Mathematical Journal of Modelling and Forecasting
Vol. 4, No. 1, June 2026, pp. 47-56

Application of Support Vector Regression (SVR) with a Radial Basis Function (RBF) Kernel for Predicting the Global Happiness Index

Aji Pandu Winata[1], Nala Kamila Azizy[2], Wina Ayu Lestari [3*], Riwi Dyah Pangesti [4]

[1,2,3,4] Department of Statistics, Universitas Bengkulu, Indonesia

*walestari@unib.ac.id

Abstract. The Global Happiness Index is widely used to measure countries' well-being across social, economic, and health-related dimensions. The complex and non-linear relationships among these dimensions often limit the predictive performance of conventional linear regression models. This study aims to evaluate the effectiveness of Support Vector Regression (SVR) with a Radial Basis Function (RBF) kernel in predicting Global Happiness Index scores. The study used data from 158 countries obtained from Kaggle, including GDP per capita, social support, healthy life expectancy, freedom, trust in government, and generosity as predictor variables. Data preprocessing was performed before splitting the dataset into training and testing sets, and the optimal SVR parameters were determined using Grid Search with K-fold cross-validation. The optimal SVR-RBF model produced an RMSE of 0.4462 and an MAE of 0.3829 on the testing data. In addition, the model achieved an R^2 value of 0.8328, indicating that it explained 83.28% of the variation in Global Happiness Index scores. These results suggest that SVR with an RBF kernel is an effective approach for modeling complex nonlinear relationships and can be used as a reliable tool for predicting national happiness levels.

Keywords: Global Happiness Index, Support Vector Regression, Radial Basis Function, Grid Search, Machine Learning.

1 Introduction

Happiness is an important aspect of human life that is closely related to physical, mental, and social well-being. Various factors, such as economic conditions, health, social support, freedom, and the surrounding environment, influence the level of happiness of individuals or societies. Therefore, measuring happiness is important for evaluating the quality of life in society and for supporting the formulation of policies oriented toward public welfare. One of the indicators widely used to measure the level of happiness is the World Happiness Index, which describes the level of societal well-being across countries based on social and economic factors.

The relationship between socio-economic factors and happiness levels is often complex and nonlinear. Factors such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption may interact to influence a country's level of happiness. Previous research using linear regression showed that the model explained variation in happiness levels with an R-squared value of 0.7737 [1]. However, the remaining prediction errors indicate that the linear model has not fully captured the relationship patterns among the variables. In addition, linear regression methods such as Ordinary Least Squares (OLS) generally assume linear relationships among variables. They are sensitive to outliers, leading to a decline in performance when applied to data with complex relationship patterns [2]. Therefore, a method capable of capturing nonlinear relationships and possessing good generalization ability is needed to improve prediction accuracy.

Various machine learning approaches have been used to model and predict the global happiness index. Previous studies have applied linear regression, Random Forest, XGBoost, Neural Network, and Multivariate Adaptive Regression Splines (MARS) methods to capture the relationship between socio-economic factors and happiness levels. Recent research [3] showed that machine learning methods are better at capturing complex interactions among variables than traditional linear methods. Random Forest and XGBoost are known to be strong at modeling multidimensional relationships in the happiness index. In addition, research by [4], using the MARS method, demonstrated that a nonlinear approach improved predictive performance compared to conventional linear regression.

In Support Vector Regression (SVR), several types of kernels are commonly used, including linear, polynomial, and RBF, which map data into higher dimensions to capture nonlinear relationships. Research by [5] applied SVR with an RBF kernel to predict nonlinear patterns in data, and the results showed that the method could model complex relationships among variables with good accuracy. The application of SVR with linear and polynomial kernels to predict the exchange rate of the Indonesian Rupiah against the United States Dollar by [6] showed that the SVR model was able to produce predictions with a very high level of accuracy based on MAPE and (R^2) metrics on the tested data.

Although various studies have applied machine learning approaches to predict the happiness index, the use of SVR with a RBF kernel to predict global happiness scores remains relatively limited. In addition, the performance evaluation of SVR in capturing nonlinear relationships among socio-economic factors has not been widely discussed in depth. Previous studies have generally focused more on methods such as Random Forest, XGBoost, and MARS, while the optimization of SVR parameters, such as cost (C), gamma (γ), and epsilon (ε) Grid Search and K-Fold Cross-Validation have rarely been used for predicting global happiness indices. In fact, the selection of appropriate parameters significantly affects the SVR model's ability to produce accurate and stable predictions.

Given these issues, this study aims to apply SVR with a RBF kernel to predict global happiness scores from socio-economic factors. In addition, this study optimizes model parameters using Grid Search and K-Fold Cross Validation to identify optimal settings and improve predictive performance. The model performance is evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). This study is expected to contribute to the development of machine learning-based methods for predicting happiness indices, particularly by applying SVR with an RBF kernel to capture nonlinear relationships among socio-economic variables. Furthermore, the results of this study are also expected to serve as a reference for the development of community welfare prediction models and to support data-driven policymaking at the global level.

2 Research Methods

The analysis was conducted in R using SVR with a RBF kernel to predict the Global Happiness Index from explanatory features. The stages of the research are as follows:

1. Data source and variables.
2. Descriptive statistics.
3. Data preprocessing.
4. Splitting data into training and testing data.
5. Hyperparameter optimization using Grid Search with K-Fold cross-validation.
6. Modeling the testing data using Ordinary Least Squares and SVR.
7. Model performance evaluation using RMSE, MAE, and R^2.
8. Prediction analysis on testing data.
9. Visualization of actual and predicted values.
10. Permutation feature importance analysis.
11. Residual validation tests.

This study uses secondary data from the 2015 World Happiness Report dataset obtained from Kaggle.com. The dataset comprises 158 countries worldwide and is cross-sectional, describing countries' happiness conditions at a single point in time. The target variable is the Happiness Score (Global Happiness Index). In contrast, the predictor variables include GDP per capita, family social support, life expectancy, individual freedom, trust in government, and generosity level.

2.1 Preprocessing Data

The data preprocessing stage is an important step before applying machine learning algorithms. The purpose of this stage is to transform raw data into cleaner features ready for use in the modeling process. The preprocessing process involves several steps. The first step is data cleaning, which involves identifying and filling missing values, detecting and resolving inconsistencies, removing noisy data or outliers, identifying duplicates, and handling incomplete data. This data cleaning process can significantly impact data mining, as the amount of data may decrease [7].

The next process in this stage is data standardization using the Z-Score Normalization method. This method is used to equalize the scales across variables, making prediction results more accurate and fair. Data standardization is performed by subtracting each value from the variable mean and then dividing by the standard deviation [8].

2.2 Training and Testing Data

In machine learning, data splitting is commonly used to partition a dataset into two main components: training data and testing data [9]. Training data is the data used to train and learn the model, while testing data is used to evaluate the model's performance. The proportion of data split is one of the factors that determines accuracy; therefore, the composition of both datasets must be appropriate to improve accuracy and optimally represent model performance [10]. A data classification model is built using the training dataset, and its classification performance is then measured based on the testing dataset. Common proportions for dividing training and testing data are 80:20 (80% training, 20% testing) and 50:50 (50% training, 50% testing). Optimal classification results depend on the training data; if the training data can cover most of the patterns required in the testing process, the results will be better [11].

In this study, an 80:20 data split was used, with 80% of the data used for training and 20% for testing. This proportion was chosen because it provides more training data, allowing the model to learn data patterns more effectively while still maintaining a representative testing dataset for evaluating the performance of the SVR model.

2.3 Hyperparameter Optimization Using Grid Search

The optimal hyperparameters are typically identified using the Grid Search approach combined with k-fold cross-validation [12]. This validation technique is widely employed to assess model performance and involves several sequential procedures [5].

1. Dividing the data into (k) subsets of equal size.
2. Using $(k$-1) subsets as training data and one subset as testing data.
3. Repeating this process (k) times for each combination of training and testing data.

In this study, K-Fold Cross Validation testing was conducted using several values of (k), namely k= 3, 5, 7, and 10. The selection of these (k) values was intended to obtain a validation configuration that produces the smallest prediction error. Furthermore, the parameter optimization process was carried out using the Grid Search method, a machine learning technique for determining the best combination of model parameters. In this method, several candidate values are first determined for each parameter to be optimized. Afterward, Grid Search evaluates all possible parameter combinations from the predefined values to obtain the parameter configuration that yields the optimal model performance [13].

2.4 SVR with RBF kernel

SVR is an application of Support Vector Machine (SVM) used for regression cases [14]. SVR works by providing an error tolerance limit called ε. Errors are still accepted if the difference between the predicted result and the actual value remains within the ε margin [15]. The SVR method focuses on finding the optimal hyperplane by minimizing errors from the training data and the insensitive loss function, thereby producing continuous real-valued outputs. In this study, the hyperparameters used are Cost (C), gamma (γ), and epsilon (ε) [16].

The main advantage of SVR is that its computational complexity does not depend on the dimensionality of the input space, and it has excellent generalization ability with high prediction accuracy [17]. The regression function of the SVR method is as follows [5]:

$$f(x) = w\varphi(x) + b \tag{1}$$

where w is the weight vector, φ(x) is a function that maps x into a certain dimension, and b represents the bias. The coefficients w and b function to minimize the risk function, as shown in Equation (2) below [18]:

$$R = \min \frac{1}{2}\|w\|^2 + C\frac{1}{l}\left(\left(\sum_{i=1}^{l} L_\varepsilon(y_i, f(x_i)\right)\right) \tag{2}$$

that has limitations

$$y_i - w\varphi(x_i) - b \leq \varepsilon$$

$$w\varphi(x_i) - y_i + b \leq \varepsilon,$$

with $i = 1, 2, 3, \ldots., l$. The loss function used is the ε-insensitive loss function, which is defined as follows:

$$L_\varepsilon(y_i, f(x_i) = \begin{cases} 0 & for \quad |y_i - f(x_i)| < \varepsilon \\ |y_i - f(x_i)| - \varepsilon & for \quad |y_i - f(x_i)| \geq \varepsilon \end{cases} \tag{3}$$

In solving nonlinear function problems, data can be mapped into a higher-dimensional space, called kernel space. Suppose there are n training data points, (x_i, y_i) where $i = 1, 2, \ldots, n$, with input data $x = \{x_1, \ldots, x_n\} \in R^n$ and $y = \{y_1, \ldots, y_n\} \in R$, and n is the number of training data points [19]. In general, the SVR function using a kernel function approach is given by the following equation [20]:

$$y = \sum_{i=1}^{n}(a_i - a_i^*).K(x_i, x) + b \tag{4}$$

The working mechanism of nonlinear data is determined by the type of kernel function and its parameter settings. In this study, the RBF kernel function is used and is expressed as follows [14].

$$K(x_i, x) = exp\,\{-\gamma\|x - x_i\|^2\} \tag{5}$$

where γ is a kernel parameter that controls the influence of the distance between data points.

2.5 Model Evaluation

Several evaluation metrics can be used to assess the accuracy and predictive performance of a model. In this study, RMSE, MAE, and R^2 were used, with the following formulas:

1. *Root Mean Square Error* (RMSE) is used to measure the average prediction error of the model based on the square root of the mean squared differences between the actual values and the predicted values. A smaller RMSE value indicates that the model's predictions are closer to the actual values. The RMSE formula is expressed as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

2. Mean Absolute Error (MAE) is used to measure the average absolute value of the prediction errors between the actual values and the predicted values of the model. A smaller MAE value indicates a lower level of prediction error.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

3. The coefficient of determination (R^2) is used to measure the model's ability to explain the variation in the actual data. An (R^2) value that is close to 1 indicates that the model has better predictive performance.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

Where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values, and n is the number of observations. In this study, the best model is selected based on the smallest RMSE and MAE values and the largest (R^2) value, to obtain a model with optimal predictive performance.

3 Results and Discussion

The analysis began with a preprocessing stage to ensure data quality. The examination results showed no missing values across all country observations. Subsequently, descriptive statistics were computed to characterize the dataset's initial characteristics, as presented in Table 1.

Table 1. Descriptive statistics

Features	Mean	Standard deviation	Minimum	Maximum
Happiness.Score	5.38	1.15	2.84	7.59
Economy	0.85	0.40	0	1.69
Family	0.99	0.27	0	1.40
Health	0.63	0.25	0	1.03
Freedom	0.43	0.15	0	0.67
Trust	0.14	0.12	0	0.55
Generosity	0.24	0.13	0	0.80

Based on Table 1, the Economy feature exhibits the greatest heterogeneity, with a standard deviation of 0.40. Furthermore, there are significant scale differences among features, underscoring the need for a data standardization step before SVR modeling. Overall, the Family and Economy variables are the features with the largest average contribution to the global Happiness Score. The data standardization process was performed using Z-score normalization to align the scales of all explanatory features to a common range. Subsequently, the dataset was partitioned into training and testing sets in an 80:20 ratio, resulting in 128 countries in the training set and 30 in the testing set.

SVR modeling with a RBF kernel began with a search for the optimal parameter values, consisting of the cost (C), gamma (γ), and epsilon (ε) parameters. The cost parameter controls the penalty level for prediction errors, the epsilon parameter determines the error tolerance threshold, and the gamma parameter regulates the range of data influence within the RBF kernel. Various combinations of these parameters were tested using Grid Search with K-fold cross-validation, and the results were compared to achieve optimal prediction performance. The following are several parameter combinations that produced the best model performance.

Table 2. Top 10 parameter combinations from grid search results

K_Fold	Cost (C)	Gamma (γ)	Epsilon (ε)	Error
5	1	0.01	0.05	0.322395
5	10	0.01	0.05	0.335528
5	50	0.01	0.05	0.375807
5	1	0.05	0.05	0.348169
5	1	0.10	0.05	0.395953
5	1	0.01	0.10	0.333948
5	1	0.01	0.20	0.343173
3	1	0.01	0.05	0.326592
7	1	0.01	0.05	0.326273
10	1	0.01	0.05	0.326905

Based on Table 2, the sensitivity analysis indicates that the SVR-RBF model is significantly influenced by the gamma (γ) and *cost* (C) parameters. An increase in the gamma (γ) value from 0.01 to 0.10 leads to higher error rates, suggesting that a broad range of support vector influence is crucial for capturing global data patterns. Meanwhile, increasing the *cost* (C) values up to 50 tend to decrease model performance due to overfitting. The stability of error values across K-Fold schemes (3, 7, and 10) indicates that the model is consistent. This confirms that the selected optimal parameter combination cost (C), gamma (γ), and epsilon (ε) of 1, 0.01, and 0.05, respectively, with K-fold = 5 exhibits high reliability.

To test the effectiveness of the proposed SVR-RBF model, a performance metric comparison was conducted against the Ordinary Least Squares (OLS) linear regression model. This comparison aims to determine whether the non-linear approach using a kernel function yields a significant increase in accuracy over the standard linear approach. The results of the performance metric comparison on the testing data are presented in Table 3.

Table 3. Performance comparison between OLS and SVR-RBF models

Model	RMSE	MAE	R^2
OLS	0.47970	0.386262	0.825038
SVR RBF	0.46624	0.382984	0.833527

Based on Table 3, the SVR-RBF model demonstrates competitive performance relative to the OLS model, although the difference between the two is not substantial. This indicates that the relationship patterns in the Global Happiness Index data are predominantly linear; however, a small non-linear variance component can only be captured by the SVR-RBF mechanism. Once the SVR-RBF model was declared optimal based on evaluation metric testing, a predictive analysis was performed on the test data to assess the model's ability to estimate happiness index values.

Table 4. Happiness index prediction results

No	Country	Actual	Predicted	Residual	Error
1	Switzerland	7.587	7.229515	0.357485	0.357485
2	Iceland	7.561	6.912221	0.648779	0.648779
3	Sweden	7.364	7.107523	0.256477	0.256477
4	Ireland	6.940	7.032535	-0.092540	0.092535
5	Belgium	6.937	6.725176	0.211824	0.211824
6	Oman	6.853	6.514749	0.338251	0.338251
7	Qatar	6.611	6.989193	-0.378190	0.378193
8	Kuwait	6.295	6.548709	-0.253710	0.253709
9	Trinidad and Tobago	6.168	5.940468	0.227532	0.227532
10	Bahrain	5.960	6.405311	-0.445310	0.445311
11	Bolivia	5.890	5.209017	0.680983	0.680983
12	Nicaragua	5.828	5.711625	0.116375	0.116375
13	Peru	5.824	5.310221	0.513779	0.513779
14	Cyprus	5.689	5.669156	0.019844	0.019844
15	Bhutan	5.253	5.616466	-0.363470	0.363466
16	Jordan	5.192	5.473358	-0.281360	0.281358
17	Montenegro	5.192	5.065281	0.126719	0.126719
18	Bosnia and Herzegovina	4.949	4.805290	0.143710	0.143710
19	Dominican Republic	4.885	5.883956	-0.998960	0.998956
20	Bangladesh	4.694	4.094600	0.599400	0.599400
21	Ghana	4.633	4.312033	0.320967	0.320967
22	India	4.565	4.104497	0.460503	0.460503
23	Haiti	4.518	4.194712	0.323288	0.323288
24	Congo (Kinshasa)	4.517	3.994650	0.522350	0.522350
25	Mauritania	4.436	4.394856	0.041144	0.041144
26	Bulgaria	4.218	5.413409	-1.195410	1.195409
27	Yemen	4.077	4.238698	-0.161700	0.161698
28	Mali	3.995	4.471594	-0.476590	0.476594
29	Senegal	3.904	4.618650	-0.714650	0.714650
30	Madagascar	3.681	3.899214	-0.218210	0.218214

Based on Table 4, the prediction results closely approximate the actual values; for instance, Switzerland has an actual value of 7.587 and a predicted value of 7.229, while Sweden has an actual value of 7.364 and a predicted value of 7.107. These small discrepancies indicate the SVR model's ability to capture the relationship patterns between the explanatory features and the Global Happiness Index. Nevertheless, certain countries exhibit an error value exceeding 1.0, such as Bulgaria. This indicates the presence of unique characteristics in that country's data that fall outside the ε-insensitive band, thus not being fully captured by the RBF kernel function. The predictive performance of the SVR-RBF model on the test data is visualized in Figure 1, which shows a scatter plot of actual versus predicted values.

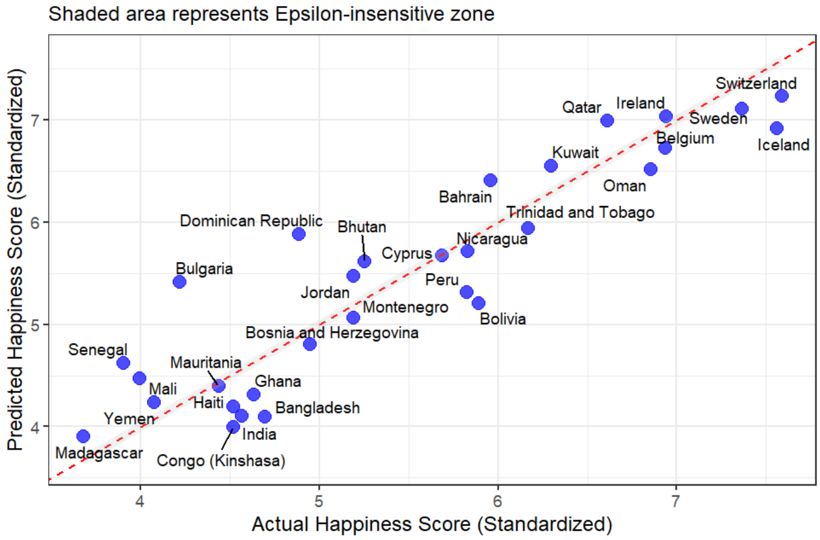

Fig 1. Plot comparing actual and forecast values.

Figure 1 presents a comparison plot of the actual and predicted standardized happiness scores. Based on the visualization, most country observations are tightly clustered and follow the dashed diagonal line, which represents the ideal prediction condition. The distribution of points dominating the area around the shaded region indicates that the SVR-RBF model possesses high precision in capturing data patterns. Although several countries, such as Bulgaria and the Dominican Republic, lie slightly outside the tolerance zone, most countries with high happiness scores, such as Switzerland and Iceland, as well as those with low scores, such as Madagascar, were successfully estimated with good accuracy.

Next, to assess a feature's importance in predicting happiness scores, a feature-importance analysis was conducted using the permutation feature-importance method. This method evaluates the contribution of each variable by measuring the increase in the model's prediction error after a variable is randomly permuted, thereby severing its relationship with the target. The greater the resulting increase in error, the more important that feature is in the SVR-RBF prediction model [21]. Based on the analysis results, a visualization of the importance levels of the predictor variables is presented in the following figure.

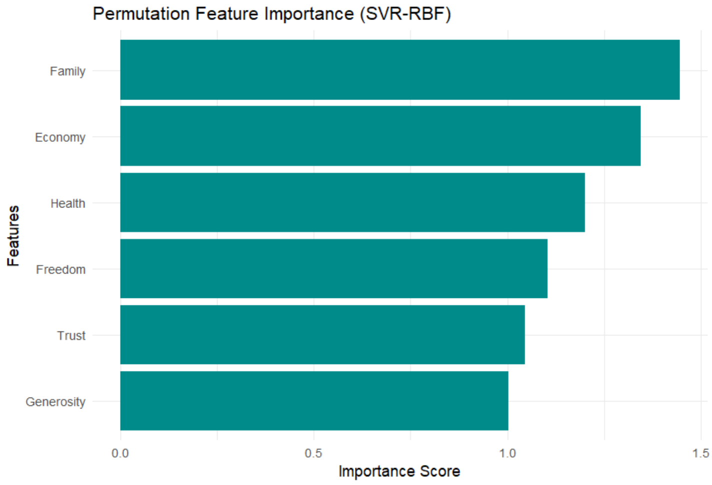

Fig 2. Plot of the importance levels of explanatory features

Figure 2 shows that Family is the feature with the greatest influence on the Happiness Score, followed by Economy, Health, Freedom, Trust, and Generosity. This finding indicates that social support plays a more dominant role in shaping happiness levels across countries compared to purely economic factors. Countries with stronger family and social relationships tend to exhibit higher levels of life satisfaction because social support contributes to emotional security, social stability, and overall well-being. Meanwhile, Economy and Health also demonstrate substantial influence, suggesting that financial stability and access to good health conditions remain important components in improving quality of life. However, the lower contribution of Freedom, Trust, and Generosity implies that these factors may function as complementary rather than primary determinants of happiness at the global level. In addition, the relatively balanced importance distribution across several variables suggests that happiness is a multidimensional phenomenon influenced by both social and economic conditions. These results are consistent with the descriptive statistics, which revealed variations among predictor variables and justified the need for data standardization before modeling.

To ensure that the prediction errors meet the basic statistical assumptions, a residual analysis was conducted. The residual analysis included a normality test to examine the distribution and a homoscedasticity test to ensure the stability of the variance. The test results are presented in Table 5.

Table 5. Residual validation test

Type of test	Test statistic	*p-value*
Normality	0.1213	0.7244
Homoscedasticity	0.9167	0.5009

Based on Table 5, the p-values obtained for the normality test and the homoscedasticity test were 0.7244 and 0.5009, respectively. Since both values exceed the significance level of 0.05, the residuals of the SVR-RBF model satisfy the assumptions of normality and homoscedasticity. This indicates that the resulting model is statistically valid. It also indicates that the residuals are random, show no systematic patterns, and contain no spatial effects or specific spatial patterns between countries.

4 Conclusion

In conclusion, the SVR method with a RBF kernel demonstrates competitive performance in modeling the relationship between socio-economic features and the Global Happiness Index. The Grid Search process with K-fold cross-validation successfully identified the optimal parameter combination cost (C), gamma (γ), and epsilon (ε), respectively, 1, 0.01, 0.05, with k-fold = 5, yielding an R^2value of 0.8328, indicating that the model can explain 83.28% of the data variance with low error rates.

However, this study is subject to several limitations. First, the model's effectiveness is highly dependent on hyperparameter tuning, and the predefined search space in Grid Search may not capture a broader global optimum. Second, the analysis is limited to six explanatory variables from the World Happiness Report, which may not fully represent all multidimensional factors affecting happiness. Therefore, future research is encouraged to incorporate additional relevant variables, such as political stability and income inequality, and to explore advanced meta-heuristic optimization techniques. This study is expected to serve as a reference for developing machine learning-based welfare prediction models.

References

[1] K. Hamidah and A. Voutama, “Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier ,” *Explore IT : Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 15, no. 1, pp. 1–7, Dec. 2023, doi: 10.35891/explorit.v11i2.1793.

[2] Maryam, M. Nosheen, S. Ahmad, and S. Naz, “Ordinary Least Square (OLS) Methods and Issues: Theoretical Foundations, Challenges, and Methodological Remedies,” *Pakistan Journal of Social Sciences Review*, no. 2, pp. 2959–8044, 2025, [Online]. Available: https://pjssr.com.pk/index.php/Journal/issue/archive

[3] B. Yang and X. Xie, “Analyzing and predicting global happiness index via integrated multilayer clustering and machine learning models,” *PLoS One*, vol. 20, no. 4, pp. 1–20, 2025.

[4] Ulfitasari, A. Yonita, Siswanto, and A. Kalondeng, “Pemodelan Indeks Kebahagiaan Negara Dengan Metode Multivariate Adaptive Regression Spline,” *MATHunesa: Jurnal Ilmiah Matematika*, vol. 13, no. 1, pp. 209–216, 2025.

[5] Isnaeni, Sudarmin, and Z. Rais, “Analisis Support Vector Regression (Svr) Dengan Kernel Radial Basis Function (Rbf) Untuk Memprediksi Laju Inflasi Di Indonesia,” *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 4, no. 1, pp. 30–38, 2022, doi: 10.35580/variansiunm13.

[6] R. Amanda, H. Yasin, and A. Prahutama, “Analisis Support Vector Regression (Svr) Dalam Memprediksi Kurs Rupiah Terhadap Dollar Amerika Serikat,” vol. 3, no. 4, pp. 849–857, 2014, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[7] M. Y. Putra and D. I. Putra, “Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI,” *Jurnal TEKNO KOMPAK*, vol. 16, no. 2, pp. 176–187, 2022.

[8] M. D. Sianturi, M. C. Lubis, S. D. Payana, A. N. Putri, and H. R. A. Panjaitan, “Analisis Pengelompokan Jenis Kejahatan di Sumatera Utara Berdasarkan Pola Kejadian Tahunan Menggunakan Algoritma K Means Clustering,” *Jurnal Sains Ekonomi dan Edukasi* , vol. 2, no. 6, pp. 1308–1324, 2025.

[9] T. Gori, A. Sunyoto, and H. Al Fatta, “Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[10] B. N. Azmi, A. Hermawan, and D. Avianto, “Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver,” *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 281–290, 2023, doi: 10.35746/jtim.v4i4.298.

[11] R. Darmawan and S. Amini, “Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Twitter,” Jakarta: Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI), 2022.

[12] A. W. Ishlah, Sudarno, and P. Kartikasari, “Implementasi Gridsearchcv Pada Support Vector Regression (SVR) Untuk Peramalan Harga Saham (Studi Kasus: harga Saham PT Anabatic Technologies Tbk),” *Jurnal Gaussian*, vol. 12, no. 2, pp. 276–286, 2023, doi: 10.14710/j.gauss.12.2.276-286.

[13] R. A. Maulana, I. Permana, F. N. Salisah, T. K. Ahsyar, and M. Jazman, “Perbandingan Kernel Algoritma Support Vector Regression Terhadap Performa Prediksi Produksi Kelapa Sawit,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science* , vol. 5, no. 1, pp. 405–413, 2025.

[14] A. N. Safira, B. Warsito, and A. Rusgiyono, “Analisis Support Vector Regression (Svr) Dengan Algoritma Grid Search Time Series Cross Validation Untuk Prediksi Jumlah Kasus Terkonfirmasi Covid-19 Di Indonesia,” *Jurnal Gaussian*, vol. 11, no. 4, pp. 512–521, 2023, doi: 10.14710/j.gauss.11.4.512-521.

[15] Fahmi, SuhermanSuryati, Mawarni, and Hanafi, “Model Support Vector Regression (SVR) Berdimensi Tinggi dengan Pendekatan Fungsi Kernel Berbeda untuk Peramalan Harga Saham TLKM: Sebuah Pemodelan Deret

Waktu Selama Masa Pandemi Covid-19," *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, vol. 5, no. 2, pp. 44–49, 2020.

[16] M. R. Apriyadi, Ermatita, and D. P. Rini, "Hyperparameter Optimization of Support Vector Regression Algorithm using Metaheuristic Algorithm for Student Performance Prediction," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 144–150, 2023, [Online]. Available: www.ijacsa.thesai.org

[17] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress Open, 2015.

[18] R. E. Cahyono, J. P. Sugiono, and S. Tjandra, "Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 1, no. 2, pp. 106–116, 2019, [Online]. Available: www.siskaperbapo.com

[19] Ruliana, Z. Rais, Marni, and A. S. Ahmar, "Implementation of the Support Vector Regression (SVR) Method in Inflation Prediction in Makassar City," *ARRUS Journal of Mathematics and Applied Science*, vol. 4, no. 1, pp. 28–35, 2024, doi: 10.35877/mathscience2608.

[20] Trivusi, "Algoritma Support Vector Regression (SVR): Jenis SVM untuk Regresi."

[21] W. Zhou, Z. Yan, and L. Zhang, "A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction," *Sci. Rep.*, vol. 14, no. 5905, 2024.