Journal of Renewable Energy, Electrical, and Computer Engineering
Volume 4, Number 2, September 2024. 83-89
eISSN 2776-0049
DOI: https://doi.org/10.29103/jreece.v4i2.17197
Research Original Article

Sentiment Analysis on Digital Korlantas POLRI Application Reviews Using the Distilbert Model

Nabila Rizky Amalia Putri✉[1], Trimono[2], Aviolla Terza Damaliana[3]
[1] Data Science, Faculty of Computer, University of Pembangungan Nasional "Veteran" Jawa Timur, Surabaya, 60294, Indonesia, 20083010018@student.upnjatim.ac.id
[2] Data Science, Faculty of Computer, University of Pembangungan Nasional "Veteran" Jawa Timur, Surabaya, 60294, Indonesia, trimono.stat@upnjatim.ac.id
[3] Data Science, Faculty of Computer, University of Pembangungan Nasional "Veteran" Jawa Timur, Surabaya, 60294, Indonesia, aviolla.terza.sada@upnjatim.ac.id
✉Corresponding Author: 20083010018@student.upnjatim.ac.id | Phone: +6288228219178

Received: July 03, 2024 **Revision: August 18, 2024** **Accepted: September 20, 2024**

Abstract
The implementation of digitalization in public services by Korlantas Polri has facilitated faster administration, wider access, and improved service quality. The Korlantas Polri Digital app has garnered more than 5 million downloads on the Google Play Store, with a rating of 3.7 and around 110 thousand reviews. Given that an app's reputation can be significantly affected by criticism, sentiment analysis becomes very important to categorize user reviews as positive, negative, or neutral, thus assisting developers in identifying app shortcomings. This study uses DistilBERT, a deep learning model distilled from BERT, to assess the effectiveness of sentiment analysis on reviews. Data was collected from user reviews on the Google Play Store between September 1, 2023 and May 31, 2024, resulting in 8,752 reviews retained for analysis. Model performance was evaluated at three data ratios: 60:40, 70:30, and 80:20, with the best performance results seen at a ratio of 80:20, achieving 88% accuracy. Increasing the training data ratio from 60:20 to 80:20 has a positive impact on the model, suggesting that the model can learn better with larger training data.
Keywords: Digitalization; Digital Korlantas Polri; Sentiment Analysis; DistilBERT

Introduction
Digital transformation in government involves the application of digital technology across various sectors to enhance efficiency, transparency, public participation, and the delivery of quality services to society (Wiranti & Frinaldi, 2023). Digitalizing public services is instrumental in optimizing service effectiveness, enabling people to access information and conduct transactions without physically visiting public offices. Therefore, the digitalization of public services represents a key solution for improving efficiency and accountability to the public. (Yulanda & Fachri Adnan, 2023).

The application of digitalization to public services enables the government to expedite administrative processes, broaden access to services, and enhance service quality (Yulanda & Fachri Adnan, 2023). Digitalization streamlines access to public services, reducing bureaucratic complexities. This innovation facilitates more efficient, effective, and transparent public services, thereby enhancing public satisfaction (Wiranti & Frinaldi, 2023). The assessment of public satisfaction serves as a critical measure of the success of public services, reflecting public trust in the government (Prasetiya & Niswah, 2020). The government has undertaken numerous initiatives to bolster public trust and improve the quality of public services. One agency dedicated to enhancing public trust through improved service performance is the Indonesian National Police (POLRI).

According to Law Number 2 of 2002, the police of the Republic of Indonesia are an integral part of the government tasked with maintaining security, and public order, enforcing the law, and providing protection, guidance, and service to the community. Additionally, the police are responsible for motor vehicle administration matters, including issuing the Vehicle Registration Certificate, renewing the Vehicle Registration Certificate, and issuing and renewing the Driving License (Putra & Febriawan, 2024).

The service for issuing and renewing SIMs provided by the Police serves as proof of registration and identification for individuals who meet administrative requirements, possess an understanding of traffic regulations, maintain physical and mental health, and demonstrate proficiency in driving (Cahyani et al., 2021). This service is often perceived as complex and cumbersome, prompting the National Police to strive for enhanced service optimization and quality improvement. Therefore, *Korlantas Polri* launched the "Digital *Korlantas Polri*" application in April 2021, incorporating features such as SINAR for SIM registration and renewal, SIGNAL, NTMC POLRI, and ETLE, although currently only SINAR is operational for SIM renewals (Maharani & Meiliana, 2021). This application has been implemented across 54 SATPAS nationwide, downloaded 5 million times from the Google Play Store, with a rating of 3.7 and approximately 110 thousand reviews, aimed at reducing crowds and preventing intermediary practices (Aulia & Maulana, 2023). As an innovation, this application has garnered public attention with diverse reviews offering suggestions and criticism. These reviews play a crucial role for developers in identifying application shortcomings, necessitating sentiment analysis to categorize reviews based on positive or negative sentiments.

Sentiment analysis is a computational study that analyzes text containing individuals' thoughts, opinions, views, judgments, and emotions on a particular subject (Pasek et al., 2022). Opinion data can be sourced from various digital platforms, such as social media and online market reviews. This analysis aids decision-making by categorizing sentences based on positive or negative sentiment (Sutrisno & Amini, 2023). Several methods support the classification process in sentiment analysis, including Naive Bayes, Linear Regression, Support Vector Machine (SVM), and deep learning methods such as BERT and DistilBERT. While Transformer models like BERT necessitate substantial computation and time, DistilBERT, a lighter and more efficient variant, maintains comparable accuracy (Sanh et al., 2019).

Previous research demonstrates the effectiveness of the DistilBERT model in sentiment analysis. For instance, research by Fajri et al. (2022) found that DistilBERT achieved 97% accuracy in analyzing Twitter sentiments related to COVID-19, outperforming BERT, which achieved 87%. Similarly, research by Mahira Putri et al. (2023) indicates that DistilBERT achieved 89% accuracy in sentiment analysis for the Indonesian presidential election, demonstrating more efficient performance in GPU memory usage compared to BERT. These findings suggest that DistilBERT represents a more efficient alternative to BERT in sentiment analysis, without compromising accuracy.

This research aims to evaluate the effectiveness of the DistilBERT text classification method in sentiment analysis of reviews. Based on recent research, researchers want to identify the extent of DistilBERT's effectiveness in text classification. It is hoped that the results of this research provide valuable insight into the effectiveness of these methods in sentiment analysis, which can have a major impact on developing more sophisticated and reliable sentiment analysis systems in the future.

Literature Review

Sentiment Analysis

Sentiment analysis is a computational study that focuses on opinions, evaluations, and emotions expressed in text. The objective of sentiment analysis is to assess the entities discussed in each document and determine whether the opinions expressed are positive, negative, or neutral (Yunitasari et al., 2019), (Wankhade et al., 2022).

Natural Language Processing (NLP)

Natural Language Processing (NLP) is a branch of computer science that encompasses artificial intelligence and linguistics, focusing on how computers interact with natural human languages such as English or Indonesian (Mhd. Zamil, 2019), (Kang et al., 2020).

Text Preprocessing

Text preprocessing is one of the most important stages and must be conducted before data is used to create a model (Hermawan et al., 2023). The objective is to transform raw data into a useful and efficient format. Generally, text preprocessing involves several steps, including:

1. Case folding is the process of converting all uppercase letters to lowercase to standardize the text and avoid discrepancies due to letter case during analysis. This stage is crucial for facilitating information retrieval.
2. Data cleaning involves removing elements from the data that could impact the analysis results, such as punctuation, symbols, duplicated characters, extra spaces, and numbers.
3. Normalization is the process of standardizing word forms, such as replacing abbreviations with full words or correcting spelling errors to ensure consistent meanings.
4. Tokenization is the process of segmenting sentences in a dataset into individual words using spaces or punctuation. Irrelevant characters and numbers are typically excluded.
5. Stopword removal entails eliminating words that are irrelevant and deemed to have no significant impact on data processing, including question words, conjunctions, and meaningless words.
6. Stemming is the process of reducing words with affixes to their base or root form, commonly applied to nouns, verbs, and adjectives.

Lexicon based labeling

Lexicon-based labeling is the process of analyzing each word in a sentence within a dataset to ascertain the sentiment polarity of the sentence. This method does not necessitate training data and employs two approaches: dictionary-based and corpus-based. One example of the corpus-based approach is the Indonesian Sentiment Lexicon, which contains 3,609 positive words and 6,609 negative words. The corpus identifies positive or negative words in the dataset and assigns a corresponding value to each word (Khatoon et al., 2020).

Transformers

The Deep Learning model introduced in 2017, described in the paper "Attention is All You Need," is known as Transformers. The Transformers architecture was developed to process sequential data, particularly text, and can be applied to various NLP tasks such as machine translation and text summarization (Vaswani et al., 2017)

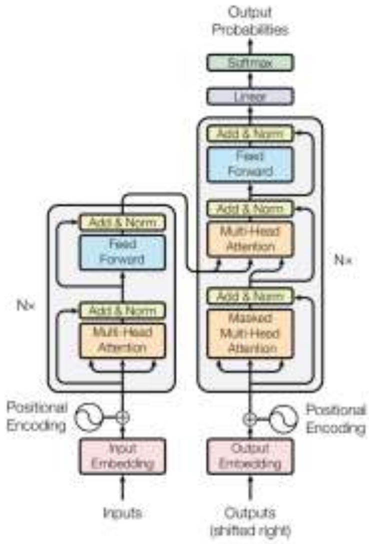


Figure 1. Transformers Architecture

Transformers consist of two main components: the encoder and the decoder. The encoder processes input iteratively through a series of encoding layers, while the decoder performs a similar function on the output generated by the encoder (Vaswani et al., 2017).

The encoder comprises two sub-layers: multi-head attention in the first layer and a fully connected feed-forward network in the second layer, with a normalization layer between them. The output of each layer is the sum of the layer's normalization and the result of the sub-layer, expressed as Norm (x + sub-layer(x)). The decoder, on the other hand, consists of three sub-layers: the first two sub-layers mirror those in the encoder (multi-head attention and fully connected feed-forward network), while the third sub-layer is masked multi-head attention. Each sub-layer also has a normalization layer. Masking in the third sub-layer ensures that predictions for position i depend only on the output of positions lower than I (Vaswani et al., 2017).

BERT

BERT stands for Bidirectional Encoder Representations from Transformers and utilizes the Transformers architecture to generate precise text representations. Its primary feature is the capability to comprehend the contextual meaning of words within lengthy sentences. BERT has been trained across various language understanding tasks, including word completion, text classification, and question answering, where it predicts missing words or answers questions based on context. Once trained, BERT can be applied to diverse Natural Language Processing (NLP) tasks, including question comprehension and sentiment analysis (Devlin et al., 2019).

The BERT architecture consists of a multi-layer bi-directional Transformers encoder and has two forms: BERT BASE and BERT LARGE. BERT BASE has 12 Transformer blocks, 12 attention layers, and 768 hidden layers, while BERT LARGE has 24 Transformer blocks, 16 attention heads, and 1024 hidden layers (Devlin et al., 2019).

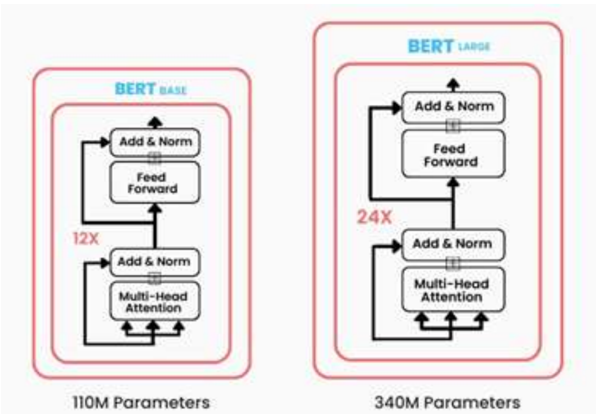


Figure 2. BERT Architecture

DistilBERT

DistilBERT is a model in the field of Natural Language Processing (NLP) developed as a lighter and more efficient version of the BERT (Bidirectional Encoder Representations from Transformers) model. DistilBERT is designed to offer nearly equivalent performance to BERT but with reduced resource requirements. The goal of DistilBERT is to decrease the size of BERT parameters and enhance the training speed of bidirectional encoder representations of transformers (BERT) models. Using the DistilBERT method, the size of the BERT parameter model can be reduced by 40% while increasing the inference speed by 60%.

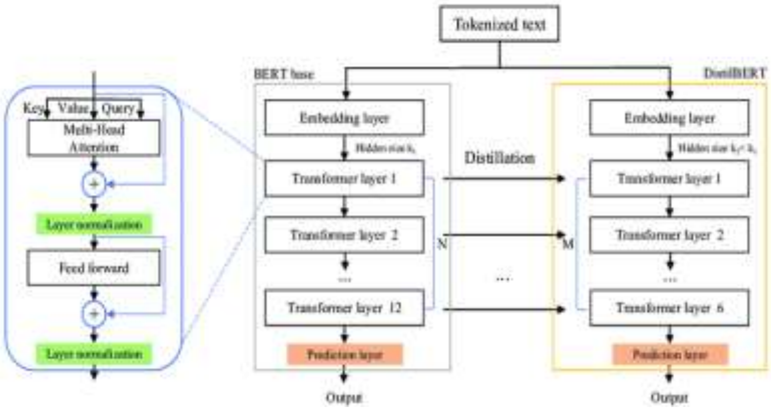


Figure 3. DistilBERT Architecture

DistilBERT comprises only 6 layers, which amounts to 50% of the layers in BERT. During the simplification process, information from multiple layers in BERT is consolidated. The weights of the deeper BERT layers are adjusted and transferred to the shallower layers in DistilBERT, resulting in a model with only 66 million parameters.

Confusion Matrix

A Confusion Matrix is a table used to assess the performance of a classification model by analyzing its effectiveness in classifying the test dataset. The table in the confusion matrix shows the amount of data in the actual class, while the column shows the amount of data in the predicted class.

Table 1. Confusion Matrix

Predicted	**Actual**	
	Positives	Negatives
Positives	True Positive (TP)	False Positive (FP)
Negatives	False Negative (FN)	True Negative (TN)

The Confusion Matrix compares the classification results of actual data by measuring the level of accuracy, precision, recall, and F1 score.

a) Accuracy is a metric obtained by comparing correctly classified data with the total dataset. It measures the model's precision in classifying results accurately. Mathematically, accuracy is defined as follows.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

b) Precision is a metric derived from the ratio of correctly predicted positive instances to the total predicted positive instances. Mathematically, precision can be defined as follows.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

c) Recall is a metric representing the ratio of correctly predicted positive instances to the total actual positive instances. Mathematically, recall can be defined as follows.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

d) The F1-Score is the harmonic mean of precision and recall. Mathematically, the F1-Score can be defined as follows.

$$F1\ Score = 2 \times \frac{precision \times recall}{precision+recall} \tag{4}$$

Materials & Methods

This research extracted data from user reviews of the National Police Traffic Corps Digital Application posted on the Google Play Store. The data was scraped and saved in a Comma Separated Value (CSV) format file. Data collection occurred between September 2023 and May 2024, resulting in a total of 8752 reviews. These reviews encompass various aspects of the application and were gathered for sentiment analysis. The dataset comprises 11 attributes from the scraping process, including reviewID, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replay

Content, revisedAt, and appVersion.

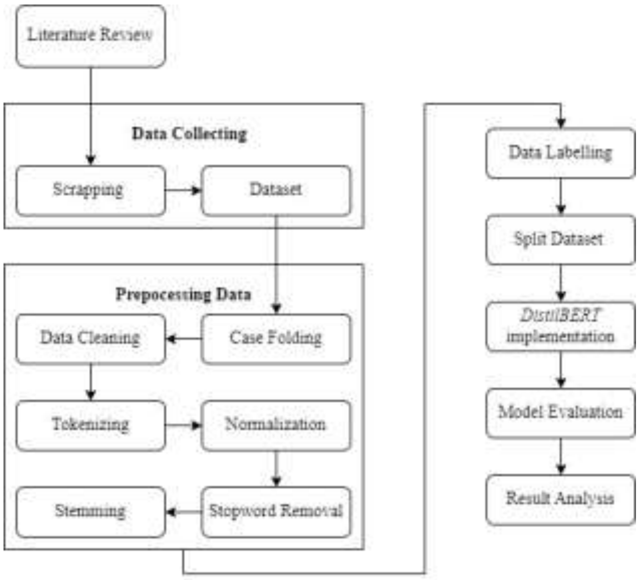


Figure 4. Research Flow

User review data are captured using web scraping techniques and saved in CSV format for further analysis. The collected data undergoes preprocessing to ensure quality and consistency, which includes removing duplicates, handling missing values, and normalizing the review text. Sentiment analysis is conducted using the DistilBERT model, a distilled variant of BERT known for its lighter and more efficient architecture. This model was selected due to its accuracy in understanding the context and nuances of natural language, coupled with enhanced computational efficiency. The sentiment analysis categorizes reviews into three distinct categories: positive sentiment, neutral sentiment, and negative sentiment.

The results of the sentiment analysis will be visualized using graphs and diagrams to facilitate interpretation, aiding in the comprehension of patterns and trends in user sentiment toward the application over the studied period. All findings from this research are comprehensively documented in the final report, encompassing methodology, analysis results, discussion, and recommendations for further development of the application.

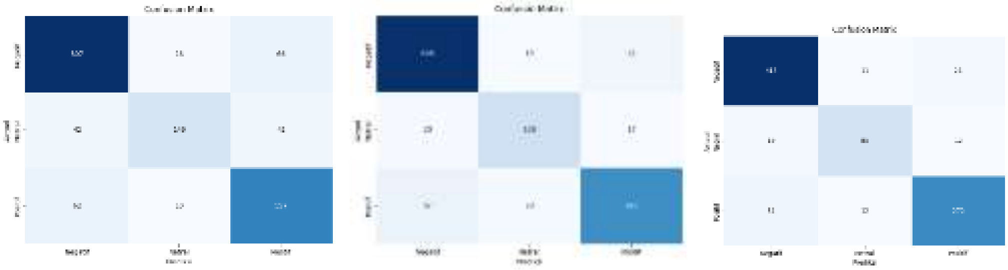
Results and Discussion

This research used data derived from user reviews of the National Police Traffic Corps Digital application, available on the Google Play Store. The data was collected through data scraping techniques using the Google-play-scraper library. Data scraping occurred within a time frame from 1 September 2023 to 31 May 2024. The scraping process resulted in a total of 48,245 rows and 11 columns. Subsequently, the data was filtered, yielding 8752 review entries which constituted the dataset for analysis. For data classification, the dataset was divided into three trials with ratios of 60:40, 70:30, and 80:20. Each trial's training data was further divided, maintaining the respective ratios, into training data, validation data, and test data.

Table 2. Dataset Split Ratio

Distribution Ratio	Training Data	Validation Data	Test Data
60:40	5251	1750	1751
70:30	6126	1313	1313
80:20	7001	875	867

After dividing the data, the next stage involved testing the dataset using DistilBERT for classification, and evaluation was conducted using Confusion Matrix calculations. This research employed three testing scenarios to calculate accuracy, precision, recall, and F1-Score, with data ratios of 60:40, 70:30, and 80:20. The evaluation results of the Confusion Matrix calculations are as follows:



(1) Ratio 60:40 (2) Ratio 70:30 (3) Ratio 80:30

Figure 5. Confusion Matrix Results

Based on the information provided, the model successfully classified the three sentiments. Regarding the negative class (0), the model correctly classified 413 negative samples as negative. It incorrectly classified 19 negative samples as neutral and 31 negative samples as positive. For the neutral class (1), the model correctly classified 85 neutral samples as neutral. It incorrectly classified 11 neutral samples as negative and 12 neutral samples as positive. In terms of the positive class (2), the model correctly classified 272 positive samples as positive. 21 positive samples were incorrectly classified as negative and 12 positive samples were incorrectly classified as neutral.

The results of the performance evaluation for the three scenarios were summarized and placed into one table as shown below.

Table 3. DistilBERT Model Evaluation Result

	Ratio 60:40	Ratio 70:30	Ratio 80:20
Accuracy (%)	86.40%	86.44%	88%
Precision (%)	86.39%	86.46%	87.80%
Recall (%)	86.40%	86.44%	88%
F1-Score (%)	86.16%	86.42%	87.81%

Based on the table above, it was concluded that generally, the model with batch size 16, learning rate 1e-5, and 15 epochs exhibited good performance in classifying sentiment, achieving high accuracy, precision, recall, and F1-score. At an 80:20 ratio, the model demonstrated the best performance with an accuracy of 88%, surpassing other data ratios. Increasing the training data ratio from 60:40 to 80:20 positively impacted the model, suggesting improved learning capabilities with larger training datasets. However, it should be noted that the model still has room for improvement, particularly in correctly classifying neutral sentiment samples, indicating a need for enhancements in detecting this class.

Conclusions

Based on the research conducted across all data ratios (60:40, 70:30, 80:20), the DistilBERT model exhibited improved loss and achieved high accuracy on both training and validation data as the epochs progressed. The evaluation results using the confusion matrix showed that the highest accuracy of 88% was achieved with a data ratio of 80:20, batch size 16, learning rate 1e-5, and 15 epochs, indicating excellent performance for the NLP model. However, there were minor fluctuations in validation loss and accuracy observed across several epochs (60:40 from the 4th to the 8th epoch, 70:30 from the 5th to the 7th epoch, and 80:20 from the 7th to the 8th epoch). Increasing the training data ratio had a positive impact on the model's performance. Recommendations include acquiring more training data, particularly for the Neutral class, improving neutral class handling through oversampling or undersampling techniques, exploring different model architectures, and adjusting hyperparameters to discover the optimal configuration and prevent overfitting.

Acknowledgments

Both parents and the writer's family have always provided prayers, motivation, and moral support in completing this final project. Mr. Trimono, S.Si., M. Si, as the Main Supervisor, and Ms. Aviolla Terza Damaliana, S.Si., M. Stat, as the Supporting Supervisor, have dedicated much time to guide, direct, and provide input to the writer throughout the completion of this final project. All fellow Data Science cohort members from the 2020 batch have greatly assisted in the writing process of this final project. Apologies for any shortcomings. May the grace, guidance, and protection of Allah SWT be bestowed upon us all. Ameen.

References

Aulia, S., & Maulana, A. (2023). Daftar Satpas Yang Menerima Pengajuan SIM Secara Online. Kompas.Com.

Cahyani, L. I., Fiiki Nurrohman, M., Setiowati, M. I., Taufiq, R. I., & Mukti, A. (2021). Transformasi Manajemen Kepolisian Melalui Pelayanan Publik Berbasis Aplikasi Sim Nasional Presisi (Sinar). Jurnal Mahasiswa Administrasi Negara (JMAN), 5(2), 34–41.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-Training Of Deep Bidirectional Transformers For Language Understanding. NAACL HLT 2019 - 2019 Conference Of The North American Chapter Of The Association For Computational Linguistics: Human Language Technologies - Proceedings Of The Conference, 1(Mlm), 4171–4186.

Fajri, F., Tutuko, B., & Sukemi, S. (2022). Membandingkan Nilai Akurasi BERT Dan Distilbert Pada Dataset Twitter. JUSIFO

(Jurnal Sistem Informasi), 8(2), 71–80. Https://Doi.Org/10.19109/Jusifo.V8i2.13885

Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining Untuk Analisis Sentimen Pada Twitter Dengan Algoritma Support Vector Machine. JST (Jurnal Sains Dan Teknologi), 12(1), 129–137. Https://Doi.Org/10.23887/Jstundiksha.V12i1.52358

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural Language Processing (NLP) In Management Research: A Literature Review. Journal Of Management Analytics, 7(2), 139–172.

Khatoon, S., Romman, L. A., & Hasan, M. M. (2020). Domain Independent Automatic Labeling System For Large-Scale Social Data Using Lexicon And Web-Based Augmentation. Information Technology And Control, 49(1), 36–54.

Maharani, T., & Meiliana, D. (2021). Korlantas Polri Luncurkan Aplikasi SINAR, Perpanjang SIM Cukup Lewat Handphone. Kompas.Com.

Mahira Putri, Sutanto, T. E., & Inna, S. (2023). Studi Empiris Model BERT Dan Distilbert Analisis Sentimen Pada Pemilihan Presiden Indonesia. Indonesian Journal Of Computer Science, 12(5), 2972–2980. Https://Doi.Org/10.33022/Ijcs.V12i5.3445

Mhd. Zamil. (2019). Klasifikasi Kalimat Ofensif Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier. Universitas Islam Negeri Sultan Syarif Kasim Riau, 2(5), 107–110.

Pasek, P., Mahawardana, O., Arya, G., Agus, I. P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini Dari Media Sosial Twitter Terhadap “ Figure Pemimpin ” Menggunakan Python. JITTER- Jurnal Ilmiah Teknologi Dan Komputer, 3(1).

Prasetiya, A. Y., & Niswah, F. (2020). Strategi Peningkatan Pelayanan Pembayaran Surat Izin Mengemudi Melalui Program Cashless Payment System (CPS) Di Kantor Satlantas Polres Gresik. Publika, 8(4), 1–10.

Putra, S. H. W., & Febriawan, D. (2024). Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes Pada Google Play Store. KLIK: Kajian Ilmiah Informatika Dan Komputer, 4, 1962–1971. Https://Doi.Org/10.30865/Klik.V4i4.1600

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). Distilbert, A Distilled Version Of BERT: Smaller, Faster, Cheaper And Lighter. 2–6.

Sutrisno, E. P., & Amini, S. (2023). Implementasi Algoritma K-Nearest Neighbor Pada Implementation Of K-Nearest Neighbor Algorithm In Sentiment Analysis Of User Reviews For Digital. 2(April 2021), 687–695.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances In Neural Information Processing Systems, 2017-Decem(Nips), 5999–6009.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A Survey On Sentiment Analysis Methods, Applications, And Challenges. Artificial Intelligence Review, 55(7), 5731–5780.

Wiranti, N. E., & Frinaldi, A. (2023). Meningkatkan Efisiensi Pelayanan Publik Dengan Teknologi Di Era Digital. JIM: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah, 8(2), 748–754.

Yulanda, A., & Fachri Adnan, M. (2023). Transformasi Digital: Meningkatkan Efisiensi Pelayanan Publik Ditinjau Dari Perspektif Administrasi Publik. Jurnal Ilmu Sosial Dan Humaniora (Isora), 1(3), 103–110.

Yunitasari, Y., Musdholifah, A., & Sari, A. K. (2019). Sarcasm Detection For Sentiment Analysis In Indonesian Tweets. IJCCS (Indonesian Journal Of Computing And Cybernetics Systems), 13(1), 53. Https://Doi.Org/10.22146/Ijccs.41136