Prediction of Foreign Tourist Arrivals in Bali Using Support Vector Machine Algorithm and Linear Regression

Arif Lukman Hakim, Indra Wibisono, Yuanita Octoria, Andi Nugroho

Abstract- The island of Bali, a famous tourist destination in Indonesia, has experienced an increase and decline in its tourism sector. To anticipate the arrival of international tourists, a thorough analysis and prediction approach is needed. The main objective of this study is to analyze and predict the arrival of foreign tourists in the coming year. The use of machine learning algorithms can be important in analyzing data and forecasting the development of international tourist arrivals. This research details the use of several machine learning algorithms, namely *Support Vector Machine (SVM),* and *Linear Regression*. The test results show that the SVM program evaluation provides accurate predictions with a low error rate in the test data. Meanwhile, the error rate in the evaluation of the Linear Regression program is slightly higher. However, the model was still able to make good predictions on the test data.

Keywords- Prediction, Tourism, Bali, *Support Vector Machine (SVM), Linear Regression*

I. INTRODUCTION

The tourism sector of Bali Island is one of the famous tourist destinations in Indonesia [1]. The island of Bali is the main destination for foreign tourists because of its stunning natural scenery, cultural richness and hospitality to the community. Natural wealth plays a crucial role in Indonesia's tourism sector. This country has a very large area, equipped with a variety of natural resources that have the potential to be exploited and utilized [2]. According to the Central Bureau of Statistics (BPS) Bali over the past three years, foreign tourist arrivals to the island of Bali have decreased and increased significantly, recorded from 1,069,473 people in 2020 to 51 people in 2021 and increased again in 2022 which was 282,457 foreign tourists. With a significant decrease and increase, analysis and prediction are needed to estimate the number of foreign tourist arrivals who will come to Bali [3].

One way to estimate the number of foreign tourist arrivals is to use the ETL method (*Extract, Transform, Load*). ETL methods are planned to execute three types of data transformations: value changes, rule-based transformations, and concept mapping [4]. Process *Extract-Transform-Load* (ETL) in data warehouse development has a key role in taking data from various sources, changing the data format according to needs, and filling it into data warehouse storage [5]. In this study, the ETL method was used to take data on foreign tourist arrivals according to their country of origin from trusted sources, then process it and make informative visualizations using algorithms to analyze and make predictions about future tourist arrivals.

Furthermore, the data that has been collected will be processed and visualized. Data visualization is an important component to help information recipients understand the results of data analysis, especially when the data to be analyzed has a very large volume [6][7]. The purpose of data visualization is to facilitate understanding of data, recognize patterns and trends contained in data, and also support the data-based decision-making process [8]. The visualization in this study will take from the data and will be classified or grouped based on the geography of the tourist country visiting Bali. The visualization that will be displayed will be in the form of a scatter plot. Then, the results of the predictions made will also be visualized into Matrix Evaluation. It is expected that the results of data visualization are able to provide concrete solutions to overcome existing problems. The use of data visualizations, such as responsive graphs and tables with the use of supportive colors, allows the available information to be quickly understood [9][10].

In addition, in data analysis and prediction of the development of foreign tourist arrivals, machine learning algorithms can be a very useful tool. In this study, two machine learning algorithms will be used, namely *Support Vector Machine* (SVM) and Linear Regression, to make predictions based on historical data and factors that influence tourist arrivals.

Through the incorporation of ETL methods, informative data visualization, and the use of SVM and Linear Regression algorithms, this study aims to analyze and predict the arrival of foreign tourists in the following year.

II. RESEARCH METHOD

This research was conducted using data on foreign tourist visits to the island of Bali from 2019 to 2022 obtained from *website* Central Bureau of Statistics of Bali Province (https://bali.bps.go.id/). Time series data itself comes from the level of patterns and changes caused by randomness and seasonal factors [11]. The flow of research conducted to analyze and predict the arrival of foreign tourists on the island of Bali is as follows.

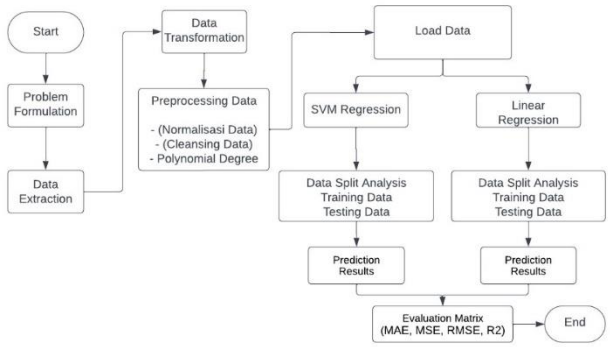

Figure 1. Research Flow

a. Problem Formulation (Perumusan Masalah)
Problem formulation is the first step that can be done by researchers in order to be able to identify problems and describe and determine the research problem to be studied. At this stage the formulation of the problem is carefully negotiated so that the direction of the research carried out is clear and specific.

b. Data Extraction
Data extraction is a process of taking data or information from the source either in the form of documents or other forms that will be analyzed based on the research conducted [12]. Researchers used data on foreign tourist visits to Bali from 2019 to 2022 obtained from *website* Central Bureau of Statistics of Bali Province (https://bali.bps.go.id/).

c. Data Transformation
Data transformation is important to do in preparing data before analysis, to ensure that the data is ready to be used in accordance with the expected analysis needs [13]. At this stage, the data that has been collected is combined and changes to the data structure are made to get the desired results.

d. Pre-processing Data
Data pre-processing is the process of preparing data for use in further analysis in order to ensure good data quality [14]. The main goal is to ensure that the data used in the classification process becomes more optimal. Preprocessing of research data is carried out in two stages, namely the data cleaning stage and the data transformation stage [15]. There are six components that must be met in this process: accuracy, completeness, consistency, timeliness, reliability, and the ability to be clearly interpreted [16]. Pre-Processing conducted in this study includes:

- Data normalization, the process of changing the scale or range of values of variables in a dataset to have a standard range of values [17]. The goal is to ensure that variables increase the accuracy of the algorithm used.

  StandardScaler normalizes it by changing the values in the dataset so that the mean of each feature becomes 0 and the standard deviation becomes 1. This makes the distribution of each feature have a uniform scale.
  Rumus: $z = \sigma x - \mu$
  x is the original value.
  μ is average
  σ is the standard deviation.

- *Cleansing* Data, the process of identification, data handling, and deleting incomplete data, redundant data or duplicate data, and irrelevant data [18]. There are six components that must be met in this process: accuracy, completeness, consistency, timeliness, reliability, and the ability to be clearly interpreted. The goal is to clean the data from data inaccuracies and inconsistencies [19]. If not done *cleansing* data, it will cause an error when *machine learning* Read the dataset. Errors in reading datasets cause *machine learning* cannot read multiple datasets, so the data becomes NaN [20].

e. Load Data
The data that has been extracted and transformed will then be analyzed. In this data load process, it ensures that the data needed for research analysis is available and well structured in the data storage [21].

f. Support Vector Machine (SVM)
The object of this study is forecasting the arrival of foreign tourists to the island of Bali, using an algorithm *Support Vector Machine* (SVM). SVM algorithms are one of the powerful classification and regression approaches in *machine learning*. SVM is designed to solve the problem of classification of two classes with the aim of finding *hyperplane* optimal [22][23].
Hyperplane It is a function used to separate two classes in the input space, thus allowing classification and regression analysis of the given data [24][25]. SVM can also address non-linear data by using the kernel, which allows the transformation of data into higher dimensions, where linear separators can be found [26]. The advantages of SVM include its ability to cope *overfitting*, good performance in datasets with a large number of features, and flexibility in handling different types of data [27]. Therefore, SVM is used in various applications such as in regression problem solving.

$$f(x) = w^T \varphi(x) + b\,,$$ [27][28]

Where w is the weighting vector, is a function that maps x into a dimension, and b is the refractive factor.$\varphi(x)$

g. Regresi Linear

Linear regression algorithm is one of the statistical methods used to predict outcomes and to find linear relationships between one or more independent and dependent variables of a data [29][30]. Regression analysis has at least three important benefits. First, it is used to describe the phenomenon of the data or case under investigation. Secondly, it is used as a means of control. Lastly, it is used to predict future outcomes [31]. With the aim of identifying correlations or relationships between these variables by using the most appropriate straight line to describe the relationship between these variables [32].

$$y = w_1 x_1 + b\,,$$ [33]

y = variable depend
x1 = variable independent
w1 = slope
b = intercept the intersection point of the regression line with the y-axis when *x=0*.

h. Evaluation Metrics

Evaluation metrics is a measurement process carried out to evaluate and measure performance and effectiveness, models *machine learning* in predicting or classifying data. The existence of several types of evaluation metrics that are commonly used, it allows researchers to select evaluation metrics that are in accordance with the specific problems and objectives of using the model or system used [34][35].
The goal is that researchers can assess how well the model used performs in making accurate predictions [36][37].

- *Mean Squared Error* (MSE)

MSE is a method for finding parameters that minimize the amount of residual squares between target and predicted values [33]. MSE can be used to calculate the difference in the square of the error between the predicted value and the actual value [36].

$$MSE = n1\sum i = 1\mathrm{n}(\mathrm{yi} - y\hat{}\mathrm{i})2\,,$$ [29]

n = number of observations
yi = observed value
y^i = value predicted by the model.

- *Mean Absolute Error* (MAE)

MAE is an evaluation metric used to average the absolute error between the predicted value and the actual value [38]. MAE can be calculated by the following formula:

$$MAE = n1\sum i = 1\mathrm{n}\left|\mathrm{yi} - \mathrm{y}^{\mathrm{i}}\right|\,,$$ [33]

n = number of observations
yi = observed value
y^i = value predicted by the model.

- *Root Mean Squared Error* (RMSE)

RMSE is the square root of MSE. RMSEs are often used to evaluate the performance results of the predictions used. Although rooting produces a measure equal to y, it does not affect the relative ratings of the models. This metric clearly describes ordinary errors or for normally distributed errors [39]. When the RMSE is low, it indicates that the change in value produced by the prediction model is significantly close to its original form. Conversely, if the RMSE is large enough, it indicates that the accuracy of the prediction model is further from its original form [29].

$$, RMSE = \sqrt{\left(\frac{1}{n}\right)} * \sum(y_1 - \bar{y}\,)^2$$ [29][40]

n = the amount of data used in the calculation.
yi = the actual value of the i-th data.
$\bar{y}$ = predicted value for the i-th data.
Σ = indicates the addition operation, which is performed for each data from i = 1 to n.

- *R-Squared* (R2)

R-Squared (R2) can be interpreted as a predictable comparison of the variance of the dependent variable with the independent variable. When the R^2 value approaches 1, it signifies that the model is able to explain most of the variation, while a lower value indicates a limitation in the model's ability to explain the variation. [41]

$$R2 = 1 - \sum i = 1n(yi - y\bar{}\,)2\ \sum i = 1n(yi - y\,\hat{}\mathrm{i})2\,,$$ [33]

yi = observed value
y^i = value predicted by the model
$y\bar{}$ = *mean* of the observed values.

III. Result and Analysis

This study focuses on analyzing and predicting the development of foreign tourist arrivals on the island of Bali using SVM (*Support Vector Machine*) and Linear Regression algorithms. In the testing process we used 60% of the training data as well as 40% of the test data. In SVM and Linear Regression programs we extract data from the variable "time", where we create new variables in the data "year" and "month". After that, normalize the data using min max scaling. Then we normalize the feature to make predictions for 2024 every month.

A. SVM Algorithm Results and Analysis

0.46888162438535325

Mean Absolute Error (MAE)	0.3570286470194035
Mean Squared Error (MSE)	0.21984997768624748
Root Mean Squared Error (RMSE)	0.46888162438535325

Table 1. SVM Evaluation Metrics Results

In this testing process produces a Mean Absolute Error (MAE) value of 0.3570286470194035, which indicates the average absolute error between the predicted value and the actual value in the test data. The test also produces a Mean Squared Error (MSE) value of 0.21984997768624748, which represents the average of the squared differences between predictions and actual values. In addition, this test process also calculates the result of the Root Mean Squared Error (RMSE) value with a result of 0.46888162438535325, which is the square root of MSE, and indicates the average error rate of the model in predicting the actual value of the test data using polynomial kernels and predetermined parameters. From the test results, the results for predictions in 2024 are as follows.

	Date	Predicted_Australia
0	2024-01-01	82069.677501
1	2024-02-01	86300.268395
2	2024-03-01	90710.113155
3	2024-04-01	95335.331694
4	2024-05-01	100212.043926
5	2024-06-01	105376.369763
6	2024-07-01	110864.429119
7	2024-08-01	116712.341907
8	2024-09-01	122956.228039
9	2024-10-01	129632.207428
10	2024-11-01	136776.399989
11	2024-12-01	144424.925633

Figure 2. SVM Prediction Results

We also attach the prediction results that have been obtained in the form of a graph as below.

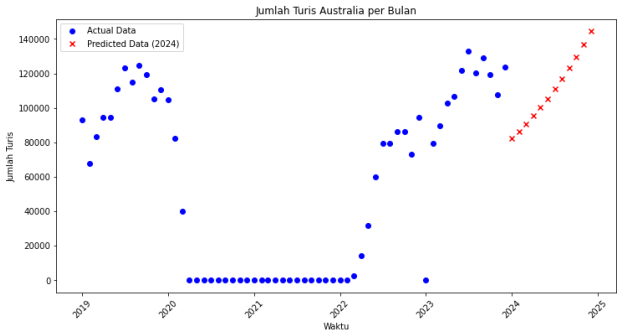

Figure 3. SVM Prediction Results Graph

Based on the graphic image above, it can be explained that the blue circle is the actual data used to test predictions in 2024. As for the red cross, it shows the prediction results that have been obtained from the tests that have been done.

B. Results and Analysis of Linear Regression Algorithms

0.42926572911591354*Support Vector Machine (SVM)* algorithm model, it was continued by testing using the Linear Regression Algorithm as a form of comparison of the two algorithms. Testing using the Linear Regression algorithm produces results as shown below.

Mean Absolute Error (MAE)	0.39294471916373813
Mean Squared Error (MSE)	0.18426906619341685
Root Mean Squared Error (RMSE)	0.42926572911591354

Table 2. Results of Linear Regression Evaluation Metrics

Tests conducted using this Linear Regression algorithm model resulted in an absolute average error value (MAE) of 0.39294471916373813, which shows the mean squared error between the predicted value and the actual value in the test data. In addition, the mean squared error (MSE) value reaches 0.18426906619341685, which indicates the mean squared error between the predicted value and the actual value. Next, the square root value of the mean error (RMSE) is also calculated, with a result of 0.42926572911591354. which is the square root of MSE. The average error rate the model shows in predicting actual values on test data in 2024 using a linear regression approach. The predicted results of the test are as follows.

	Date	Predicted_Australia
0	2024-01-01	56793.265591
1	2024-02-01	59583.634265
2	2024-03-01	62374.002940
3	2024-04-01	65164.371614
4	2024-05-01	67954.740288
5	2024-06-01	70745.108962
6	2024-07-01	73535.477637
7	2024-08-01	76325.846311
8	2024-09-01	79116.214985
9	2024-10-01	81906.583660
10	2024-11-01	84696.952334
11	2024-12-01	87487.321008

Figure 4. Linear Regression Prediction Results

The attachment of the prediction results that have been obtained in the form of a graph as shown below.

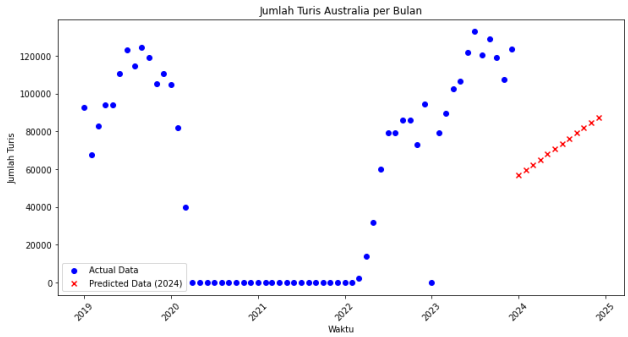

Figure 5. Linear Regression Prediction Results Graph

Just like the explanation in Figure 3. The SVM Prediction Results graph, which explains the blue circle, is the actual data used to test predictions in 2024. As for the red cross, it shows the prediction results that have been obtained from the tests that have been done.

However, when viewed from a distance, the number of Australian tourist arrivals to Bali between 2021 and 2022 has undergone a striking change. In 2021, there was a significant decrease in the number of Australian tourists visiting Bali. That's because factors such as international travel restrictions imposed due to the COVID-19 pandemic may have contributed to the decline. This drastic change highlights how the dynamics of international travel can be affected by a variety of factors, including the global health situation and travel policies imposed by destination countries. However, in 2022, there has been a marked increase in Australian tourist arrivals to Bali, signalling a potential recovery in the post-pandemic tourism industry.

IV. CONCLUSION

The conclusion that can be drawn from this are:

1. SVM programs evaluated under specific conditions, such as polynomial kernels, C=100, and epsilon=0.1, have satisfactory results in predicting test data with low error rates. As shown by the low values of Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), this model is able to provide accurate predictions of test data.
2. The evaluation results of the Linear Regression program in 2024 also show good performance; however, the Linear Regression model is still able to provide accurate predictions of the test data, albeit with a slightly higher error rate compared to the SVM model.

ACKNOWLEDGEMENTS

This research was supported by the Research Center of Mercu Buana University. We thank colleagues from the Faculty of Computer Science and the Central Bureau of Statistics (BPS) of Bali Province who have provided insight, expertise, and permission to use data that greatly helped this research.

REFERENCES

[1] I Wayan Budi Suryawan, Nengah Widya Utami, and Ketut Queena Fredlina, "Analysis of Sentiment of Tourist Reviews on Ubud Attractions Using Support Vector Machine Algorithm," *J. Inform. Techno. and Science*, vol. 5, no. 1, pp. 133–140, 2023, doi: 10.51401/jinteks.v5i1.2242.

[2] S. A. Salimu and Y. Yunus, "Prediction of Foreign Tourist Arrival Rate Using Backpropagation Method (Case Study: Mentawai Islands)," *J. Inform. Ekon. Business*, vol. 2, pp. 98–103, 2020, doi: 10.37034/infeb.v2i4.50.

[3] V. N. Putriwijaya, Novi Nur &; Wijayaningrum, "Support Vector Regression to Predict the Number of Foreign Tourist Visits on the Island of Bali," *Mach. Learn. Methods Appl. to Brain Disord.*, vol. 11, no. June, 123–140 (2019).

[4] Y. Yu *et al.*, "Developing an ETL tool for converting the PCORnet CDM into the OMOP CDM to facilitate the COVID-19 data integration," *J. Biomed. Inform.*, vol. 127, no. October 2021, p. 104002, 2022, doi: 10.1016/j.jbi.2022.104002.

[5] I. Hidayat, "Timely graduation prediction using svm algorithm and k-nearest neighbor based on particle swarm optimization in STIMK Eresha," *J. ESIT (E-Business, Sist. Inf., Teknol. Information)*, vol. 16, no. 01, pp. 55–65, 2021.

[6] W. Irmayani, "Data Visualization in Data Mining Using Classification Method," *J. Khatulistiwa Inform.*, vol. IX, no. I, pp. 68–72, 2021.

[7] S. Royal, "MAYA WIDYANTI'S LINEAR REGRESSION APPROACH," VOL. 1, NO. 1, pp. 10–18, 2024.

[8] R. Guntara, "Data Visualization of Online Store Sales Reports through Data Science Approach Using Google Colab," *J. Ilm. Multidisciplinary*, vol. 2, no. 6, pp. 2091–2100, 2023.

[9] W. I. Loka and F. Natalia, "Design and Making Data Visualization of Internal Research Fund and Higher Education Grant LPPM Universitas Multimedia Nusantara," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 10, no. 1, pp. 61–68, 2019, doi: 10.31937/si.v10i1.867.

[10] B. Priambodo *et al.*, "Predicting Employee Turnover in IT Industries using Correlation and Chi-Square

Visualization," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, pp. 71–75, 2022, doi: 10.14569/IJACSA.2022.0131210.

[11] N. Sulthan, N. Eltivia, and N. I. Riwajanti, "Implementation of Exponential Smoothing Method for Forecasting Foreign Tourist Arrivals on the Island of Bali," *Media Mahard.*, vol. 18, no. 2, pp. 171–178, 2020, doi: 10.29062/mahardika.v18i2.145.

[12] I. P. W. Prasetia and I. N. H. Kurniawan, "Implementation of ETL (Extract, Transform, Load) in Sales Data warehouse Using Pentaho Tools," *TIERS Inf. Technol. J.*, vol. 2, no. 1, pp. 1–8, 2021, doi: 10.38043/tiers.v2i1.2844.

[13] S. Widaningsih, "Comparison of Data Mining Methods for Prediction of Grades and Graduation Time of Informatics Engineering Study Program Students with C4.5, Naïve Bayes, Knn and SVM Algorithms," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[14] N. C. Gosari and R. Rismayani, "Application of Data Mining in Grouping Foreign Tourist Visits in Prov. South Sulawesi with K-Means and SVM," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 174–180, 2023, doi: 10.30591/jpit.v8i3.4554.

[15] G. Sanhaji and A. Febrianti, "DIATECT Application for Diabetes Disease Prediction Using Web-Based SVM," vol. 18, no. 1, pp. 150–163, 2020.

[16] H. Nuraliza, O. N. Pratiwi, and F. Hamami, "IMBd Film Review Dataset Sentiment Analysis Using Support Vector Machine (SVM) and Feature Importance Selection," *J. Mirai Manj.*, vol. 7, no. 1, pp. 1–17, 2022.

[17] L. Surtiningsih, M. T. Furqon, and S. Adinugroho, "Prediction of the Number of Foreign Tourist Visits to Bali Using Support Vector Regression with Genetic Algorithm," *J. Developer. Techno. Inf. and Computal Science.*, vol. 2, no. 8, pp. 2578–2586, 2018, [Online]. Available: http://j-ptiik.ub.ac.id

[18] S. Lolitha, S. Lukas, and F. Panduwinata, "Flight Delay Prediction Using Decision Tree Method for Determination of Travel Insurance Premiums," *Pros. Semin. Us. Teknoka*, vol. 5, no. 2502, pp. 50–58, 2020, doi: 10.22236/tekk.v5i.326.

[19] R. Wijaya and B. Pudjoatmodjo, "Application of Extraction-Transformation-Loading (ETL) in Data Warehouse (Case Study: Ministry of Agriculture)," *J. Nas. Educators. Tech. Inform.*, vol. 5, no. 2, p. 61, 2016, doi: 10.23887/jan.v5i2.9855.

[20] D. Nurfauzan and T. Fatimah, "Implementation of K-Nearest Neighbors Regression Algorithm in Predicting Stock Prices," *Semin. Nas. Mhs. …*, no. September, pp. 576–584, 2022.

[21] M. M. Singh, "Extraction Transformation and Loading (ETL) of Data Using ETL Tools," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 6, pp. 4415–4420, 2022, doi: 10.22214/ijraset.2022.44939.

[22] R. Umar, I. Riadi, and D. A. Faroek, "A Komparasi Image Matching Menggunakan Metode K-Nearest Neightbor (KNN) dan Support Vector Machine (SVM)," *J. Appl. Informatics Comput.*, vol. 4, no. 2, pp. 124–131, 2020, doi: 10.30871/jaic.v4i2.2226.

[23] A. A. Karim, M. A. Prasetyo, and M. R. Saputro, "Comparison of Random Forest, K-Nearest Neighbor, and SVM Methods in Predicting Italian League Match Accuracy," *Stain. (Seminar Nas. Techno. Science)*, vol. 2, no. 1, pp. 377–382, 2023.

[24] Ramadhika Dwi Poetra, "CHAPTER II Literature Review CHAPTER II LITERATURE REVIEW 2.1. 1–64," *Gastron. Ecuadorian and Tur. local.*, vol. 1, no. 69, pp. 5–24, 2019.

[25] S. Adi and A. Wintarti, "Comparison of Support Vector Machine (SVM), K-Nearest Neighbors (Knn), and Random Forest (Rf) Methods for Heart Failure Prediction," *MATHunesa J. Ilm. Mat.*Vol. 10 No. 2 pp. 258–268, 2022, doi: 10.26740/mathunessa.v10n2.p258–268.

[26] O. Arifin and T. B. Sasongko, "Comparative analysis of performance levels of support vector machine and naïve bayes classifier methods," *Semin. Nas. Techno. Inf. and Multimed. 2018*, vol. 6, no. 1, pp. 67–72, 2018.

[27] Y. A. Singgalen, "Performance Analysis of NBC, DT, SVM Algorithms in the Data Classification of Borobudur Temple Visitor Reviews Based on CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi: 10.47065/bits.v4i3.2766.

[28] B. A. Nugroho, A. K. A. Pradana, and E. Nurfarida, "Prediction of Arrival Time of Motor Vehicle Service Customers Based on Historical Data using Support Vector Machine," *J. Education and Research. Inform.*, vol. 7, no. 1, p. 25, 2021, doi: 10.26418/jp.v7i1.42964.

[29] A. Vermaysha and N. Nurmalitasari, "House Price Prediction in Karanganyar Regency Using Linear Regression Method," *Pros. Semin. Nas. Techno. Inf. and Business*, pp. 6–11, 2023.

[30] Miftahuljannah, Aswan Supriyadi Sunge, and Ahmad Turmudi Zy, "Sales Prediction Analysis with Linear Regression Method at Pt. Eagle Industry Indonesia," *J. Inform. Techno. and Science*, vol. 5, no. 3, pp. 398–403, 2023, doi: 10.51401/jinteks.v5i3.3325.

[31] N. R. Setyoningrum, P. J. Rahimma, S. T. Teknologi, I. Tanjungpinang, and K. Tanjungpinang, "Implementation of Linear Regression Algorithm in the Prediction System for New Student Registrants of the Indonesian College of Technology Tanjungpinang," *Pros. Semin. Nas. Social Science. and Techno.*, no. 4, pp. 13–18, 2022.

[32] D. Novianty, N. D. Palasara, and M. Qomaruddin, "Linear Regression Algorithm on Prediction of Registered Patent Applications in Indonesia," *J. Sist. and Techno. Inf.*, vol. 9, no. 2, p. 81, 2021, doi: 10.26418/justin.v9i2.43664.

[33] S. J. Kim, S. J. Bae, and M. W. Jang, "Linear Regression Machine Learning Algorithms for Estimating Reference Evapotranspiration Using Limited Climate Data," *Sustain.*, Vol. 14, No. 18, 2022, doi: 10.3390/su141811674.

[34] M. A. Musababa, "Implementation of Linear Regression Algorithm for Prediction of Rice Production in Grobogan District," *Data Sci. Indones.*, vol. 3, no. 2, pp. 68–78, 2023, [Online]. Available: https://jurnal.itscience.org/index.php/dsi/article/view/3118%0Ahttps://jurnal.itscience.org/index.php/dsi/article/download/3118/2711

[35] M. Fadli and R. A. Saputra, "Classification and Performance Evaluation of Random Forest Model for Stroke Prediction," *J.T. J. Tech.*, vol. 12, no. 2, pp. 72–80, 2023.

[36] I. A. Tarigan, I. P. A. Bayupati, and G. A. A. Putri, "Comparison of support vector machine and backpropagation models in forecasting the number of foreign tourists in Bali province," *J. Techno. and Sist. Comput.*, vol. 9, no. 2, pp. 90–95, 2021, doi: 10.14710/jtsiskom.2021.13847.

[37] D. A. Mukhsinin, M. Rafliansyah, and S. A. Ibrahim, "Implementation of Decision Tree Algorithm for Movie Recommendation and Rating Classification on the Netflix Platform Implementasi Algoritma Decision Tree

untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix," vol. 4, no. April, pp. 570–579, 2024.

[38] A. A. Suryanto, "Application of Mean Absolute Error (MEA) Method in Linear Regression Algorithm for Rice Production Prediction," *Saintekbu*, vol. 11, no. 1, pp. 78–83, 2019, doi: 10.32764/saintekbu.v11i1.298.

[39] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, Vol. 15, No. 14, pp. 5481–5487, 2022, Yogurt: 10.5194/GMD-15-5481-2022.

[40] M. Ćalasan, S. H. E. Abdel Aleem, and A. F. Zobaa, "On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function," *Energy Convers. Manag.*, vol. 210, 2020, doi: 10.1016/j.enconman.2020.112716.

[41] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

Andi Nugroho,
Phd Student in Computer Sciences – Computer Science Department, Mercu Buana University, Jakarta, Indonesia
Email : andi.nugroho@mercubuana.ac.id
Scopus Author ID : 57208427717
ORCID : orcidID= https://orcid.org/0000-0002-1713-035X.

Arif Lukman Hakim
Computer Science Student - Department of Information Systems, Universitas Mercu Buana, Jakarta, Indonesia
Email : 41820010018@mercubuana.ac.id

Indra Wibisono
Computer Science Student - Department of Information Systems, Universitas Mercu Buana, Jakarta, Indonesia
Email : 41820010066@mercubuana.ac.id

Yuanita Octoria
Computer Science Student - Department of Information Systems, Universitas Mercu Buana, Jakarta, Indonesia
Email : 41820010094@mercubuana.ac.id