PREDICTION OF CRIME VIRALITY BY INDONESIA NATIONAL POLICE ON SOCIAL MEDIA

By
Yudhi Prasongko[1], Abba Suganda Girsang[2], Andry Yayogi[3], Deny Prasetyo[4]
[1,2,3,4,5]Computer Science Department, BINUS Graduate Program - Master of Computer Science Bina Nusantara University Jakarta, Indonesia 11480
Email: [1]yudhi.prasongko@binus.ac.id

Abstract

Twitter is the most generally involved online entertainment stage for sharing data. Furthermore, Twitter is quite possibly of the biggest social medium in Indonesia. The vast majority of the data or news that is distributed comes from issues or occasions that are creating in the public eye. The issue of police officers committing crimes is one that quickly becomes viral. This examination is a continuation of past exploration to foresee virality in view of Twitter information. Predicting and classifying text in the form of tweets and posts (socmed Twitter) using BERT. IndoBERT as a safeguarded model in the characterization of kinds of wrongdoing. ViralBERT to foresee tweet virality utilizing text includes and adding numeric highlights to enhance virality forecasts. Using the baseline from this study, the data set was used to train ViralBert and validate model results. The dataset was gathered utilizing the Twitter Programming interface with the watchwords 'oknum polisi' and oknum polri'. The consequences of this virality expectation will be utilized as a kind of perspective by the police in following up a case that is viral locally**.**

***Keywords*:Bert, Twitter, ViralBert, Vira, IndoBERT, Crime**

INTRODUCTION

It has become a necessity to keep up with the most recent developments in the news. News passed on to the public should be authentic and quick. The most recent reality is the news that is distributed in all media channels, its majority comes from issues or occasions that are creating in the public eye. Twitter virtual entertainment stage is the greatest stage, particularly in Indonesia. Videos and images can be followed by up to 280 characters of text written by users. 500 million tweets are posted consistently on twitter. Clients can speak with others, similar to, share tweets and remarks on a post. Through Twitter online entertainment, data can be spread rapidly and hugely. This demonstrates that Twitter significantly influences the spread of a problem. Social media will play a significant role as a source of information, ideas, and inspiration in the future. [1]. Predicting the popularity of tweets is crucial because there are numerous studies on how viral posts can influence political outcomes, social views, and economics [2]. As of late, issues or occasions that have frequently turned into the worry of the general population are violations dedicated by state executives, including the police. Issues connected with police wrongdoings become a web sensation rapidly and adversely affect public confidence in the police. Knowing how likely it is for negative news to spread throughout society necessitates predicting its virality. With the goal that the police can quickly do whatever it takes to authorize the law against cops who perpetrate wrongdoings and can lessen negative virality connected with the police. Tweets connected with cops are typically made straight by Twitter clients who are casualties of wrongdoings by thesecops.

By recording from the casualty's cellphone and making it viral through a Twitter account. This is likewise upheld by referencing powerhouse accounts that have countless supporters and are dynamic on Twitter to create ideal commitment.

Notwithstanding a desire to bettering the lives of ordinary In Jakarta, the situation is regrettably very paradoxical. The ascent to mass organization dominance in the majority of the states coincided with a record-high level of violence against public. The numerous attacks, murders committed in, and rioting in certain locations, all of which were invariably sensitive communal areas, were the worst effects of the recent saffron group. Due to this persistent systemic violence, minority populations are marginalized in numerous ways structurally. It went on in a well-organized manner by examining the historical background of the persecution of mass organization[3]. To develop an ensemble forecast system based on machine learning methods to predict the number of crimes in Indonesia are important. [4]

Anticipating virality on the Twitter stage has its own difficulties. Since it is impacted by many variables, some can't be estimated, for example, the innovative substance of content, the pertinence of a socio-political environment, and the connection between clients. The vast majority of the examinations on tweet commitment center around the Twitter network structure model and how tweets spread to different clients [5]. In this review, we fabricated an expectation model for the virality of Twitter virtual entertainment posts, utilizing a dataset connected with the issue of violations committed by cops of 133,023 tweets. We utilize a profound learning model for this expectation issue involving a similar engineering as in a past report called ViralBERT. By modifying a tweet and changing the sentiment feature to match the type of crime, we can create a predictive model that can use textual features, content-based features (hashtag, mention, text-length), and user-based features (followers, following, verified). furthermore, added another component, specifically the position of the culprit of the wrongdoing. This concentrate likewise explores whether the highlights of the kind of wrongdoing and the position of the culprit have a relationship with the degree of virality.

LITERATURE REVIEW

Related Work

There has been a ton of exploration on pattern investigation and displaying client communication on Twitter. Lymperopoulos, (2021), Thakur et al., (2022) by following the development in the quantity of retweets and distinguishing examples of virality for all tweets - beginning with quick development and arriving at viral over the long run [7], [8]. Investigate multiple ways of demonstrating Twitter web-based entertainment clients. Pagerank contrasted with adherents and retweets, it was found that pagerank is practically equivalent to devotees yet having huge supporters doesn't ensure that posts from that record will get high retweets. Retweets are mean a lot to expand how much openness. There is likewise research connected with the tweet model to foresee whether a post will be viral or famous. With a client based way to deal with virality forecast as the primary reference in this review ViralBERT [9], [10]

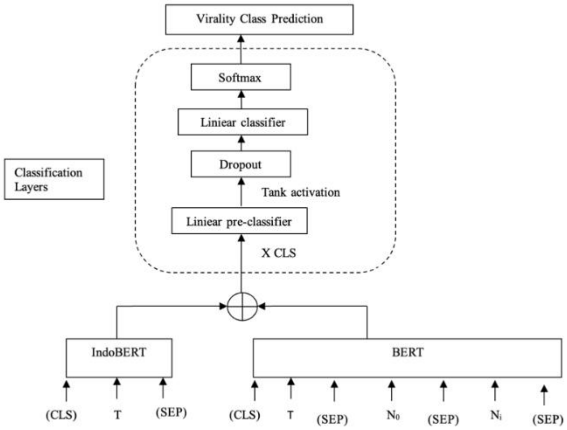

Fig. 1. ViralBERT Architecture

3. Related Work

A methodology for discovering behavior rules, correlations, and patterns of information attending large-scale public events by analyzing social media posts[11].In this section, we explain the method we propose to predict virality, the same method in previous research, namely ViralBERT, using text features and numerical features. Pointed out the growing significance of how information is arranged and used for decision-making[12]with the model shown in figure 2.

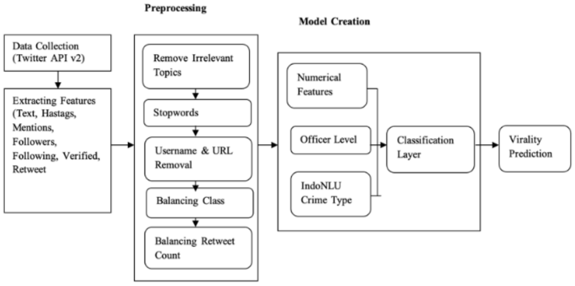

Fig. 2. Research Model

BERT

Posts on Twitter that are often found using informal language and abbreviated words (hashtags, abbreviations, regional names, an event) are a challenge for us in choosing a model. The model we use is BERT[13]. There are two steps in the BERT model, namely pre-training and fine-tuning. we use BERT because the model can represent a text that does not have a label in depth by adjusting both contexts (left context and right context) in all layers[14],[15]. BERT is based on 12 transformer blocks, 768 hidden units, 12 self attention with resulting 768 embeddings. BERT is a multi-layer transformer encoder based on the original implementation described.However, the pre-train model of BERT is only limited to use in English. So in this study, IndoBERT will be used in making news recommendations based on the category. This dataset uses an Indonesian news dataset that has 5 categories, including football, news, business, technology, and automotive[16].

IndoBERT

Models such as BERT are widely used in research recently [17]because they have good performance, but these models focus on training in English and not those that focus on limited resources such as Indonesian. Therefore, this study uses IndoBERT to leverage Indonesian-focused resources in a transformation-based model. For the classification of types of crime we use IndoBERT as a pre-trained model. This is a pre-trained model that is trained with about 4 billion word corpus (Indo4B), more than 20 GB of text data. The output of this model is 7 classes of types of crime, namely: drugs, immorality, abuse, corruption, murder, abuse of power, and error in persona.

Table 1 Type of Crime

Class	Keyword
Narcotics	Drugs, marijuana
immorality	Obscenity, adultery
Persecution	beatings, acts of violence
Corruption	Bribery, extortion, nepotism
Murder	Shootings, kill
Abuse of authority	Arrogance, falsification, obstructing investigations. eliminate evidence
Error in Persona	Wrong catch, wrong shot

Analysis of the type of crime is used as an additional ViralBERT task which is an additional numerical input for the final result of the virality classification layer and is a fine-tuned task when carrying out the training model.

ViralBERT

The quality of this information becomes a crucial issue(Torres et al., 2023). BERT is used to predict virality using layer classification - using fine-tuning provides ease of implementation, faster training process and increased performance when we build a pre-trained model [19]. The classification layer used connects the MLP structure with tanh activation – where the dimensions in this layer are based on XCLS, so there are 771 hidden units. Then a dropout value of 0.1 is implemented before the last linear prediction layer, giving rise to a suspicion of ownership of the virality class after being entered into the softmax function. Experiments were carried out to increase the number of layers to classify and use CNN instead of ANN, but the performance also did not improve.

RESEARCH METHOD

Data Collection

Twitter has beeninstrumental in studying human behavior with social me-dia data and the entire field of Computational Social Science (CSS) has heavily relied on data from Twitter[20].Society's reliance on social media for information is enormous, unlike conventional news sources. The volume of data accessed daily led to the adoption of natural language processing (NLP) for text analytics [21]. The dataset used in this study was collected using the Twitter API version 2 using Python. Twitter's API makes it possible to search for matching tweets using keywords. In this study, the keywords 'police personnel' and 'police personnel' were used to find tweets that were relevant to the case study. The fields of a tweet that are collected include: text, post date, number of hashtags, number of mentions, while the user information used includes: verified status, number of followers, number of following. Meanwhile, the fields for the number of retweets, the number of likes and the number of quotes are only collected as a supporting dataset, but the number of retweets will be used to determine the level of virality. The dataset collected spanned from March 2021 to March 2023, a total of 133,023 tweets related to 'police personnel' were collected.

Preprocessing

The collected tweet data is first removed by tweets that have duplicate text, tweet texts whose contents post police performance in suppressing elements (not posts that want to spread news/crime incidents by elements).We leveraged Twitter's trending topics and hashtags to identify and curate a list of keywords[22]

Step Preprocessing:

1. Change the characters in the tweet text.
   - From : RT @BeritaSubang_PR: Mine in South Sumatra is Above The Law , IPW: Check The Police Officer https://t.co/Bwc4UwbwnF
   - To : RT [USER]: Mine in South Sumatra is Above The Law , IPW: Check The Police Officer [URL]
2. Balancing rank datasets.
3. Balancing the crime type dataset.
4. Balancing number of retweet = 0.

Data Labelling

Four virality classes were used: 0 retweets (1.907), 1 retweets (511), 2-20 retweets (1.000) dan 21+ retweets (423). 4 classes divided into:

Table 2 Rank Class

Police General
Police Commissioner General
Police Inspector General
Police Brigadier General
Middle-rank officers
Police Senior Superintendent
Police Superintendent
Police Assistant Superintendent
Inspector
Police senior inspector
Police first inspector
Police second inspector
Non-Commissined Officer
Police Assistant First Inspector
Police Assistant Second Inspector
Police Senior Brigadier
Police Brigadier

Police First Brigadier
Police Second Brigadier
Private
Police Senior Assistant Brigadier
Police Assistant 1st Brigadier
Police Assistant 2nd Brigadier
Police Senior Private
Police First Private
Police Second Private

Baseline Model

In this study we compare the performance of the modified ViralBERT with several other baseline models.

- Logistic Regression: Utilises Newton's method for gradient optimisation. This method has been used for predicting popular messages in Twitter[23].
- Support Vector Machine (SVM): Uses hinge loss and SGD optimization. This method used for predicting the popularity of newly emerging hashtags in Twitter, as well as for assessing the retweet proneness[24].
- Decision Tree : Uses Gini impurity score to measure quality of a split with no max depth. This method has also been used for assessing retweet proneness[25].
- Random Forest: Uses 100 trees with no max depths. This baseline is based on previous work, which focused on the problem of temporal prediction of retweet count and likelihood of a retweet [26].

Table 3 Evaluation Metric On Models

Method	F1 Score	Precision	Recall	Accuracy
Logistic Regression	0.274	0.271	0.208	0.271
KNN	0.349	0.328	0.322	0.328
SVM	0.391	0.278	0.219	0.278
Decision Tree	0.396	0.396	0.395	0.396
Random forest	0.420	0.407	0.405	0.407
Naive Bayes	0.320	0.279	0.299	0.279
MLP Num	0.187	0.247	0.135	0.246
ViralBERT Text	0.296	0.352	0.308	0.352
ViralBERT	0.53	0.513	0.539	0.539

Evaluation

There is no explicit theoretical support that the clicked regions can be properly activated and correctly classified [27]. To improve the ability of language models to handle Natural Language Processing (NLP) tasks and intermediate step of pre-training has recently been introduced [28]. In this setup, one takes a pre-trained language model, trains it on a (set of) NLP dataset(s), and then finetunes it for a target task. It is known that the selection of relevant transfer tasks is important, but recently some work has shown substantial performance gains by doing intermediate training on a very large set of datasets. In general, detecting AI-generated text using machine learning concerns two types: black-box and white-box detection. Blackbox detection relies on API-level access to language models, limiting its capability to detect synthetic texts [29]. Although some researchers have proposed different strategies to evaluate the effectiveness of our classifiers we employ the following metrics on a holdout test set after training: accuracy, which is the ratio of correct predictions over total number of instances evaluated; recall, which is the fraction of actual classifications for a class that were correctly identified; precision, which is the proportion of model predictions for a class that were actually correct; F1 Score, which is the harmonic mean of precision and recall. As accuracy and F1 scores consider both the positive and negative classifications for evaluation, these are ideal for discriminating our optimal solution. These metrics are balanced using the macro weighting, so there is an equal weighting and importance for each class allowing metrics to be more indicative of predictions of all classes.

RESULT AND DISCUSSION

Table 3 shows the results of several architectural models used as a comparison (baseline model). when only using ViralBERT with text features it does not give optimal results when compared to the baseline model (lower than Random Forest). Where if adding numeric features to the ViralBERT model can improve performance quite significantly when compared to only using text features, this proves that these numerical features are also very important for predicting virality. By combining input in the form of numeric features and text features into the classification layer, we can achieve a higher level of model performance compared to the existing baseline model. This means that by adding its numeric features to the prediction feature, it allows ViralBERT to get additional context that can be used as input to the model and can display can provide optimal results when compared to the baseline model.

We also conduct experiments by measuring whether other features play a significant role in the model, by removing those numeric features and numeric teks from the model input, and calculating the performance of the model as well. Table 4 shows the results of removingfeatures from the input model when compared to the model produced in this study. If we remove the following feature or the type of crime or the number of hashtags or the number of mentions, it can significantly reduce the accuracy of the model when compared to other features. This shows that the type of crime to predict the virality of crime issues by police officers has a very important role in model formation.

Table 4 Performance After Removing Features of ViralBERT

Feature removed	F1 Score	Accuracy
ViralBERT	0.53	0.539
Crime Type	0.393	0.436
Officer Level	0.509	0.51
Hashtags	0.433	0.45
Mensions	0.455	0.455
Followers	0.444	0.455
Following	0.424	0.454
Verified	0.489	0.509
Text Length	0.489	0.455

What is surprising, is that the rank of the perpetrator gives better performance results if the feature is removed from the input model, this can indicate that for a tweet related to a police officer it can go viral regardless of the rank of the perpetrator, or you could say whatever the rank of the perpetrator a tweet is related to a person. policing can go viral. Contrary to previous research, the number of hashtags and the number of mentions are very important in determining the virality of tweets related to the issue of crimes committed by police officers.

CONCLUSION

This study investigates whether the features of the type of crime and the rank of the perpetrator have a correlation with the level of virality with the results of an accuracy of the F1 model score of 0.53 and an accuracy of 0.539. It was found that the features of the type of crime are important in predicting virality. Suggestions for further research are to further develop the model used, taking into account the network of friends, such as whether a tweet made by a user follows or is followed or another dimensioned account that has a large follower base or has a high level of engagement. Or tweets made using hashtags that have a high engagement value.

REFERENCES

[1] W. Stassen, “Your news in 140 characters: exploring the role of social media in journalism,” *Glob. Media J. African Ed.*, vol. 4, no. 1, pp. 116–131, 2011, doi: 10.5789/4-1-15.

[2] M. T. Borges-Tiago, F. Tiago, and C. Cosme, “Exploring users’ motivations to participate in viral communication on social media,” *J. Bus. Res.*, vol. 101, no. August, pp. 574–582, 2019, doi: 10.1016/j.jbusres.2018.11.011.

[3] M. Afroz, “Rise of Violence against Minorities, Fragile governance or Structured Marginalization?,” *Res. Rev. Int. J. Multidiscip.*, vol. 4, no. 1, 2019, [Online]. Available: www.rrjournals.com

[4] V. Ivanyuk, “Forecasting of digital financial crimes in Russia based on machine learning methods,” *J. Comput. Virol. Hacking Tech.*, vol. 1, no. 1, pp. 1–14, 2023, doi: 10.1007/s11416-023-00480-3.

[5] I. N. Lymperopoulos, “RC-Tweet: Modeling and predicting the popularity of tweets through the dynamics of a capacitor,” *Expert Syst. Appl.*, vol. 163, 2021, doi: 10.1016/j.eswa.2020.113785.

[6] A. K. Thakur *et al.*, “Multimodal and Explainable Internet Meme Classification,” *arXiv*, vol. 2, no. 2, pp. 23–32, 2022, [Online]. Available: http://arxiv.org/abs/2212.05612

[7] A. Al-Hashedi *et al.*, “Ensemble Classifiers for Arabic Sentiment Analysis of Social Network (Twitter Data) towards COVID-19-Related Conspiracy Theories,” *Appl. Comput. Intell. Soft Comput.*, vol. 2022, 2022, doi: 10.1155/2022/6614730.

[8] R. Kora and A. Mohammed, “An enhanced approach for sentiment analysis based on meta-ensemble deep learning,” *Soc. Netw. Anal. Min.*, vol. 13, no. 1, pp. 37–52, 2023, doi: 10.1007/s13278-023-01043-6.

[9] T. Elmas, S. Stephane, and C. Houssiaux, *Measuring and Detecting Virality on Social Media: The Case of Twitter’s Viral Tweets Topic*, vol. 1, no. 1. Association for Computing Machinery, 2023. doi: 10.1145/3543873.3587373.

[10] M. Benson, “Predicting Virality of Online News Articles using Textual Content,” *EECS*, vol. 1, no. 1, pp. 1–8, 2020.

[11] E. Cesario, P. Lindia, and A. Vinci, “Detecting Multi-Density Urban Hotspots in a Smart City: Approaches, Challenges and Applications,” *Big Data Cogn. Comput.*, vol. 7, no. 1, pp. 1–18, 2023, doi: 10.3390/bdcc7010029.

[12] D. Darwish, *Big Data Issues*. Egypt: Ahram Canadian University, 2023. doi: 10.4018/978-1-6684-7366-5.ch020.

[13] D. Q. Nguyen, T. Vu, and A. Tuan Nguyen, “BERTweet: A pre-trained language model for English Tweets,” in *Proceedings of the 2020 EMNLP (Systems Demonstrations)*, 2020, pp. 9–14. doi: 10.18653/v1/2020.emnlp-demos.2.

[14] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[15] M. I. Sinapoy, K. Sibaroni, Y. Prasetyowati, and S. Suryani, “Comparison of LSTM and IndoBERT Method,” *J. RESTI*, vol. 5, no. 158, pp. 2–6, 2023.

[16] B. Juarto and Yulianto, “Indonesian News Classification Using IndoBert,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 2, pp. 454–460, 2023.

[17] J. Mutinda, W. Mwangi, and G. Okeyo, “Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network,” *Appl. Sci.*, vol. 13, no. 3, 2023, doi: 10.3390/app13031445.

[18] Sepúlveda-Torres *et al.*, “Leveraging relevant summarized information and multi-layer classification to generalize the detection of misleading headlines,” *Data Knowl. Eng.*, vol. 145, no. December 2022, p. 102176, 2023, doi: 10.1016/j.datak.2023.102176.

[19] L. Huang, C. Sun, X. Qiu, and X. Huang, “Glossbert: BERT for word sense disambiguation with gloss knowledge,” in *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural*

Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 2019, pp. 3509–3514. doi: 10.18653/v1/d19-1355.

[20] J. Pfeffer *et al.*, "Just Another Day on Twitter: A Complete 24 Hours of Twitter Data," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2023, vol. 17, no. Icwsm, pp. 1073–1081. doi: 10.1609/icwsm.v17i1.22215.

[21] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, no. January, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.

[22] E. Chen and E. Ferrara, "Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War between Ukraine and Russia," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 17, no. ICWSM, pp. 1006–1013, 2023, doi: 10.1609/icwsm.v17i1.22208.

[23] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in Twitter," *Proc. 20th Int. Conf. Companion World Wide Web, WWW 2011*, pp. 57–58, 2011, doi: 10.1145/1963192.1963222.

[24] Z. Ma, A. Sun, and G. Cong, "On Predicting the Popularity of Newly Emerging Hashtags in Twitter," *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. July, pp. 1852–1863, 2013, doi: 10.1002/asi.

[25] P. Nesi, G. Pantaleo, I. Paoli, and I. Zaza, "Assessing the reTweet proneness of tweets: predictive models for retweeting," *Multimed. Tools Appl.*, vol. 77, no. 20, pp. 26371–26396, 2018, doi: 10.1007/s11042-018-5865-0.

[26] S. Sharma and V. Gupta, "Role of twitter user profile features in retweet prediction for big data streams," *Multimed. Tools Appl.*, vol. 81, no. 19, pp. 27309–27338, 2022, doi: 10.1007/s11042-022-12815-1.

[27] M. Zhou *et al.*, "Interactive Segmentation as Gaussian Process Classification," *Comput. Vis. Found.*, vol. 1, no. 2, pp. 19488–19497, 2023, [Online]. Available: http://arxiv.org/abs/2302.14578

[28] R. Van Der Goot, "Effectiveness of Intermediate Training on an Uncurated Collection of," in *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, 2023, pp. 230–245.

[29] A. Pegoraro, K. Kumari, H. Fereidooni, and A.-R. Sadeghi, "To ChatGPT, or not to ChatGPT: That is the question!," *arXiv*, vol. 2, no. April, pp. 1–6, 2023, [Online]. Available: http://arxiv.org/abs/2304.01487

[30] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," *A WHITE Pap.*, vol. 14, no. August, pp. 1–17, 2020, [Online]. Available: http://arxiv.org/abs/2008.05756