VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v9i2.5046

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

IMPROVING TRAFFIC DENSITY PREDICTION USING LSTM WITH PARAMETRIC ReLU (PReLU) ACTIVATION

Nur Alamsyah[1*]; Titan Parama Yoga[2]; Budiman[3]

Information Systems[1,2,3]
Universitas Informatika Dan Bisnis Indonesia[1,2,3]
https://unibi.ac.id/[1,2,3]
nuralamsyah@unibi.ac.id[1*], titanparamayoga@gmail.com[2], budiman@unibi.ac.id[3]

(*) Corresponding Author

Abstract—*In the presence of complex traffic flow patterns, this research responds to the challenge by proposing the application of the Long Short-Term Memory (LSTM) model and comparing four different activation functions, namely tanh, ReLU, sigmoid, and PReLU. This research aims to improve the accuracy of traffic flow prediction through LSTM model by finding the best activation function among tanh, relu, sigmoid, and PReLU. The method used starts from the collection of traffic flow datasets covering the period 2015-2017 used to train and evaluate the LSTM model with the four activation functions. Tests were conducted by observing the Train Mean Squared Error (MSE) and Validation MSE. The experimental results show that PReLU provides the best results with a Train MSE of 0.000505 and Validation MSE of 0.000755. Although tanh, ReLU, and sigmoid provided competitive results, PReLU stood out as the optimal choice to improve the adaptability of the model to complex traffic flow patterns.*

Keywords: *activation function, LSTM, prediction, PReLU, traffic flow.*

Intisari—*Dalam menghadapi kompleksitas pola arus lalu lintas, penelitian ini menanggapi tantangan tersebut dengan mengusulkan penerapan model Long Short-Term Memory (LSTM) dan membandingkan empat fungsi aktivasi berbeda, yaitu tanh, ReLU, sigmoid, dan PReLU. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi arus lalu lintas melalui model LSTM dengan mencari fungsi aktivasi terbaik di antara tanh, relu, sigmoid, dan PReLU. Metode yang digunakan dalam penelitian ini dimulai dari pengumpulan dataset arus lalu lintas yang mencakup periode 2015-2017 digunakan untuk melatih dan mengevaluasi model LSTM dengan keempat fungsi aktivasi. Pengujian dilakukan dengan mengamati Train Mean Squared Error (MSE) dan Validation MSE. Hasil eksperimen menunjukkan bahwa PReLU memberikan hasil terbaik dengan Train MSE sebesar 0.000505 dan Validation MSE 0.000755. Meskipun tanh, ReLU, dan sigmoid memberikan hasil yang kompetitif, PReLU menonjol sebagai pilihan optimal untuk meningkatkan adaptabilitas model terhadap pola arus lalu lintas yang kompleks.*

Kata Kunci: *fungsi aktivasi, LSTM, prediksi, PReLU, arus lalu lintas.*

INTRODUCTION

The rapid growth in the number of motor vehicles that occurs over time has significant consequences for urban transportation systems [1]. The increasing number of vehicles presents new challenges in traffic management, resulting in congestion, increased travel time, and even safety risks. In this context, the need for traffic density prediction has become increasingly urgent [2]. Accurate predictions allow authorities to take timely preventive and responsive actions, optimize traffic flow, and minimize its negative impacts [3]. Therefore, this research emerged as a response to the real-world problems associated with the rapid growth of motorized vehicles [4]. The importance of traffic flow prediction is closely related to the efficient use of transportation infrastructure and the improvement of urban quality of life. By investigating the potential use of the Long Short-Term Memory (LSTM) method with Parametric Rectified Linear Unit (PReLU). activation function, this research seeks to make a significant contribution in improving the accuracy and reliability of traffic density prediction amidst the increasing vehicle growth [5].

This research is conducted in the context of the increasing need for traffic density prediction in urban transportation management. The success of an efficient prediction system has a direct impact on resource allocation, route planning, and traffic safety improvement. In response to the complexity

Accredited Rank 2 (Sinta 2) based on the Decree of the Directorate General of Higher Education, Research and Technology. Ministry of Education, Culture, Research and Technology of the Republic of Indonesia No.225/E/KPT/2022, 07 December 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i2.5046

of urban traffic patterns, traditional prediction methods are often limited. Therefore, this research explores the potential use of Long Short-Term Memory (LSTM) as one approach capable of capturing complex time patterns [6]. Special emphasis is given to the application of the PReLU activation function, which is expected to improve the model's ability to understand the data structure [7]. This research involves an in-depth understanding of the literature review on the use of LSTMs in traffic density prediction, and draws inspiration from the best approaches that have been exemplified by previous studies [8]. By combining current literature and past research, this study aims to make a significant contribution to optimizing traffic density prediction systems in urban environments.

The literature review investigates recent developments in traffic density prediction, focusing on the application of Long Short-Term Memory (LSTM) and PReLU activation [9]. Several previous studies have revealed the potential of LSTM in improving the accuracy of traffic density prediction, especially in coping with complex dynamic patterns in urban environments [10]. This talk is based on a number of prominent journals that have investigated the application of LSTM in the context of urban traffic prediction [11]. These studies highlight the advantages of LSTMs in capturing and modeling complex time patterns, which are the hallmark of urban traffic density. Detailing the methods that have been applied in these studies, this research involves the application of LSTM with ReLU activation function as an innovative approach in forecasting traffic density. By building on previous findings [12], this research aims to make a significant contribution towards the development of more advanced traffic prediction methods.

This study aims to implement the LSTM model with ReLU activation in forecasting urban traffic density. Using a dataset that includes both time and vehicle count information, the goal of this research is to improve the accuracy of traffic density prediction to support more efficient traffic management. Through this approach, it is expected that this research can make a significant contribution to the understanding and prediction of traffic density, supporting sustainable urban transport development systems.

MATERIALS AND METHODS

Our research involves a series of systematic stages, which include dataset collection, data preprocessing, Exploratory Data Analysis (EDA), Train-Test Data Splitting, LSTM Model Building, Prediction with Tuned LSTM Model, Evaluation of Prediction Results. The metrics we use here are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) The methods we use are shown in Figure 1.

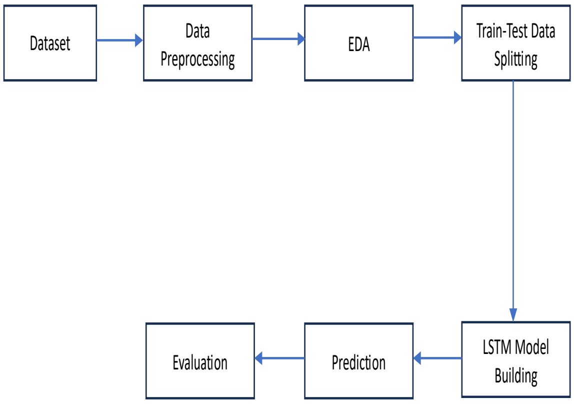

Figure 1. Proposed Method

A. Dataset

We obtained the traffic flow dataset from Kaggle with a time span from 2015 to 2017 totaling 48,120 data [13]. The features contained in this dataset include DateTime, Juction, Vehicles and ID. An explanation of these features is shown in Table 1.

Table 1. Dataset features

Feature	Description
DateTime	Menunjukkan waktu pengukuran arus lalu lintas (format tanggal dan jam).
Junction	Identifikasi persimpangan lalu lintas yang diamati.
Vehicles	Jumlah kendaraan yang terdeteksi pada waktu tertentu.
ID	Identifikasi unik untuk setiap entri data dalam dataset.

B. Data Preprocessing

In the data preprocessing stage, we perform DateTime conversion and data normalization. DateTime columns in the dataset are converted to datetime type to facilitate time manipulation. This is so that we can perform more effective time analysis in the context of traffic flow prediction. Next, data normalization is performed using MinMaxScaler to convert variable values into a range between 0 and 1. This normalization process is done so that the LSTM model can learn better from data that has a uniform scale, improving the performance of the model in making predictions. We present MinMaxscaler in formula (1).

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v9i2.5046

Where X is the original value of the dataset, X_{Min} is the minimum value of the dataset, X_{Max} is the maximum value of the dataset, X_{Scaled} is the normalized value.

C. Exploratory Data Analysis (EDA)

An Exploratory Data Analysis (EDA) was conducted to gain a deeper understanding of the traffic flow dataset used [14]. The two main aspects analyzed involved the intersections with the most number of vehicles and the analysis of the densest traffic flow times. The first is the analysis of Intersections with the Most Vehicles. We analyzed this in order to identify the intersections or locations that have the highest traffic levels. Knowing the intersections with the highest volume of vehicles can help in determining the focus of traffic prediction and planning. By analyzing the Intersections with the Most Vehicles we are able to understand the traffic patterns at a particular intersection which can provide insights into the congestion and mobility levels in the area. This is key for the development of accurate and relevant prediction models. We visualize this analysis in Figure 2. The result is as shown in Figure 2 that only 1 intersection is found to have the highest number of vehicles, meaning that the more intersections there are, the fewer vehicles there will be.

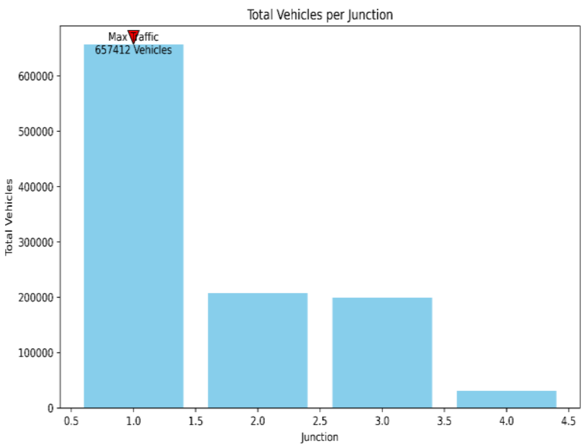

Figure 2. Total Vehicles per Junction

The next one is the analysis of the Time of Heaviest Traffic Flow. The purpose of this analysis is to identify the time pattern where traffic flow reaches its highest density. This analysis helps determine critical time periods that need more attention in traffic prediction and management. This analysis is conducted to understand the daily or seasonal variability in traffic, as it can affect the performance of prediction models. Capturing these time patterns enables optimization of resources and actions required at specific times. The results of this analysis are presented in Figure 3.

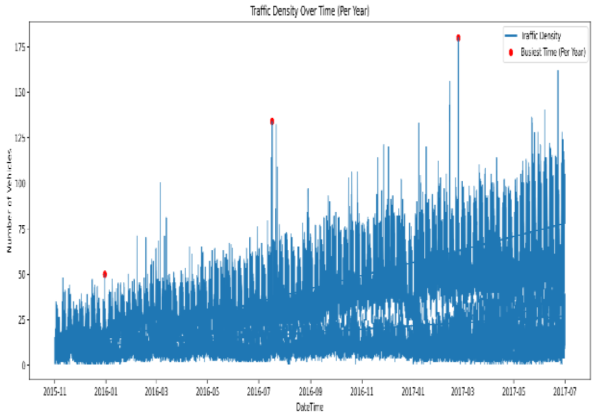

Figure 2. Traffic density over time (per year)

D. Train / Test Data Splitting

The traffic data obtained from 2015 to 2017 amounted to 48,120 entries. For the purposes of training and testing the LSTM model, the data was divided as follows: In 2015, 80% of 4392 entries were used as training data, while the remaining 20% became testing data. In 2016, 80% of the 26,352 entries became training data, and the remaining 20% became testing data. Finally, in 2017, 80% of the 17,376 entries became training data, while the remaining 20% became testing data. This division aims to train the model with historical data and test its performance on data that has never been seen before. Information on the split amount of training and testing data is presented in Table 2.

Table 2. Split Training and testing data

Year	Total Data	Data Training (80%)	Data Testing (20%)
2015	4392	3514	878
2016	26352	21082	5269
2017	17376	13901	3475

E. LSTM Model Building

First of all, in the LSTM model building stage, we chose four different activation functions: Hyperbolic Tangent (Tanh), Rectified Linear Unit (ReLU), Sigmoid, and Parametric Rectified Linear Unit (PReLU). We used the TensorFlow and Keras libraries to build this model. This LSTM model has a basic structure consisting of one LSTM layer with 50 units and one Dense layer with one output unit, corresponding to the traffic flow prediction task. In the model initiation process, we create a loop for each activation function to be tested. If the activation function to be tested is PReLU, we use the PReLU module of Keras to activate it. We present the tanh activation function in formula (2), ReLU in

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i2.5046

formula (3), Sigmoid in formula (4) and PReLU in formula (5).

$$\text{Tanh}(x) = \frac{e^{2x}-1}{e^{2x}+1} \quad (2)$$

The tanh function produces an output in the range (-1, 1), and is similar to sigmoid but has a larger range of values.

$$\text{ReLU}(x) = \max(0, x) \quad (3)$$

The ReLU function returns the input value if positive, and zero if negative. This helps in solving the vanishing gradient problem and improves the convergence speed.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (4)$$

The sigmoid function is used to produce an output in the range (0, 1). It is commonly used in the output layer of models for binary classification tasks.

$$\text{PReLU}(x) = \max(\alpha x, x) \quad (5)$$

PReLU is a variation of ReLU where the gradient for negative values can also be learned during training. Parameters α is a parameter that can be changed.

Next, we determine the training configuration of the model. We used the 'adam' optimizer which is a popular optimization algorithm in deep learning. This optimizer helps to update the model weights based on the loss function gradient, thus accelerating convergence and improving model performance. adam optimizer is presented in formula (6). We also use the loss function 'mse' or Mean Squared Error, which is measured between the prediction and the true value, and is the main evaluation parameter for training regression models such as LSTM. The training process was conducted for 50 epochs, where one epoch is one iteration through the entire training data. A batch size of 32 was used to speed up the training by processing the data in small groups.

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t}+\epsilon} \quad (6)$$

Adam's optimizer uses the gradient of the loss function to update the weights θ model. η is the learning rate, m_t is the moving average gradient, v_t is the moving gradient of the mean square, and ϵ is a small value added to avoid division by zero.

During training, the learned model adjusts its weights to minimize the MSE value, thereby improving the prediction accuracy on the training data and validation data. The training results of each model are recorded and a comparison table is generated that includes the MSE on the training data and validation data. We present the MSE in formula (7). Thus, we can analyze and select the activation function that provides the best performance for traffic flow prediction.

$$\text{MSE}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \quad (7)$$

MSE measures the mean square of the difference between the true value of the y and predictions $\hat{y}$. This is commonly used for regression problems.

F. Prediction

After the training process, the resulting LSTM model has the ability to understand the temporal patterns contained in the traffic data. In the prediction stage, validation data is presented to the model, allowing the algorithm to recognize and respond to the patterns. During this process, the model utilizes the knowledge gained during training to provide traffic flow predictions. The initial prediction results generated by the model are on a normalized scale, which is a range of values between 0 and 1. To obtain interpretable values, the prediction results are returned to the original scale using the inverse normalization process. This allows us to compare and evaluate the model predictions with actual data in a more familiar and relevant context.

G. Evaluation

After predicting the traffic flow with the LSTM model, the next step is to evaluate the performance of the model to ensure its accuracy. The evaluation is done by utilizing several performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). MSE gives an idea of how close the prediction is to the true value, while RMSE gives the average error value in the same units as the original data. Furthermore, MAE gives an indication of how far our prediction is from the true value overall. Through these three metrics, we can get a comprehensive picture of the model's ability to generalize from training data to validation data.

It is important to note that this evaluation aims to assess the extent to which the model is reliable in providing accurate and useful predictions. In addition, the results of this evaluation can help us to understand whether the model has overfitting or underfitting tendencies, as well as whether further adjustments to the model architecture or parameters are required. This evaluation is a critical step in ensuring that the built model can be relied upon to provide accurate forecasts in real scenarios.

RESULTS AND DISCUSSION

In the construction of the LSTM model, we consider four main activation functions: Tanh, ReLU, Sigmoid, and PReLU. The use of Tanh activation function was chosen because it has non-linear properties and is able to effectively handle gradient changes, which is beneficial in modeling complex data patterns [15]. ReLU, or Rectified Linear Unit, was chosen for its effectiveness in capturing non-linear features and preventing the problem of vanishing gradients [16]. Furthermore, a Sigmoid activation function is used to produce an output that is limited between 0 and 1 [17], very useful in the case of probability prediction. In addition, we consider the use of PReLU (Parametric ReLU) activation function which allows parameter learning to adjust the convergence speed and overcome the problem of "dying ReLU." PReLU can adaptively set zero gradients for some neurons, optimizing the capacity of the model. The selection of these four activation functions is based on their respective advantages and characteristics, according to the needs of modeling complex and dynamic traffic flow data.

In the construction of the LSTM model, we chose the 'adam' optimizer and the 'mse' loss function to handle the model training. The optimizer 'adam' (Adaptive Moment Estimation) was chosen for its ability to adaptively adjust the learning rate based on momentum and adaptive learning rate [18]. Adam combines the advantages of the stochastic gradient descent (SGD) optimizer with momentum correction and learning rate adaptation techniques, making it an efficient and fast convergence option [19]. Furthermore, the loss function 'mse' or Mean Squared Error was chosen as the model evaluation criterion. MSE is a commonly used metric for regression problems, and in the context of traffic flow prediction, it measures how well our model captures the error between predicted and actual values. The goal of using MSE is to minimize the difference between the prediction and the target as much as possible, making it a suitable choice for traffic flow prediction problems, where precision needs to be carefully considered [20].

In Table 3 we present the results of training the model with four different activation functions, Train MSE and Validation MSE.

Table 3 Results of training the model

Activation Function	Train MSE	Validation MSE
Tanh	0.000538	0.000792
ReLU	0.000510	0.000766
Sigmoid	0.000768	0.001098
PReLU	0.000505	0.000755

From the table 3, the model with the activation function 'PReLU' shows better performance with a Train MSE of 0.000505 and Validation MSE of 0.000755 which is lower than the other activation functions. This indicates that the model with PReLU is able to better adapt to the training data and has a better ability to make predictions on validation data not seen during training. Therefore, for this case, the use of 'PReLU' activation function can be considered more optimal and recommended to improve the performance of the LSTM model in predicting traffic flow.

After creating a model by comparing 4 activation functions, we predicted using the PReLU activation function because it has the smallest Train MSE and Validation MSE values among other activation functions, and the results are presented in Figure 3.

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i2.5046

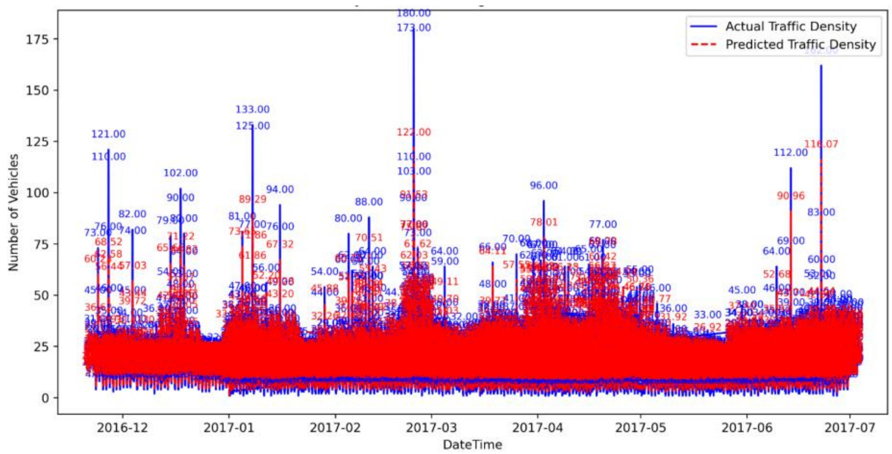

Figure 3. Traffic Density Prediction using LSTM with PReLU Activation

The prediction results are presented in Figure 3 show that the blue graph is the actual traffic density and the red graph is the predicted traffic density. The red graph almost partially covers the blue graph so it can be considered that the creation of the LSTM model to predict traffic flow using the PReLU activation function is successful. Research conducted by Yuanzheng *et al*, [21]. In particular, the use of PReLU instead of ReLU reduces MAPE by 0.79% and 1.58%, proving that replacing ReLU with PReLU can improve model performance. This research also shows good performance using PReLU.

CONCLUSION

Our research aimed to enhance traffic density prediction using LSTM models with various activation functions. We investigated the impact of tanh, ReLU, sigmoid, and PReLU activation functions on the model's performance. The primary objective was to identify the most effective activation function for accurate traffic flow prediction. Our study revealed that among the evaluated activation functions, PReLU yielded the best results with a Train MSE of 0.000505 and Validation MSE of 0.000755. This finding underscores PReLU's superior ability to capture intricate traffic patterns, providing a more robust solution for traffic flow prediction compared to other activation functions.
Furthermore, the selection of the 'adam' optimizer and 'mse' loss function played a crucial role in optimizing the LSTM model. The 'adam' optimizer dynamically adjusted the learning rate, enhancing the model's convergence, while the 'mse' loss function quantified the model's ability to replicate target values accurately. The combined use of 'adam' and 'mse' contributed to the efficiency and precision of our LSTM model in predicting traffic flow patterns. In addressing the research question on the optimal activation function and considering the title 'Improving Traffic Density Prediction Using LSTM With ReLU Activation,' our findings not only highlight the superior performance of PReLU but also emphasize the pivotal role of the optimizer and loss function configuration in achieving superior LSTM model performance for traffic density prediction. These insights contribute to advancing the field of traffic flow prediction and offer practical guidance for researchers and practitioners alike.

REFERENCE

[1] T. Afrin and N. Yodo, "A Survey Of Road Traffic Congestion Measures Towards A Sustainable and Resilient Transportation System," *Sustainability*, vol. 12, no. 11, p. 4660, 2020, doi: https://doi.org/10.3390/su12114660.

[2] F. Outay, H. A. Mengash, and M. Adnan, "Applications Of Unmanned Aerial Vehicle (UAV) In Road Safety, Traffic and Highway Infrastructure Management: Recent Advances and Challenges," *Transp. Res. Part Policy Pract.*, vol. 141, pp. 116–129, 2020, doi: https://doi.org/10.1016/j.tra.2020.09.018.

[3] Y. Berhanu, E. Alemayehu, D. Schröder, and others, "Examining Car Accident Prediction Techniques and Road Traffic Congestion: A Comparative Analysis of Road Safety and Prevention of World Challenges In Low-Income and High-Income Countries," *J. Adv. Transp.*, vol. 2023, 2023 doi: https://doi.org/10.1155/2023/6643412.

[4] S. M. Abdullah *et al.*, "Optimizing Traffic Flow in Smart Cities: Soft GRU-Based Recurrent Neural Networks for Enhanced Congestion Prediction Using Deep Learning," *Sustainability*, vol. 15, no. 7, p. 5949, 2023, doi: https://doi.org/10.3390/su15075949.

[5] S. F. Pane, J. Ramdan, A. G. Putrada, M. N. Fauzan, R. M. Awangga, and N. Alamsyah, "A Hybrid CNN-LSTM Model With Word-Emoji Embedding For Improving The Twitter Sentiment Analysis on Indonesia's PPKM Policy," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2022, pp. 51–56. doi: 10.1109/ICITISEE57756.2022.10057720.

[6] A. G. Putrada, N. Alamsyah, S. F. Pane, and M. Nurkamal Fauzan, "Feature Importance on Text Analysis for a Novel Indonesian Movie Recommender System," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, 2023, pp. 34–39. doi: 10.1109/ICoICT58202.2023.10262504.

[7] Y. Yu, K. Adu, N. Tashi, P. Anokye, X. Wang, and M. A. Ayidzoe, "RMAF: Relu-Memristor-Like Activation Function For Deep Learning," *IEEE Access*, vol. 8, pp. 72727–72741, 2020, doi: 10.1109/ACCESS.2020.2987829.

[8] D. Macêdo, C. Zanchettin, A. L. Oliveira, and T. Ludermir, "Enhancing Batch Normalized Convolutional Networks Using Displaced Rectifier Linear Units: A Systematic Comparative Study," *Expert Syst. Appl.*, vol. 124, pp. 271–281, 2019, doi: https://doi.org/10.1016/j.eswa.2019.01.066.

[9] S. Poornima and M. Pushpalatha, "Prediction Of Rainfall Using Intensified LSTM Based Recurrent Neural Network With Weighted Linear Units," *Atmosphere*, vol. 10, no. 11, p. 668, 2019, doi: https://doi.org/10.3390/atmos10110668.

[10] F. Wirthmüller, M. Klimke, J. Schlechtriemen, J. Hipp, and M. Reichert, "Predicting The Time Until A Vehicle Changes The Lane Using LSTM-Based Recurrent Neural Networks," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2357–2364, 2021, doi: 10.1109/LRA.2021.3058930.

[11] W. Wei, H. Wu, and H. Ma, "An Autoencoder and LSTM-Based Traffic Flow Prediction Method," *Sensors*, vol. 19, no. 13, p. 2946, 2019, doi: https://doi.org/10.3390/s19132946.

[12] Z. Wang, X. Su, and Z. Ding, "Long-Term Traffic Prediction Based on LSTM Encoder-Decoder Architecture," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6561–6571, 2020 doi: 10.1109/TITS.2020.2995546.

[13] FEDESORIANO, "Traffic Prediction Dataset." Accessed: Dec. 15, 2023. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset

[14] D. Hungness and R. Bridgelall, "Exploratory Spatial Data Analysis of Traffic Forecasting: A Case Study," *Sustainability*, vol. 14, no. 2, p. 964, 2022 doi: https://doi.org/10.3390/su14020964.

[15] D.-H. Lee, Y.-T. Kim, and S.-R. Lee, "Shallow Landslide Susceptibility Models Based on Artificial Neural Networks Considering The Factor Selection Method and Various Non-Linear Activation Functions," *Remote Sens.*, vol. 12, no. 7, p. 1194, 2020, https://doi.org/10.3390/rs12071194.

[16] A. Ciuparu, A. Nagy-Dăbâcan, and R. C. Mureşan, "Soft++, A Multi-Parametric Non-Saturating Non-Linearity That Improves Convergence In Deep Neural Architectures," *Neurocomputing*, vol. 384, pp. 376–388, 2020, https://doi.org/10.1016/j.neucom.2019.12.014.

[17] Y. Koçak and G. Ü. Şiray, "New Activation Functions for Single Layer Feedforward Neural Network," *Expert Syst. Appl.*, vol. 164, p. 113977, 2021, doi: https://doi.org/10.1016/j.eswa.2020.113977.

[18] A. G. Putrada, N. Alamsyah, S. F. Pane, M. N. Fauzan, and D. Perdana, "Knowledge Distillation for a Lightweight Deep Learning-Based Indoor Positioning System on Edge Environments," in *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2023, pp. 370–375. doi: 10.1109/ISITIA59021.2023.10220996.

[19] W. Yuan, F. Hu, and L. Lu, "A New Non-Adaptive Optimization Method: Stochastic Gradient Descent With Momentum And Difference," *Appl. Intell.*, pp. 1–15, 2022, doi: https://doi.org/10.1007/s10489-021-02224-6.

[20] N. Alamsyah, Saparudin, and A. P. Kurniati, "A Novel Airfare Dataset To Predict Travel Agent Profits Based On Dynamic Pricing," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, 2023, pp. 575–581. doi: 10.1109/ICoICT58202.2023.10262694.

[21] Y. Li *et al.*, "Dense Skip Attention Based Deep Learning For Day-Ahead Electricity Price Forecasting," *IEEE Trans. Power Syst.*, 2022, doi: 10.1109/TPWRS.2022.3217579.