Classification and Prediction of Theft Criminal Court Decision Documents Using the BERT Model

Galih Wasis Wicaksono1,a), Lusy Rohmadhoni1,b), Nur Putri Hidayah2,c), Uswatun Khasanah1,d)

1Department of Informatics, Universitas Muhammadiyah Malang, Kota Malang, Indonesia

2Department of Law, Universitas Muhammadiyah Malang, Kota Malang, Indonesia

a)Corresponding author: [galih.w.w@umm.ac.id](mailto:galih.w.w@umm.ac.id)

b) [lusyrohmadhoniummacid@webmail.umm.ac.id](mailto:lusyrohmadhoniummacid@webmail.umm.ac.id)

c) [nurputri@umm.ac.id](mailto:nurputri@umm.ac.id)

d) [uswatunkhasanah209@webmail.umm.ac.id](mailto:uswatunkhasanah209@webmail.umm.ac.id)

**Abstract.**  One of the most common types of crime in Indonesia is theft, which has significant social consequences for victims, society, and increases the responsibility of the justice system. Variations in the pattern of judges' decisions and the complexity of the language structure in legal documents are common problems that make the process of legal analysis more difficult to perform manually. This study aims to create a natural language processing-based classification and prediction system that utilizes transformation-based models. This study will focus on comparing the IndoBERT LongFormer and LegalBERT models. Data extraction from the text of the record, pre-processing, data sharing, model training, and performance evaluation are all parts of the research process. Two scenarios were used to test the three models: the classification of theft crime types and the prediction of estimated prison sentence lengths. Classification and prediction methods were chosen to capture the urgency of criminal acts, which often contain burdens and the need for more accurate and fair legal decision analysis. The results of the evaluation showed that IndoBERT performed best in classification, achieving a validation accuracy of 69% and a precision of 71%. The IndoBERT model can also predict prison crime estimates with an accuracy of up to 90%. To avoid biased results, model comparisons were conducted using consistent experimental scenarios. The results of this study demonstrate that the BERT-based model can be effectively applied in the Indonesian justice system. It can be a valuable tool for data-driven decision-making in the legal domain.

**Keywords:** Theft, Classification, Prediction, Court Decision, Criminal.

# INTRODUCTION

Crime can occur at all levels of society because crime can be committed by anyone, anywhere, and at any time without restrictions. The high level of crime against property, such as theft, is the primary concern of the community. This type of crime can be a significant threat to social norms and values. These norms and values form the basis of an orderly social order, and if violations are ignored without consequence, the social order will be damaged [1]. Theft breaches social norms and values. Theft as a criminal act causes disturbance for individuals and society as a whole. The crime of theft in aggravating circumstances will have a detrimental impact on the community, including disturbing order, peace, and security. This can result in significant losses to the community, both physically and financially [2]. In Indonesia, every criminal case is examined, tried, and decided by the court under Law No. 48 of 2009 concerning Judicial Power.

Jurisprudence in Indonesia is used to determine the results of constitutional decisions and written laws and regulations [3]. Past decisions made by other judges can help maintain legal certainty in cases where a legal vacuum or unclear rules exist, thereby ensuring consistency and predictability in the law.

The Indonesian judicial system faces a significant challenge in managing the increasing volume of criminal cases, particularly those involving theft. The crime of theft is influenced by the perpetrator's daily living conditions, such as economic conditions or low-income levels, which make it difficult for them to afford necessities and often result in a low level of education. The number of theft crimes in Indonesia continues to rise annually. According to statistical data from the Supreme Court of Indonesia's website, accessed on October 6, 2024, 246,647 theft verdicts are expected to be decided in 2024. The large number of verdicts generally makes it difficult for law enforcement to understand and analyze the verdict documents. Analysis of decisions is crucial for assessing and evaluating a case [4]. The theft case is one that has garnered the attention of judges due to its complexity and legal implications [5]. Manual classification and analysis of these court documents are time-consuming and prone to human error, which can lead to inconsistencies in case assessment and potentially affect the fairness of verdicts. Judges often struggle to classify and measure the impact of a theft case, especially in cases with aggravating circumstances [6]. As a result, there's a growing need for automated systems to classify and process legal documents efficiently, ensuring transparency and consistency in judicial proceedings.

Proper classification of judgments is essential to ensure a fair and proportionate verdict. The manual classification process of hundreds of thousands of judgment documents is time-consuming and has the potential to result in inconsistent analysis. Therefore, efforts to assist the judicial system by using automatic classification technology are becoming increasingly important [7]. Other studies suggest that legal document classification technology can improve the efficiency and accuracy of the judicial process, enabling judges to handle complex legal cases more effectively [8]. In conclusion, a system is needed that can automatically process the compilation of judicial documents and extract the necessary information quickly. This can only be done with the use of information technology.

This research focuses on two types of theft cases, namely Criminal Theft (CT) and Criminal Theft with Aggravating Circumstances (CTAC). Restrictions on this type of case are designed to enhance the accuracy of the information extraction process. Thus, an automatic classification system based on information technology is needed to ensure transparency, consistency, and speed in handling theft cases in Indonesia.

Natural Language Processing (NLP) has emerged as a crucial technology for addressing this challenge. Recent advancements in transformer-based models, such as Bidirectional Encoder Representations from Transformers (BERT), have revolutionized text processing by capturing complex semantic relationships in language more effectively than traditional methods [8], [9], [10]. Transformer models, especially Bidirectional Encoder Representations from Transformer (BERT), are more accurate and consistent in classification than previous text processing methods [11], [12], [13]. BERT's ability to process context bidirectionally makes it particularly well-suited for understanding the nuances of legal texts and factual presentations in court documents [14]. These models have demonstrated exceptional performance in various classification tasks, including those involving complex legal language and document structures [15], [16], [17]. In addition, studies have shown that BERT-based models can achieve good results in a wide range of classification tasks, including those involving legal texts by refining specific datasets [18].

Research on the prediction of court decisions has been conducted using various models [18], [19], [20]. The standard BERT model's limitation in analyzing long manuscripts—typically a maximum of 512 tokens—is a notable weakness. This study aims to address this issue by developing a classification model for theft cases, specifically distinguishing between theft and theft with aggravating circumstances. Our primary objective is to create a model that efficiently processes and classifies these lengthy verdict documents. Additionally, we will propose a sentence prediction model based on our classification results. This research is expected to contribute as a reference in predicting court decisions using Transformer models.

A diagram of a process

AI-generated content may be incorrect.

**FIGURE 1**. Research Workflow

# METHODS

This research employs a systematic methodology for data collection and processing, ensuring the quality and relevance of the dataset for our classification model. Figure 1 illustrates the workflow of our research, starting with data acquisition and moving through the essential preprocessing steps. It visually represents the transition from raw data obtained from the Supreme Court's website to a clean and structured dataset, ready for model training and evaluation.

**Data Collection**

This study utilizes primary data obtained directly by the researcher from the data source [21]. Data was obtained from the website of the Directory of Decisions of the Supreme Court of the Republic of Indonesia (<https://putusan3.mahkamahagung.go.id/>). Data collection was carried out using the web scraping method, specifically by selecting the type of verdict with the classification of “General Criminal-Theft” verdict, then choosing the judicial institution of the Surabaya District Court, and finally selecting the years 2020-2023. Crawl data using library requests and the Supreme Court's BeautifulSoup Directory. The pdfminer library is used to extract text from the downloaded verdict PDF file. OS and IO libraries help in file management. Library Re for the extraction of specific patterns from text. Library time and concurrent.futures, especially ThreadPoolExecutor, to ensure the process runs efficiently and is unblocked by managing requests simultaneously.

After the data of the theft criminal verdict is successfully collected through web scraping, the next step is to label each verdict. This process involves analyzing the content of the extracted verdict text to identify and classify the length of the prison sentence imposed, as well as the circumstances or modus operandi of theft (for example, ordinary theft, with aggravation, or involving violence). From the combination of these two pieces of information, researchers will generate specific types of theft categories, and this labeled data will serve as the basis for the BERT model to automatically study and classify new theft criminal convictions in the future.

**Preprocessing**

The preprocessing stage aims to extract and filter important information from criminal convictions, the results of which will be used for model training and significantly affect the accuracy and efficiency of the model [22]. To achieve this, the data cleaning process is carried out to eliminate duplication and data that is not relevant to the crime of theft, thereby ensuring high dataset quality. Furthermore, case folding uniformizes the text format, making the model recognize the same word regardless of its capitalization. Finally, feature selection focuses on selecting the most relevant attributes—especially the content of the affidavit—to group the types of crimes and sentences imposed, effectively reducing the data dimension and increasing the model's effectiveness in classifying criminal theft verdicts.

1. Data cleaning

Data cleaning aims to remove data from court decision documents that are not relevant to the classification purpose [23]. The process carried out involves deleting duplicate data and data unrelated to theft.

1. Case folding

Case folding aims to standardize data formats by standardizing text [25].

1. Feature selection

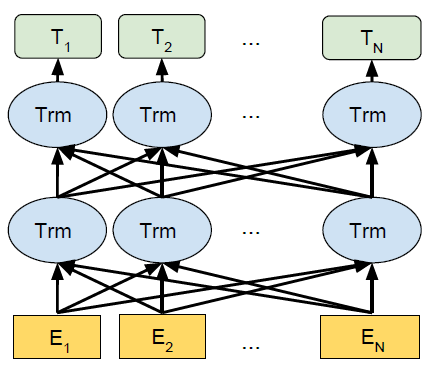
Feature selection aims to select which attributes will be used in the analysis. This preprocessing is performed to reduce the model's complexity [24]. There are 19 attribute columns available in the dataset. The selection of these attributes is based on their relevance to the model's classification of criminal theft verdicts. In this study, a classification was conducted based on the content of the verdict records to categorize the types of theft crimes and determine the punishments imposed.

**Split Data**

Data splitting is a technique used to partition datasets, which affects the performance of classification models in machine learning algorithms. Data splitting is a process of dividing data into training data and test data [25]. The training data is used to train the model to understand patterns in the dataset. In contrast, the test data is used to evaluate the model's performance in predicting patterns in the dataset that have never been seen before. The dataset used in this study is divided into 80% training data and 20% test data.

**Model**

The BERT model was released by Google in 2018 as an NLP model [26]. To process text input, the BERT model uses several interconnected layers of Transformers (Trm). Illustrate in Figure 2. Each Transformer layer features an attention mechanism, enabling the model to consider the context of each word in a sentence from both the left and right [27].



**FIGURE 2**. Model BERT

The models used in this study are IndoBERT Base Phase 2, LegalBERT, and LongFormer. The selection of these models is based on the data used related to the law. AutoTokenizer and AutoModel Huggingface Transformers are used to load models and tokenizers. Then, the tokenizer converts the text data into an ID token and an attention mask that matches the BERT input format. Once the data is tokenized, the BERT model is put in evaluation mode (without training) to generate embedding output only from text input.

**Fine-tuning**

Fine-tuning is a process in which a pre-trained neural network model is modified to fit a new and more specific task. The concept is to leverage some of the knowledge and features learned from the previously trained model and adapt them to suit the needs of the new task better. With this process, much less time and resources are required than would be necessary to train the model from scratch [28].

Additionally, fine-tuning enables the model to adapt to specific linguistic patterns present in the datasets, thereby improving classification accuracy and F1-score values [29]. Fine-tuning the BERT architecture effectively enhances the performance of classifier models, a technique that has been common practice in previous research. Additionally, this method is widely used to evaluate benchmarks on various tasks, demonstrating the advantages of BERT compared to earlier models [30].

**Evaluation**

Model evaluation is a systematic assessment of the model's performance. Predicted data is compared with actual data to ensure accurate (true) and inaccurate (false) prediction quantities [31]. The performance evaluation of the BERT model was conducted using a Confusion Matrix. The Confusion Matrix is a statistical table used to summarize the predictions of the applied model. In a confusion matrix, each row shows the actual data class, while each column shows the class predicted by the model [32]. From this Confusion Matrix, metrics such as accuracy, precision, recall, and F1-score are calculated to provide a comprehensive picture of how well the model differentiates between different types of theft and jail sentences, with the intention of comparing the performance of each model.

# RESULTS AND DISCUSSION

**Data Collection Results**

The results of data collection through web scraping from 2020 to 2023 at the Surabaya District Court were obtained, 2.804 datasets. This dataset is then preprocessed to eliminate irrelevant data, address missing values, and select the relevant attributes, specifically numbers and records. The result after preprocessing was 2.362 datasets, which were then labeled to classify the type of theft crime based on the information contained in the 'Amar Record' section in each verdict document, as shown in Table 1.

**TABLE 1**. Crawling Data Results

|  |  |
| --- | --- |
| **Theft Categories** | **Total** |
| Criminal theft (CT) | 646 |
| Criminal theft in aggravating circumstances (CTAC) | 1493 |
| Criminal act of theft with violence | 223 |

Based on the results of labeling the data that has been processed in Table 1. This study uses 1200 data points from the two main categories of theft crimes, with an equal number of data points in each category. To avoid data distribution inequality that can affect model performance, the amount of data for each category is arranged in a balanced manner. The purpose of this selection is to ensure that the model runs both categories fairly without dominating either class. This will enhance the accuracy of the classification results.

**TABLE 2**. The Final Dataset

|  |  |
| --- | --- |
| **Theft Categories** | **Total** |
| Criminal Theft (CT) | 600 |
| Criminal Theft in Aggravating Circumstances (CTAC) | 600 |

The relevance and clarity of the legal classification in the verdict record also determine the selection of the two categories of theft in Table 2. To avoid the model's tendency to recognize better patterns from larger categories, which can lead to biased prediction results, the training process of the classification model must be optimally carried out, taking into account the balance of data present in each category. Therefore, a balanced dataset is used to enhance the model's generalization ability in distinguishing between types of theft crimes based on existing categories.

**Preprocessing Results**

The first stage of preprocessing involves data cleaning to filter out data unrelated to theft and remove duplicate data. The next stage is case folding, which converts the entire document text to lowercase letters. After case folding, the next stage is feature selection. The last stage is labeling for the grouping of theft categories. The labeling includes the length of the prison sentence imposed, the circumstances of the theft crime, and the category of theft.

**Test Results**

The entire training process for each model was carried out by dividing the dataset into training data and test data, using an 80:20 ratio. This enables a fair evaluation of each model's performance on data that has never been seen before.

1. IndoBERT Base Phase 2 (IB)

To adapt the IndoBERT model (indobenchmark/indobert-base-p2), which serves as the primary feature extractor, for classifying criminal theft convictions, the architecture was modified by adding a custom Bert Layer to extract the token representation [CLS]. This representation is then passed through two dense layers, where the first layer has 128 units with ReLU activation and is followed by a dropout layer (0.3) to reduce overfitting. It is then fed into the last dense layer with softmax activation, corresponding to the number of predicted categories. At the same time, the original pre-training head of BERT (MLM and NSP) is implicitly released to adjust the classification task. The results of the training showed that the model achieved an accuracy of 69.17%. The results are presented in Table 3.

**TABLE 3**. Classification Report IndoBERT Base Phase 2 (IB), LegalBERT (LB), and LongFormer (LF)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Label** | **Precision** | | | **Recall** | | | **F1-score** | | |
| **IB** | **LB** | **LF** | **IB** | **LB** | **LF** | **IB** | **LB** | **LF** |
| CT | 0.68 | 0.69 | 0.66 | 0.72 | 0.72 | 0.77 | 0.70 | 0.70 | 0.71 |
| CTAC | 0.71 | 0.71 | 0.72 | 0.66 | 0.66 | 0.61 | 0.68 | 0.68 | 0.66 |
| Accuracy |  |  |  |  |  |  | 0.69 | 0.70 | 0.69 |
| Macro avg | 0.69 | 0.70 | 0.69 | 0.69 | 0.70 | 0.69 | 0.69 | 0.70 | 0.69 |
| Weighted avg | 0.69 | 0.70 | 0.69 | 0.69 | 0.70 | 0.69 | 0.69 | 0.70 | 0.69 |

1. LegalBERT (LB)

The LegalBERT (nlpaueb/legal-bert-base-uncased) model, pre-trained specifically on the English-language legal corpus, was modified to understand the terminology and unique sentence structure in legal documents, enabling it to classify criminal theft verdicts effectively. These modifications include the addition of a BertLayer custom layer to extract the token representation [CLS], which is then routed through two dense layers—the first layer with 64 ReLU units and a dropout layer (0.3) for overfitting mitigation—before reaching the final dense output layer with softmax activation corresponding to the number of predicted categories. At the same time, the original LegalBERT pre-training heads (MLM and NSP) are implicitly released for use in classification tasks. The results of the training showed that the model achieved an accuracy of 69.58%. Detailed classification reports are presented in Table 3.

1. LongFormer (LF)

The LongFormer model (allenai/longformer-base-4096) is specifically used to overcome the input length limitations of standard BERT models. The LongFormer architecture is modified by adding a custom BertLayer layer to extract the token representation [CLS], which is then passed to two dense layers—the first layer with 64 ReLU units and a dropout (0.3) to reduce overfitting—before reaching the last dense output layer with softmax activation as per the number of predicted categories.

The transfer learning strategy was implemented by freezing the initial nine layers of the LongFormer encoder to retain pre-trained knowledge while adapting the model to specific legal data. LongFormer's original pre-training head (MLM) is implicitly released for adjustments to this classification task. During training, the validation accuracy was stable at 0.69-0.73, while the training accuracy only increased slowly to 0.56. The Loss Model graph shows that the value of training losses continues to decline consistently, and validation losses also decline steadily with no signs of overfitting. The results of the training showed the model achieved an accuracy of 68.75%, as shown in Table 3.

**TABLE 4**. Comparison of Model Classification Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Classification Results** | | | |
| **Precision** | **Recall** | **F1-score** | **Accuracy** |
| IndoBERT Base Phase 2 (IB) | 0.70 | 0.69 | 0.69 | 0.69 |
| LegalBERT (LB) | 0.70 | 0.69 | 0.69 | 0.70 |
| LongFormer (LF) | 0.69 | 0.69 | 0.69 | 0.69 |

Based on the classification results in Table 4, the three models successfully categorized the types of theft crimes. This table displays evaluation metrics, including Precision, Recall, F1-score, and Accuracy, for each model. LegalBERT shows slightly higher classification accuracy (70%). This suggests that LegalBERT, designed explicitly for legal texts, is slightly superior in categorizing judgment documents based on the type of crime. However, overall, the performance of the three models' classification is quite similar.

**TABLE 5**. Comparison of Model Prediction Results

|  |  |  |
| --- | --- | --- |
| **Model** | **Probability** | |
| **CT** | **CTAC** |
| IndoBERT Base Phase 2 | 10% | 90% |
| LegalBERT | 62% | 38% |
| LongFormer | 51% | 49% |

Table 5, showing the probability prediction results for the theft crime categories, highlights the superior performance of the IndoBERT Base Phase 2 model. The model's ability to predict the CTAC category with a high probability of 90% is a key finding. This high confidence score suggests that IndoBERT effectively identifies specific linguistic and factual patterns within the Indonesian legal documents that correspond to aggravating circumstances, such as the use of violence, a repeated offense, or a sophisticated modus operandi. From a legal standpoint, this high accuracy is crucial because it indicates the model's potential to significantly assist judges and legal professionals in quickly identifying complex cases that warrant a more severe sentence, thereby promoting greater consistency and fairness in judicial decisions. While LegalBERT is tailored for general legal language, IndoBERT’s strong performance confirms that its pre-training on a large Indonesian corpus.

# Conclusion

This study successfully implemented a classification method using the IndoBERT Base Phase 2, LegalBERT, and LongFormer models to categorize theft criminal verdict documents in Indonesian. While LegalBERT showed a slightly higher classification accuracy, IndoBERT Base Phase 2 proved to be superior for probability prediction tasks due to its compatibility with the Indonesian language dataset, making it the most suitable model for this research. Despite these successes, the study's primary limitation is its focus on a single court, the Surabaya District Court, which may introduce geographical and institutional bias. Therefore, for future work, we recommend expanding the dataset to include verdicts from various other courts to improve the model's generalizability and reduce this bias. We also suggest broadening the scope to include more specific theft classifications beyond the current categories. Furthermore, applying advanced data augmentation and fine-tuning techniques specifically to the IndoBERT model could significantly enhance its accuracy and robustness, paving the way for a more comprehensive and effective digitalization of the Indonesian legal system.

# References

[1] H. Hamdiyah, “ANALISIS UNSUR-UNSUR TINDAK PIDANA PENCURIAN: TINJAUAN HUKUM,” *Jurnal Tahqiqa : Jurnal Ilmiah Pemikiran Hukum Islam*, vol. 18, no. 1, pp. 98–108, Jan. 2024, doi: 10.61393/tahqiqa.v18i1.216.

[2] M. D. Pane and G. Endang Renika Siahaan, “TINJAUAN YURIDIS MENGENAI TANGGUNG JAWAB PELAKU TINDAK PIDANA PENCURIAN: ANALISIS BERDASARKAN PASAL 362 KITAB UNDANG-UNDANG HUKUM PIDANA (KUHP),” *Jurnal Ilmiah Dinamika Hukum*, vol. 26, no. 1, pp. 20–30, Apr. 2025, doi: 10.35315/dh.v26i1.10097.

[3] E. Simanjuntak, “Peran Yurisprudensi dalam Sistem Hukum di Indonesia,” *Jurnal Konstitusi*, vol. 16, no. 1, p. 83, Apr. 2019, doi: 10.31078/jk1615.

[4] E. Q. Nuranti, E. Yulianti, and H. S. Husin, “Predicting the Category and the Length of Punishment in Indonesian Courts Based on Previous Court Decision Documents,” *Computers*, vol. 11, no. 6, p. 88, May 2022, doi: 10.3390/computers11060088.

[5] Nazwa Aziza Berliana Putri, Dona Raisa Monica, and Nikmah Rosidah, “Urgensi Penegakan Hukum Terhadap Pelaku ResidiveTindak Pidana Pencurian Dengan Pemberatan,” *Judge : Jurnal Hukum*, vol. 4, no. 01, Feb. 2025.

[6] L. Hakim and S. Manullang, “Pertimbangan Hakim Terhadap Tindak Pidana Pencurian Dalam Keadaan Memberatkan (Putusan Nomor:108/PID.B/2023/PN LIW),” *JALAKOTEK: Journal of Accounting Law Communication and Technology*, vol. 1, no. 2, pp. 114–122, Jul. 2024, doi: 10.57235/jalakotek.v1i2.2161.

[7] L. Ikawati *et al.*, “Masa Depan Penegakan Hukum Indonesia: Sistem Peradilan Pidana Berbasis Kecerdasan Buatan (AI),” 2024, doi: 10.62383/prosemnashuk.v1i1.19.

[8] Maulidya Prastita Syah, Ajeng Puspa Wardani, Mohammad Idhom, and Trimono, “Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks,” *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, pp. 47–59, Jul. 2025, doi: 10.47709/dsi.v5i1.6021.

[9] T. D. Salma, M. F. Kurniawan, R. Darmawan, and A. Basri, “Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang Pertama,” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 2, pp. 757–764, Jan. 2025, doi: 10.35870/jtik.v9i2.3535.

[10] M. Amien, G. Frendi Gunawan, and K. Kunci, “ELANG: Journal of Interdisciplinary Research BERT dan Bahasa Indonesia: Studi tentang Efektivitas Model NLP Berbasis Transformer”.

[11] E. A. Junita and R. R. Suryono, “ANALISIS SENTIMEN HATE SPEECH MENGENAI CALON WAKIL PRESIDEN INDONESIA MENGGUNAKAN ALGORITMA BERT,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2042–2053, Nov. 2024, doi: 10.29100/jipi.v9i4.5625.

[12] M. Adrinta Abdurrazzaq and E. Lesmana Tjiong, “Analisis Sentimen KUHP Baru Pada Data Twitter Menggunakan BERT,” *Jurnal Komunikasi, Sains dan Teknologi*, vol. 1, no. 2, pp. 83–88, Dec. 2022, doi: 10.61098/jkst.v1i2.10.

[13] D. F. Sjoraida, B. W. K. Guna, and D. Yudhakusuma, “Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers),” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 8, no. 2, pp. 393–404, Apr. 2024, doi: 10.35870/jtik.v8i2.1580.

[14] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, “IndoBERT Model Analysis: Twitter Sentiments on Indonesia’s 2024 Presidential Election,” *Journal of Applied Informatics and Computing*, vol. 8, no. 1, pp. 7–12, Jul. 2024, doi: 10.30871/jaic.v8i1.7440.

[15] F. Basbeth and D. H. Fudholi, “Klasifikasi Emosi Pada Data Text Bahasa Indonesia Menggunakan Algoritma BERT, RoBERTa, dan Distil-BERT,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 1160, Apr. 2024, doi: 10.30865/mib.v8i2.7472.

[16] A. R. Hanum *et al.*, “Analisis Kinerja Algoritma Klasifikasi Teks Bert dalam Mendeteksi Berita Hoaks,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 3, pp. 537–546, Jul. 2024, doi: 10.25126/jtiik.938093.

[17] F. Fajri, B. Tutuko, and S. Sukemi, “Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter,” *JUSIFO (Jurnal Sistem Informasi)*, vol. 8, no. 2, pp. 71–80, Dec. 2022, doi: 10.19109/jusifo.v8i2.13885.

[18] N. Limsopatham, “Effectively Leveraging BERT for Legal Document Classification,” in *Proceedings of the Natural Legal Language Processing Workshop 2021*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 210–216. doi: 10.18653/v1/2021.nllp-1.22.

[19] A. R. Hanum *et al.*, “Analisis Kinerja Algoritma Klasifikasi Teks Bert dalam Mendeteksi Berita Hoaks,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 3, pp. 537–546, Jul. 2024, doi: 10.25126/jtiik.938093.

[20] E. Subowo, “Implementasi Pembelajaran Mendalam dalam Klasifikasi Sentimen Ulasan Aplikasi: Evaluasi Model BERT, LSTM, dan CNN,” *Jurnal Surya Informatika*, vol. 14, no. 2, pp. 66–70, Nov. 2024, doi: 10.48144/suryainformatika.v14i2.1973.

[21] B. Betesda, H. Purwanto, H. Nuryadi, D. Sinaga, S. J. Al Din, and D. Y. Al Afghani, “ANALISA SENTIMEN DATA ULASAN PADA GOOGLE PLAY DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE,” *Jurnal Sains dan Teknologi ISTP*, vol. 22, no. 01, pp. 08–15, Dec. 2024, doi: 10.59637/jsti.v22i01.423.

[22] C. Y, P. Kiran, and M. P B, “The Novel Method for Data Preprocessing CLI,” *Advances in Intelligent Systems and Technologies*, pp. 117–120, Dec. 2022, doi: 10.53759/aist/978-9914-9946-1-2\_21.

[23] R. A. A. Renal, Syariful Alam, and Moch Hafid T, “KOMPARASI PAYMENT DIGITAL UNTUK ANALISIS SENTIMEN BERDASARKAN ULASAN DI GOOGLE PLAYSTORE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE,” *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 118–128, Aug. 2023, doi: 10.55123/storage.v2i3.2337.

[24] N. C. Ramadhan, H. H. H, T. Rohana, and A. M. Siregar, “Optimasi Algoritma Machine Learning Menggunakan Seleksi Fitur Xgboost Untuk Klasifikasi Kanker Payudara,” *TIN: Terapan Informatika Nusantara*, vol. 5, no. 2, pp. 162–171, Jul. 2024, doi: 10.47065/tin.v5i2.5408.

[25] R. Oktafiani, A. Hermawan, and D. Avianto, “Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning,” *Jurnal Sains dan Informatika*, pp. 19–28, Jun. 2023, doi: 10.34128/jsi.v9i1.622.

[26] M. S. Islam and L. Zhang, “A Review on BERT: Language Understanding for Different Types of NLP Task,” Jan. 26, 2024. doi: 10.20944/preprints202401.1857.v1.

[27] G. F. Situmorang and R. Purba, “Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan IndoBERT,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 649–661, Sep. 2024, doi: 10.47065/bits.v6i2.5496.

[28] D. Nuryadi *et al.*, “FINE TUNING INDOBERT UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI TIKET.COM DI GOOGLE PLAY STORE,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3577–3583, Apr. 2025, doi: 10.36040/jati.v9i2.13204.

[29] M. I. Salih, S. M. Mohammed, A. Kh. Ibrahim, O. M. Ahmed, and L. M. Haji, “Fine-Tuning BERT for Automated News Classification,” *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22953–22959, Jun. 2025, doi: 10.48084/etasr.10625.

[30] Q. T. Nguyen, T. L. Nguyen, N. H. Luong, and Q. H. Ngo, “Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews,” in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, IEEE, Nov. 2020, pp. 302–307. doi: 10.1109/NICS51282.2020.9335899.

[31] A. F. AlShammari, “Implementation of Model Evaluation Using Confusion Matrix in Python,” *Int J Comput Appl*, vol. 186, no. 50, pp. 42–48, Nov. 2024, doi: 10.5120/ijca2024924236.

[32] M. G. Putra, I. Rizal Setiawan, and D. Indrayana, “Analisis Sentimen Terhadap Isu Kecurangan Pemilu 2024 Pada Platfom Twitter (X) Dengan Metode Naive Bayes Multinomial Dan Cosine Similiarity,” *Jurnal Sintaks Logika*, vol. 5, no. 1, pp. 20–31, Feb. 2025, doi: 10.31850/jsilog.v5i1.3562.