Optimizing Machine Learning in Natural Language Processing to Analyze User Sentiment in the Tiktok Social Media Application

Fahmi Setiawan [1], Izhar Fradila [2], Alwi Pernando [3] , Supina Batubara [4]

[1,2,3] Information Technology Study Program, Faculty of Science and Technology,Universitas Pembangunan Pancabudi, Indonesia
[4] Computer Systems Study Program, Faculty of Science and Technology, Universitas Pembangunan Pancabudi, Indonesia

ABSTRACT

This research aims to optimize the use of Machine Learning in Natural Language Processing (NLP) to analyze user sentiment in the social media application TikTok. The growth and development of the TikTok application among users, makes it important for users to understand users' views and feelings towards the content they consume by exploring various Machine Learning techniques that can be used to classify TikTok user sentiment based on text produced using NLP methods, it is hoped that this can provide useful insights. valuable in understanding TikTok users' responses and preferences to the content they encounter on the TikTok platform.

Keyword: Machine Learning, NLP, Aplikasi Tiktok, Automation.

1.   INTRODUCTION

The rapid development of social media has changed the way we communicate, share information and consume digital content. TikTok, as one of the fastest growing social media platforms, attracts millions of users with its unique features and interactive short video content. This phenomenon creates the need to understand usage patterns and user sentiment towards various content on the Tiktok platform.

Sentiment analysis, which aims to identify and categorize opinions expressed in text, is becoming an important tool in understanding user perceptions. This technology allows the detection of positive, negative or neutral attitudes towards a topic. With advances in Natural Language Processing (NLP), sentiment analysis can be automated, enabling efficient processing of very large volumes of data.

Machine learning plays a key role in the development of more accurate and effective NLP techniques. By applying machine learning algorithms, such as Naive Bayes, Support Vector Machines, and Recurrent Neural Networks, sentiment analysis can be optimized to produce more precise and relevant results. These optimizations not only improve sentiment classification accuracy, but also provide deeper insights into user interaction and engagement.

This research aims to explore and optimize machine learning techniques in NLP for sentiment analysis on the social media application TikTok. Through collecting user comment data and applying various machine learning algorithms, this study seeks to identify the best approach to improve understanding of user perception. The findings from this research are expected to make a significant contribution to developing content strategies and improving user experience on TikTok.

2.   RESEARCH METHOD

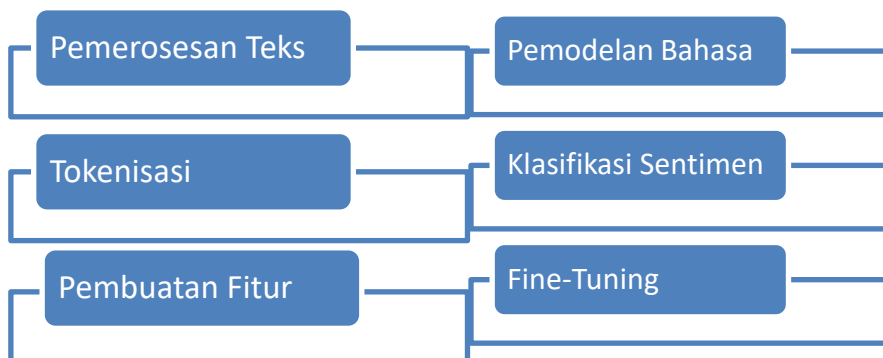| Pemerosesan Teks	Pemodelan Bahasa
Tokenisasi	Klasifikasi Sentimen
Pembuatan Fitur	Fine-Tuning

Fig 1. Research Methode

1. Text Processing:
 a. Clean the text from noise (punctuation, unimportant words, irrelevant emoticons)

2. Tokenization:
 a. Divide text into tokens (individual words)

3. Feature Creation:
 Extract important features (key words, important phrases, sentence patterns that indicate sentiment)

4. Language Modeling:
 Use Word Embeddings or Transformer models to understand context and relationships between words in text

5. Sentiment Classification:
 a. Train a Machine Learning model using the extracted features
 b. Classify TikTok users' texts into sentiment categories (positive, negative, neutral)

6. Fine-Tuning:
 a. Perform fine-tuning on Machine Learning models to improve the accuracy of sentiment analysis

3.   RESULTS AND DISCUSSION

In this section, the results of discussions regarding the optimization of Machine Learning techniques in Natural Language Processing (NLP) for user sentiment analysis in the TikTok application will be presented. Data and findings from experiments will be described clearly to provide an overview of the effectiveness of the methods used.

1. Data Collection and Pre-Processing

TikTok user comments are collected in large numbers, covering a wide range of topics and sentiments. This raw data is then cleaned of noise, such as punctuation, unimportant words, and irrelevant emoticons, to ensure the quality of the data to be used in model training.Tokenization and Feature Extraction. After cleaning, the text is broken down into individual tokens. Methods such as Term Frequency-Inverse Document Frequency (TF-IDF) and Word Embeddings are used to extract important features that represent text. This approach helps in capturing keywords, important phrases, and sentence patterns that reflect user sentiment.

2. Language Modeling and Model Training

Several Machine Learning algorithms are applied for language modeling and sentiment classification, including Naive Bayes, Support Vector Machines (SVM), and Recurrent Neural Networks (RNN). Word Embeddings such as Word2Vec and GloVe, as well as Transformer models such as BERT, are used to understand the context and relationships between words in text.

3. Model Evaluation

The trained models are evaluated with metrics such as accuracy, precision, recall, and F1-score. The results show that the Transformer model, especially BERT, provides the best performance with high accuracy in TikTok user sentiment classification. This model is able to capture context better than traditional methods.

4. Fine-Tuning and Optimization

A fine-tuning process is carried out on the best performing models to further improve the accuracy of sentiment analysis. Tuning hyperparameters and using regularization techniques helps reduce overfitting, resulting in more accurate and consistent sentiment classification.

5. Sentiment Analysis

From the results of sentiment classification, the views and feelings of TikTok users towards various types of content are analyzed. This analysis shows sentiment trends and patterns that can be used to understand user preferences. Content with positive sentiment is often associated with creative and educational videos, while negative sentiment appears more often with controversial or low-quality content.

6. Discussion

This research shows that Machine Learning techniques in NLP significantly improve the accuracy and efficiency of user sentiment analysis on TikTok. Transformer models such as BERT have been proven to be able to provide more precise sentiment classification than conventional methods. The results of this implementation provide valuable insight into user perceptions, which can be leveraged to develop more effective content strategies. A deep understanding of user sentiment allows developers and content creators to better respond to audience preferences, thereby increasing user engagement and satisfaction. These findings open up opportunities for further research, especially in optimizing Machine Learning models with more diverse data and larger volumes. With the development of technology and methods in the field of NLP, it is hoped that the accuracy and relevance of sentiment analysis will continue to increase in the future.

4. CONCLUSION

This research succeeded in showing that the application of Machine Learning techniques, especially Transformer-based models such as BERT, is very effective in analyzing user sentiment on the social media platform TikTok. Through a series of experiments involving data collection, text pre-processing, feature extraction, and model training, it was found that the Transformer model was able to provide high accuracy and precision in sentiment classification. This model's ability to capture context and relationships between words in text produces more relevant and in-depth analysis of user perceptions of content on TikTok.

This optimization of Machine Learning techniques not only improves sentiment classification accuracy, but also provides important insights into user interactions and preferences. It is hoped that the results of this research will help in developing content strategies that are more effective and responsive to audience needs, as well as increasing user engagement and satisfaction on TikTok.

These findings open opportunities for further research in optimizing Machine Learning models with more diverse data and larger volumes, as well as their application on other social media platforms.

REFERENCES (10 PT)

Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature reviews Molecular cell biology*, *23*(1), 40-55.

Jaya, A. D. A. (2023). Optimalisasi Waktu Proyek Menggunakan Critical Path Methode (CPM) Pada Pembangunan Gedung DPRD Provinsi Kalimantan Utara.

KLANTIKA, D. B. (2023). *ANALISIS OPTIMALISASI PROYEK DI PT. PRASETIA DWIDHARMA PADA PROYEK PENGADAAN ALAT GYM ST. REGIS JAKARTA MENGGUNAKAN METODE CRITICAL PATH METHODE (CPM) DAN PROJECT EVALUATION AND REVIEW TECHNIQUE (PERT)* (Doctoral dissertation, Universitas Mercu Buana Jakarta).

Li, W., Wang, C. H., Cheng, G., & Song, Q. (2023). International conference on machine learning. *Transactions on machine learning research*.

Mandias, G. F. (2017). Analisis pengaruh pemanfaatan smartphone terhadap prestasi akademik mahasiswa Fakultas Ilmu Komputer Universitas Klabat. *Cogito Smart Journal*, *3*(1), 83-90.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., ... & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1-40.

Nadhif, I., Budiarto, U., & Iqbal, M. (2020). Optimasi Bentuk Buritan Kapal Perintis 750 DWT Menggunakan Response Surface Methode (RSM) untuk Mengurangi Hambatan. *Jurnal Teknik Perkapalan*, *9*(1), 105-111.

Nizar Saeful Muslim, N. S. (2021). *Pengaruh Kombinasi Suhu dan Lama Pengeringan Laru Tepung Jagung Putih dengan Tray Dryer Hasil Optimasi Response Surface Methode (RSM)* (Doctoral dissertation, Universitas Sahid Jakarta).

Nugraha, A. C. (2020). Penerapan Teknologi Blockchain dalam Lingkungan Pendidikan: Studi Kasus Jurusan Teknik Komputer dan Informatika POLBAN. *Produktif: Jurnal Ilmiah Pendidikan Teknologi Informasi*, *4*(1), 302-307.

Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D. (2023). Is chatgpt a general-purpose natural language processing task solver?. *arXiv preprint arXiv:2302.06476*.

Rahmanto, Y., Ulum, F., & Priyopradono, B. (2020). Aplikasi pembelajaran audit sistem informasi dan tata kelola teknologi informasi berbasis Mobile. *Jurnal Tekno Kompak*, *14*(2), 62-67.

Ramadhani, M. R., Aryadita, H., & Wicaksono, S. A. (2018). Analisis dan Perancangan Sistem Informasi Manajemen Donasi, Kegiatan, dan Relawan bagi Komunitas Sosial di Kota Malang (Studi Kasus: Komunitas TurunTangan Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *2*(9), 3102-3109. Tedjasuksmana, B. (2021, March). Optimalisasi Teknologi Dimasa Pandemi Melalui Audit Jarak Jauh Dalam Profesi Audit Internal. In *Seminar Nasional Akuntansi dan Call for Paper* (Vol. 1, No. 1, pp. 313-322).

Widyono, S. F., Hendrakusma, N., & Akbar, M. A. (2019). Perancangan User Interface Aplikasi Travelingyuk Berbasis Mobile Menggunakan Metode Human Centered Design (HCD). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *3*(8), 7415-7424.

Wu, L., Chen, Y., Shen, K., Guo, X., Gao, H., Li, S., ... & Long, B. (2023). Graph neural networks for natural language processing: A survey. *Foundations and Trends® in Machine Learning*, *16*(2), 119-328.