ensemble technique is applied after training the individual models. This ensemble method integrates the outputs of the three base learners to construct a more accurate and robust prediction model. By capitalizing on the strengths of each algorithm, the ensemble approach aims to minimize prediction errors and deliver the highest level of accuracy in forecasting desired scores.

$$\hat{y} = g(f_1(x), f_2(x), \ldots, f_n(x)) \qquad (2)$$

Equation (2) illustrates the Stacking Ensemble formula, where y^\hat{y}y^ denotes the final predicted value produced by the ensemble model. This prediction is based on the outputs f1(x),f2(x),…,fn(x)f_1(x), f_2(x), \dots, f_n(x)f1(x),f2(x),…,fn(x) from the base learners—such as Random Forest Regression (RFR), XGBoost Regression (XGBR), and Decision Tree (DT). These outputs are then combined by a meta-learner, represented by the function ggg, which integrates the base model predictions to generate the final output.

3.4. Performance Evaluation and Result Analysis

The final phases of this study—performance evaluation and result analysis—aim to assess the accuracy and effectiveness of the developed prediction model. This process involves three key steps and utilizes several evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Squared (R^2). These metrics provide a comprehensive understanding of the model's predictive

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

performance. Specifically, MAE measures the average magnitude of prediction errors, offering insight into how accurately the model estimates happiness levels based on human development indicators.

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - x_i| \qquad (3)$$

Equation (3) explains that y_i is *predicted value*. x_i is *true value*. n is *amount of data*[17].

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

Equation (4) shows the MSE formula, averages the error square between the prediction and actual values. Smaller MSE values improve model prediction[18].

$$RMSE = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(y_i - x_i)^2} \qquad (4)$$

Equation (4) explains that y_i is *predicted value*. x_i is *true value*. n is *amount of data* [19].

$$R - Square = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \qquad (5)$$

Equation (5) defines the coefficient of determination (R^2), which quantifies the proportion of variance in the dependent variable that can be explained by the model [20]. This metric serves as an indicator of the model's predictive performance relative to a baseline model that simply predicts the mean of the target variable.

To further understand variable contributions after model training, **Permutation Importance** is applied. This technique evaluates how the model's performance deteriorates when the values of a particular variable are randomly shuffled. The resulting change in performance helps identify how crucial each feature is in predicting the target variable.

Additionally, **SHAP (Shapley Additive Explanations)** is employed to interpret model predictions. SHAP assigns a contribution value to each input feature, allowing for a detailed explanation of how each variable influences the model's output. It quantifies both positive and negative effects and captures the marginal impact of each attribute on individual predictions.

SHAP offers a significant advantage over Pearson correlation. While **Pearson correlation** provides a global, statistical measure of the linear relationship between a variable and the target, **SHAP delivers instance-level interpretability**, capturing both linear and non-linear interactions. Pearson is useful for identifying general trends, whereas SHAP reveals how specific features affect individual model predictions, offering a more granular and model-aware interpretation.

4. RESULTS AND DISCUSSION

This section presents the experimental results, including data preparation, descriptive and statistical analysis, and regression model evaluation.

The study uses HDI-related data from the Indonesian Central Statistics Agency for the years 2018–2020. Each dataset contains annual observations of variables such as Life Expectancy, Gender Development Index (GDI), Expected and Mean Years of Schooling, Male and Female Labor Force Participation Rates, and HDI. Although direct Happiness Scores for Indonesian cities are not publicly available, the study uses proxy variables to estimate happiness levels.

Data pre-processing involved normalization using **StandardScaler**, converting percentages into decimal form and handling missing values to ensure consistency. Exploratory Data Analysis (EDA), conducted in Python, revealed that cities with higher HDI tend to report higher estimated Happiness Scores, suggesting a strong association between the two.

A **Pearson Correlation Analysis** was then conducted to assess the linear relationships among the variables. The Human Development Index (HDI) showed the strongest correlation with happiness and was assigned the highest weight (0.30) in the composite score. Mean Years of Schooling and Life Expectancy followed closely with weights of 0.25 each. Female and Male Labor Force Participation Rates were each assigned weights of 0.15, as they appeared to negatively influence schooling indicators. Due to its low correlation with the other variables, the Gender Development Index (GDI) received the lowest weight.

This weighted approach ensures that variables with the strongest associations have the greatest influence on the predicted Happiness Score. The results not only support the development of an accurate prediction

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

model but also highlight the most influential factors shaping happiness in Indonesia.

Table 1 Weight of Happiness Index Calculation

No.	Parameters	Value
1.	w_1	0.25
2.	w_2	0.3
3.	w_3	0.25
4.	w_4	0.15
5.	w_5	0.1

$$H_{Index} = w_1 \cdot a + w_2 \cdot b + w_3 \cdot c + w_4 \cdot \frac{d+e}{2} + w_5 \cdot f \quad (6)$$

The Happiness Index is calculated using Equation (6), where "a" is Life Expectancy, "b" is the Human Development Index, "c" is the Mean Years of Schooling, d and e are the Female and Male Labor Force Rates, and "f" is the Gender Development Index. First Pearson correlation analysis multiplies variables by predefined weights. The average is calculated first for female and male Labor Force Rate variables before multiplying by weight. The Happiness Index is the sum of each variable's product.

The correlation analysis shows that the Human Development Index has a very strong positive correlation with the Happiness Score (correlation value of 0.88), indicating that the higher the Development Index value, the higher the Happiness Score in a city. Life Expectancy (0.76), Mean Years of Schooling (0.72), and Expected Years of Schooling (0.64) also positively affect the Happiness Score.

The Pearson correlation analysis examines Indonesian Happiness Scores and HDI indicators. The heatmap in Figure 2, and 3 depicts the degree and direction of each variable's link, helping pinpoint the primary factors affecting happiness in different cities.

Regression models employing Random Forest (RF), XGBoost (XGB), and Decision Tree (DT) predicted Happiness. Figure 4 shows R^2 values of 97.66% for RF, 96.26% for XGB, and 92.65% for DT in the regression findings of the trained models. This method predicts the Happiness index well, but Stacking Ensemble can improve it. The three models are combined using a stacking ensemble method to improve forecast accuracy.

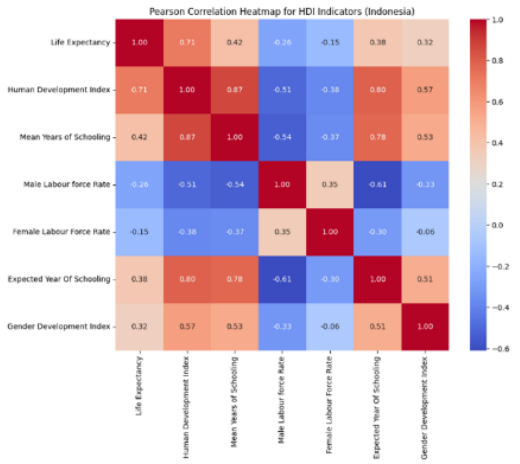

Figure 2 Pearson Correlation Before Happiness Index Calculated

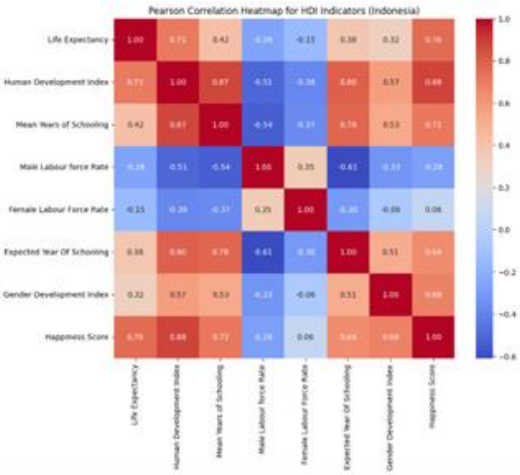

Figure 3 Pearson Correlation After Happiness Index Calculated

Stacking Ensemble combines models using Linear Regression as the main estimator. The Stacking Regressor excels with an MAE of 0.002900, MSE of 0.000021, RMSE of 0.004604, and R^2 of 97.68%, outperforming comparable models These outcomes surpass models such as Random Forest Regression (R^2 = 97.66%), XGBoost Regression (R^2 = 96.26%), and Decision Tree Regression (R^2 = 92.65%). The stacking ensemble model outperforms individual models by 0.02%.

This shows that Stacking Ensemble enhances prediction accuracy by combining regression model strengths. Less prediction error means stacking models forecast better than individual models.

Algorithm 1 : HDI Regression Analysis for Happiness Score Prediction and Feature Importance

1 Dataset HDI (Life Expectancy, Gender

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

Development Index, Expected Year of Schooling, Female and Male Labour Force Rate, Mean Years of Schooling)
Predicted happiness scores, feature importance rankings
2 *Load dataset:* ***hdi_data = read_csv("HDI_2019.csv")***
3 *Display first 10 rows of hdi_data*
4 *Handle missing values (missing values, handle types)*
5 *If hdi_data is not empty then;*
6 *Identify percentage columns: columns = list of columns containing %;*
7 *Convert percentage columns to decimals for easier computation;*
8 *Fill missing values with appropriate strategy;*
9 *Calculate composite score from features;*
10 *Construct formula for Happiness Score:*

$$H_{Index} = w_1 \cdot a + w_2 \cdot b + w_3 \cdot c + w_4 \cdot \frac{d+e}{2} + w_5 \cdot f$$

11 *Split data into features X and target Y (Happiness Score);*
12 *Train-test split: train_test_split(X, Y, train_size=0.8);*
13 *Initialize regression models:*
- *LinearRegression*
- *Decision Tree*
- *Random Forest Regressor*
- *XGB Regressor*

14 *Fit each model on training data: model.fit(X_train, Y_train);*
15 *Predict on test data: Y_pred = model.predict(X_test);*
16 *Calculate evaluation metrics: MAE, MSE, RMSE, R^2;*
17 *For each model in models do;*
18 *Perform cross-validation: cross_val_score(model, X, Y);*
19 *Calculate permutation importance;*
20 *Plot permutation importance graph;*
21 *Display cross-validation score;*
22 *Ensemble combination using estimators:*
- *('rf', RandomForestRegressor())*
- *('xgb', XGBRegressor())*
- *('dt', DecisionTree())*

23 *Fit ensemble model on training data;*
24 *Save trained model: joblib.dump(stacked_model, "stacked_model.pkl");*
25 *Return: Predicted happiness scores, feature importance ranking;*

Algorithm 1 shows the basic steps in creating a Stacking Regressor model to predict Indonesian Happiness Scores.

Feature selection, data separation, estimator definition, model training, performance evaluation, and evaluation findings are covered in this pseudocode. The goal is to summarize the model implementation workflow.

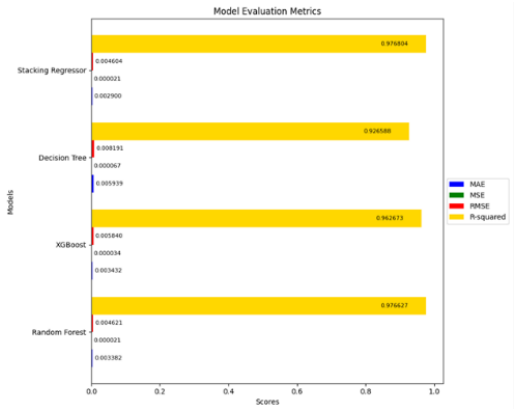

Figure 4 Evaluation Metrics from Model Result

Performance evaluation and result analysis use Feature Importance on each regression base-model to discover the most important factors predicting happiness. The data context and kind determine the best technique. This study uses Random Forest, XGBoost, and Decision Tree base-models' built-in feature importance characteristics and Permutation Importance. Combining these methods helps determine which model features most affect happiness prediction.

Permutation Importance (PI) and SHAP were utilized at this level. Decision Tree, XGBoost, Random Forest, and Stacking Ensemble Regressor Permutation Importance results are shown in Figure 5 -8. The data shows that the Human Development Index (HDI) predicts happiness the greatest, with a score of 1.2 to 1.7. With values around 0.2 to 0.3, the Female Labor Force Rate and Life Expectancy variables likewise rank second and substantially affect all models. The happiness prediction model emphasizes female labor force involvement and life expectancy. Mean Years of Schooling, Gender Development Index, Male Labor Force Rate, and Expected Year of Schooling also affect HDI and Female Labor Force Rate, albeit less so. These findings suggest that human development, female labor force involvement, and life expectancy drive happiness in Indonesia, while education plays a minor influence.

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

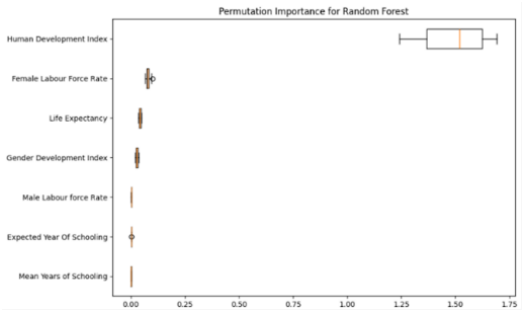

Figure 5 Permutation Importance Result RandomForest

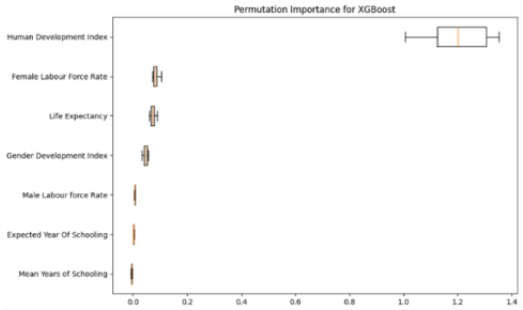

Figure 6 Permutation Importance Result XGB Regressor

The four SHAP plots (Decision Tree, XGBoost, Random Forest, and Stacking Ensemble Regressor) reveal that the Human Development Index (HDI) is the most influential variable in predicting happiness across all models (Figure 9 - 12). The Decision Tree and Random Forest models' SHAP HDI values range from -0.10 to 0.05, but XGBoost's are -0.08 to 0.06. While HDI remains dominant, the Stacking Ensemble model distributes SHAP values more evenly. Along with HDI, the Female Labor Force Rate and Life Expectancy variables contribute to happiness prediction, although less so. Mean Years of Schooling and Expected Years of Schooling have low SHAP values, indicating little impact.

Permutation Importance is easier to implement and suitable for preliminary assessments of complex models, but SHAP has better interpretative depth and visualization. SHAP clarifies prediction results, while Permutation Importance evaluates features' entire relevance.

Among the models tested in this study, the Stacking Ensemble Regressor performs best. Figure 4 shows that the Stacking Ensemble Regressor model has an MAE of 0.002900, MSE of 0.000021, RMSE of 0.004604, and R^2 of 97.68%.

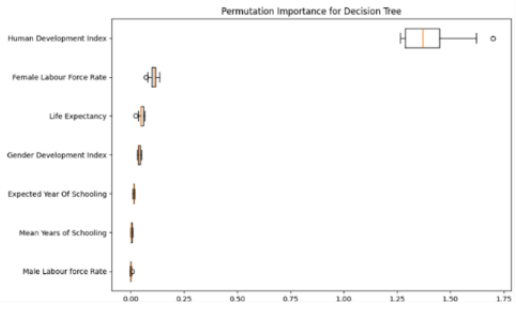

Figure 7 Permutation Importance Result Decision Tree

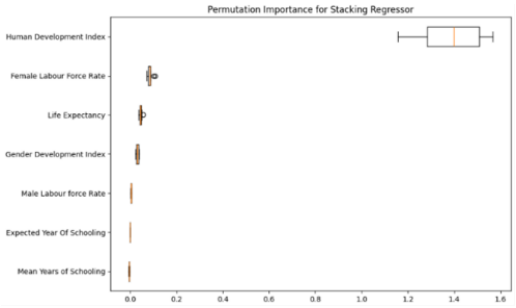

Figure 8 Permutation Importance Result Stacking Regressor

This high R-squared value shows that the model explains data variability well. Stacking Ensemble Regressor predicts happiness scores best based on indicators.

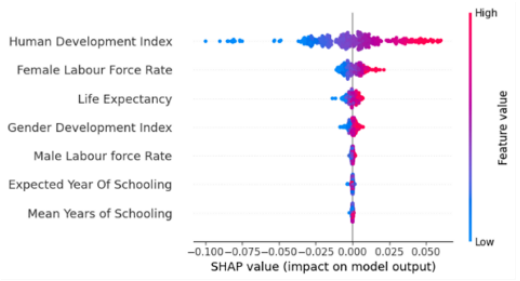

Figure 9 SHAP Result Random Forest

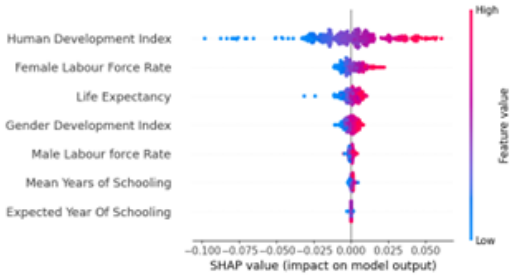

Figure 10 SHAP Result XGB Regressor

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

4. DISCUSSION

The proposed Stacking Ensemble Regressor significantly outperforms individual models in predicting the Happiness Index based on HDI-related indicators, achieving an R-Squared value of 97.68%, slightly higher than Random Forest (97.66%), XGBoost (96.26%), and Decision Tree (92.65%). This improvement is primarily attributed to the ensemble's ability to integrate the strengths of diverse base learners, thereby reducing variance and minimizing prediction errors, as reflected by the lowest MAE (0.002900) and RMSE (0.004604). To enhance model interpretability, both Permutation Importance and SHAP (SHapley Additive exPlanations) analyses were employed. These analyses consistently highlight the Human Development Index (HDI), Life Expectancy, and Female Labour Force Rate as the most impactful predictors, whereas variables such as the Gender Development Index and Expected Years of Schooling exhibit comparatively marginal effects.

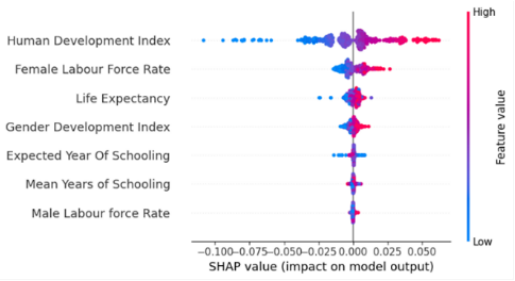

Figure 11 SHAP Result Decision Tree

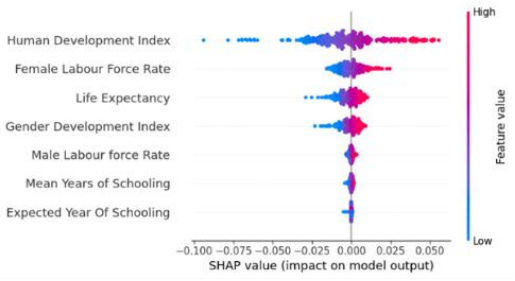

Figure 12 SHAP Result Stacking Regressor

The SHAP summary plot provides granular insight into how each feature contributes to the model's prediction. In this context, a positive SHAP value implies that the feature increases the predicted value of the Happiness Index for a specific instance, indicating a supportive or enhancing effect. Conversely, a negative SHAP value suggests that the feature contributes to lowering the predicted Happiness Index, exerting a suppressive influence. For instance, a high HDI value (depicted in red) tends to have a positive SHAP value, meaning that higher HDI consistently leads to higher predicted happiness levels. Meanwhile, lower feature values (blue) may contribute negatively depending on the indicator's nature and direction of influence. This dual-axis interpretation—magnitude and direction—enables stakeholders to understand not only which variables are important but also how and in what direction they affect prediction outcomes. Compared to prior models such as that of Jannani et al., where Random Forest reached 93.66% accuracy, the integration of ensemble learning with SHAP-based interpretability in this study offers both superior performance and greater transparency, contributing a valuable methodology for evidence-based urban policy and happiness modeling.

5. CONCLUSION

A comparative examination of literature about Happiness Index forecasts hinges on the employed model, the assessment methodology, and the analytical approach utilized. Diverse regression models have been employed by numerous academics to forecast the happiness index with differing degrees of accuracy. Research [10] employing the Random Forest model attained a R^2 value of 93.66%, with model assessment encompassing Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). This study used Permutation Importance (PI) for feature analysis, demonstrating that the GNI per capita variable significantly influences the model. A separate study [11] employed the Elastic Net Regression model, achieving a R^2 accuracy of 44.05%, however no additional evaluation methods were specified. Similarly, the study[21] employed the XGBoost Regression model, achieving a R^2 of 32%. This study demonstrated that the Stacking Ensemble Regressor approach (Proposed approach/PM) attained the maximum accuracy, evidenced by a R^2 value of 97.68%, a Mean Absolute Error (MAE) of 0.002900, a Mean Squared Error (MSE) of 0.000021, and a Root Mean Squared Error (RMSE) of 0.004604. This approach integrates analysis using Permutation Importance and SHAP, demonstrating that the Human Development Index significantly impacts happiness predictions.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by Universitas Logistik dan Bisnis Internasional in facilitating the completion of this research.

REFERENCES

[1] V. Susanti and A. Fitri, "Economics of Happiness: What Really Counts?," *Kne Social Sciences*, 2024, doi: 10.18502/kss.v9i16.16258.

[2] Kenzo, A. Yudiarso, M. A. Nugroho, and J. S. Mustika, "Analyzing Oxford Happiness Questionnaire Indonesian Version Using the Generalized Partial Credit Model," *Psyche 165 Journal*, 2024, doi: 10.35134/jpsy165.v17i2.353.

[3] Q. Liu, "Can Happiness Be Measured?," *Advances in Education Humanities and Social Science Research*, 2023, doi: 10.56028/aehssr.7.1.423.2023.

[4] K. Ruggeri, E. Garcia-Garzon, Á. Maguire, S. Matz, and F. A. Huppert, "Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries," *Health Qual Life Outcomes*, vol. 18, pp. 1–16, 2020.

[5] A. \. I. Akgun, S. P. Türkoğlu, and S. Erikli, "Investigating the determinants of happiness index in EU-27 countries: a quantile regression approach," *International Journal of Sociology and Social Policy*, vol. 43, no. 1/2, pp. 156–177, 2022.

[6] J. Tanuwijaya, C. Krisanti, and A. W. Gunawan, "The Impact of Perceived Inclusion Climate for Leader Diversity (PICLD) on Organizational Justice and Its Effects on Employee Engagement, Emotional Wage, and Happiness at Work," *Ajesh*, 2024, doi: 10.46799/ajesh.v3i10.423.

[7] P. R. Sihombing, "Comparison of Regression Analysis With Machine Learning Supervised Predictive Model Techniques," *Jurnal Ekonomi Dan Statistik Indonesia*, 2023, doi: 10.11594/jesi.03.02.03.

[8] S. C. Br Ginting, S. Afifuddin, and R. RAHMANTA, "Analysis of the Effect of Macroeconomic Variables on Happiness in Indonesia," *International Journal of Research and Review*, 2021, doi: 10.52403/ijrr.20211009.

[9] M. Acharya, A. Dumre, P. Poudel, and S. C. Dhakal, "Happiness Index; Current Status, Global Ranking and Its Importance in the Government's Agenda 'Prosperous Nepal, Happy Nepali,'" *Acta Scientific Agriculture*, 2020, doi: 10.31080/asag.2020.04.0789.

[10] A. Jannani, N. Sael, and F. Benabbou, "Machine learning for the analysis of quality of life using the World Happiness Index and Human Development Indicators," *Mathematical Modeling and Computing*, vol. 10, no. 2, pp. 534–546, 2023, doi: 10.23939/mmc2023.02.534.

[11] N. Çiftci and M. Yldz, "The relationship between social media addiction, happiness, and life satisfaction in adults: analysis with machine learning approach," *Int J Ment Health Addict*, vol. 21, no. 5, pp. 3500–3516, 2023.

[12] Z. Jiang, "Prediction and industrial structure analysis of local GDP economy based on machine learning," *Math Probl Eng*, vol. 2022, no. 1, p. 7089914, 2022.

[13] V. Çetin and O. Yldz, "A comprehensive review on data preprocessing techniques in data analysis," *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 28, no. 2, pp. 299–312, 2022.

[14] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022.

[15] I. Izonin, R. Tkachenko, N. Shakhovska, B. Ilchyshyn, and K. K. Singh, "A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain," *Mathematics*, vol. 10, no. 11, p. 1942, 2022.

[16] N. Pessanha Santos, "The Expansion of Data Science: Dataset Standardization," *Standards*, vol. 3, no. 4, pp. 400–410, 2023.

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

[17] S. F. Pane, A. G. Putrada, N. Alamsyah, and M. N. Fauzan, "A PSO-GBR solution for association rule optimization on supermarket sales," in *2022 seventh international conference on informatics and computing (ICIC)*, 2022, pp. 1–6.

[18] S. F. Pane, A. Adiwijaya, M. D. Sulistiyo, and A. A. Gozali, "Multi-Temporal Factors to Analyze Indonesian Government Policies regarding Restrictions on Community Activities during COVID-19 Pandemic," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2263–2269, 2023.

[19] S. F. Pane, A. G. Putrada, N. Alamsyah, M. N. Fauzan, and others, "The Influence of The COVID-19 Pandemics in Indonesia On Predicting Economic Sectors," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 2022, pp. 1–6.

[20] S. F. Pane, F. Abdullah, and R. Habibi, "DETEKSI EMOSI PADA TEKS BERBAHASA INDONESIA MENGGUNAKAN PENDEKATAN ENSEMBLE," *JTT (Jurnal Teknologi Terapan)*, vol. 10, no. 2, pp. 80–90, 2024.

[21] E.-J. Kim, H.-W. Kang, and S.-M. Park, "Leisure and Happiness of the Elderly: A Machine Learning Approach," *Sustainability*, vol. 16, no. 7, p. 2730, 2024.