Automatic Summarization of Court Decision Documents Over Narcotic Cases Using BERT

Galih Wasis Wicaksono [a, *], Sheila Fitria Al'asqalani [a], Yufis Azhar [a], Nur Putri Hidayah [b], Andreawan [a]

[a] *Department of Informatics, University of Muhammadiyah Malang, Malang, Indonesia*
[b] *Department of Law, University of Muhammadiyah Malang, Malang, Indonesia*
Corresponding author: *[*]galih.w.w@umm.ac.id*

Abstract— **Reviewing court decision documents for references in handling similar cases can be time-consuming. From this perspective, we need a system that can allow the summarization of court decision documents to enable adequate information extraction. This study used 50 court decision documents taken from the official website of the Supreme Court of the Republic of Indonesia, with the cases raised being Narcotics and Psychotropics. The court decision document dataset was divided into two types, court decision documents with the identity of the defendant and court decision documents without the defendant's identity. We used BERT specific to the IndoBERT model to summarize the court decision documents. This study uses four types of IndoBert models: IndoBERT-Base-Phase 1, IndoBERT-Lite-Bas-Phase 1, IndoBERT-Large-Phase 1, and IndoBERT-Lite-Large-Phase 1. This study also uses three types of ratios and ROUGE-N in summarizing court decision documents consisting of ratios of 20%, 30%, and 40% ratios, as well as ROUGE1, ROUGE2, and ROUGE3. The results have found that IndoBERT pre-trained model had a better performance in summarizing court decision documents with or without the defendant's identity with a 40% summarizing ratio. The highest ROUGE score produced by IndoBERT was found in the INDOBERT-LITE-BASE PHASE 1 model with a ROUGE value of 1.00 for documents with the defendant's identity and 0.970 for documents without the defendant's identity at a ratio of 40% in R-1. For future research, it is expected to be able to use other types of Bert models such as IndoBERT Phase-2, LegalBert, etc.**

Keywords— **Court decision documents; BERT; document summarization; extractive summarization.**

I. INTRODUCTION

In the legal context, the term court decision refers to the process by which the court decides on legal disputes and records this process. The term case is also used synonymously. The source is the judge's correct decision. Based on data from the website of the Indonesian supreme court, www.mahkamahagung.go.id, the total number of court decisions is 2,400,121, with an average of 206,832 new decisions per year. Until December 2021, the number of decision documents on narcotic and psychotropic cases on the Indonesian Supreme Court's decision directory website had reached 276349 decisions and is still growing. Data on the growth in the number of court decisions shows that as long as there is a legal process, the number of decision documents will continue to increase. Many case decision documents show many narcotic and psychotropic abuse in Indonesia, and people still need to understand the law's impact. Decision documents also have lengthy systematics that takes more

work to understand. Therefore, it is necessary to summarize the narcotic case decision document [1].

Automatic document summarization aims to transform the entered text into a condensed form to present the most critical information to users [2]. The undertaking is regularly partitioned into two ideal models: abstractive summarization and extractive summarization. In the abstractive summarization model, target summaries contain words or expressions that were not in the first content. This model required different content-changing tasks to create words or phrases that were not in the original text. In contrast, extractive approaches form summaries by copying and concatenating the most important sentences in a document [3].

Automatic summarization has previously been carried out in a study [4]. In this case, the data used in the automated operation is a single Arabic document collected independently and translated from English to Arabic using Google translate. This research summary uses a combination of statistical and semantic features by utilizing several pre-processing such as Tokenization, Normalization, Stop-word

removal, and Stemming. The results given to each feature are measured using ROUGE-2, the result of the F-score of 0.617. This result shows that the proposed machine learning method has a better impact when compared to Naive Bayes, SVM (with RBF kernel), two-layer neural network, J48, and Random Forest in recall and F-score. The average increase of respectively 33% and 14%.

A research conducted by Meena et al [5] also uses supervised and unsupervised learning algorithms. In this case, the data used for summarizing is collected from Amazon Product reviews using the TFRSP (Text Frequency Ranking Sentence Prediction) method. After the data set is obtained, the data will be processed using the TF-IDF-TR (Term Frequency – Inverse Document Frequency – Text Rank) algorithm, an unsupervised learning algorithm to produce extractive summaries in the first phase. The extractive summary involves the seq2seq model (supervised learning algorithm) to obtain the second phase of the abstractive summary, which includes the training and test dataset. The second phase aims to obtain a practical summary when performance is calculated using the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) scores. The results obtained through the TFRSP algorithm by combining unsupervised (extractive summary) and supervised (abstract summary) techniques provide an accuracy increase of 87.58% % when compared to traditional methods in text summarization.

Automatic summarization of multiple documents sharing the same information was carried out in a study [6]. In multi-document summaries, two or more introductory sentences may share similar information. Incorporating all the meaningful sentences into the summary result will give you too much information and may lead to repetition of information. This study aims to reduce similar sentences from multi-documents that share identical information to obtain a more concise text summary. The data used in this study are online news articles sourced from Tribunnews (www.tribunnews.com) and Detik (detik.com). The number of articles used was thirty, divided into six categories. Each article category consists of 200 to 700 words. After the data is ready, the next step is the TextRank algorithm to extract important sentences using text similarity measurements. The TextRank process will provide summarized results, but the summary results still contain similar sentences. This study uses Maximal Marginal Relevance (MMR) calculation to reduce similar sentences. The results of the MMR are the final summary which will then be evaluated using ROUGE-1 and ROUGE-2 with an average F-score of 0.5103 and 0.4257, respectively.

In this research, the summary method is extractive single-document summarization of court decision documents on narcotic and psychotropic substances. However, due to many narcotic types and crimes, this study focused on court decision documents legally binding with the evidence of ecstasy and the indictment of Article 114 of Law of the Republic of Indonesia No.35 of 2009.

The basis of this study's algorithm model is referred to as BERT (Bidirectional Encoder Representation form Transformer). BERT is conceptually simple and empirically robust, and it obtains new cutting-edge results on eleven natural language processing tasks. BERT is designed for practicing deep bidirectional representations of unlabeled text by co-conditioning the left and right context across all layers. As a result, the pre-existing BERT model can enhance various information with just one additional layer to create a sophisticated model for a task [7]. BERT is also beneficial for applying pre-trained language models such as BERT to extractive summary models [8]. This study aims to implement IndoBert pre-trained in Indonesian automatic document summarization, which is applied to narcotic and psychotropic drug decisions.

II. MATERIALS AND METHOD

The sophisticated extractive text summarization method considers a document (or collection of documents) as a set of textual units (e.g., sentences, clauses, phrases). It formulates summaries as a combinatorial optimization problem, i.e., selecting subsets of a group of textual units that maximize purpose without breaking any length restrictions [9]. For example, Table I shows that a court decision document has a structure considering a set of textual units [10].

TABLE I
THE STRUCTURE OF THE COURT DECISION DOCUMENT

Head of verdict and title Identity of the accused	Opening
Case history History of filing cassations Cassation reasons Legal considerations and dissenting opinions Verdict	Contents
Closing paragraph and signature of the panel of judges and the substitute clerk	Closing

The method implemented in this research was extractive summarization with the BERT algorithm. BERT is a new language representation model that generates a pre-train model of bidirectional (two-way) representations of unlabeled text by co-conditioning from both contexts across all layers [7]. BERT was developed on layers of the two-way Transformer encoder [11], or, in other words, BERT is a stack of encoder transformers so that the trained BERT model can be adjusted by adding only one output layer. This study chooses BERT because it is simple to apply, empirically powerful [7], highly beneficial [8], and it can further boost the performance of extractive summarization [3].

Some studies have applied BERT in text summarization. For example, the first research summarised the CNN / Daily Mail and New York Times dataset using BERT and Transformer Based Decoder. This study has resulted in the CNN / Daily Mail dataset, with the value of R1 accounting for 41.71 and R2 scoring 19.49, while in the New York Times dataset, the value of R1 accounted for 45.33 and R2 was 26.5 [12].

Another research was focused on abstractive and extractive single-document summarization on the CNN / Daily Mail dataset using BERT as the pre-trained encoder. Evaluation using ROUGE F1 obtained ROUGE-1 results of 41.76, ROUGE-2 amounting to 19.31, and ROUGE-L of 38.86 [13]. Other research tried to combine summarization models using BERT and OpenAI GPT-2, providing abstract and comprehensive keyword-based information from a collection

of scientific articles on the COVID-19 Open Research Dataset Challenge. The results of extractive summarization produced a higher ROUGE score than abstractive summarization [14].

The last research was creating a model to summarize the lecturer's material using BERT for extractive summarization and K-Means clustering to identify sentence selection choices. The extractive summarization is still imperfect, but compared to TextRank, BERT has a steady increase in summary quality and the integration of context with meaningful sentences [15]. Various other studies using BERT for text summarization in various languages have also been carried out [16]–[19].

In case summarization in the Indonesian language, IndoBERT provides pre-trained models for document summarization [20]. Some research has been done to prove the IndoBERT accuracy [21], but few studies have used BERT to summarize legal documents. Referring to these problems, the authors developed this research based on previous research using different datasets and pre-training models. Court decision documents were used as datasets, and the monolingual pre-trained BERT algorithm was used to summarize documents. This study aims to see the performance of IndoBERT in summarizing legal documents. The research was conducted in several stages, as shown in Figure 1 below.
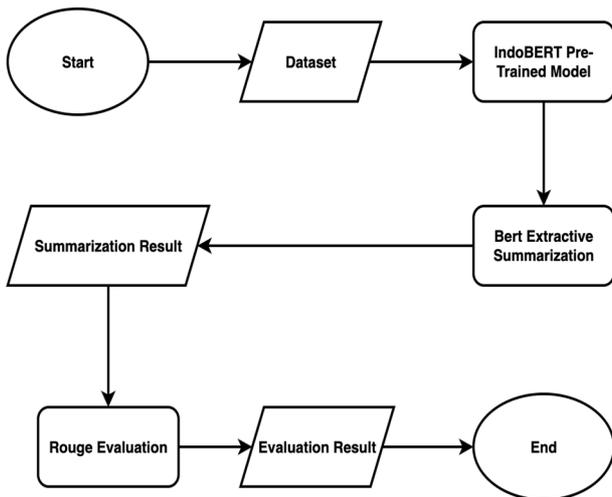
Fig. 1 Research flow

A. Dataset

The dataset used in this study uses legal documents dealing with Narcotics and Psychotropics cases. Narcotics and psychotropics are a class of drugs that are managed under strict laws by the government. The law is strict on these drugs because they have a great potential for abuse and dependency effects. Narcotics and Psychotropics are included in special criminal acts according to Indonesian law. In addition, corruption, money laundering, terrorism, and others are also included in special crimes. This research focuses on narcotics and psychotropics because these crimes are quite numerous and can be committed by all ages, both young and old. The strict law regarding narcotics and psychotropics also aims to prevent damage to the next generation of Indonesian.

The legal basis for narcotics is regulated in the Law of the Republic of Indonesia No. 35 of 2009. According to article 1,

paragraph 1 of Law no. 35 of 2009 concerning narcotics, "Narcotics are substances or drugs derived from plants or non-plants, both synthetic and semi-synthetic, which can cause a decrease or change in consciousness, loss of taste, reduce to eliminate pain, and can cause dependence, which is divided into groups as attached to this Law." Narcotics crime consists of 4 groups based on the type of narcotics, and each has provisions for sanctions. Article 127, paragraph 1 explains the punishment that the perpetrator will serve. The four categories are Narcotics Group I will get a maximum prison sentence of 4 years. Narcotics class II for a maximum of 2 years, and Group III for a maximum of 1 year in prison.

The legal basis for Psychotropics is regulated in the Law of the Republic of Indonesia No. 5 of 1997. According to article 1, paragraph 1 of Law no. 5 of 1997 concerning Psychotropics, "Psychotropics are substances or drugs, both natural and synthetic, which are not narcotics, which have psychoactive properties through a selective influence on the central nervous system which causes certain changes in mental activity and behavior." Psychotropic users are not immediately imprisoned but will be rehabilitated, and this provision is stated in article 39, paragraph 1. Apart from the disadvantages of narcotics and psychotropics, both also have good benefits in the medical field if used according to the provisions.

A dataset of court decision documents was taken from the official website of the Supreme Court of the Republic of Indonesia. However, due to a large number of narcotics decision documents, this study only took a sample of 50 documents resulting from decisions on special criminal cases of narcotic and psychotropic substances containing the criminal charges under Article 114 and evidence of ecstasy, which has permanent legal force.

PDF is the format document utilized as a dataset in this research. Then the document is converted into TXT format using phyton before being processed into the model. Changing data format utilizes a library provided by Pdfminer in PDFResourceManager, PDFPageInterpreter, PDFPage, TextConverter, and LAParams. The purpose of using this library is to change the file format from pdf to TXT and organize the sentence structure.
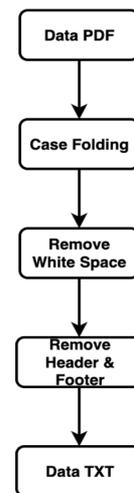
Fig. 2 PDF to TXT

The converting data PDF into TXT goes through several stages, as shown in Figure 2. The initial step is case folding. It is a simple text processing, and Despite simple, case folding works effectively. The purpose of case folding is to change all uppercase letters in a document to lowercase. In this process, only the letters a to z can be accepted in the case folding system. Characters other than these letters will be omitted and considered a delimiter. In the case folding process, we don't use external libraries but use the functions and modules available in Python.

The extracted data sometimes has double spacing that can ruin the structure and interface of the document that will be processed in a model. Therefore, this approach can affect model performance in summarizing an automatic document. This research utilizes a function to remove white space to solve the double spacing problem.

Not only white space and case folding, headers, and footers on the documents used in this study were also removed. The reason for eliminating headers and footers is that every legal document must have them. The purpose of automatic summarization of decision documents focuses on the document's contents.

Data already in TXT format is then pre-processed again before input into the BERT model. The second pre-processing stage in this study was divided into two types. Version 1 used pre-processing, included in the IndoBERT library, and version 2 added stemming and case folding to existing pre-processing. Pre-processing was carried out on a dataset that was separated into a document with the defendant's identity and without the defendant's identity.

1) Tokenization: Tokenization is the first step in most text-processing jobs [22]., tokenization is the task of separating the full-text string into a separate list of words [23]. Another study defines tokenization as a type of lexical analysis that breaks down text into words, phrases, symbols or other meaningful elements called tokens [24].

2) Word Embedding: IndoBERT was trained with 4 billion vocabularies and 250 million sentences incorporated in the Indo4B dataset [25]. First, this dataset was used to build the fastText model. Then the fastText embeddings were pre-trained using a skip-gram word representation and generated a 300-dimensional embedding vector [26]. Then all the embeddings needed for each task were created from the previous pre-trained FastText and included all the vocabulary [20]. The result of tokenization was then converted into a vector using word embedding.

B. BERT Extractive Summarization

This study uses a pre-trained model modified to monolingual (Indonesian). We used four pre-trained models in this research, namely 1) IndoBERT-BASE Phase 1; 2) IndoBERT-Lite-Base Phase 1; 3) IndoBERT-Large Phase 1; 4) IndoBERT-Lite-Large Phase 1. This pre-training model was chosen because the monolingual model learns sentiment-level semantics better in everyday and formal languages than the multilingual model. The discrepancy between one pre-trained model and others is in the number of vocabularies. The experimental results also showed that the larger model had a performance advantage over the smaller model [20].

In addition to the number of vocabularies, there are several characteristics of Bert's pre-trained model, which are also differentiators for each model. The different characteristics of each model are the number of layers, hidden units, attention heads, and parameters [27]. These characteristics can be seen in Table II.

TABLE II
BERT MODEL CHARACTERISTICS

Model	Layers	Hidden Units	Attention Head	Parameters
Indobert-base	12	768	12	124.5 M
Indobert-large	24	1024	16	335.2 M
Indobert-lite-base	12	768	12	11.7 M
Indobert-lite-large	24	1024	16	17.7 M

The process of summarizing using BERT was split into two cycles. The first was summarizing the court decision document with the defendant's identity, while the second involved the summary of court decision documents without the defendant's identity. Documents were divided into two types because identity is patent information that cannot be summarized. The separation was performed to see the performance of BERT in extracting meaningful information in the document unless there was identity in it.

This study tries three types of summary ratios (20%, 30%, and 40%) to see the best performance of BERT in extracting legal documents. The ratio value was determined to avoid the difference in the number of sentences that were too large between the system summary results and the expert summary reference.

C. Testing Scenario

The summarization result measurement was performed using Recall-Oriented Understanding for Gisting Evaluation (ROUGE) [28]. The ROUGE score will measure the similarity between the two summary objects produced by adding the number of overlapping units [29]. Therefore, we chose ROUGE-N, as represented in (1).

$$ROUGE - N = \frac{\sum_{S \in Summ_{ref}} \sum_{gram_n \in S} count_{match}(gram_n)}{\sum_{S \in Summ_{ref}} \sum_{gram_n \in S} count(gram_n)} \quad (1)$$

Equation (1) n is the length of $n - grams$, Ref represents a pair (set) of summary references. $Countmatch(gram_n)$ represents the maximum number of n-grams co-occurring in the summary model results and sets of summary texts. $Cou(gram_n)$ represents the n-gram number in the summary text. The scoring process was performed separately based on the number of n [R1 (unigram), R2 (bigram), R3, R4 (multigram)] and the ratio [30].

III. RESULTS AND DISCUSSION

A. Data Processing

The total data collected from the court's decision directory website was 50 court decision documents on specific crimes of Narcotics and Psychotropics. Before converting the document dataset from pdf to txt, the watermark, header, and footer were manually removed. Furthermore, the dataset was divided into two. Document A represents a document with an

identity, and document B represents a document without an identity. After the data processing, the data were converted before being submitted into the system.

D. Evaluation

The evaluation scenario used in this study was intended to test each pre-trained model for each summary ratio on both types of documents. The number of NGRAM used was N=4. The tests were conducted using ROUGE-N and carried out on both document datasets. Doc A represents the court decision document with the defendant's identity, and Doc B is the court decision document without the defendant's identity. The ideal ROUGE score was identified as a balanced ROUGE score between Doc A and Doc B, occurring at the same summarization ratio and the same number of NGRAMs.

TABLE III
ROUGE SCORE COMPARISON ON INDOBERT-BASE PHASE 1

ROUGE AVG	RATIO 20%		RATIO 30 %		RATIO 40%	
	DOC A	DOC B	DOC A	DOC B	DOC A	DOC B
R-1	0.454	0.435	0.614	0.603	0.806	0.785
R-2	0.311	0.306	0.414	0.417	0.530	0.533
R-3	0.255	0.251	0.339	0.341	0.428	0.432
R-4	0.219	0.216	0.290	0.293	0.365	0.369

TABLE IV
ROUGE SCORE COMPARISON ON INDOBERT-LITE-LARGE PHASE 1

ROUGE AVG	RATIO 20%		RATIO 30 %		RATIO 40%	
	DOC A	DOC B	DOC A	DOC B	DOC A	DOC B
R-1	0.478	0.462	0.680	0.664	0.895	0.875
R-2	0.327	0.327	0.463	0.463	0.607	0.611
R-3	0.267	0.268	0.376	0.377	0.494	0.499
R-4	0.227	0.230	0.320	0.323	0.421	0.427

TABLE V
ROUGE SCORE COMPARISON ON INDOBERT-LARGE PHASE 1

ROUGE AVG	RATIO 20%		RATIO 30 %		RATIO 40%	
	DOC A	DOC B	DOC A	DOC B	DOC A	DOC B
R-1	0.495	0.480	0.696	0.674	0.913	0.889
R-2	0.327	0.333	0.459	0.460	0.600	0.606
R-3	0.263	0.272	0.367	0.371	0.478	0.489
R-4	0.223	0.232	0.311	0.315	0.405	0.412

TABLE VI
ROUGE SCORE COMPARISON ON INDOBERT-LITE-BASE PHASE 1

ROUGE AVG	RATIO 20%		RATIO 30 %		RATIO 40%	
	DOC A	DOC B	DOC A	DOC B	DOC A	DOC B
R-1	0.527	0.510	0.757	0.726	**1.00**	**0.970**
R-2	0.361	0.362	0.506	0.504	0.668	0.666
R-3	0.296	0.299	0.411	0.412	0.541	0.543
R-4	0.253	0.258	0.351	0.354	0.463	0.467

Based on testing on four pre-trained models in the IndoBERT-Base Phase 1 model, as shown in Table III, the ideal ROUGE score was found at all summary ratios of 20%, 30%, and 40% for R-1. The highest ROUGE score for Doc A also fell under this criterion. However, Doc A's highest ROUGE score was at R-1 at a 40% summary ratio. In the IndoBERT-Lite-Large Phase 1 model, as shown in Table IV, an increase occurred in the ideal ROUGE score, which still occurred at a summary ratio of 40% in R-1. Meanwhile, the highest ROUGE score increased for each Doc A and Doc B but was still within the same criteria.

In the IndoBERT-Large Phase 1 model shown in Table V, there was a decrease in the ideal ROUGE score; however, the decreased score was insignificant and still occurred at the 30% summary ratio in R-1. In the ROUGE scores generated in Table VI, the pre-trained model used was IndoBERT-Lite-Base Phase 1. This model found the ideal and highest ROUGE score at a summary ratio of 40% at R-1.

The 30% and 40% summary ratios had the pattern to see the best NGRAM that can be used to evaluate the system summary results with expert summaries for both types of documents. The greater the value of NGRAM, the evaluation result will get bigger. R-1 had the best evaluation value for these two ratios in each record because ROUGE-N paid attention to the similarity of words and the order of words in the sentence. Meanwhile, the 40% ratio did not have a specific pattern to see which test NGRAM had the best value for evaluating the document summary results with the expert summary. However, the best result from the ROUGE-N evaluation in the 40% ratio was R-1 for the court decision document with identity and without the defendant's identity.

Furthermore, we will examine which document types had the best BERT summary performance. In the ratio of 20% and 30%, the ROUGE score had a random pattern, making it difficult to determine what type of decision document was to be used as input for the BERT summarization program. Meanwhile, at the 40% ratio, it is clear that the ROUGE score for doc A (documents with the defendant's identity) was higher than doc B (documents without the defendant's identity). With a 40% ratio of BERT summarization, Doc A had the highest ROUGE score among the other ratios and document types.

From the results of table VI analysis, it can be concluded that 40% is the most proper ratio to use in summarizing court decision documents using IndoBERT pre-trained model. At the same time, R-1 is the best NGRAM evaluation if ROUGE-N is used as the evaluation method. To improve the summarization result, separation of the defendant's identity from the document's contents is necessary before entering the system to avoid unimportant information getting extracted and included in the summary results.

B. Summary Result

The summary A and B results in every ratio can still be read well and extract essential points in the court decision document. However, there is some unimportant information that goes into the summary results. Based on the results of this study, BERT can be implemented in court to improve performance in digitizing document filtering based on evidence or certain types of cases.

Some errors in extracting sentences in summarizing can mess up the summary results. It may occur because of the document watermark, header, footer removal, and manual document conversion. In addition, some unimportant

information was not completely erased and mixed after being converted to a .txt file.

If you want to do similar research, you should be more careful about cleaning and converting data processes. For further similar research, it is essential to bear in mind that cleaning and converting should be carefully performed in data processing, recalling that the template of the court decision document of the Republic of Indonesia has a watermark on every page. Without proper cleaning, the system will fail to read the data properly, and summary results will be compromised.

IV. Conclusion

The rapid development of digital storage technology has triggered a surge in electronic documents, one of which is court decision documents. Many documents can be a problem for law enforcement officers such as lawyers or judges who intend to find or compare similar cases. This research used the BERT (Bidirectional Encoder Representation Form Transformer) algorithm to summarize documents automatically. This study used a dataset of 50 documents of criminal court decisions specifically for narcotic and psychotropic substances containing criminal charges under Article 114 and evidence of ecstasy. The documents used involved only Indonesian language documents that have permanent legal force. From the results of this study, it was found that IndoBERT-Lite-Base Phase 1 had a better performance in summarizing court decision documents with or without the defendant's identity with a 40% summarizing ratio, and the best evaluation method to evaluate the summarization result is ROUGE-N with N=1. This research is expected to serve as a reference to create an excellent automatic document summary system and to further serve as a basis for decision-making for law enforcement officers conducting investigations over similar cases.

There are many opportunities for further research using a monolingual pre-trained encoder such as IndoBert. BERT algorithm is still classified as a new language processing system and can be fine-tuned according to tasks such as summarizing, classifying, and answering questions. In the case of conducting similar research, it is important to carefully clean watermarks, headers, and footers at the early stages and convert documents from .pdf to .txt files. Adding processes in the pre-processing stage and different types of datasets can also be performed to see BERT's performance summarizing other types of text or documents.

References

[1] F. Solihin and I. Budi, "Recording of law enforcement based on court decision document using rule-based information extraction," in *2018 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2018*, 2019, pp. 349–354, doi: 10.1109/ICACSIS.2018.8618187.

[2] M. Campr and K. Ježek, "Comparing semantic models for evaluating automatic document summarization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9302, no. 1, pp. 252–260, doi: 10.1007/978-3-319-24033-6_29.

[3] Y. Liu, "Fine-tune BERT for Extractive Summarization," *Cornell Univ.*, vol. 1, no. 1, pp. 1–6, 2019.

[4] N. Baruah, S. K. Sarma, and S. Borkotokey, "A Single Document Assamese Text Summarization Using a Combination of Statistical Features and Assamese WordNet," in *Advances in Intelligent Systems and Computing*, 2021, vol. 1199, pp. 125–136, doi: 10.1007/978-981-15-6353-9_12.

[5] S. Meena, M. Ramkumar, R. Asmitha, and G. Emil Selvan, "Text Summarization Using Text Frequency Ranking Sentence Prediction," in *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Sep. 2020, pp. 1–5, doi: 10.1109/ICCCSP49186.2020.9315203.

[6] D. Gunawan, S. H. Harahap, and R. Fadillah Rahmat, "Multi-document Summarization by using TextRank and Maximal Marginal Relevance for Text in Bahasa Indonesia," in *Proceeding - 2019 International Conference on ICT for Smart Society: Innovation and Transformation Toward Smart Region, ICISS 2019*, 2019, pp. 1–5, doi: 10.1109/ICISS48059.2019.8969785.

[7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186.

[8] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," 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*, 2020, pp. 3730–3740, doi: 10.18653/v1/d19-1387.

[9] T. Hirao, Y. Yoshida, M. Nishino, N. Yasuda, and M. Nagata, "Single-document summarization as a tree knapsack problem," in *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2013, pp. 1515–1520.

[10] D. Djamal, S. Ono, and D. A. Wicaksana, "Laporan Penelitian Penyederhanaan Format Putusan," 2016.

[11] L. Gong, D. He, Z. Li, T. Qin, L. Wang, and T. Y. Liu, "Efficient training of BERT by progressively stacking," in *36th International Conference on Machine Learning, ICML 2019*, 2019, vol. 2019-June, pp. 4202–4211.

[12] H. Zhang, J. Cai, J. Xu, and J. Wang, "Pretraining-based natural language generation for text summarization," in *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*, 2019, pp. 789–797, doi: 10.18653/v1/k19-1074.

[13] R. Wei, H. Huang, and Y. Gao, "Sharing Pre-trained BERT Decoder for a Hybrid Summarization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11856 LNAI, pp. 169–180, doi: 10.1007/978-3-030-32381-3_14.

[14] V. Kieuvongngam, B. Tan, and Y. Niu, "Automatic Text Summarization of COVID-19 Medical Research Articles using BERT and GPT-2," pp. 1–13, 2020.

[15] D. Miller, "Leveraging BERT for Extractive Text Summarization on Lectures," pp. 1–7, 2019.

[16] W. Kai and Z. Lingyu, "Research on Text Summary Generation Based on Bidirectional Encoder Representation from Transformers," in *Proceedings - 2020 2nd International Conference on Information Technology and Computer Application, ITCA 2020*, Dec. 2020, pp. 317–321, doi: 10.1109/ITCA52113.2020.00074.

[17] H. F. Mahdi, R. Dagli, A. Mustufa, and S. Nanivadekar, "Job Descriptions Keyword Extraction using Attention based Deep Learning Models with BERT," Jun. 2021, doi: 10.1109/HORA52670.2021.9461296.

[18] S. R. K. Harinatha, B. T. Tasara, and N. N. Qomariyah, "Evaluating Extractive Summarization Techniques on News Articles," in *Proceedings - 2021 International Seminar on Intelligent Technology and Its Application: Intelligent Systems for the New Normal Era, ISITIA 2021*, Jul. 2021, pp. 88–94, doi: 10.1109/ISITIA52817.2021.9502230.

[19] H. Gupta and M. Patel, "Method of Text Summarization Using Lsa and Sentence Based Topic Modelling with Bert," in *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, Mar. 2021, pp. 511–517, doi: 10.1109/ICAIS50930.2021.9395976.

[20] S. Cahyawijaya *et al.*, "IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation," pp. 843–857, 2021, doi: 10.18653/v1/2021.emnlp-main.699.

[21] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Nov. 2021, pp. 757–770, doi: 10.18653/v1/2020.coling-main.66.

[22] K. Park, J. Lee, S. Jang, and D. Jung, "An Empirical Study of Tokenization Strategies for Various Korean NLP Tasks," pp. 1–10, 2020.

[23] J. A. Balazs and J. D. Velásquez, "Opinion Mining and Information Fusion: A survey," *Inf. Fusion*, vol. 27, no. 1, pp. 95–110, 2016, doi: 10.1016/j.inffus.2015.06.002.

[24] J. Atkinson, G. Salas, and A. Figueroa, "Improving opinion retrieval in social media by combining features-based coreferencing and memory-based learning," *Inf. Sci. (Ny).*, vol. 299, no. 1, pp. 20–31, 2015, doi: 10.1016/j.ins.2014.12.021.

[25] B. Richardson and A. Wicaksana, "Comparison of Indobert-Lite and Roberta in Text Mining for Indonesian Language Question Answering Application," *Int. J. Innov. Comput. Inf. Control*, vol. 18, no. 6, pp. 1719–1734, 2022, doi: 10.24507/ijicic.18.06.1719.

[26] A. Alwehaibi and K. Roy, "Comparison of Pre-Trained Word Vectors for Arabic Text Classification Using Deep Learning Approach," in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 2019, pp. 1471–1474, doi: 10.1109/ICMLA.2018.00239.

[27] H. R. Salim, C. De, N. D. Pratamaputra, and D. Suhartono, "Indonesian automatic short answer grading system," *Bull. Electr. Eng. Informatics*, vol. 11, no. 3, pp. 1586–1603, 2022, doi: 10.11591/eei.v11i3.3531.

[28] U. Naseem, I. Razzak, and P. W. Eklund, "A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter," *Multimed. Tools Appl.*, vol. 80, no. 28–29, pp. 35239–35266, 2021, doi: 10.1007/s11042-020-10082-6.

[29] J. M. Sanchez-Gomez, M. A. Vega-Rodríguez, and C. J. Pérez, "Sentiment-oriented query-focused text summarization addressed with a multi-objective optimization approach," *Appl. Soft Comput.*, vol. 113, p. 107915, 2021, doi: 10.1016/j.asoc.2021.107915.

[30] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 874, no. 1, pp. 1–7, doi: 10.1088/1757-899X/874/1/012017.