Sentiment Analysis on Online Shopping Store Product by Product Review

Jovan Amarta Liem
Computer Science Department,
School of Computer Science,
Bina Nusantara University,
Jakarta 11480, Indonesia
jovan.liem@binus.ac.id

Wilbert Yang
Computer Science Department,
School of Computer Science,
Bina Nusantara University,
Jakarta 11480, Indonesia
wilbert.yang@binus.ac.id

Eko Setyo Purwanto
Computer Science Department,
School of Computer Science,
Bina Nusantara University,
Jakarta 11480, Indonesia
eko.purwanto@binus.ac.id

Muhamad Keenan Ario
Computer Science Department,
School of Computer Science,
Bina Nusantara University,
Jakarta 11480, Indonesia
muhamad.ario@binus.ac.id

***Abstract*—Many online stores provide a customer review feature, where customers who buy goods at a particular store can provide a review of the goods. Reviews of these products can be categorized as positive or negative reviews. With so many reviews on a product, this can make it difficult for customers to choose products. The objective to be achieved in this research is to perform sentiment analysis on e-commerce product reviews using machine learning algorithms and models. Then compare the accuracy, recall, and precision. Choosing a model that has the best accuracy in performing sentiment analysis will certainly have a significant impact on classifying public sentiment towards a product. The XGBoost model has an overall accuracy of 81%, and the BERT model has an overall accuracy of 92%.**

***Keywords*—XGBoost, BERT, Sentiment Analysis, Product Review.**

I. Introduction

Technological advances, especially the internet, have affected various aspects of life with the aim of making everything practical to do, one of which is people's shopping habits [1]. Now, people can use a platform to shop online called E-commerce. E-commerce has been increasingly used and has developed a lot until now [1, 2]. Many online stores provide a customer review feature, where customers who buy goods at a particular store can provide a review of the goods. Reviews of these products can be categorized as positive or negative reviews. With so many reviews on a product, this can make it difficult for customers to choose products. This is because there are some reviews that are less relevant to the product itself, for example reviews about shipping goods or even reviews that have sarcasm sentences.

One method that can be used to classify these reviews is sentiment analysis. Sentiment analysis is a field of research that examines people's opinions, sentiments, emotions and evaluations of things such as products, services and other events. The main focus of sentiment analysis is to extract writing or opinion into a sentiment that is positive or negative [3, 4].

The background of the research in sentiment analysis is to accurately classify the comments, especially in E-commerce platforms. The information that can be obtained from product reviews can educate consumers about the product and help them make a decision to buy it or not. So analyzing sentiment from product reviews is an important aspect and a strategy in increasing the marketability of a store's products.

There are several studies that have discussed sentiment analysis and various methods for performing sentiment analysis as found in research [2, 5, 6]. Sentiment analysis is a classification method to determine whether a text has a positive, negative, or neutral sentiment [6]. To perform sentiment analysis, a machine learning model can be created that is trained using customer review data on e-commerce. To make the process faster, feature extraction can be done first. Feature extraction is the process of extracting features contained in the data so that the data used is suitable for modeling and speeding up the process [7].

After reading several papers, researchers found that the most widely used techniques for feature extraction are TF-IDF, GloVe, and word2vec. Then, because the dataset used has labels, the model that researcher uses is a supervised learning model. To minimize reviews that are less relevant to the product, researchers will also remove reviews such as shipping reviews, reviews that use sarcasm, and reviews that have rating errors.

The objective to be achieved in this research is to perform sentiment analysis on e-commerce product reviews using machine learning algorithms and models. Then compare the accuracy, recall, precision. Choosing a model that has the best accuracy in performing sentiment analysis will certainly have a significant impact in classifying public sentiment towards a product. With this, store owners can easily find out whether their products are well received by customers.

This paper will give a complete look at the topic through five sections. Section 1 is an introduction, section 2 reviews previous research on the topic, section 3 explains the methods used for the study, section 4 shows the results, section 5 summarizes the main points.

II. Literature Review

To perform sentiment analysis, it is necessary to create a machine learning or deep learning model. Before creating a machine learning or deep learning model, there is a method to improve the accuracy of the model, namely word embedding. The word embedding model is a mapping between words and vectors to see the similarity between words [8].

It is known that the results obtained when doing word embedding using BERT will give better results than using word2vec or glove [9]. This happens because BERT for word embedding can capture the contextual representation of each token in the sentence. There is a method to improve the accuracy of models that use Word2Vec or GloVe word embedding, called S-EWE. However, adding S-EWE method only increases the accuracy by about 1% [10].

There are various models that can be used to perform sentiment analysis, including Naive Bayes, and Support Vector Machine (SVM). Nikmah, et al. [11] use the model to do classification tasks, but their model also can be used for sentiment analysis. Another example is from [12, 13], both studies use Naive Bayes for sentiment analysis and get a good accuracy.

There has also a research on Long-Short Term Memory (LSTM) and Convolutional Neural Network (CNN). In [9], the models used are LSTM, CNN, and LSTM-CNN models. They try to combine the LSTM and CNN models and get a fairly high accuracy of 88.06% for LSTM, 90.04% for CNN, and 93.91% for LSTM-CNN. Other studies have used CNN models in their research to perform various tasks [14-20]. There is also LSTM model that has been used for sentiment analysis for product review tasks [21], as well as a combination of LSTM and CNN which forms a hybrid model [22].

There is research on detecting sarcasm [23-25], detecting sarcasm in the dataset used can affect the accuracy of the model built. Using Random Forest Algorithm to detect sarcasm, the accuracy of sentiment analysis using Naive Bayes model built by Yunitasari, et al. increased by 5.49% [24].

There are studies that use other models as well such as Gradient Boosting, eXtreme Gradient Boosting, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest conducted by research [26]. The results they got were 89.72% for Gradient Boosting, 91.25% for eXtreme Gradient Boosting and Logistic Regression 91.52%, 92.14% for Support Vector Machine, 90.12% for Decision Tree, 91.86% for Random Forest. The highest result from research [26] is Bi-LSTM + Max Pooling with 93.32% accuracy.

Bidirectional Encoder Representations from Transformers (BERT), BERT is a machine learning model created to improve the accuracy of NLP (Natural Language Processing) [27]. BERT is one of the most used model for doing NLP tasks [28], because BERT can be utilized for NLP tasks such as rating prediction [29], sentiment analysis, and other tasks related to NLP. There is research that uses BERT for sentiment analysis product reviews, and the accuracy they get for the BERT model is also quite high at 94%. The results of the BERT model they used were the highest compared to other models they used [30].

To resolve sentiment analysis tasks, there are many BERT models that can be used. For example, IndoBERT [31], and RoBERT [32, 33]. According to Suhartono et al. [33] BERT model has better accuracy than RoBERTa they built. There are other models such as DomBERT which is used [34], and BERT-Base-Chinese [35]. Many BERT models also can be used to do sentiment analysis tasks such as comparative sentiment analysis in airline customer reviews [36], also sentiment analysis towards COVID-19 using data in social media [37]. BERT model also shows the best result among other models to do sentiment analysis in the stock market [38].

A couple of studies utilize BERT model to perform sentiment analysis on a product review [39, 40]. The BERT model used shows that it results an excellent accuracy score and outperforms other machine learning model in accuracy scores [40]

XGBoost or Extreme Gradient Boosting is a machine learning technique to do classification and regression tasks based on Gradient Boosting Decision Tree (GBDT). This model is a decision tree based on machine learning algorithm that uses a gradient enhancer framework [41]. XGBoost is already used in many cases. Unlike BERT, XGBoost is not a pre-trained model. There are some studies that use XGBoost to do sentiment analysis [42-49].

Hendrawan, et al [41]. used an unbalanced dataset to XGBoost model proved to do sentiment analysis on product review tasks better than the Naive Bayes model. Hendrawan, et al. got high accuracy which is equal to 89% with Word2Vec and TF-IDF, the XGBoost model accuracy surpasses their naive bayes model which has 88% accuracy for NB + TF-IDF, and 83% for NB + Word2Vec. Similar to the previous study, Alghazzawi, et al. [42] use a large and balanced dataset to further prove that the XGBoost model can output a good result, although if their model is compared to their ERF-XGB model, their ERF-XGB model has a better result.

This research paper aims to compare the XGBoost model and the BERT model.

III. Method

A. Dataset

The dataset used in this research is about product amazon reviews and sourced from Kaggle. Researchers randomly selected 30000 samples from 34 million data which has 15063 positive sentiment and 14937 negative sentiment. The dataset was divided into 80% for training and 20% for testing.

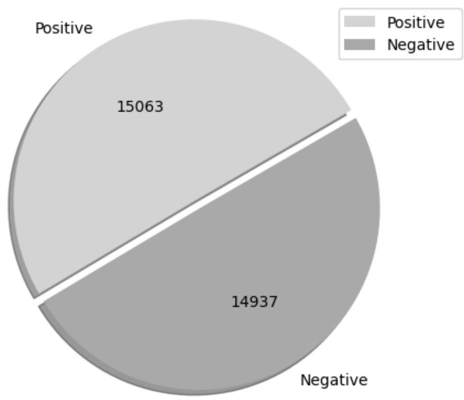

Fig 1: Balanced dataset visualization using pie chart

Word Cloud

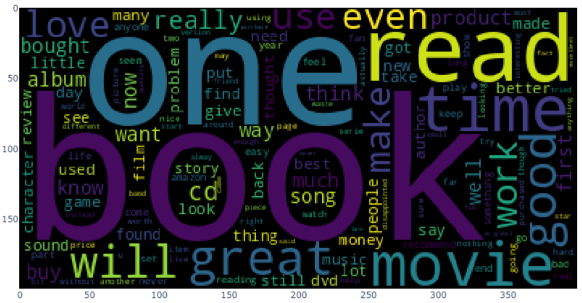

Fig 2: Word Cloud

B. Pre-Processing Data

Before the dataset is used to train the model, the researcher performs several data preprocessing steps first. These stages include data cleaning, removing stopwords, removing punctuations, and lemmatizing.

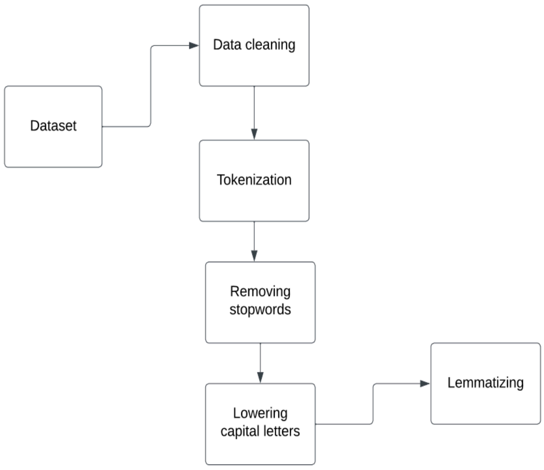

Fig 3: Flowchart of data preprocessing

The data cleaning process includes removing data containing numbers, symbols and so on that can interfere with the results of sentiment analysis and make the data alphanumeric, then the researcher will also remove stopwords using stopwords from the NLTK library in python. Finally, lemmatizing will be carried out on the data using the NLTK library, namely WordNetLemmatizer in python.

C. BERT

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model that utilizes the Transformer architecture. This architecture enables BERT to comprehend the meaning of words based on their context, which includes the surrounding words that come before and after. Consequently, BERT can generate word representations that align with the contextual meaning within a sentence.

To create these contextual representations, BERT uses an encoder on the input text. The text is first tokenized into smaller units, which are then converted into word vector embeddings. The BERT encoder, composed of multiple Transformer layers, processes these tokens through multi-head self-attention operations and feed-forward neural networks, generating contextual representations. A key aspect of BERT is its bidirectionality, allowing it to attend to both the preceding and succeeding tokens simultaneously. This process iterates across each encoder layer, facilitating a deeper understanding of the text. Ultimately, BERT produces a vector representation derived from the tokens, which can be leveraged for various natural language tasks, such as text classification.

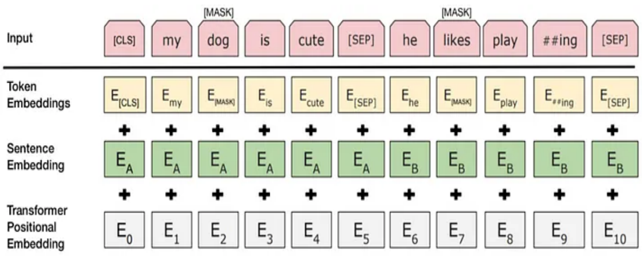

Fig 4: BERT input representation [50].

TABLE I. BERT Parameter

Parameter	**Parameter Used**
Optimizer	AdamW
Model Name	Bert-Base-Uncased
Batch Size	32
Epochs	15
Learning Rate	2e-5

D. XGBOOST

In addition to the BERT model, there exists another highly effective machine learning model for classification and regression tasks known as eXtreme Gradient Boosting (XGBoost). The fundamental structure of XGBoost is an ensemble learning model comprised of a sequence of Gradient Boosting Decision Trees (GBDT). In this approach, each decision tree constructed has the objective of minimizing the residual error from the previous decision tree model. Consequently, these decision tree models work in a complementary manner, collectively forming a more powerful and precise model for predicting target variables, whether it is for classification or regression problems [51].

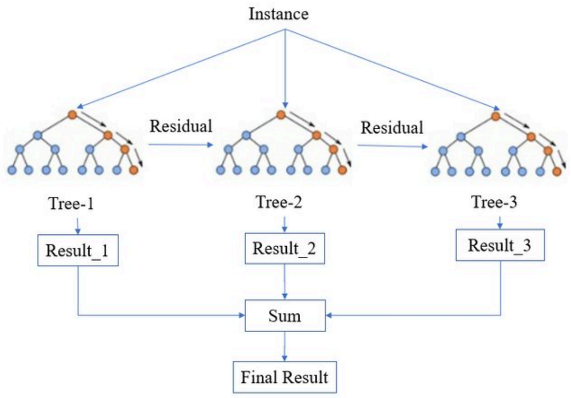

Fig 5: Simplified structure of XGBoost [51].

The BERT model was built using the tokenizer and word embedding from BERT, while the XGBoost model used TF-IDF. The maximum token that the BERT model used was 150.

E. Evaluation

Accuracy is the metric used to evaluate how the model perform. This metric can describe how frequent the model predicts an output correctly. To calculate the Accuracy metric, a confusion matrix is needed to obtain the essentials for the accuracy formula.

Accuracy is a measure of how often the classifier makes the correct prediction. The formula is the number of correct predictions divided by the total number of predictions.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

Recall is a measure of actual observations which are predicted correctly.

$$Recall = \frac{TP}{TP + FN}$$

Precision is a measure of correctness that is achieved in true prediction. It tells how many predictions are actually positive out of all the total positive predicted.

$$Precision = \frac{TP}{TP + FP}$$

F1 is the harmonic mean of precision and recall. Unlike the simple mean, the purpose of the harmonic mean is that it is not sensitive to extremely large values.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

IV. Results

TABLE II. Result of The Model

Model Name	Test Size	Classification Report			
		Accuracy	*F1*	*Recall*	*Precision*
XGBoost	6000	81%	81%	83%	80%
BERT	6000	92%	93%	95%	90%

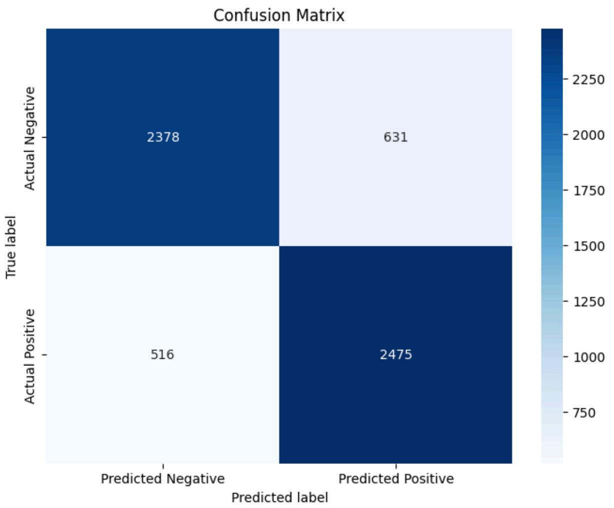

Fig 6: XGBoost Confusion Matrix

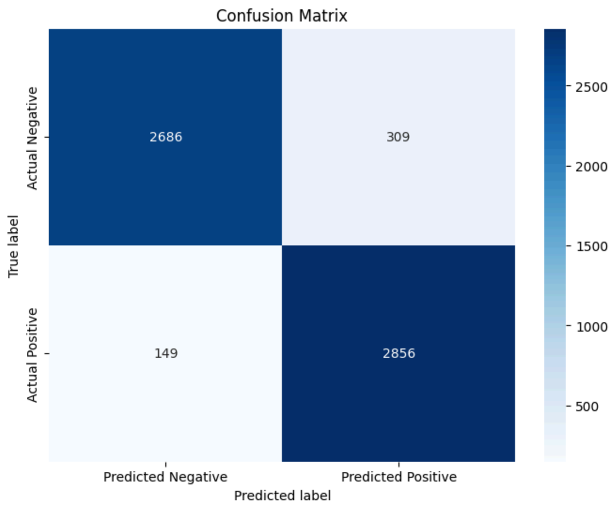

Fig 7: BERT Confusion Matrix

V. Conclusion

The results showed that the BERT model performed better than XGBoost for sentiment analysis on product review. BERT with batch size 32, 15 epochs and 2×10^{-5} learning rate has 92% accuracy. Meanwhile, the XGBoost model has 81% accuracy. The BERT model has a precision of 90%, a recall of 95% and 93% f1 score. The XGBoost model has a precision of 80%, a recall of 83% and 81% f1 score. For the future research, it is better to try another word embedding such as Word2Vec and GloVe to the model, also sarcasm detection can be applied to increase the performance of the model.

References

[1] M. H. D. Bangsa, S. Priyanta, and Y. Suyanto, "Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, p. 123, Apr. 2020, doi: 10.22146/ijccs.51646.

[2] I. C. Juanatas, R. R. Fajardo, E. T. Manansala, A. A. Pasilan, J. R. Tabor, and H. D. A. Balmeo, "Sentiment Analysis Platform of Customer Product Reviews," in *Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019*, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 230–234. doi: 10.1109/ICCIKE47802.2019.9004415.

[3] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham: Springer International Publishing, 2012. doi: 10.1007/978-3-031-02145-9.

[4] M. v Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018, doi: https://doi.org/10.1016/j.cosrev.2017.10.002.

[5] S. Wang and C. Manning, "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, H. Li, C.-Y. Lin, M. Osborne, G. G. Lee, and J. C. Park, Eds., Jeju Island, Korea: Association for Computational Linguistics, Jul. 2012, pp. 90–94. [Online]. Available: https://aclanthology.org/P12-2018

[6] T. A. Le, D. Moeljadi, Y. Miura, and T. Ohkuma, "Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian Tweets," in *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, K. Hasida, K.-F. Wong, N. Calzorari, and K.-S. Choi, Eds., Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 123–131. [Online]. Available: https://aclanthology.org/W16-5415

[7] M. F. bin Harunasir, N. Palanichamy, S.-C. Haw, and K.-W. Ng, "Sentiment Analysis of Amazon Product Reviews by Supervised Machine Learning Models," *Journal of Advances in Information Technology*, vol. 14, no. 4, pp. 857–862, 2023, doi: 10.12720/jait.14.4.857-862.

[8] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *Journal of Biomedical Informatics*, vol. 110, p. 103539, 2020, doi: https://doi.org/10.1016/j.jbi.2020.103539.

[9] Syaiful Imron, E. I. Setiawan, Joan Santoso, and Mauridhi Hery Purnomo, "Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 586–591, Jun. 2023, doi: 10.29207/resti.v7i3.4751.

[10] Q. Li, X. Li, Y. Du, Y. Fan, and X. Chen, "A New Sentiment-Enhanced Word Embedding Method for Sentiment Analysis," *Applied Sciences*, vol. 12, no. 20, p. 10236, Oct. 2022, doi: 10.3390/app122010236.

[11] T. L. Nikmah, M. Z. Ammar, Y. R. Allatif, R. M. P. Husna, P. A. Kurniasari, and A. S. Bahri, "Comparison of LSTM, SVM, and naive bayes for classifying sexual harassment tweets," *Journal of Soft Computing Exploration*, vol. 3, no. 2, Sep. 2022, doi: 10.52465/joscex.v3i2.85.

[12] K. A. F. A. Samah, N. S. Jailani, R. Hamzah, R. Aminuddin, N. A. Z. Abidin, and L. S. Riza, "Aspect-Based Classification and Visualization of Twitter Sentiment Analysis Towards Online Food Delivery Services in Malaysia," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 37, no. 1, pp. 139–150, Jul. 2024, doi: 10.37934/araset.37.1.139150.

[13] M. Shenify, "Sentiment analysis of Saudi e-commerce using naïve bayes algorithm and support vector machine," *International Journal of Data and Network Science*, vol. 8, no. 3, pp. 1607–1612, Jun. 2024, doi: 10.5267/j.ijdns.2024.3.006.

[14] A. Muzakir, K. Adi, and R. Kusumaningrum, "Advancements in Semantic Expansion Techniques for Short Text Classification and Hate Speech Detection," *Ingénierie des systèmes d information*, vol. 28, no. 3, pp. 545–556, Jun. 2023, doi: 10.18280/isi.280302.

[15] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," *IEEE Access*, vol. 6, pp. 23253–23260, Jan. 2018, doi: 10.1109/ACCESS.2017.2776930.

[16] F. Ratnawati and E. Winarko, "Sentiment Analysis of Movie Opinion in Twitter Using Dynamic Convolutional Neural Network Algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 12, no. 1, p. 1, Jan. 2018, doi: 10.22146/ijccs.19237.

[17] W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, "Aspect Based Sentiment Analysis with Feature Enhanced Attention CNN-BiLSTM," *IEEE Access*, vol. 7, pp. 167240–167249, 2019, doi: 10.1109/ACCESS.2019.2952888.

[18] R. Man and K. Lin, "Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network," in *Proceedings of IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 769–772. doi: 10.1109/IPEC51340.2021.9421110.

[19] I. Priyadarshini and C. Cotton, "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis," *The Journal of Supercomputing*, vol. 77, no. 12, pp. 13911–13932, Dec. 2021, doi: 10.1007/s11227-021-03838-w.

[20] A. Bello, S. C. Ng, and M. F. Leung, "A BERT Framework to Sentiment Analysis of Tweets," *Sensors*, vol. 23, no. 1, Jan. 2023, doi: 10.3390/s23010506.

[21] S. Xiao, H. Wang, Z. Ling, L. Wang, and Z. Tang, "Sentiment Analysis for Product Reviews Based on Deep Learning," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1651/1/012103.

[22] 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)*, IEEE, Dec. 2022, pp. 51–56. doi: 10.1109/ICITISEE57756.2022.10057720.

[23] M. Bouazizi and T. Otsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016, doi: 10.1109/ACCESS.2016.2594194.

[24] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 1, p. 53, Jan. 2019, doi: 10.22146/ijccs.41136.

[25] D. Šandor and M. Bagić Babac, "Sarcasm detection in online comments using machine learning," *Information Discovery and Delivery*, vol. 52, no. 2, pp. 213–226, Mar. 2024, doi: 10.1108/IDD-01-2023-0002.

[26] Z. Tao and Z. Wu, "Sentiment Analysis of Product Reviews Based on Bi-LSTM and Max Pooling," 2024. doi: 10.3233/FAIA231407.

[27] Fransiscus and A. S. Girsang, "Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method," *International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 281–288, Dec. 2022, doi: 10.14445/22315381/IJETT-V70I12P226.

[28] M. S. Sayeed, V. Mohan, and K. S. Muthu, "BERT: A Review of Applications in Sentiment Analysis," *HighTech and Innovation Journal*, vol. 4, no. 2. Ital Publication, pp. 453–462, Jun. 01, 2023. doi: 10.28991/HIJ-2023-04-02-015.

[29] X. Zhao and Y. Sun, "Amazon Fine Food Reviews with BERT Model," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 401–406. doi: 10.1016/j.procs.2022.10.056.

[30] M. Mostafa and A. AlSaeed, "Sentiment Analysis Based on Bert for Amazon Reviewer," *Journal of the ACS Advances in Computer Science*, vol. 13, no. 1, pp. 1–10, 2022, doi: 10.21608/asc.2023.171559.1007.

[31] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian

reviews using fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[32] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.

[33] D. Suhartono, K. Purwandari, N. H. Jeremy, S. Philip, P. Arisaputra, and I. H. Parmonangan, "Deep neural networks and weighted word embeddings for sentiment analysis of drug product reviews," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 664–671. doi: 10.1016/j.procs.2022.12.182.

[34] H. Xu, B. Liu, L. Shu, and P. Yu, "DomBERT: Domain-oriented Language Model for Aspect-based Sentiment Analysis," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 1725–1731. doi: 10.18653/v1/2020.findings-emnlp.156.

[35] F. Zhou, "Sentiment analysis and implementation in film evaluation utilizing BERT," *Applied and Computational Engineering*, vol. 33, no. 1, pp. 224–233, Feb. 2024, doi: 10.54254/2755-2721/33/20230274.

[36] Z. Li, C. Yang, and C. Huang, "A Comparative Sentiment Analysis of Airline Customer Reviews Using Bidirectional Encoder Representations from Transformers (BERT) and Its Variants," *Mathematics*, vol. 12, no. 1, p. 53, Dec. 2023, doi: 10.3390/math12010053.

[37] N. Rifat, M. Ahsan, R. Gomes, and Md. Chowdhury, "COVID-19 Sentiment Analysis applying BERT," in *2022 IEEE International Conference on Electro Information Technology (eIT)*, IEEE, May 2022, pp. 417–422. doi: 10.1109/eIT53891.2022.9813777.

[38] M. G. Sousa, K. Sakiyama, L. de S. Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "BERT for Stock Market Sentiment Analysis," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, Nov. 2019, pp. 1597–1601. doi: 10.1109/ICTAI.2019.00231.

[39] Y. Cao, Z. Sun, L. Li, and W. Mo, "A Study of Sentiment Analysis Algorithms for Agricultural Product Reviews Based on Improved BERT Model," *Symmetry*, vol. 14, no. 8, Aug. 2022, doi: 10.3390/sym14081604.

[40] A. S. M. AlQahtani, "Product Sentiment Analysis for Amazon Reviews," *International Journal of Computer Science and Information Technology*, vol. 13, no. 3, pp. 15–30, Jun. 2021, doi: 10.5121/ijcsit.2021.13302.

[41] I. R. Hendrawan, E. Utami, and A. D. Hartanto, "Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 143–149, Jun. 2022, doi: 10.29408/edumatic.v6i1.5613.

[42] D. M. Alghazzawi, A. G. A. Alquraishee, S. K. Badri, and S. H. Hasan, "ERF-XGB: Ensemble Random Forest-Based XG Boost for Accurate Prediction and Classification of E-Commerce Product Review," *Sustainability (Switzerland)*, vol. 15, no. 9, May 2023, doi: 10.3390/su15097076.

[43] A. Athar, S. Ali, M. M. Sheeraz, S. Bhattachariee, and H. C. Kim, "Sentimental Analysis of Movie Reviews using Soft Voting Ensemble-based Machine Learning," in *2021 8th International Conference on Social Network Analysis, Management and Security, SNAMS 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/SNAMS53716.2021.9732159.

[44] M. N. Fakhruzzaman, S. Z. Jannah, S. W. Gunawan, A. I. Pratama, and D. A. Ardanty, "IndoPolicyStats: sentiment analyzer for public policy issues," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 482–489, Feb. 2024, doi: 10.11591/eei.v13i1.5263.

[45] A. Samih, A. Ghadi, and A. Fennan, "Enhanced sentiment analysis based on improved word embeddings and XGboost," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1827–1836, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1827-1836.

[46] A. Verma, "Quantitative Evaluation of Sentiment Analysis on E-Commerce Data," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 11, no. 1, pp. 901–907, Apr. 2020, doi: 10.17762/turcomat.v11i1.13588.

[47] A. Irwanto and L. Goeirmanto, "Sentiment Analysis From Twitter About Covid-19 Vaccination in Indonesia Using Naive Bayes and Xgboost Classifier Algorithm," *SINERGI*, vol. 27, no. 2, p. 145, Feb. 2023, doi: 10.22441/sinergi.2023.2.001.

[48] A. Angdresey, I. Y. Kairupan, and K. G. Emor, "Classification and Sentiment Analysis on Tweets of the Ministry of Health Republic of Indonesia," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/ICIC56845.2022.10007008.

[49] M. S. Homaid and I. Moulitsas, "Measuring Airport Service Quality Using Machine Learning Algorithms," in *2022 The 6th International Conference on Advances in Artificial Intelligence*, New York, NY, USA: ACM, Oct. 2022, pp. 8–14. doi: 10.1145/3571560.3571562.

[50] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[51] W. Wang, G. Chakraborty, and B. Chakraborty, "Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm," *Applied Sciences*, vol. 11, no. 1, 2021, doi: 10.3390/app11010202.