Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

June 2026. Vol. 5. No. 3; e-ISSN: 2808-4519

Sentiment Analysis of Pre-Loved Shoe Product Sales Based on X Reviews with a Comparison of Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) Algorithms

Setyo Harry Nugroho[1*], Al-khowarizmi[2]

[1,2] *Muhammadiyah University of North Sumatra*
setyoharrynugroho@gmail.com[1]*

Abstract

The rapid growth of social media enables consumers to express opinions about products openly, including preloved shoes. These reviews are crucial as they can influence purchase intentions and brand perception. This study aims to analyze user reviews on the X (Twitter) platform using Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) algorithms. A total of 1,005 reviews were collected, then preprocessed and balanced into 738 data consisting of positive and negative sentiments. The results show that SVM achieved an accuracy of 68%, while LSTM obtained 61.49% in its best configuration. Thus, SVM demonstrates better efficiency in classifying simple text, whereas LSTM requires more complex parameters to achieve optimal performance. This research is expected to serve as a reference for utilizing sentiment analysis to support business decision-making in the preloved product market.

***Keywords**: Sentiment Analysis, Preloved Shoes, SVM, LSTM, X (Twitter).*

1. Introduction

The development of the internet in Indonesia has been increasing year by year, proving that the internet plays an important role in various aspects of life. Initially, the internet was developed and used for military purposes in the 19th century, but developments and technological needs have continued to demand the use of the internet in various sectors such as commerce, social media, and even government systems [1]. Twitter is a social media platform that allows users to create and share short messages called “tweets,” which can be up to 280 characters long and include text, videos, or links to other websites. Twitter is a social media platform that allows users to create and share short messages called “tweets,” which can be up to 280 characters long and include text, videos, or links to other websites [2]. People use Twitter to communicate, ask questions, request directions, support, advice, and validate interpretations or open ideas through discussion. Twitter has combined personal publishing and communication, resulting in a new type of real-time publishing [3].

Shoes are one of the most sought-after fashion items, serving as foot protection but now also supporting appearance and reflecting the wearer's personality. Some consumers are willing to spend more to get shoes with the best brands, attractive designs, and comfort. However, the large number of reviews left by buyers, ranging from praise to complaints, makes it difficult for businesses to monitor consumer responses. Product reviews are crucial because they directly impact consumer purchasing interest and brand image. Reviews not only Product reviews are crucial because they directly impact consumer purchasing interest and brand image. Reviews not only reflect individual experiences with the product but also contain information about consumer perceptions that can be analyzed through sentiment analysis [4].

Through this approach, opinions spread in the form of text can be identified and classified into positive or negative sentiment categories. This provides buyers with an overview of product and service quality. This analysis enables manufacturers and sellers to understand consumer expectations by improving product deficiencies and developing appropriate marketing strategies. In addition, this analysis can be used as a tool to anticipate a decline in market interest arising from negative sentiment. To obtain accurate classification results, researchers used the Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) algorithms. Researchers compared the two algorithms to determine which method was more accurate and efficient in classifying sentiment, thereby contributing to more targeted business decision-making for both sellers and buyers [5], [6].

2. Research Metodology

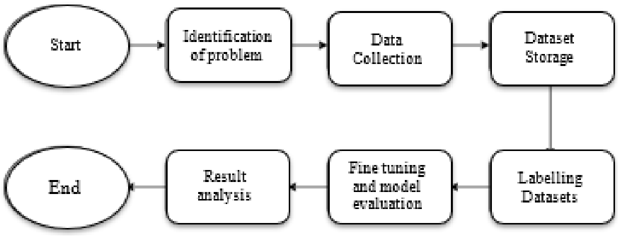


Fig. 1: Flowchart Design

2.1. Type of Research

This research uses a quantitative research method that compares SVM and LSTM algorithms for sentiment analysis of preloved shoe sales based on X reviews. This research begins with a literature study, dataset collection, data preparation, dataset labeling, model fine tuning & evaluation, and analysis of results [7].

2.2. Data Collection

Data collection uses web scraping or crawling techniques. Compared to manual surveys, crawling or scraping is more efficient because it can quickly retrieve large amounts of data. By using crawling, researchers can collect the data needed for their research easily and effectively.

The crawling process begins by importing the nscrape.modules.twitter module required to access Twitter-related functions. Next, an empty list named tweets is created to store the tweet content that has been successfully retrieved. A for loop is then used to iterate through each tweet found by Twitter Search Scraper with the specified search parameters. Each tweet that meets these criteria will have its content (tweet.content) retrieved and added to the tweets list. This code is a simple but effective example of automatically collecting tweet data, which can later be used for various analysis purposes, such as sentiment analysis in research on preloved shoes. By using the lang:id parameter, the code ensures that only tweets in Indonesian are collected, making them relevant for specific analysis targets.

2.3. Labelling Datasets

The labeling process was carried out after the data collection process from platform X (Twitter) was completed. Each comment or review obtained was analyzed and labeled based on the meaning of the sentence and the expression of opinion conveyed by the user. This labeling process was carried out manually by referring to an Indonesian sentiment dictionary to determine whether a text was classified as positive or negative sentiment. Reviews containing praise, satisfaction, or positive experiences with preloved shoes are labeled as positive, while comments containing complaints, criticism, or dissatisfaction are labeled as negative. Of the total 1,005 data collected, only 738 data were used after the balancing process, with 369 positive data and 369 negative data. This step aims to ensure that the model is not biased towards either sentiment class and can produce more accurate classifications during training and testing.

2.4. Text Preprocessing

After the labeling process, the text data undergoes preprocessing to convert the raw data into clean data that is ready to be processed by machine learning algorithms. This stage begins with case folding to standardize the text by converting all letters to lowercase. Then, data cleaning is performed to remove irrelevant characters such as punctuation marks, numbers, emoticons, hashtags, URLs, and mentions. Next is tokenization, which is the process of breaking down text into word fragments (tokens) so that they can be analyzed separately. The next stage is normalization, which aims to convert non-standard words into standard words according to the Big Indonesian Dictionary, because text on social media often uses informal words. After that, stop word removal is performed to remove common words that do not contribute significantly to the meaning, such as “yang”, ‘dan’, or “di”. Finally, stemming is performed to convert inflected words into their basic form so that words with the same root meaning can be considered equivalent by the model. The end result of this preprocessing stage is clean, uniform text that can be converted into numerical representations using techniques such as TF-IDF before being used in SVM and LSTM models.

3. Result and Discussion

3.1. Result

The dataset used in this study consists of a collection of texts taken from the social media platform X (formerly known as Twitter). This platform was chosen because it is often used by the public to express opinions, feelings, and assessments on various topics, including products such as preloved shoes. The amount of text data collected was 1,005, each representing one opinion or statement from a user that explicitly or implicitly expressed sentiment towards preloved shoes.

This labeling process considers the overall context of the sentence, the style of language, and words that express the user's emotions or attitudes toward the topic being discussed. The labels used are divided into two main categories: positive and negative sentiment. Positive sentiment indicates a favorable assessment or feeling toward preloved shoes, while negative sentiment indicates dissatisfaction, rejection, or criticism of them.

Table 1: Datasets

No	Comment	Product Name	Label
1	The glue is peeling off	Sneaker Casual	Negatif
2	The model fits well and feels nice and soft, but is slightly heavy and smaller than shoes of the same size.	Aerostreat Casual 37	Positif
3	The shoe size is actually really small	Onit Onsu Tiger	Negatif
4	Cheap, Good, Awesome	Tuta Casual Shoes	Positif

The data used in the processing stage amounted to 1,005 data points. Then, the data was manually labeled with positive and negative values. From the labeling process, 636 positive class data points and 369 negative class data points were obtained. The balance between positive and negative classes greatly affects the accuracy obtained. Therefore, the number of positive classes was reduced to 369 data points to balance the negative classes. The total dataset used ultimately amounted to 738 data points. The dataset will be divided into training data and testing data, with 80% allocated to training data and 20% to testing data. There are 147 testing data points and 519 training data points.

The models used in this study are Long Short-Term Memory (LSTM) and Support Vector Machine (SVM). The LSTM model was tested with several parameter variations, such as batch size, number of neurons, and epoch. From the test results using a batch size of 16, the best results were obtained with an accuracy of 71.62%, precision of 72.22%, recall of 70.27%, and F-measure of 71.23%.

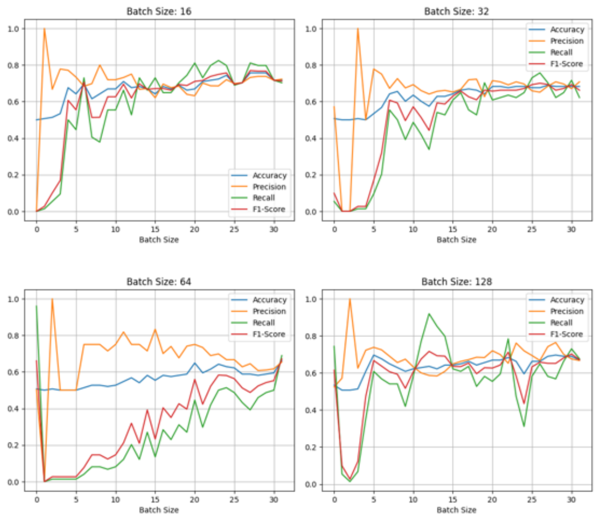


Fig . 2: History of LSTM Testing Based on Batch Size

Table 2: LSTM Testing Based on Batch Size

Batch Size	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Time (Second)
16	71.62	70.27	72.22	71.23	73
32	68.24	62.16	70.77	66.19	54
64	66.22	68.92	65.38	67.11	40
128	66.89	67.57	66.67	67.11	40

The highest accuracy was obtained at a batch size of 16 with an accuracy of 71.62%. The highest recall value was also obtained at the same batch size, which was 70.27%. The highest precision value was also obtained at the same batch size, which was 72.22%. The highest f-measure value was also obtained at the same batch size, which was 71.23%. Batch size 16 took the longest time, which was 73 seconds, while batch size 128 only took 40 seconds. This shows that increasing the batch size can speed up the training process, although it should be noted that accuracy and other evaluation metrics do not always increase with increasing batch size.

Pengujian ini juga menunjukkan bahwa pengujian neuron dilakukan dengan mengubah besar nilai neuron dalam rentang tertentu untuk mendapatkan nilai optimal. Pada pengujian ini, ditetapkan batch size sebesar 16 dan epoch sebanyak 32 karena berdasarkan pengujian sebelumnya mendapatkan nilai akurasi tertinggi pada neuron 64. Pada pengujian ini, nilai neuron akan ditetapkan sebanyak 16, 32, 64, 128.

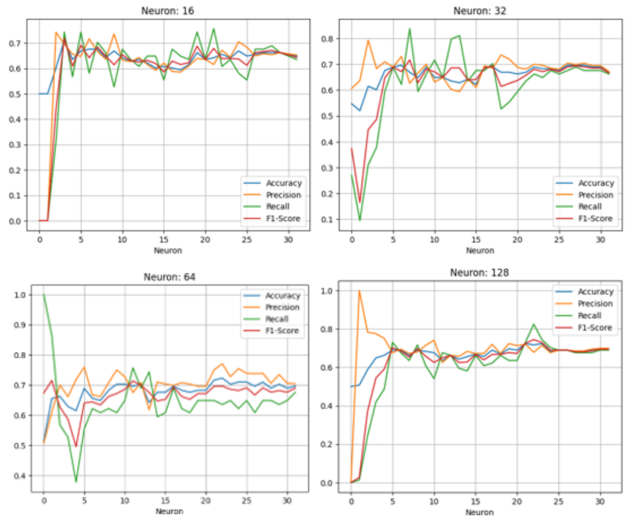


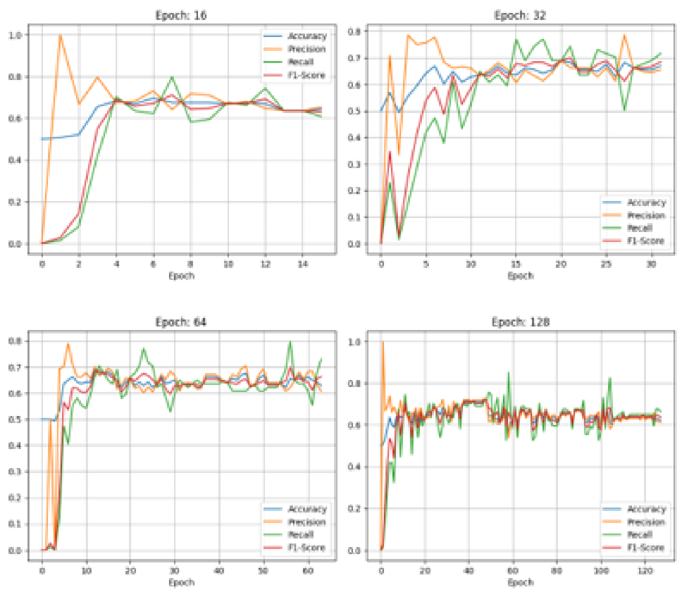
Fig. 3: History of LSTM Testing Based on Neuron Values

Table 3: Neuron Value-Based LSTM Testing

Neuron	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Time (second)
16	64.86	63.51	65.28	64.38	51
32	66.89	66.22	67.12	66.67	66
64	69.59	67.57	70.42	68.97	74
128	69.59	68.92	69.86	69.39	131

The highest accuracy and precision results were obtained with 64 neurons, with an accuracy of 69.59% and precision of 70.42%. Meanwhile, the highest recall and f-measure were obtained with 128 neurons, with a recall value of 68.92% and an f-measure of 69.39%. Based on the time required, a larger number of neurons tends to require longer processing time. Neuron 16 requires 51 seconds and neuron 128 requires the longest time, which is 131 seconds. This shows that increasing the number of neurons can increase the complexity of the model and result in longer training time, although it does not always guarantee a significant improvement in performance.

Testing based on epoch values is done by changing the epoch value within a certain range to obtain the optimal value. In this test, 16 neurons and a batch size of 16 are set, and the epoch values will start from 16, 32, 64, and 128.



Fig, 4: History of LSTM Testing Based on Epoch Values

Table 4. LSTM Testing Based on Epoch Values

Epoch	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (detik)
16	64.19	60.81	65.22	62.94	41
32	66.89	71.62	65.43	68.39	73
64	62.84	72.97	60.67	66.26	141
128	62.16	66.22	61.25	63.64	286

LSTM testing based on epoch values shows the model evaluation results based on variations in the number of epochs. It can be seen that the highest accuracy and precision values were obtained at epoch 32, with an accuracy of 66.89% and a precision of 65.43%. Meanwhile, the highest recall and f-measure values were also achieved at epoch 32, at 71.62% and 68.39%, respectively. In terms of training time, it can be seen that the greater the number of epochs, the longer the time required. Epoch 16 only took 41 seconds, while epoch 128 took the longest time, namely 286 seconds. This shows that adding epochs can indeed improve model performance to a certain extent, but it also causes a significant increase in training time. Therefore, the selection of the number of epochs needs to consider the trade-off between model performance and training time efficiency.

The optimal combination of parameters for the LSTM model resulted in a final accuracy of 61.49% with a training time of 73 seconds. Although LSTM is capable of recognizing sentence context well, its performance is still limited due to the relatively small amount of data and the need for more complex parameters.

Meanwhile, Support Vector Machine testing was conducted to utilize the sklearn library in Python. The data used in this classification was divided into 80% training data and 20% testing data. This data division scenario was able to produce high accuracy values in the Support Vector Machine method. Furthermore, the method used in feature extraction was TF-IDF.

Table 5: Support Vector Machine Algorithm Testing

Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Time (Second)
68	68	68	68	17

Based on the test results, the SVM model showed more stable performance than LSTM with an accuracy of 68%, balanced precision and recall values, and a much shorter training time of 17 seconds. This shows that SVM is more efficient in processing simple text data with fairly accurate results, while LSTM is more suitable for analysis with more complex language contexts and large data sets.

Table 6: Comparison of Sentiment Results

Metode	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Time (Second)
LSTM	61.49	62.16	61.33	61.74	73
SVM	68	68	68	68	17

A comparison between the two models shows that SVM has advantages in terms of speed and stability of results, while LSTM excels in understanding the sequential context of text. Implementations of both models are also applied in the form of a sentiment analysis website that can process user reviews directly and display the results of positive or negative sentiment classification. This website is designed to be responsive and displays prediction results along with the confidence level of the model used.

Fig. 5: SVM and LSTM Sentiment Analysis Website

3.2. Discussion

The SVM algorithm shows better performance than LSTM in sentiment classification of preloved shoe reviews on platform X. The LSTM model with the best configuration produced an accuracy of 61.49%, while SVM achieved an accuracy of 68% with faster training time. This confirms that SVM is more efficient and stable for datasets with simple text structures, while LSTM requires more in-depth parameter settings and large amounts of data to achieve optimal results.

This study suggests that future research should expand the amount and variety of data to make the analysis results more representative. In addition, the use of more advanced deep learning models such as BiLSTM, GRU, or CNN-LSTM is also recommended to improve accuracy. The use of Indonesian language embeddings such as IndoBERT or FastText can be an alternative to improve language context understanding. Comprehensive hyperparameter optimization is also important to make the model more stable. Based on the results of this study, the SVM algorithm is recommended for practical implementation because it has a balance between accuracy, efficiency, and speed in analyzing sentiment in preloved product review texts.

References

[1] S. Akmala, “Perkembangan Internet Pada Generasi Muda Di Indonesia Dengan Kaitan Undang-Undang Ite Yang Berlaku,” *Cyber Secur. dan Forensik Digit.*, vol. 1, no. 2, pp. 45–49, 2019, doi: 10.14421/csecurity.2018.1.2.1371.

[2] S. Wulandari and F. N. Hasan, “Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 768, 2024, doi: 10.30865/mib.v8i2.7520.

[3] G. GROSSECK and C. HOLOTESCU, “Can we use Twitter in educational activities ?,” *4Th Int. Sci. Conf. …*, no. October, 2015, [Online]. Available: http://www.researchgate.net/profile/Gabriela_Grosseck/publication/228685205_Can_we_use_Twitter_for_educational_activities/links/0fcfd50aa46c8394e6000000.pdf

[4] PARDEDE, A., FAUZI, A., MAULITA, Y., MARPAUNG, P., & ARIANTO, S. (2025). NON-LINEAR SMART MODEL SUPPORT VECTOR MACHINE (SVM) AS A HOAX NEWS CLASSIFIER. *Journal of Theoretical and Applied Information Technology*, *103*(22).

[5] Suhada, S., & Pardede, A. M. H. (2019). PENGKLASTERAN DOKUMEN DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR CLUSTERING. *Jurnal Sistem Informasi Kaputama (JSIK)*, *3*(2), 1-6.

[6] Nouval, M., Habibi, F. M., Rahmi, A., Amru Bittaqwa, M. D., & Agustianto, R. (2026). Sentiment Analysis of Indonesian National Team Failure in the 2026 World Cup Qualifications Using Support Vector Machine Algorithm. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *5*(2), 2223–2228. https://doi.org/10.59934/jaiea.v5i2.1830

[7] Setyawan, J. (2026). Sentiment Analysis of Honda Esaf Frame Quality Based on Reviews on Platform X using Support Vector Machine Algorithm. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *5*(2), 2477–2483. https://doi.org/10.59934/jaiea.v5i2.1901Darusman, A. (2026). Sentiment Analysis of Mie Gacoan Pemuda Cirebon Restaurant Reviews Using Support Vector Machine. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *5*(2), 2490–2494. https://doi.org/10.59934/jaiea.v5i2.1905