
APPLICATION OF THE NAIVE BAYES METHOD FOR DETERMINING THE ELIGIBILITY OF TOLL FARE REFUND REPORTS FOR MISCLASSIFIED VEHICLES

Fitri Pranita Nasution^{1*}, Wulan Dari², Dian Mayasari³, Fetty Ade Putri⁴

^{1,2,3,4} Universitas Potensi Utama, North Sumatra, Indonesia

E-mail: fitrinasution126@gmail.com^{1*}, ulandari2796@gmail.com², dian.tjan84@gmail.com³,
echiputri12@gmail.com⁴

Received: 11/08/2026 | Revised: 23/08/2026 | Accepted: 19/09/2026 | Published: 19/10/2026

Abstract

Vehicle misclassification at toll gates results in fare discrepancies for toll road users. The verification of fare refund reports, which is currently performed manually, is time-consuming and prone to human error. To address this problem, this study aims to design and implement a system based on the Naive Bayes method that automatically classifies and verifies the eligibility of vehicle fare reports. The Naive Bayes method was selected for its ability to process heterogeneous data and to produce accurate probability-based classifications. The system utilizes evidence such as payment receipts, e-toll transaction histories, and vehicle data to determine report eligibility. The collected data were analyzed to identify relevant patterns, enabling the system to generate appropriate decisions within a short time. The results indicate that the application of the Naive Bayes method improves the efficiency and accuracy of the report verification process. The developed system significantly reduces verification time and produces more consistent results than the manual process. The implementation of this system is expected to improve the quality of service for toll road users, minimize fare calculation errors, and increase customer satisfaction.

Keywords: Naive Bayes, Reporting Data, Vehicle Fares.

INTRODUCTION

The continuous growth of human needs has driven the demand for increasingly advanced technology. Technological progress facilitates and accelerates the fulfillment of the needs of individuals at all levels; hence, its benefits in solving various problems in today's highly mobile era are undeniable. One of the companies that play an important role in the management and operation of toll roads in Indonesia is responsible, in its operations, for ensuring smooth transactions at toll gates, including determining vehicle fares based on predefined vehicle classes. Accurate vehicle classification is a crucial aspect, as it directly affects the fairness of the fares charged to toll road users as well as the company's service image. Vehicle misclassification may occur due to technical factors in the toll gate system, limitations of automatic vehicle detection, or inconsistencies in vehicle data. When users submit refund reports for fares deemed inconsistent with their actual vehicle class, the current reporting process still relies on manual verification through the examination of transaction receipts, e-toll transaction histories, and CCTV footage at toll gates. Consequently, the process is relatively time-consuming and susceptible to subjectivity in decision-making. These problems indicate the need for a system-based approach capable of supporting objective and consistent decision-making in determining the eligibility of vehicle fare refund reports. The required system should not merely function as a data recorder but should also be able to classify reports based on a number of relevant variables, thereby automatically providing decision recommendations as to whether a report should be accepted or rejected. To address this issue, the Naive Bayes method is employed to classify vehicle fare refund reports into two classes, namely accepted and rejected, based on four variables: receipt evidence validity, e-toll transaction history, CCTV data verification, and vehicle class data conformity.

By applying the Naive Bayes method to the classification system for vehicle fare refund reports, the verification process is expected to become faster, more accurate, and more objective. The system is also expected to assist officers and supervisors in decision-making, reduce reliance on manual verification, and improve operational efficiency and service quality at the company.

LITERATURE REVIEW

1. Information Systems

A system is a group or set of closely interrelated elements that function together to achieve a particular objective. A system has several characteristics, including its components, environment, boundaries, interfaces, inputs, processing, outputs, and goals. Information, meanwhile, refers to a collection of data that has been processed into something more useful and meaningful to its users or recipients and that reduces uncertainty in the decision-making process regarding a particular situation. Accordingly, an information system can be defined as an activity that collects, analyzes, processes, stores, and disseminates information for a specific purpose. An information system can also be defined as a combination of technologies that supports operations and management. A system is an essential tool for an organization or company, as an integrated system enables the company's performance to be more focused and systematic.

2. The Naive Bayes Method

Naive Bayes is an algorithm used to predict probabilities in order to facilitate the classification of a given subject. The formula of the Naive Bayes algorithm can be written as follows:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} .$$

Where:

X = data with an unknown class

Y = the hypothesis that data X belongs to a specific class

P(Y|X) = the probability of hypothesis Y given condition X

P(Y) = the probability of hypothesis Y

P(X|Y) = the probability of X given hypothesis Y

P(X) = the probability of X.

3. Database

A database is a collection of information stored systematically in a computer so that it can be managed by a computer program to retrieve information. The term “database” originated in computer science, although its meaning was later extended to include non-electronic records. Database-like records existed before the Industrial Revolution in the form of ledgers, receipts, and collections of business data. A database is an arrangement or collection of data records stored in a computer, and the relationships between entries in a database can serve as a source of information for users. To date, many database records are still displayed as plain text to users, which constitutes one of the vulnerabilities that cryptanalysts may exploit to access, manipulate, leak, and distribute database records. A database is a shared collection of logically related data and their descriptions, designed to meet the information needs of a particular domain.

4. MySQL

MySQL is a DBMS application widely used by web application programmers. In MySQL, a database contains one or more tables. A table consists of a number of rows, and each row contains one or more columns. SQL is the standard language used to communicate with relational databases and is also the language used by many applications or tools to interact with database servers. SQL uses commands composed of simple words that resemble everyday human language.

Its functions include creating database objects, manipulating objects, populating database tables with data, updating existing data in tables, deleting data, performing database queries, controlling database access, and performing overall database administration. The first step in performing any database-related activity is to establish a connection to the database to be used. Establishing a database connection is required so that programmers can perform other database-related operations, such as adding, correcting, and deleting data. Since PHP does not provide specific functions for data operations, SQL (Structured Query Language) syntax is used.

5. PHP

PHP is a programming language used to build dynamic, static, and interactive websites. A website application should ideally be dynamic and interactive. Dynamic means that the website can display content

according to certain conditions (for example, displaying different products for each visitor). Interactive means that the website can provide feedback to the user (for example, displaying product search results). PHP is a server-side programming language; thus, PHP code is processed by the server, and the output is sent back to the browser. Therefore, one of the tools that must be available before starting PHP programming is a server.

METHOD

1. Research Stages

The waterfall method is a system development approach carried out sequentially or linearly. Thus, if the first stage has not been completed, the second, third, and subsequent stages cannot be performed. In other words, the third stage can only be carried out once the first and second stages have been completed.

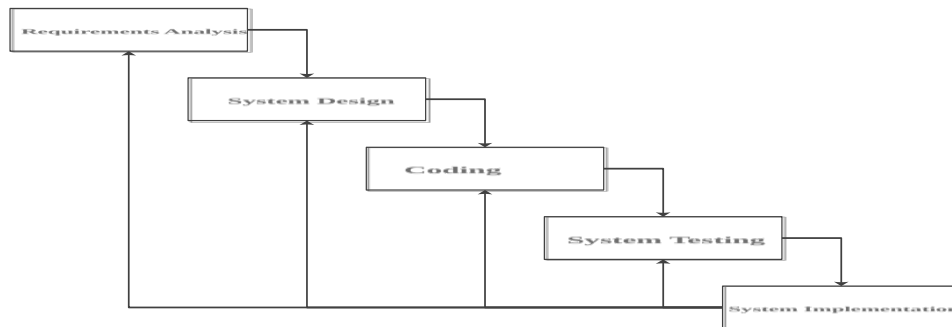


Figure 1. Waterfall Diagram

1. Requirements Analysis

At this stage, system requirements were identified to support the verification process for vehicle fare refund reports. The analyzed data included information from payment receipts, e-toll transaction histories, and the needs of officers and supervisors in the reporting process. The objective was to ensure that the designed system could effectively accelerate the classification and verification of reports.

2. System Design

Based on the results of the requirements analysis, the system was designed using an approach based on the Naive Bayes method. This stage included the design of the system architecture, the report verification workflow, and the user interface. The system design also involved the preparation of a relevant dataset for the classification process.

3. Coding

The coding stage involved implementing the system in accordance with the design. The Naive Bayes algorithm was developed to process report data and to make decisions regarding fare refund eligibility. In addition, a user interface was built to facilitate system operation by officers and supervisors.

4. System Testing

After the system was developed, testing was conducted to ensure that all functions operated according to specifications. Testing included validation of the dataset, evaluation of the accuracy of the Naive Bayes algorithm in classifying reports, and functional testing of the user interface. Testing was also performed to ensure that the system could handle data in real time and produce accurate decisions.

5. System Implementation

The final stage was the deployment of the system in the operational environment. Training was provided to officers and supervisors to ensure that they understood how to use the system. Following implementation, initial monitoring was carried out to evaluate system performance and to make adjustments where necessary.

2. Web-Based Implementation

This diagram illustrates the interactions between users (actors) and the system. The Use Case Diagram helps identify system functionality from the users' perspective, including the use cases (usage scenarios) and the relationships between the actors and the system, as follows:

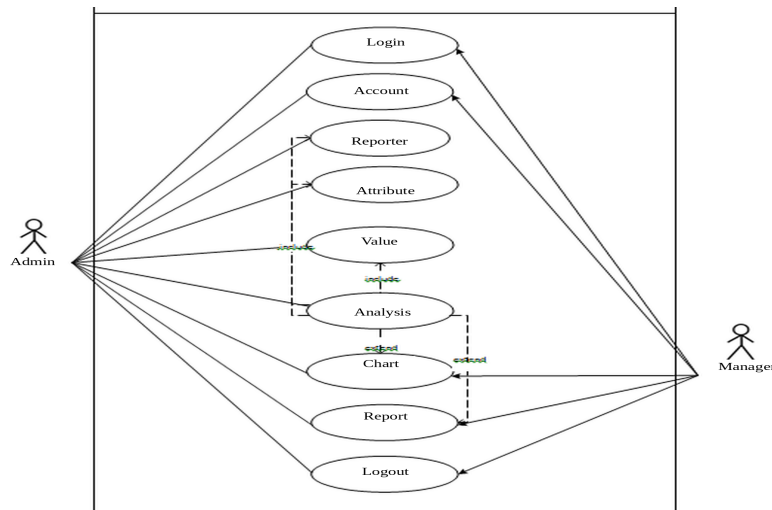


Figure 2. Use Case Diagram

3. Research Results Analysis

Through the application of the Naive Bayes method to vehicle fare refund report transaction data, this study demonstrates that a probability-based classification approach can be used effectively to support decision-making. The classification results generated by the system can assist officers and supervisors in verifying reports more objectively and consistently on the basis of historical data, thereby reducing reliance on manual verification. Furthermore, this study contributes a model for applying the Naive Bayes method in toll road services, particularly in the verification of vehicle fare refund reports. The resulting model and system are expected to serve as a reference for the development of similar decision support systems and as a basis for further research that develops other classification methods or incorporates additional supporting variables.

RESULTS AND DISCUSSION

At the company, the reporting process for fare refunds of misclassified vehicles is still carried out through manual verification by examining transaction receipts, e-Toll transaction histories, and toll gate CCTV footage. This process is relatively time-consuming and susceptible to subjectivity in decision-making. However, such report data exhibit patterns and characteristics that can be analyzed on the basis of historical data to determine report eligibility more objectively. The existing system is not yet able to integrate and classify report data automatically, which makes the verification process less efficient and less consistent. Therefore, a classification system based on the Naive Bayes method is required to process report data and generate accept or reject decisions based on probability calculations, thereby helping officers make fast, accurate, and data-driven decisions.

1. Determination Using the Naive Bayes Algorithm

The steps of the Naive Bayes calculation are as follows:

a. Variable Data

Variable data refer to a concept in research or analysis denoting characteristics, attributes, or values that can change or vary. The variables used in this study are as follows:

Table 1. Variable Data

No	Variabl	Subvariable	Weight
V1	Receipt Evidence Validity (0.30)	Information Legibility	0.08
		Data Completeness	0.08
		Information Accuracy	0.07
		Evidence Authenticity	0.07
V2	e-Toll Transaction History (0.30)	Transaction Pattern	0.08
		Fare Conformity	0.08
		Transaction Frequency	0.07
		Transaction Time	0.07
V3	CCTV Data Verification (0.20)	Vehicle Identification	0.05
		Vehicle Type	0.05
		Recording Time	0.05
		Recording Quality	0.05
V4	Vehicle Class Data Conformity (0.20)	Receipt and e-Toll Data	0.05
		Verification via CCTV	0.05
		Data Consistency Evaluation	0.05
		Fare Regulation Compliance	0.05

b. Training Data

Training data are a collection of data used to train models in machine learning or artificial intelligence. The dataset contains 100 samples consisting of inputs and their expected outputs, of which 15 samples are presented below. The algorithm uses these data to learn specific patterns, relationships, or characteristics, as described below:

Table 2. Training Data

No.	V1	V2	V3	V4	Total	Class
1	0.95	0.82	0.60	0.45	2.82	Rejected
2	1.12	0.97	0.55	0.40	3.04	Accepted
3	0.59	1.13	0.60	0.80	3.12	Accepted
4	0.62	0.98	0.40	0.60	2.60	Rejected
5	1.11	0.68	0.35	0.30	2.44	Rejected
6	0.60	0.75	0.50	0.40	2.25	Rejected
7	1.13	0.58	0.75	0.40	2.86	Rejected
8	0.75	0.46	0.65	0.60	2.46	Rejected
9	0.61	0.65	0.60	0.55	2.41	Rejected
10	0.83	0.67	0.45	0.55	2.50	Rejected
11	0.83	0.45	0.65	0.75	2.68	Rejected
12	0.98	0.83	0.75	0.80	3.36	Accepted
13	0.96	0.84	0.60	0.65	3.05	Accepted
14	0.84	1.10	0.65	0.60	3.19	Accepted
15	0.81	0.97	0.75	0.70	3.23	Accepted

c. Calculation of Probability P(H)

Probability Calculation for the Receipt Evidence Validity Variable

$$P(V1 = "0 \text{ to } 0.75" \mid \text{class} = "Accepted")$$

$$P(V1 = 3/40 = 0.08$$

$$P(V1 = "0.76 \text{ to } 1.5" \mid \text{class} = "Accepted")$$

$$P(V1 = 37/40 = 0.93$$

$$P(V1 = "0 \text{ to } 0.75" \mid \text{class} = "Rejected")$$

Probability Calculation for the e-Toll Transaction History Variable

$$P(V2 = "0 \text{ to } 0.75" \mid \text{class} = "Accepted")$$

$$P(V2 = 5/40 = 0.13$$

$$P(V2 = "0.76 \text{ to } 1.5" \mid \text{class} = "Accepted")$$

$$P(V2 = 35/40 = 0.88$$

$$P(V2 = "0 \text{ to } 0.75" \mid \text{class} = "Rejected")$$

$$P(V2 = 26/60 = 0.43$$

$$P(V_2 = "0.76 \text{ to } 1.5" \mid \text{class} = "Rejected")$$

$$P(V_2 = 34/60 = 0.57$$

Probability Calculation for the CCTV Data Verification Variable

$$P(V_3 = "0 \text{ to } 0.5" \mid \text{class} = "Accepted")$$

$$P(V_3 = 11/40 = 0.28$$

$$P(V_3 = "0.6 \text{ to } 1" \mid \text{class} = "Accepted")$$

$$P(V_3 = 29/40 = 0.73$$

$$P(V_3 = "0 \text{ to } 0.5" \mid \text{class} = "Rejected")$$

$$P(V_3 = 37/60 = 0.62$$

$$P(V_3 = "0.6 \text{ to } 1" \mid \text{class} = "Rejected")$$

$$P(V_3 = 23/60 = 0.38$$

Probability Calculation for the Vehicle Class Data Conformity Variable

$$P(V_4 = "0 \text{ to } 0.5" \mid \text{class} = "Accepted")$$

$$P(V_4 = 13/40 = 0.33$$

$$P(V_4 = "0.6 \text{ to } 1" \mid \text{class} = "Accepted")$$

$$P(V_4 = 27/40 = 0.68$$

$$P(V_4 = "0 \text{ to } 0.5" \mid \text{class} = "Rejected")$$

$$P(V_4 = 33/60 = 0.55$$

$$P(V_4 = "0.6 \text{ to } 1" \mid \text{class} = "Rejected")$$

$$P(V_4 = 27/60 = 0.45$$

d. Calculating the Number of Matching Cases in the Same Class, $P(X|H)$

Receipt Evidence Validity = 0.91

$$P(X \mid \text{Class "Accepted"})$$

$$= 37 / 40 = 0.93$$

$$P(X \mid \text{Class "Rejected"})$$

$$= 31 / 60 = 0.52$$

e-Toll Transaction History = 1.31

$$P(X \mid \text{Class "Accepted"})$$

$$= 35 / 40 = 0.88$$

$$P(X \mid \text{Class "Rejected"})$$

$$= 34 / 60 = 0.57$$

CCTV Data Verification = 0.6

$$P(X \mid \text{Class "Accepted"})$$

$$= 29 / 40 = 0.73$$

$$P(X \mid \text{Class "Rejected"})$$

$$= 23 / 60 = 0.38$$

Vehicle Class Data Conformity = 0.7

$$P(X \mid \text{Class "Accepted"})$$

$$= 27 / 40 = 0.68$$

$$P(X \mid \text{Class "Rejected"})$$

$$= 27 / 60 = 0.45$$

e. Multiplying All Variables by Class, $P(X|H) \times P(H)$

$$P(X \mid \text{Class "Accepted"}) \times P(\text{Class "Accepted"})$$

$$= (0.93 \times 0.88 \times 0.73 \times 0.68) \times 0.40$$

$$= 0.158$$

$$P(X \mid \text{Class "Rejected"}) \times P(\text{Class "Rejected"})$$

$$= (0.52 \times 0.57 \times 0.38 \times 0.45) \times 0.60$$

$$= 0.030$$

f. Probability Normalization:

$$= P(\text{Class "Accepted"}) \frac{0.158}{0.158 + 0.030}$$

$$= 0.839$$

$$= P(\text{Class "Rejected"}) \frac{0.030}{0.158 + 0.030}$$

$$= 0.161$$

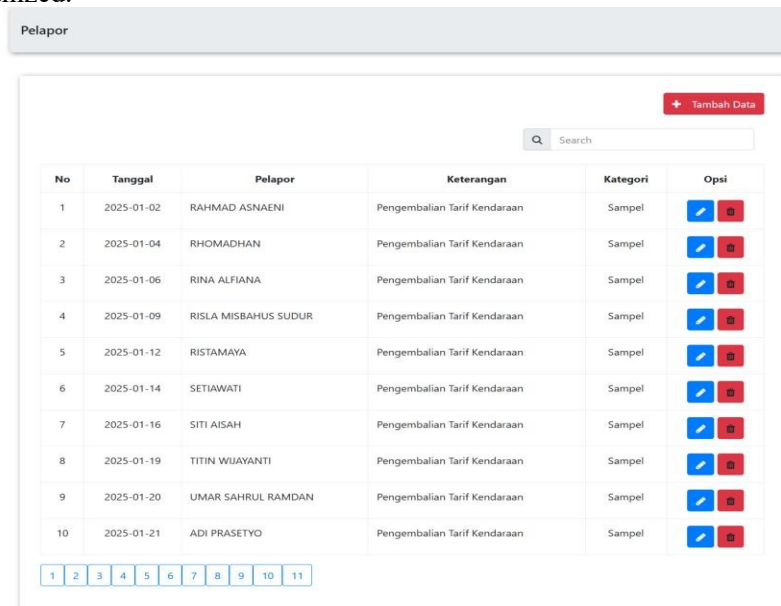
Hence, based on the Naive Bayes calculation, the highest value is 0.839, obtained for the "Accepted" class. Therefore, the fare refund for the misclassified vehicle can be "Accepted", with a percentage of $0.8394 \times 100 = 83.94\%$.

2. Web-Based Implementation

The interfaces resulting from the developed system design are presented below:

a. Reporter Data Page

This menu can be accessed only by the administrator to manage reporter information. In this menu, the administrator can add, edit, and delete reporter data. The managed data are organized into several fields, namely date, reporter name, description, and category. This page serves as the central management hub for reporter data to ensure that the recorded information is always accurate, detailed, and well organized.



No	Tanggal	Pelapor	Keterangan	Kategori	Opsi
1	2025-01-02	RAHMAD ASNAENI	Pengembalian Tarif Kendaraan	Sampel	 
2	2025-01-04	RHOMADHAN	Pengembalian Tarif Kendaraan	Sampel	 
3	2025-01-06	RINA ALFIANA	Pengembalian Tarif Kendaraan	Sampel	 
4	2025-01-09	RISLA MISBAHUS SUDUR	Pengembalian Tarif Kendaraan	Sampel	 
5	2025-01-12	RISTAMAYA	Pengembalian Tarif Kendaraan	Sampel	 
6	2025-01-14	SETIAWATI	Pengembalian Tarif Kendaraan	Sampel	 
7	2025-01-16	SITI AISAH	Pengembalian Tarif Kendaraan	Sampel	 
8	2025-01-19	TITIN WIJAYANTI	Pengembalian Tarif Kendaraan	Sampel	 
9	2025-01-20	UMAR SAHRUL RAMDAN	Pengembalian Tarif Kendaraan	Sampel	 
10	2025-01-21	ADI PRASETYO	Pengembalian Tarif Kendaraan	Sampel	 

Figure 3. Reporter Data Page

b. Analysis Process Page

This menu can be accessed only by the administrator to perform classification calculations. When the administrator clicks the "Proses Analisa" (Process Analysis) button, the system displays the complete calculation details of the Naive Bayes method, including the probability of each class based on the entered attribute data and values. This menu also includes a "Simpan Hasil Analisa" (Save Analysis Results) button, which stores the computed conclusions or classification results in the system. This page serves as the core module that transforms data into statistically based information for decision-making.

APPLICATION OF THE NAIVE BAYES METHOD FOR DETERMINING THE ELIGIBILITY OF TOLL FARE REFUND REPORTS FOR MISCLASSIFIED VEHICLES

Fitri Pranita Nasution et al

Data Uji: MIFTAH GUFRANI					
Nama pelapor	Validitas Bukti Struk	Riwayat Transaksi e-Toll	Verifikasi Data dari CCTV	Kesesuaian Data Golongan	Kelas
MIFTAH GUFRANI	0.91	1.13	0.60	0.70	?

Perhitungan Probabilitas untuk MIFTAH GUFRANI

Atribut	Nilai Uji	Diterima		Ditolak	
		Jumlah	Probabilitas	Jumlah	Probabilitas
Validitas Bukti Struk	0.91	37	0.925	31	0.517
Riwayat Transaksi e-Toll	1.13	35	0.875	34	0.567
Verifikasi Data dari CCTV	0.60	29	0.725	23	0.383
Kesesuaian Data Golongan	0.70	27	0.675	27	0.450
Hasil Perkalian (PI)		0.396088		0.050504	
Probabilitas Kelas		0.400		0.600	
Hasil Final (PIC) x (PI)		0.158435		0.030303	
Hasil Normalisasi		0.839446		0.160554	

Kesimpulan untuk MIFTAH GUFRANI

Hasil dari perhitungan menggunakan metode **Naive Bayes** berdasarkan nilai tertinggi yaitu sebesar **0.8394465** untuk kategori **"Diterima"**. Sehingga data uji ini masuk kedalam kategori **"Diterima"**, dengan tingkat persentase **83.94 %**.

Simpan Hasil Analisis

Figure 4. Analysis Process Page

CONCLUSION

This study successfully developed a system for classifying the eligibility of fare refund reports for misclassified vehicles based on receipt evidence data, e-toll transaction history, CCTV verification, and vehicle class data conformity. The Naive Bayes method was able to classify vehicle fare refund reports into accepted and rejected classes objectively and on a probabilistic basis, thereby making the decision-making process faster and more consistent than manual verification. For future development, it is recommended to increase the quantity and variety of report data so that the Naive Bayes classification model can achieve higher accuracy that is more representative of actual conditions. The system can be further developed by integrating other classification methods for comparison in order to identify the most suitable method for determining the eligibility of vehicle fare refund reports.

REFERENCES

- Aswiputri, M. (2022). Literature Review Determinasi Sistem Informasi Manajemen: Database, Cctv Dan Brainware. *Jurnal Ekonomi Manajemen Sistem Informasi*, 3(3), 312-322.
- Cahyo, A.D. (2023). Metode Naïve Bayes Untuk Klasifikasi Masa Studi Sarjana. *Jurnal Teknologi Pintar*, 3(4).
- Dari, W., & Hanayah, E. T. (2023). Analisis Tingkat Kepuasan Pengguna Aplikasi Ojek Online Dengan Metode Naive Bayes. *INSOLOGI: Jurnal Sains dan Teknologi*, 2(1), 221-232.
- Faulina, S. T., & Lestari, N. (2021). Penerapan Metode Waterfall pada Aplikasi Pemesanan Soundsystem Dan Organ Tunggal Jefri. *JIK: Jurnal Informatika dan Komputer*, 12(2), 65-79.
- Fitriani, Y., Utami, S., & Junadi, B. (2022). Perancangan Sistem Informasi Human Capital Management Berbasis Website. *Journal of Information System, Applied, Management, Accounting and Research*, 6(4), 792-803.
- Ke, B., Khandelwal, et al. (2021). Rock-burst Occurrence Prediction Based On Optimized Naïve Bayes Models. *IEEE Access*, 9, 91347-91360.
- Kewsuwun, N., & Kajornkasirat, S. (2022). A Sentiment Analysis Model Of Agritech Startup On Facebook Comments Using Naive Bayes Classifier. *International Journal of Electrical & Computer Engineering* (2088-8708), 12(3).
- Nasution, F.P. (2024). Penerapan Metode Additive Ratio Assessment Dalam Sistem Pendukung Keputusan Untuk Menentukan Pemberian Bonus Delivery Pada CV Inside Computer House Medan. *Jurnal JUREKSI*, 2(3), 906-919.