Detecting Online Gambling Promotion on Social Media with Random Forest

Danny Akhmad Zulfikar[1], Vivi Aida Fitria[1*]

[1,2]Institut Teknologi dan Bisnis Asia Malang, Jl. Soekarno Hatta Rembuksari 1 A Malang, Indonesia

Keywords

Comment Detection; Online Gambling; Random Forest; Social Media; Text Mining; TF-IDF; Twscrape

***Correspondence Email:**
viviaida@asia.ac.id*

Abstract

The main problem in today's digital era is the increasing spread of comments containing online gambling promotions on social media. This activity not only disrupts other users but also has the potential to normalize gambling behavior in the digital public sphere. Therefore, it is necessary to develop an intelligent system capable of automatically detecting and filtering comments that contain online gambling promotions accurately and efficiently. This study aims to develop a detection model for online gambling promotion comments on social media using a *text mining* approach with the Random Forest algorithm. The dataset was collected using the twscrape library without the official API, resulting in 10,607 comments, consisting of 5,139 non-gambling and 5,468 gambling-related comments, making the dataset relatively balanced. The preprocessing steps included text cleaning, case folding, tokenizing, stopword removal, and stemming in the Indonesian language. The TF-IDF method was used for feature extraction, and the Random Forest algorithm was applied to classify comments into two categories: gambling promotion and non-gambling. The experimental results show that the Random Forest model achieved an accuracy of 92%, with consistently high precision, recall, and F1-score across all classes. These findings indicate that the text mining approach using Random Forest is effective in detecting online gambling promotion content. The developed model can serve as a foundation for automated detection systems that support efforts to prevent the spread of gambling-related activities on social media platforms.

1. Introduction

The rapid development of information technology has had a major impact on various aspects of human life. Social media has now become one of the main means of communication for sharing information, promotion, and even digital economic activities. However, behind this convenience, various forms of abuse have emerged, one of which is the increase in online gambling promotion on social media. This phenomenon not only disturbs the comfort of other users but also has the potential to normalize gambling behavior in digital public spaces (Firmansyah, 2024).

Twitter is a social media and information medium that connects its users with information about topics that match their interests (Aprilizdihar et al., 2022). Twitter emerged after the rise in popularity of Facebook. Twitter has a new format that is different from Facebook, namely microblogging, with 280 characters for each tweet. Initially, there were only 140 characters per tweet, but this was considered too few. This allows users to share information (Mustaqlilah et al., 2023).

Online gambling promotion activities are increasingly developing through various social media platforms, especially Twitter (X), where comments, retweets, and the use of hashtags are utilized to expand the reach of promotions (Ghoni et al., 2024). The methods used are usually hidden in the form of comments containing links, promotional codes, or figurative words such as "salam gacor" and "maxwin" (Julianti et al., 2024). These promotions are often carried out by bot accounts or fake accounts, making them difficult to detect manually and causing massive dissemination. In addition to violating the law, these activities can also trigger addictive behavior and cause negative social and economic impacts on society (Firmansyah, 2024).

Manual efforts to detect online gambling promotional comments are clearly inefficient given the large and dynamic volume of data. Therefore, a technology-based approach, particularly machine learning, is needed to automate the detection process. One algorithm that performs well in text classification is Random Forest, which is an extension of the Decision Tree method that combines multiple decision trees randomly to improve model stability and accuracy (Santoso et al., 2023). This algorithm has been proven effective in detecting various forms of text content such as cyberbullying, hate speech, and negative comments on social media (Julianti et al., 2024).

Although Random Forest has been proven effective for general text classification, its specific application to detect Indonesian-language online gambling promotions, with all their linguistic complexities and slang, remains a challenging area. Previous research may have focused on other negative content or used datasets that did not capture the specific jargon used by gambling promoters. To address this gap, this study proposes an automatic detection model using the Random Forest algorithm to classify online gambling promotion comments on the Twitter (X) platform. The main objective of this study is to develop a robust text classification model by utilizing TF-IDF feature representations to handle the unique vocabulary of online gambling promotions and to comprehensively evaluate the performance of the Random Forest model in accurately identifying such content. This research is expected to provide practical contributions to platform providers and regulators in their efforts to filter harmful content, as well as serve as an academic reference for NLP studies related to cybercrime detection in Indonesia.

1.1 Literature Review

Research related to the analysis of online gambling content on social media has been conducted extensively using various machine learning approaches. The main focus of this research is to analyze public sentiment and understand people's perceptions of the phenomenon of online gambling.

Prastiko and Wiranata (2024) researched public sentiment analysis of the online gambling phenomenon on social media X using the Support Vector Machine (SVM) algorithm. The results showed that the SVM algorithm was able to provide the highest accuracy of 90.59% compared to other models tested. However, the study still used an unbalanced dataset, so the model tended to be biased towards the majority class and was less than optimal in detecting variations in comments with hidden promotional contexts.

Furthermore, Agustia and Suryono (2025) in the Inovtek Polbeng Seri Informatika journal compared the Naïve Bayes, Random Forest, and Logistic Regression algorithms for sentiment analysis related to online gambling. This study used 4,592 data points that had been processed using the SMOTE technique to balance the classes. The test results showed that Random Forest produced the highest accuracy of 78%, followed by Naïve Bayes and Logistic Regression at 77% each. This study proves that ensemble learning methods such as Random Forest have more stable performance in text classification than linear models.

In addition, Anam and Rusdiana (2020) in the Merdeka Pasuruan Informatics Journal emphasize that the use of unbalanced datasets can cause classification results to be biased and unrepresentative. They show that accuracy alone is not sufficient to assess model quality and that data balancing is crucial for models to learn fairly from both classes and produce more meaningful evaluation metrics such as precision, recall, and AUC.

A review of the three studies shows that although machine learning algorithms have been successfully applied to analyze texts related to online gambling, most studies still focus on public sentiment analysis rather than the direct detection of online gambling promotional comments. In addition, data imbalance remains a major obstacle that affects the accuracy and generalization ability of the model.

Based on a review of the literature, there is a clear gap in research on online gambling content detection. Previous studies tend to focus on sentiment and are often hampered by unbalanced datasets. Therefore, this study was designed to overcome these limitations by balancing the dataset and focusing on the task of classifying promotional comments (rather than sentiment) using the Random Forest algorithm, which has been proven to have stable performance (Agustia & Suryono, 2025).

2. Research Methods

This research was conducted in several stages as shown in Figure 1, consisting of data collection, preprocessing, feature extraction, modeling, and evaluation.

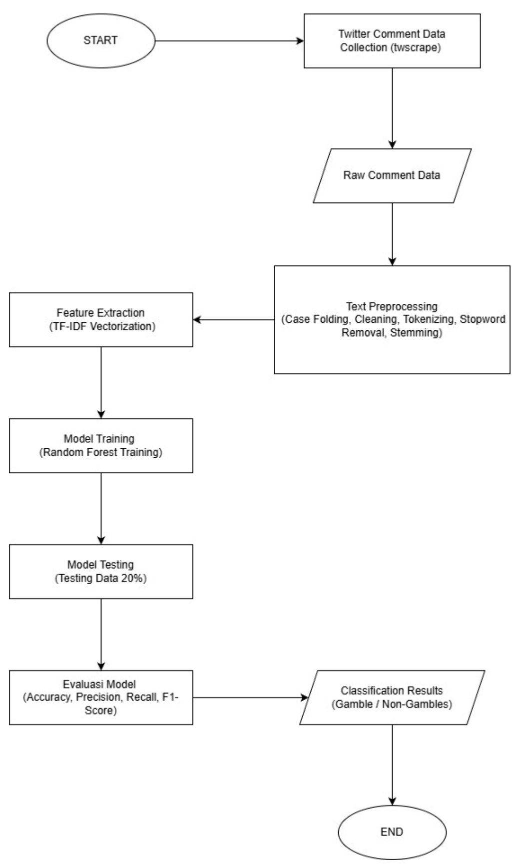

Fig 1. Flowchart

A. Data Collection

Data was collected from the Twitter (X) platform using a scraping method with the Twscrape library in the Python programming language, which allows data retrieval without limiting the number of tweets (Dalam et al., 2024). The collection was carried out using keywords such as "online gambling," "slots," "2D," "4D," and "togel." From this process, 10,607 comments were obtained, consisting of 5,139 non-gambling comments and 5,468 gambling comments, resulting in a balanced dataset.

B. Preprocessing data

Data preprocessing is an important process in data mining analysis that aims to clean, reformat, and prepare data so that it is easier and more accurate in the analysis process.

This stage also aims to convert raw data into a clean, uniform, and structured form so that it is easier to process computationally. The preprocessing process is carried out through the following systematic steps (Shidiq and Alita, 2025):

1. Text Cleaning: Remove URLs, tags, numbers, punctuation marks, emojis, and repeated characters.
2. Case Folding: Change all text to lowercase for consistency.
3. Tokenizing: Splitting text into word tokens.
4. Stopword Removal: Removing common words such as "yang", "dan", dan "itu".
5. Stemming: Return the word to its base form using the Indonesian language library.
6. Selection dan labeling data

C. Feature Extraction

In this study, the Term Frequency–Inverse Document Frequency (TF-IDF) method was used to extract features from pre-processed text data. TF-IDF is a word weighting technique that represents text in the form of numerical vectors based on its level of importance in a document compared to all analyzed tweet data. Words that are more unique and rarely appear in the overall data will receive a higher weight, as they are considered to be more representative of the content of a particular tweet. Conversely, words that are too common and appear frequently in many tweets will be given a low weight because they are not discriminatory. This process is performed automatically using the TfidfVectorizer function from the scikit-learn library, which converts text into numerical features based on word frequency and distribution. This method gives greater weight to important words that appear infrequently but contribute significantly to classification (Prastiko and Wiranata, 2025).

Text extraction can be calculated using the following formula (Shevira et al., 2022).

$$w_{ij} = tf_{ij}\, x \log\left(\frac{D}{df_j}\right) \tag{1}$$

The weight of each word (w) is determined by multiplying the number of words i in a document j (tf) and idf, where (idf) is obtained from the logarithm of the division of the total number of documents (D) by the number of documents containing word j (df).

D. Random Forest Classification

Random Forest is an ensemble method that combines several decision trees trained on random data and features to improve accuracy. Here is how it works (Raff and Ratnawati, 2025):

1. Data Preparation: Labeling and dividing data to form a decision tree.
2. Bootstrap Sampling (Bagging): Create several subsets of data from the original dataset by sampling randomly, allowing for sample duplication.
3. Random Feature Selection: Randomly selecting a subset of features for each node to split the data.

4. Decision Tree Construction: Build decision trees for each data subset using the CART (Classification and Regression Tree) algorithm.
5. Voting (Agregasi): The prediction results from all trees are combined. For classification, majority voting is used; for regression, the average of the prediction results is used.

The Random Forest visualization is shown in Figure 2.

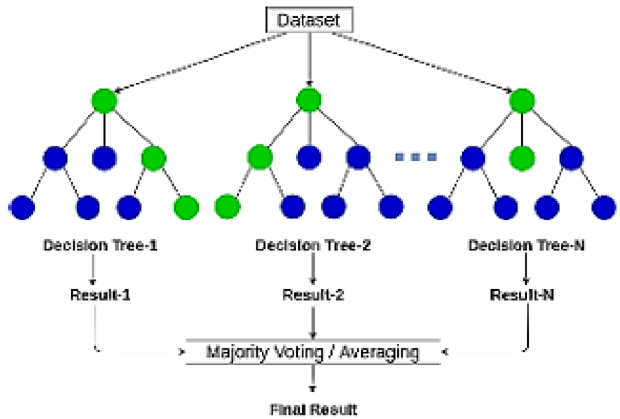

Fig 2. Random Forest Visualization

Entropy€ Formula:

$$E\ (S)\ =\ \sum_{i\,=\,1}^{k} Pi\ log2\ (Pi) \tag{2}$$

Description:

S : A collection of samples (examples, reviews in a subset of data)
Pi : Probability of class III occurrence (650issal, positive or negative)

Information Gain (IG) Formula:

$$IG\ (S, A)\ = E(S)\ -\ \sum_{v\epsilon Values(A)} \frac{|Sv|}{|S|} E(Sv) \tag{3}$$

Description:

A : Features for separating data
$Values(A)$: Set of possible values for feature A
Sv : Subset of data S separated by feature A that has value v
$|Sv|$: Amount of data in the sv subset
S : Total amount of data

E. Evaluation

The evaluation was conducted to determine the performance of the classification model, using several evaluation metrics, namely accuracy, precision, recall, and F1-score. In addition, a confusion matrix was used to assess how many of the model's predictions were correct and incorrect for each class. This evaluation was carried out by comparing the model's prediction results with the actual labels from the previously labeled test data (Husada and Paramita, 2021).

3. Result and Discussion

3.1 Dataset

The dataset used in this study consists of 10,607 comment data obtained from the Twitter (X) platform through the twscrape library. The data is divided into two main categories, namely comments containing online gambling promotions (class 1) and those that do not contain promotions (class 0).

This dataset is balanced, allowing the model to learn proportionally from both classes without bias towards either category. The data then underwent text cleaning, tokenization, stopword removal, and stemming before being converted into numerical form using TF-IDF.

3.2 *Preprocessing Data*

A. Text Cleaning

The data cleaning stage is carried out to clean raw data from unnecessary elements. This process is a very important first step because raw scraping data usually contains various characters that can interfere with text processing, such as symbols, punctuation marks, URLs, emojis, and numbers.

This stage also involves identifying and handling missing (null), outlier, or invalid data. Missing data can be deleted or filled in with appropriate values. Outliers can be deleted or handled using statistical methods such as trimming or winsorizing. At this stage, duplicate data will also be deleted (Agung et al., 2023).

The steps taken in the cleaning process include:

1. Remove punctuation marks, numbers, and special characters such as “!”, “?”, “#”, “@”, and so on.
2. Delete links that usually lead to gambling promotion sites.
3. Remove mentions (@username) and hashtags (#*slotgacor*) because they do not add contextual meaning.
4. Delete empty comments, duplicate comments, and comments containing only symbols or emojis.

The results of the cleaning process are shown in Table 1.

Table 1. Data Cleaning

No	Comments Before Cleaning	Comments After Cleaning
1	Depo 10 k kak sini https://t.co/F1N9RNJCFs	Depo 10k kak sini
2	S3H4T1T0T0 SANGAT BERKUALITAS AMAN & TERPERCAYA PELAYANAN CEPAT SECEPAT KILAT BERGABUNG SEKARANG JUGA !! 💦 Bocoran Slot Gacor Hari Ini 💦 Info RTP Tertinggi Malam Minggu 💦 Situs Resmi Slot Online Paling Hoki Tahun Ini	S3H4T1T0T0 SANGAT BERKUALITAS AMAN TERPERCAYA PELAYANAN CEPAT SECEPAT KILAT BERGABUNG SEKARANG JUGA Bocoran Slot Gacor Hari Ini Info RTP Tertinggi Malam Minggu Situs Resmi Slot Online Paling Hoki Tahun Ini

No	Comments Before Cleaning	Comments After Cleaning
3	Kamu suka denger lagu dangdut apa?	Kamu suka denger lagu dangdut apa
4	Terima kasih atas partisipasinya semoga beruntung ya. ~Gita	Terima kasih atas partisipasinya semoga beruntung ya Gita
5	Paling suka rasa apa?	Paling suka rasa apa

Thus, the cleaning process produces cleaner text that focuses on the main content of the sentence to be analyzed by the system.

B. Case Folding

The case folding stage is carried out to standardize the letters in the entire text to lowercase. This process aims to ensure that the system does not distinguish between identical words simply because of differences in capitalization. For example, the words "Slot," "slot," and "SLOT" will be considered identical.

This stage also helps reduce the number of unique features in the text data, thereby speeding up the model training process and reducing the complexity of calculations.

The results of the case folding process are shown in Table 2.

Table 2. Data Case Folding

No	Cleaned Comments	Case Folding Comments
1	Depo 10k kak sini	depo 10k kak sini
2	S3H4T1T0T0 SANGAT BERKUALITAS AMAN TERPERCAYA PELAYANAN CEPAT SECEPAT KILAT BERGABUNG SEKARANG JUGA Bocoran Slot Gacor Hari Ini Info RTP Tertinggi Malam Minggu Situs Resmi Slot Online Paling Hoki Tahun Ini	s3h4t1t0t0 sangat berkualitas aman terpercaya pelayanan cepat secepat kilat bergabung sekarang juga bocoran slot gacor hari ini info rtp tertinggi malam minggu situs resmi slot online paling hoki tahun ini
3	Kamu suka denger lagu dangdut apa	kamu suka denger lagu dangdut apa
4	Terima kasih atas partisipasinya semoga beruntung ya Gita	terima kasih atas partisipasinya semoga beruntung ya gita
5	Paling suka rasa apa	paling suka rasa apa

C. Tokenizing

The next step is tokenizing, which is the process of breaking sentences down into units of words, or tokens. Each token will become a basic entity that is analyzed in the feature extraction stage using TF-IDF.

With tokenizing, the system can calculate the frequency of each word's appearance and recognize patterns of words that frequently appear in online gambling promotional comments, such as "slot", "bonus", "deposit", or "*gacor*".

The results of the tokenizing process are shown in Table 3.

Table 3. Data Tokenizing

No	Case Folding Comments	Tokenizing Comments
1	depo 10k kak sini	[“depo”, “10k”, “kak”, “sini”]
2	s3h4t1t0t0 sangat berkualitas aman terpercaya pelayanan cepat secepat kilat bergabung sekarang juga bocoran slot gacor hari ini info rtp tertinggi malam minggu situs resmi slot online paling hoki tahun ini	[“s3h4t1t0t0”, “sangat”, “berkualitas”, “aman”, “terpercaya”, “pelayanan”, “cepat”, “secepat”, “kilat”, “bergabung”, “sekarang”, “juga”, “bocoran”, “slot”, “gacor”, “hari”, “ini”, “info”, “rtp”, “tertinggi”, “malam”, “minggu”, “situs”, “resmi”, “slot”, “online”, “paling”, “hoki”, “tahun”, “ini”]
3	kamu suka denger lagu dangdut apa	[“kamu”, “suka”, “denger”, “lagu”, “dangdut”, “apa”]
4	terima kasih atas partisipasinya semoga beruntung ya gita	[“terima”, “kasih”, “atas”, “partisipasinya”, “semoga”, “beruntung”, “ya”, “gita”]
5	paling suka rasa apa	[“paling”, “suka”, “rasa”, “apa”]

D. Stopword Removal

The stopword removal process aims to remove common words (stopwords) that do not contribute significantly to the context of the sentence. In Indonesian, words such as “*yang*” (which), “*di*” (in), “*ke*” (to), “*dan*” (and), “*ini*” (this), “*itu*” (that), or “*dengan*” (with) often appear in texts but do not affect the identification of topics or the semantic meaning of comments.

This step is carried out using a list of Indonesian stopwords from the Sastrawi library, which contains hundreds of common words that can be removed automatically. By removing these irrelevant words, the model becomes more focused on words that contribute to class determination (gambling or non-gambling promotion).

E. Stemming

Stemming is the process of returning each word to its root form. This process is important because in Indonesian, many words have prefixes, infixes, or suffixes. With stemming, words such as “*bermain*” (play), “*dimainkan*” (played), and ‘*memainkan’* (play) will all be returned to “*main*” (play).

The stemming stage ensures that the system does not treat the same word in different inflected forms as separate features. This improves efficiency in feature extraction and improves the performance of the classification model.

F. Selection dan Labeling data

The selection stage was carried out to select data relevant to the research context, namely comments containing elements of online gambling promotion or general non-gambling comments. After that, the data was labeled into two categories:

1. Label 1: Comments containing online gambling promotions (e.g., containing the words “slot”, “bonus”, “depo”, “*daftar*”, and “*gacor*”).
2. Label 0: Comments do not contain promotional elements (general user comments).

Labeling was performed automatically using an application that reads the content of comments to ensure that the classification is truly contextually appropriate. After the labeling stage was completed, a dataset of 10,607 data points was obtained, consisting of 5,468 gambling promotion comments (label 1) and 5,139 non-gambling comments (label 0).

The results of the data selection and labeling process are shown in Table 4.

Table 4. Data Labeling

Label	Comment Category	Amount of data	Percentage
0	Non-Gamble	5.139	48.5%
1	Gamble	5.468	51.5%
Total		10.607	100%

3.3 Algorithm Evaluation

After going through the preprocessing stage, the text is converted into a numerical representation using Term Frequency–Inverse Document Frequency (TF-IDF). The TF-IDF method calculates the weight of each word based on how often it appears in the comments compared to the entire document in the dataset. The more often a word appears in one document but rarely appears in other documents, the higher its weight will be.

An example of TF-IDF feature extraction results is shown in Table 5.

Table 5. Sample TF-IDF Feature Values

No	Word	TF-IDF
1	slot	0.8735
2	depo	0.7628
3	bonus	0.7410
4	gacor	0.6932
5	judi	0.6033

Words such as "slot", "*bonus*", and "*gacor*" have the highest weight, indicating that these words play an important role in identifying comments that contain elements of online gambling promotion. The classification model was developed using the Random Forest algorithm, which falls under the ensemble learning category. This algorithm works by building a number of decision trees and combining their prediction results through a majority voting process. The model was trained using 80% of the training data (8,478 data points) and tested using 20% of the test data (2,120 data points), with the main parameters being n_estimators = 100, criterion = 'gini', max_depth = None, and random_state = 42. The Random Forest algorithm was chosen based on it's ability to handle high-dimensional and complex text data without easily overfitting.

Performance evaluation was carried out using four main metrics, namely accuracy, precision, recall, and F1-score. The model testing results are shown in Table 6.

Table 6. Model Performance Metrics

No	Metrics	Nilai (%)
1	Accuracy	92.97
2	Precision	98.46
3	Recall	87.75
4	F1-Score	92.80

An accuracy value of 92.97% indicates that the model performs very well in detecting online gambling promotional comments. In addition, high precision (98.46%) indicates that most comments classified as "gambling" are indeed promotional comments, while a recall of 87.75% indicates that the model is able to detect most promotional comments effectively.

The distribution of classification results between gambling and non-gambling comments is visualized through a confusion matrix as shown in Figure 3.

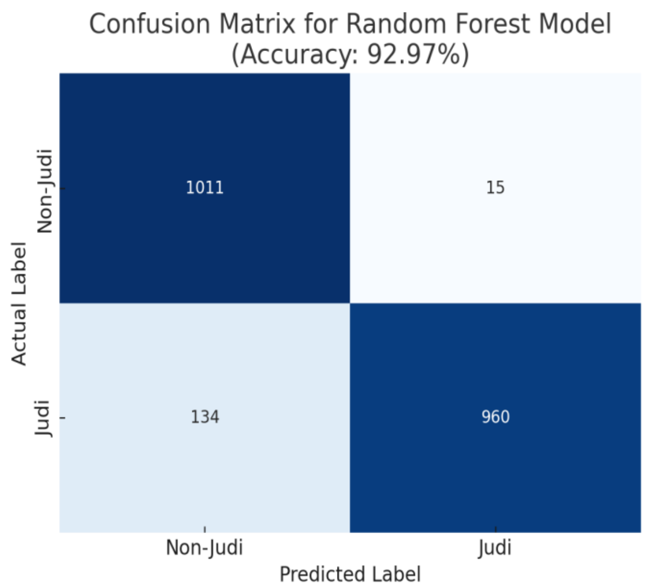

Fig 3. Confusion Matrix for Random Forest Model

From these results, it can be seen that there were 15 false positives (non-gambling comments classified as gambling) and 134 false negatives (gambling comments not detected). The higher false negative value indicates that some comments with ambiguous words, such as "*slot waktu*" or "*bonus belajar*" can still be misclassified. However, this error rate is still relatively low and does not significantly affect the model's performance.

In addition, the training results show that some words contribute significantly to the model's predictions. The words with the highest importance scores are shown in Table 7.

Table 7. Most Important Features

No	Word	Importance
1	menang	0.091
2	slot	0.072
3	judi	0.060
4	banget	0.030

The words “*menang*”, “slot”, and “*judi*” are the strongest indicators in detecting online gambling promotional comments, while the words “bonus,” “*gacor*,” and “*bandar*” also frequently appear in the context of invitations to play or promotions for gambling sites. This indicates that the model is capable of recognizing linguistic patterns commonly used in online gambling promotions, such as invitations (“*bergabung*”, “*join*”, “*daftar*”), rewards (“bonus”, “*menang besar*”), and general terms in the context of gambling (“slot”, “*gacor*”, “*togel*”).

To verify the model's performance, several real comments taken from the test data were tested. The prediction results are shown in Table 8.

Table 8. Example Predictions

Comment	Prediction	Probability
Main slot gacor hari ini, bonus new member 100%!	Gamble	0.85
Prediksi togel hk malam ini pasti tembus.	Gamble	0.72
Saya sedang mengatur slot waktu untuk rapat besok.	Non-Gamble	0.63
Bonus belajar diberikan untuk siswa berprestasi.	Non-Gamble	0.54

The prediction results show that the model is able to distinguish semantic contexts even when ambiguous words such as “slot” or “bonus” are present. This indicates that the model does not rely solely on single keywords but also pays attention to the combination of words in the sentence as a whole. Thus, the combination of the TF-IDF method and the Random Forest algorithm has proven to be effective in automatically detecting online gambling promotional comments with a high level of accuracy, minimal errors, and good contextual ability to distinguish between different meanings in sentences.

4. Conclusions

This study successfully developed a classification model to detect online gambling promotional comments on social media using the Random Forest algorithm combined with the Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction method. Based on the results of testing 10,607 balanced comment data collected through the twscrape library from the Twitter (X) platform, the model achieved excellent performance with an accuracy value of 92.97%, precision of 98.46%, recall of 87.75%, and an F1-score of 92.80%. These results indicate that the approach used is effective in recognizing comments containing elements of online gambling promotion with a low classification error rate.

Feature importance analysis shows that the words “win,” “slot,” and ‘gambling’ have the greatest influence in the classification process, followed by words such as “bonus,” “bookie,” and “gacor.” The model is also able to understand the context of sentences well, as evidenced by its ability to distinguish ambiguous words such as "slot" and "bonus" that can be used in non-gambling contexts. This proves that the model does not only rely on the appearance of keywords but also pays attention to the semantic patterns between words in a sentence.

Compared to previous studies that still faced data imbalance issues and produced lower accuracy, this study shows that the use of balanced datasets and comprehensive preprocessing stages can significantly improve model performance. In addition, the use of twscrape as a free data collection tool is an efficient alternative for obtaining large amounts of data without relying on paid APIs.

5. References

Firmansyah, F. N. F. (2024). JOS | Universitas Jenderal Soedirman. *Soedirman Accounting, Auditing and Public Sector Journal (SAAP)*.

Ghoni, A., Khalilullah, M. F., Pratama, A. A., & Rakhmawati, N. A. (2024). *KOMDIGI and Online Gambling in Indonesia X Tweets Dataset for Sentiment Analysis* (Versi 1.0) [Dataset]. Zenodo.

Julianti, O. N., Suarna, N., & Prihartono, W. (2024). PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANALISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER. *JATI (Jurnal Mahasiswa Teknik Informatika)*.

Santoso, H., Putri, R. A., & Sahbandi, S. (2023). Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest. *Jurnal Manajemen* Informatika *(JAMIKA)*.

Agustia, D. N., Suryono, R. R., Indonesia, U. T., Ratu, L., & Lampung, K. B. (2025). COMPARISON OF NAÏVE BAYES , RANDOM FOREST , AND LOGISTIC REGRESSION ALGORITHMS FOR SENTIMENT ANALYSIS ONLINE GAMBLING KOMPARASI ALGORITMA NAÏVE BAYES , RANDOM FOREST , DAN LOGISTIC REGRESION UNTUK ANALISIS. 10(1), 284–295.

Prastiko, A. D., & Wiranata, A. D. (2025). Analisis Sentimen Publik terhadap Fenomena Judi Online di Media Sosial X dengan SVM. 5–8.

Shidiq, M & Alita, Debby. (2025). ANALISIS SENTIMEN MASYARAKAT TERHADAP KASUS JUDI ONLINE MENGGUNAKAN DATA DARI MEDIA SOSIAL X PENDEKATAN NAIVE BAYES DAN SVM. Jurnal Sistem Informasi dan Informatika (Simika). 8. 24-35. 10.47080/simika.v8i1.3624.

Husada, Hendry & Paramita, Adi. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). Teknika. 10. 18-26. 10.34148/teknika.v10i1.311.

Raff, Hafizah & Ratnawati, Fajar. (2025). Klasifikasi Sentimen Ulasan Pengguna terhadap Aplikasi Video Editing Play Store Menggunakan Random Forest. Techno.Com. 24. 646-657. 10.62411/tc.v24i3.12873.

Anam, C., & Rusdiana, N. (2020). *Analisis Pemeringkatan Kualitas Klasifier Pada Dataset Tidak Seimbang*. *5*(1), 38–44.

Dalam, M., Web, T., Animasi, S., & Nussa, A. (2024). *Dinamika Opini Publik terhadap Undang-Undang Pelindungan Data Pribadi (Kasus Percakapan Media Sosial X)*. *5*(September).

Agung, A., Daniswara, A., & Nuryana, I. K. D. (2023). *Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru*. *05*, 97–100.

Shevira, S., Agus, I. M., Suarjaya, D., & Wira, P. (2022). *Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia*. *3*(2).

Mustaqlilah, R., Widyaningtyas, O., & Wantoro, T. (2023). *Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi*. *2*(1), 18–28. https://doi.org/10.54259/mukasi.v2i1.1346

Pitaloka, E. D., Aprilizdihar, M., & Dewi, S. (2022). *PEMANFAATAN SOSIAL MEDIA SEBAGAI SARANA PEMBELAJARAN*. *5*(1), 40–49.