Published online on the journal's webpage: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

# Journal of Dinda

**Data Science, Information Technology, and Data Analytics**

Vol. 5 No. 2 (2025) 164 - 175 E-ISSN: 2809-8064

# Enhancing Prediction Accuracy of the Happiness Index Using Multi-Estimator Stacking Regressor and Web Application Integration

**Rofi Nafiis Zain[1], Nisa Hanum Harani[2], Syafrial Fachri Pane[3*]**

[1, 2, 3*]Applied Informatics Engineering, Vocational School, Universitas Logistik dan Bisnis Internasional

[1]rofinafiisr@ulbi.ac.id, [2]nisa@ulbi.ac.id, [3*]syafrial.fachri@ulbi.ac.id

**Abstract**

This study presents a novel approach to predicting the Happiness Index using a multi-estimator stacking regressor integrated into a web-based application. The dataset, sourced from Badan Pusat Statistik Indonesia (BPSI), includes six socioeconomic features: Life Expectancy, Gender Development Index, Expected Years of Schooling, Human Development Index, Labor Force Rate (female and male), and Mean Years of Schooling. The proposed model combines five regression algorithms—Decision Tree, Random Forest, Gradient Boosting, LGBM, and Support Vector Regressor (SVR)—to improve prediction accuracy. The ensemble achieved strong performance with an $R^2$ score of 0.981398 and low error rates (MAE = 0.002668, MSE = 0.000017, RMSE = 0.004123). A custom Happiness Score was computed using Pearson correlation values as feature weights. To enhance interpretability, SHapley Additive exPlanations (SHAP) were applied, identifying the Human Development Index, Female Labor Force Rate, and Life Expectancy as the most influential variables. The trained model was deployed through a Flask-based dashboard, allowing users to interact with dynamic visualizations by filtering regions and years. These results demonstrate that stacking-based regression, coupled with explainability and practical deployment, can serve as a reliable tool for modeling socioeconomic well-being and supporting data-driven policy decisions.

Keywords: Prediction, Happiness, Stacking Regressor, SHAP, Web-Based Application



## 1. Introduction

Research on the niche of happiness has garnered significant traction in the pursuit of understanding and enhancing human well-being. Happiness, often equated with subjective well-being, is a multidimensional construct influenced by various factors, such as emotional state, life satisfaction, and cognitive appraisal. Financial literacy has emerged as a key predictor of happiness, with studies indicating that it may surpass income in its ability to predict overall life satisfaction [1]. This conceptualization aligns with broader definitions within psychology that examine happiness as not merely a transient emotion, but a durable quality-of-life assessment encompassing both affective and cognitive dimensions [2].

Recent methodologies employed to gauge happiness have increasingly led to advanced machine learning techniques, highlighting a paradigm shift in which traditional measures are complemented or replaced by data-driven approaches [3]. Researchers have demonstrated the efficacy of machine learning frameworks, such as Random Forest and XGBoost, in predicting happiness indices with notable accuracy. These techniques provide insights into the variables that influence happiness and quantify their relative importance, thus reframing happiness research in the context of computational science [4]. This burgeoning field highlights the potential of harnessing big data and algorithmic modeling to develop nuanced insights into what constitutes happiness across diverse populations.

Despite the growing success of these approaches, limitations in prior studies reveal a pressing need for more powerful, generalizable, and deployable predictive models. For instance, Jannani et al. (2023) [3] applied a Random Forest Regression model with standardized input features and permutation importance analysis, achieving an $R^2$ of 93.66%, but without integration into practical tools or deeper ensemble strategies. Yang et al. (2025) [4] explored similar approaches with Random Forest and ANOVA-based feature importance, yielding a lower accuracy of 84.1%. Other studies such as Han et al. (2023) [5] applied a fusion-weighted average with basic imputation and feature engineering but reported extremely low predictive power with an $R^2$ of only

Received: 07-06-2025 | Accepted: 14-07-2025 | Published: 05-08-2025

20.8%. Meanwhile, Kim et al. (2024)[6] used Lasso Regression with one-hot encoding and standardization, achieving just 31.9% $R^2$ despite applying RMSE for evaluation. These works were largely constrained by limited model complexity, lack of interpretability, and absence of real-world integration—factors that significantly limit their applicability and impact.

This study seeks to address significant gaps in the literature by integrating a multi-estimator stacking regressor approach, which amalgamates various predictive models to enhance the accuracy of happiness index predictions. The principal hypothesis posits that employing a synthesis of multiple estimators—specifically Decision Tree, Random Forest, Gradient Boosting, LGBM, and SVR—will yield superior predictive performance compared to isolated models. This method outperforms the previous stacking configuration (Decision Tree, Random Forest, XGBoost), demonstrating its improved ability to capture complex, nonlinear patterns in socioeconomic data. The proposed model also retains interpretability through the use of SHAP analysis, which identifies key variables such as Human Development Index, female labor force rate, and life expectancy as primary contributors to happiness scores.

The overarching aim of this article is to construct a robust predictive model for the happiness index that not only advances theoretical understanding but also translates these insights into practical applications, exemplified through web application deployment. This dual focus on precision and accessibility underscores the commitment to making complex predictive analyses user-friendly and impactful in everyday decision-making[7]. By augmenting the existing knowledge base and introducing innovative methodologies, this study contributes significantly to happiness studies, engendering a more nuanced understanding of the factors that foster or hinder human well-being in contemporary societies.

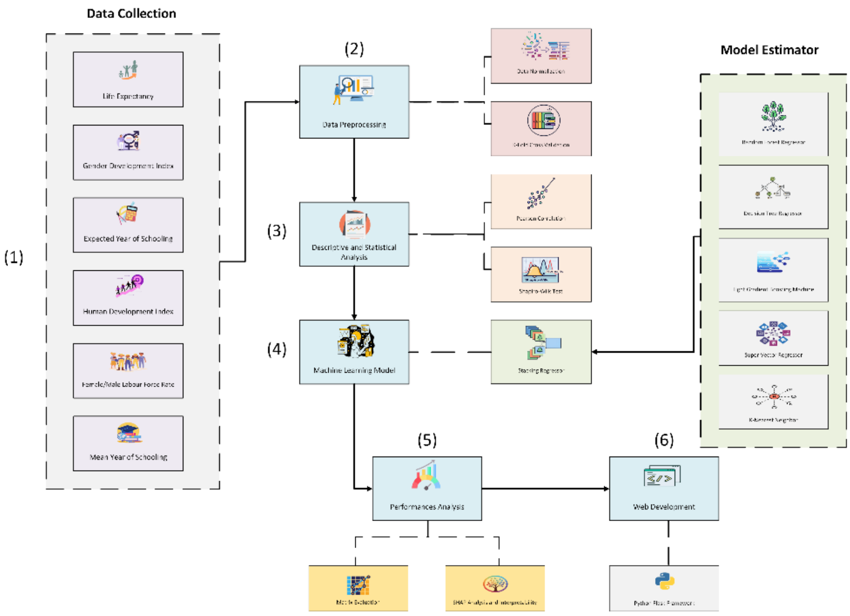



Figure 1 Propose Method

## 2. Research Methods

### 2.1. *Data Collection*

The first stage of the proposed methodology involves collecting relevant socioeconomic and developmental indicators that are hypothesized to influence the Happiness Index (H-index) across urban areas in Indonesia. As illustrated in Figure 1, six key features were selected: Life Expectancy, Gender Development Index, Expected Years of Schooling, Human Development Index (HDI), labor force participation rate (female/male), and Mean Years of Schooling. These variables were chosen based on their strong theoretical and empirical relevance to previous happiness and development studies.

The primary data used in this research were sourced from Badan Pusat Statistik Indonesia (BPSI), Indonesia's official national statistics agency, along with supporting data from global development repositories to ensure reliability, accuracy, and adequate regional coverage. The dataset consists of 549 rows of city-level data, representing a wide range of socio-economic conditions across different regions. Each feature was carefully cleaned, standardized, and formatted to ensure consistency and completeness, forming the foundational input for the subsequent preprocessing and machine learning modeling pipeline.

Table 1 provides a detailed overview of the dataset features used in this study, including their variable names, units, and definitions. This structured summary highlights the role of each indicator in shaping the predictive model for the Happiness Index.

Table 1 Description and Definitions of Dataset Features

| Variable | Description | Unit/Scale | Definition |
|---|---|---|---|
| $x_1$ | Life Expectancy | Percentage (%) | A composite index measuring average achievement in health, education, and income. |
| $x_2$ | Human Development Index (HDI) | Percentage (%) | A composite index measuring average achievement in health, education, and income. |
| $x_3$ | Mean Years of Schooling | Percentage (%) | The average number of completed years of education of a country's population aged 25 years and older. |
| $x_4$ | Male Labor Force Participation Rate | Percentage (%) | The proportion of the male population aged 15+ that is economically active (employed or actively seeking employment). |
| $x_5$ | Female Labor Force Participation Rate | Percentage (%) | The proportion of the female population aged 15+ that is economically active. |
| $x_6$ | Expected Years of Schooling | Percentage (%) | The total number of years of schooling a child entering the education system can expect to receive. |
| $x_7$ | Gender Development Index (GDI) | Percentage (%) | Measures gender equality by comparing female and male HDI values. |

## 2.2. *Data Preprocessing*

The second stage of the methodology focused on preparing the collected dataset to ensure its quality, consistency, and suitability for modeling. The preprocessing phase began with data normalization using the Min-Max Scaling technique, which transformed all percentage-based and index-based indicators from a 0–100 range into a standardized 0–1 scale. This approach ensured uniform scaling across features and reduced the influence of large value disparities on the learning process. Subsequently, missing value detection was conducted using the *.isnull()* function to identify any incomplete entries. However, the dataset was found to be complete, with no missing values detected, thereby eliminating the need for imputation or record removal. This clean and uniformly scaled dataset provided a reliable foundation for the subsequent modeling phase.

To support robust model evaluation, the dataset was split into training and testing sets using an 80:20 split, and a K-fold cross-validation procedure was implemented to mitigate overfitting and provide generalized performance metrics. This step enhances the reliability of the modeling outcomes by ensuring that the model's performance is validated across multiple subsets of data[8]. All preprocessing operations were carried out using Python libraries such as Pandas, Scikit-learn, and NumPy, providing a structured foundation for the statistical analysis and machine learning stages that follow.

## 2.3. *Descriptive and Statistical Analysis*

Descriptive and statistical analyses were performed to better understand the overall structure of and the relationships within the dataset. This process involved examining the central tendencies and variability of each development indicator to detect potential irregularities or outliers that could influence the modeling results. As presented in Table 1, the descriptive statistics include the mean, standard deviation (SD), and minimum and maximum values for each feature, providing a clear overview of their distribution across the observed cities. For instance, the Mean Years of Schooling (x3) has a relatively low mean of 8.63 years, while the Gender Development Index (x7) shows a high average of 89.76, indicating varying levels of education and gender equality across different regions.

In addition to descriptive metrics, the Shapiro-Wilk normality test was employed to assess whether each variable followed a normal distribution[9]. The Shapiro-Wilk test is a widely used method to test the null hypothesis that a sample comes from a normally distributed population[10]. A p-value less than 0.05 typically indicates a significant deviation from normality. As shown in Table 2, most features, such as Life Expectancy, Human Development Index, and Expected Years of Schooling, produced p-values below 0.001, suggesting non-normal distribution. Only the male labor force rate (x4) has a p-value above 0.05, indicating that it does not significantly deviate from a normal distribution and may be appropriate for parametric modeling.

Table 2 Descriptive statistics and normality test results for each feature

| Feature description | Variable | Mean (SD) | [Min, Max] | P-value |
|---|---|---|---|---|
| *Life Expectancy* | $x_1$ | 69.3861 (3.4407) | [55.08, 77.645] | <0.001 |
| *Human Development Index (HDI)* | $x_2$ | 69.6224 (6.4193) | [30.75, 86.65] | <0.001 |
| *Mean Years of Schooling* | $x_3$ | 8.2367 (1.6144) | [0.97, 12.64] | <0.001 |
| *Male Labor Force Participation Rate* | $x_4$ | 83.5473 (3.823) | [71.17, 97.0] | 0.0189 |
| *Female Labor Force Participation Rate* | $x_5$ | 53.9134 (10.2206) | [30.28, 97.04] | <0.001 |
| *Expected Years of Schooling* | $x_6$ | 12.9006 (1.3121) | [3.29, 17.39] | <0.001 |
| *Gender Development Index (GDI)* | $x_7$ | 89.7599 (5.9311) | [53.71, 99.05] | <0.001 |

*Note:* The values shown represent the mean, standard deviation (SD), minimum, and maximum for each feature. P-values were calculated using the Shapiro-Wilk test to assess data normality. A P-value less than 0.001 indicates a significant deviation from the normal distribution.

Furthermore, Pearson's correlation analysis was conducted to examine the linear relationships among the selected indicators[11]. The resulting correlation coefficients were directly used as weights for each variable in constructing the composite Happiness Score. In this approach, each indicator was multiplied by its corresponding Pearson correlation value, and the weighted results were summed to compute the final score. This method ensures that features with stronger linear relationships to the target variable contribute more significantly to the predicted Happiness Index, resulting in a robust, data-driven, and interpretable metric. The Happiness Score is computed using the following formula:

$$H_{Index} = w_1 \cdot g + w_2 \cdot h + w_3 \cdot i + w_4 \cdot \frac{j+k}{2} + w_5 \cdot l \qquad (1)$$

The Happiness Index is computed based on Equation (1), where each variable represents a key socio-economic indicator: "g" corresponds to Life Expectancy, "h" to the Human Development Index (HDI), "i" to the Mean Years of Schooling, "j" and "k" to the Female and Male Labor Force Participation Rates respectively, and "l" to the Gender Development Index (GDI). Each variable is multiplied by a corresponding weight $w$, which reflects the relative importance or contribution of that variable, as determined through correlation analysis. These weights ensure that indicators with stronger statistical influence on happiness have a proportionally greater impact on the final index score.

### 2.4. *Machine Learning Modelling*

In the fourth stage of the methodology, a machine learning modeling pipeline was implemented to predict the Happiness Score based on the processed socio-economic indicators. Several regression models were selected to compare their performance and robustness, including Decision Tree Regressor, Random Forest Regressor, Light Gradient Boosting Machine (LGBM), and Support Vector Regressor (SVR). These models were chosen because of their proven effectiveness in handling nonlinear relationships, varying data distributions, and high-dimensional datasets.

After the individual models were trained and evaluated, a more advanced ensemble strategy was introduced using a Stacking Regressor, which combines the predictive strengths of multiple base learners [12]. This ensemble architecture integrates the predictions of individual models as inputs for a meta-learner, thereby enabling it to learn complex interactions and enhance generalization. Two versions of the stacking regressor were evaluated: the baseline model, which included three base estimators—Random Forest, Decision Tree, and XGBoost—and the enhanced version, which incorporated Gradient Boosting, LGBM, and SVR alongside Random Forest and Decision Tree. In the enhanced version, XGBoost was removed to reduce redundancy with Gradient Boosting and to allow for greater diversity among the base learners, especially through the inclusion of LGBM and kernel-based SVR, which offer complementary modeling capabilities.

The training process involved fitting each model on the training set and validating performance using 5-fold cross-validation to ensure robustness and minimize overfitting. Hyperparameter tuning for each base learner was conducted using a grid search strategy and empirical optimization based on validation scores. The final stacking ensemble was implemented using the Scikit-learn and XGBoost libraries, with carefully selected settings for each estimator to achieve optimal predictive performance. This improved model demonstrated consistent superiority across all evaluation metrics and was deployed as the core prediction engine for the interactive web-based dashboard developed in the final stage.

### 2.5. *Performances Analysis*

A comprehensive performance analysis was conducted to evaluate and interpret the predictive performance of the machine learning models. Standard regression evaluation metrics were employed, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared ($R^2$), to assess each model's accuracy and reliability[13]. These metrics provide quantitative insights into how closely the predicted values align with the actual Happiness Scores and help identify which models generalize best to unseen data.

In addition to performance metrics, SHapley Additive exPlanations (SHAP) analysis was applied to improve the interpretability of the best-performing model[14]. SHAP is a model-agnostic explanation technique based on cooperative game theory that assigns each feature an importance value representing its contribution to the final prediction[15]. It computes the Shapley values, which reflect the marginal contribution of a feature across all possible feature combinations. This enables a clear understanding of how each input feature impacts the model's output, both globally (overall importance across the dataset) and locally (on a per-instance basis).

In this study, SHAP was instrumental in visualizing the effect of socioeconomic indicators on the predicted Happiness Score. It helped uncover not only which variables were most influential, but also how changes in those variables influenced predictions. The SHAP bar, summary, and dependence plots collectively provide a comprehensive explanation of the model's decision-making process. Thus, the use of SHAP strengthened the transparency and trustworthiness of machine learning predictions.

The mathematical formulations of the evaluation metrics used in this study are described in detail in the following equations:

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - x_i| \quad (2)$$

Equation (2) explains that $y_i$ is predicted value. $x_i$ is true value. $n$ is amount of data [16]

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

Equation (3) shows the MSE formula, averages the error square between the prediction and actual values. Smaller MSE values improve model prediction[17]

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

Equation (4) explains that $y_i$ is predicted value. $x_i$ is true value. $n$ is amount of data [18]

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

Equation (5) represents the coefficient of determination ($R^2$) formula, which quantifies the proportion of total variation in the dependent variable that can be elucidated by the model[19]. This illustrates the model's predictive performance relative to a basic model that forecasts only the mean value.

### 2.6. *Web Development*

The final stage of the proposed methodology involves deploying the trained machine learning model into a web-based dashboard to provide real-time visualization of happiness index predictions in Indonesian cities. The application was developed using the Python Flask framework, which serves as the back-end logic controller and API handler to dynamically deliver model outputs[20]. Unlike interactive systems that rely on user-submitted inputs, this dashboard is designed to receive updated data exclusively through a manual process conducted via an admin panel, where authorized personnel can input and modify socioeconomic indicators as new information becomes available. Users are provided with a read-only, interactive visualization of the Happiness Index, ensuring controlled data management and consistency across the dashboard. These inputs are automatically processed by the trained Enhanced Stacking Regressor, and the resulting predictions are visualized on the front end of the dashboard for public monitoring and decision-making purposes.

As shown in Figure 6, the main interface of the dashboard displays insightful visual summaries, including city-level rankings for variables such as Life Expectancy, Human Development Index, Years of Schooling, and the final Happiness Score. Users can also interact with dynamic graphs that illustrate temporal trends for selected cities, such as year-over-year progress in HDI and happiness levels. This interactive design ensures that users, from policymakers to the general public, can easily explore and interpret socioeconomic trends.

By integrating a machine learning workflow into a web application, this study bridges the gap between predictive analytics and practical decision support. It transforms complex statistical results into an intuitive and actionable interface, further enhancing the societal value of research.

## 3. Results and Discussion

### 3.1. *Correlation Analysis*

The Pearson correlation heatmap in Figure 2 illustrates the strength and direction of the linear relationships between selected development indicators. Several strong positive correlations can be observed, particularly between the *Human Development Index* (HDI) and both *Mean Years of Schooling* (r = 0.87) and *expected years of schooling* (r = 0.80). Similarly, *Life Expectancy* showed a moderately strong correlation with HDI (r = 0.71), highlighting the interconnectedness of education and health with human development metrics.

In contrast, the *Male Labor Force Rate* demonstrates negative correlations with key indicators such as HDI (r = -0.51), expected years of schooling (r = -0.61), and *Mean Years of Schooling* (r = -0.54), suggesting that higher male labor participation is inversely related to education and development levels in certain urban contexts. Meanwhile, the *Gender Development Index* shows positive but moderate correlations with most variables, indicating its complementary role in assessing development from an equality perspective.

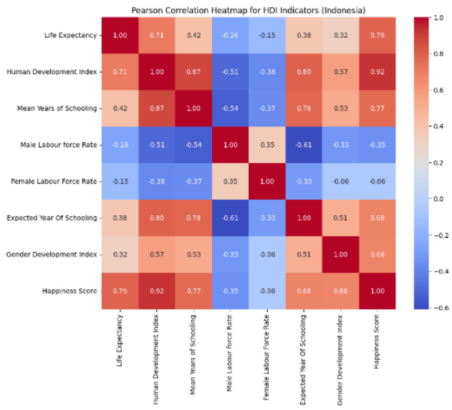



Figure 2 Pearson Correlation

Based on the strength of these correlations, a set of weights is assigned to each feature to compute the Happiness Score. Indicators with stronger and more direct associations with development and well-being were assigned a higher weight. Specifically, the *Human Development Index* ($w_2 = 0.30$) and *Life Expectancy* ($w_1 = 0.25$) received the highest weights because of their substantial positive correlations. *Mean Years of Schooling* was also weighted at $w_3 = 0.25$, reflecting its strong correlation with HDI and Expected Years of Schooling. The combined *Labor Force Rate*, averaged between males and females, was assigned a lower weight of $w_4 = 0.15$, as it demonstrated weaker and, in some cases, negative correlations with other features. Finally, the *Gender Development Index* received a weight of $w_5 = 0.10$, acknowledging its relevance while balancing its moderate statistical influence.

These weights were then integrated into the custom formula used to calculate the Happiness Score for each city, ensuring that the resulting metric was both data-driven and interpretable based on statistical relationships.

### 3.2. *Modelling*

Algorithm 1 shows the complete workflow used in this study to predict the Happiness Index based on HDI-related indicators and integrate the results into a web-based application. The algorithm outlines all essential stages, including data preprocessing, happiness score computation, model training, cross-validation, and final deployment. The trained model is saved as a *.pkl* file and integrated into the Flask application using the Joblib library, allowing real-time predictions to be served through a Flask REST API. The backend system is built using the Flask Python framework, with MongoDB employed to store and manage the socioeconomic data, and the entire application is containerized using Docker for scalable deployment.

From the user side, the application provides an interactive dashboard where users can explore dynamic visualizations of the predicted Happiness Index. While users cannot modify or input data, they can interactively filter the dashboard by year and city, as well as view the rankings of cities by happiness level. All data updates are performed manually by administrators via a secured admin panel, ensuring controlled data management. This architecture enhances usability and supports accessible, real-time dissemination of predictive insights to stakeholders and decision-makers.

In this section, we compare the performance of various machine learning models implemented in Steps 5–8 of Algorithm 1. Four models were evaluated: Decision Tree, Gradient Boosting (LGBM), Random Forest, and the proposed Enhanced Stacking Regressor. The comparison was based on both cross-validation metrics and the final test performance, with a particular focus on the accuracy improvement resulting from the enhanced ensemble configuration.

To assess the effectiveness of different machine learning models in predicting happiness scores, a comparative analysis was conducted using four models: Decision Tree, Gradient Boosting (LGBM), Random Forest, and Stacking Regressor. The evaluation was performed using both 5-Fold Cross Validation and testing on the unseen data.

As shown in Figure 3, the R-squared values across the five folds indicate consistent performance among ensemble-based models, with the Enhanced Stacking Regressor (dashed cyan line) outperforming all others across most folds, achieving the highest and most stable scores. In contrast, the Decision Tree (blue line) showed the lowest performance and highest variability across folds, with a gradual decline in R-squared values, indicating a lower generalization capability.

The second image provides a summary of the model performance on the testing dataset using four key evaluation metrics. MAE, MSE, RMSE, and R-squared values The Enhanced Stacking Regressor, which integrates additional estimators such as LGBM and SVR, achieved the best overall results with the lowest error rates (MAE = 0.002668, RMSE = 0.004123) and the highest accuracy (R-squared = 0.981398). This represents a significant improvement compared to the previously developed stacking model, which had an R-squared value of only 0.944 (94.4%). The performance boost confirms that the inclusion of more diverse and powerful base learners in the stacking architecture, especially models with complementary learning characteristics, such as LGBM and SVR, enhances the model's ability to capture complex patterns in the data. These findings demonstrate that model stacking, when properly optimized with well-chosen base estimators, can significantly improve the prediction accuracy in socioeconomic modeling tasks such as happiness index estimation.

To further validate the effectiveness of the enhanced Stacking Regressor, a comparison between the actual and predicted Happiness Scores was conducted, as shown in Figure 4.

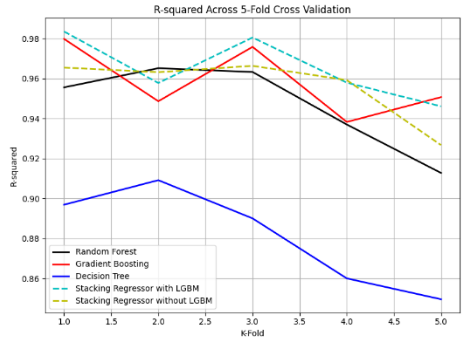



Figure 3 Performances Models

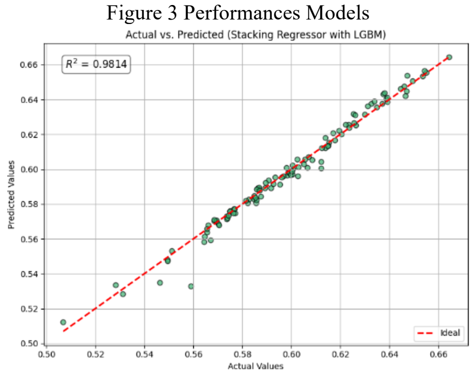



Figure 4 Actual vs Predicted

Each point in the scatter plot represents a city with its actual Happiness Score plotted on the x-axis and the predicted score on the y-axis. The dashed red line indicates the ideal prediction line in which the predicted values exactly match the actual values.

The results clearly show that the predicted values from the Stacking Regressor with LGBM and SVR are tightly clustered around the ideal line, indicating a high prediction accuracy. This visual confirmation is consistent with the quantitative metrics reported earlier, where the model achieved an $R^2$ score of 0.9814, which was the highest among all tested models. The minimal dispersion of points around the diagonal suggests that the model generalizes well and maintains low bias and variance, further validating the effectiveness of the enhanced estimator configuration.

This strong alignment between the predicted and actual values reinforces the findings discussed in the previous section and confirms that the enhanced Stacking Regressor can capture complex relationships in the data while delivering highly reliable predictions for the Happiness Index.

### 3.3. *SHAP Analysis and Interpretablity*

To better understand how individual features contribute to the predictions made by the Enhanced Stacking Regressor, SHAP (SHapley Additive exPlanations) analysis was conducted. As illustrated in Figure 5, three types of SHAP visualizations were used to interpret the model output: bar, summary, and dependence plots.

Figure 5(a) presents the SHAP bar plot, which illustrates the average absolute SHAP values for each feature, reflecting their overall contribution to the model's output across all observations. The Human Development Index (HDI) emerges as the most influential predictor of the Happiness Score, followed by the Female Labor Force Rate and Life Expectancy. These findings are consistent with the Pearson correlation results and weight assignment discussed in Section 3.1, reinforcing the significance of these features in shaping the predicted happiness levels.

**Algorithm 1:** HDI-Based Happiness Prediction and Web Application Integration

**Input:** *HDI dataset (Life Expectancy, Gender Development Index, Schooling, Labor Force Rates)*
**Output:** *Predicted Happiness Score, Feature Importance Rankings, REST API endpoint*

1: Load dataset *HDI_2019.csv*; clean and handle missing values;
2: Normalize percentage columns to decimals; convert relevant columns to numeric;
3: Compute the happiness score:

$$H_{Index} = w_1 \cdot g + w_2 \cdot h + w_3 \cdot i + w_4 \cdot \frac{j+k}{2} + w_5 \cdot l$$

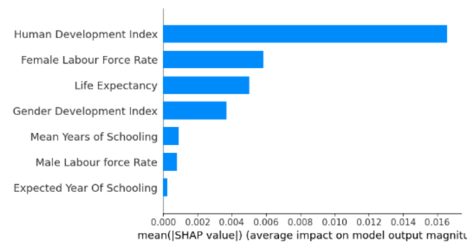
4: Split data into features $X$ and target $Y$; perform a train-test split (80:20);
5: Initialize models: LinearRegression, DecisionTree, RandomForest, LGBM;
6: Train models on X_train; evaluate them on X_test with MAE, MSE, RMSE, $R^2$;
7: Apply cross-validation and calculate permutation importance for each model\;
8: Train the final ensemble using StackingRegressor; save with joblib.dump();

**Web Application Integration using Flask:**

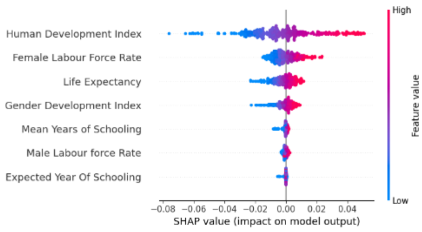
9: Define Flask route: @app.route('/predict', methods=['POST'])} ;
10: Load model: model = joblib.load("stacked_model.pkl");
11: Receive input JSON via POST, parse into feature vector $X_{new}$;
12: Predict using the trained model: pred = model.predict(X_new);
13: Return prediction as JSON response: {"prediction": pred};

Figure 5(b) displays the SHAP summary plot, which extends the interpretation by showing both the distribution and direction of each feature's impact across all data points. For instance, high values of HDI and Female Labor Force Rate (colored in red) tend to have positive SHAP values, indicating that they increase the predicted Happiness Score. Conversely, low values of features such as Male Labor Force Rate and Expected Years of Schooling (in blue) are associated with negative SHAP impacts, reducing the predicted score. This plot also reveals the variability of each feature's influence, highlighting that even features with lower average importance may have strong localized effects on specific predictions.

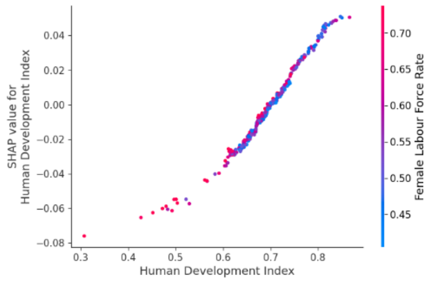
Figure 5(c) presents the SHAP dependence plot for HDI, illustrating a clear positive monotonic relationship between the HDI value and its SHAP contribution. This means that as HDI increases, its positive effect on the model's prediction also increases in a consistent manner. Moreover, the color gradient representing the Female Labor Force Rate reveals a second-order interaction: the positive effect of HDI is more pronounced when the Female Labor Force Rate is high (red points), suggesting that the benefit of human development on happiness is amplified in regions with greater female economic participation. This interaction highlights the multidimensional nature of happiness determinants, where socioeconomic progress and gender inclusion jointly influence well-being outcomes.




(a) SHAP Bar Plot




(b) SHAP Summary Plot




(c) SHAP Dependence Plot

Figure 5 SHAP visualizations for feature impact analysis: bar plot, summary plot, and dependence plot.

## 3.4. *Web App Development*

To bridge the results of the machine learning model with real-world accessibility, a web-based dashboard was developed using the Python Flask framework. This application serves as a visual interface for stakeholders to explore the predicted Happiness Index across Indonesian cities in an intuitive and informative manner. The prediction values displayed on the dashboard were generated by the Enhanced Stacking Regressor model discussed in Section 3.2 and are presented through a REST API integration. Rather than accepting user input, the dashboard provides interactive visualizations and curated data outputs, allowing users to engage with model insights without altering the underlying prediction pipeline.

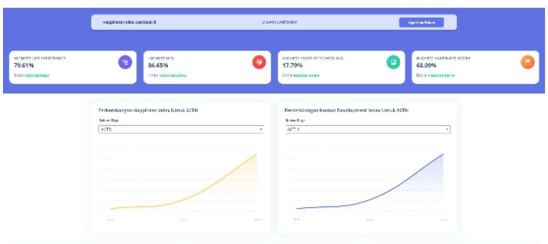


Figure 6 Web App Dashboard

As shown in Figure 6, the homepage of the dashboard presents key insights, such as the cities with the highest values for Life Expectancy, Human Development Index, Years of Schooling, and overall Happiness Score. Users can also explore detailed visualizations showing the temporal progression of both the Happiness Index and HDI for each selected city, as illustrated in the charts for the region of Aceh. These dynamic plots are rendered based on processed data and offer an interactive way to understand socioeconomic development trends over time.

This application not only demonstrates the applicability of machine learning in public policy and urban planning but also highlights the benefits of model interpretability and explainable AI in real-world decision support systems.

## 3.5. *Discussion*

The results of this study demonstrate that the Enhanced Stacking Regressor, which integrates diverse base learners including Decision Tree, Random Forest, Gradient Boosting, LGBM, and Support Vector Regressor (SVR), significantly improves the prediction accuracy of the Happiness Index in Indonesian cities. The enhancement lies in the expansion of the model architecture compared to the previous stacking version, which only utilized three estimators (Decision Tree, Random Forest, and XGBoost). By incorporating additional models—particularly LGBM and SVR, which are capable of capturing both gradient-based and kernel-based nonlinearities—the improved ensemble achieves a more comprehensive representation of the underlying data structure. With an R-squared value of 0.981, the model outperformed both individual regressors and the earlier stacking model used in prior research. This improvement highlights the advantage of ensemble learning with heterogeneous models in capturing nonlinear and high-dimensional relationships within socioeconomic data. Moreover, the use of SHAP analysis provided deeper insights into feature importance, validating that the Human Development Index, Female Labor Force Rate, and Life Expectancy are the most influential variables, supporting previous assumptions drawn from correlation and weight-based analyses.

To further assess the effectiveness of the proposed approach, a comparative analysis with related studies is presented in Table 3. The referenced works by [3] and [4] both utilized Random Forest Regression but relied on simpler preprocessing techniques such as standardization and missing value imputation. While [3] achieved an $R^2$ of 93.66% using Permutation Importance (PI) and [4] reported an $R^2$ of 84.1% with ANOVA for feature analysis, neither approach incorporated model interpretability frameworks as comprehensive as SHAP. In contrast, the proposed method combined Shapiro-Wilk testing, data normalization, and multi-metric evaluation (MSE, MAE, and RMSE), leading to a notably higher $R^2$ of

Table 3 Comparison of Related Works and Proposed Method

| Author | Model | Pre-processing | Evaluation Model | | | $R^2$ | Feature Importance | Web App Integration |
|---|---|---|---|---|---|---|---|---|
| | | | MSE | MAE | RMSE | | | |
| [3] | Random Forest Regression | StandarScaler | ✓ | ✓ | ✓ | 93.66% | PI | - |
| [4] | Random Forest Regression | Missing vaue in population, Standarization | ✓ | - | - | 84.1% | ANOVA | - |
| [5] | Fusion Weighted Average | Imputation, Outlier Removal, Feature Engineering | ✓ | - | - | 20.8% | - | - |
| [6] | Lasso Regression | One-hot encoding, standardization | - | - | ✓ | 31.9% | - | - |
| **PM (Propose Method)** | Stacking Regressor | **Shapiro-Wilk** Test, Data Normalization | ✓ | ✓ | ✓ | 98.14% | SHAP | ✓ |

98.14%. This demonstrates not only a more accurate prediction model but also greater transparency and explainability in the analysis of contributing features.

Additionally, the work by [5] focused on predicting national happiness scores using a hybrid of three regression models—XGBoost, CatBoost, and Gradient Boosting—achieving a strong performance with a minimized MSE of 0.46704, although it placed less emphasis on $R^2$ as a central evaluation metric. Meanwhile, [6] applied eight regression models to predict elderly happiness in South Korea, ultimately selecting Lasso Regression as the optimal model with an $R^2$ of 0.319, highlighting the influence of leisure satisfaction and public facility use. These complementary findings reinforce the value of using interpretable, ensemble-based regression strategies for modeling happiness outcomes across diverse populations and feature sets.

Compared to previous research, which employed a Stacking Regressor with three estimators—Decision Tree, Random Forest, and XGBoost—achieving an accuracy of 97.4%, the current study improves the model by incorporating additional estimators, namely Gradient Boosting, LGBM, and SVR, resulting in a higher accuracy of 98.1%. While both studies used Permutation Importance and SHAP for model interpretability, the expanded ensemble in the current approach proved more effective in capturing complex feature interactions. Despite these promising results, this study had several limitations that warrant further investigation. The current model was trained on a static dataset without accounting for temporal factors, such as policy changes or economic shocks. Additionally, the Happiness Score was computed using fixed weights derived from linear correlations, which may overlook deeper nonlinear feature interactions.

In future work, the integration of time-series modeling or longitudinal datasets could provide a more dynamic and realistic predictive framework. Furthermore, exploring automated feature selection, neural-network-based regressors, and context-aware ensemble methods may enhance model robustness. On the application side, expanding the web dashboard to support interactive queries, real-time updates, or scenario simulations could transform it into a comprehensive decision-support tool for policymakers, urban planners, and the public.

## 4. Conclusion

This study demonstrates the effectiveness of a multi-estimator stacking regressor—combining Decision Tree, Random Forest, Gradient Boosting, LGBM, and SVR—in significantly improving Happiness Index prediction for Indonesian cities, achieving an $R^2$ score of 0.9814, surpassing previous Stacking Regressor models. SHAP analysis revealed the Human Development Index, female labor force rate, and life expectancy as the most influential features, adding interpretability to the model. The resulting system was deployed via a Python Flask web application, providing interactive access to city-level happiness insights. These results highlight the potential of machine learning for accurate socioeconomic prediction and practical decision-support, with future directions including time-series modeling and real-time data integration.

**Acknowledgments**

I would like to express my deepest gratitude to the Department of Informatics Engineering at Universitas Logistik dan Bisnis Internasional for the continuous support, guidance, and resources provided throughout this research. Their commitment to academic excellence and innovation is a significant source of inspiration. I am especially thankful for the opportunity to grow and contribute to such a dynamic and forward-thinking academic environment.

**References**

[1] H. Xiao and Z. Xin, "Financial Literacy Is Better Than Income to Predict Happiness.," *J Neurosci Psychol Econ*, 2022, doi: 10.1037/npe0000161.

[2] E. J. Kim, S. K. Kim, S. H. Jung, and Y. S. Ryu, "Predictive Factors of Adolescents' Happiness: A Random Forest Analysis of the 2023 Korea Youth Risk Behavior Survey," *Child Health Nursing Research*, 2025, doi: 10.4094/chnr.2024.049.

[3] 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*, no. 10, Num. 2, pp. 534–546, 2023.

[4] B. Yang and X. Xie, "Analyzing and Predicting Global Happiness Index via Integrated Multilayer Clustering and Machine Learning Models," *PLoS One*, 2025, doi: 10.1371/journal.pone.0322287.

[5] Y. Han, Y. Shao, and Y. Zhang, "Happiness index prediction using hybrid regression model," in *proceedings of the 2nd international academic conference on blockchain, information technology and smart finance (ICBIS 2023)*, 2023, pp. 76–87.

[6] 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.

[7] N. V. Sailaja, K. L. Reddy, G. V Aditya, B. Shashank, and V. H. Sai, *Happiness Index Prediction of Students Using Machine Learning*. 2023. doi: 10.2991/978-94-6463-314-6_9.

[8] 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.

[9] E. González-Estrada, J. A. Villaseñor, and R. Acosta-Pech, "Shapiro-Wilk test for multivariate skew-normality," *Comput Stat*, vol. 37, no. 4, pp. 1985–2001, 2022.

[10] J. Rochon, M. Gondan, and M. Kieser, "To test or not to test: Preliminary assessment of normality when comparing two independent samples," *BMC Med Res Methodol*, vol. 12, pp. 1–11, 2012.

[11] T.-P. Teng and W.-J. Chen, "Using Pearson correlation coefficient as a performance indicator in the compensation algorithm of asynchronous temperature-humidity sensor pair," *Case Studies in Thermal Engineering*, vol. 53, p. 103924, 2024.

[12] N. P. Jakowiec and K. M. Cramer, "More Than Oneself: Cultural Values as Predictors of Happiness," *Journal of Interpersonal Relations Intergroup Relations and Identity*, 2022, doi: 10.33921/onai9138.

[13] Y. Altork, "Comparative Analysis of Machine Learning Models for Wind Speed Forecasting: Support Vector Machines, Fine Tree, and Linear Regression Approaches," *International Journal of Thermofluids*, p. 101217, 2025.

[14] K. K. P. M. Kannangara, W. Zhou, Z. Ding, and Z. Hong, "Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 4, pp. 1052–1063, 2022.

[15] M. Li, H. Sun, Y. Huang, and H. Chen, "Shapley value: from cooperative game to explainable artificial intelligence," *Autonomous Intelligent Systems*, vol. 4, no. 1, p. 2, 2024.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] R. Chauhan, A. Goel, B. Alankar, and H. Kaur, "Predictive modeling and web-based tool for cervical cancer risk assessment: A comparative study of machine learning models," *MethodsX*, vol. 12, p. 102653, 2024.