IJIE (Indonesian Journal of Informatics Education)

Research Paper ISSN: 2549-0389

A Comparative Study of Machine Learning Models for Sentiment Analysis of Dana App Reviews

Yudianto Sujana
Informatics Education Program, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Indonesia
Corresponding Email: yudianto.sujana@staff.uns.ac.id

Abstract:

Sentiment analysis of user reviews has become increasingly important for mobile app developers, as it can provide valuable insights into customer satisfaction and guide the improvement of app features. In this study, we compared the performance of three machine learning models - Support Vector Machine, Neural Network, and Bidirectional Long Short-Term Memory - in classifying the sentiment of user reviews for the Dana mobile application. Our results showed that the Bi-LSTM model outperformed the other models, achieving the highest accuracy, precision, recall, and F1-score. The superior performance of the Bi-LSTM model can be attributed to its ability to capture long-term dependencies and contextual information within the review text, which is crucial for accurate sentiment analysis. These findings highlight the effectiveness of deep learning techniques in handling the complexities of language and sentiment analysis, particularly in user-generated content. The insights from this study can inform the development of more accurate and efficient sentiment analysis tools for mobile app reviews, ultimately benefiting both app developers and users.

Keywords: *Bi-LSTM, Deep Learning, Machine Learning, Sentiment Analysis.*

IJIE (Indonesian Journal of Informatics Education)
Vol 7 Issue 2 2023
DOI: http://dx.doi.org/10.20961/ijie.v7i2.93132
Received: 06/09/23 Revised: 09/09/23 Accepted: 09/09/23 Online: 31/12/23

Introduction

The rapid advancement of digital technology has significantly transformed the financial services landscape (Kristiyanti et al., 2020). One manifestation of this is the proliferation of digital wallet applications such as Dana, which offer a variety of features ranging from money transfers to bill payments. The popularity of these applications is due in part to their ease of access and various attractive promotions offered. However, behind this convenience lies a fundamental question about the quality of service and user perceptions of the application.

Sentiment analysis of user reviews on platforms like the Google Play Store is one effective way to measure user satisfaction (Fu et al., 2013; Islam, 2014). By analyzing the words and sentences used in reviews, we can identify a product or service's positive and negative aspects (Singla et al., 2017). In the context of the Dana application, sentiment analysis can provide a more comprehensive picture of the application's strengths and weaknesses from the user's perspective. The insights gained from sentiment analysis of user reviews have broad implications for the technology industry, businesses, and consumers (Pai et al., 2017; Ranjan & Mishra, 2020).

Researchers have employed various machine learning models to perform sentiment analysis on app reviews. The Support Vector Machine and Naïve Bayes algorithms have been commonly used, with studies demonstrating their effectiveness in classifying reviews as positive, negative, or neutral. For instance, researchers collected OVO and DANA review data from the Google Play store and compared the performance of the Naïve Bayes and SVM models (Kristiyanti et al., 2020). Similarly, another study utilised four supervised learning algorithms, including SVM and Naïve Bayes, to analyse sentiment and emotions in reviews for the Boost mobile payment app (Balakrishnan et al., 2020). Additionally, a separate study leveraged the SVM method to conduct a sentiment analysis of online transportation application reviews (Handani et al., 2019; Hermanto et al., 2020), revealing that SVM could accurately classify sentiment.

In addition to traditional machine learning models, more advanced techniques, such as Neural Networks, Recurrent Neural Networks, and Convolutional Neural Network have also been explored for sentiment analysis. These models have the potential to capture more complex relationships within the text and provide even more accurate sentiment classification. For example, Neural Networks can learn to recognize intricate patterns and dependencies in the text, allowing them to better understand the nuanced meaning and sentiment expressed by users (Alzhrani et al., 2022; Y. Wang et al., 2014). Similarly, Recurrent Neural Networks, which are specialized for processing sequential data like text, have shown impressive performance in sentiment analysis tasks by accounting for the contextual information and flow of language (Sari et al., 2020; X. Wang et al., 2016). Convolutional Neural Networks have also demonstrated efficacy in sentiment analysis by extracting local features and patterns from the text. By applying convolutional filters to the input text, Convolutional Neural Networks are able to identify salient features and extract meaningful representations that can be effectively leveraged for sentiment classification tasks (H. Kim & Jeong, 2019; Y. Kim, 2014; Ruder et al., 2016; Severyn & Moschitti, 2015; Van et al., 2017).

While these studies provide valuable insights, there is a need for more in-depth analysis of the performance of various machine learning models in the context of sentiment analysis for the Dana application specifically. This research paper aims to bridge this gap by conducting a comparative study of multiple machine learning models, including Support Vector Machine, Neural Network, and Recurrent Neural Network, for sentiment analysis of Dana app reviews. It is expected that these techniques will yield accurate sentiment classification. The results of this study are anticipated to assist Dana application developers in improving the quality of their services and to provide users with insights to aid their decision-making before using the application.

Research Method

Data Collection

The primary data for this study were collected from Google Play Store reviews of the Dana application. Using the Google-play-scraper library (JoMingyu, 2019), 30,000 recent Indonesian language reviews were gathered. These reviews were then pre-processed to extract the review text and corresponding star ratings, which served as the target variable for the sentiment analysis task. A dataset sample is provided in Table 1, while the data distribution is illustrated in Figure 1.

Table 1. Dataset sample

Review	English Translation	Label
Sangat bagus, dana membantu yg tidak punya rek bank. cpt dlam transaksi. puas banget aplikasi dana mantap.	Very good, Dana helps those who don't have a bank account. Fast transactions. Really satisfied, Dana application is great.	Positive
Semoga lebih ditingkatkan lagi keamanan nya.	Hopefully the security will be further improved.	Neutral
Keamanannya tidak terjamin, buruk, banyak penipu. Uang saya hilang tanpa transaksi apa pun dan tidak dapat mengajukan keluhan.	Security is not guaranteed, bad, many scammers. My money is gone without any transaction and can't file a complaint.	Negative

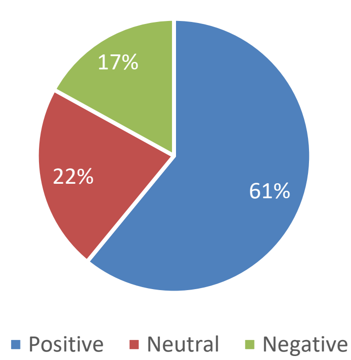

Figure 1. Data distribution

Data Pre-processing

The collected data underwent several pre-processing steps to prepare it for modelling. These steps included:

- **Cleaning:** Removing non-alphabetic characters, numbers, and punctuation.
- **Tokenization:** Breaking down text into individual words or tokens.
- **Lowercasing:** Converting all text to lowercase for consistency.
- **Removing stop words:** Eliminating common words that have little semantic value (e.g., "ada", "di", "yang").
- **Vectorization:** Converting the text data into numerical representations using the TF-IDF technique so that machine learning algorithms can process it.

Data Splitting

The pre-processed dataset was split into training and testing sets. The training set was used to train the machine learning models, while the testing set was used to evaluate their performance. A split of 80% of the data was allocated for training, and the remaining 20% was used for testing.

Model Selection

Three machine learning algorithms were selected for this study:

- **Support Vector Machine (SVM):** We utilized the scikit-learn library with a linear kernel and set the probability parameter to True (Pedregosa et al., 2011).
- **Neural Network:** We implemented a multi-layer perceptron neural network model with two hidden layers and ReLU activation function, using the Keras library (Chollet & others, 2015). The model was optimized using the Adam algorithm with a learning rate of 1e-3.
- **Bidirectional Long Short-Term Memory (Bi-LSTM):** We leveraged the Keras library to construct a Bidirectional LSTM-based architecture (Schuster & Paliwal, 1997). The model utilized Adam optimization with a learning rate of 1e-3 and a dropout rate of 0.5.

Model Training and Evaluation

The selected models were trained on the training dataset, and their performance was optimized through hyper parameter tuning. The models were then evaluated using the testing set, and their performance was measured using metrics such as accuracy, precision, recall, and F1-score.

Result and Discussion

Result

The performance of the SVM, Neural Network, and Bi-LSTM models on the sentiment classification task is presented in Table 2 and Figure 1.

Table 2. Performance comparison of SVM, Neural Network, and Bi-LSTM models

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.90	0.88	0.87	0.87
Neural Network	0.93	0.90	0.91	0.91
Bi-LSTM	0.94	0.92	0.92	0.92

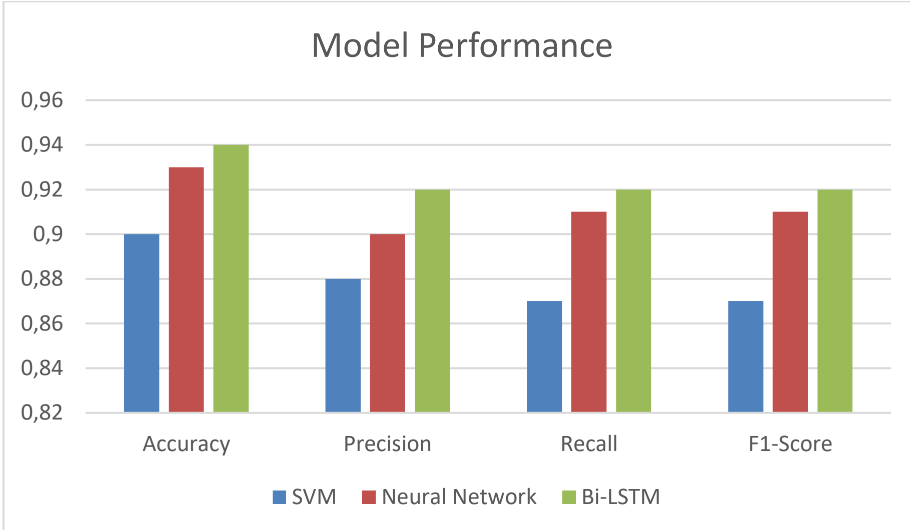

Figure 2. Performance comparison of SVM, Neural Network, and Bi-LSTM models

As shown in Table 2 and Figure 2, the Bi-LSTM model exhibited the highest performance across all metrics. This suggests that the Bi-LSTM architecture is particularly well-suited for capturing the contextual information and semantic relationships inherent in the Dana app reviews, thereby leading to more accurate sentiment classification.

The SVM model, while performing reasonably well, exhibited lower performance compared to the Bi-LSTM. This could be attributed to the ability of the Bi-LSTM to better capture the temporal dependencies and complex linguistic patterns within the review text, which may be challenging for the SVM model to learn effectively.

The neural network model also demonstrated strong performance, but it fell short of the Bi-LSTM in terms of accuracy, precision, and F1-score.

The superior performance of the Bi-LSTM model can be attributed to its ability to capture long-term dependencies and contextual information within the review text, which is crucial for accurate sentiment analysis. Additionally, the incorporation of bidirectional processing allows the model to consider both past and future information, further enhancing its ability to understand the sentiment conveyed in the reviews.

Discussion

The results of this study demonstrate that the Bi-LSTM model effectively classified the sentiment of user reviews for the Dana application. This indicates that machine learning techniques can be successfully applied to analyze sentiment in the context of mobile app reviews (Ranjan & Mishra, 2020).

Several factors may have contributed to the performance of the models. Firstly, the quality and quantity of the training data played a significant role, as the use of a large and diverse dataset likely helped the models to generalize well to unseen data (Paullada et al., 2021). Secondly, the pre-processing steps, such as tokenization and stemming, were crucial for preparing the text data for modelling (Alshdaifat et al., 2021). Finally, the choice of hyper parameters for each model also had a substantial impact on their performance(Kandhro et al., 2020).

The findings of this study provide valuable insights for developers and researchers interested in analyzing sentiment within app reviews. The superior performance of the Bi-LSTM model highlights the effectiveness of deep learning techniques in handling the complexities of language and sentiment analysis, particularly in the context of user-generated content (Baccouche et al., 2018; Chakraborty et al., 2022).

Furthermore, the comparative analysis of the different machine learning models suggests that the choice of algorithm can significantly impact the accuracy of sentiment classification. Developers and researchers should carefully evaluate the trade-offs between model complexity, interpretability, and performance when selecting the most appropriate technique for their specific use case.

Limitations and Future Work

While this study provides valuable insights into the performance of different machine learning models for sentiment analysis of Dana app reviews, a few limitations could be addressed in future research.

Firstly, expanding the dataset by collecting more reviews from a wider range of sources, such as other app stores or social media platforms, could lead to more robust and generalizable results. Secondly, incorporating additional feature engineering techniques, such as pre-trained word embedding or domain-specific lexicons, may further improve the performance of the models. Finally, exploring ensemble methods, which combine multiple models' strengths, could yield even higher accuracy in sentiment classification.

Future research could expand upon the current document-level sentiment classification approach by exploring aspect-based sentiment analysis techniques (Brauwers & Frasincar, 2023). This would enable identifying specific app features or components that elicit positive or negative user sentiments, providing more granular insights to application developers.

Additionally, investigating the application of transfer learning or few-shot learning techniques could be a promising direction, as these approaches can leverage knowledge gained from other sentiment analysis domains to improve performance on the Dana app reviews (Ruder et al., 2019).

Conclusion

This study compared the performance of three machine learning models - SVM, Neural Network, and Bi-LSTM - for sentiment analysis of user reviews for the Dana mobile application. The results showed that the Bi-LSTM model outperformed the other models, achieving the highest accuracy, precision, recall, and F1-score. The superior performance of the Bi-LSTM model highlights the effectiveness of deep learning techniques in handling the complexities of language and sentiment analysis, particularly in the context of user-generated content. The findings of this study provide valuable insights for developers and researchers interested in analyzing sentiment within app reviews and contribute to the growing body of literature on the application of machine learning for sentiment analysis.

References

Alshdaifat, E., Alshdaifat, D., Alsarhan, A., Hussein, F., & El-Salhi, S. M. F. S. (2021). The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance. *Data*, *6*(2), 11. https://doi.org/10.3390/data6020011

Alzhrani, A., Alatawi, A., Alsharari, B., & Albalawi, U. (2022). Towards security awareness of mobile applications using semantic-based sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *13*(4), 800–809. https://doi.org/10.14569/IJACSA.2022.0130493

Baccouche, A., Garcia-Zapirain, B., & Elmaghraby, A. (2018). Annotation technique for health-related tweets sentiment analysis. *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 382–387. https://doi.org/10.1109/ISSPIT.2018.8642685

Balakrishnan, V., Selvanayagam, P. K., & Yin, L. P. (2020). Sentiment and emotion analyses for Malaysian mobile digital payment applications. *ACM International Conference Proceeding Series*, 67–71. https://doi.org/10.1145/3388142.3388144

Brauwers, G., & Frasincar, F. (2023). A survey on aspect-based sentiment classification. *ACM Computing Surveys*, *55*(4), 1–37. https://doi.org/10.1145/3503044

Chakraborty, I., Kim, M., & Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, *59*(3), 600–622. https://doi.org/10.1177/00222437211052500

Chollet, F., & others. (2015). Keras. https://keras.io/

Fu, B., Lin, J., Liy, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013). Why people hate your app: Making sense of user feedback in a mobile app store. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1276–1284. https://doi.org/10.1145/2487575.2488202

Handani, S. W., Saputra, D. I. S., Hasirun, Arino, R. M., & Ramadhan, G. F. A. (2019). Sentiment analysis for Go-Jek on Google Play Store. *Journal of Physics: Conference Series*, *1196*(1), 012032. https://doi.org/10.1088/1742-6596/1196/1/012032

Hermanto, Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab user sentiment analysis on Google Play using Naive Bayes algorithm and support vector machine based SMOTE technique. *Journal of Physics: Conference Series*, *1641*(1), 012102. https://doi.org/10.1088/1742-6596/1641/1/012102

Islam, M. R. (2014). Numeric rating of apps on Google Play Store by sentiment analysis on user reviews. *2014 International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 1–4. https://doi.org/10.1109/ICEEICT.2014.6919058

JoMingyu. (2019). google-play-scraper. https://github.com/JoMingyu/google-play-scraper

Kandhro, I. A., Jumani, S. Z., Ali, F., Shaikh, Z. U., Arain, M. A., & Shaikh, A. A. (2020). Performance analysis of hyperparameters on a sentiment analysis model. *Engineering, Technology & Applied Science Research*, *10*(4), 6016–6020. https://doi.org/10.48084/etasr.3549

Kim, H., & Jeong, Y.-S. (2019). Sentiment classification using convolutional neural networks. *Applied Sciences*, *9*(11), 2347. https://doi.org/10.3390/app9112347

Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. https://doi.org/10.3115/v1/D14-1181

Kristiyanti, D. A., Putri, D. A., Indrayuni, E., Nurhadi, A., & Umam, A. H. (2020). E-wallet sentiment analysis using Naïve Bayes and support vector machine algorithm. *Journal of Physics: Conference Series*, *1641*(1), 012079. https://doi.org/10.1088/1742-6596/1641/1/012079

Pai, H.-T., Lai, H.-W., Wang, S.-L., Wu, M.-F., & Chuang, Y.-T. (2017). Recommendations for mobile applications. *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*, 1–6. https://doi.org/10.1145/3109761.3109771

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, *2*(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Ranjan, S., & Mishra, S. (2020). Comparative sentiment analysis of app reviews. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–7. https://doi.org/10.1109/ICCCNT49239.2020.9225348

Ruder, S., Ghaffari, P., & Breslin, J. G. (2016). INSIGHT-1 at SemEval-2016 task 4: Convolutional neural networks for sentiment classification and quantification. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 178–182. https://doi.org/10.18653/v1/S16-1026

Ruder, S., Peters, M. E., Swayamdipta, S., & Wolf, T. (2019). Transfer learning in natural language processing. *Proceedings of the 2019 Conference of the North*, 15–18. https://doi.org/10.18653/v1/N19-5004

Sari, W. K., Rini, D. P., Malik, R. F., & Azhar, I. S. B. (2020). Sequential models for text classification using recurrent neural network. *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, 333–340. https://doi.org/10.2991/aisr.k.200424.050

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681. https://doi.org/10.1109/78.650093

Severyn, A., & Moschitti, A. (2015). Twitter sentiment analysis with deep convolutional neural networks. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 959–962. https://doi.org/10.1145/2766462.2767830

Singla, Z., Randhawa, S., & Jain, S. (2017). Statistical and sentiment analysis of consumer product reviews. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT.2017.8203960

Van, V. D., Thai, T., & Nghiem, M.-Q. (2017). Combining convolution and recursive neural networks for sentiment analysis. *Proceedings of the Eighth International Symposium on Information and Communication Technology*, 151–158. https://doi.org/10.1145/3155133.3155158

Wang, X., Jiang, W., & Luo, Z. (2016). Combination of convolutional and recurrent neural network for sentiment analysis of short texts. https://aclanthology.org/C16-1229

Wang, Y., Li, Z., Liu, J., He, Z., Huang, Y., & Li, D. (2014). Word vector modeling for sentiment analysis of product reviews. *Communications in Computer and Information Science, 496*, 168–180. https://doi.org/10.1007/978-3-662-45924-9_16