JOURNAL OF WORLD SCIENCE
https://jws.rivierapublishing.id/index.php/jws
Volume 3 No. 6 June 2024
P-ISSN: 2828-8726
E-ISSN: 2828-9307

ANALYSIS OF THE PLACEMENT OF DISASTER EARLY WARNING FACILITIES BASED ON VILLAGE DATA IN WEST JAVA WITH A CLASSIFICATION APPROACH UTILIZING THE NAIVE BAYES ALGORITHM

Prafangasta Geo Ginantaka[1], Doddi Sudartha[2], Christina Juliane[3]
STMIK LIKMI, Jawa Barat, Indonesia

geoginantala@gmail.com[1], doddi.sudartha@gmail.com[2] , christina.juliane@likmi.ac.id[3]

ABSTRACT
West Java is one of the regions in Indonesia that is prone to various natural disasters such as earthquakes, floods, and landslides. These disasters are frequent and difficult to predict, such as the tornado that hit Rancaekek, Bandung on February 2, 2024, which caused significant damage. According to data from the West Java Regional Disaster Management Agency (BPBD), this disaster resulted in many damaged buildings and injuries. An early warning system is essential to reduce the impact of disasters. This study aims to place early warning facilities based on village data in West Java using the Naive Bayes method. The method used in this study is a data mining approach to extract patterns and valuable information from data that will be used in strategic decision-making related to the placement of early warning facilities. The data used was obtained from the West Java government's open data site, which includes attributes such as codes and names of provinces, districts, sub-districts, villages/sub-districts, as well as the availability status of disaster mitigation facilities. The results of the study show that many areas in West Java still do not have adequate early warning facilities. The use of Naive Bayes' algorithm aids in data classification and provides insights into the placement of more effective early warning facilities. The implication of this study is the need for more serious and coordinated efforts from the government, non-governmental organizations, and the community to increase the availability of disaster mitigation facilities in West Java.

Keywords: West Java, Natural Disasters, Data Mining, Naive Bayes, Early Warning Facilities.

Corresponding Author: Prafangasta Geo Ginantaka
E-mail: geoginantala@gmail.com

INTRODUCTION

West Java is prone to various natural disasters such as earthquakes, floods, and landslides, like other parts of Indonesia according to (Ismana et al., 2022). Specifically, there is a report that highlights Indonesia in addressing climate change, which includes adaptation measures, mitigation, and so on; the report analyzes that Indonesia is discussed as one of the prominent G-20 members because it is included in the list of the top ten countries that have the most significant emissions in the world (Urfah, 2017), this is at risk of weather changes and causing unwanted things such as natural disasters. There are several times that disasters in West Java are quite frequent events and are also difficult to predict in advance; for example, the tornado that occurred in Rancaekek Bandung on February 2, 2024, quoted from detik.com states that based on West Java BPBD records, the strong winds that occurred on the border of Bandung and Sumedang Regencies had damaged 534 buildings. (Detik, 2024).

BPBD data, 835 families were affected by the incident. BPBD said there were 33 people injured. Then, five other places were affected, namely Rancaekek, Cicalengka, and Cileunyi in Bandung Regency, Jatinangor, and Cimanggung; besides the tornado, several other events in West

Java have been mentioned previously where the most frequent natural disaster events or the most occurring in one area are natural disasters related to earthquakes, earthquakes also occur most frequently in West Java, especially in the Tasikmalaya area, which in 2020 alone recorded 563 occurrences. Therefore, preparing early warning facilities is very important to improve disaster preparedness and reduce the risk of losses arising from natural disasters, but are the existing facilities sufficient? It is a question that we need to analyze for future needs. This research will discuss the placement of natural disaster mitigation using the naive Bayes approach to study natural disaster mitigation penetration patterns. It will then be analyzed to determine whether it is appropriate.

Classification

Classification is one of the models in data mining. The classification model is a technique of predicting data, making predictions of the value of data whose results have been found to come from different data. (Nugraha & Juliane, 2022). The purpose of this model is to predict the value of an unknown variable from other variables that have been given; for example, in a study on the classification of emails as spam or not spam, the given variables may include the email title, sender, and email content. According to (Padri et al., 2023), the development of software that uses the Naive Bayes method can be applied to various fields, including email spam detection and text classification, among others.

The importance of this classification model lies in its ability to help make future decisions for the community or related parties or understand the patterns underlying the data on the placement of natural disaster warning facilities; using this technique, it is hoped that in the future, we can automate the process of classifying data that can be implemented in the field in various fields, one of which is to provide knowledge of the placement of natural disaster mitigation, precisely for those in West Java.

Data Mining

Data mining is the process of extracting knowledge obtained from data that is processed or inputted repeatedly to produce new knowledge or useful and valuable patterns from large amounts of data stored in various forms, such as databases, data warehousing, or even collections of text documents. (Swastika et al., 2023). The main goal is identifying relationships, patterns, or trends not directly visible in the raw data (Zai, 2022). Data mining methods involve statistical, mathematical, artificial intelligence, and machine learning techniques to analyze data and uncover helpful information. (Mardiani et al., 2023). Data mining is used in various fields, including business, science, health, finance, and others, to support better decision-making, understand customer behavior, forecast market trends, and more. (Hartatik et al., 2023).

Naive Bayes

According to (Azhar et al., 2022), Naive Bayes is a relatively uncomplicated use of statistical principles based on Naive Bayes to assume the existence or absence of a class based on features and attributes related to one another, with a simple assumption called "naive" or "plain." This method is very efficient and is often used to estimate an observation's class category based on features and attributes.

The basis of the Naive Bayes algorithm is Bayes' Theorem (Putra et al., 2016), which states that the probability of an event A occurring if we know that event B has occurred can be calculated using a formula:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

This algorithm is also relatively fast in the training and prediction process, making it suitable for use on large datasets. Naive Bayes has been used in various applications, including text classification; this time, the data used is approximately fifteen thousand columns.

Orange Application

Orange is an integrated and easy-to-use data analytics and machine learning platform. (Samad, 2022). Here are some of Orange's capabilities:

1. Data processing: Easily clean, prepare, and explore your data.
2. Data analysis: Perform statistical analysis, data visualization, and machine learning to gain insights from your data.
3. Machine learning: Building prediction, classification, and clustering models to solve various business problems.
4. Automation: Automate your data analysis and machine learning tasks to save time and effort.
5. Collaboration: Work with your team in real-time to complete data analysis projects.

In this research, we are assisted by tools to process data and implement algorithms that have been previously determined. We use Rapidminer to implement algorithms on the data we have with applicable provisions.

RESEARCH METHODS

For more organized research, we use a research method that will be described and described next, which aims to make the results obtained by the objectives of our research on the Placement of Disaster Early Warning Facilities Based on Village / Village Data in West Java with Naive Bayes. The stages are described in Figure 1.

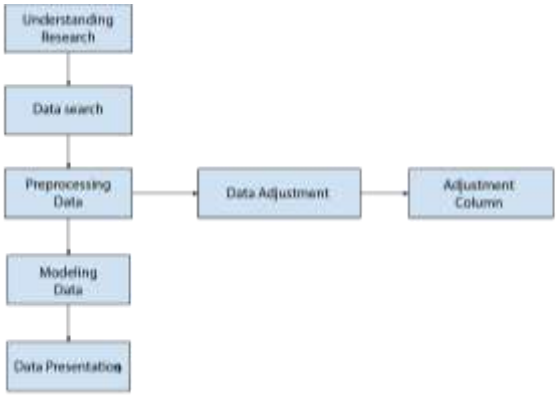

Figure 1. Research method

Research Understanding

At this stage we aim to identify where this research is aimed or can be called the ultimate goal, as for the steps or stages we take are the identification of research then we look for data that we get from public data or open data that we get from the West Java government website which presents data to be accessed by the broader community by the people of West Java or Indonesia, after we search and get data we do data preprocessing where at this stage we select and sort data between one of them cleansing data or cleaning data which aims to eliminate noice or attributes that are not used in our research, After preprocessing the data, we start or enter the data modeling stage, in this stage we will process the data we get with the methods or algorithms that will be used

in this research, our research uses a data mining approach, we use a classification approach, after we have successfully obtained the data, the data will be presented in the form of a report or report from the results of applying the approaches and algorithms we use.

Data Search

We searched the West Java Provincial government website at the data search stage, which we accessed at the following link or address: https://opendata.jabarprov.go.id/. This address contains data that can be accessed publicly; the address or link presents data uploaded by the West Java government, one of which is the data "Status of Availability of Natural Disaster Mitigation Facilities Based on Regency / City in West Java," which we use in this study. We use the data as a dataset, among others, has the following attributes:

1. code_province: states the code of West Java Province according to BPS provisions referring to the Central Statistics Agency Regulation Number 3 of 2019 rules with the numeric data type.
2. name_province: states that the data scope comes from the West Java Province area according to BPS regulations referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with the text data type.
3. bps_code_kabupaten_kota: states the code of each regency and city in West Java Province according to BPS regulations referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with the numeric data type.
4. bps_name_kabupaten_kota: states that the data scope comes from each district and city in West Java Province according to the naming of BPS referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with the text data type.
5. bps_kode_kecamatan: states the code of each sub-district in West Java Province according to BPS regulations referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with the numeric data type.
6. bps_name_kecamatan: states that the data scope comes from each sub-district in West Java Province according to the naming of BPS referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with the text data type.
7. bps_kode_desa_kelurahan: states the code of each village and kelurahan in West Java Province according to BPS provisions referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with the numeric data type.
8. bps_name_desa_kelurahan: states that the scope of the data comes from each village and kelurahan in West Java Province according to BPS naming referring to the rules of the Central Statistics Agency Regulation Number 3 of 2019 with text data type.
9. kemendagri_kode_kecamatan: states the code of each sub-district in West Java Province according to the provisions of the Ministry of Home Affairs referring to the rules of the Minister of Home Affairs Regulation No. 137 of 2017 with the numeric data type.
10. kemendagri_nama_kecamatan: states that the data scope comes from each sub-district in West Java Province according to the naming of the Ministry of Home Affairs referring to the rules of the Minister of Home Affairs Regulation No. 137 of 2017 with the text data type.
11. kemendagri_kode_desa_kelurahan: states the code of each village and kelurahan in West Java Province according to the provisions of the Ministry of Home Affairs referring to the rules of the Minister of Home Affairs Regulation No. 137 of 2017 with the numeric data type.

12. kemendagri_name_desa_kelurahan: states that the data scope comes from each village and kelurahan in West Java Province according to the naming of the Ministry of Home Affairs referring to the rules of the Minister of Home Affairs Regulation No. 137 of 2017 with the text data type.
13. status_availability_facility: states the availability status of the facility using the text data type.
14. Exists: states the availability status of the facility.
15. None indicates that there is no status of facility availability.
16. Year specifies the year of data production using a numeric data type.

The dataset has several attributes ranging from province code to ID, but the data of one attribute becomes a class attribute where the status attribute of facility availability has 2 values: none and none, which makes the approach we use classification.

Data Preprocessing

In this process or stage, we will perform data adjustments such as data cleansing, where we will sort and select which data will be processed further, such as performing data cleansing or identifying and resolving missing, incomplete, or invalid data. This can involve deleting rows or columns that are not needed or relevant to this research or filling in missing values with reasonable estimates so that there are no attribute blanks in the data to be processed in the future; we will reduce the number of features or attributes in the data to eliminate excess information or features that are not relevant to this research, which will then be processed by utilizing the classification approach by utilizing its algorithm with the help of available tools or those that are compatible to process the available data after the data preprocessing process or stage.

Data Modeling

At this data modeling stage, we will carry out or implement a classification approach by utilizing data that has gone through the data preprocessing stage; then, the data will be processed using an algorithm; in this study, we use or utilize the Naive Bayes algorithm and are assisted by compatible tools, namely the tool we use is rapid miner where the data will be presented in the form of a report or report where the results can become new knowledge.

Data Presentation

Data presentation is the stage where we present the results of the data that has been processed and has gone through the previous stages of the data search process (Wahyudi, 2017). Then, the data processing stage is passed, which will be processed with the help of tools and algorithms in the data modeling stage and then presented at this stage or data presentation (Arhami et al., 2020).

RapidMiner is the tool we use in this research where the data will be trained and tested (Munawar et al., 2023), which will then get results that can be presented. The picture can be seen in Figure 2.

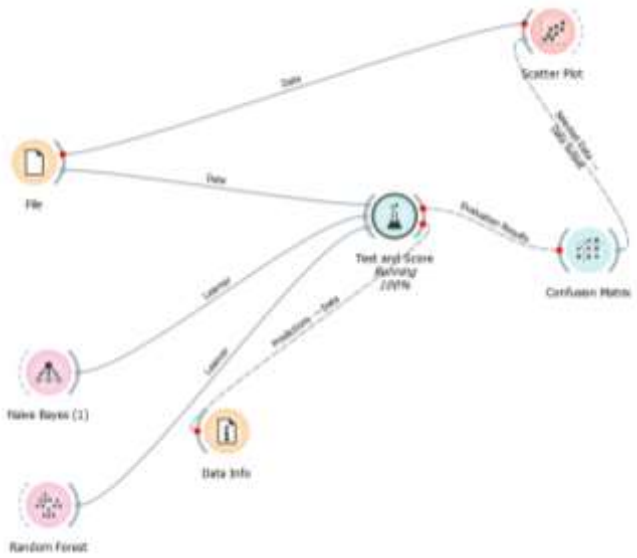

Figure 2. Orange Data Processing Process

RESULT AND DISCUSSION

Since this research uses the Naive Bayes algorithm, the formula used by us to implement the algorithm is:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

With the following information:

B = Data with unknown class
A = Hypothetical data B is a specific Class
P(A|B) = Probability of hypothesis A based on condition B (a posteriori prob.)
P(A) = Probability of hypothesis A (prior prob.)
P(B|A) = Probability of B based on the condition
P(B) = Probability of B

From the dataset attribute "status_availability," which contains NONE and ADA, will be adjusted to no for NONE and yes for ADA. Then the resulting 2203/15936 for when we compare with no or NONE, a comparison will be produced with NONE or no data, which results in data as much as 13733/15936. This data shows that in West Java Province, there still needs to be locations that have early warning facilities for natural disasters for the entire range, namely 2019-2021. The data we use in this study is in the range of 2020-2021, so the resulting data is 1591/10624 for yes and 9033/10624 for no. After preprocessing the data, we do a correlation where the correlation is by the use of the information mentioned earlier, among others, as follows:

P(yes) = 2203/15936 all
P(no) = 13733/15936 all
P(yes) = 1591/10624
P(no) = 9033/10624

After processing, the data will display the conclusion data as shown in Figure 3, which we can compare with the results using other methods, which can be seen in Figure 4, which can then be

seen in the matrix in Figure 5 and if displayed with a plot we take the x-axis and y-axis can be seen in Figure 6.

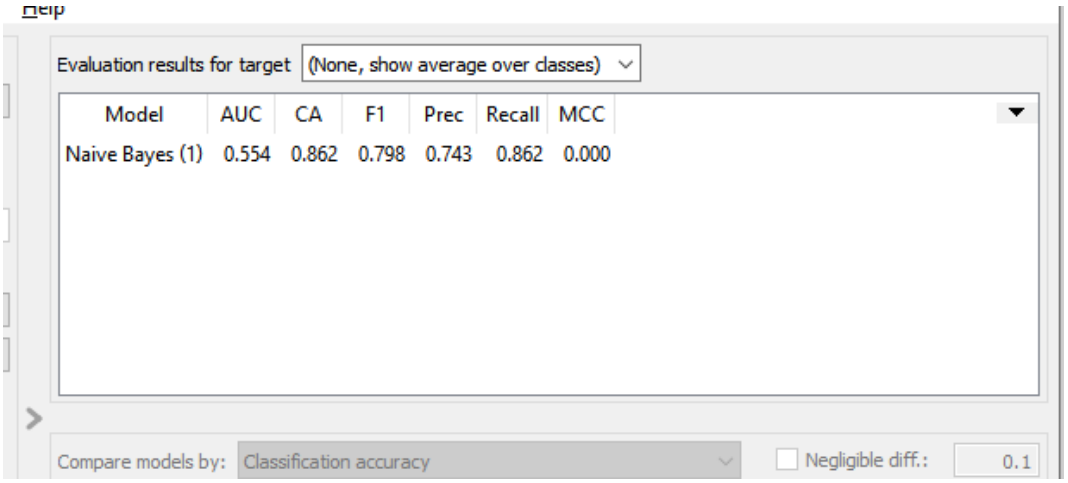

Figure 3. Naïve Bayes processing results

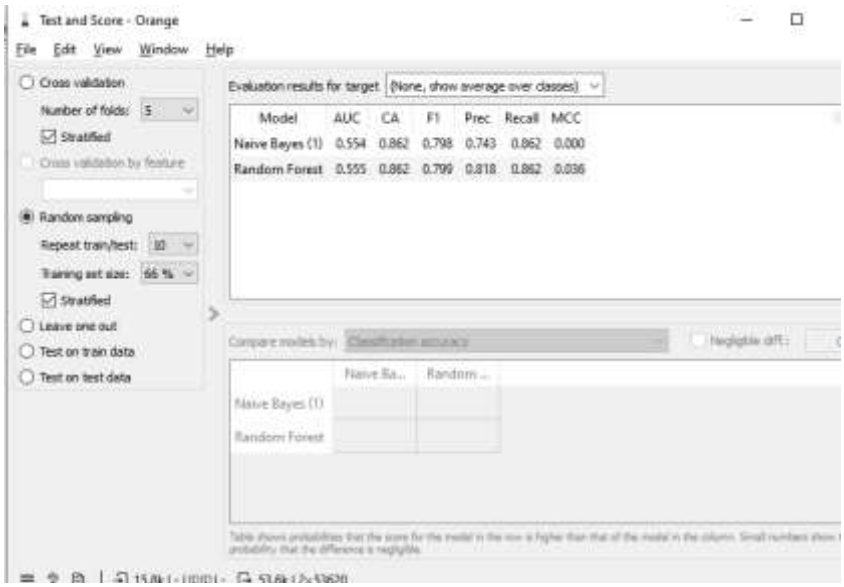

Figure 4. Naïve Bayes and Random Forest

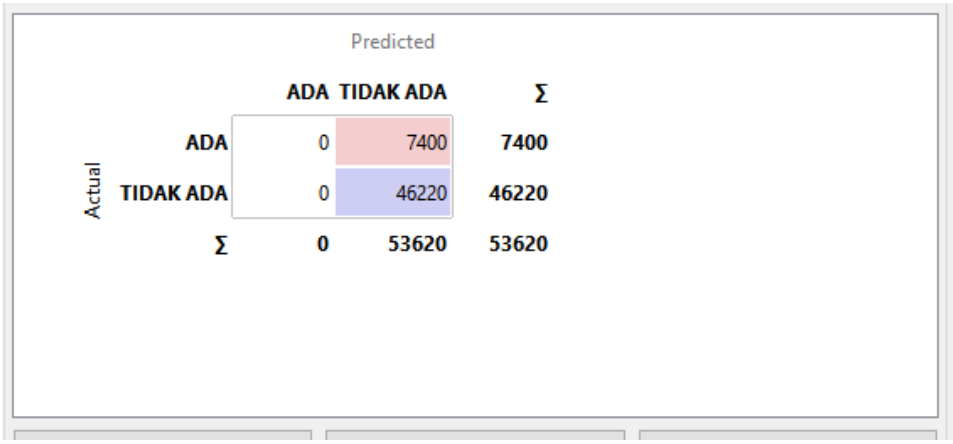

Figure 5. Matrix

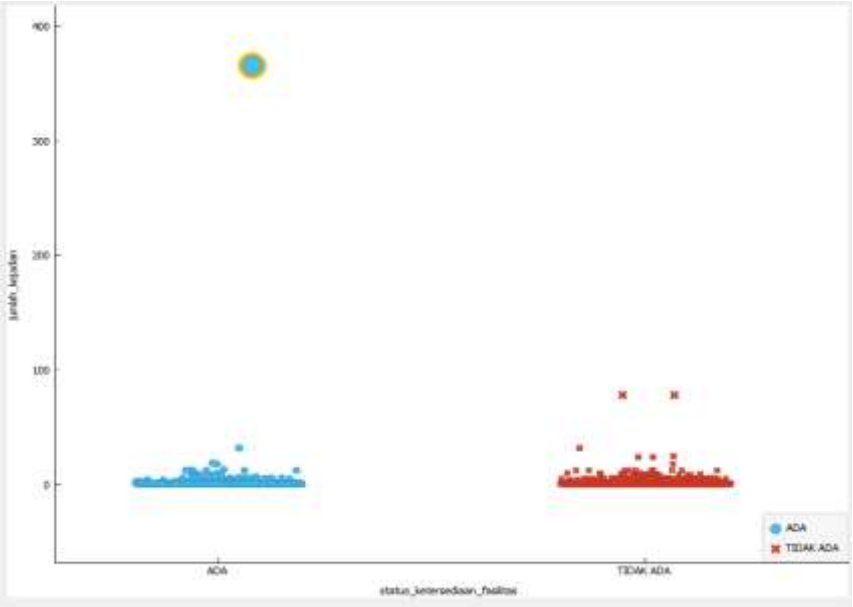

Figure 6. x and y axes

CONCLUSIONS

From the results of the data analysis that has been conducted, conclusions can be drawn regarding the availability of natural disaster mitigation in West Java. In general, the data shows that the availability of natural disaster mitigation in West Java still needs to improve, with many regions still needing to have or implement measures to reduce the impact of natural disasters. Natural disaster mitigation must be considered, especially considering West Java is prone to various natural disasters, such as earthquakes, floods, landslides, and tornadoes. Appropriate mitigation measures can help reduce the risk of loss and protect communities and infrastructure from damaging impacts. However, from the analysis, many West Java places still have not adequately implemented disaster mitigation efforts. This can be caused by various factors, including limited resources, lack of awareness of the importance of disaster mitigation, and administrative or policy constraints. Therefore, more severe and coordinated efforts are needed from the government, non-governmental organizations, and the community to improve the availability of disaster mitigation in West Java. Measures that can be taken include improved understanding of disaster risks, increased investment in mitigation infrastructure, development of early warning systems, and training and education to the community on emergency actions in the event of a disaster. In addition, it is also essential to strengthen inter-agency and inter-regional cooperation in disaster mitigation efforts. Cross-sectoral cooperation and good coordination between stakeholders will accelerate the implementation of mitigation measures and increase resilience to natural disasters in West Java. Thus, the results of this analysis call for all relevant parties to improve natural disaster mitigation efforts in West Java to reduce the risks and losses caused by natural disasters in the future. With the proper steps and strong support from all parties, West Java can become a more resilient and safe region from the threat of natural disasters.

REFERENCES

Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi.

Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *6*(1), 267–281.

Detik. (2024). *Angin Tornado Rancaekek yang Membawa Nestapa*. Www.Detik.Com. https://www.detik.com/jabar/berita/d-7209962/angin-tornado-rancaekek-yang-membawa-nestapa

Hartatik, H., Kwintiana, B., Nengsih, T. A., Baradja, A., Harto, B., Sudipa, I. G. I., Handika, I. P. S., Adhicandra, I., & Gugat, R. M. D. (2023). *DATA SCIENCE FOR BUSINESS: Pengantar & Penerapan Berbagai Sektor*. PT. Sonpedia Publishing Indonesia.

Ismana, D. R., Baehera, S., Fitrianto, A., Sartono, B., & Oktarina, S. D. (2022). Penggerombolan Desa di Jawa Barat Berdasarkan Daerah Rawan Bencana. *Jurnal Statistika Dan Aplikasinya*, *6*(2), 243–252.

Mardiani, E., Rahmansyah, N., Kurniati, I., Matondang, N., Zanitha, D. A., & Romzy, I. (2023). Membandingkan Algoritma Data Mining Dengan Tools Orange untuk Social Economy. *Digital Transformation Technology*, *3*(2), 686–693.

Munawar, Z., Muliantara, A., Kmurawak, R. M. B., Reba, F., Sroyer, A., Sukmawan, D., Rahman, A., Insany, G. P., Mandowen, S. A., & Toyib, W. (2023). *Big Data Analytics: Konsep, Implementasi, dan Aplikasi Terkini*. Kaizen Media Publishing.

Nugraha, F. F., & Juliane, C. (2022). Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4. 5. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(4), 2862–2869.

Padri, A. R., Asro, A., & Indra, I. (2023). Classification of Traffic Congestion in Indonesia Using the Naive Bayes Classification Method. *Journal of World Science*, *2*(6), 877–888.

Putra, I., Sudarma, M., & Kumara, I. N. S. (2016). Klasifikasi Teks Bahasa Bali dengan Metode Supervised Learning Naive Bayes Classifier. *Teknologi Elektro*, *15*(2), 81–86.

Samad, A. W. (2022). Analisis Data Sumber Daya Manusia Dalam Isu-Isu Global. *Indonesian Journal of Business Analytics*, *2*(1), 99–110.

Swastika, R., Mukodimah, S., Susanto, F., Muslihudin, M., & Adab, S. I. P. (2023). *IMPLEMENTASI DATA MINING (Clastering, Association, Prediction, Estimation, Classification)*. Penerbit Adab.

Urfah, S. (2017). *Pengaruh Keanggotaan Indonesia Dalam Forum G-20 Terhadap Pembangunan Infrastruktur Di Indonesia*. PERPUSTAKAAN.

Wahyudi, S. T. (2017). *Statistika Ekonomi: Konsep, Teori, dan Penerapan*. Universitas Brawijaya Press.

Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, *2*(3).