Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer Vol. 21, No. 1, Februari 2026 1
e-ISSN 2597-4963 dan p-ISSN 1858-4853 DOI : 10.30872/jim.v21i1.25392

Traffic Accident Prediction Using Machine Learning Based on PT Jasa Raharja Data

Muhammad Fikri Utomo[1*)], Muhammad Fikry[2)], Defry Hamdhana[3)], Dahlan Abdullah[4)], Nurdin[5)]

[1,2,3,4,5] Teknologi Informasi, Universitas Malikussaleh, Lhokseumawe, 24355, Indonesia
E-Mail : muhammad.237110201006@mhs.unimal.ac.id *[1)]; muh.fikry@unimal.ac.id[2)]; defryhamdhana@unimal.ac.id[3)]; dahlan@unimal.ac.id[4)]; nurdin@unimal.ac.id[5)];

ABSTRAK

Traffic accidents represent a critical issue that significantly affects public safety and generates substantial social and economic impacts, particularly within the operational area of PT. Jasa Raharja Lhokseumawe Branch. The lack of predictive information regarding accident occurrences often results in reactive policy making. This study aims to develop a machine learning–based forecasting model for traffic accident rates using a combination of K-Means Clustering and Recurrent Neural Network (RNN). The dataset consists of historical traffic accident records from 2022 to 2024, which were preprocessed and aggregated on a weekly basis at the district level. K-Means Clustering was employed to group districts according to weekly accident patterns, resulting in two optimal clusters based on silhouette score evaluation. Subsequently, separate RNN models were developed for each cluster to forecast weekly accident occurrences. Model performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The results indicate that the RNN model achieved higher prediction accuracy for clusters with more stable accident patterns compared to clusters exhibiting higher fluctuation. Overall, the proposed combination of clustering and RNN demonstrates strong potential in producing accurate traffic accident forecasts

Kata Kunci –*Traffic Accidents, Forecasting, Machine Learning, K-Means Clustering, Recurrent Neural Network (RNN)*

1. INTRODUCTION

A traffic accident is an event that occurs on a highway involving vehicles, pedestrians, or both, resulting in damage, injury, or death (Apriyanti et al., 2020). The average traffic accident rate in 2020–2023 in the Lhokseumawe Branch of PT. Jasa Raharja reached 2.3 cases per day. Indonesia is one of the countries with the highest traffic accident fatality rate in the ASEAN region, with 103,645 cases in the 2019–2021 period. Globally, the WHO estimates 1.17 million deaths per year due to traffic accidents.

Aceh Province ranks in the top six accident-prone areas with a death toll reaching 6,000 by 2023. The high number of accidents creates obstacles in the form of limited and delayed accurate and predictive information, thus hampering decision-making and policy formulation by the government, police, and PT. Jasa Raharja.

The main problem faced by stakeholders—the government, the police, and PT. Jasa Raharja—is the lack of fast, accurate, and data-based accident prediction information. The government requires predictive data for accurate road safety program planning and budget allocation. The police require predictive information to determine patrol patterns and address accident-prone areas.

Meanwhile, PT. Jasa Raharja requires accident predictions for compensation budget planning, financial risk management, and accident prevention programs. Without sound forecasting, policies tend to be reactive and inefficient.

This study applies machine learning methods, such as K-Means Clustering and Recurrent Neural Network, to predict the number of traffic accidents more accurately to support data-driven decision-making by stakeholders. This prediction is expected to help formulate more effective policies and preventative measures through forecasting techniques. Previous studies used Support Vector Regression based on annual data and have limitations in capturing short-term fluctuations.

Therefore, this study proposes accident forecasting with a weekly period in the work area of PT. Jasa Raharja Lhokseumawe Branch so that the resulting information is more relevant, detailed, and useful for rapid and appropriate preventive actions by the government, police, and PT. Jasa Raharja.

This study first grouped regions using a clustering method to identify sub-districts with similar accident characteristics. Clustering, an unsupervised learning method, groups data based on similar features such as location, time, and accident frequency.

This grouping aims to improve the focus and accuracy of the forecasting model, as areas within a cluster have a high degree of similarity (Chusyairi & Ramadar Noor Saputra, 2019).

RNN is an effective neural network for sequential data because it is able to store previous information, making it suitable for predicting time-based accident trends. In this study, clustering is used to group areas with similar characteristics, while RNN is applied to each cluster to perform more specific forecasting.

The model is trained using historical accident data from PT. Jasa Raharja Lhokseumawe Branch to predict the number of accidents and identify priority areas for handling. This study aims to support data-based accident prevention and more appropriate

DOI : 10.30872/jim.v21i1.25392 e-ISSN 2597-4963 dan p-ISSN 1858-4853

decision-making. The proposed hypothesis is that the combination of Clustering and RNN produces more accurate accident forecasting than the Polynomial kernel SVR method, making it more effective for the government, police, and PT. Jasa Raharja in formulating preventive policies.

2. HERITAGE REVIEW

Traffic is a crucial infrastructure that plays a major role in supporting smooth development. Its presence facilitates public access to various activities, including economic ones. Imagine how difficult it would be to get to work or reach a work location without traffic. Almost all activities, especially those related to mobility, depend heavily on a good traffic system (Enggarsasi & Sa'diyah, 2017).

Government Regulation Number 43 of 1993, Article 93 concerning Road Infrastructure and Traffic, states that a traffic accident is an unexpected event that accidentally involves vehicles and/or other road users that can result in human casualties or property damage.

PT Jasa Raharja (Persero) is a state-owned enterprise engaged in social insurance to provide protection to victims of traffic accidents. Established in 1960 and incorporated in the holding Indonesia Financial Group (IFG), Jasa Raharja provides compensation for public transportation passengers and road users based on Law No. 33 of 1964 and Law No. 34 of 1964.

This protection includes passengers of official public transportation as well as victims of road traffic accidents outside vehicles, such as pedestrians, with funding sources coming from the Compulsory Contribution of Public Motor Vehicles (IWKBU) and Compulsory Contributions to Road Traffic Accident Funds (SWDKLLJ).

Table 1. Data on Compensation Distribution

WILAYAH	S.D. DESEMBER 2023 - Korban (Orang)	S.D. DESEMBER 2023 - Santunan (Milyar)	Aktivitas % - Korban	Aktivitas % - Santunan	Kontribusi Nasional % - Korban	Kontribusi Nasional % - Santunan
Jakarta	7,877	93.12	7.74%	6.26%	3.51%	3.02%
JaBar	19,068	356.34	2.51%	1.67%	8.50%	11.57%
JaTeng	41,247	629.95	3.19%	4.62%	18.38%	20.45%
JaTim	48,792	661.59	7.90%	9.01%	21.74%	21.47%
Bali	10,267	71.57	75.44%	21.70%	4.58%	2.32%
Aceh	6,113	72.54	1.16%	-1.23%	2.72%	2.35%
SumSel	3,966	63.5	1.96%	1.80%	1.32%	2.06%
SulSel	13,337	121.23	13.43%	0.96%	5.94%	3.93%
DIY	9,562	109.08	-12.14%	0.96%	4.26%	3.54%
Papua	4,365	38.13	6.08%	22.79%	1.95%	1.24%
SulTeng	2,577	30.18	17.40%	1.59%	1.15%	0.98%
KalTeng	1,848	24.23	26.75%	16.36%	0.82%	0.79%
Banten	5,261	96.26	4.26%	3.56%	2.34%	3.12%
Kep. Riau	1,880	21.86	25.50%	23.77%	0.84%	0.71%
Total	224,402	3080.87	6.65%	4.41%		

A. *Data Mining*

Data mining is a process that aims to extract hidden knowledge and information from large data sets, even if the data is incomplete, contains noise, is ambiguous, or is random (Huang et al., 2020). In general, data mining includes various analytical approaches such as machine learning, artificial intelligence (AI), statistics, and deep learning in its processing (Simon, 2025).

B. *Forecasting*

Forecasting is a method for planning and controlling production to address future uncertainty. More specifically, it is used to predict future product demand (Apriyanti et al., 2020).

Forecasting is the prediction, projection, or estimation of the extent of uncertain future events. This forecasting method emphasizes the exponential de-prioritization of previously observed objects. Explanatory smoothing involves one or more explicitly specified smoothing parameters, and these results determine the weights assigned to the observed values (Yuniarsyih R.A et al., 2025).

Forecasting is the estimation of events that have not yet occurred. This process requires an appropriate forecasting method. Forecasting is based on the analysis of historical data using specific techniques. The accuracy of research results depends heavily on the accuracy of the predictions made.

To develop a system for predicting the number of prospective new students, a reliable forecasting method and accurate calculations are required to estimate the number of applicants. In the forecasting process, calculations are performed repeatedly using the latest data to improve prediction accuracy (Lusiana & Yuliarty, 2020).

C. *Clustering*

Clustering is a grouping process based on the similarities of members in each partition in a particular matrix (Ananta & Sofro, 2025). The K-Means method is one of the most widely used clustering methods and has been applied in various fields of science and technology. One of the main problems of the K-Means algorithm is its possibility of producing empty clusters depending on the initial centroid vector.

The K-Means method is a partitioning algorithm because K-Means is based on determining the initial number of clusters by determining the initial centroid value. The K-Means algorithm has been widely used because of its simple concept, easy implementation, and high efficiency when processing large-scale data (Matdoan et al., 2023).

e-ISSN 2597-4963 dan p-ISSN 1858-4853 DOI : 10.30872/jim.v21i1.25392

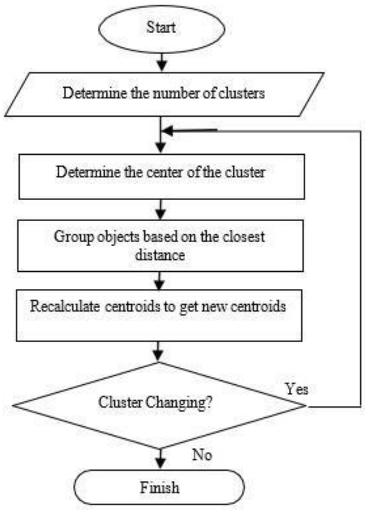

Figure 1. Flow Process of K-Means Clustering

D. *Recurrent neural networks* (RNN)

A Recurrent Neural Network (RNN) is a type of neural network commonly used for processing sequential input data. RNNs are highly flexible and have been used to solve a wide range of problems, such as time series problems, speech recognition, language modeling, machine translation, sentiment analysis, and image captioning, to name a few (Ramadhan et al., 2025).

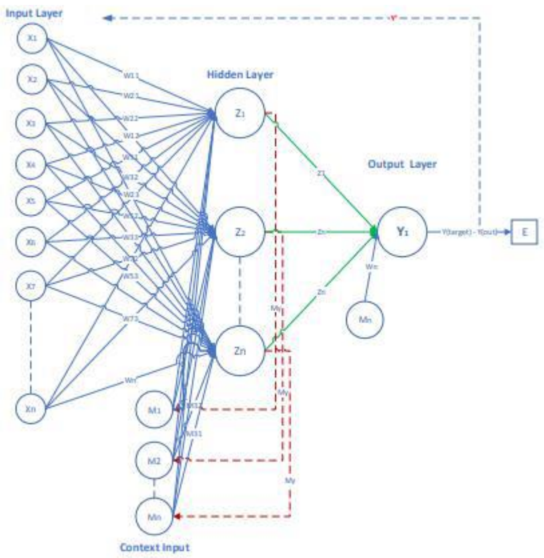

Figure 2. Illustration of RNN Architecture

3. RESEARCH METHODOLOGY

This study aims to forecast the number of traffic accidents in the work area of PT. Jasa Raharja Lhokseumawe Branch using the K-Means Clustering and Recurrent Neural Network methods. The research stages include data selection and pre-processing, clustering, forecasting, and evaluation. Accident data is processed and grouped by sub-district using K-Means with the number of clusters determined through the silhouette score.

The clustering results then become the basis for applying RNN to forecast the number of accidents in each cluster. The methodology is arranged systematically so that the data processing and modeling process takes place in a structured, measurable, and scientifically accountable manner in accordance with the research objectives.

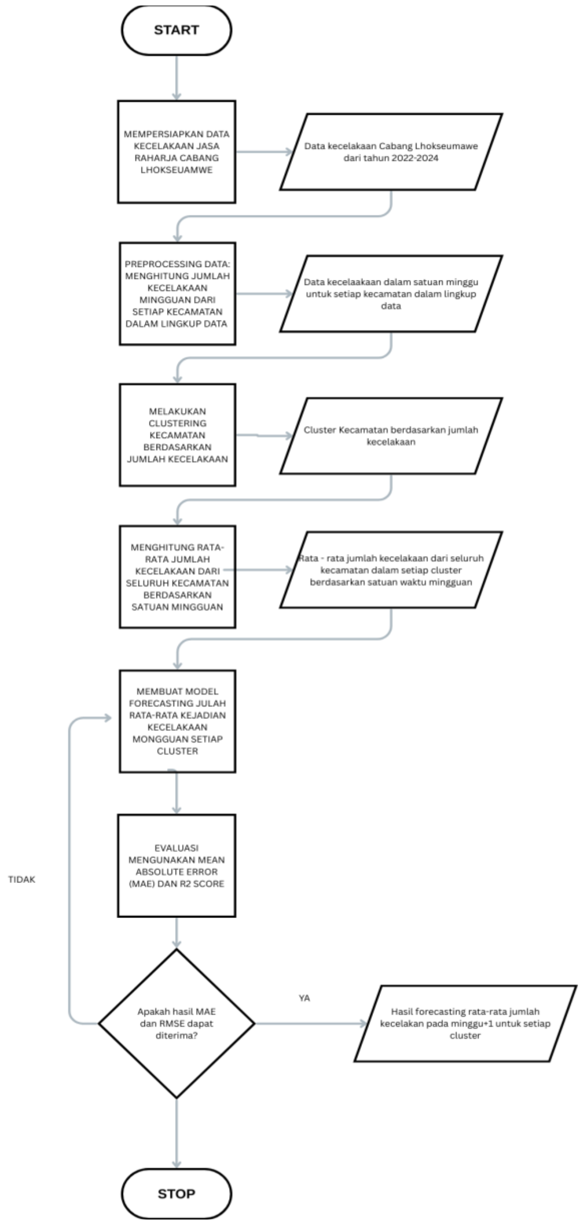

Figure 3. Research steps for forecasting the number of weekly accidents based on data from PT. Jasa Raharja Lhokseumawe

The research flow encompasses an integrated process from accident data collection to forecasting model evaluation by combining clustering and forecasting. The research stages include data pre-processing to ensure quality and consistency, sub-district clustering using K-Means based on accident patterns, compilation of weekly time series datasets per cluster, and RNN modeling and training to capture temporal patterns of accidents. The final stage is model performance evaluation using MAE, MSE, and RMSE to ensure the accuracy and reliability of the forecasting results.

DOI : 10.30872/jim.v21i1.25392 e-ISSN 2597-4963 dan p-ISSN 1858-4853

A. Data Preparation and Pre-processing

This study uses traffic accident data from PT. Jasa Raharja Lhokseumawe Branch for the period 2022 to 2024. The accident data covers Lhokseumawe City, Bireuen Regency, North Aceh Regency, Central Aceh Regency, and Bener Meriah Regency, recorded in accident case units. An example of the data is shown in Table 2 below.

Table 2. Sample Accident Case Data

Lokasi	Kecamatan	Kabupaten	Nama Polres	No Laporan Polisi	Waktu Kejadian	Tanggal Kejadian	Deskripsi Lokasi	Kasus Kecelakaan	Nama Korban	Umur Korban	Alamat Korban	Jenis Kelamin Korban	Sifat Cidera
0800201	KEC. JANGKA	KAB.BIREUEN	POLRES BIREUEN	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES BIREUEN/POLDA ACEH	20.00	01/02/24	JL COT RABO - LINGGONG DESA PULO IBOIH KEC. JANGKA KAB. BIREUEN. , RT -, RW -, TITIK KOORDINAT 5°1410.6"N 96°4557.5"E 5.236266, 96.765969	NORMAL/ TABRAKAN BELAKANG-SAMPING	MUSTAQIM	11th	DS. JANGKA KEUTAPANG KEC. JANGKA KAB. BIREUEN	P	LL
0800201	KEC. JANGKA	KAB.BIREUEN	POLRES BIREUEN	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES BIREUEN/POLDA ACEH	20.00	01/02/24	JL COT RABO - LINGGONG DESA PULO IBOIH KEC. JANGKA KAB. BIREUEN. , RT -, RW -, TITIK KOORDINAT 5°1410.6"N 96°4557.5"E 5.236266, 96.765969	NORMAL/ TABRAKAN BELAKANG-SAMPING	AULIA RAHMAT	23th	DS. MATA MAMPLAM KEC. JANGKA KAB. BIREUEN	P	LL
0800201	KEC. JANGKA	KAB.BIREUEN	POLRES BIREUEN	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES BIREUEN/POLDA ACEH	20.00	01/02/24	JL COT RABO - LINGGONG DESA PULO IBOIH KEC. JANGKA KAB. BIREUEN. , RT -, RW -, TITIK KOORDINAT 5°1410.6"N 96°4557.5"E 5.236266, 96.765969	NORMAL/ TABRAKAN BELAKANG-SAMPING	MUKHLIS	19th	DS. JANGKA KEUTAPANG KEC. JANGKA KAB. BIREUEN	P	MD
0800201	KEC. BEBESEN	KAB. ACEH TENGAH	POLRES ACEH TENGAH	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES ACEH TENGAH/POLDA ACEH	23.30	01/03/24	JL TAKENGON-BIREUEN DESA MONGAL KECAMATAN BEBESEN KABUPATEN ACEH TENGAH, RT -, RW -, TITIK KOORDINAT -	NORMAL/ KECELAKAAN TUNGGAL /SLIP SENDIRI	NAZRIEL ANSYAR	15th	DESA MONGGAL KEC. BEBESEN KAB. ACEH TENGAH	P	LL
0800201	KEC. BEBESEN	KAB. ACEH TENGAH	POLRES ACEH TENGAH	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES ACEH TENGAH/POLDA ACEH	23.30	01/03/24	JL TAKENGON-BIREUEN DESA MONGAL KECAMATAN BEBESEN KABUPATEN ACEH TENGAH, RT -, RW -, TITIK KOORDINAT -	NORMAL/ KECELAKAAN TUNGGAL /SLIP SENDIRI	KHAIRUL AMRI	26th	DESA MONGGAL KEC. BEBESEN KAB. ACEH TENGAH	P	LL
0800201	KEC. BAKTIYA	KAB. ACEH UTARA	POLRES ACEH UTARA	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES ACEH UTARA/POLDA ACEH	20.15	01/01/24	JL BANDA ACEH MENUJU MEDAN,KEL. ALUE IE PUTEH KEC. BAKTIYA KAB. ACEH UTARA	NORMAL/ TABRAKAN DEPAN-DEPAN	MIRZA FAHMI	23th	KEL. RAWANG ITIK , KEC. TANAH JAMBO AYE , KAB ACEH UTARA	P	LL
0800201	KEC. JEUMPA	KAB.BIREUEN	POLRES BIREUEN	LP/A/1/I/2025/SPKT.SATLANTAS/POLRES BIREUEN/POLDA ACEH	12.00	27/12/24	JL DI N LINTAS MEDAN - BANDA ACEH DESA BLANG BLADEH KEC. JEUMPA KAB. BIREUEN., RT -, RW -, TITIK KOORDINAT 5°1222.4"N 96°3960.0"E 5.206211, 96.666657	NORMAL/ TABRAKAN DEPAN-SAMPING	MUSTAFA	25th	KEL. KULU KUTA , KEC. KUTA BLANG , KOTA BIREUEN	P	MD
0800201	KEC. SILIH NARA	KAB. ACEH TENGAH	POLRES ACEH TENGAH	LP/A/1/I/2025/SPKT.SATLANTAS/POLRES ACEH TENGAH/POLDA ACEH	16.30	31/12/24	JL TAKENGON - BEUTONG DESA SIMPANG KEMILI KECAMATAN SILIH NARA KABUPATEN ACEH TENGAH, RT -, RW -, TITIK KOORDINAT	NORMAL/ TABRAKAN DEPAN-DEPAN	YASIR	18th	DESA SIMPANG KEMILI , KEL. SIMPANG KEMILI , KEC. SILIH NARA , KOTA ACEH TENGAH , PROVINSI ACEH	P	LL
0800201	KEC. SILIH NARA	KAB. ACEH TENGAH	POLRES ACEH TENGAH	LP/A/1/I/2025/SPKT.SATLANTAS/POLRES ACEH TENGAH/POLDA ACEH	16.30	31/12/24	JL TAKENGON - BEUTONG DESA SIMPANG KEMILI KECAMATAN SILIH NARA KABUPATEN ACEH TENGAH, RT -, RW -, TITIK KOORDINAT	NORMAL/ TABRAKAN DEPAN-DEPAN	JUWITA	23th	DESA SIMPANG KEMILI , KEL. SIMPANG KEMILI , KEC. SILIH NARA , KOTA ACEH TENGAH , PROVINSI ACEH	W	LL
...	...	...	...	...	...	...	...	...	...	...	...	...	...
0800201	KEC. PINTU RIME GAYO	KAB. BENER MERIAH	POLRES BENER MERIAH	LP/B/89/VIII/2024/SPKT/POLRES BENER MERIAH/POLDA ACEH	17.00	21/08/24	JL TAKENGON - BIREUEN DESA ALUR CINCIN KEC PINTU RIME GAYO KAB BENER MERIAH	NORMAL/ TABRAKAN DEPAN-DEPAN	MUHAMMAD HANIF	19th	DSN. BALEE LABANG DS. BLANG COT BAROH KEC. JEUMPA KAB. BIREUN	P	LL
0800201	KEC. PINTU RIME GAYO	KAB. BENER MERIAH	POLRES BENER MERIAH	LP/B/93/IX/2024/SPKT/POLRES BENER MERIAH/POLDA ACEH	18.00	29/08/24	JL SOSIAL - MENDEREK KEC PINTU RIME GAYO KAB BENER MERIAH	NORMAL/ TABRAKAN DEPAN-DEPAN	WIN SIMAHATE	16th	JL. MESJID TUHA DSN. MESJID TUHA DS. IE MASEN ULEE KARENG KEC. ULEE KARENG KOTA BANDA ACEH	P	LL
0800201	KEC. PINTU RIME GAYO	KAB. BENER MERIAH	POLRES BENER MERIAH	LP/B/93/IX/2024/SPKT/POLRES BENER MERIAH/POLDA ACEH	18.00	29/08/24	JL SOSIAL - MENDEREK KEC PINTU RIME GAYO KAB BENER MERIAH	NORMAL/ TABRAKAN DEPAN-DEPAN	RAMLAH	64th	DS. SINGGAH MULO KEC. PINTU RIME GAYO KAB. BENER MERIAH	W	LL
0800201	KEC. WIH PESAM	KAB.BENER MERIAH	POLRES BENER MERIAH	LP/B/96/IX/2024/SPKT/POLRES BENER MERIAH/POLDA ACEH	18.30	09/10/24	JL SIMPANG BALEK - BLANG MANCUNG DESA BLANG PAKU KEC WIH PESAM KAB BENE MERIAH	NORMAL/ KECELAKAAN TUNGGAL /SLIP SENDIRI	ANDI PRIANTO	27th	DS. WONO SOBO KEC. WIH PESAM KAB. BENER MERIAH	P	LL
0800201	KEC. WIH PESAM	KAB.BENER MERIAH	POLRES BENER MERIAH	LP/B/97/IX/2024/SPKT/POLRES BENER MERIAH/POLDA ACEH	08.00	13/09/24	JL SIMPANG LAPANGAN BOLA - BANDARA REMBELE DESA PANTE RAYA KEC WIH PESAM KAB BENER MERIAH	NORMAL/ TABRAKAN DEPAN-SAMPING	DESSY SALFIANA RAHMAWATI	41th	DS. MEKAR AYU KEC. TIMANG GAJAH KAB. BENER MERIAH	W	LL
0800201	KEC. TIMANG GAJAH	KAB. BENER MERIAH	POLRES BENER MERIAH	LP/B/99/IX/2024/SPKT/POLRES BENER MERIAH/POLDA ACEH	18.00	15/09/24	JL TAKENGON - BIREUN DESA LAMPAHAN BARAT KEC TIMANG GAJAH KAB BENER MERIAH	NORMAL/ TABRAKAN DEPAN-DEPAN	SAFRIZAL EPENDI	20th	DSN. SARA PAKAT DS. LINUNG BALE KEC. TIMANG GAJAH KAB. BENER MERIAH	P	LL
0800201	KEC. BAKTIYA	KAB. ACEH UTARA	POLRES ACEH UTARA	LP/A/1/I/2024/SPKT.SATLANTAS/POLRES ACEH UTARA/POLDA ACEH	20.15	01/01/24	JL BANDA ACEH MENUJU MEDAN,KEL. ALUE IE PUTEH KEC. BAKTIYA KAB. ACEH UTARA	NORMAL/ TABRAKAN DEPAN-DEPAN	INTAN GUSMALA WATI LUBIS	17th	KEL. GAMPONG KEUDE (KAMPUNG KEUDE) , KEC. DARUL AMAN , KAB ACEH TIMUR	W	LL
0800201	KEC. WIH PESAM	KAB.BENER MERIAH	POLRES BENER MERIAH	LP/A/14/XII/2024/SPKT.SATLANTAS/POLRES BENER MERIAH/POLDA ACEH	17.00	12/06/24	JL SIMPANG BALIK - BLANG MANCUNG DESA KEBUN BARU KEC WIH PESAM KAB BENER MERIAH	NORMAL/ TABRAKAN DEPAN-DEPAN	BAHTIAR	48th	DS. BALE ATU KEC. BUKIT KAB. BENER MERIAH	P	LL
0800201	KEC. SIMPANG MAMPLAM	KAB.BIREUEN	POLRES BIREUEN	LP/A/36/II/2024/SPKT.SATLANTAS/POLRES BIREUEN	03.10	02/01/24	JL. MEDAN BANDA ACEH DESA MNS MAMPLAM KEC. SIMPANG MAMPLAM KAB. BIREUEN	NORMAL/ TABRAKAN DEPAN-BELAKANG	AFZALUL AKBAR	15th	DS. COT TRIENG KEC. SIMPANG MAMPLAM KAB. BIREUEN	P	MD
0800201	KEC. DEWANTARA	KAB.ACEH UTARA	ADIAN/I/PAM UTARA	PT KAI/LAKA KA, PWK LHOKSEUMAWE 02/LP-KEJADIAN/I/PAM.SUBDIV I.1-2024	16.00	24/12/24	KM 10+600 PETAK JALAN STA. KRUENG GEUKEH (KRG) - STA. BUNGKAIH (BKH)	NORMAL/ TABRAKAN DENGAN KERETA API	MUHAMMAD WALI LINGE	21th	DUSUN SUKA DAMAI KEL. BANGKA JAYA KEC. DEWANTARA KAB. ACEH UTARA	P	LL

Table 2. provides information regarding traffic accident case data reports, including the time of the incident, location, victim's name, victim's address,

e-ISSN 2597-4963 dan p-ISSN 1858-4853 DOI : 10.30872/jim.v21i1.25392

and the police station handling the case. This data covers 72 sub-districts across five regencies/cities. This data will be aggregated based on the number of accidents per week for each sub-district.

A sample of this aggregated data is provided in Table 3. Data preparation and pre-processing were conducted to ensure that the accident data used was clean, structured, and ready for processing in both clustering and forecasting. The base data used were traffic accident case reports from PT. Jasa Raharja, Lhokseumawe Branch, for the period 2022 to 2024, covering 30 sub-districts across five regencies/cities: Lhokseumawe City, Bireuen Regency, North Aceh Regency, Central Aceh Regency, and Bener Meriah Regency. This raw data includes information on the time of the incident, location of the accident, victim's identity, victim's address, and the police station handling the case.

Table 3. Aggregate Sample of Weekly Accident Number Data for each District

Tahun	Bulan	Minggu	Kecamatan	Total Jumlah Kecelakaan Mingguan
2022	Januari	1	Dewantara	3
2022	Januari	1	Muara Batu	2
2022	Januari	1	Lhoksukon	2
2022	Januari	2	Dewantara	2
2022	Januari	2	Muara Batu	1
2022	Januari	2	Lhoksukon	4
...	...	...	...	...
2024	Desember	156	Dewantara	3
2024	Desember	156	Muara Batu	1
2024	Desember	156	Lhoksukon	1

For the needs of aggregate clustering, the data available in table 3.2 is reprocessed to obtain an overview of the number of weekly accidents for sub-district units, because the clustering to be built is a sub-district clustering with an overview of the number of accidents that occur. An overview of the input data for clustering is given in Figure 5. The clustering results obtained will be the basis for building a forecasting model; each cluster will have its own forecasting model.

Tabel 4. Clustering Input Data

Kecamatan	2022-00	2022-01	2022-02	2022-03	2022-04	2022-05	2022-06	2023-01	2023-02	2023-03	...	2024-43	2024-44	2024-45	2024-46	2024-47	2024-48	2024-49	2024-50	2024-51	2024-52
ATU LINTANG	0	0	1	0	0	0	1	0	0	0	...	1	0	0	1	0	0	1	0	0	0
BAKTIYA	0	0	1	0	0	0	1	0	0	0	...	0	2	1	4	0	4	1	2	3	0
BAKTIYA BARAT	0	0	2	0	0	1	0	0	0	0	...	0	0	1	0	0	0	1	1	0	0
BANDA BARO	0	0	2	0	0	0	1	0	0	0	...	0	1	1	1	0	0	1	1	0	0
BANDA SAKTI	0	1	0	0	0	0	1	0	1	0	...	5	0	1	0	0	0	1	1	0	0

B. Subdistrict Clustering Modeling Based on Weekly Accident Data

This study uses the K-Means algorithm to cluster sub-districts based on the number of weekly accidents. K-Means is an unsupervised learning method that works by determining the number of clusters (K), randomly selecting cluster centers, calculating the distance of the data to each center using Euclidean Distance, assigning data to the nearest cluster, and updating the cluster centers based on the average data until convergence. The feature used in clustering is the number of weekly accidents per sub-district, so that each sub-district has a number of weekly time series features that are used as the basis for clustering.

Table 5. Overview of data that will be used for sub-district clustering based on the number of weekly accident cases

Kecamatan	Minggu 0 Jan 2022	Saturday 1 Jan 2022	...	Minggu 51 Des 2024	Minggu 52 Des 2024	Total
Atu Lintang	0	0	...	0	0	5
Baktiya	0	0	...	2	0	103
Baktiya Barat	0	0	...	0	0	59
Banda Baro	0	0	...	0	0	37
Banda Sakti	0	1	...	0	5	114
...	...	...	... ...	...	...	
Bener Kelipah	0	0	...	0	0	3
Kute Panang	0	0	...	0	0	2
Paya Bakong	0	0	...	0	0	6
Pirak Timur	0	0	...	0	0	2
Lapang	0	0	...	0	0	3
Total	1	27	...	34	6	4589

C. Weekly Accident Forecasting Modeling

Forecasting modeling is performed for each cluster that has been formed. The feature used for forecasting

DOI : 10.30872/jim.v21i1.25392 e-ISSN 2597-4963 dan p-ISSN 1858-4853

is the average number of accidents from sub-districts within the same cluster per week, as shown in Table 6.

Table 6. Description of data on the average number of accidents from sub-districts in a Cluster

Time Step	**Rata-rata Jumlah Kecelakaan mingguan *cluster* 1**
Minggu 1 Januari 2022	3
Minggu 2 Januari 2022	2
...	...
Minggu 2 Desember 2024	3
Minggu 3 Desember 2024	1
Minggu 3 Desember 2024	1

In this study, the forecasting model uses a recurrent neural network (RNN) approach. RNNs were chosen because they have proven to be able to capture sequential features of the number of accidents occurring in a single week. For this forecast, the window size used will be tuned and evaluated to achieve optimal performance.

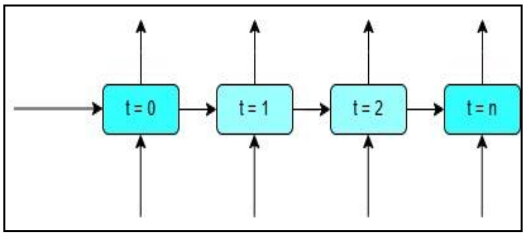

Figure 4. Recurrent Neural Network (RNN) Architecture

D. Evaluation of Clustering and Forecasting Models

Clustering evaluation was conducted using the Silhouette Score to assess the accuracy of cluster formation based on the distance between data within and between clusters. Meanwhile, forecasting evaluation used the Mean Absolute Error (MAE) to measure the average difference between actual and predicted values, and the R^2 Score to assess the model's ability to represent the actual data. A small MAE value and an R^2 value close to 1 indicate good forecasting model performance.

E. Implementation Tools and Environment

This research was implemented using the Python programming language with the support of the Pandas and NumPy libraries for data processing, Scikit-learn for K-Means Clustering, and Keras/TensorFlow for building and training the RNN model. The RNN model was built with two hidden layers (32 and 16 units) and a ReLU activation function at the output to predict the number of accidents. The training process used the Adam optimizer with a learning rate of 0.001 and a MAE loss function, as well as batch size, epoch, and window size parameters determined through hyperparameter tuning according to the characteristics of each cluster. All experiments were run on Google Colab to take advantage of CPU/GPU support and ease of data management.

4. RESULTS AND DISCUSSION

A. Data Preparation and Aggregation

The data used in this study is traffic accident data collected from PT. Jasa Raharja Lhokseumawe Branch, covering five administrative areas: Lhokseumawe City, Bireuen Regency, North Aceh Regency, Central Aceh Regency, and Bener Meriah Regency. The data covers the period from January 2022 to December 2024 and consists of individual accident case records, including the date of the incident, location, victim's name, and the police unit responsible.

Table 7. Overview of PT. Jasa Raharja Accident Data

Kecamatan	**Kabupaten**	**Nama Polres**	**No. Laporan Polisi**	**Tanggal Kejadian**	**Deskripsi Lokasi Kejadian**	**Nama Korban**	**Cidera**
KEC. DEWANTARA	KAB. ACEH UTARA	POLRES LHOKSEUMAWE	LRA/*********/MAWE/POLDA ACEH	30/08/2022	*********KEC. DEWANTARA, KABUPATEN ACEH UTARA, ACEH, INDONESIA	HAS*********ID	MD
KEC. GANDAPURA	KAB. BIREUEN	POLRES BIREUEN	LRA/*********/POLDA ACEH	08/11/2022	*********KAB. BIREUEN	BUS*********RI	LL
KEC. PEULIMBANG	KAB. BIREUEN	POLRES BIREUEN	LRA/*********/POLDA ACEH	07/11/2022	*********LIMBANG KEC. PEULIMBANG KAB. BIREUEN	MUS*********RA	LL
KEC. PEUSANGAN	KAB. BIREUEN	POLRES BIREUEN	LRA/*********/POLDA ACEH	25/12/2022	*********JAH KEC. PEUSANGAN KAB. BIREUEN	KHA*********IDA	LL
KEC. PEUSANGAN	KAB. BIREUEN	POLRES BIREUEN	LRA/*********/POLDA ACEH	25/12/2022	*********JAM KEC. PEUSANGAN KAB. BIREUEN	AFKA*********ILMI	LL
KEC. PEUDADA	KAB. BIREUEN	POLRES BIREUEN	LRA/*********/POLDA ACEH	24/06/2023	*********NASAH TUNONG KEC. PEUDADA KAB. BIREUEN	RAHM*********RA	LL
KEC. PEUDADA	KAB. BIREUEN	POLRES BIREUEN	LRA/*********/POLDA ACEH	24/06/2023	*********NASAH TUNONG KEC.	RAIS*********IL	LL

e-ISSN 2597-4963 dan p-ISSN 1858-4853 DOI : 10.30872/jim.v21i1.25392

					PEUDADA KAB. BIREUEN		
KEC. TANAH JAMBO AYE	KAB. ACEH UTARA	POLRES ACEH UTARA	LRA/*********/POLDA ACEH	24/06/2023	*********K KEC. TANAH JAMBO AYE KAB. ACEH UTARA	KHAI*********IR	MD
KEC. BANDAR	KAB. BENER MERIAH	POLRES BENER MERIAH	LP/*********/POLDA ACEH	04/01/2024	*********AKAT JEROH KEC BANDAR KAB BENER MERIAH	AMAL*********DA	LL
KEC. WIH PESAM	KAB. BENER MERIAH	POLRES BENER MERIAH	LP/*********/POLDA ACEH	30/04/2024	*********R UJUNG KEC. WIH PESAM KAB BENER MERIAH	INDR*********HA	LL

Table 7. shows an example of raw traffic accident data from PT. Jasa Raharja Lhokseumawe Branch converted from PDF to Excel without aggregation, where each row represents a single accident. The 4,593 entries from the 2022–2024 period served as the basis for analysis before preprocessing. The preprocessing stage included importing data from Excel, checking the data structure and completeness, and cleaning and standardizing sub-district names using regular expression (regex) techniques, so that the data was ready for use in clustering and forecasting.

```
df_2022["Kecamatan"] = df_2022["Kecamatan"].str.extract(r"KEC[.\s]+([\w\s]+)")
```

Figure 5. Regex process code

Figure 5. shows the process of cleaning sub-district data, where if the sub-district column contains the word “KEC”, then the sub-district name is taken from the text after the word, for example “KEC. GANDAPURA” becomes “GANDAPURA”. In addition, newline characters and excess spaces are removed to prevent them from interfering with the grouping and subsequent analysis process.

```
df_2022["Kecamatan"] = [re.sub(r"\n", "", i) for i in df_2022["Kecamatan"]]

df_2022["Kecamatan"] = [re.sub(r"\s+", " ", i).strip() for i in df_2022["Kecamatan"]]
```

Figure 6. Process code for cleaning excess characters and spaces

After each year's data was cleaned, all data was combined into a single data frame. Weekly aggregation was then performed by sub-district by converting the incident dates to week format. This allowed the number of accidents to be calculated from the total police reports in each sub-district for a given week, allowing for temporal analysis.

Table 8. Examples of Data Aggregation

No	Week Ke-	Year	Subdistrict
1	2022	Kota Juang	1
1	2022	Banda Sakti	1
1	2022	Dewantara	2
...	...	...	...
52	2022	Muara Batu	3
52	2022	Peudada	2
52	2022	Peusangan	5
...	...	...	...
1	2023	Bukit	2
1	2023	Gandapura	1
1	2023	Jeumpa	3
...	...	...	...
52	2023	Nibong	2
52	2023	Nisam	4
52	2023	Peulimbang	1
...	...	...	...
1	2024	Baktiya	1
1	2024	Banda Sakti	2
1	2024	Bandar	2
...	...	...	...
51	2024	Syamtalira Aron	2
51	2024	Tanah Luas	1
52	2024	Kota Juang	1

The aggregation results yielded 159 weeks of data used as clustering input. The data preparation phase produced two main outputs: accident data by date and sub-district for general analysis, and weekly aggregated data per sub-district, which served as the primary input for the K-Means clustering process and forecasting using RNN.

B. *Clustering*

The clustering process in this study uses aggregated data on the number of weekly accidents per sub-district for the 2022–2024 period, where each sub-district is represented as a time series. Clustering aims to group sub-districts based on similar accident patterns to obtain regional segmentation that can support the forecasting process and determine safety policies. The method used is K-Means Clustering, an unsupervised learning algorithm that is efficient and effective in grouping numerical data based on proximity to the cluster center.

1) Data Retrieval and Preparation

The data used is the results of weekly aggregated traffic accidents, which have been cleaned and stored in the Data_Laka_Complete.csv file. The data was accessed via Google Drive, read using pandas, then checked for empty values, which were then filled with zeros to maintain the quality and stability of the calculations. Next, the clustering process was carried out using the K-Means method to

DOI : 10.30872/jim.v21i1.25392 e-ISSN 2597-4963 dan p-ISSN 1858-4853

group sub-districts based on the pattern of weekly accidents.

Table 9. Overview of Data to be Clustered

Kecamatan	2022-00	2022-01	2022-02	2022-03	2022-04	2022-05	2022-06	2022-07	2022-08	...	2024-43	2024-44	2024-45	2024-46	2024-47	2024-48	2024-49	2024-50	2024-51	2024-52
ATU LINTANG	0	0	1	0	0	2	0	0	0	...	1	0	0	0	0	0	0	0	0	0
BAKTIYA	0	0	1	0	0	0	0	0	0	...	0	2	0	4	0	4	0	0	2	0
BAKTIYA BARAT	0	0	0	0	0	1	0	0	0	...	0	0	1	0	0	0	0	0	0	0
BANDA BARO	0	0	0	0	0	0	0	5	0	...	0	0	1	0	0	0	0	0	0	0
BANDA SAKTI	0	1	0	0	0	0	0	0	1	...	5	1	0	0	0	0	0	0	0	0

Before clustering, the dataset consisted of cleaned, weekly aggregated traffic accident data in CSV format. Non-numeric columns were removed so that only numerical data were processed, and missing values were replaced with zeros to ensure stable model training.

Table 10. Data Snippets

No	Week Ke-	Year	Subdistrict	Number of Accidents
1	1	2022	Kuta Makmur	2
2	2	2022	Muara Batu	3
3	3	2022	Dewantara	1
4	4	2022	Syamtalira Aron	4
5	5	2022	Banda Sakti	5
6	1	2023	Bukit	2
7	2	2023	Gandapura	1
8	3	2023	Jeumpa	3
9	1	2024	Baktiya	1
10	2	2024	Banda Sakti	2
11	3	2024	Bandar	2
12	1	2022	Kuta Makmur	2

Table 5. shows the structure of the weekly numeric data used as clustering input, where each row represents the number of accidents per week in each sub-district, and non-numeric columns are removed during the feature selection stage. The clustering results show that each cluster has a distinct accident pattern, demonstrating the ability of K-Means to identify variations between regions. These results are then used as the basis for compiling forecasting datasets for each cluster.

2) Determining the Optimal Number of Clusters

The optimal number of clusters was selected using the Silhouette Score by testing k values from 2 to 9. Each k value was evaluated based on the average Silhouette Score, and the highest value was selected as the best number of clusters. The evaluation results showed that k = 2 was the optimal number of clusters. Before clustering, the data was standardized using MinMaxScaler to avoid scale bias, and the evaluation results were visualized in graphs.

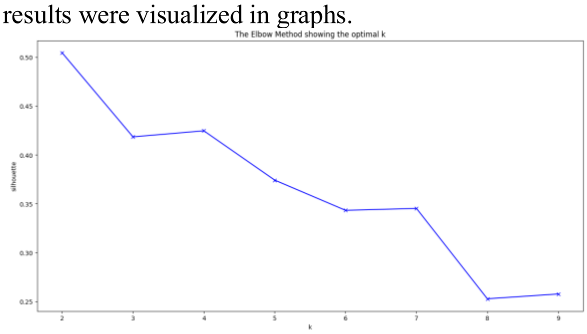

Figure 7. The Optimal Elbow Showing Method K

Based on the Silhouette Score evaluation using a K-value ranging from 2 to 10, the optimal number of clusters was set at 2, as it yielded the highest score (0.5). This result indicates adequate data separation and that two clusters effectively represent the data structure. The Elbow/Silhouette visualization demonstrates this optimal point, thus the clustering results are considered objective and suitable for use as a basis for forecasting.

3) K-Means Model Training

The K-Means model was applied with two clusters (n_clusters = 2) to ensure consistent results. The process involves initializing centroids, calculating Euclidean distances, assigning data to the nearest cluster, updating centroids, and iterating until convergence. After training, each data point is assigned a cluster label stored in a new column. The clustering features are weekly accident count values by region, which represent accident intensity and form the basis of the grouping process.

4) Interpretation of Clustering Results

The clustering results formed two sub-district groups: one with a low weekly accident rate and one with a high weekly accident rate. This grouping provides a snapshot of the different risk levels in each region and can be used as a basis for decision-making in traffic accident mitigation efforts.

e-ISSN 2597-4963 dan p-ISSN 1858-4853 DOI : 10.30872/jim.v21i1.25392

Table 11. K-Means Clustering Results

Cluster	Amount	Criteria
0	3	kecamatan dengan kecelakaan tinggi dan tidak stabil.
1	69	kecamatan dengan tren sedang dan cenderung fluktuatif.

The 2D PCA visualization shows a fairly clear separation of the two clusters, although there is slight overlap. The K-Means results with K = 2 divide the sub-districts into two groups: Cluster 0, which contains 3 sub-districts with a relatively high and fluctuating weekly accident pattern, and Cluster 1, which consists of 69 sub-districts with a lower and more stable accident pattern. This grouping demonstrates the ability of K-Means to identify differences in accident patterns between regions and is subsequently used as a basis for dataset formation and forecasting modeling using RNN.

5) Clustering Results Visualization

Clustering results are visualized using a Silhouette Score chart to assess the quality of cluster separation, as well as a colored scatterplot to show the distribution and boundaries between clusters. If spatial data is available, the visualization can also be presented as a regional map to facilitate geographic interpretation.

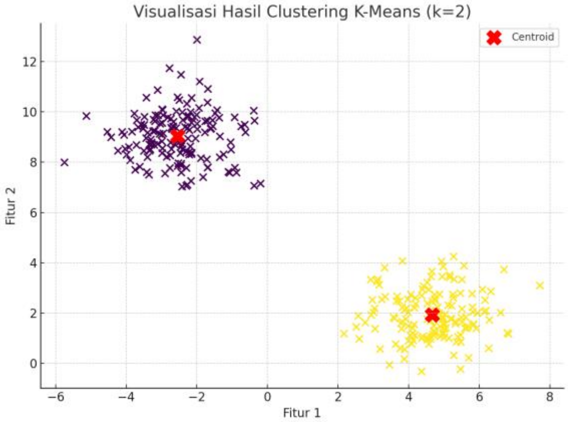

Figure 8. Visualization of Clustering Results

The 2D PCA visualization in Figure 4.5 shows that the K-Means results form two distinctly separate clusters. Cluster 0 has a wider and more fluctuating distribution, indicating a higher accident rate, while Cluster 1 is more homogeneous and stable. This result aligns with the Silhouette Score of 0.5, thus supporting the appropriate selection of the two clusters and the validity of the sub-district grouping structure.

C. *Clustering Process Conclusion*

This study successfully grouped sub-districts into two clusters based on weekly accident intensity using K-Means with a K value of 2 and a Silhouette Score of 0.5. These clustering results serve as an important basis for subsequent forecasting and provide strategic input for stakeholders in formulating policies and allocating resources for accident management more precisely, accurately, and data-driven.

D. *Forecasting*

The forecasting stage is performed after clustering is complete, using weekly accident data that has been grouped into clusters. Each cluster forms a time series dataset that represents accident patterns in areas with similar characteristics. Forecasting is performed using the Recurrent Neural Network (RNN) method to predict the number of accidents in each cluster, thus supporting more accurate data-driven prevention strategy planning for PT. Jasa Raharja Lhokseumawe Branch.

1) Forecasting Data Preparation

The forecasting stage uses the average weekly accident data for each cluster as a univariate time series. Previous weekly values are used to predict the next week based on a defined window size. The average weekly accident count for each cluster is calculated from the aggregated weekly data and stored in a new dataset, which becomes the input for the forecasting model.

Table 12. Descriptive Statistical Analysis

Statistik	Acc avg
Count	159
Mean	1,662
Std	1,336
Min	0
25%	0,666
50%	1,333
75%	2,333
Max	8,666

The weekly average data index was changed to week, and the average accident value column was renamed acc_avg, while the k_means column was removed as it was unnecessary for the time series analysis. Descriptive statistical analysis was performed to describe the data characteristics, which showed 159 weeks of observations with an average of 1.56 accidents per sub-district and significant

DOI : 10.30872/jim.v21i1.25392 e-ISSN 2597-4963 dan p-ISSN 1858-4853

variation across regions. This variation underscores the importance of the clustering process. The dataset was then divided into training, validation, and test data based on time sequence to maintain the time series nature and ensure objective model evaluation.

2) Dataset Division into Training, Validation, and Test Data

During the preparation phase, the time series dataset is divided into training, validation, and test data. This division allows the model to be trained, hyperparameter tuned, and performance objectively evaluated on previously unused data, thus helping to prevent overfitting and provide more accurate estimates of model performance.

The split_time_series function is used to split time series data into training, validation, and test data based on the ratio and time sequence, without randomization, for each cluster. The time range and number of data sets in each set are displayed for verification and visualized in a time series plot to ensure the data is split properly.

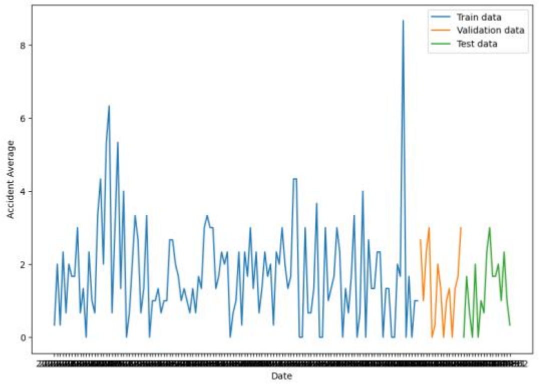

Figure 14. Visualization of training, validation, and test data in the form of a time series plot on Cluster 0

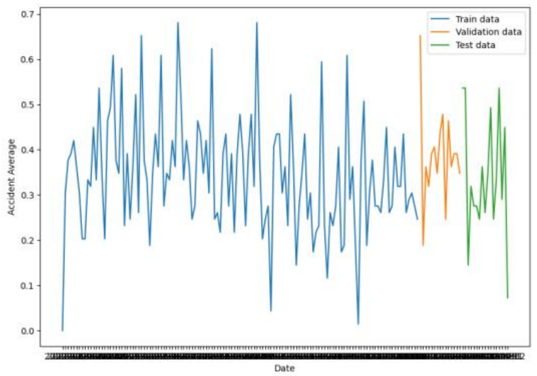

Figure 15. Visualization of training, validation, and test data in the form of a time series plot in Cluster 1

The visualization of the training, validation, and test data in Cluster 0 shows a fluctuating weekly crash time series with several peaks and dips, but the distribution pattern remains consistent across all segments, indicating a successful time-based data partitioning. In Cluster 1, a similar visualization shows a lower, more stable range of crash values with small amplitude fluctuations. These differences in characteristics between the clusters confirm that the time series partitioning is consistent and allows the model to learn different crash patterns more specifically and realistically.

Tabel 13. Sub-districts with the highest number of accidents

No	District	Number of Accidents
1	Peusangan	335
2	Kota Juang	290
3	Juli	220
4	Lhoksukon	217
5	Peudada	199
6	Jeumpa	190
7	Dewantara	187
8	Muara Dua	177
9	Muara Batu	168
10	Jeunib	134

3) Dataset Creation for RNN Model

The time series data is converted into a format suitable for RNN model training using the create_dataset function. This function creates pairs of input (feature) and output (target) data using the concept of time windows (steps). The input consists of a sequence of 'acc_avg' values over a 'step' of weeks, and the target is the 'acc_avg' value for the following week.

The input and output data were converted to NumPy arrays and reshaped to fit the requirements of the SimpleRNN and Dense layers. The RNN model was built with two SimpleRNN layers of 32 and 16 units, ending with a Dense layer for regression, and using ReLU activation on the output to ensure positive predictions.

4) SimpleRNN Model Construction and Training

The forecasting model is built using the SimpleRNN architecture consisting of two layers and one Dense layer. The first SimpleRNN layer uses return_sequences=True to generate sequential outputs, while the second layer produces a single output that is passed to the Dense layer as the final prediction with the ReLU activation function. The model is compiled using the Adam optimizer and the Mean Absolute Error (MAE) loss function, then trained using training data and validated with validation and test data to monitor performance and detect overfitting.

Data Set	**Start**	**End**
Train	Minggu 0 Tahun 2022	Minggu 20 Tahun 2024
Validasi	Minggu 21 Tahun 2024	Minggu 35 Tahuh 2024
Test	Minggu 36 Tahun 2024	Minggu 52 Tahun 2024

5) Forecasting Model Evaluation

After training, the model is evaluated using test data to assess its generalizability. Model performance is measured using MAE, MSE, and RMSE metrics to determine the level of prediction

e-ISSN 2597-4963 dan p-ISSN 1858-4853 DOI : 10.30872/jim.v21i1.25392

error between actual values and forecast results, and compared with the error values in the training and validation data.

No	Step	Hyperparameter	Train MAE	Val MAE	Test MAE	Test RMSE	Test MSE
1	3	Hidden RNN: 2 Unit RNN 1: 32	0.85	0.99	0.74	0.88	0.77
		Unit RNN 2: 16 Epoch: 20 Batch size: 16	01.01	0.86	0.75	0.89	0.79
		Learning rate: 0.001	01.01	0.84	0.77	0.89	0.79
2	5	Hidden RNN: 2 Unit RNN 1: 64	0.89	0.69	0.73	0.97	0.93
		Unit RNN 2: 32 Epoch: 30 Batch size: 25	0.89	0.72	0.73	0.96	0.94
		Learning rate: 0.001	0.89	0.73	0.81	01.07	1.15

6) Forecasting Forecasting Model Prediction Results

Model evaluation was performed using MAE, MSE, and RMSE. The results showed that Cluster 1 had better predictive performance than Cluster 0, with MAE values of 0.09, MSE 0.01, and RMSE 0.11, while Cluster 0 had MAE 0.62, MSE 0.58, and RMSE 0.77. The smaller error value in Cluster 1 indicates that the RNN model is more accurate and stable in relatively consistent accident patterns compared to fluctuating patterns.

7) Prediction Evaluation Results

Based on the evaluation using MAE, MSE, and RMSE, the model in Cluster 0 produced an MAE of 0.62, MSE of 0.58, and RMSE of 0.77, indicating a relatively higher level of prediction error and still quite large error variations. In contrast, the model in Cluster 1 showed better performance with an MAE of 0.09, MSE of 0.01, and RMSE of 0.11, indicating more accurate and consistent predictions. This difference in performance indicates that the more stable data characteristics in Cluster 1 are easier to model than Cluster 0, which has higher accident fluctuations.

8) Model Training and Testing Process Analysis

The Recurrent Neural Network (RNN) model training process in this study was carried out by dividing the data into training, validation, and test data to optimize performance and prevent overfitting. The Train MAE results show that the model in Cluster 0 has a value of 0.62, while Cluster 1 is lower at 0.09, indicating that Cluster 1's data patterns are easier to learn. In the Validation MAE, Cluster 0 ranges from 0.62–0.99 and Cluster 1 0.06–0.09, indicating that the Cluster 1 model is more consistent when tested on the validation data. Meanwhile, in the Test MAE, Cluster 0 reaches 0.62 (RMSE 0.77; MSE 0.58) and Cluster 1 only 0.09 (RMSE 0.11; MSE 0.01), confirming that the Cluster 1 model is able to predict with high accuracy and stability, while Cluster 0 has difficulty learning more fluctuating accident patterns.

Table 14. Test Results

Minggu	Rata-rata Kasus Kecelakaan
2024-36	0.536232
2024-37	0.536232
2024-38	0.144928
2024-39	0.318841
2024-40	0.275362
2024-41	0.275362
2024-42	0.246377
2024-43	0.362319
2024-44	0.260870
2024-45	0.347826
2024-46	0.492754
2024-47	0.246377
2024-48	0.333333
2024-49	0.536232
2024-50	0.289855
2024-51	0.449275

5. CONCLUSION

Based on a study of traffic accident forecasting in the work area of PT. Jasa Raharja Lhokseumawe Branch for the period 2022–2024, data from 72 sub-districts in 5 regencies/cities were processed into weekly accident counts for clustering and forecasting purposes. K-Means Clustering produced two clusters, namely Cluster 0 with sub-districts that have a high and fluctuating accident pattern, and Cluster 1 with a lower and stable pattern, with clear cluster separation in the PCA visualization. Forecasting using RNN showed that Cluster 0 had an MAE of 0.62, MSE of 0.58, and RMSE of 0.77, while Cluster 1 was more accurate with an MAE of 0.09, MSE of 0.01, and RMSE of 0.11, confirming that the characteristics of the data patterns of each cluster significantly affect the model performance.

6. DAFTAR PUSTAKA

Ananta, M. A. F., & Sofro, A. (2025). *Comparison of K-Means and K-Medoids Algorithms in Clustering Indonesian Provinces Using Stunting Handling Index*. *9*.

Apriyanti, N. P. R., Putra, I. K. G. D., & Putra, I. M. S. (2020a). Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 72. https://doi.org/10.24843/JIM.2020.v08.i02.p01

Apriyanti, N. P. R., Putra, I. K. G. D., & Putra, I. M. S. (2020b). Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression. *Jurnal Ilmiah Merpati*

DOI : 10.30872/jim.v21i1.25392 e-ISSN 2597-4963 dan p-ISSN 1858-4853

(Menara Penelitian Akademika Teknologi Informasi), 72. https://doi.org/10.24843/JIM.2020.v08.i02.p01

Chusyairi, A., & Ramadar Noor Saputra, P. (2019). Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering. *Telematika*, *12*(2), 139–148. https://doi.org/10.35671/telematika.v12i2.848

Enggarsasi, U., & Sa'diyah, N. K. (2017). Kajian Terhadap Faktor-Faktor Penyebab Kecelakaan Lalu Lintas Dalam Upaya Perbaikan Pencegahan Kecelakaan Lalu Lintas. *Perspektif*, *22*(3), 228. https://doi.org/10.30742/perspektif.v22i3.632

Huang, Y., Cheng, Z., Zhou, Q., Xiang, Y., & Zhao, R. (2020). Data Mining Algorithm for Cloud Network Information Based on Artificial Intelligence Decision Mechanism. *IEEE Access*, *8*, 53394–53407. https://doi.org/10.1109/ACCESS.2020.2981632

Lusiana, A., & Yuliarty, P. (2020). Penerapan Metode Peramalan (Forecasting) Pada Permintaan Atap di PT X. *Industri Inovatif : Jurnal Teknik Industri*, *10*(1), 11–20. https://doi.org/10.36040/industri.v10i1.2530

Matdoan, M. Y., Purnamasari, N. A., & Laamena, N. S. (2023). Application of the K-Means Algorithm for Clustering Production of Capture Fisheries in Maluku Province. *Pattimura International Journal of Mathematics (PIJMath)*, *2*(2), 63–70. https://doi.org/10.30598/pijmathvol2iss2pp63-70

Ramadhan, A. Z. H., Rahayudi, B., & Ratnawati, D. E. (n.d.). *Prediksi Polusi Udara Di Dki Jakarta Dengan Menggunakan Metode Long-Short Term Memory (LSTM)*.

Simon, G. (2025). Peramalan pendaftar mahasiswa baru dengan menggunakan metode moving average, weighted moving average dan exponential smoothing. *Jurnal Teknik Industri Terintegrasi*, *8*(1), 13–21. https://doi.org/10.31004/jutin.v8i1.36423