JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.7 No.1 January 2025: 79-87
e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.11580

Classifying Viral Indonesian Twitter with Transformer Models and Multi-Layer Perceptron

Jeffrey Junior Tedjasulaksana[1*], Alexander Agung Santoso Gunawan[2]

[1]Computer Science Department, BINUS Graduate Program - Master of Computer Science,
[2]Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
jeffrey.tedjasulaksana@binus.ac.id; aagung@binus.edu

*Correspondence: jeffrey.tedjasulaksana@binus.ac.id

Abstract *– This research explores the classification of virality levels in Indonesian tweets, employing advanced natural language processing (NLP) techniques and machine learning algorithms to predict tweet engagement and influence. The study leverages state-of-the-art transformer models, including RoBERTa for sentiment analysis and XLNet for text embedding, combined with Multi-Layer Perceptron (MLP) classifiers. By incorporating emotion features and implementing cost-sensitive strategies to address class imbalance, the model is good enough to predict tweet virality. The integration of sentiment analysis and emotion detection allows for a deeper understanding of the factors influencing virality, revealing significant correlations between tweet sentiment, and emotion distribution. Our findings show that XLNet outperforms BERTweet in capturing contextual nuances, leading to improved classification performance. Additionally, incorporating emotion-related features and applying cost-sensitive methods further enhances the model's ability to predict tweet virality accurately, achieving an impressive 95% accuracy and an F1-score of 59%. These results offer valuable insights into how emotions and sentiment can influence viral content on social media platforms, providing actionable recommendations for marketers, businesses, and content creators. By optimizing their social media strategies based on these findings, they can better understand audience behavior and engagement trends. The proposed model paves the way for more effective social media analytics, contributing to the growing field of NLP applications for social media content classification.*

Keywords: *Cost-Sensitive; Multi-Layer Perceptron; Twitter; Virality Classification; XLNet*

I. INTRODUCTION

Social media platforms have become ubiquitous channels of communication, information dissemination and public discourse (Aichner et al., 2021), with Twitter standing out as a platform for real-time interaction and content sharing (Karami et al., 2020). Figure 1 shows that Indonesia is the 5th largest social media user in the world with approximately 24 million users (Statista, 2024). The vast amount of user-generated content on Twitter provides an opportunity to explore and analyze various aspects of online behavior, including factors that contribute to the virality of tweets.

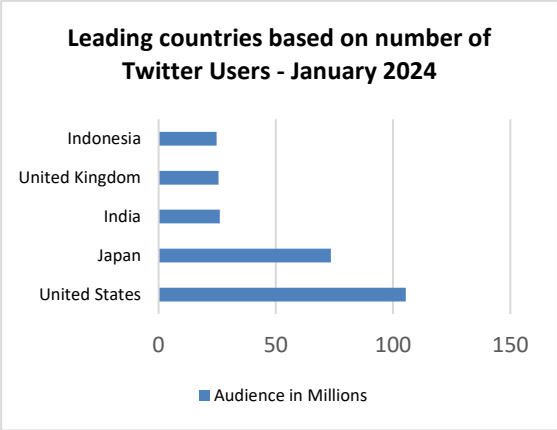

Figure 1. Most Twitter Users in The World

In today's digital marketplace, the ability to create viral content is coveted, yet elusive (Zadeh & Sharda, 2022). Without reliable methods to forecast which tweets will capture the attention of users and spread rapidly, businesses risk investing time and resources into campaigns that may not resonate with their target audience. This uncertainty poses a significant obstacle for marketers seeking to maximize the impact of their messaging and drive engagement with their brand (Vrontis et al., 2021).

Moreover, in the highly competitive landscape of social media marketing, the inability to predict virality can result in missed opportunities to capitalize on trending topics or cultural moments. By harnessing the power of AI to predict virality on Twitter, businesses can gain a competitive edge by identifying the content most likely to resonate with their audience and achieve widespread visibility.

NLP involves a wide range of methods and approaches aimed at teaching computers to understand, interpret, and even generate human-like language. By integrating machine learning algorithms, deep learning models, and linguistic principles, With the explosive growth of digital content and the prevalence of online communication platforms, the importance of NLP in extracting meaningful information from unstructured text data continues to soar (Khurana et al., 2023).

Recent studies employing deep learning models and transformer architectures to assess virality levels include the research conducted by (Rameez et al., 2022). They utilized the RoBERTa method to forecast sentiment using English tweet data, while considering crucial features such as the count of hashtags, mentions, followers, following, verification status, and text length, alongside the sentiment expressed in tweets. Furthermore, they integrated tweet text as an additional feature through BERTweet text embedding. These features were then amalgamated into a Multi-Layer Perceptron (MLP) model for the classification process, utilizing retweet count as a determinant of virality level.

Prior research has neglected to thoroughly investigate certain parameters, such as the incorporation of emotion detection from tweets, as factors in classifying viral levels. According to a study conducted by (Nanath & Joy, 2023), there appears to be a correlation between emotion categories and the levels of virality observed.

In previous research, BERTweet was used to perform text embedding which has the main architecture of BERT. However, one limitation of BERT lies in its approach to token masking during training, which may hinder its ability to capture subtle linguistic dependencies (Devlin et al., 2019). However, XLNet addresses this weakness by adopting permutation language modeling (PLM), which considers all possible word permutations within a sentence. This enables XLNet to capture dependencies more effectively and produce richer contextual representations. Additionally, XLNet's autoregressive formulation allows it to leverage both left and right context during pretraining, further enhancing its ability to understand and generate coherent text. Consequently, XLNet demonstrates superior performance

in tasks requiring nuanced language understanding, making it a more robust choice for text embedding compared to BERT (Li et al., 2020).

In the dataset, imbalanced data poses a challenge, with instances of high virality comprising only 2% and medium virality just 1% of the total data. To address this, this research employs a cost-sensitive approach in handling the imbalance. (Zhu et al., 2017) conducted research to analyze and compare various methods for addressing imbalanced data. Their findings revealed that cost-sensitive approaches tended to enhance results compared to standard classifiers, although the extent of improvement varied. This involves adjusting the learning process to ensure that instances of minority classes receive greater emphasis during model training. By assigning higher weights or penalties to these minority class instances, the aim is to mitigate the effects of class imbalance and improve the model's ability to accurately classify instances across all virality levels.

There are several studies related to virality level classification, one of which is research to predict the popularity of Weibo social media content by utilizing representations of homophily, social groups, and social influence which become input for the Multi-layer perceptron (MLP) classifier to predict the popularity of content (Shang et al., 2022). Other research also classifies the level of virality of Tweet by extracting features from text using BERT and extracting image features using VGG16 whose features will be input for MLP (Amitani et al., 2021). There is also research related to classifying virality levels based on the number of retweets on the topic COVID-19 using a neural network classifier (Qu & Ding, 2020). There is also research that carries out virality classification by utilizing BERT and machine learning in the process (Prasongko et al., 2023).

This research will focus on comparing the baseline model conducted by (Rameez et al., 2022) with the proposed model with the main differences being the addition of emotion features, imbalanced handling methods, and the type of text embedding used. The data used in this research also uses Indonesian Tweets, while the previous research used English Tweets.

II. METHODS

In this section, we delineate our methodology for classifying virality on Twitter, amalgamating advanced natural language processing techniques with machine learning algorithms. Leveraging state-of-the-art transformer models, including RoBERTa for sentiment analysis, BERT for emotion classification, and XLNet for text embedding, we formulate a comprehensive approach to discerning the viral potential of tweets. Our methodology encompasses multiple stages: preprocessing and tokenization of the Twitter data, sentiment analysis to gauge the emotional tone, emotion classification to discern underlying sentiments, and embedding using XLNet to capture contextual information. Subsequently, we integrate these diverse insights into a feature matrix and employ a Multi-Layer Perceptron (MLP) classifier for virality prediction. By delineating our methodology in detail, we aim to provide a robust framework for understanding and predicting the virality of Twitter content. Figure 2 shows the flow of the methodology carried out in this study.

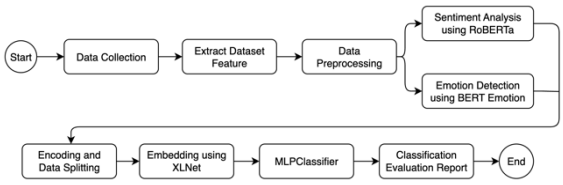

Figure 2. Proposed Methodology

2.1 Data

The dataset we're working with comprises tweet data collected via the Apache Solr server from Twitter in July 2021, totaling 4953 entries written in Indonesian. To fetch the data, we established a connection between Twitter's JDBC (Java Database Connectivity Driver) and the Apache Solr server, facilitating seamless access to the Twitter database. Given the abundance of metadata associated with each tweet, we carefully selected relevant metadata features to drive the classification of tweet virality. These chosen features play pivotal roles in determining the virality level of tweets. For a comprehensive understanding, the specific features of the tweet data utilized in our analysis are outlined in Table 1.

Table 1. Features Details

No	Features	DataType
1.	Topic	object
2.	Tweet	object
3.	Follower	int
4.	Following	int
5.	Like	int
6.	Retweet	int
7.	Reply	int
8.	Verified	object

Our dataset exhibits a significant imbalance in the distribution of viral and non-viral tweets, with a vast majority belonging to the latter category. Such an imbalance poses a challenge for our classification task, as it may lead to biased model predictions favoring the majority class. To address this issue and ensure the model's sensitivity to minority classes, we employ weight initialization techniques during model training. By assigning higher weights to samples from underrepresented classes, we effectively amplify their influence during the learning process. This strategy aims to mitigate the impact of class imbalance and improve the model's ability to accurately classify viral tweets, thereby enhancing its overall performance and reliability in real-world scenarios.

2.2 Data Preprocessing

We start with raw tweet data in JSON format and focus on user-generated text. Our preprocessing aims to structure the data for better model performance. We remove duplicates, compute tweet length, count hashtags and mentions, and discard non-essential elements like account mentions and hashtags. We also standardize text to lowercase, eliminate non-ASCII characters, and convert slang to formal language. Finally, we categorize tweets into three virality classes based on retweet count: low, medium, and high.

2.3 Sentiment Analysis

After preprocessing, we conduct sentiment analysis on the tweet text, which helps classify virality levels. To use RoBERTa for sentiment analysis, we first translate the preprocessed tweets into English, as RoBERTa operates in English and relies on a large data corpus called OpenWebText. Sentiment analysis with RoBERTa involves tokenization, converting text into word-level representations. These tokens are then embedded with position encoding to retain word order and fed into RoBERTa for sentiment classification. The analysis outputs three sentiment classes—positive, negative, and neutral—which are used as additional features in the MLP model.

2.4 Emotion Detection

Emotion detection results are another key feature for classifying virality levels. We use BERT, which boasts a transformer architecture and is pre-trained on vast datasets like BookCorpus and English Wikipedia. BERT's fine-tuned emotion model is capable of predicting emotions, with input consisting of preprocessed tweet text translated into English. The emotion detection process yields six emotion

classes—sadness, joy, love, anger, fear, and surprise.

2.5 Encoding and Data Splitting

For features still in string or categorical formats, we'll use one-hot encoding, except for the virality feature, which will be encoded using ordinal encoding. This encoding process retains the information within these variables while converting them into a format algorithms can process.

We split the data into training, validation, and test sets following a common practice. Specifically, 70% of the data is allocated to the training set, enabling the model to learn patterns from a substantial portion of the dataset. The validation set comprises 10% of the data, serving as a checkpoint during training to fine-tune model parameters and prevent overfitting. Finally, the remaining 20% of the data forms the test set, used to evaluate the model's performance on unseen data and assess its generalization capabilities. This division ensures that the model is trained, validated, and tested on distinct subsets of the data, facilitating robust performance evaluation.

2.6 Text Embedding

We utilize XLNet for tweet embedding, renowned for its profound contextual understanding. XLNet tokenizes the preprocessed tweet text, capturing detailed contextual nuances. These tokens are transformed into numerical representations, preserving their order and relationships. XLNet's advanced architecture effectively captures the tweet's essence, providing rich embeddings. These embeddings, compact yet comprehensive, empower downstream tasks like virality classification with deep contextual insights, enhancing accuracy.

2.6 Classification

In the classification approach, weight initialization is employed to address the inherent imbalance in the data, ensuring fair treatment of minority classes. Tackling imbalanced data classification, cost-sensitive learning emerges as a valuable approach. This technique comes into play when one class dominates the data distribution, posing a significant challenge for machine learning models that prioritize equal treatment of all classes. Cost-sensitive learning introduces varied misclassification costs for different classes, enabling the model to adjust to the skewed data distribution. In cost-sensitive learning, the aim is to reduce the overall expense of misclassification, where the cost incurred for misclassifying instances varies between different classes. This method proves especially potent in handling imbalanced data classification, as it empowers the model to prioritize the minority class, which often holds greater significance in practical scenarios (Zhang et al., 2019).

The loss function chosen for the MLP model is sparse categorical cross entropy, apt for multiclass classification tasks, as it quantifies the disparity between predicted and actual targets. Meanwhile, we utilize the Adaptive Moment Estimation (Adam) optimizer to expedite convergence during training. Adam's adaptive learning from past gradients and ability to escape local minima accelerate convergence, alongside effective bias correction. To optimize model performance, we conduct hyperparameter tuning, systematically testing different combinations on both training and testing data. This process ensures the selection of optimal parameters for our MLP model architecture, as detailed in Table 2. Through these strategies, we aim to enhance classification accuracy and robustness in handling imbalanced data distributions.

The proposed model will undergo hyperparameter tuning using grid search method, where adjustments will be made to the model’s parameters. This process

involves testing various combinations of hyperparameters using both training and validation data. The optimal set of hyperparameters will be selected based on the results of this tuning process. Table 2 provides a detailed overview of the hyperparameters used to design the MLP model architecture in order to achieve the best performance outcomes.

Table 2. Hyperparameters Tuning

Hyperparameter	Hyperparameter Value
Batch sizes	[4, 8, 16]
Learning rates	[0.00001, 0.0001, 0.001, 0.001

2.7 Evaluation

In our evaluation process, we employ the confusion matrix to derive key performance metrics, including accuracy, precision, recall, and F1-score. The confusion matrix provides a comprehensive overview of our model's classification performance by summarizing the counts of true positive, true negative, false positive, and false negative predictions across different classes. From this matrix, we calculate accuracy, representing the proportion of correctly classified instances over the total number of instances. Precision measures the ratio of correctly predicted positive instances to the total predicted positive instances, while recall assesses the proportion of correctly predicted positive instances to the total actual positive instances. F1-score, the harmonic mean of precision and recall, offers a balanced evaluation metric that considers both false positives and false negatives. By leveraging these metrics derived from the confusion matrix, we gain valuable insights into the effectiveness of our classification model in accurately predicting tweet virality levels.

III. RESULTS AND DISCUSSION

Sentiment prediction carried out in this study using RoBERTa with the results of sentiment distribution across virality level can be seen in Figure 3.

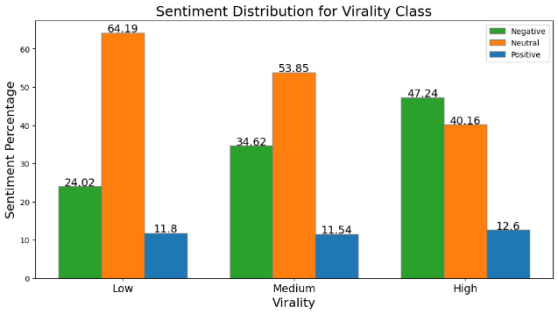

Figure 3 Sentiment Distribution

The findings revealed tweets with negative sentiment exhibited a higher probability of achieving high virality with a percentage of negative sentiment of more than 47% of tweets that have high virality. This intriguing observation suggests a potential correlation between the emotional tone of tweets and their virality. Further investigation into the underlying reasons behind this phenomenon could provide valuable insights into the dynamics of online engagement and information dissemination.

The hyperparameters that are tuned in this study are batch sizes of 4, 8, and 16 and also learning rates which are 0.00001, 0.0001, 0.001, and 0.001. Hyperparameters are tuned using validation data. The result of hyperparameter tuning produces a batch size value of 16 and learning rate 0.0001, it is because model lead to more stable training dynamics and smoother convergence during the optimization process. Table 3 presents the hyperparameters selected for the MLP model, based on the outcomes of the grid search.

Table 3. MLP selected hyperparameters

Hyperparameter	Hyperparameter Value
Batch sizes	16
Learning rates	0.0001

In this research, the model was also tested using testing data and the results were compared with the baseline model. The results of the test comparison can be seen in Table 4.

Table 4 Classification Result

Method	Precision	Recall	F1-Score	Accuracy
Baseline (Rameez et al., 2022)	0.49	0.35	0.37	0.35
Proposed Model (No Cos-Sent + no Emotion)	0.55	0.52	0.54	0.97
Proposed Model (Cos-Sent + no emotion)	0.52	0.61	0.55	0.95
Proposed Model (no Cos-Sent + emotion)	0.66	0.5	0.55	0.98
Proposed Model (Cos-Sent + emotion)	0.59	0.65	0.59	0.95

The proposed model gets better performance results than the baseline with the main differentiators being the imbalance handling method, text embedding, the addition of emotion features, and the language of the data used.

XLNet outperforms BERTweet due to its compatibility with formal English language, which is essential for analyzing Indonesian language data translated into English. Given that the data in this study involves Indonesian language content translated into English, a model pre-trained using formal English is more suitable for capturing the nuances and subtleties of the language. XLNet's pre-training on formal English text ensures that it is adept at understanding and processing such linguistic nuances, resulting in superior performance compared to BERTweet.

The great performance of the model in this research can be attributed to two key factors which are the implementation of cost-sensitive methods for handling class imbalance and the incorporation of emotion features, both tailored to the nuances of Indonesian tweets. The application of cost-sensitive techniques effectively addresses the class imbalance present in the dataset, where certain classes are severely underrepresented. By assigning different weights to instances based on their class distribution, the model is able to give greater emphasis to the minority classes—specifically the medium and high labels, which constitute less than 5% of the data. This approach mitigates the effects of imbalance, thereby enhancing the overall classification performance. As demonstrated in Table 4, the models incorporating cost-sensitive strategies exhibit notable improvements in F1-Score across all categories, compared to those that do not utilize such techniques. This results in higher overall accuracy, as the model performs better at classifying majority class instances, while the performance for minority class instances remains relatively weaker. Furthermore, the addition of emotion features enriches the semantic understanding of the text, providing the model with valuable contextual information about the sentiment and emotional tone of the tweets.

IV. CONCLUSION

The proposed model in this study has produced satisfactory performance with a precision value of 0.59, recall 0.65, and F1-Score 059, and accuracy 0.95. The model is good enough to classify the virality level of Indonesian twitter data when compared to the baseline. The model's accuracy results are very good. However, the precision, recall, and F1-Score remain relatively low. This can be attributed to the imbalance in the dataset, particularly with respect to the underrepresentation of data corresponding to medium and high virality levels. Research development that can be done is to increase the amount of data that is minority in the high and medium classes, then also develop the model by testing using ensemble learning.

REFERENCES

Aichner, T., Grünfelder, M., Maurer, O., & Jegeni, D. (2021). Twenty-Five Years of Social Media: A Review of Social Media Applications and Definitions from 1994 to 2019. In *Cyberpsychology, Behavior, and Social Networking* (Vol. 24, Issue 4, pp. 215–222). Mary Ann Liebert Inc. https://doi.org/10.1089/cyber.2020.0134

Amitani, R., Matsumoto, K., Yoshida, M., & Kita, K. (2021). Buzz tweet classification based on text and image features of tweets using multi-task learning. *Applied Sciences (Switzerland)*, *11*(22). https://doi.org/10.3390/app112210567

Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*, 4171–4186. https://github.com/tensorflow/tensor2tensor

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and Research: A Systematic Literature Review through Text Mining. *IEEE Access*, *8*, 67698–67717. https://doi.org/10.1109/ACCESS.2020.2983656

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Li, H., Choi, J., Lee, S., & Ho Ahn, J. (2020). Comparing BERT and XLNet from the Perspective of Computational Characteristics. *2020 International Conference on Electronics, Information, and Communication (ICEIC)*.

Nanath, K., & Joy, G. (2023). Leveraging Twitter data to analyze the virality of Covid-19 tweets: a text mining approach. *Behaviour and Information Technology*, *42*(2), 196–214. https://doi.org/10.1080/0144929X.2021.1941259

Prasongko, Y., Girsang, A., Yayogi, A., & Prasetyo, D. (2023). Prediction of Crime Virality by Indonesia National Police on Social Media. *Media Bina Ilmiah*, *17*.

Qu, Z., & Ding, Z. (2020). Predicting the retweet level of covid-19 tweets with neural network classifier. *Proceedings of 2020 IEEE 19th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2020*, 15–20. https://doi.org/10.1109/ICCICC50026.2020.9450271

Rameez, R., Rahmani, H. A., & Yilmaz, E. (2022). ViralBERT: A User Focused BERT-Based Approach to Virality Prediction. *UMAP2022 - Adjunct Proceedings of the 30th ACM Conference on User Modeling,*

Adaptation and Personalization, 85–89. https://doi.org/10.1145/3511047.3536415

Shang, Y., Zhou, B., Zeng, X., Wang, Y., Yu, H., & Zhang, Z. (2022). Predicting the Popularity of Online Content by Modeling the Social Influence and Homophily Features. *Frontiers in Physics*, *10*. https://doi.org/10.3389/fphy.2022.915756

Statista. (2024, April). *Leading countries based on number of X (formerly Twitter) users as of April 2024*. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Vrontis, D., Makrides, A., Christofi, M., & Thrassou, A. (2021). Social media influencer marketing: A systematic review, integrative framework and future research agenda. *International Journal of Consumer Studies*, *45*(4), 617–644. https://doi.org/10.1111/ijcs.12647

Zadeh, A., & Sharda, R. (2022). How Can Our Tweets Go Viral? Point-Process Modelling of Brand Content. *Information & Management*, *59*(2), 103594. https://doi.org/10.1016/j.im.2022.103594

Zhang, C., Tan, K. C., Li, H., & Hong, G. S. (2019). A Cost-Sensitive Deep Belief Network for Imbalanced Classification. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(1), 109–122. https://doi.org/10.1109/TNNLS.2018.2832648

Zhu, B., Baesens, B., & vanden Broucke, S. K. L. M. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, *408*, 84–99. https://doi.org/10.1016/j.ins.2017.04.015