

## Article

# A Statistical Approach to Crime Rate Prediction Using Multiple Linear Regression

Setiawan Ardi Wijaya<sup>1,3</sup>, Edo Arribe<sup>2,3</sup>, Muhitauldi<sup>3,\*</sup><sup>1</sup> Faculty of Business Management and Information Technology, Universiti Muhammadiyah Malaysia, Perlis 02100, Malaysia<sup>2</sup> Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Kelantan 16300, Malaysia<sup>3</sup> Department of Information Systems, Universitas Muhammadiyah Riau, Pekanbaru 28294, Indonesia; setiawanardiwijaya@umri.ac.id; edoarribe@umri.ac.id; 210402051@student.umri.ac.id

\* Correspondence

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**Abstract:** The high crime rate in Riau Province poses a serious threat to social stability and public safety, requiring accurate prediction strategies to support crime prevention efforts. Based on data from the Central Statistics Agency (BPS), Riau ranked seventh among the provinces with the highest crime rates in Indonesia in 2022, indicating that conventional prevention efforts remain insufficient. However, studies applying statistical data-based prediction models to crime in Riau are still limited, creating a gap in data-driven decision making. This study aims to develop a crime rate prediction model in Riau Province using the Multiple Linear Regression (MLR) method with BPS crime data from 2019–2023. The independent variables include six types of crime: corruption, drug dealers, drug users, terrorism, illegal logging, and human trafficking, while the dependent variable is the total number of crimes per district or city. The research process involved data collection, understanding, preprocessing, application of linear regression algorithms, model training and testing, and evaluation using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The results show that Pekanbaru City recorded the highest number of cases, mostly related to drug crimes. The model predicts an increase in Pekanbaru's cases from 3,331 in 2024 to 5,852 in 2027, while Dumai City is projected to decline from 543 to 397 cases. The model demonstrates high accuracy in most areas, particularly in Kampar (MAPE 0.28%), Siak (0.52%), and Rokan Hilir (0.94%), though less accurate in the Meranti Islands (565.99%) due to data instability. These findings prove that the Multiple Linear Regression method effectively predicts crime trends and can serve as a quantitative decision-making tool for law enforcement and local governments. Further research should include socioeconomic factors such as poverty and unemployment, and compare results with alternative forecasting methods like ARIMA and Exponential Smoothing to enhance prediction accuracy.

**Keywords:** Prediction; Crime; Multiple Linear Regression; Riau; MAPE**Copyright:** © 2025 by the authors. This is an open-access article under the CC-BY-SA license.

## 1. Introduction

The crime rate is one of the primary indicators of public security [1-4]. Globally, Indonesia faces a serious challenge related to organized crime, reflected by its position at 20th place in 2023, according to the Global [5]. One region in Indonesia with a high level of vulnerability to criminality is Riau Province, which ranked seventh out of 38 provinces in Indonesia with a crime figure of 12,389 reported cases in 2022 [6]. The high rate of crime in strategic areas like Pekanbaru and its surroundings [7]

necessitates accurate predictive methods to support law enforcement agencies in formulating preventive policies.

Prediction is the process of estimating something that will happen in the future [8-11]. Another definition states that prediction is the process of estimating future needs, covering the aspects of time, quality, quantity, and location required to meet the demand for goods or services [10-13]. The Multiple Linear Regression (MLR) method is proposed as an approach to model and predict crime trends. MLR is an analysis method used to model the

relationship between one dependent variable (Y) and two or more independent variables (X) [14-18]. This technique allows researchers to assess the extent to which the independent variables influence the dependent variable and to predict changes in the value of Y based on variations in X. MLR was chosen based on its proven effectiveness in various comparative studies. Research shows that MLR has superior accuracy compared to the Decision Tree and Autoregressive Integrated Moving Average (ARIMA) methods in different prediction cases [8]. This performance surpasses other methods, such as Triple Exponential Smoothing, which was previously applied to crime data with a lower level of accuracy [19].

This research addresses a literature gap by applying the MLR method for the first time to crime data in Riau Province. The model's performance will be evaluated using the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) metrics to measure prediction precision and accuracy [20-23]. Thus, this study aims to build and validate a predictive model that can serve as a strategic decision-support instrument in efforts to mitigate criminality in Riau Province.

## 2. Research Method

In this study, the Multiple Linear Regression algorithm was used. The evaluation metrics MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Squared Error) were applied as shown in the following procedure, consisting of the stages illustrated in Figure 1.

In Figure 1, this study began with data collection followed by data understanding to identify the characteristics of crime data from the Riau Provincial Statistics Agency for the period 2019–2023, including the identification of independent variables (types of crime) and dependent variables (number of crimes). Next, data pre-processing was carried out by cleaning duplicate data, handling missing values, and standardizing the format so that the data was ready for use. The cleaned data was then used to apply the algorithm and train the Multiple Linear Regression model, where the regression coefficients representing the relationship between variables were calculated. The resulting model was then tested using Mean Absolute Percentage Error (MAPE) and Root Mean

Square Error (RMSE) to measure the accuracy of the predictions. Models that met the accuracy criteria were then used to predict the number of crimes for the 2024–2027 period.

### 2.1. Data Collection

Data collection is a process or activity carried out by researchers with the aim of obtaining and collecting various phenomena, information, and conditions at the research location that are relevant to the scope of the research [8]. This study uses quantitative data, which is data that can be measured and calculated directly, containing information or descriptions in the form of numbers and statistics [9, 24, 25]. The data is sourced from Riau Provincial Statistics Agency in figures for 2021-2024: <https://riau.bps.go.id/id/publication/2021/02/26/984c872d808a74964a550e1e/provinsi-riau-dalam-angka-2021.html>.

### 2.2. Data Understanding

This stage involves identifying and understanding the quantitative data used, with the aim of recognizing patterns and structures within the data [26, 27].

### 2.3. Preprocessing Data

During the data preparation stage, several important steps are taken. First, data duplication is checked to maintain data quality and accuracy, thereby avoiding redundancy and errors in the analysis process. Next, missing values in each attribute are examined. This step aims to ensure that the data used in the study is complete and does not contain gaps that could potentially affect the analysis results. Its function is to prepare and improve the quality or readiness of data to meet requirements [10, 28, 29].

### 2.4. Application of Algorithms Linier Regression

Multiple linear regression is used to predict changes in a dependent variable based on variations in several independent variables. It is called multiple because it involves more than one predictor [16-18, 30]. This regression model is formulated through the following equation:

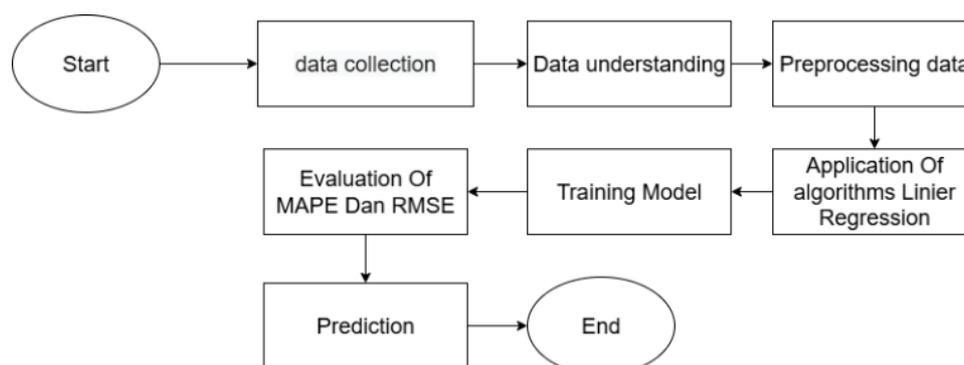


Figure 1. Research Flow.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

Information:

Y (1) : dependent variable  
 X<sub>1</sub>, ..., X<sub>n</sub> : independent variable on observation to -i  
 β<sub>0</sub>..., β<sub>n</sub> : Regression Coefficient

## 2.5. Model Training and Testing

This stage involves selecting data mining techniques by determining the algorithms to be used. Modeling is selected and applied to the prepared dataset to meet requirements. In addition, model creation also includes evaluation of the resulting model [31, 32].

A total of 80% of the data was used to train the model, while the remaining 20% was used to test it on unseen data. The 80:20 ratio is a commonly used optimal balance for dividing data. Most of the data (80%) is used to train the model so that it can recognize patterns in depth, while the remaining data (20%) is used to test the model's performance on data it has never seen before. This ratio is important to ensure that the model does not just memorize the training data (overfitting) but is also able to perform well when encountering new data in the real world [32, 33].

## 2.6. Evaluation Metrics

This study uses MAPE and RMSE evaluations. MAPE is used as a measure to assess the accuracy of a forecast result. MAPE is calculated by finding the absolute difference between the actual value and the predicted value for each period, then dividing it by the actual value for that period. The results of these calculations are then averaged as a percentage to obtain a measure of forecast accuracy [21-23, 34].

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (2)$$

Formula 2 MAPE is obtained by calculating the difference between the predicted value ( $\hat{y}_i$ ) and the actual value ( $y_i$ ), then dividing it by the actual value to obtain the relative error. The absolute value of the error is taken so that the result is always positive, then added and divided by the number of data (n) to obtain the average error. The final result is multiplied by 100 to be expressed as a percentage. A smaller MAPE value indicates that the model has a higher level of accuracy, while a large MAPE value indicates that the model's predictions are less accurate. Model performance can be classified as excellent if the MAPE value is below 10%, while a MAPE range between 10% and 20% indicates good performance [35, 36].

Table 1 shows the MAPE (Mean Absolute Percentage Error) performance categories based on prediction accuracy levels. A smaller MAPE value indicates a more accurate prediction. If the MAPE value is less than 10%,

the prediction result is Predictions are very accurate. A MAPE range of 10–20% indicates that the prediction is good, while a value between 20–50% indicates moderate accuracy. Meanwhile, if the MAPE value exceeds 50%, the prediction results are considered inaccurate and the model needs to be improved. RMSE is a square-based assessment method that measures the average error. RMSE is the square root of the average difference between the predicted data and the actual data. the lower the RMSE value, the better the prediction quality [37, 38]. Here is the formula for calculating RMSE:

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

RMSE is calculated by summing the squared differences between the predicted values ( $\hat{y}_i$ ) and the actual values ( $y_i$ ), dividing the result by the total number of data points (n), and then taking the square root of that value. The RMSE value indicates how far, on average, the predictions deviate from the actual values. A smaller RMSE value signifies a more accurate model, as it means the predicted results are closer to the actual observations.

## 2.7. Prediction

After the model evaluation stage is complete, the next step is to make predictions. Prediction is the process of estimating something that will happen in the future [9-11, 39]. In this study, predictions use multiple linear regression algorithms.

## 3. Result and Discussion

The research process follows the steps described in the methodology section. In this study, the data used to build the crime prediction model came from the Badan Pusat Statistika (BPS). The data collected included dependent and independent variables in prediction processing using multiple linear regression algorithms. The software used was Python programming with data processing using Google Colab.

Based on data analysis in Figure 2 for the 2019–2023 period, there are three main patterns of crime in Riau Province. First, from a temporal perspective, the number of cases fluctuated sharply, peaking at around 9,500 cases in 2021, then declining in 2022 and 2023. Second, in terms of the type of crime, narcotics cases dominated, especially

**Table 1.** MAPE Performance.

MAPE%	Interpretation
<10	Predictions are very accurate
10-20	good prediction
20-50	decent predictions
>50	inaccurate predictions

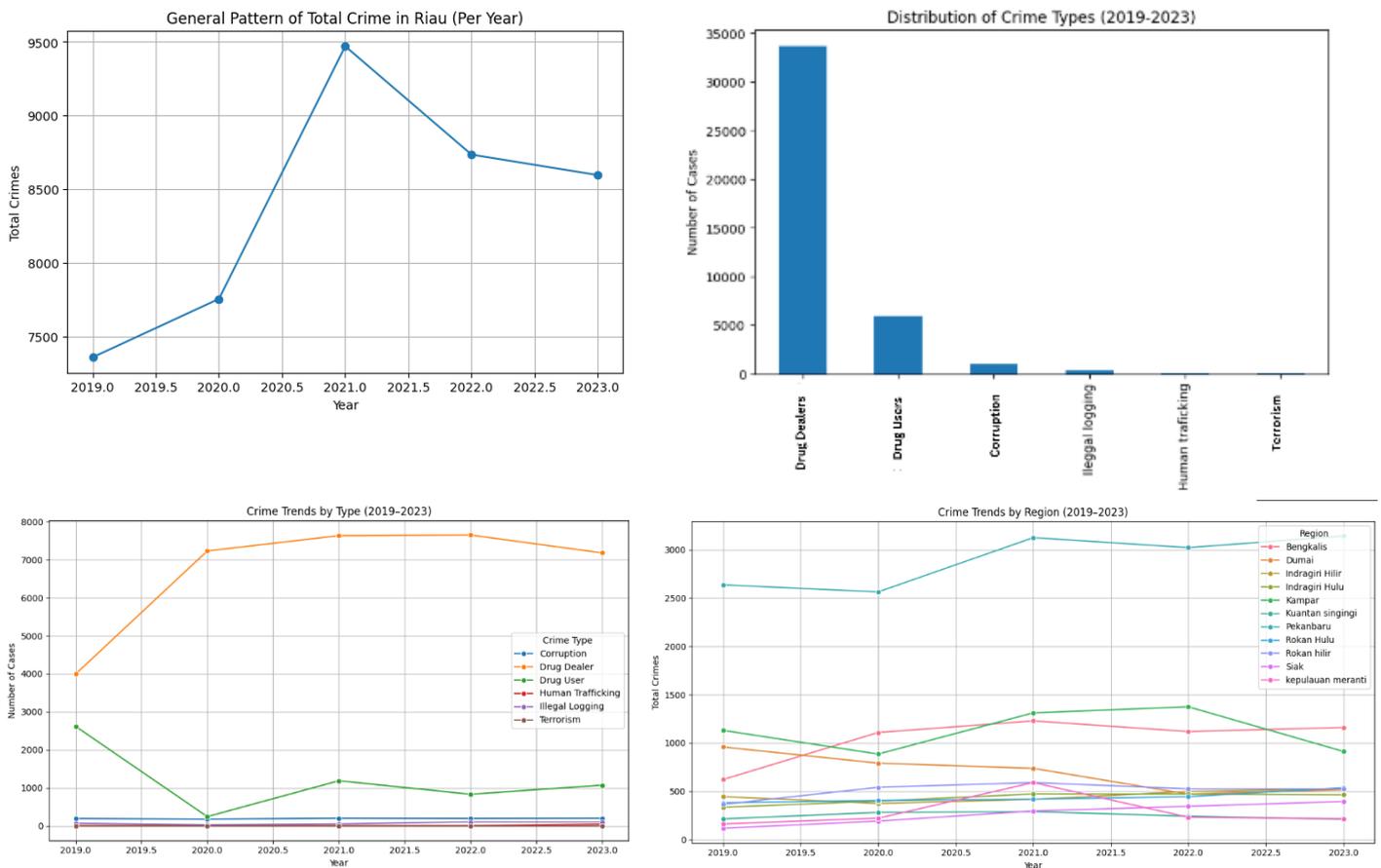


Figure 2. Crime Patterns in Riau Province.

drug dealers with 33,680 cases, followed by drug users with 5,950 cases. Third, from a regional perspective, Pekanbaru was the center of criminal activity with more than 14,000 cases, far higher than Kampar (5,700 cases) and Bengkalis (5,300 cases). The next step is to preprocess the data.

Data verification shows that there is no duplication based on the combination of region and type of crime, and there are no empty values in any column. Thus, the dataset is declared unique, complete, and clean. so it can be used to apply the linear regression algorithm.

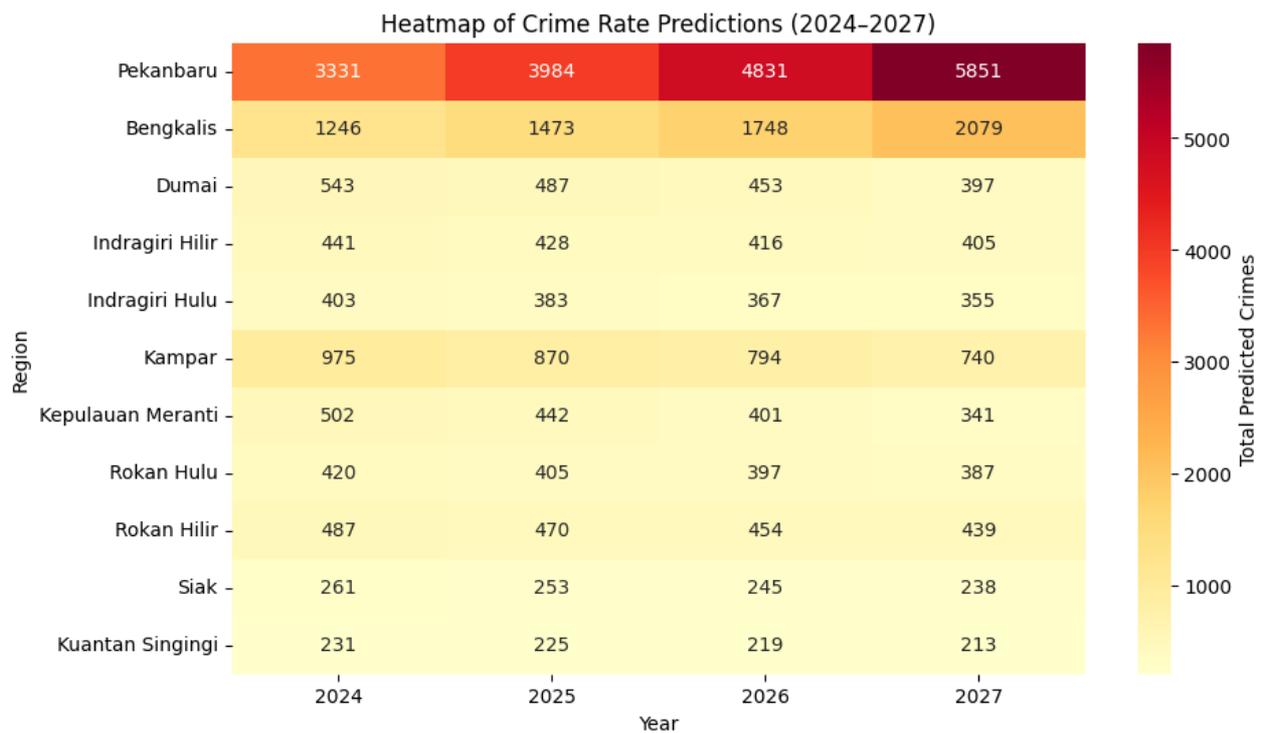
The process of creating a linear regression model using the Python programming language. In the line

model = LinearRegression(), the program creates or initializes a linear regression model that will later be used to analyze the relationship between input and output variables. Meanwhile, the command print(model) is used to display information that the model has been successfully created. At this stage, the model is still empty and has not been trained using data, so the next step is to train the model so that it can make predictions.

The process of dividing data into two parts, namely training data and testing data. This division is done using the train\_test\_split() function from the scikit-learn library. In the code, the data is divided with a proportion of 80% for training and 20% for testing, which is determined by

Table 2. MAPE and RMSE Performance Table.

Region	MAPE (%)	RMSE	Interpretation
Pekanbaru	12.41	318.10	good prediction
Bengkalis	3.96	43.85	Predictions are very accurate
Dumai	20.20	159.76	Decent predictions
Indragiri Hilir	3.73	13.93	Predictions are very accurate
Indragiri Hulu	1.36	5.43	Predictions are very accurate
Kampar	0.28	2.47	Predictions are very accurate
Kep. Meranti	565.99	1250.83	inaccurate predictions
Rokan Hulu	5.21	21.05	Predictions are very accurate
Rokan Hilir	0.94	5.06	Predictions are very accurate
Siak	0.52	0.99	Predictions are very accurate
Kuantan Singingi	17.62	49.50	good prediction



**Figure 3.** Prediction Results.

the `test_size=0.2` parameter. The `random_state=42` parameter is used so that the data division process is consistent every time the code is run. This division results in four variables: `X_train` and `y_train` for the training data, and `X_test` and `y_test` for the test data. Based on the output displayed, the training data (`X_train`) has a size of (4, 7) and the test data (`X_test`) has a size of (1, 7), which indicates the number of rows and columns in each data set. This allows us to proceed with algorithm evaluation.

In Table 2, this model shows the highest accuracy in areas such as Kampar (MAPE 0.28%), Siak (MAPE 0.52%), and Rokan Hilir (MAPE 0.94%), with very low RMSE values (below 6 cases), indicating high prediction reliability. However, this model fails to predict accurately for the Meranti Islands region, with a MAPE of 565.99%, which is likely due to significant data instability and fluctuations. For Pekanbaru, although the MAPE is relatively good (12.41%), the high RMSE value (318.10) indicates that the absolute deviation is quite large, which is reasonable considering the very high volume of cases in that region. The next step is to predict crime rates.

In Figure 4, crime prediction results for the 2024–2027 period show varying trends in each region. Pekanbaru is predicted to remain a crime-prone area with a significant increase from 3,331 cases in 2024 to 5,851 cases in 2027. Bengkalis also shows an upward trend from 1,246 cases to 2,079 cases in the same period. Conversely, several regions are expected to experience a decline, such as Dumai (from 543 cases to 397 cases), Indragiri Hilir (from 441 to 405 cases), and Kampar (from 975 to 740 cases). A similar downward trend is also seen in the Meranti Islands, Rokan Hulu, Rokan Hilir, Siak, and Kuantan Singingi, although

in relatively small numbers. Overall, these projections confirm the dominance of Pekanbaru and Bengkalis in crime rates, while most other regions are likely to experience a decline.

#### 4. Conclusion

This study successfully developed a crime prediction model in Riau Province using the Multiple Linear Regression algorithm with BPS data from 2019–2023. The findings indicate that the model achieved high accuracy in most regions, particularly in Kampar, Siak, and Rokan Hilir with MAPE values below 1% and very low RMSE. Predictions for 2024–2027 suggest that Pekanbaru will remain the primary crime hotspot with a significant upward trend, followed by Bengkalis, while most other regions are projected to experience a decline in crime cases.

The implication of these findings is that the model can serve as a decision-support tool for local governments and law enforcement in designing more targeted crime prevention strategies, especially in high-crime regions. Nevertheless, this study has limitations, particularly in Kepulauan Meranti, where the MAPE value was extremely high (565.99%), indicating possible data anomalies or unaccounted socioeconomic factors.

Future research is recommended to incorporate additional variables such as socioeconomic and demographic factors (e.g., poverty, unemployment, or population density) and to compare the performance with other forecasting methods such as ARIMA or Exponential Smoothing, in order to enhance accuracy in regions with unstable data patterns.

## 5. Conflicts of Interest

The authors declare no conflicts of interest.

## 6. References

- [1] F. Duarte, Á. Jiménez-Molina, and I. Sarmiento, "Crime and mental health: Examining the associations between perceived insecurity, crime victimization, and psychological distress in Chile," *Wellbeing, Space and Society*, vol. 9, Dec. 2025, doi: 10.1016/j.wss.2025.100303.
- [2] I. Adamse, A. Blokland, and V. Eichelsheim, "The geographical aspect of offending across crime types: A study on the journey to crime and co-offender dispersion," *J Crim Justice*, vol. 101, p. 102529, Nov. 2025, doi: 10.1016/j.jcrimjus.2025.102529.
- [3] R. E. A. Churchill, E. Amankwah, S. Awaworyi Churchill, and L. Farrell, "Crime and chance: Understanding the relationship between neighbourhood crime and gambling," *Public Health*, vol. 248, Nov. 2025, doi: 10.1016/j.puhe.2025.105953.
- [4] P. Burlian, *Patologi Sosial*. Bumi Aksara, 2022. Accessed: Oct. 30, 2025. [Online]. Available: <https://books.google.co.id/books?id=0L5mEAAAQBAJ>
- [5] global organized crime index, "Indeks Kriminalitas Global," Aug. 15, 2024. Accessed: Feb. 23, 2025. [Online]. Available: <https://goodstats.id/article/indonesia-masuk-20-besar-negara-dengan-indeks-kriminalitas-tertinggi-di-dunia-3ktwl>
- [6] Linda, "Terjadi 5.764 Kasus Kejahatan di Riau, Pekanbaru Paling Rawan," *Metro Riau*. Accessed: Oct. 30, 2025. [Online]. Available: <https://metroriau.com/berita/17910-terjadi-5-764-kasus-kejahatan-di-riau-pekanbaru-paling-rawan.html>
- [7] F. S. Pratiwi, "Data Sebaran Kasus Kejahatan Menurut Provinsi di Indonesia pada 2022," *Data Indonesia*. Accessed: Oct. 30, 2025. [Online]. Available: <https://dataindonesia.id/varia/detail/data-sebaran-kasus-kejahatan-menurut-provinsi-di-indonesia-pada-2022>
- [8] T. Hall and K. Rasheed, "A Survey of Machine Learning Methods for Time Series Prediction," *Applied Sciences (Switzerland)*, vol. 15, no. 11, Jun. 2025, doi: 10.3390/app15115957.
- [9] J. Chen, K. Tong, Q. Yu, S. Chen, T. Balezentis, and D. Streimikiene, "Innovative knowledge-based system for forecasting daily hotel operations amid external events using multi-source data: A time-varying parameter state-space model," *Journal of Innovation & Knowledge*, vol. 11, p. 100858, Jan. 2026, doi: 10.1016/j.jik.2025.100858.
- [10] E. Spiliotis and E. Theodorou, "Improving wind power forecasting accuracy through bias correction of wind speed predictions," *Sustainable Energy Technologies and Assessments*, vol. 83, Nov. 2025, doi: 10.1016/j.seta.2025.104599.
- [11] J. Wang, S. Xu, Y. Wang, and H. EL BOUHISSI, "An Analytics Framework for Healthcare Expenditure Forecasting with Machine Learning," *Healthcare Analytics*, p. 100428, Oct. 2025, doi: 10.1016/j.health.2025.100428.
- [12] A. Lusiana and P. Yuliarty, "Penerapan Metode Peramalan (Forecasting) pada Permintaan Atap di PT X," *Industri Inovatif: Jurnal Teknik Industri*, vol. 10, no. 1, pp. 11–20, Jun. 2020, doi: 10.36040/industri.v10i1.2530.
- [13] M. A. Kalhor et al., "Machine learning-based prediction and forecasting of chlorophyll-a in the northern Indian Ocean using satellite data," *Ecol Inform*, p. 103482, Oct. 2025, doi: 10.1016/j.ecoinf.2025.103482.
- [14] P. Roback and J. Legler, *Beyond Multiple Linear Regression*, 1st Edition. New York: CRC Press, 2021. doi: 10.1201/9780429066665.
- [15] H. Winnos and R. Septima, "Perbandingan Metode Regresi Linier Berganda dan Autoregressive Integrated Moving Average (ARIMA) Untuk Prediksi Saham PT. BSI, Tbk," Dec. 2022. doi: 10.58192/ocean.v1i4.350.
- [16] A. Sobral, J. Folgado, and C. Quental, "A predictive method for estimating the glenohumeral joint center from palpable landmarks using multiple linear regression trained on CT data," *J Biomech*, vol. 192, Nov. 2025, doi: 10.1016/j.jbiomech.2025.112954.
- [17] Y. Huang, Y. Wei, Z. Xia, C. Qu, Y. Suo, and X. Zhou, "Effect of subinhibitory concentrations of ascorbic and lactic acids on Salmonella Enteritidis surface adhesion: A multiple linear regression analysis," *Journal of Future Foods*, Sep. 2025, doi: 10.1016/j.jfutfo.2025.09.004.
- [18] T. Nguyen Luu Minh et al., "Multiple linear regression in adsorption capacity prediction: Application in plastic waste pyrolysis oil purification," *Sep Purif Technol*, vol. 378, Dec. 2025, doi: 10.1016/j.seppur.2025.134651.

- [19] M. Warahmah, Risnita, and M. S. Jailani, "Pendekatan Dan Tahapan Penelitian Dalam Kajian Pendidikan Anak Usia Dini," *Jurnal DZURRIYAT Jurnal Pendidikan Islam Anak Usia Dini*, vol. 1, no. 2, pp. 72–81, Sep. 2023, doi: 10.61104/jd.v1i2.32.
- [20] R. N. T. Siregar, V. Sitorus, and W.P. Ananta, "Analisis Prediksi Harga Rumah di Bandung Menggunakan Regresi Linear Berganda," *Journal of Creative Student Research*, vol. 1, no. 6, pp. 395–404, Dec. 2023, doi: 10.55606/jcsrpolitama.v1i6.3038.
- [21] Q. Xiao et al., "Standard molar entropy estimation of binary composite oxides: A novel hybrid approach based on deep learning and hyperparameters optimization," *Alexandria Engineering Journal*, vol. 131, pp. 16–32, Nov. 2025, doi: 10.1016/j.aej.2025.10.002.
- [22] A. Shukla, R. K. Pachauri, A. Hussain, A. Ali, and B. Khan, "Comparative analysis dust accumulation impact on PV performance using artificial neural network and machine learning algorithms," *Results in Engineering*, vol. 26, Jun. 2025, doi: 10.1016/j.rineng.2025.105024.
- [23] K. Ullah et al., "Hybrid CNN–BiGRU model with attention mechanism for enhanced short-term load forecasting," *Energy Reports*, vol. 14, pp. 2570–2577, Dec. 2025, doi: 10.1016/j.egyr.2025.09.035.
- [24] M. Sonderegger and M. Sóskuthy, "Advancements of phonetics in the 21st century: Quantitative data analysis," *J Phon*, vol. 111, Jul. 2025, doi: 10.1016/j.wocn.2025.101415.
- [25] S. Huang and Q. Ma, "A systematic review of data-driven learning research on language learning and teaching for pre-tertiary learners: Balancing qualitative and quantitative research," *Learn Individ Differ*, vol. 122, Aug. 2025, doi: 10.1016/j.lindif.2025.102752.
- [26] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Pencemaran Nama Baik di Media Sosial Twitter," *KOMPUTA: Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, May 2023, doi: 10.34010/komputa.v12i1.9418.
- [27] X. Cai, K. Xiong, Z. Luo, D. Weng, S. Ye, and Y. Wu, "CodeLin: An in situ visualization method for understanding data transformation scripts," *Visual Informatics*, vol. 9, no. 2, Jun. 2025, doi: 10.1016/j.visinf.2025.03.002.
- [28] W. G. Shin, J. S. Lee, Y. C. Ju, H. M. Hwang, and S. W. Ko, "Data preprocessing and machine learning method based on ameliorated mathematical models for inferring the power generation of photovoltaic system," *Energy Convers Manag*, vol. 333, Jun. 2025, doi: 10.1016/j.enconman.2025.119793.
- [29] A. Lazcano and M. A. Jaramillo-Morán, "Data preprocessing techniques and neural networks for trended time series forecasting," *Appl Soft Comput*, vol. 174, Apr. 2025, doi: 10.1016/j.asoc.2025.113063.
- [30] D. Alita, A. D. Putra, and D. Darwis, "Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, p. 295, Jul. 2021, doi: 10.22146/ijccs.65586.
- [31] A. Damayanti, F. D. Marleny, and A. A. Ningrum, "Implementasi Regresi Linear Berganda Untuk Prediksi Penjualan pada PT Trimandiri Sarana Propetindo Banjarmasin," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.6679.
- [32] H. K. Chinmayi, K. C. Flynn, G. S. Baath, P. Gowda, B. Northup, and A. Ashworth, "Monitoring legume nutrition with machine learning: The impact of splits in training and testing data," *Appl Soft Comput*, vol. 176, May 2025, doi: 10.1016/j.asoc.2025.113186.
- [33] Ernianti Hasibuan and n Aldian Karim, "Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web," *Jurnal Ilmiah Komputasi*, vol. 21, no. 4, Dec. 2022, doi: 10.32409/jikstik.21.4.3327.
- [34] D. C. I. Astuti, D. M. Khairina, and S. Maharani, "Peramalan Nilai Ekspor Nonmigas Kalimantan Timur dengan Metode Double Moving Average (DMA)," *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 2, no. 1, pp. 20–34, Jun. 2023, doi: 10.30872/atasi.v2i1.393.
- [35] R. Novita, I. Yani, and G. Ali, "Sistem Prediksi untuk Penentuan Jumlah Pemesanan Obat Menggunakan Regresi Linier," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 62–70, May 2022, doi: 10.57152/malcom.v2i1.198.
- [36] M. Walaszek et al., "Machine learning prediction of suicide attempt counts in Poland: Insights from Google trends and historical data," *International Journal of Clinical and Health Psychology*, vol. 25, no. 4, p. 100644, Oct. 2025, doi: 10.1016/j.ijchp.2025.100644.

- [37] M. Čalasan, I. Radonjić, M. Micev, M. Petronijević, and L. Pantić, "Voltage root mean square error calculation for solar cell parameter estimation: A novel g-function approach," *Heliyon*, vol. 10, no. 18, Sep. 2024, doi: 10.1016/j.heliyon.2024.e37887.
- [38] A. Villasante, Á. Fernández-Serrano, C. Osuna-Sequera, and E. Hermoso, "Methodology for stiffness prediction in structural timber using cross-validation RMSE analysis," *Journal of Building Engineering*, vol. 107, Aug. 2025, doi: 10.1016/j.job.2025.112767.
- [39] F. Andrian, S. Martha, and S. Rahmayuda, "Sistem Peramalan Jumlah Mahasiswa Baru Menggunakan Metode Triple Exponential Smoothing," *Coding Jurnal Komputer dan Aplikasi*, vol. 8, no. 1, Jan. 2020, doi: 10.26418/coding.v8i1.39199.