

Implementation of Least Square Method to Predict Crime in Indonesia Based on the Web

Nita Cahyani^{1*}, Rahmat Irsyada², Diva Anggi³

¹Universitas Padjadjaran, ²Politeknik Negeri Subang, ³Universitas Nahdlatul Ulama Sunan Giri, Indonesia

¹nita.cahyani@unpad.ac.id, ²irsyada.rahmat@gmail.com, ³divaanggi@gmail.com



*Corresponding Author

Article History:

Submitted: 07-12-2025

Accepted: 24-12-2025

Published: 29-12-2025

Keywords:

Least Square; crime prediction;
prediction accuracy; web.

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

This study was initiated by the need to apply the Least Square method to project the number of crimes in Indonesia using historical data from 2018 to 2022. Crime is a crucial issue in maintaining public security and supporting law enforcement, so accurate prediction results can assist the government in formulating public policies and optimizing resource use. The main problem of this study is how to apply the Least Square method to predict various categories of crimes in Indonesia, such as crimes against life, physical violence, morality, individual freedom, property rights with or without violence, and narcotics crimes. The purpose of this study is to develop a prediction model that can provide an accurate picture of future crime trends. The Least Square method was chosen because it can minimize prediction errors and process data with diverse variations, resulting in more stable and reliable estimates. The data used covers various types of crimes within the study period, with accuracy checked through the Mean Absolute Percentage Error (MAPE) value. The results show that the Least Square method is able to produce highly precise predictions with a MAPE value of 1.21%, thus proving effective in predicting crime rates in Indonesia with a very low error rate.

INTRODUCTION

Crime in Indonesia has increased in line with the complexity of modern society, driven by abundant material needs and non-constructive social ambitions, leading individuals or groups to seek wealth through illegal means. Crime is defined as behavior that violates the law, deviates from social norms, and causes harm or inconvenience to society (Budiman, 2023). Common examples of crimes include theft, assault, murder, fraud, and corruption. Based on the Indonesian Criminal Code (KUHP) and the International Classification of Crimes for Statistical Purposes (ICCS), crimes are classified into nine groups, including crimes against life, physical injury, morality, property/property (with or without violence), narcotics, fraud/corruption, and violations of public order.

Crime statistics from the Central Statistics Agency (BPS, 2024) show that crime rates in Indonesia over the past five years (2018–2022) have fluctuated significantly. Geographically, 2022 police administrative data shows the highest concentration of incidents in densely populated areas: the East Java Regional Police recorded the highest number of crimes (51,905 incidents), followed by the North Sumatra Regional Police (43,555 incidents), and the Jakarta Metropolitan Police (32,534 incidents). Conversely, the North Maluku, North Kalimantan, and West Sulawesi Regional Police recorded the lowest crime rates.

Analysis of specific crime types reveals distinct trends. Crimes against morality (rape and molestation) fluctuated, peaking in 2020 (6,872 incidents) and declining significantly to 4,336 incidents in 2022. That year, North Sumatra was the province with the highest number of incidents (453 incidents). Meanwhile, violent property crimes showed a downward trend, from 8,423 incidents in 2018 to 4,335 incidents in 2022, with North Sumatra also dominating the number of incidents. However, non-violent property crimes saw a sharp spike in 2022 (91,892 incidents) after a decline, making it the highest number in the past five years, with the North Sumatra Regional Police again recording the highest number of incidents.

In the context of data analysis and prediction, the Least Square method can be used because it offers analytical solutions that are easy to calculate and interpret, and is effective in handling data contaminated with noise or high variation (Octavia, 2022). The application of this method, particularly to forecast time series data, as demonstrated in research on LED lamp sales (Fauzi & Mulyana, 2021), demonstrated its excellent predictive ability with a low Mean Absolute Percentage Error (MAPE) value (8.0744%). This demonstrates the potential of the Least Square method as an appropriate approach for predicting future crime trends, thus enabling more effective planning in crime prevention and response efforts.



LITERATURE REVIEW

Forecasting

Forecasting, according to various experts such as Jay Heizer and Barry Render, and William J. Stevenson, is a process that combines art and science to predict future events or demand. This process involves analyzing historical data and projecting into the future using a systematic approach. In the context of operations management decision-making, forecasting is a crucial input that provides information about future demand. Its purpose is to assist management in determining required production capacity, inventory holdings, and other decisions related to the company's long-term planning. (Lusiana, 2020)

Essentially, forecasting is a crucial tool for planning the needs, inventory, and budget required to meet future demand. By analyzing patterns in historical sales data and other events, forecasting provides a solid foundation for making informed decisions regarding the management of operations, the supply chain, and the company's overall resources (AB Santoso, 2021).

Forecasting is a crucial activity in business management, aiming to predict or estimate future values based on historical data and other factors. In practice, forecasts are typically classified based on the time horizon they cover. According to previous research, there are several categories of forecasts based on this time horizon: long-term forecasts, medium-term forecasts, and short-term forecasts.

The three main categories of forecasts are based on their time span: Long-Term Forecasts, which cover 2 to 10 years and are used for strategic planning such as product development and large-scale resource allocation; Medium-Term Forecasts, which range from 1 to 24 months and are more specific, typically used for cash flow determination, production planning, and budgeting; and Short-Term Forecasts, which focus on the next 1 to 5 weeks and are highly operational, used for daily decision-making such as work scheduling, overtime requirements, and other short-term controls.

Decision Support System

A Decision Support System (DSS) is a computer-based system integrated with three main components: a language system, a knowledge system, and a problem processing system. The language system acts as a communication bridge, facilitating user interaction to ask questions and receive information in the decision-making process (Septilia, 2020). Furthermore, the knowledge system acts as a repository that stores data and procedures regarding the problem domain, which are crucial for analyzing and solving problems. The problem processing system, meanwhile, serves as the link between these two components, providing the problem manipulation tools and techniques necessary to analyze the data and generate recommendations or desired solutions.

The fundamental purpose of DSS is to support the decision-making process by providing relevant, accurate, complete, and timely information. DSS is designed to support decision-makers in addressing managerial problems, particularly those in semi-structured or unstructured situations. A typical application example is the use of DSS in selecting scholarship recipients, which is classified as a semi-structured problem where superior information quality is vital to supporting effective decisions (Pasaribu, 2023).

Least Squares

The Least Squares method is an approach used to find trend equations from data. This method is often used in time series analysis to determine patterns or trends in data, both for even and odd data. This approach searches for the linear equation that best fits the existing data, using the least squares technique to find the coefficients a and b in the trend equation $Y = a + bX$ (A. Purwanto, 2021).

Dividing the data into two groups, even and odd, is done to determine the appropriate X scores or values. For example, for even data, the X scores can be set as ..., -5, -3, -1, 1, 3, 5, ..., while for odd data, the X scores can be set as ..., -3, -2, -1, 0, 1, 2, 3,

The formula for finding the coefficients a and b using the Least Squares method is as follows:

$$\begin{aligned} Y &= a + bX \\ a &= \frac{\sum Y}{n} \\ b &= \frac{\sum xY}{\sum x^2} \end{aligned}$$

In this formula:

- YY is the time series data, which is an estimate of the trend data.
- XX is the time variable (day, week, month, or year).
- aa is the trend value at the base time point.
- bb is the average growth rate of the trend value per unit of time.

If the number of time periods is odd, the midpoint of the time period is defined as $X=0X=0$, so the sum of the positive and negative values equals zero.

The Least Squares method is a commonly used method in forecasting due to its simplicity and applicability to various types of data. Using this technique, a trend equation can be obtained that can be used to predict future values based on patterns or trends in existing historical data (A. Mihalik, 2022).

Crime

Crime is an act that violates the laws or rules prevailing in a society. It encompasses a variety of behaviors that can harm other individuals or disrupt social order in general. Crimes can be physical, such as committing violence against others or stealing property, or they can be non-physical, such as fraud or embezzlement. In society, crime is often considered a serious offense because it can disrupt the safety and well-being of others and affect social stability. Therefore, crimes are usually subject to sanctions or punishments, whether in the form of prevention, rehabilitation, or imprisonment, in accordance with applicable law in a country. Law enforcement against crime is crucial in maintaining order and justice in society and providing protection for individuals from harmful actions (N. Marufah, 2020). In general, crimes are classified as follows (Polri Public Relations, 2018):

METHOD

This study focuses on the implementation of the Least Squares method to predict crime rates in Indonesia. The subject of the study is crime rates in Indonesia, and the data used were obtained from the Central Statistics Agency's journal publication number 04300.2306 on crime statistics. This study utilizes secondary data, namely information already available from other sources such as previous research reports, public databases, archives, official documents, and related scientific literature.

The method used in this study is Least Squares, a statistical technique for finding the best solution to overdetermined problems, often used in linear regression and other data processing. The goal of this method is to find model parameters that minimize the squared error between observed values and predicted values. In linear regression, the Least Squares method is used to find the line or plane that best fits a set of data points.

The steps for using the Least Square method begin with dividing the data into even and odd groups, calculating the x^2 value, the xy value, and finally determining the a and b values. The a value is the intercept or point where the regression line intersects the y -axis, while b is the slope or incline of the regression line. Crime rate prediction is done using the formula $y = a + bX$, where y is the predicted variable and X is the independent variable. The accuracy of the prediction is measured by the Mean Absolute Percentage Error (MAPE), with a lower MAPE value indicating a better prediction model.

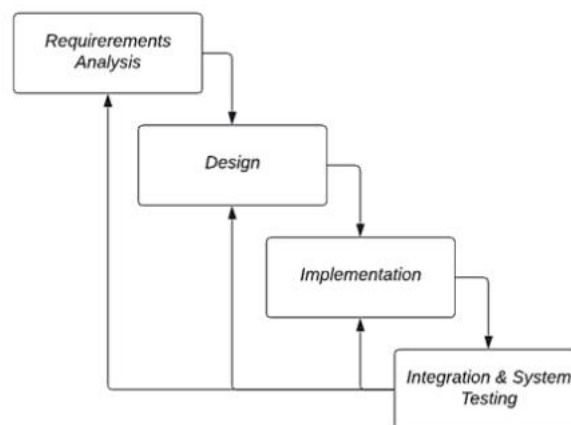


Fig. 1 SDLC Waterfall

The software development method used in this study is the waterfall model, which is tested using a blackbox model. The SDLC (Software Development Life Cycle) waterfall includes steps from requirements analysis, system design, implementation, testing, to maintenance. System requirements analysis includes functional and non-functional requirements. The main features of the system include prediction using the Least Square method, crime data management, and user account management. Hardware requirements include an Intel Core i3 or AMD A8 processor, a minimum of 4GB of RAM, and a minimum of 128GB of hard disk space, while software requirements include XAMPP as a web server, SQL database, and Windows OS.

RESULT

The following are the results of the implementation of the Least Square Method to predict crime in Indonesia based on the website :

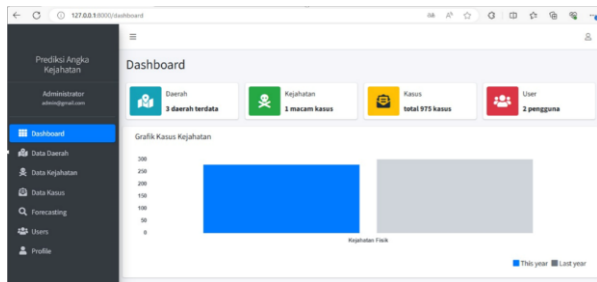


Fig. 2 Dashboard Page

The dashboard is the initial display after logging in which presents a summary of important information (area, crime, cases, users) and graphs to monitor crime developments.

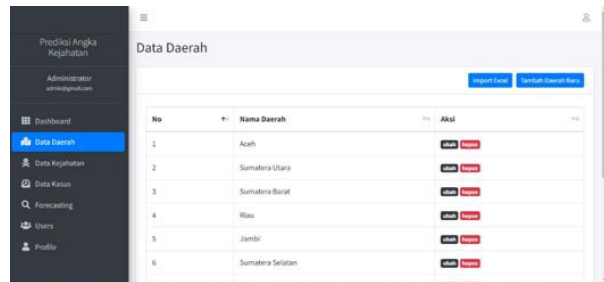


Fig. 3 Regional Data Page

The Regional Data menu in the crime rate forecasting application functions to manage regional information to be analyzed, displaying it in a table that includes the number, regional name, and an action column that provides options for deleting, changing, and adding new regional data.

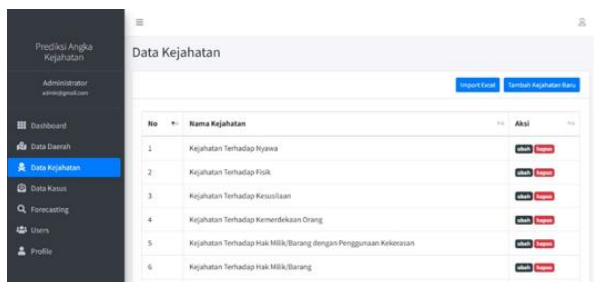


Fig. 4 Crime Data Page

The crime data menu functions to manage information on various types of crimes through a table that displays the serial number, crime name, and action buttons to delete, change, and add new crime entries.

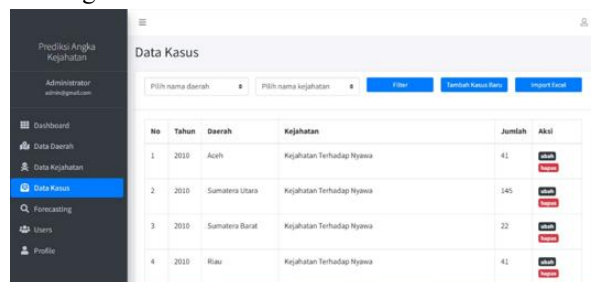


Fig. 5 Case Data Page

In the crime forecasting application, the Case Data menu presents a table containing detailed case information (Number, Year, Region, Crime Type, Total) with Delete and Edit options, and is equipped with the Add New Case feature and filters based on Crime Name or Region Name.

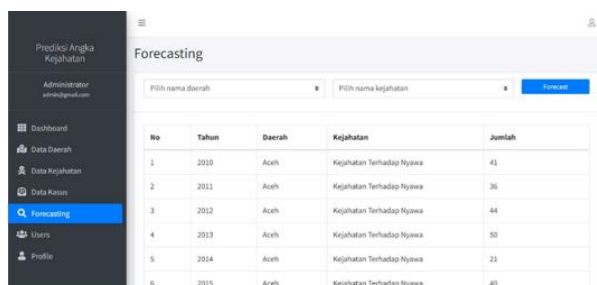


Fig. 6 Forecasting Page

In the crime forecasting application, the forecasting menu allows users to select the area name and crime name from two interface columns and then press the Forecast button to have the system process historical data and display a prediction of future crime rates for the selected combination.

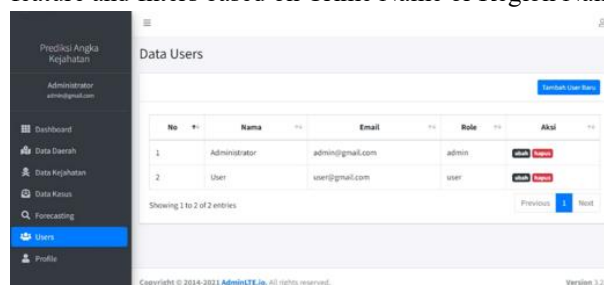


Fig. 7 User Data Page

The user management page in the crime forecasting application displays a list of registered users along with information (number, name, user, email, role), change and delete options for each user, and a button to add a new user to the system.

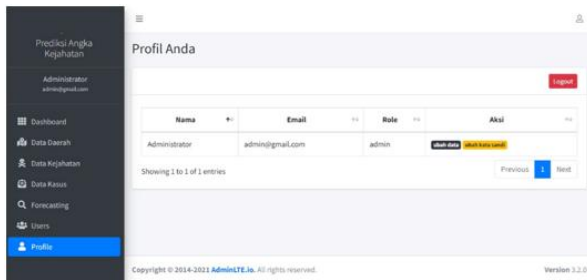


Fig. 8 Profil Page

The profile menu in the crime prediction application displays user information (number, name, user, email, and role) and provides action buttons to change data, change password, and log out so that users can manage their accounts safely and easily.

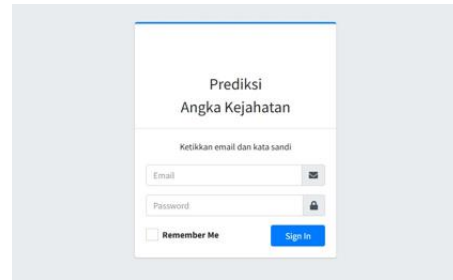


Fig. 9 Login Page

The login form, which requires a username and password for authentication, is an opening interface that grants different access rights between admin (can manage users) and user (can only use the application) before access to the main page is granted.

DISCUSSION

Based on the proposed method, the explanation uses the least square method with previously provided crime data.

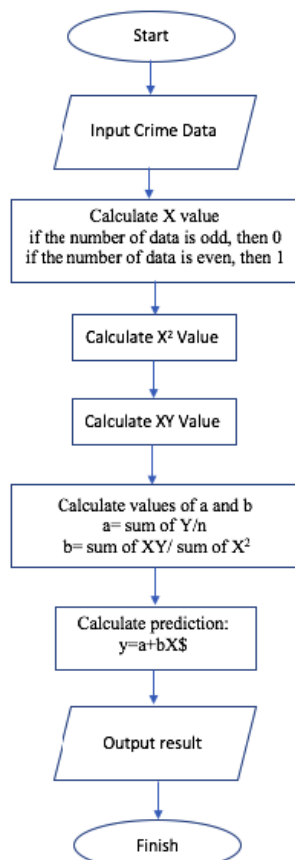


Fig. 9 Least square method flowchart

The forecasting process starts with the Start step, followed by Entering the Crime Data that will be used in the calculation. After the data is entered, the process moves on to Calculating the Value of X: the total number of data points is evaluated to determine if it is odd or even; if odd, X is set to 0, and if even, X is set to 1. This determined value of X is then squared to obtain the X Squared Value. Next, the XY Value Calculation is performed by multiplying the X value by the corresponding Y value (the crime data). Using these values, the process then Calculates the Values of a and b: a is calculated as the average of all Y values divided by the number of data points, while b is calculated as the average of all XY values divided by the average of all X values. The calculated a and b values are used to Calculate the

Y Prediction using the simple linear regression formula: $Y=a+bX$. Finally, the predicted Y Result Output is displayed, marking the end of the forecasting process with the Finish step.

In this chapter, manual calculations are performed to predict crime rates as follows:

Manual Calculations

Sample Data

In this calculation, the researcher used sample crime data on crimes against life taken from the Aceh Regional Police over 13 periods.

1. Trend Calculation

Table 1. Trend calculation

No	Period	Amount (Y)	X	XY	X2
1	2017	80	-2	- 160	4
2	2018	87	-1	-87	1
3	2019	79	0	0	0
4	2020	94	1	188	1
5	2021	102	2	204	4
	Σ	442	0	145	10

The table above outlines annual crime data, detailing the number of crimes recorded over a specific period. For example, 2017 recorded 80 incidents, which then increased to 87 in 2018 and reached a total of 102 in 2021. This annual breakdown is important for analyzing crime trends and rates in Indonesia.

The respective values of a and b are determined for the crime rate as follows:

$$\alpha = \frac{\Sigma y}{n} \quad b = \frac{\Sigma xy}{x^2}$$

Calculation of the values of a and b:

$$\alpha = \frac{\Sigma y}{n} = \frac{418}{13} = 32.1538461538$$

$$b = \frac{\Sigma xy}{x^2} = \frac{-336}{182} = -1.84615384615$$

1. Forecast (2023)

$$\begin{aligned} y &= a + bx \\ &= 32.1538461538 + -1.84615384615 (7) \\ &= 19.230769 \end{aligned}$$

2. Accuracy Test

$$a = 32.1538461538$$

$$b = -1.84615384615$$

$$F (2023) = 19.2307692308$$

$$MAPE = \frac{|Y - F_{2023}|}{Y} \times 100\%$$

$$MAPE = \frac{|19 - 19.2307692308|}{19} \times 100\%$$

$$MAPE = 1.21\%$$

Thus, the MAPE for predictions using the least squares method with the existing parameters is approximately 1.21%. The lower the MAPE value, the better the prediction model's performance.

Based on manual calculations, the predicted crime rate in Aceh in 2023 is 19,230,769. This figure aligns with the application's calculations, thus concluding that the calculations are accurate. Black-box testing is a testing method that focuses on testing the functionality of a system without examining its internal structure or program code. In this context, black-box testing is conducted to ensure the entire information system is tested and operates according to the functionality desired by administrators and users. This system testing is carried out by directly accessing the information system and running a series of pre-designed test stages. These stages include testing the input and output of various system modules to ensure each function operates properly according to specifications. This testing process includes checking the validity of input data, system responses to input, and interactions between system components.

The application test results showed that all key features functioned properly as expected. During the login test, the application successfully displayed the login page as intended, ensuring that the initial user process began correctly. After successfully logging in, the user was directed to the main display or dashboard of the application, indicating that the transition from the login page to the main display was smooth. Testing of the crime data management feature showed that users could add, edit, and delete data effectively, indicating that the data management function was operating as needed. Furthermore, the crime prediction feature functioned properly, allowing users to view and predict crime rates, indicating that the prediction model was working as expected. Testing also confirmed that administrators could manage user account data by successfully adding, editing, and deleting user accounts. The profile feature allowed users to change their own biodata and password, indicating that the profile personalization function was functioning properly. Finally, the logout feature functioned correctly, ensuring that users were automatically logged out of the system after pressing the logout button. Overall, the application test results showed that all tested features operated successfully and according to expectations.

CONCLUSION

The Least Squares method is a statistical technique used to analyze the relationship between independent and dependent variables by minimizing the sum of squared errors between observed and predicted values. In the context of predicting crime rates in Indonesia, this method begins with collecting historical data on crime rates in various regions, including the number of crimes per year and the types of crimes. After the data is collected, the next step is to determine the independent variables that influence the crime rate and the dependent variable, namely the crime rate itself. This data is then used to build a linear equation model, namely $Y = a + bX$ where Y is the crime rate, X is the independent variable, a is a constant, and b is the coefficient of influence of the independent variable. The Least Squares method is then used to predict parameters a and b by minimizing the sum of squares of the differences between the predicted and actual values. After the model is formed, an evaluation is carried out to measure accuracy with metrics such as the Mean Absolute Percentage Error (MAPE). In this study, the use of the least squares method has an accuracy of 1.21%, which is a very good result.

REFERENCES

- Adam, Fachrul Kurnia, A. Ferico Octaviansyah Pasaribu, and Agung Deni Wahyudi. "Aplikasi Monitoring Absensi Karyawan Ditlantas Dengan Penerapan Teknologi GPS (Studi Kasus: Ditlantas Polda Lampung)." *Jurnal Informatika dan Rekayasa Perangkat Lunak* 4.1 (2023): 1-9.
- Alghawazi, Maha, Daniyal Alghazzawi, and Suaad Alarifi. "Detection of sql injection attack using machine learning techniques: a systematic literature review." *Journal of Cybersecurity and Privacy* 2.4 (2022): 764-777.
- Badan Pusat Statistik. (2024). Statistik Kriminal. *Jurnal Publikasi BPS*, Nomor 04300.2306.
- Budiman, Arief. "Implementasi algoritma Autoregressive Moving Average dalam mengukur insiden kriminalitas berdasarkan data media online." *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)* 6.1 (2023): 24-33.
- Divisi Humas Polri. (2018). Jenis-jenis kejahatan). Diakses dari <https://divisi.humaspolri.go.id>
- Fauzi, Riki Maulana, and Dadang Iskandar Mulyana. "Implementasi Data Mining Menggunakan Metode Least Square untuk Memprediksi Penjualan Lampu LED pada PT. Sumber Dinamika Solusitama." *Jurnal Sosial Teknologi* 1.8 (2021): 907-919.
- Handojo, Andreas, Tanti Octavia, and Widya Arditanti. "Classification of Potential Blood Donors Using Artificial Neural Networks and Alternating Least Squares." 2023 2nd International Conference on Computational Modelling, Simulation and Optimization (ICCMO). IEEE, 2023.
- Kurniawan, Muhammad, and Budi Kurniawan. "Implementasi Pemrograman Python Menggunakan Visual Studio Code." *Jurnal Informatika dan Komputer (JIK)* 11.2 (2020): 1-9.
- Lusiana, Anna, and Popy Yuliarty. "Penerapan Metode Prediksi (Forecasting) pada Permintaan Atap di PT X." *Industri Inovatif: Jurnal Teknik Industri* 10.1 (2020): 11-20.
- Marufah, Nurbaiti, Hayatul Khairul Rahmat, and I. Dewa Ketut Kerta Widana. "Degradasi Moral sebagai Dampak Kejahatan Siber pada Generasi Millennial di Indonesia." *NUSANTARA: Jurnal Ilmu Pengetahuan Sosial* 7.1 (2020): 191-201.
- Mihalik, Agoston, et al. "Canonical correlation analysis and partial least squares for identifying brain-behavior associations: A tutorial and a comparative study." *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 7.11 (2022): 1055-1067.
- Octavia, Ranti Wilda Nur, and Umi Chotijah. "Implementasi Metode Least Square Untuk Prediksi Penjualan Kue Donat dan Bomboloni." *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi* 11.1 (2022): 251-262.
- Pambudiyatno, Nyaris, and Wasito Utomo. "Rancangan Sistem Presensi Dan Rekapitulasi Jurnal Kegiatan OJT Menggunakan Visual Studio Code Berbasis Web Di Airnav Cabang Matsc." *Prosiding SNITP (Seminar Nasional Inovasi Teknologi Penerbangan)*. Vol. 7. No. 1. 2023.



- Pasaribu, A. Ferico Octaviansyah, et al. "Sistem Pendukung Keputusan Seleksi Penerimaan Guru Menggunakan SAW." *Journal of Artificial Intelligence and Technology Information* 1.1 (2023): 13-19.
- Praba, Ardian Dwi, and Maryanah Safitri. "Studi perbandingan performansi antara mysql dan postgresql." *jurnal khatulistiwa informatika* 8.2 (2020).
- Purnomo, Kevin Rezananta, and Berto Mulia Wibawa. "Analisis tingkat kepuasan pengguna jalan Tol di wilayah unit Jatim 02 atas layanan satuan patroli jalan raya ditlantas Polda Jatim." *Jurnal Sains dan Seni ITS* 9.2 (2021): D256-D261.
- Purwanto, Agus. "Partial least squares structural squation modeling (PLS-SEM) analysis for social and management research: a literature review." *Journal of Industrial Engineering & Management Research* (2021).
- Ramadhan, Rizky Fajar, and Riki Mukhaiyar. "Penggunaan Database Mysql dengan Interface PhpMyAdmin sebagai Pengontrolan Smarthome Berbasis Raspberry Pi." *JTEIN: Jurnal Teknik Elektro Indonesia* 1.2 (2020): 129-134.
- Risqiati, Risqiati, Indrayanti Indrayanti, and Devi Sugianti. "Algoritma Apriori Untuk Asosiasi Transaksi Penjualan Benang di PT Hana Text." *Krea-TIF: Jurnal Teknik Informatika* 9.1 (2021): 1-8.
- Santoso, Agustinus Budi, Matheus Supriyanto Rumetna, and Kristy Isnaningtyas. "Penerapan Metode Single Exponential Smoothing Untuk Analisa Prediksi Penjualan." *Jurnal Media Informatika Budidarma* 5.2 (2021): 756-761.
- Septilia, Heni Ayu, Parjito Parjito, and Styawati Styawati. "Sistem pendukung keputusan pemberian dana bantuan menggunakan metode ahp." *Jurnal Teknologi Dan Sistem Informasi* 1.2 (2020): 34-41.
- Shiddieq, Diqy Fakhrun, and Andini Nur'aeni. "Implementasi Algoritma Least Square Untuk Memprediksi Penjualan (Studi Kasus Di Java Seven Cibaduyut Bandung)." *Jurnal Komputer Bisnis* 13.2 (2020).