Institut Riset dan Publikasi Indonesia (IRPI)
IJATIS: Indonesian Journal of Applied Technology and Innovation Science
Journal Homepage: https://journal.irpi.or.id/index.php/ijatis
Vol. 1 Iss. 2 August 2024, pp: 75-84
ISSN(P): 3032-7466 | ISSN(E): 3032-7474

Performance Evaluation of Machine Learning Algorithms in Predicting Global Warming: A Comparative Study of Random Forest, K-Nearest Neighbors and Support Vector Machine

Anisa Putri[1*], Refri Martiansah[2], Qhairani Frilla F. Safiesza[3], Muhammad Fahri Abduh[4]

[1,2,3]Department of Information Systems, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[4]Department of Psychology, Faculty of Science and Letters, Kastamonu University, Türkiye

Email: [1]12050326104@students.uin-suska.ac.id, [2]12050313004@students.uin-suska.ac.id, [3]12050322811@students.uin-suska.ac.id , [4]kuyos@kastamonu.edu.tr

Received Jan 7th 2024; Revised 16th Jun 2024; Accepted Jul 21th 2024
Corresponding Author: Anisa Putri

Abstract

Global Warming is a global warming phenomenon that has a significant impact on human health and the environment. This research aims to apply Machine Learning algorithms, namely the Random Forest algorithm, K-Nearest Neighbors (K-NN), and Support Vector Machines (SVM) in predicting global warming. First, global warming data downloaded from Kaggle via dataset is used as research material. Then, a global warming prediction model is built using this algorithm and then evaluated using criteria such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean squared error (RMSE), R2, and Confusion Matrix. Finally, based on the evaluation results, research confirms that the K-NN algorithm shows the best performance, with the highest R2 value and low prediction error compared to other algorithms, such as Random Forest which shows the lowest performance. In terms of classification, K-NN achieved the highest accuracy (96.55%) and excellent performance in the confusion matrix and classification report. Overall, the findings of this study emphasize the dominance of K-NN in this context, thereby providing a strong basis for selecting models for predicting global warming phenomena.

Keyword: Global Warming, K-Nearest Neighbors, Machine Learning, Random Forest, Support Vector Machine

1. INTRODUCTION

The main problem that has arisen from increasing global temperatures since the Industrial Revolution is a serious threat to the earth's ecosystem and environmental balance [1]. In a relatively short period, global temperatures have risen about one degree Celsius since 1880, creating major impacts on weather, ecosystems, and global climate patterns [2][3]. This phenomenon continues, and scientists express concern that global temperatures could increase by around 0.3 to 0.7 degrees Celsius by 2035 [1]. The threat of global warming not only includes an immediate increase in temperature, but also triggers various consequences such as rising water levels sea, intensification of natural disasters, and drastic changes to the earth's ecosystem [4]. These conditions demand an urgent response to understand the mechanisms behind climate change, reduce negative impacts, and adapt to inevitable changes [2]. Faced with this threat, the need to understand, reduce the impact of, and adapt to climate change becomes increasingly urgent [1][5].

In this context, the need to take advantage of technological advances, especially in the field of Machine Learning, becomes increasingly important [4][6]. The potential for global warming predictions using machine learning holds out hope for understanding the dynamics of climate change more accurately [2]. Analysis of historical data and factors influencing climate change form the basis of this approach, with the aim of not only predicting future global temperatures but also providing deep insight into the complexity and consequences of climate change [7][8]. Problems arising from climate change are not only limited to environmental aspects but also include significant social and economic impacts [1][5]. Therefore, a deep understanding of climate change is needed as well as concrete efforts in designing policies and actions that can help reduce the risks and negative impacts [1].

In research conducted by Malakouti in 2023, he used Machine Learning to predict global temperature changes. Algorithms such as K-Nearest Neighbors (K-NN), Extra Trees, Light Gradient Boosting Machine, Gradient Boosting, Bayesian Ridge, and Random Forest were used in the experiments, and Extra Trees showed the best performance [1]. E. Syed Mohamed et al. (2023) conducted research involving a comparison of three different algorithms Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Random Forest to predict anxiety stages. Among them, Random Forest shows an accuracy of 98.13% [9]. In research conducted by Babak Azari in 2022, he said Machine Learning (ML) techniques for forecasting time series data are getting better and better, especially when it comes to understanding climate change [6]. With so many new methods popping up, it's getting tougher to tell them apart. That's why we decided to test and compare the predictive power of different ML methods [10]. According to Anggista Oktavia Praneswara (2023), the results of comparing predictions from the three algorithms, namely K-NN, Random Forest, and SVM, can be concluded that K-NN has the lowest error value on test data, while Random Forest has more errors on test than K-NN and SVM. The highest error value compared to K-NN and Random Forest [11].

Based on the topic and previous research that has been presented. So, the aim of this research is not only to predict future global temperatures but also to provide in-depth insight into the dynamics of climate change [1][12][13]. The focus of this research is on global warming prediction analysis, with special emphasis on the Random Forest, K-NN and SVM. These three algorithms have their respective advantages in managing complex data and can provide valuable contributions to understanding the dynamics of global climate change [14][1].

Thus, this research is also expected to provide a better understanding of possible climate change scenarios in the future so that it can identify risks and open up opportunities for better and more effective solutions in dealing with the impacts of ongoing climate change and can produce prediction models that provide significant impacts [1][5]. contribution to understanding and reducing the impacts of global warming [15].

2. MATERIALS AND METHODS

This study aims to apply Machine Learning, namely Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (K-NN) in predicting global warming.

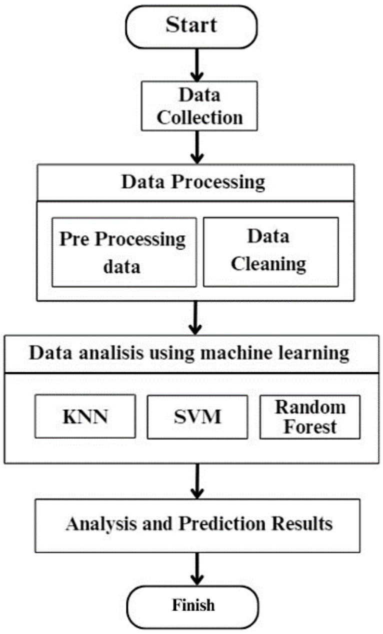

Figure 1. Research Method

2.1. Dataset

A dataset is a collection of data stored in digital format. Data has a key role in every Machine Learning [16]. These data sets can involve various types, such as images, text, audio, video, numeric data points, and so on [17][11][10]. The dataset used in this research is global warming data on Earth, namely a collection of data about warming temperatures on Earth downloaded from Kaggle.

2.2. Data Processing Stages

First, global warming data downloaded from Kaggle via dataset is used as research material. The data is then processed through the data processing stage, where the row containing the NAN value is deleted, and the data is converted into 2 columns, namely the Year and DN columns (average temperature per year). Next, the data is divided into 80 percent training data and 20 percent test data [18] [19].

After the data processing stage, data analysis is carried out using Machine Learning algorithms. The three algorithms used in this research are Random Forest, SVM, and K-NN. We used this algorithm to create a model that predicts global warming. To see how well it works, we tested it using a few different measures: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2), and the Confusion Matrix [20]. Thus, the data processing and analysis steps in this research include processing global warming data, building a prediction model using Machine Learning algorithms, and evaluating model performance based on the criteria mentioned.

2.3. Data Analysis Using Machine Learning

Machine Learning (ML) is a research field in a branch of science that combines ideas from various scientific disciplines[21], including artificial intelligence, statistics, information theory, mathematics, and so on [19]. Next, at this stage, we use prediction models, namely Random Forest, SVM, and K-NN.

1. K-Nearest Neighbors (K-NN)

K-NN observes the characteristics of data collected from training and test data. In general, K-NN calculates the distance between data points [22], using a simple Euclidean formula, as in Eq. Meanwhile shows a simple illustration of K-NN [23] Distance search formula using the Euclidean formula [24].

$$d(x, x_i) = \sqrt{\sum_{j=1}^{p} (x_j - x_{ij})^2} \tag{1}$$

In the context of K-NN, a crucial and commonly used formula is the Euclidean distance formula for computing the distance between a new data point and existing data points in the dataset. The Euclidean distance $d(x, x_i)$ is calculated as the square root of the sum of squared differences across each dimension between the new data x and the *i*-th data point in the dataset. This formula allows us to assess how close or far the new data point is from each existing data point, which is then used to determine the nearest neighbors. By applying this formula, K-NN predicts the class or target value of the new data based on the majority or average of the values from its nearest neighbors, depending on whether it is used for classification or regression tasks.

2. Support Vector Machine (SVM)

SVM are versatile tools that can be used for both predicting continuous values (regression) and categorizing data (classification) [25][26], although SVM is more commonly used in context classification. The main idea behind SVM is to find the best possible dividing line (called a hyperplane) in a multi-dimensional space (where each dimension represents a different feature of the data). This line should be able to accurately separate the different categories of data points [3]. The core idea of Support Vector Machines is to find a dividing line (called a hyperplane) that can neatly split the data into two groups. The way we calculate this line is shown in the formula below [27].

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i \mathrm{K}(x_i, x) + b \tag{2}$$

The formula (2) is used in SVM for makin prediction. α_i are weights indicating the importance of each data point in determining the hyperplane, y_i are the class labels of data points x_i and $\mathrm{K}(x_i, x)$ s the kernel function that measures the similarity between training data points x_i and a new data point x. This sum, added to the bias *b*, gives the value *f(x)*. The Value of *f(x)* is then used to predict the class of x, where a positive value indicates one class and a negative value indicates the other class.

3. Random Forest

Random Forest works by creating decision trees and combining them [28]. If used correctly, the RF algorithm is very useful for regression (numerical) and categorical data types [29]. RF is relatively easy to implement, requires a short training time, and can produce an accurate representation of the decision tree used [23]. Random forests work by building a bunch of decision trees. Each tree is made up of branches (internal nodes), starting points (root nodes), and final outcomes (leaf nodes). To create these trees, we randomly select features and data points. We use the formula in Equation 1 to calculate the entropy value, and Equation 2 to figure out the information gain [27].

$$F(x) = \frac{1}{J}\sum_{j=1}^{J} f_j\,(x) \tag{3}$$

The formula (3), $F(x)$ is the final prediction for the data point x, J is the number of decision trees in the Random Forest. $f_j(x)$ is the prediction of the j -th decision tree for the data point x.

2.4. Confusion Matrix

Confusion matrix as output and describes the complete performance of the model, show on figure 2.

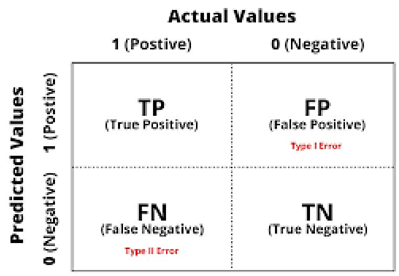

Figure 2. Confusion Matrix model

To calculate accuracy, can use the following formula 4.

$$\text{Accuracy} = \frac{TP + FN}{N} \tag{4}$$

Precision: Represents the sum of the divisions between actual positive outcomes divided by the number of predicted positive outcomes, the following formula 5.

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{5}$$

Recall: This is the division between true positive results and the results of all matching data, the following formula 6.

$$\text{Recall} = \frac{TP}{(TP + FP)} \tag{6}$$

Where,

TP	: True Positive
FP	: False Positive
FN	: False Negative
TN	: True Negative

3. RESULTS AND DISCUSSION

After testing using 3 algorithms, namely Random Forest, K-NN, and SVM, it will be seen which algorithm is better at predicting Global Warming. The following is a graphical depiction of a prediction plot using numbers. For residual plot analysis, can be seen in Figure 2 Global Warming Prediction Error Using K-Nearest Neighbor (a), Support Vector Machine (b), and Random Forest (c).

Figure 3 displays the Prediction Error of Global Warming Forecasting using K-NN. Based on the form assessment criteria, an R2 value of 0.92 was obtained. R2 provides the accuracy value of the algorithm for predicting regression results. The results obtained were 0.92 for R2 with this special algorithm indicating that this algorithm can predict Global Warming with a value of 0.92. This form of assessment obtained an R2 of 0.80. This algorithm has less strong performance than the K-NN algorithm in predicting Global Warming and can be seen in Figure 3.

Figure 4 Error Prediction Global Warming display Make predictions using the SVM algorithm. Based on the criteria in Figure 4, all errors were made, SVM, and Random Forest algorithm, on test data in predicting Global Warming is displayed to show that the relationship between the year and the average temperature each year is an effective pattern model. Then in Figure 5, all those errors in K-NN, SVM, and the Random Forest algorithm have The dataset provided for Global Warming prediction includes both training and testing phases.

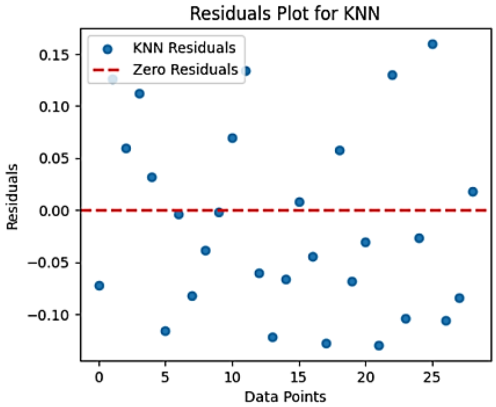

Figure 3. Residual Plot for K-NN

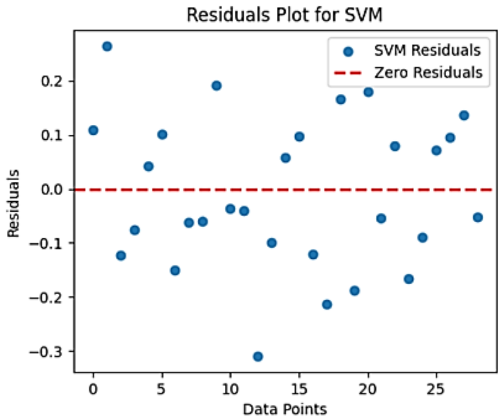

Figure 4. Residual Plot for SVM

Figure 5 displays Error Prediction Global Warming Make predictions using the Random Forest algorithm. Based on the shape assessment criteria, an R2 value of 0.88 was obtained. This algorithm has higher performance than the SVM algorithm but lower than the K-NN algorithm in predicting Global Warming.

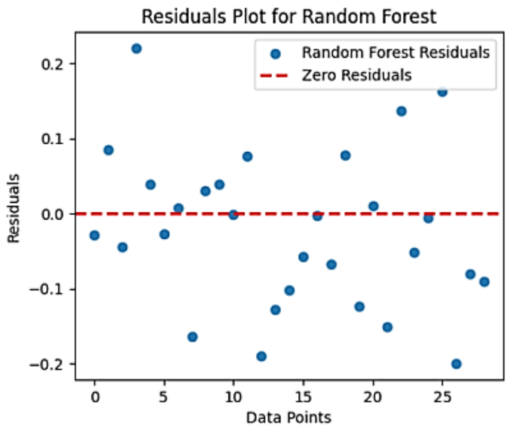

Figure 5 Residual Plots for Random Forest

Next in Figure 5, we will show how to compare the algorithm's capabilities on Train and Test data and also show in what range the algorithm operates showing the highest number of errors on both training and test data.

Figure 6 shows the results of using the K-NN method to predict global warming. When we tested the model on the training data, we got an R2 score of 0.95, meaning it was very accurate. This tells us the model learned well from the training data. We also tested the model on new data (test data) and got an R2 score of

0.92, showing it can predict global warming with good accuracy. Figure 6 specifically shows this R2 score of 0.92 for the test data, confirming the model's performance.

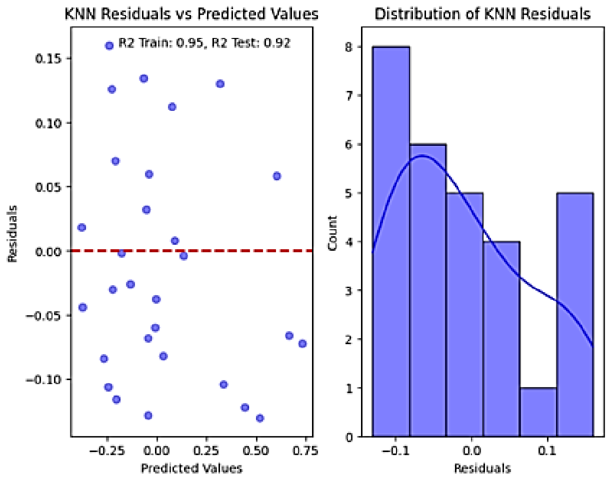

Figure 6 Residual Plot with K-NN

Figure 7 shows the results of using the SVM method to predict global warming. When we tested the model on the training data, we got an R2 score of 0.76, meaning it was fairly accurate. This tells us the model learned well from the training data. We also tested the model on new data (test data) and got an R2 score of 0.80, showing it can predict global warming with good accuracy. Figure 7 specifically shows this R2 score of 0.80 for the test data, confirming the model's performance.

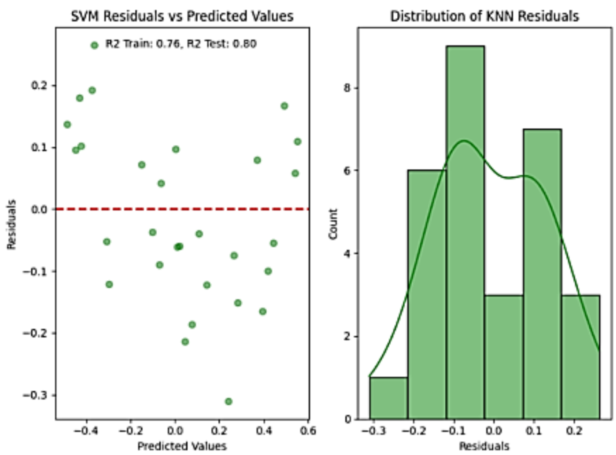

Figure 7. Residual Plot with SVM

Figure 8 displays a residual plot diagram for predicting global warming using the Random Forest method. R2 on the training data, the accuracy value reaches 0.99, and the test data value is 0.88. So it has been trained on training data with an accuracy of 0.88. In Figure 2 c, the R2 evaluation criteria for test data is 0.88. In this illustration, the evaluation criterion depicted in Figure 2 c is 0.88 which can be concluded. The accuracy of the simulation is confirmed.

Figure 9 below, the R2 evaluation criteria of the K-NN, SVM, and RF algorithms are compared and the algorithm with the best performance is obtained, namely using K-NN and the worst is the Support-Vector Machine.

Figure 10 below shows the comparison of time compared to the algorithm used. The smallest time that can be obtained between the algorithms used is SVM, namely 0.0048 seconds, and the largest time obtained with the K-NN algorithm is 0.2891 seconds.

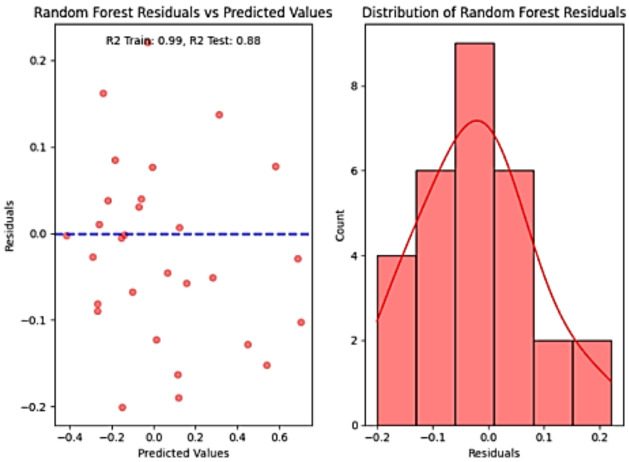

Figure 8. Residual Plot with Random Forest

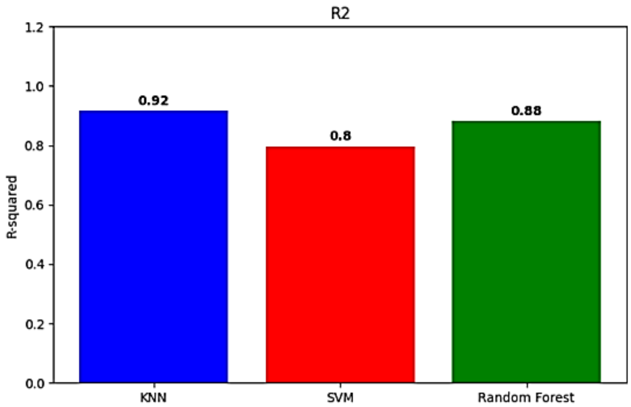

Figure 9. Evaluation of R2 on 3 Algorithms

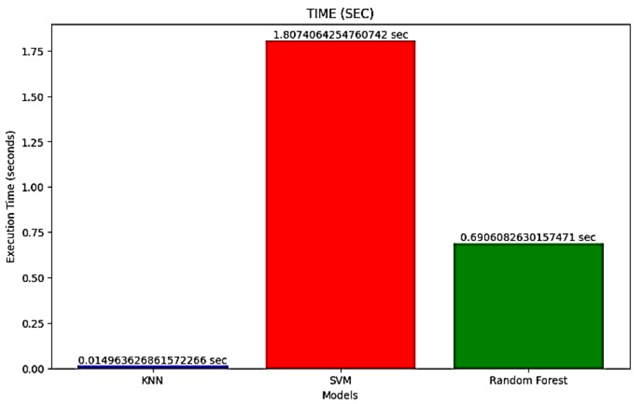

Figure 10. Comparison of Execution Time on 3 Algorithms

Machine learning algorithms and evaluation criteria were show table 1.

Table 1. Criteria Evaluation

Model	MAE	MSE	RMSE	R2	Time (sec)
K-NN	0.075517	0.007667	0.087562	0.916257	0.016797
SVM	0.118415	0.018600	0.136384	0.796841	1.836989
Random Forest	0.082779	0.010815	0.103995	0.881876	0.471458

From the table 2 comparing thesippp results of the three algorithms, it can be seen that the K-NN algorithm is superior in several evaluation criteria, such as MAE, MSE, and RMSE, with the smallest value, indicating the best performance in prediction. In the R2 criterion, the K-NN algorithm also achieved the highest score, confirming its superiority as the best algorithm. Apart from that, in terms of time, K-NN has the shortest duration. Overall, the model evaluation shows that the K-NN algorithm has the best performance, with the highest R2 and lower prediction error compared to other algorithms, such as Random Forest which shows the lowest performance. Graphical analysis also supports these findings, confirming that the K-NN algorithm is more reliable in predicting global warming. The confusion matrix results are based on data processing with the K-NN algorithm, show as table 2.

Table 2. K-NN Confusion Matrix

	True Positif	True Negatif	Class Precision
Pred. Positif	18	0	100%
Pred. Negatives	1	10	92%
Class Recall	94%	100%	

Looking at the table, we see that out of 30 years of data, the K-NN algorithm correctly predicted 18 years as negative. It also correctly predicted 10 years as positive. However, there was one year where it predicted a positive outcome but the actual result was negative. Overall, the K-NN algorithm was accurate 96.55% of the time.

The confusion matrix results are based on data processing with the SVM algorithm, show as table 3.

Table 3. SVM Confusion Matrix

	True Positif	True Negatif	Class Precision
Pred. Positif	15	3	94%
Pred. Negatives	1	10	77%
Class Recall	83%	91%	

Based on the table 3 of the 30 years of data, 15 years were predicted to be negative, so the correct result was negative, 3 years were predicted to be negative and the results were positive, and 13 years were predicted to be positive but showed that 1 year was predicted to be positive but the results were negative. and 10 years which are predicted to be positive with the appropriate results, namely positive as predicted. The accuracy results obtained using the SVM algorithm were 86.20%.

The confusion matrix results are based on data processing with the Random Forest algorithm show as table 4.

Table 4. Random Forest Confusion Matrix

	True Positif	True Negatif	Class Precision
Pred. Positif	16	2	89%
Pred. Negatives	2	9	82%
Class Recall	89%	82%	

Based on the table. From 30 years of data, 16 years are predicted to be negative, so the correct result is negative, 2 years positive negative predictions, and 11 years that are predicted to be positive but display 2 years predicted to be positive but the results were negative, and 9 years were predicted to be positive with the corresponding results, namely positive as predicted. The accuracy results obtained using the SVM algorithm were 86.20%.

In this study, three classification models, namely K-NN, SVM, and Random Forest were evaluated based on accuracy, confusion matrix, and classification report. The results show that K-NN has the highest accuracy of 96.55%, with a confusion matrix that reflects excellent performance (18 true positives for class 0 and 10 true positives for class 1). The K-NN classification report shows good precision, recall, and F1-score for both classes. On the other hand, SVM and Random Forest have similar accuracy (86.21%), with differences in the confusion matrix and classification reports. Although K-NN leads in accuracy, the final decision to choose a model depends on certain criteria, such as preference for detecting false positives or false negatives.

4. CONCLUSION

Based on the evaluation results, this research confirms that the K-NN algorithm shows the best performance in predicting global warming, with the highest R2 value and low prediction error. Graphical analysis also confirms K-NN's superiority in prediction ability. In terms of classification, K-NN achieved the highest accuracy (96.55%) and excellent performance in confusion matrix and classification report. Although SVM and Random Forest have comparable accuracy (86.21%), the model selection decision depends on the preference for detecting false positives or false negatives. Overall, the findings of this study emphasize the dominance of K-NN in this context, thereby providing a strong basis for model selection in predicting global warming phenomena. In addition, the conclusions of this study clearly answer the research objective, namely to identify the best algorithm for predicting global warming.

REFERENCES

[1] S. M. Malakouti, "Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a Global Temperature Change Prediction Model," *Case Stud. Chem. Environ. Eng.*, vol. 7, no. December 2022, p. 100312, 2023, doi: 10.1016/j.cscee.2023.100312.

[2] F. M. Viola, S. L. D. Paiva, and M. A. Savi, "Analysis of the global warming dynamics from temperature time series," *Ecol. Modell.*, vol. 221, no. 16, pp. 1964–1978, 2010, doi: 10.1016/j.ecolmodel.2010.05.001.

[3] W. C. Leong, R. O. Kelani, and Z. Ahmad, "Prediction of air pollution index (API) using support vector machine (SVM)," *J. Environ. Chem. Eng.*, vol. 8, no. 3, p. 103208, 2020, doi: 10.1016/j.jece.2019.103208.

[4] K. Fabrice and Y. Emmanuel, "Exploring the nexus of climate variability , population dynamics , and maize production in Togo : Implications for global warming and food security," *Farming Syst.*, vol. 1, no. 3, p. 100053, 2023, doi: 10.1016/j.farsys.2023.100053.

[5] H. Ghasemi-Mobtaker, A. Kaab, S. Rafiee, and A. Nabavi-Pelesaraei, "A comparative of modeling techniques and life cycle assessment for prediction of output energy, economic profit, and global warming potential for wheat farms," *Energy Reports*, vol. 8, pp. 4922–4934, 2022, doi: 10.1016/j.egyr.2022.03.184.

[6] O. Access, A. P. Mishra, and U. P. Singh, "Global warming prediction using machine learning," no. 07, pp. 1670–1675, 2021.

[7] A. Sanjurjo-de-No, A. M. Pérez-Zuriaga, and A. García, "Analysis and prediction of injury severity in single micromobility crashes with Random Forest," *Heliyon*, vol. 9, no. 12, 2023, doi: 10.1016/j.heliyon.2023.e23062.

[8] L. A. Demidova, I. A. Klyueva, and A. N. Pylkin, "Hybrid approach to improving the results of the SVM classification using the random forest algorithm," *Procedia Comput. Sci.*, vol. 150, pp. 455–461, 2019, doi: 10.1016/j.procs.2019.02.077.

[9] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, and A. Hussain, "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms," *Healthc. Anal.*, vol. 3, no. July 2022, p. 100185, 2023, doi: 10.1016/j.health.2023.100185.

[10] B. Azari, K. Hassan, J. Pierce, and S. Ebrahimi, "Evaluation of Machine Learning Methods Application in Temperature Prediction," *Comput. Res. Prog. Appl. Sci. Eng.*, vol. 8, no. 1, pp. 1–12, 2022, doi: 10.52547/crpase.8.1.2747.

[11] Anggista Oktavia Praneswara, "Perbandingan K-Nearest Neighbors, Support Vector Dan Random Forest Pada Prediksi Medical Cost," *Indones. J. Comput. Sci.*, vol. 12, no. 4, pp. 2035–2048, 2023, doi: 10.33022/ijcs.v12i4.3298.

[12] R. Hruška, M. Kmetík, and J. Chocholáč, "Selection of the Transport Mode Using the Ahp Method Within Distribution Logistics of Motor Fuels," *Promet - Traffic - Traffico*, vol. 33, no. 6, pp. 905–917, 2021, doi: 10.7307/ptt.v33i6.3940.

[13] L. Wang, L. Wang, Y. Li, and J. Wang, "A century-long analysis of global warming and earth temperature using a random walk with drift approach," *Decis. Anal. J.*, vol. 7, no. April, p. 100237, 2023, doi: 10.1016/j.dajour.2023.100237.

[14] C. J. Varshney, A. Sharma, and D. P. Yadav, "Sentiment analysis using ensemble classification technique," *2020 IEEE Students' Conf. Eng. Syst. SCES 2020*, pp. 12–17, 2020, doi: 10.1109/SCES50439.2020.9236754.

[15] W. Chen, L. Yi, J. Wang, J. Zhang, and Y. Jiang, "Evaluation of the livability of arid urban environments under global warming: A multi-parameter approach," *Sustain. Cities Soc.*, vol. 99, no. March, p. 104931, 2023, doi: 10.1016/j.scs.2023.104931.

[16] A. Deolika, K. Kusrini, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.

[17] T. Zulhaq Jasman, E. Hasmin, C. Susanto, and W. Musu, "Perbandingan Logistic Regression, Random Forest, dan Perceptron pada Klasifikasi Pasien Gagal Jantung," *CSRID J.*, vol. Vol. 14 No, no. 3, pp. 271–286, 2022.

[18] A. Satria, R. M. Badri, and I. Safitri, "Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning," *Digit. Transform. Technol.* |, vol. 3, no. 2, pp. 389–398, 2023.

[19] Y. I. Sulistya, "Analisis perbandingan Reduction Technique dengan metode Dimentional Reduction dan Cross Validation pada dataset Breast Cancer," *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 82–88, 2022, doi: 10.56705/ijodas.v3i2.41.

[20] Winarno, A. Prasetyo, and A. Wijayanto, "Decision Support System for Indonesia Smart Card (KIP)

Scholarship Selection Using The AHP And VIKOR Method Integrated with EKTP," *E3S Web Conf.*, vol. 448, p. 02035, 2023, doi: 10.1051/e3sconf/202344802035.

[21] G. Battineni, N. Chintalapudi, and F. Amenta, "Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM)," *Informatics Med. Unlocked*, vol. 16, no. May, p. 100200, 2019, doi: 10.1016/j.imu.2019.100200.

[22] V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 1314–1320, 2022, doi: 10.1016/j.procs.2023.01.110.

[23] V. A. Nugroho, D. P. Adi, A. T. Wibowo, M. T. Sulistyono, and A. B. Gumelar, "Klasifikasi Jenis Pemeliharaan dan Perawatan Container Crane menggunakan Algoritma Machine Learning," *Matics*, vol. 13, no. 1, pp. 21–27, 2021, doi: 10.18860/mat.v13i1.11525.

[24] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode K-NN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.

[25] K. C. Onyelowe, C. B. Mahesh, B. Srikanth, C. Nwa-David, J. Obimba-Wogu, and J. Shakeri, "Support vector machine (SVM) prediction of coefficients of curvature and uniformity of hybrid cement modified unsaturated soil with NQF inclusion," *Clean. Eng. Technol.*, vol. 5, p. 100290, 2021, doi: 10.1016/j.clet.2021.100290.

[26] C. Fan, X. Lai, H. Wen, and L. Yang, "Coal and gas outburst prediction model based on principal component analysis and improved support vector machine," *Geohazard Mech.*, no. November, 2023, doi: 10.1016/j.ghm.2023.11.003.

[27] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 578, 2021, doi: 10.32493/informatika.v5i4.7575.

[28] M. Nachouki, E. A. Mohamed, R. Mehdi, and M. Abou Naaj, "Student course grade prediction using the random forest algorithm: Analysis of predictors' importance," *Trends Neurosci. Educ.*, vol. 33, p. 100214, 2023, doi: 10.1016/j.tine.2023.100214.

[29] A. Gatera, M. Kuradusenge, G. Bajpai, C. Mikeka, and S. Shrivastava, "Comparison of random forest and support vector machine regression models for forecasting road accidents," *Sci. African*, vol. 21, p. e01739, 2023, doi: 10.1016/j.sciaf.2023.e01739.