Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm

Reza Bayu Perdana[1], Ardin[2], Indra Budi[3], Aris Budi Santoso[4], Amanah Ramadiah[5], Prabu Kresna Putra[6]
Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia[1, 2, 3, 4, 5]
Research Center for Data and Information Science, National Research and Innovation Agency, Bandung, Indonesia[6]

***Abstract*—This study addresses the pressing challenge of detecting online gambling promotions on Indonesian Twitter using text mining algorithms for text classification and analytics. Amid limited research on this subject, especially in the Indonesian context, we aim to identify common textual features used in gambling promotions and determine the most effective classification models. By analyzing a dataset of 6038 tweets collected and using methods such as Random Forest, Logistic Regression, and Convolutional Neural Networks, complemented by a comparison analysis of text representation methods, we identified frequently occurring words such as 'link', 'situs', 'prediksi', 'jackpot', 'maxwin', and 'togel'. The results indicate that the combination of TF-IDF and Random Forest is the most effective method for detecting online gambling promotion content on Indonesian Twitter, achieving a recall value of 0.958 and a precision value of 0.966. These findings can contribute to cybersecurity and support law enforcement in mitigating the negative effects of such promotions, particularly on the Twitter platform in Indonesia.**

Keywords*—*Social media; analytics; online gambling; intention classification

I. Introduction

In today's digital age, the landscape of social media platforms has undergone significant evolution, becoming an integral part of people's daily lives. Prominent social media accounts now have millions of followers [1], and they are utilized by various elements of society [2]. With billions of users worldwide, platforms such as Facebook, Twitter, Instagram serve as virtual hubs where individuals connect, share, and consume content. According to Digital 2024 Global Overview Report [3], as of January 2024, internet users in Indonesia spent an average of 3 hours 11 minutes a day accessing social media. This duration is above the global average of 2 hours 23 minutes, highlighting the strong attraction of social media for internet users in Indonesia.

Unfortunately, the strong appeal of social media in Indonesia has also led to a dramatic increase in exploitation of social media accounts for the promotion of online gambling in recent years, posing substantial challenges to regulatory authorities and law enforcement agencies. Online gambling promotion, particularly through social media platforms, has emerged as a crucial strategy for reaching potential players [4]. The widespread accessibility of online gambling sites, operating 24/7, is evident [5]. Furthermore, the prevalence of online gambling promotions on social media is notable, with influencers, social media personalities, and even celebrities actively engaging in such campaigns [6]. The ease of access and widespread reach of social media platforms have facilitated the dissemination of these illicit promotions, exacerbating concerns surrounding the detrimental effects of online gambling on vulnerable individuals, including minors and those with gambling addiction issues [7].

Over the past 30 years, the internet has sparked profound changes, presenting three primary challenges for gambling marketing and public policy: (1) significant transformations in the gambling industry's scale, scope, and nature; (2) a surge in gambling advertising on social media; and (3) inadequate methodologies for analyzing the extensive volume of online advertising data [8]. Detecting online gambling promotions on social media presents multifaceted challenges, ranging from identifying the numerous tactics employed by promoters to evade detection to pinpointing the specific accounts engaging in such activities. Moreover, discerning the nuanced interactions and strategies utilized by these accounts adds another layer of complexity to the detection process. Traditional gambling advertising techniques have historically been inspected through manual content analysis [9] [10] [11]. Despite offering detailed insights into specific content characteristics, this approach is laborious and resource-intensive, leading to significant constraints on sample sizes [12].

In the face of these challenges, there exists an opportunity to utilize cutting-edge technologies, especially artificial intelligence algorithms such as machine learning or deep learning with a text classification or text analytics approach, to effectively address the widespread dissemination of online gambling promotions [13]. By harnessing the power of machine learning, researchers can develop robust detection mechanisms capable of identifying and categorizing online gambling promotions with high accuracy. Therefore, the aim of this study is to answer these research questions (RQ):

- What textual features are commonly utilized in online gambling promotions on Indonesian Twitter?
- How do classification models effectively perform in detecting online gambling promotion content in Indonesian Twitter data?

This includes investigating the frequently used textual features in online gambling promotions on Indonesian Twitter and determining the effectiveness of various text mining classification methods for detecting such content. Through this study, we aim to provide significant insights into the common characteristics and patterns found in such promotional content, as well as to identify the most effective classification algorithms for this context. We expect that the results will contribute to various aspects, including the development of prevention strategies and regulations, enhancing cybersecurity measures, and preventing the spread of illegal content on social media platforms. Furthermore, these findings are anticipated to aid law enforcement efforts and protect social media users from the negative impacts of online gambling promotion.

II. RELATED WORKS

Based on our literature review, discussions on the detection of online gambling promotion content in social media are scarcely found. Therefore, we have endeavored to compile knowledge related to the characteristics of similar content and techniques used in previous studies. The literature we gathered covers topics including online gambling, spam detection, as well as text classification and the detection of harmful content users.

In our exploration of online gambling, we have gained insights into textual features and common patterns found in gambling advertisements. Although studies [14] and [15] lack detailed evaluation reports on advertisement classification, study [14] reveals high-frequency words used in gambling ads, suggesting a trend of positive language that highlights benefits, bonuses, and special deals, as observed across studies [7], [16]–[18]. Complementing this, study [7] specifically sheds light on the characteristics of online gambling ads in Indonesia, pointing out features such as bonuses, luck, financial gains, and the ease of joining. This finding aligns with our earlier literature review, which identified a scarcity of discussions on the detection of online gambling content, especially in Indonesia, underscoring the need for our comprehensive approach to understanding these promotional strategies.

In our study, we have reviewed various approaches for text and spam classification. Predominant algorithms such as Random Forest, Support Vector Machine (SVM), and Naïve Bayes, particularly when integrated with N-Gram analysis [19][20], are favored for their reported high accuracy rates, often exceeding 95%. Furthermore, applications of LSTM and Word Sequence [21] are also considered effective alternatives, achieving an accuracy rate of 88%. Other studies [22]–[24] have explored RNN, manual analysis, and topic modeling, though their effectiveness evaluations are not explicitly reported.

Focusing on spam classification, especially content promoting gambling, many papers compare different algorithms, preprocessing stages, and their evaluations. Studies [25] and [20] advocate the use of the SVM algorithm for spam detection, with study [20] achieving an F-Score of 96% through bi-gram and TF-IDF preprocessing. Study [26] employs a Thai BERT derivative, reaching an F-Score of 0.8, while other research, like study [27] using neural networks and GloVe, and study [28] employing logistic regression and Word2Vec, report high F-Scores of 99.73% and 93%, respectively. An alternative approach in study [29] utilizes regex-based rules for spam detection, although it lacks detailed evaluation.

Most of the studies we analyzed use English text data. For text analysis in the Indonesian context, several studies have demonstrated the application of various classification models and feature extraction techniques on Indonesian-language datasets. One notable study [30] utilized a Convolutional Neural Network (CNN) alongside GloVe (Global Vectors) for intent classification within the ATIS (Airline Travel Information System) dataset, achieving a commendable accuracy of 95.84%. In another study [31], the implementation of Naive Bayes combined with TF-IDF was employed for the classification of terrorism-related content on Indonesian Twitter, attaining an F-measure of 77%. Setiawan's research [32] applied Logistic Regression with Word2Vec for feature extraction to detect spam posts on Indonesian Twitter, resulting in an accuracy of 93.67%. Additional research efforts in spam detection on Indonesian Twitter include [33], which used CountVectorizer and KNN to achieve an accuracy of 79%, and another research [34], which employed CountVectorizer and Random Forest, achieving an accuracy of 85.1%. These studies underscore the versatility and effectiveness of different text classification approaches within the Indonesian linguistic framework.

In advancing our research on the detection of online gambling promotion content in social media, we encounter several challenges that are crucial to address. These include the determination of effective keywords for data retrieval from social media queries, streamlining the labeling processes, and the selection of appropriate models and textual features. The need for precise keyword selection and labeling is particularly acute due to the limited data availability from Twitter following recent policy changes. To address this issue, we conducted descriptive analysis of frequently occurring words in online gambling promotions on Indonesian Twitter to provide insights into more effective search keyword identification. This analysis will precede the predictive analysis of training on classification models to detect such online gambling promotional content, aiming to develop an optimized approach for our text mining processes.

III. METHODOLOGY

In this research, we aim to conduct both descriptive and predictive analyses to gain a comprehensive understanding of online gambling promotion on Indonesian Twitter, thus enhancing detection knowledge and capabilities. Our research process, depicted in Fig. 1, initiates the identification of query keywords for Twitter data collection. Once keywords are selected, Twitter data retrieval is carried out. The accumulated data are labeled and then undergo pre-processing steps such as tokenization, normalization, noise removal, and stemming. Post-preprocessing, experiments are conducted as specified in Table II to identify the optimal model for detecting online gambling content on Indonesian Twitter. The experimental outcomes are further discussed to interpret the findings. Finally, conclusions and reports are drafted based on the discussions and analyses conducted, culminating in the research.

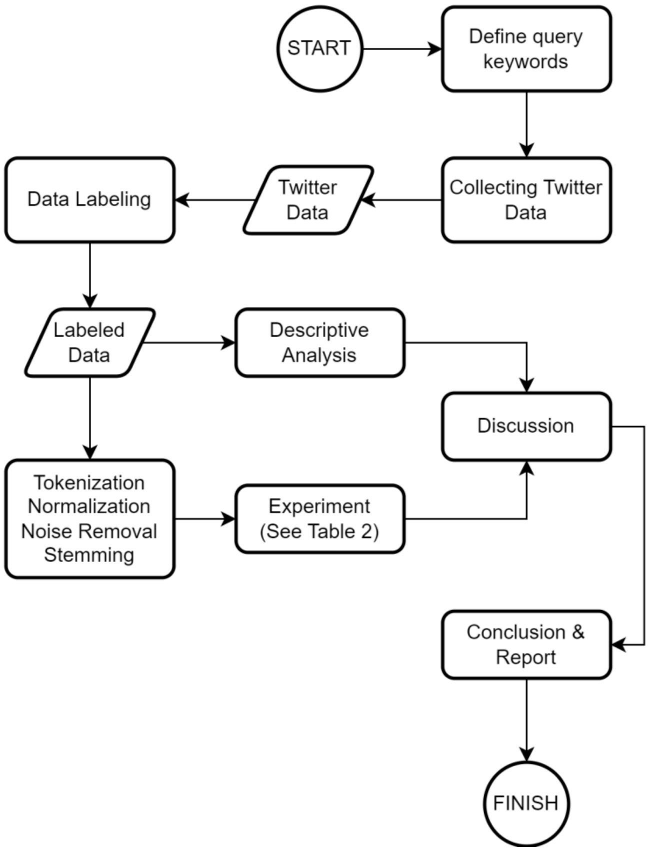

Fig. 1. Proposed workflow of this study.

Keyword query determination aims to ensure effective retrieval from Twitter searches. In identifying keywords, we refer to previous research on general online gambling promotion, characterized by a positive tone and suggestive of benefits and ease [7], [14]. Based on [7] and other studies published online [35], we use the query string *"(slot OR gacor OR untung OR cuan) lang:id"* to collect Indonesian Twitter samples containing online gambling content. We target a minimum of 5000 Indonesian language tweet samples for further analysis.

Upon sample data acquisition, data labeling ensues. We employ two annotators who have agreed upon the criteria for what constitutes online gambling promotion content, with examples outlined in Table I. The labeled data is then utilized for both descriptive and predictive analyses. In determining the label of a tweet, we provided clear labeling criteria to both annotators to minimize inconsistencies. A tweet was labeled as containing online gambling promotional content if it met one or more of the following criteria:

- Explicitly inviting individuals to participate in online gambling activities,
- Displaying links or names of online gambling platforms,
- Showing results from online gambling activities,
- Providing instructions on how to join online gambling activities.

TABLE I. EXAMPLE OF POSTS WITH ONLINE GAMBLING PROMOTION

No	Tweet
1	DIMANA LAGI CUMA MODAL KECIL DAPAT UNTUNG BESAR KALAU BUKAN DI SGA338, TUNGGU APA LAGI ? AYO ... BURUAN DAFTAR SEKARANG JUGA DAN DAPATKAN KEMENANGANNYA ! INFO LEBIH LANUT DAPAT HUBUNGI KONTAK KAMI NO HP WA : +6287815585xxx https://t.co/Y7OeIF7S35 https://t.co/qiJuNc88af
2	INFO SLOT GACOR HARI INI DAFTAR SLOT GACOR ➡□ https://t.co/hnMWbSLliD https://t.co/2bCt8SCFyB
3	388HERO - Slot Paling Gacor Di Indonesia !! ✅Bonus New Member 100% ✅Garansi Kekalahan 100% Daftar Slot Gacor Dan Claim Sekarang Juga □□□ https://t.co/jog4l9Jt0S https://t.co/RZTBrEyIYJ
4	Gotobet88 berbagi thr lewat maxwin yang di persembahkan spesial kepada member setia kami daftarkan diri anda segera di situs gacor gampang maxwin kami https://t.co/LgaCegaamG lewat @pinterest
5	Jadwal pertandingan CSGO. Ayo dukung jagoanmu dengan cara gabung di WINGSLOTS77, boskiuh. Raih kemenangan besar dengan cuan tiada batasnya. Link Alternatif : https://t.co/jdlSYjoZVS #wingslots77 #csgo #pgl #pglmajorcopenhagen https://t.co/YffbGd4Ebe

Furthermore, we calculated the inter-annotator agreement using Cohen's Kappa coefficient.

To address any disagreements between annotators, a reconciliation process was implemented. Disagreements were discussed in a joint session between the annotators, and if consensus could not be reached, a third expert annotator was brought in to make the final decision. This process ensured a thorough review and maintained consistency in labeling.

Descriptive analysis is conducted to depict the nature of online gambling content more comprehensively on Indonesian Twitter and aim to answer RQ 1. The steps for descriptive analysis derive from the study of online gambling content on Facebook [35], including temporal analysis, interaction count analysis, and examination of frequently occurring words besides search keywords, analyzed using a word cloud.

Predictive analysis aims to develop a classification model that detects online gambling promotional posts from the available dataset as the answer to RQ 2. After labeling, the data is tokenized, normalized, noise is removed, and stemming is conducted. The feature extraction and modeling phases, which determine the best algorithm for detecting online gambling promotion, are conducted through experiments listed in Table II. Algorithm selection for experimentation is based on previous research [30], [32], [34].

The selection of algorithms for this study was carefully considered to match the unique challenges presented by the task of detecting online gambling promotions on Twitter. Given the informal and varied nature of Twitter text, including the use of slang and abbreviations, Convolutional Neural Networks (CNNs) were chosen for their ability to capture complex patterns and contextual nuances within text. Random Forest was selected due to its effectiveness in handling high-dimensional data and its resistance to overfitting, which is essential when dealing with diverse textual features. Logistic Regression was included for its interpretability and as a baseline model to compare performance, as it is widely used in text classification tasks. This combination of models was intended to leverage the strengths of each algorithm in addressing different facets of the data, ensuring a robust analysis of the factors contributing to the detection of online gambling promotions.

Evaluations are performed using a confusion matrix, focusing on recall and F-measure to minimize the potential oversight of online gambling content by the model. Due to the lack of reported training times in related studies, a direct comparison of computational efficiency between our model and those presented in [30], [32], and [34] cannot be conducted. Nevertheless, our results highlight the importance of including training time as a key metric for evaluating model performance, especially in contexts where computational resources are limited. We chose the F1 score and Recall as our primary evaluation metrics due to their relevance in contexts with imbalanced data. Recall is particularly important because it measures the model's ability to correctly identify all instances of the minority class, which is crucial in detecting online gambling promotions. The F1 score provides a balanced measure of both precision and recall, offering insight into the model's performance in minimizing false positives while still capturing true positives. These metrics were selected to ensure that the chosen model performs effectively even in the presence of class imbalance.

Once the optimal model is identified and descriptive analysis is complete, the report writing is finalized with discussions and conclusions.

TABLE II. EXPERIMENTAL LIST

Experiment	Text Representation	Algorithm
1	GloVe	CNN
2	GloVe	LR
3	GloVe	RF
4	Word2Vec	CNN
5	Word2Vec	LR
6	Word2Vec	RF
7	TF-IDF	CNN
8	TF-IDF	LR
9	TF-IDF	RF

These steps form the basis of our proposed research methodology, which aims to address the intricacies of detecting online gambling promotions in social media content effectively.

IV. RESULTS AND DISCUSSIONS

We successfully collected 6,038 Twitter data points based on the predetermined keywords. This result exceeds our target of 5,000 Twitter data points, allowing us to proceed with data annotation.

The data annotation was carried out by two annotators working independently. The annotation process involved assigning a label of "1" if the post contained online gambling promotion content, and a label of "0" if the post did not contain such content. The annotation results were then evaluated using Cohen's Kappa coefficient, with the matrix presented in Table III.

TABLE III. ANNOTATION EVALUATION MATRIX

Label	Label 0 (Annotator B)	Label 1 (Annotator B)
Label 0 (Annotator A)	4308	41
Label 1 (Annotator A)	42	1647

Based on Table III, the Cohen's Kappa coefficient from the evaluation is 0.9659, indicating an almost perfect agreement between the annotators [36]. This result suggests that the labeled data can be used for the next steps in the analysis, which include descriptive analysis and data preprocessing. The comparison of the label counts for each category is shown in Fig. 2 that shows the imbalanced dataset as labeling results.

Our approach reflects a realistic scenario where data imbalance is common, and practitioners may not always have the opportunity or resources to apply complex rebalancing techniques. By evaluating model performance on the raw, imbalanced data, we aimed to identify algorithms that are naturally resilient to such conditions. This decision was based on the need to assess the practical utility of the models in environments where preprocessing capabilities may be limited.

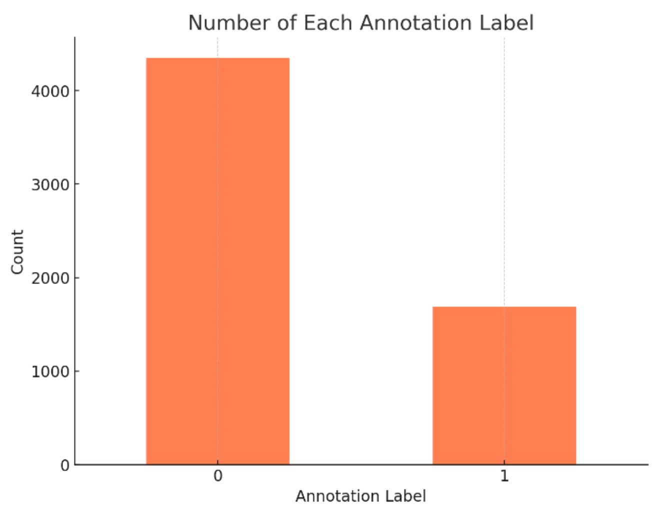

Fig. 2. Number of labels.

We then proceeded with temporal analysis by visualizing the number of tweets labeled "1" per hour. The visualization results, as shown in Fig. 3, indicate that online gambling promotion content is often posted during the early morning and late afternoon to evening hours.

Interactions with tweets containing online gambling promotion content are relatively high, averaging 1.68 interactions per tweet, including quotes, replies, retweets, or favorites. This is five times higher than the average interactions per tweet labeled "0," which is 0.38 interactions per tweet. Three accounts with usernames indicative of online gambling operators have average interactions of 36, 29, and 24 interactions per tweet. This indicates that accounts and posts containing online gambling promotion content attract significant attention from Twitter users.

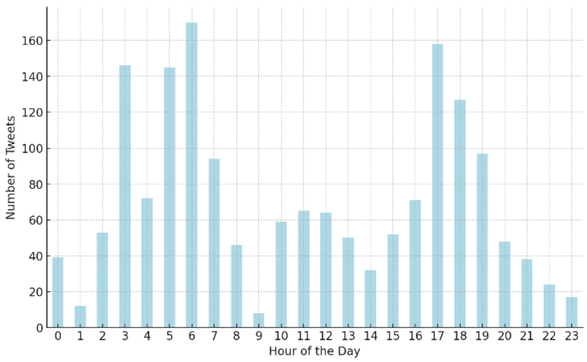

Fig. 3. Temporal distribution of labeled "1" data.

We then conducted a word cloud analysis to identify the most frequently occurring words in tweets containing online gambling promotion content. As shown in Fig. 4, in addition to the search keywords, words such as "link," "situs," "prediksi," "jackpot," "maxwin," and "togel" were identified as being affiliated with posts containing online gambling promotion content.

Fig. 4. Wordcloud of labeled "1" data.

We then proceeded with data preprocessing for modeling. We employed a back-translation method for data normalization [37]. In this process, we translated the documents into English and then retranslated them back into Indonesian using the Google Translate API. This approach ensures that words with the same meaning are standardized and helps avoid spelling errors in Indonesian. We present an example of the outcomes from the back-translation process in Table IV. This table includes an example of data labeled as '1' or 'online gambling promotion,' due to its explicit promotion of an online gambling site. Our back-translation process successfully reduced the noise present in the raw data.

TABLE IV. RESULT EXAMPLE OF BACK-TRANSLATION METHOD

Raw Data (Indonesian)	English Translation Data	Indonesian Back-Translation Data
Situs SABDA4D t3rbaik dan t3rperc4ya di 1ndon3sia m3ny3di4kan g4mes t3rlengkap d4n t3rpopuler s3rta RTP slot &amp pol4 g4cornya t3rupd4te s3tiap j4m ! bonu5 D3po 200k p4ka1 l1nk m4xwinny4 : {LINK} {LINK}	The best and most trusted SABDA4D site in Indonesia offers the most complete and popular games and RTP slots & pol4 g4cor every hour! D3po bonus 200k p4ka1 l1nk m4xwinny4: {LINK} {LINK}	Situs SABDA4D terbaik dan terpercaya di Indonesia menawarkan permainan dan RTP slot & pol4 g4cor terlengkap dan terpopuler setiap jamnya! Bonus D3po 200k p4ka1 l1nk m4xwinny4: {LINK} {LINK}

The next steps we undertook included removing stopwords, replacing slang words, removing extra characters in words, converting emojis to phrases, removing usernames, removing numbers, removing punctuation, and converting text to lowercase. During this data cleaning process, we encountered several challenges, including the use of slang, symbols, and local languages that were occasionally embedded in sentences written in Indonesian. As a result, we also conducted manual cleaning to ensure that the data included in the modeling process was accurate. Additionally, we applied stemming to revert words to their root forms. An example of the outcomes from these processes are presented in Table V.

TABLE V. RESULT EXAMPLE OF DATA PREPROCESSING

Initial Data	Cleaned Data
Situs SABDA4D terbaik dan terpercaya di Indonesia menawarkan permainan dan RTP slot & pol4 g4cor terlengkap dan terpopuler setiap jamnya! Bonus D3po 200k p4ka1 l1nk m4xwinny4: {LINK} {LINK}	situs sabda d baik percaya indonesia sedia game rtp slot pola gacor update per jam lengkap populer bonus

After completing these processes, we initiated the modeling process as outlined in Table II, starting with document vectorization.

The experimental results using nine scenarios, as shown in Table VI, indicate that the combination of GloVe and CNN demonstrates excellent performance across all evaluation metrics, albeit with a relatively high training time. The high recall and precision values indicate that this model is highly effective in detecting tweets that genuinely promote online gambling and minimizing false positives. In contrast, the combination of GloVe and Logistic Regression performs very poorly, especially in terms of recall and F1 score, indicating that this model is inadequate at detecting online gambling promotion tweets. Random Forest with GloVe shows good performance but is not as effective as CNN. This model has lower recall but is still reasonably good at detecting promotional tweets.

TABLE VI. RESULTS OF EXPERIMENT'S SCENARIO

Vectorizer	Algorithm	Accuracy	AUC	Recall	Precision	F1 Score	Kappa	MCC	Training Time
GloVe	CNN	0.97462	0.99517	0.94268	0.96493	0.95363	0.93612	0.93632	7.81112
GloVe	LR	0.72162	0.61804	0.08757	0.51136	0.14749	0.07424	0.11495	0.07246
GloVe	RF	0.93234	0.97619	0.80093	0.9492	0.86824	0.82316	0.82899	0.54267
TF-IDF	CNN	0.71799	0.74627	0.16888	0.48671	0.24517	0.12366	0.15008	8.037
TF-IDF	LR	0.97463	0.99434	0.94542	0.96326	0.95414	0.93655	0.93677	0.02478
TF-IDF	RF	0.97878	0.99419	0.95829	0.96572	0.96183	0.94711	0.94726	0.63892
Word2Vec	CNN	0.97082	0.9939	0.94344	0.95303	0.94759	0.92736	0.92798	4.12421
Word2Vec	LR	0.93732	0.98075	0.86085	0.911	0.88469	0.84168	0.8428	0.04422
Word2Vec	RF	0.96119	0.9902	0.92247	0.9376	0.92967	0.90286	0.9032	2.06679

The combination of TF-IDF and CNN shows poor performance, particularly in recall and F1 score, indicating that this model is less effective in detecting online gambling promotion tweets. The combination of TF-IDF and Logistic Regression demonstrates excellent performance, similar to GloVe-CNN, but with significantly lower training time, making it more efficient. Random Forest with TF-IDF shows the best performance among all combinations, with very high evaluation metrics and still relatively efficient training time.

The combination of Word2Vec and CNN demonstrates excellent performance, approaching the performance of GloVe-CNN but with lower training time. Logistic Regression with Word2Vec shows good performance, though not as strong as the other top combinations. It's very low training time makes this model highly efficient. Random Forest with Word2Vec shows excellent performance, slightly below the performance of TF-IDF with Random Forest. This model is still very effective in detecting promotional tweets, with lower training time compared to TF-IDF with CNN.

Overall, the combination of TF-IDF with the Random Forest algorithm provides the best performance in identifying online gambling promotion content. The Recall value of 95.8% indicates that the TF-IDF with RF combination can minimize false negatives. The high Precision value of 96.6% shows that this combination effectively avoids false positives. This is also supported by a high MCC value and moderate training time.

TF-IDF with Logistic Regression and Word2Vec with Logistic Regression show a good balance between high performance and very low training time. GloVe with Logistic Regression and TF-IDF with CNN demonstrate significantly lower performance and are not recommended for detecting online gambling promotion tweets.

We conducted hypothesis testing based on the results presented in Table IV to determine if there is a significant difference between the performance of the TF-IDF with Random Forest (RF) combination and the Word2Vec with Convolutional Neural Network (CNN) combination. This analysis was undertaken because both combinations demonstrated high Recall and F1 Score with acceptable training times, along with the difference in text vectorization's method. The null hypothesis (H0) was "There is no significant difference between the performance of TF-IDF with RF and Word2Vec with CNN," and the alternative hypothesis (H1) was "There is a significant difference between the performance of TF-IDF with RF and Word2Vec with CNN."

We performed hypothesis tests on two model evaluation metrics: Recall and F1 Score. The results of these hypothesis tests are presented in Table VII.

TABLE VII. HYPOTHESIS TEST RESULTS FOR RECALL AND F1 SCORE

	alpha	p-value	Results
Recall	0.05	0.130	Fail to reject H0: No significant difference
F1 Score	0.05	0.047	Reject H0: Significant difference

Table VII shows that the p-value for the recall metric is 0.130, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis. This indicates that there is no statistically significant difference in recall between the TF-IDF + RF model and the Word2Vec + CNN model. In other words, both models perform similarly in terms of recall. The recall metric measures the ability of the model to correctly identify all relevant instances of online gambling promotion. The lack of a significant difference in recall suggests that both models are equally effective in identifying true positive instances of online gambling promotions on Indonesian Twitter. This could imply that both text representation methods (TF-IDF and Word2Vec) and classification algorithms (RF and CNN) are similarly proficient in capturing the relevant features needed for high recall in this context.

The p-value for the F1 score metric is 0.047, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis. This indicates that there is a statistically significant difference in F1 score between the TF-IDF + RF model and the Word2Vec + CNN model. Specifically, the TF-IDF + RF model has a significantly better F1 score compared to the Word2Vec + CNN model. The F1 score is a harmonic means of precision and recall, providing a balance between the two. The significant difference in F1 score, favoring the TF-IDF + RF model, indicates that this model not only identifies true positives effectively (recall) but also minimizes false positives (precision). The higher F1 score suggests that TF-IDF + RF strikes a better balance between precision and recall compared to Word2Vec + CNN, making it a more reliable model for this task.

Compared to the previous studies we use for reference, although the exact training times were not reported, the studies [30], [32], and [34] utilized models such as Random Forest and SVM, which are known for their differing computational complexities. Random Forest, for example, often requires more computational resources due to the ensemble nature of the algorithm, compared to a single-layer SVM model. By contrast, our study's use of a TF-IDF combined with Random Forest may offer a more balanced trade-off between accuracy and computational efficiency.

V. Conclusion

The conclusion of this study shows that, based on the analysis, textual features commonly utilized in online gambling promotions on Indonesian Twitter include words such as "link," "situs," "prediksi," "jackpot," "maxwin," and "togel." These words frequently appear in online gambling content on Twitter. This indicates that these words are effective textual features for detecting online gambling promotions on Indonesian Twitter, in addition to the initial textual features we used as keywords in our query string. Furthermore, the results of the classification model training indicate that the combination of TF-IDF and Random Forest is the most effective method for detecting online gambling promotion content on Indonesian Twitter. With a recall value of 95.8% and a precision value of 96.6%, this combination significantly minimizes false negative and positive detections. This research makes an important contribution to the development of strategies for preventing and regulating illegal content on social media and supports law enforcement efforts in addressing the negative impacts of online gambling promotion.

While our study focuses on detecting online gambling promotions on Indonesian Twitter, the proposed methodology could potentially be generalized to other languages and platforms. For instance, the tokenization and normalization processes would need adjustments to account for linguistic differences in languages like Japanese or Arabic. Furthermore, the high variability in content length and user behavior across platforms like Facebook and Instagram might require a reevaluation of the feature extraction techniques to maintain model accuracy.

For future research, it is recommended to apply and test this model on other social media platforms to broaden the generalization of the results, develop more efficient approaches for data annotation to enhance labeling accuracy and consistency, integrate sentiment analysis to gain a deeper understanding of the psychological impact of online gambling promotion content on social media users, and explore the use of other deep learning algorithms that may offer better performance with larger datasets.

With these suggestions, it is hoped that future research can be more comprehensive in addressing and preventing the spread of illegal content on social media.

Acknowledgment

This paper is the original work of the authors, Reza Bayu Perdana, Ardin, Aris Budi Santoso, Prabu Kresna Putra, Amanah Ramadiah, and Indra Budi, all of whom contributed significantly to the research, analysis, and writing of this study. We declare that no part of this work has been plagiarized, and it represents the authors' authentic efforts in addressing the challenge of detecting online gambling promotions on Indonesian Twitter.

The writing of this scientific paper was supported by the University of Indonesia and the Meteorology, Climatology, and Geophysics Agency (BMKG), with funding provided by BMKG as part of a scholarship to improve the quality of human resources managed by the BMKG Education and Training Center and also provided by Faculty of Computer Science University of Indonesia by Hibah Riset Internal Fakultas Ilmu Komputer UI TA 2024-2025 No. NKB-6/UN2.F2.D/NKP.05.00/2024.

References

[1] S. Wu and D. Peng, “Pre-SMATS: A multi-task learning based prediction model for small multi-stage seasonal time series,” Expert Syst. Appl., vol. 201, 2022, doi: 10.1016/j.eswa.2022.117121.

[2] X. Y. Leung, B. Bai, and K. A. Stahura, “The marketing effectiveness of social media in the hotel industry: A comparison of Facebook and Twitter,” J. Hosp. Tour. Res., vol. 39, no. 2, pp. 147–169, 2015.

[3] Datareportal, “Digital 2024: Global Overview Report,” 2024. https://datareportal.com/reports/digital-2024-global-overview-report (accessed Jun. 01, 2024).

[4] A. Sudibyo, Media Massa Nasional Menghadapi Disrupsi Digital. Kepustakaan Populer Gramedia, 2023.

[5] I. Y. Nono, A. A. S. L. Dewi, and I. P. G. Seputra, “Penegakan Hukum Terhadap Selebgram yang Mempromosikan Situs Judi Online,” J. Analog. Huk., vol. 3, no. 2, pp. 235–239, 2021.

[6] S. Desriwaty, “Pertanggungjawaban Pidana terhadap Pelaku Promosi Judi Online yang dilakukan melalui Media Sosial ditinjau dari Perspektif Hukum Pidana.” Prodi Ilmu Hukum, 2023.

[7] M. S. Gunawan, N. Mujahidah, S. Sofyan, and M. A. M. A, “Pertanggungjawaban Platfom Media Sosial Terhadap Promosi Judi Online,” J. Plaza Huk. Indones., vol. 1, no. 19, pp. 1–15, 2023.

[8] O. Burkeman, “Forty years of the internet: How the world changed forever,” Guard., vol. 23, 2009.

[9] B. Abarbanel, S. M. Gainsbury, D. King, N. Hing, and P. H. Delfabbro, “Gambling games on social platforms: How do advertisements for social casino games target young adults?,” Policy & internet, vol. 9, no. 2, pp. 184–209, 2017.

[10] R. Cassidy and N. Ovenden, “Frequency, duration and medium of advertisements for gambling and other risky products in commercial and public service broadcasts of English Premier League football.,” 2017.

[11] S. M. Gainsbury, P. Delfabbro, D. L. King, and N. Hing, “An exploratory study of gambling operators’ use of social media and the latent messages conveyed,” J. Gambl. Stud., vol. 32, no. 1, pp. 125–141, 2016.

[12] C. Erlingsson and P. Brysiewicz, “A hands-on guide to doing content analysis,” African J. Emerg. Med., vol. 7, no. 3, pp. 93–99, 2017.

[13] R. J. E. James and A. Bradley, “The use of social media in research on gambling: A systematic review,” Curr. Addict. Reports, vol. 8, pp. 235–245, 2021.

[14] S. Choi, “Understanding Involuntary Illegal Online Gamblers in the U.S.: Framing in Misleading Information by Online Casino Reviews,” UNLV Gaming Res. Rev. J., vol. 27, no. 1, pp. 23–47, 2023, [Online]. Available: https://www.proquest.com/scholarly-journals/understanding-involuntary-illegal-online-gamblers/docview/2807105791/se-2?accountid=17242

[15] A. HernáNdez-Ruiz and Y. Gutiérrez, "Analysing the Twitter accounts of licensed Sports gambling operators in Spain: a space for responsible gambling?," Commun. Soc., vol. 34, no. 4, pp. 65–79, 2021, doi: https://doi.org/10.15581/003.34.4.65-79.

[16] K. Kolandai-Matchett and M. Wenden Abbott, "Gaming-Gambling Convergence: Trends, Emerging Risks, and Legislative Responses," Int. J. Ment. Health Addict., vol. 20, no. 4, pp. 2024–2056, Aug. 2022, doi: https://doi.org/10.1007/s11469-021-00498-y.

[17] A. Bradley and R. J. James, "How are major gambling brands using Twitter?," Int. Gambl. Stud., vol. 19, no. 3, pp. 451–470, 2019.

[18] T. Teichert, A. Graf, T. B. Swanton, and S. M. Gainsbury, "The joint influence of regulatory and social cues on consumer choice of gambling websites: preliminary evidence from a discrete choice experiment," Int. Gambl. Stud., vol. 21, no. 3, pp. 480–497, Sep. 2021, doi: 10.1080/14459795.2021.1921011.

[19] S. Aiyar and N. P. Shetty, "N-gram assisted youtube spam comment detection," Procedia Comput. Sci., vol. 132, p. 1740182, 2018.

[20] Y. Zhang et al., "Lies in the Air: Characterizing Fake-base-station Spam Ecosystem in China," in Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 2020, pp. 521–534. doi: 10.1145/3372297.3417257.

[21] E. Zhu, J. Wu, H. Liu, and K. Li, "A Sentiment Index of the Housing Market in China: Text Mining of Narratives on Social Media," J. Real Estate Financ. Econ., vol. 66, no. 1, pp. 77–118, Jan. 2023, doi: https://doi.org/10.1007/s11146-022-09900-5.

[22] S. I. Alqahtani, W. M. S. Yafooz, A. Alsaeedi, L. Syed, and R. Alluhaibi, "Children's Safety on YouTube: A Systematic Review," Appl. Sci., vol. 13, no. 6, p. 4044, Mar. 2023, doi: 10.3390/app13064044.

[23] R. Rossi, A. Nairn, J. Smith, and C. Inskip, "Get a£ 10 Free Bet Every Week!"—gambling advertising on Twitter: volume, content, followers, engagement, and regulatory compliance," J. public policy Mark., vol. 40, no. 4, pp. 487–504, 2021.

[24] J. Singer, V. Kufenko, A. Wöhr, M. Wuketich, and S. Otterbach, "How do Gambling Providers Use the Social Network Twitter in Germany? An Explorative Mixed-Methods Topic Modeling Approach," J. Gambl. Stud., vol. 39, no. 3, pp. 1371–1398, 2023.

[25] S. Tang, X. Mi, Y. Li, X. Wang, and K. Chen, "Clues in Tweets: Twitter-Guided Discovery and Analysis of SMS Spam," in Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 2022, pp. 2751–2764. doi: 10.1145/3548606.3559351.

[26] P. Chiawchansilp and P. Kantavat, "Spam Article Detection on Social Media Platform Using Deep Learning: Enhancing Content Integrity and User Experience," 2023. doi: 10.1145/3628454.3628459.

[27] S. Kaddoura, S. A. Alex, M. Itani, S. Henno, A. AlNashash, and D. J. Hemanth, "Arabic spam tweets classification using deep learning," Neural Comput. Appl., vol. 35, no. 23, pp. 17233–17246, Apr. 2023, doi: 10.1007/s00521-023-08614-w.

[28] M. Liu, Y. Zhang, B. Liu, Z. Li, H. Duan, and D. Sun, "Detecting and Characterizing SMS Spearphishing Attacks," in Proceedings of the 37th Annual Computer Security Applications Conference, 2021, pp. 930–943. doi: 10.1145/3485832.3488012.

[29] N. Nasir, F. Iqbal, M. Zaheer, M. Shahjahan, and M. Javed, "Lures for Money: A First Look into YouTube Videos Promoting Money-Making Apps," in Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security, 2022, pp. 1195–1206. doi: 10.1145/3488932.3517404.

[30] C. O. Bilah, T. B. Adji, and N. A. Setiawan, "Intent Detection on Indonesian Text Using Convolutional Neural Network," 2022.

[31] M. Aryani, E. Miranda, Y. Fernando, and T. M. Kibtiah, "An Early Warning Detection System of Terrorism in Indonesia from Twitter Contents using Naïve Bayes Algorithm," 2020.

[32] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, "Detecting Indonesian Spammer on Twitter," in 2018 6th International Conference on Information and Communication Technology (ICoICT), 2018, pp. 259–263. doi: 10.1109/ICoICT.2018.8528773.

[33] A. Pinandito, R. S. Perdana, M. C. Saputra, and H. M. Az-zahra, "Spam detection framework for Android Twitter application using Na\"{\i}ve Bayes and K-Nearest Neighbor classifiers," in Proceedings of the 6th International Conference on Software and Computer Applications, 2017, pp. 77–82. doi: 10.1145/3056662.3056704.

[34] L. Alhaura and I. Budi, "Malicious Account Detection on Indonesian Twitter Account," in 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), 2020, pp. 176–181. doi: 10.1109/IC2IE50715.2020.9274682.

[35] I. Fahmi, "Darurat Judi Online di Facebook Pages Indonesia (2) Periode Data: 1 Mei – 22 Agustus 2023," 2023. https://threadreaderapp.com/thread/1704067338306543736.html

[36] D. Altman, "Inter-rater agreement," Pract. Stat. Med. Res., pp. 403–409, 1991.

[37] R. I. Alhaqq, "Klasifikasi Ulasan Pengguna Aplikasi InfoBMKG di Google Play Store," University of Indonesia, 2023.