

Detection of Online Gambling Promotional Comments on YouTube Using IndoBERT

Yonky Pernando¹, Raymond Erz Saragih², Agus Suwandi³, Ilwan Syafrinal⁴, Eka Lia Febrianti⁵, Marfuah⁶

^{1,2,3}Department of Informatics Engineering, ^{4,5}Department of Software Engineering, ⁵Department of Information System
Universitas Universal, Batam, Riau Island, Indonesia

ABSTRACT

Online gambling promotional comments on YouTube have become increasingly widespread and are difficult to detect with conventional moderation systems because they use informal language, varied spellings, and disguised keywords. This study applied the IndoBERT Transformer model to detect such comments using a cross-category video dataset in order to improve the generalization capability of the model. The research followed the Cross-Industry Standard Process for Machine Learning framework, which covers business and data understanding, data preparation, modelling, evaluation, deployment, and monitoring and maintenance. A total of 210,461 comments were collected through the YouTube Data API from 14 video categories. Preprocessing included emoji-to-text conversion, character normalization, text cleaning with regular expressions, and stopword removal using PySastrawi, NLTK, and a custom stopword list. The learning rate, batch size, and number of epochs were optimized using Optuna. On the test data, the model achieved an accuracy of 0.99591, a precision of 0.96396, a recall of 0.93265, and an F1-score of 0.94805. These results indicated that IndoBERT was effective in detecting online gambling promotional comments on YouTube and generalized well across various comment patterns.

Keyword : IndoBERT, Online Gambling, YouTube, Transformer, Text Classification



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Yonky Pernando,
Department of Informatics Engineering,
Universitas Universal,
Batam 29444, Indonesia.
Email: yongkyfernando194@gmail.com

1. INTRODUCTION

Online gambling in Indonesia has reached an alarming level, with transactions recorded by the Indonesian Financial Transaction Reports and Analysis Center (PPATK) amounting to IDR 104.42 trillion in 2023[[1]. The socio-economic impact of this phenomenon on the economic stability of society is considerable, as the total circulation of money is estimated to reach IDR 327 trillion [2]. The government has blocked more than 500,000 gambling websites and applications to reduce crime and social conflicts that damage family life [3], [4]. However, gambling operators continue to adapt by using social media platforms with high user interaction, such as YouTube, to spread online gambling promotions on a large scale through comment sections, thereby avoiding direct blocking by the authorities [5], [6].

Automatic moderation has been explored through various solutions based on artificial intelligence, particularly machine learning and deep learning[7], [8]. Previous studies applied traditional algorithms, such as Random Forest on Twitter, which achieved an accuracy of 97.8%[9]. as well as rule-based and probabilistic methods, namely regular expressions (Regex) and Naive Bayes, for filtering YouTube comments, which achieved accuracies of 90.85% and 97%, respectively [10], [11]. Although these methods are effective at a certain scale, they often have difficulty capturing deep semantic context compared with models based on the attention mechanism [12]. In addition, architectures with recurrent layers, such as the hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM), have been used to protect the digital ecosystem from negative comments such as cyberbullying and spam. However, their performance remains limited in capturing long-range semantic relationships in short and informal comment texts [13], [14]. Transformer-based models such as BERT and IndoBERT have

proven effective, reaching an accuracy of up to 98.53% in classifying Indonesian texts, owing to their ability to understand context in both directions (bidirectional) [15].

Although IndoBERT has shown superior performance, a research gap remains in previous studies. The application of IndoBERT for detecting online gambling promotions on YouTube is still limited to narrow datasets, such as those focusing only on gaming videos [16], or small datasets that do not specify the video categories [17], [18]. This limited dataset coverage may cause representation bias, because narrow data do not reflect the characteristics of the wider platform audience [19]. Most previous studies also still rely on traditional recurrent models, such as the Bidirectional Gated Recurrent Unit (BiGRU), which are less effective in capturing long-range semantic relationships in short and informal comment texts [20].

This study contributes by implementing an IndoBERT model trained on a considerably larger dataset that covers a wide range of YouTube video categories. The cross-category dataset is intended to improve the ability of the model to recognize various patterns of online gambling promotional comments across different types of content. This approach is expected to provide a more comprehensive and adaptive moderation solution than rule-based methods or conventional architectures. The objective of this study is to evaluate the effectiveness of IndoBERT in detecting online gambling promotional comments on YouTube in order to produce a more reliable and adaptive detection system.

2. RESEARCH METHOD

The workflow of this study follows the Cross-Industry Standard Process for Machine Learning (CRISP-ML), a framework specifically designed to meet the requirements of developing and managing machine learning applications [19]. As shown in Fig. 1, CRISP-ML consists of six stages: business and data understanding, data preparation, modelling, evaluation, deployment, and monitoring and maintenance.

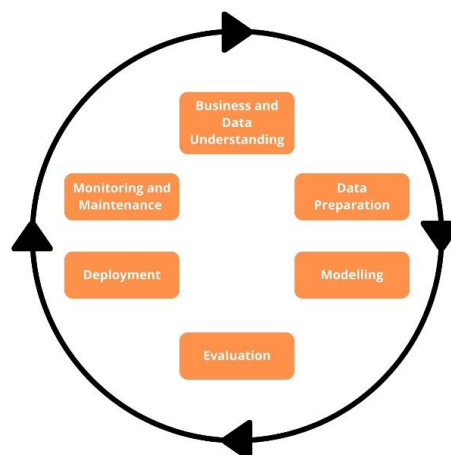


Fig 1. CRISP-ML process model [21]

The first stage is business and data understanding, in which the phenomenon of online gambling promotional comments on YouTube is analyzed. In this stage, comments from various YouTube video categories were collected and explored to overcome the dataset limitations found in previous studies.

The next stage is data preparation, which includes labelling the data into two classes and preprocessing the text to clean and normalize it. The dataset is then split into training, validation, and testing sets. Once the data are ready, the modelling stage uses IndoBERT to classify the comments. Hyperparameter optimization is performed to obtain the best model configuration, with the learning rate, batch size, and number of epochs as the optimized parameters.

The final stages cover evaluation, deployment, and monitoring and maintenance. Model performance is evaluated using a confusion matrix with accuracy, precision, recall, and F1-score as the metrics, which are calculated using (1)–(4), where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In the deployment stage, the model is integrated into a web-based interface built with the Streamlit framework. The monitoring and maintenance stage is carried out by entering various comment inputs directly and observing the classification results in real time to ensure that the system remains effective in recognizing new patterns.

3. RESULTS AND DISCUSSION

This section presents the results of each CRISP-ML stage together with a discussion of the findings. The results are presented in figures and tables to make them easier to understand.

A. Business and Data Understanding

At this stage, the problem and the characteristics of the data were examined. The problem addressed is the large number of online gambling promotional comments on YouTube that use informal terms and varied spellings, which makes them difficult to detect with conventional moderation systems, as illustrated in Fig. 2. Therefore, this study aims to apply a Transformer model to classify such comments and to evaluate its effectiveness.

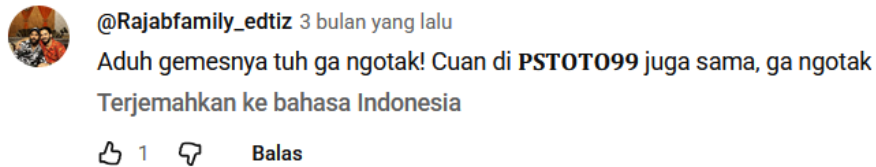


Fig 2. Examples of online gambling promotional comments on YouTube

The data were collected using the YouTube Data API from 14 video categories in the Indonesian region. The selection criteria were at least 100,000 views and 100 comments per video, with 20 videos per category taken from different channels. The resulting dataset consists of 210,461 comments in tabular format, with main columns such as comment_text, category_name, and comment_id. After the data were obtained, frequently occurring words that carry little information were identified to build a custom stopword list, as shown in Fig. 3.

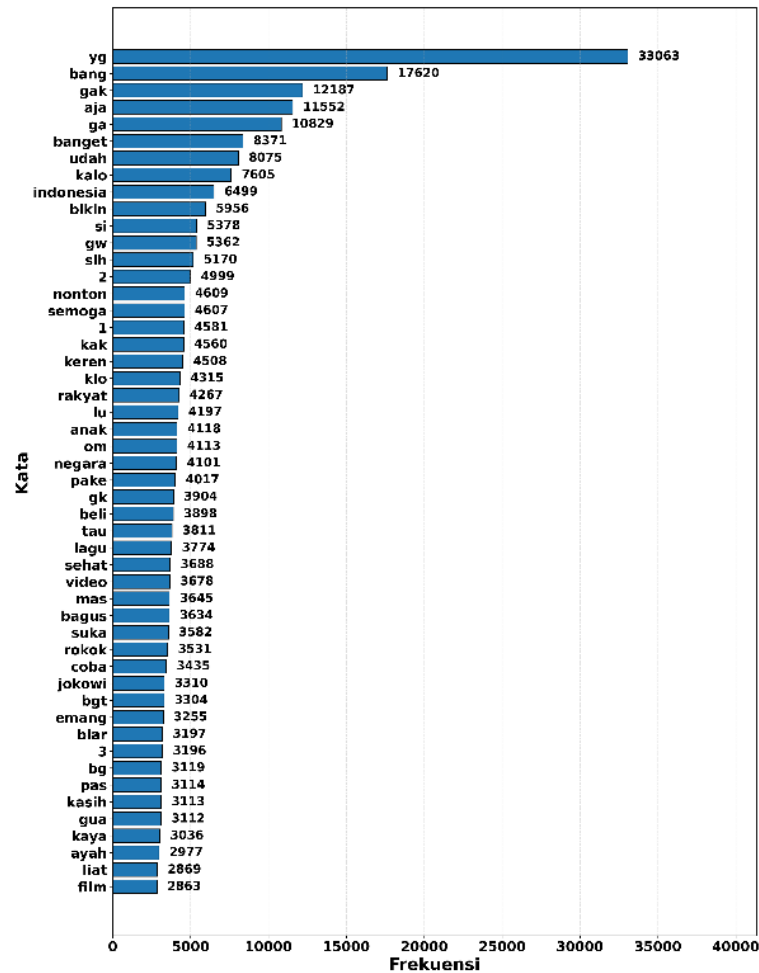


Fig 3. The 50 most frequent words in the dataset

Based on the 50 most frequent words, several words were selected as custom stopwords because they appear frequently but contribute little semantic meaning. These words are listed in Table 1.

Table 1. Words selected as custom stopwords

Word	Frequency	Word	Frequency
yg	33,063	klo	4,314
bang	17,620	lu	4,197
gak	12,187	om	4,113
aja	11,552	pake	4,017
ga	10,829	gk	3,904
banget	8,371	tau	3,811
udah	8,075	mas	3,645
kalo	7,604	bagus	3,634
bikin	5,956	suka	3,582
si	5,378	coba	3,435
gw	5,361	bgt	3,304
sih	5,170	emang	3,255
semoga	4,607	blar	3,197
kak	4,560	bg	3,119
keren	4,508		

The custom stopwords were determined by identifying high-frequency words that are general in nature and do not contribute significantly to classification, such as “yg” (an abbreviation of “yang”, meaning “which”), “gak” (“no”), “aja” (“just”), and “bang” (an informal form of address). In contrast, words that still carry contextual relevance, such as “beli” (“buy”), “rokok” (“cigarette”), and “sehat” (“healthy”), were

retained. In addition, numbers were not removed because they often appear in online gambling promotion patterns, for example in website names or specific codes.

B. Data Preparation

At this stage, the data were prepared for modelling through labelling, preprocessing, and dataset splitting. The comments were manually labelled into two classes, namely online gambling and non-online gambling. Labelling was based on the presence of promotional elements, such as the mention of gambling website names or links, invitations to play, and offers of bonuses or profits. Of the 210,461 comments, 8,428 comments (4%) were labelled as online gambling and 202,033 comments (96%) were labelled as non-online gambling.

Preprocessing was then performed, which included converting emojis into text representations using the demojize method, removing excess whitespace, normalizing characters using unicode, cleaning characters with regular expressions, and removing stopwords using a combination of PySastrawi, NLTK, and the custom stopwords list. Examples of the preprocessing results are shown in Table 2.

Table 2. Examples of text preprocessing

Before preprocessing	After preprocessing
Ini channel yang layak dapet tempat di algoritma YouTube. □ PULAUWIN □	channel layak dapet tempat algoritma youtube red heart pulauwin red heart
Keren abis idenya! MONA4D bikin video ini unik dan memorable!	abis idenya two hearts mona4d purple circle video unik memorable

The dataset was then split into training, validation, and testing sets with a ratio of 70%:15%:15%, resulting in 147,322 training samples, 31,569 validation samples, and 31,570 testing samples.

C. Modelling

At this stage, the YouTube comments were classified into two classes using IndoBERT (indobenchmark/indobert-base-p1), namely comments that contain online gambling promotions and comments that do not. The preprocessed text was tokenized using the IndoBERT tokenizer with a maximum length of 128 tokens and converted into input IDs and attention masks.

To address class imbalance, class weights were applied to the loss function. Hyperparameter optimization was carried out using Optuna with a total of three trials. The optimized hyperparameters were the learning rate in the range of 1×10^{-5} to 5×10^{-5} , a batch size of 8 or 16, and the number of epochs between 3 and 5. The F1-score on the validation data was used as the main metric for selecting the best hyperparameter combination.

The optimization results showed that the best combination was a learning rate of 2.53×10^{-5} , a batch size of 16, and 3 epochs, which produced an F1-score of 0.939382 on the validation data. The training results for each epoch are presented in Table 3.

Table 3. Training results with hyperparameter optimization

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1-score
1	0.257149	0.156189	0.993190	0.904799	0.927042	0.915785
2	0.101760	0.156567	0.994868	0.952263	0.917526	0.934572
3	0.058388	0.158073	0.995217	0.951220	0.927835	0.939382

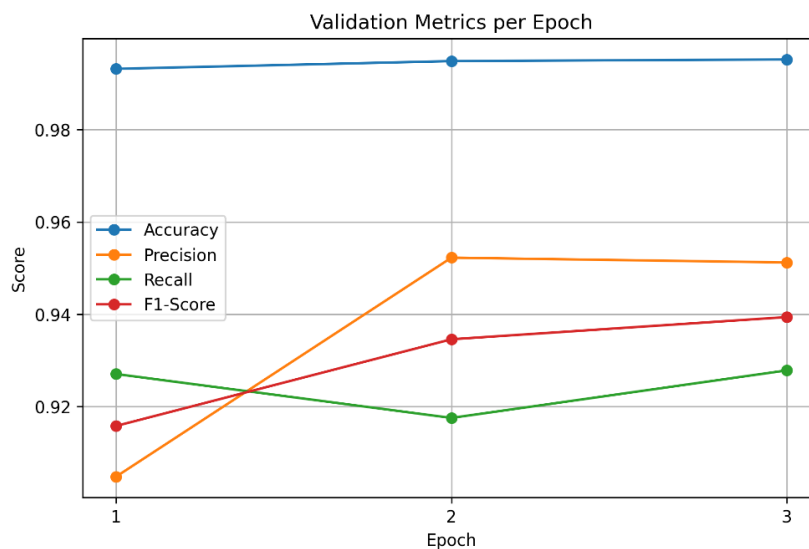


Fig 4. Accuracy, precision, recall, and F1-score curves

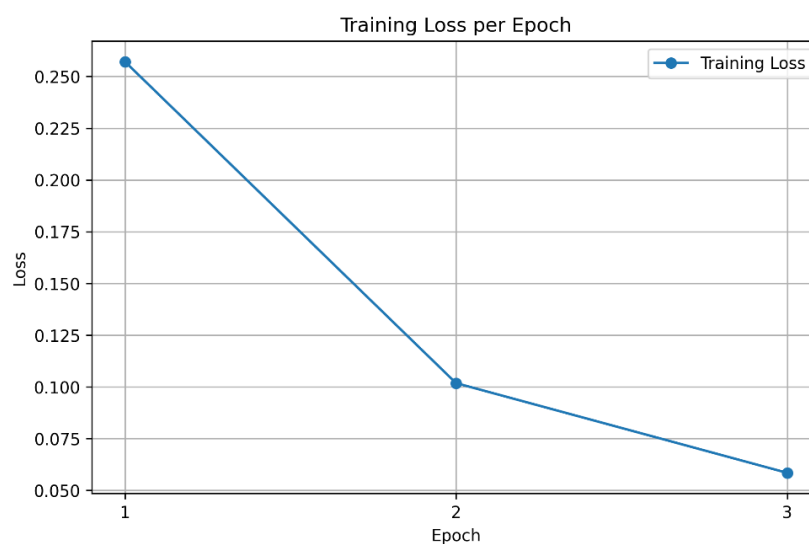


Fig 5. Training loss curve

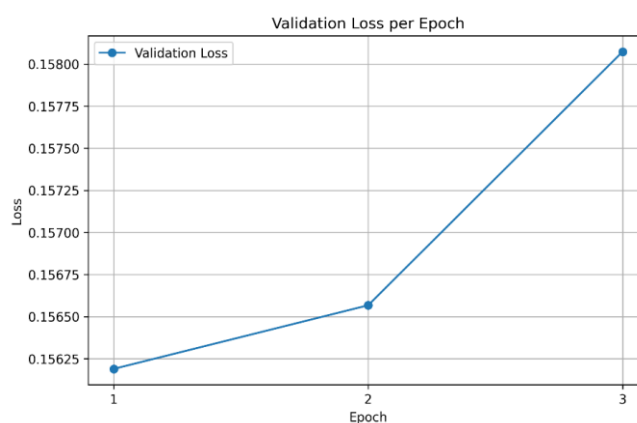


Fig 6. Validation loss curve

As shown in Table 3 and Figs. 4–6, the IndoBERT model performed well, achieving an accuracy of 0.995217, a precision of 0.951220, a recall of 0.927835, and an F1-score of 0.939382 on the validation data in the final epoch. The training loss decreased substantially from 0.257149 to 0.058388, indicating that the model learned the comment patterns effectively. Meanwhile, the validation loss remained relatively stable in the range of 0.156189–0.158073 without a notable increase, indicating that the model generalized well and showed no clear sign of overfitting. These results show that the performance improvement occurred not only on the training data but also consistently on the validation data.

D. Evaluation

At this stage, the model was evaluated using the testing data to measure its generalization ability on data that were not used during training. Performance was evaluated using the confusion matrix and the metrics of accuracy, precision, recall, and F1-score.

Based on the evaluation of the testing data, the IndoBERT model achieved an accuracy of 0.99591, with a precision of 0.96396, a recall of 0.93265, and an F1-score of 0.94805 for the online gambling class. For the non-online gambling class, the model achieved a precision of 0.99720, a recall of 0.99855, and an F1-score of 0.99787. These results are summarized in Table 4.

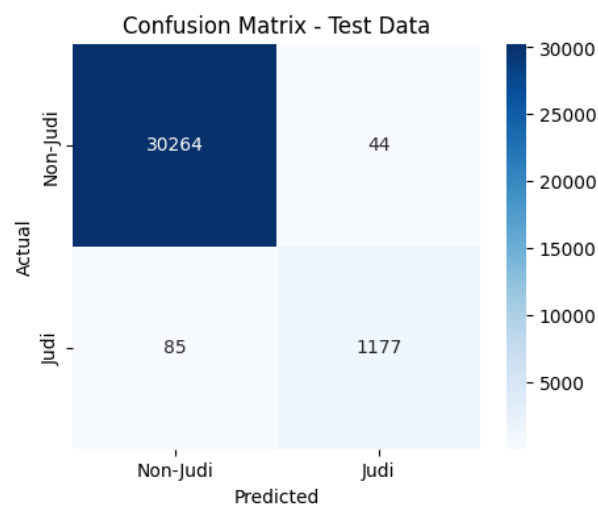


Fig 7. Confusion matrix

As shown in Fig. 7, the model produced 1,177 true positives, 30,264 true negatives, 44 false positives, and 85 false negatives. These results indicate that the model classified most of the testing data correctly. The number of false negatives, which is higher than the number of false positives, shows that some promotional comments were still not detected. This finding is consistent with the recall of the online gambling class being lower than its precision.

E. Deployment

At this stage, the trained IndoBERT model was implemented in a web-based application using the Streamlit framework. This implementation served as a simulation to demonstrate how the model detects online gambling promotional comments, as shown in Fig. 8.

Tulis komentar...

gokilll. keren om. makasih 🍕

Kirim Komentar

Hapus Semua

Deteksi Komentar

User 1
gue yakin 🍕 PUJLAU777 🍕 tuh punya sesuatu yang bikin orang penasaran terus Hapus

User 2
"Ini baru konten berkualitas! Sambil baca, aku sambil nikmati hasil dari KYT4D nih! ❤️" Hapus

User 3
gokilll. keren om. makasih 🍕 Hapus

Hasil Prediksi

No	Hasil (588 ms)
1	JUDI (99.9%)
2	JUDI (99.9%)
3	NON (100.0%)

Fig 8. Interface of the comment detection application

The application provides a text input feature for comments. Each input is processed through the preprocessing stage and then classified by the model into two classes, namely online gambling and non-online gambling. The classification results are displayed in real time to show the response of the model to various text inputs.

F. Monitoring and Maintenance

At this stage, the performance of the model was monitored after it was implemented in the web-based application. Monitoring was carried out by entering various comment inputs directly to test the response of the model to new data, as shown in Fig. 9.

User 1
PULAUWIN itu definisi "kecil tapi berdampak". ❤️ Hapus

User 2
Itu dibelakang nya ada yg kurang bagus bagian spoiler Hapus

User 3
Grafiknya di 🍕 BEM4D 🍕 keren abis! Gameplaynya juga seru! Hapus

User 4
00:59 P U L A U W I N udah kayak temen curhat Hapus

Hasil Prediksi

No	Hasil (146 ms)
1	JUDI (99.9%)
2	NON (100.0%)
3	JUDI (99.9%)
4	JUDI (99.9%)

Fig 9. Model testing with various comment inputs

Based on this test, the classification results were displayed in real time and observed to ensure that the model could still recognize online gambling promotional comments accurately. This process was also used to identify possible performance degradation and the need for model updates in the future.

4. CONCLUSION

This study showed that the IndoBERT Transformer model was successfully applied to detect online gambling promotional comments on YouTube with very good performance. On the testing data, the model achieved an accuracy of 0.99591, a precision of 0.96396, a recall of 0.93265, and an F1-score of 0.94805, which confirms the effectiveness of IndoBERT for this task. This performance was supported

by the use of a large-scale dataset covering various video categories, which improved the ability of the model to generalize across different promotional patterns. In addition, the preprocessing stage and hyperparameter optimization helped produce a model with stable performance and no clear sign of overfitting. The model was also implemented as a Streamlit web application to simulate real-time comment detection, and direct testing showed that it responded well to various comment inputs. Future research may expand the variety of the dataset and apply other techniques or more adaptive models so that more diverse promotional characteristics can be recognized.

ACKNOWLEDGEMENTS

The author would like to thank all lecturers of the Department of Informatics Engineering, Universitas Universal, for their motivation and encouragement in supporting the completion of this research.

DATA AVAILABILITY

The data used in this study are available from the author upon request.

REFERENCES

- [1] S. H. Bakhtiar and A. N. Adilah, "Fenomena judi online: Faktor, dampak, pertanggungjawaban hukum," *Innovative: Journal of Social Science Research*, vol. 4, no. 3, pp. 1016–1026, 2024, doi: 10.31004/innovative.v4i3.10547.
- [2] S. Sudaryono, P. Nugroho, and T. K. Sudaryono, "The socioeconomic impact of online gambling on economic stability in Indonesia: A policy-oriented analysis," *International Journal Multidisciplinary Science*, vol. 4, no. 2, pp. 41–47, 2025, doi: 10.56127/ijml.v4i2.2152.
- [3] L. Rafiqah and H. Rasyid, "Dampak judi online terhadap kehidupan sosial ekonomi masyarakat," *Al-Mutharahah: Jurnal Penelitian dan Kajian Sosial Keagamaan*, vol. 20, no. 2, pp. 282–290, 2023, doi: 10.46781/al-mutharahah.v20i2.763.
- [4] Suwito, "Dampak hukum dan sosial dari judi online di Indonesia: Tantangan dalam penegakan hukum," *Indonesian Journal of Intellectual Publication*, vol. 5, no. 1, pp. 82–90, 2024, doi: 10.51577/ijipublication.v5i1.623.
- [5] J. A. Liem, D. T. Djatmiko Utomo, B. Ignasio, and T. W. Cenggoro, "Deep learning and machine learning for Indonesian online gambling website promotion detection," *Procedia Comput. Sci.*, vol. 269, pp. 979–992, 2025, doi: 10.1016/j.procs.2025.09.040.
- [6] S. Samuel and D. P. Kristiadi, "Deteksi teks promosi judi online menggunakan AI dengan kombinasi NLP dan deep learning," *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, vol. 5, no. 2, pp. 179–185, 2025, doi: 10.56995/sintek.v5i2.179.
- [7] Ryan Prayoga, Adinda Tria Suci, Titus Kristanto, Samsul Lutfi, Desyanti, and Yonky Pernando, "Kombinasi Metode Multi-Objective Optimization on the Basis of Ratio Analysis (MOORA) dan Pembobotan Rank Order Centroid (ROC) dalam Pemilihan Tablet PC Terbaik," *Journal of Informatics Management and Information Technology*, vol. 3, no. 2, pp. 63–69, 2023, doi: 10.47065/jimat.v3i2.256.
- [8] M. Iqbal *et al.*, *Konsep Kecerdasan Buatan*. CV. Gita Lentera, 2024.
- [9] R. B. Perdana, B. Ardin I., A. B. Santoso, A. Ramadiah, and P. K. Putra, "Detecting online gambling promotions on Indonesian Twitter using text mining algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 8, pp. 942–949, 2024, doi: 10.14569/IJACSA.2024.0150893.
- [10] A. N. Cahyo, R. Hermawan, M. A. D. Wijaya, and A. P. Sari, "Deteksi teks promosi judi online pada kolom komentar YouTube dengan metode regex dan fuzzy matching," *Jurnal RESTIKOM: Riset Teknik Informatika dan Komputer*, vol. 7, no. 2, pp. 176–187, 2025, doi: 10.52005/restikom.v7i2.441.

- [11] F. J. M. Riza, R. G. Guntara, and M. R. Nugraha, "Implementasi algoritma naive Bayes untuk filtrasi spam komentar judi online pada YouTube," *Indonesian Journal of Digital Business*, vol. 5, no. 2, pp. 411–423, 2025, [Online]. Available: <https://ejournal.upi.edu/index.php/IJDB>
- [12] I. F. Putra and A. Purwarianti, "Improving Indonesian text classification using multilingual language model," *ArXiv*, 2020, [Online]. Available: <https://arxiv.org/abs/2009.05713>
- [13] A. J. Andika, Y. Kristian, and E. I. Setiawan, "Deteksi komentar cyberbullying pada YouTube dengan metode convolutional neural network – long short-term memory network (CNN-LSTM)," *Teknika*, vol. 12, no. 3, pp. 183–188, 2023, doi: 10.34148/teknika.v12i3.677.
- [14] M. Sam'an and K. Imaddudin, "Hybrid deep learning model for YouTube spam comment detection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 3, pp. 3313–3319, 2024, doi: 10.11591/ijece.v14i3.pp3313-3319.
- [15] M. B. M. Amin, "Deteksi spam berbahasa Indonesia berbasis teks menggunakan model BERT," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1291–1301, 2024, doi: 10.25126/jtiik.2024118121.
- [16] S. P. Santika, M. A. Saputra, A. Zahra, and D. Suhartono, "Online gambling promotion detection in Indonesian YouTube comments using semi-supervised IndoBERT classification," *Procedia Comput. Sci.*, vol. 269, pp. 1269–1278, 2025, doi: 10.1016/j.procs.2025.09.068.
- [17] K. Kamdan, M. P. Anugrah, M. J. Almutaali, R. Ramdani, and I. L. Kharisma, "Performance analysis of IndoBERT for detection of online gambling promotion in YouTube comments," *Engineering Proceedings*, vol. 107, p. 66, 2025, doi: 10.3390/engproc2025107066.
- [18] M. C. T. Manullang, A. Z. Rakhman, H. Tantriawan, and A. Setiawan, "Comparative analysis of CNN, transformers, and traditional ML for classifying online gambling spam comments in Indonesian," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 592–602, 2025, doi: 10.30871/jaic.v9i3.9468.
- [19] H. S. Dutta and B. Khan, "YTCommentVerse: A multi-category multi-lingual YouTube comment corpus," in *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM 2025)*, ACM, 2025, pp. 6351–6355. [Online]. Available: <https://arxiv.org/abs/2509.11057>
- [20] M. Z. Ammar, R. E. Putra, and Y. Yamasari, "Deep learning-based detection of online gambling promotion spam in Indonesian YouTube comments," *Journal of Applied Informatics and Computing*, vol. 9, no. 6, pp. 3632–3641, 2025, doi: 10.30871/jaic.v9i6.11240.
- [21] T. Decker, R. Gross, A. Koebler, M. Lebacher, R. Schnitzer, and S. H. Weber, "The thousand faces of explainable AI along the machine learning life cycle: Industrial reality and current state of research," in *Artificial Intelligence in HCI*, Springer, 2023, pp. 184–208. doi: 10.1007/978-3-031-35891-3_13.