Jurnal Mandiri IT
Vol. 14 No. 1, July (2025), pp. 1-11

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Comparative performance of LSTM and DNN in sentiment analysis

Jandri Tampubolon[1], Kusrini[2]
[1,2]Digital Transformation Intelligence, Informatics, Universitas Amikom Yogyakarta, Indonesia

A R T I C L E I N F O

Article history:

Received Apr 3, 2025
Revised Apr 14, 2025
Accepted Apr 30, 2025

Keywords:

Deep Neural Network;
Long Short-Term Memory;
Online Transportation;
Sentiment Analysis;
Twitter.

ABSTRACT

Understanding public sentiment toward online transportation services through social media analysis has gained increasing importance. This study provides a comparison between the effectiveness of Deep Neural Network (DNN), and Long Short-Term Memory (LSTM) models in analyzing user sentiment toward online transportation services in Indonesia using Twitter data. The dataset consists of 10,000 tweets related to Gojek, Grab, Maxim, and InDrive, collected from January to December 2023. Data preprocessing includes noise removal, case folding, tokenization, and stemming. Sentiment labeling was conducted using IndoBERTweet and manually validated. Using K-Fold Cross-Validation, both DNN and LSTM models were trained, and assessed using performance metrics such as accuration, precision, recall, and F1-score, training time, and Mean Absolute Error (MAE). The LSTM model demonstrated superior performances with accuration of 82,15%, precision of 82,21%, recall of 82,15%, specificity of 90,74%, F1-score of 82,10%, and MAE of 23,15%, compared to the DNN model which achieved an accuracy of 81,22%, precision of 81,20%, recall of 81,22%, specificity of 90,18%, F1-score of 81,12%, and MAE of 24,46%. However, DNN outperformed LSTM in training time efficiency (50,435 seconds vs. 148,765 seconds). LSTM shows significant advantages in understanding context and word relationships in sentiment analysis, while DNN offers better computational efficiency. The findings of this study can be utilized by online transportation services providers to improve service quality based on user feedback from social media.

Corresponding Author:

Jandri Tampubolon,
Digital Transformation Intelligence,
Informatics,
Universitas Amikom Yogyakarta,
Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta, 55281, Indonesia
Email: jandri.tampubolon@gmail.com

1. INTRODUCTION

Technological innovations today's digital landscape have led to substantial developments across multiple fields, notably in transportation. Online transportation services have emerged as innovative solutions that simplify mobility for the public, using only applications on mobile devices. In Indonesia, services such as Gojek, Grab, Maxim, and InDrive are becoming increasingly popular, with millions of users actively sharing their experiences, reviews, and complaints on social media.

One of the most widely used social networking sites is Twitter, where users can openly share opinions, experiences, and information with the public. This platform is chosen as a data source due to several advantages it offers. First, Twitter generates a massive volume of data daily, providing a rich and extensive source of information. Second, its real-time nature allows tweets to reflect public opinion directly and instantly on a particular issue. Moreover, the variety of topics

discussed on Twitter is also very diverse, including topics related to online transportation services (Pan et al., 2019), (Bijarnia et al., 2019), (Mamani-Coaquira & Villanueva, 2024).

The representativeness of Twitter data in reflecting public opinion toward online transportation services lies in its open-access nature and widespread usage across various user segments. Although Twitter users may not represent the entire demographic spectrum of the Indonesian population, they include a large and active portion of tech-savvy urban consumers, who are the primary users of services like Gojek, Grab, Maxim, and InDrive. Therefore, sentiment data from Twitter can provide meaningful insights into public perceptions and experiences, especially among digital-native and service-aware users.

Nevertheless, using data from Twitter comes with its own challenges. One major obstacle is the character limit for each tweet, which often results in information being conveyed briefly and in an unstructured manner. Additionally, many users employ informal language, such as slang, abbreviations, or even spelling errors (typos), which can complicate the data analysis process. Therefore, more sophisticated text preprocessing techniques are required so that the data can be accurately processed and analyzed (Koto et al., 2021), (Saymon Ahammad et al., 2024), and the deep learning models used must be capable of understanding these language complexities (Koto et al., 2021), (Šmíd & Král, 2025), (Joshi et al., 2024).

Many machine learning and deep learning techniques have been actively used and researched in many sentiment analys sectors during the past few years. Text categorization has long made use of conventional techniques like Random Forest, Naïve Bayes, and Support Vector Machine (SVM). But deep learning-based methods, like Long Short-Term Memory (LSTM) and Deep Neural Networks (DNN), have shown themselves to be better at managing intricate linguistic structures and more successfully identifying dependencies in text (Shah et al., 2018), (Li et al., 2021), (Zhou et al., 2025). LSTM, a variant of Recurrent Neural Networks (RNN), is known for its effectiveness in capturing relationships between words in a text, especially in sequential data such as tweets. Meanwhile, DNNs offer superior feature representation capabilities, enabling them to understand more complex patterns in sentiment analysis. Although both architectures have been extensively used in natural languages processing (NLP) tasks, direct comparisons of their performances in sentiment analysis of online transportation services in Indonesia remain limited in existing research (Alsini, 2023), (Manalu et al., 2020), (A. Kumar et al., 2020).

2. RESEARCH METHODS

Types, Nature, and Approach of the Research

Type of Research: Experimental, where Twitter data was collected using keywords related to online transportation services such as Gojek, Grab, Maxim, and InDrive. The DNN and LSTM models were compared based on their performance in sentiment analysis. Nature of Research: Descriptive, aiming to compare accuracy, precision, recall, F1-score, time training, and Mean Absolute Error (MAE) of both models. Research Approach: Quantitative, involving numerical data processing and visualizing results using diagrams and tables. To prevent classification bias caused by imbalanced data in the three sentiment classes (positive, negative, neutral), a data balancing process was applied. This involved oversampling the minority classes to ensure that each class had a relatively similar number of training examples, allowing fair and unbiased model evaluation.

Data Analysis Method

The data preprocessing process was carried out to improves the quality of the dataset before sentiment analysis. The steps taken included the removal of irrelevant data, such as tweets without sentiment or ambiguous ones, as well as the removal of unnecessary URLs and links. In addition, symbols and punctuation marks such as !@#$%^&()* were also deleted to make the data cleaner. Case folding was applied to convert all text to lowercase, while stopword removal used to eliminate common words that don't carry sentiment meaning, such as "dan", "di", and "yang". Next, text was tokenized by breaking it to individual words, and stemming was carried out using the Sastrawi library to return the words to their root form. Finally, duplicate data cleaning was performed by removing duplicated tweets to ensure the dataset remains unique.

The data labeling process was carried out by categorizing tweet sentiments into positive, negative, and neutral using IndoBERTweet from Hugging Face Transformers automatically, with random manual validation to ensure label accuracy. In this research, two machine learning models were used, namely Deep Neural Network (DNN) to capture complex relational patterns in data

textual and Long Short-Term Memory (LSTM), which excels at understanding word sequences in text. DNN and LSTM were trained independently for the purpose of training and evaluating the model, and K-Fold Cross Validation was used to increase accuration and prevent overfitting. To evaluate the model's effectiveness during training, training time was measured (Chen et al., 2023), (Minaee et al., 2021). Several measures were used to evaluate the model's performances, including accuration, which counts the number correct predictions, precision, which counts the percentage of right positive predictions, the model's can detect positive data is measured by recall (sensitivity), its ability to recognize negative data by specificity, and its overall performance is depicted by the F1-Score, which is harmonic mean of accuracy and recall.

Research Flow

The research flow begins with problem identification, where a literature review is conducted on sentiment analysis in online transportation services. Next, data collection is carried out through Twitter data crawling using an API to obtain relevant tweets. After the data is collected, the data preprocessing stage is conducted by cleaning and preparing the dataset for analysis. Then, sentiment labeling is applied by classifying the tweets to three categories: positive, negative, and neutral. After that, the model training process is conducted by training DNN and LSTM using the preprocessed data. The model performance is then compared through the model evaluation stage, which uses various evaluation metrics to measure the accuracy and effectiveness of each model. Finally, result interpretation is carried out by compiling a comparative model performance report and drawing conclusions from this research.

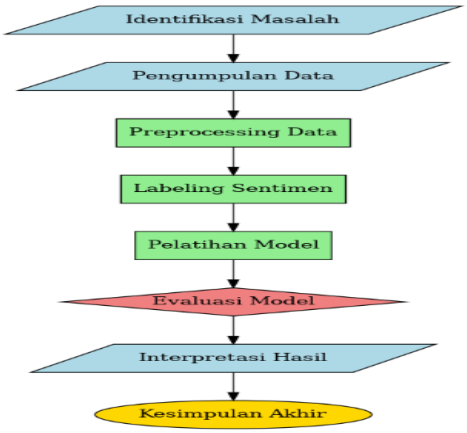

Figure 1. Research flow

Sentiment Analysis

Sentiment analysis is an automated process used to extract emotional information from text, such as positive, negative, or neutral opinions. In the context of online transportation services, sentiment analysis is employed to understand customer satisfaction, identify complaints, and improve service quality (Pan et al., 2019), (Saymon Ahammad et al., 2024), (G. L. Kumar et al., 2021). Online transportation services like Gojek and Grab have rapidly expanded in Indonesia, generating a vast amount of user opinion data, particularly through social media such as Twitter. Sentiment analysis can help service operators monitor user opinions in real-time and make strategic decisions (Chen et al., 2023), (Joshi et al., 2024), (Haque et al., 2019).

In sentiment analysis, especially with social media data like Twitter, the context of a sentence plays a crucial role in determining sentiment polarity. IndoBERT's advantage in understanding context makes it an ideal choice for this task (Koto et al., 2021), (Liu et al., 2024). Moreover, its pre-training on Indonesian language datasets enables the model to handle unique challenges, such as the use of informal words and abbreviations that frequently appear on social media (Abbas & George, 2020).

Deep Learning Models for Sentiment Analysis

With technological advancements, deep learning models have become a primary approach in sentiment analysis. The two most commonly used models for this task are Long Short-Term Memory (LSTM) and Deep Neural Network (DNN).

a. Deep Neural Network (DNN)

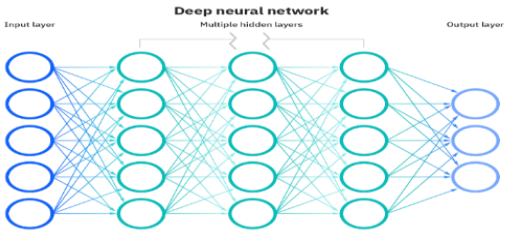

Figure 2. Deep neural network (DNN)

DNN is a neural network with multiple hidden layers capable of capturing non-linear patterns in data. This model is frequently used in sentiment analysis due to its flexibility and efficiency in processing structured data (Rishu et al., 2024), (Minaee et al., 2021).

Advantages of DNN include relatively fast training compared to LSTM and good generalization capabilities for various types of data. However, DNN is less effective at capturing temporal relationships between words, which often presents a challenge in text analysis (Mohamed Ali et al., 2019), (Șerban et al., 2019).

b. Long Short-Term Memory (LSTM)

A Recurrent Neural Network (RNN) variation called LSTM was created to solve the vanishing gradient issue, allowing the model to identify long-term dependencies or temporal linkages in text. This makes LSTM particularly effective for sentiment analysis tasks that require sequential word context (Șerban et al., 2019), (Rishu et al., 2024). Advantages of LSTM include the ability to capture context in long texts and its effectiveness in analyzing sequential data as text and speech (Șerban et al., 2019), (Ao & Fayek, 2023). However, it has drawbacks as longer training times compared to other models (Rishu et al., 2024).

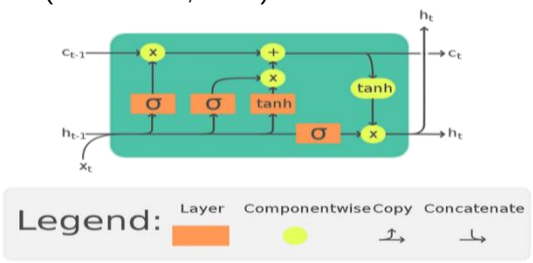

Figure 3. Long short-term memory (LSTM)

Text Preprocessing and Word Embedding Techniques

Text preprocessing includes tokenization, stopword removal, stemming, and normalization. These steps aim to clean the data and enhance the quality of input for deep learning models, leading to more accurate predictions (Saymon Ahammad et al., 2024), (Chen et al., 2023).

Word Embedding Techniques

Word embedding is the representation of words in numerical vector form, allowing models to understand semantic relationships between words. Popular methods such as Word2Vec, GloVe, and BERT are often used to encrease the proceeds of sentiment analysis models (Devlin et al., 2019), (Sengar et al., 2024). Transformer-based approaches like IndoBERT have proven to be highly effective for sentiment analysis tasks in the Indonesian language due to their ability to understand local context (Koto et al., 2021), (Liu et al., 2024).

Model Evaluation in Machine Learning

Model performance is evaluated using the following metrics: a) Accuracy: Shows the proportion of accurate forecasts to all ofinstances, indicating how often the model produces accurate results; b) Precision and Recall: Recall quantifies the model's potential to retrieve all pertinent specimen from the dataset, whereas precision shows the model's capable to accurately identify pertinent examples among the predicted positives; c) F1-Score: is a balanced statistic that is particularly helpful when addressing unequal class distributions. It is a harmonic mean of

accuracy and recall; d) Mean Absolute Error (MAE): Provides information about the model's overall prediction performance by displaying the average magnitude of errors between predicted and actual data; e) Training Time: Measures the model's efficiency during the training process (Chen et al., 2023), (Minaee et al., 2021). These evaluations are crucial for understanding the relative performance of LSTM and DNN in sentiment analysis tasks (Rishu et al., 2024), (Minaee et al., 2021).

Nilai Prediksi	Nilai Aktual: Positive	Nilai Aktual: Negative
Positive	TP	FP
Negative	FN	TN

Figure 4. Confusion matrix

Classification performance metrics such as accuracy, precision, recall, and F1-score.

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total sample}} \times 100\% \quad (1)$$

$$\text{Presisi} = \frac{\text{Jumlah True Positive}}{\text{Number of True Positive+ Number of False Positive}} \quad (2)$$

$$\text{Recall (Sensitivity)} = \frac{TP}{(TP + FN)} \quad (3)$$

$$\text{F1-Score} = \frac{2\,(\text{Presisi x Recall})}{(\text{Presisi + Recall})} \quad (4)$$

$$MAE = (1/n) * \Sigma \mid y\_i - \hat{y}\_i \mid \quad (5)$$

A machine learning model evaluation method called K-Fold Cross-Validation splits the dataset into many subsets, or folds. After that, various combinations of these subsets are used to iteratively train and test the model. This technique provides a more accurate performance evaluation compared to a simple train-test split, especially when working with limited datasets. The main advantages of this method are that it maximizes data utilization, reduces evaluation variance, and produces results that are more representative of the model's generalization ability.

3. RESULT AND DISCUSSION

Review Data

A total of 10,000 tweets were collected through the Twitter data crawling process. Below is a sample of the data retrieved using the crawling method:

Table 1. Crawling results

Create_at	Full_text	Username
Mon Jan 30 23:47:14 +0000 2023	@GrabID Makasih kak respon nya tadi barusan udah dibatalin kok. Aku udah hubungi cs lewat grab nya makasih ya	cesaaaii
Mon Feb 27 23:57:07 +0000 2023	hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	_realine
Thu Mar 30 23:48:29 +0000 2023	TERLAMPAU MURAHHH NI. Rugi kalau tak GRAB HARGA RAHMAH Pelbagai design dan warna yang menarik untuk anda ï¸□â€□ https://t.co/68oaHJ4wxh https://t.co/QAbQD5hOSQ	KakTi_AlifAisya
Sat Apr 29 23:58:17 +0000 2023	@jogmfs Waktu liburan akhir tahun kemaren mau naik maxim dari teras Malioboro sampe st tugu karena hujan ditolak terus. Sampe pada akhirnya di telp salah satu driver ditawarin harganya 100k mau ga? Langsung shock	seumseuma
Tue May 30 23:58:35 +0000 2023	ya Allah pagi pagi dapet grab selalu yg lelet	pjs2200

Data Preprocessing

Irrelevant data were removed, such as tweets without sentiment or those considered ambiguous, as well as unnecessary URLs and links. In addition, symbols and punctuation marks like !@#$%^&()* were also removed to ensure cleaner data.

Table 2. Irrelevant data removal results

Create_at	Full_text	Username
Mon Jan 30 23:47:14 +0000 2023	GrabID Makasih kak respon nya tadi barusan udah dibatalin kok. Aku udah hubungi cs lewat grab nya makasih ya	cesaaaii
Mon Feb 27 23:57:07 +0000 2023	hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	_realine
Thu Mar 30 23:48:29 +0000 2023	TERLAMPAU MURAHHH NI. Rugi kalau tak GRAB HARGA RAHMAH. Pelbagai design dan warna yang menarik untuk anda	KakTi_AlifAisya
Sat Apr 29 23:58:17 +0000 2023	jogmfs Waktu liburan akhir tahun kemaren mau naik maxim dari teras Malioboro sampe st tugu karena hujan ditolak terus. Sampe pada akhirnya di telp salah satu driver ditawarin harganya 100k mau ga? Langsung shock	seumseuma
Tue May 30 23:58:35 +0000 2023	ya Allah pagi pagi dapet grab selalu yg lelet	pjs2200

The next step was case folding, which was applied to convert all text to lowercase, as shown in Table 3.

Table 3. Case Folding results

create_at	full_text	username
mon jan 30 23:47:14 +0000 2023	grabid makasih kak respon nya tadi barusan udah dibatalin kok. aku udah hubungi cs lewat grab nya makasih ya	cesaaaii
mon feb 27 23:57:07 +0000 2023	hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	_realine
thu mar 30 23:48:29 +0000 2023	terlampau murahhh ni. rugi kalau tak grab harga rahmah. pelbagai design dan warna yang menarik untuk anda	kakti_alifaisya
sat apr 29 23:58:17 +0000 2023	jogmfs waktu liburan akhir tahun kemaren mau naik maxim dari teras malioboro sampe st tugu karena hujan ditolak terus. sampe pada akhirnya di telp salah satu driver ditawarin harganya 100k mau ga? langsung shock	seumseuma
tue may 30 23:58:35 +0000 2023	ya allah pagi pagi dapet grab selalu yg lelet	pjs2200

Afterward, stopword removal was performed to eliminate common words that do not carry sentiment meaning, such as "dan" (and), "di" (in/at), and "yang" (which/that).

Table 4. Stopword removal results

create_at	full_text	username
mon jan 30 23:47:14 +0000 2023	makasih kak respon barusan dibatalin hubungi cs grab makasih	cesaaaii
mon feb 27 23:57:07 +0000 2023	hoki pagi pagi dapetnya gini tiap hari naik gojek.	_realine
thu mar 30 23:48:29 +0000 2023	terlampau murahhh ni rugi grab harga rahmah pelbagai design warna menarik	kakti_alifaisya
sat apr 29 23:58:17 +0000 2023	liburan akhir tahun kemaren naik maxim teras malioboro sampe tugu hujan ditolak telp driver ditawarin harga langsung shock	seumseuma
tue may 30 23:58:35 +0000 2023	allah pagi pagi dapet grab lelet	pjs2200

Next, the text was tokenized by splitting it into individual words.

Table 5. Tokenization results

create_at	full_text	username
mon jan 30 23:47:14 +0000 2023	"makasih", "kak", "respon", "barusan", "dibatalin", "hubungi", "cs", "grab", "makasih"	cesaaaii
mon feb 27 23:57:07 +0000 2023	"hoki", "pagi", "pagi", "dapetnya", "gini", "tiap", "hari", "naik", "gojek"	_realine
thu mar 30 23:48:29 +0000 2023	"terlampau", "murahhh", "ni", "rugi", "grab", "harga", "rahmah", "pelbagai", "design", "warna", "menarik"	kakti_alifaisya
sat apr 29 23:58:17 +0000 2023	["liburan", "akhir", "tahun", "kemaren", "naik", "maxim", "teras", "malioboro", "sampe", "tugu", "hujan", "ditolak", "telp", "driver", "ditawarin", "harga", "langsung", "shock	seumseuma
tue may 30 23:58:35 +0000 2023	"allah", "pagi", "pagi", "dapet", "grab", "lelet"	pjs2200

Then, stemming was carried out using the Sastrawi library to reduce words to their root forms.

Table 6. Stemming results

create_at	full_text	username
mon jan 30 23:47:14 +0000 2023	"makasih", "kak", "respon", "baru", "batal", "hubung", "cs", "grab", "makasih"	cesaaaii

mon feb 27 23:57:07 +0000 2023	"hoki", "pagi", "pagi", "dapat", "gini", "tiap", "hari", "naik", "gojek"	_realine
thu mar 30 23:48:29 +0000 2023	"lampau", "murah", "ni", "rugi", "grab", "harga", "rahmah", "bagai", "design", "warna", "tarik"	kakti_alifaisya
sat apr 29 23:58:17 +0000 2023	"libur", "akhir", "tahun", "kemar", "naik", "maxim", "teras", "malioboro", "sampai", "tugu", "hujan", "tolak", "telp", "driver", "tawar", "harga", "langsung", "shock"	seumseuma
tue may 30 23:58:35 +0000 2023	"allah", "pagi", "pagi", "dapat", "grab", "lelet"	pjs2200

The final outcome of the data preprocessing process is presented in Table 7.

Table 7. Data preprocessing results

Original Tweet	After Preprocessing	username
@GrabID Makasih kak respon nya tadi barusan udah dibatalin kok. Aku udah hubungi cs lewat grab nya makasih ya	makasih, kak, respon, baru, batal, hubung, cs, grab, makasih	cesaaaii
hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	hoki, pagi, pagi, dapat, gini, tiap, hari, naik, gojek	_realine
TERLAMPAU MURAHHH NI. Rugi kalau tak GRAB HARGA RAHMAH. Pelbagai design dan warna yang menarik untuk anda ï¸□â€□ https://t.co/68oaHJ4wxh https://t.co/QAbQD5hOSQ	lampau, murah, ni, rugi, grab, harga, rahmah, bagai, design, warna, tarik	kakti_alifaisya
@jogmfs Waktu liburan akhir tahun kemaren mau naik maxim dari teras Malioboro sampe st tugu karena hujan ditolak terus. Sampe pada akhirnya di telp salah satu driver ditawarin harganya 100k mau ga? Langsung shock	libur, akhir, tahun, kemar, naik, maxim, teras, malioboro, sampai, tugu, hujan, tolak, telp, driver, tawar, harga, langsung, shock	seumseuma
ya Allah pagi pagi dapet grab selalu yg lelet	allah, pagi, pagi, dapat, grab, lelet	pjs2200

Data Labeling

The data labeling process was conducted by categorizing the tweet sentiments into positive, negative, and neutral classes, using IndoBERTweet from Hugging Face Transformers automatically.

Table 8. Labeling results

Tweet	Sentiment Category	username
makasih, kak, respon, baru, batal, hubung, cs, grab, makasih	positif	cesaaaii
hoki, pagi, pagi, dapat, gini, tiap, hari, naik, gojek	positif	_realine
lampau, murah, ni, rugi, grab, harga, rahmah, bagai, design, warna, tarik	positif	kakti_alifaisya
libur, akhir, tahun, kemar, naik, maxim, teras, malioboro, sampai, tugu, hujan, tolak, telp, driver, tawar, harga, langsung, shock	negatif	seumseuma
allah, pagi, pagi, dapat, grab, lelet	negatif	pjs2200

After labeling was completed, unnecessary variables were removed, leaving only the essential variables: content and sentiment. Below is a sample of the data after the unnecessary variables were dropped:

Table 9. Sample data after dropping unnecessary variables

No	Tweet	Sentiment Category
1	makasih, kak, respon, baru, batal, hubung, cs, grab, makasih	positif
2	hoki, pagi, pagi, dapat, gini, tiap, hari, naik, gojek	positif
3	lampau, murah, ni, rugi, grab, harga, rahmah, bagai, design, warna, tarik	positif
4	libur, akhir, tahun, kemar, naik, maxim, teras, malioboro, sampai, tugu, hujan, tolak, telp, driver, tawar, harga, langsung, shock	negatif
5	allah, pagi, pagi, dapat, grab, lelet	negatif

Model Training

Training was conducted with two deep learning models, namely DNN and LSTM, separately. To improve accuracy and avoid overfitting, the K-Fold Cross Validation technique was used.

Model Evaluation

Model evaluation as described in Subsection C.3 with K-Fold = 3 yielded results shown in Figure 5 and Table 10.

Figure 5. Model performance comparison (K-Fold=3)

Table 10. Model evaluation results (K-Fold=3)

Metric	DNN	LSTM
Accuracy	0,8100	0,8176
Precision	0,8096	0,8197
Recall	0,8100	0,8176
Specificity	0,9007	0,9051
F1-Score	0,8091	0,8164
MAE (Mean Abs. Error)	0,2449	0,2397
Training Time (seconds)	43,341	126,840

Long Short-Term Memory (LSTM) model outperformed the Deep Neural Network (DNN) in nearly every performance evaluation criterion in the first assessment using a three-fold K-Fold Cross Validation procedure (K = 3). These results further strengthen the finding that LSTM is a more accurate model in understanding the context and structure of complex text data such as tweets. In the accuracy metric, LSTM recorded a value of 81.76%, slightly higher than DNN which obtained an accuracy of 81.00%. This 0.76% difference, although numerically small, becomes significant in large-scale classification scenarios such as social media sentiment analysis, where even small accuracy improvements can affect thousands or even millions of predictions. Furthermore, LSTM's precision and recall reached 81.97% and 81.76%, respectively, slightly above DNN which recorded a precision of 80.96% and recall of 81.00%. This shows that LSTM is more effective in detecting sentiment with lower positive and negative error rates. This capability is very important in the context of public services, where accurately classifying negative sentiment can help companies detect and respond to customer complaints more proactively.

F1-score, which is the harmonic mean of precision and recall, also shows LSTM's dominance (81.64%) compared to DNN (80.91%). This difference further indicates that LSTM is more stable and balanced in its performance across various types of sentiment data. In terms of prediction errors, LSTM's MAE (Mean Absolute Error) of 0.2397 is lower than DNN's 0.2449. This means LSTM's predictions are generally closer to the actual labels. On the other hand, LSTM's specificity (90.51%) is slightly higher than DNN's (90.07%), indicating that LSTM has better capability in accurately recognizing non-positive data (neutral and negative).

However, this performance improvement comes with the consequence of longer training time. LSTM requires 126.84 seconds to complete training, nearly three times longer than DNN's training time of only 43.34 seconds. This difference needs to be considered when computational efficiency becomes a primary factor, especially in environments with limited resources or systems requiring real-time response. Next, model training was carried out with K-Fold equal to 5, and the model evaluation results were obtained as shown in Figure 6 and Table 11.

Table 11. Model evaluation results (K-Fold=5)

Metric	DNN	LSTM
Accuracy	0,8122	0,8215
Precision	0,8120	0,8221
Recall	0,8122	0,8215
Specificity	0,9018	0,9074
F1-Score	0,8112	0,8210
MAE (Mean Abs. Error)	0,2446	0,2315
Training Time (s)	50,435	148,765

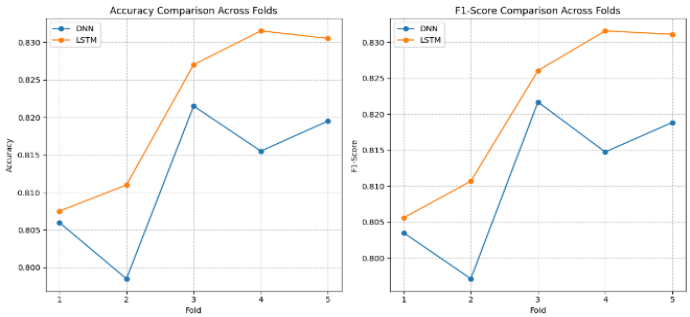

Figure 6. Model performance comparison (K-Fold=5)

Based on the evaluation results using the five-fold K-Fold Cross Validation technique (K = 5), the Long Short-Term Memory (LSTM) model again demonstrated consistently superior performance compared to the Deep Neural Network (DNN) in the sentiment classification task on Twitter data from online transportation services. All the main metrics used to evaluate model performance show LSTM's superiority, although the differences shown are moderate. Specifically, LSTM's classification accuracy reached 82.15%, slightly higher than DNN's 81.22%. Although the difference is only about 0.93%, in the context of sentiment analysis on unstructured data such as tweets, this small increase can significantly impact aggregate results, especially when the model is used for decision-making based on public opinion. The precision and recall metrics also show a similar trend, increasing by 1.01% and 0.93%, respectively, in LSTM. This shows that LSTM model is more reliable in recognizing and classifying the correct sentiment, including in detecting negative and neutral opinions which are usually more difficult to identify.

The LSTM F1-score of 82.10% is also higher than DNN's 81.12 shows better balance between precision and recall. In addition, the Mean Absolute Error (MAE) value on LSTM is lower (0.2315) compared to DNN (0.2446), which shows that LSTM's predictions are generally closer to the actual values. In other words, this model makes fewer extreme prediction errors. Nevertheless, LSTM's performance advantage must be weighed against its much longer training time. The LSTM model requires a training time of 148.76 seconds, almost three times longer than DNN, which only requires 50.43 seconds. This significant difference reflects the complexity of the LSTM architecture, which is designed to understand word sequences and temporal relationships in text but sacrifices computational efficiency. Meanwhile, DNN is simpler architecturally and can be processed in parallel, making it more efficient for real-time system implementation or on devices with limited computing resources. Overall, LSTM provides better results than DNN, especially in terms of accuracy and the ability to recognize both classes. However, if training time efficiency is a crucial factor, DNN may remain a viable option with a reasonable performance compromise.

4. CONCLUSION

Based on the evaluation results, these findings indicate that the LSTM model is more accurate for analyzing user sentiment toward Gojek/Grab/Maxim/InDrive, enabling companies to implement a more advanced automated opinion monitoring system. Based on the evaluation results, these findings indicate that the LSTM model is more accurate for analyzing user sentiment toward Gojek/Grab/Maxim/InDrive, enabling companies to implement a more advanced automated opinion monitoring system. In addition to accuracy, the training time of the LSTM model was also taken into consideration. This is particularly important for real-world deployment in environments such as contact centers, where retraining or regular model updates may be necessary. The LSTM model demonstrated acceptable training efficiency, making it a practical choice not only for its performance but also for its adaptability in resource-constrained, real-time operational settings.

ACKNOWLEDGEMENTS

We would like to expression our sincere gratitude to everyone who helped us finish this study. We would especially want to thank our colleagues for their wise counsel, unwavering encouragement, and support during the study process. We are also grateful to all participants who generously shared their time and input for this study. Our gratitude extends to the institutions and organizations that provided essential support and facilities during the implementation of this research. Every form

of assistance and contribution has played a significant role in ensuring the smooth progress and success of this work. Thank you for your dedication and collaboration

REFERENCES

Abbas, S. K., & George, L. E. (2020). The performance differences between using recurrent neural networks and feedforward neural network in sentiment analysis problem. *Iraqi Journal of Science*, *61*(6), 1512–1524. https://doi.org/10.24996/ijs.2020.61.6.31

Alsini, R. (2023). Analysis of Real Time Twitter Sentiments using Deep Learning Models. *Journal of Applied Data Sciences*, *4*(4), 480–489. https://doi.org/10.47738/jads.v4i4.146

Ao, S. I., & Fayek, H. (2023). Continual Deep Learning for Time Series Modeling. *Sensors*, *23*(16), 91–108. https://doi.org/10.3390/s23167167

Bijarnia, S., Khetan, R., Ilavarasan, P. V., & Kar, A. K. (2019). Analyzing Customer Engagement Using Twitter Analytics: A Case of Uber Car-Hailing Services. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11701 LNCS*, 404–414. https://doi.org/10.1007/978-3-030-29374-1_33

Chen, D., Su, W., Wu, P., & Hua, B. (2023). Joint multimodal sentiment analysis based on information relevance. *Information Processing and Management*, *60*(2), 1567–1589. https://doi.org/10.1016/j.ipm.2022.103193

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. https://doi.org/https://doi.org/10.48550/arXiv.1810.04805

Haque, M. R., Akter Lima, S., & Mishu, S. Z. (2019). Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews. *3rd International Conference on Electrical, Computer and Telecommunication Engineering, ICECTE 2019*, 161–164. https://doi.org/10.1109/ICECTE48615.2019.9303573

Joshi, A., Dabre, R., Kanojia, D., Li, Z., Zhan, H., Haffari, G., & Dippold, D. (2024). Natural Language Processing for Dialects of a Language: A Survey. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(3), 1–28. https://doi.org/10.1145/3712060

Koto, F., Lau, J. H., & Baldwin, T. (2021). INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 10660–10668. https://doi.org/10.18653/v1/2021.emnlp-main.833

Kumar, A., Srinivasan, K., Cheng, W. H., & Zomaya, A. Y. (2020). Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Information Processing and Management*, *57*(1). https://doi.org/10.1016/j.ipm.2019.102141

Kumar, G. L., Karan, I., Pati, R., & Koduru, S. (2021). Sentiment Analysis in Transportation System 1 1. *Internatonal Journal of Creative Research Thoughts (IJCRT)*, *9*(11), 618–627.

Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction. *Information Processing and Management*, *58*(5), 8234–8245. https://doi.org/10.1016/j.ipm.2021.102673

Liu, J., Li, K., Zhu, A., Hong, B., Zhao, P., Dai, S., Wei, C., Huang, W., & Su, H. (2024). Application of Deep Learning-Based Natural Language Processing in Multilingual Sentiment Analysis. *Mediterranean Journal of Basic and Applied Sciences*, *08*(02), 243–260. https://doi.org/10.46382/mjbas.2024.8219

Mamani-Coaquira, Y., & Villanueva, E. (2024). A Review on Text Sentiment Analysis with Machine Learning and Deep Learning Techniques. *IEEE Access*, *60*(2), 103123. https://doi.org/10.1109/ACCESS.2024.3513321

Manalu, B. U., Tulus, & Efendi, S. (2020). Deep learning performance in sentiment analysis. *2020 4th International Conference on Electrical, Telecommunication and Computer Engineering, ELTICOM 2020 - Proceedings*, 97–102. https://doi.org/10.1109/ELTICOM50775.2020.9230488

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad Khasmakhi, N., Asgari-Chenaghlu, M., & Gao, J. (2021). Deep Learning--based Text Classification: A Comprehensive Review. *ACM Computing Surveys*, *54*, 1–40. https://doi.org/10.1145/3439726

Mohamed Ali, N., El Hamid, M. M. A., & Youssif, A. (2019). Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models. *International Journal of Data Mining & Knowledge Management Process*, *09*(03), 19–27. https://doi.org/10.5121/ijdkp.2019.9302

Pan, D., Yuan, J., Li, L., & Sheng, D. (2019). Deep neural network-based classification model for Sentiment Analysis. *BESC 2019 - 6th International Conference on Behavioral, Economic and Socio-Cultural Computing, Proceedings*, *35*(3), 1–12. https://doi.org/10.1109/BESC48373.2019.8963171

Rishu, Singh, A., & Tanwar, S. (2024). Unveiling Sentiments: CNN-LSTM Based Social Media Sentiment Analysis. *2024 Asia Pacific Conference on Innovation in Technology, APCIT 2024*, *34*(2), 891–907. https://doi.org/10.1109/APCIT62007.2024.10673560

Saymon Ahammad, M., Sinthia, S. A., Muaj Chowdhury, M., Asif, N.-A.-A., & Nurul AfsarIkram, M. (2024).

Sentiment Analysis of Various Ride Sharing Applications Reviews: A Comparative Analysis Between Deep Learning and Machine Learning Algorithms. *Transportation Research Part C*, *148*, 434–448. https://doi.org/10.1007/978-3-031-69986-3_33

Sengar, S. S., Hasan, A. Bin, Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*, *66*(1), 1–27. https://doi.org/10.1007/s11042-024-20016-1

Șerban, O., Thapen, N., Maginnis, B., Hankin, C., & Foot, V. (2019). Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Information Processing and Management*, *56*(3), 1166–1184. https://doi.org/10.1016/j.ipm.2018.04.011

Shah, D., Campbell, W., & Zulkernine, F. H. (2018). A Comparative Study of LSTM and DNN for Stock Market Forecasting. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 4148–4155. https://doi.org/10.1109/BigData.2018.8622462

Šmíd, J., & Král, P. (2025). Cross-lingual aspect-based sentiment analysis: A survey on tasks, approaches, and challenges. *Information Fusion*, *120*, 119422. https://doi.org/10.1016/j.inffus.2025.103073

Zhou, W., An, L., Han, R., & Li, G. (2025). Classification and severity assessment of disaster losses based on multi-modal information in social media. *Information Processing and Management*, *62*(5). https://doi.org/10.1016/j.ipm.2025.104179