International Journal of Informatics and Computation (IJICOM)
Vol.7, No.2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Sentiment Analysis of Animated Film “JUMBO” on Twitter Using Random Forest and Semi-Supervised Learning

Eko Rahmat Slamet Hidayat Saputra[1], Arvin Claudy Frobenius[2]

Abstract

This study investigates public sentiment toward the Indonesian animated film "JUMBO" using Twitter data and a semi-supervised machine learning approach. Two thousand fifty tweets were collected and preprocessed to remove noise, standardize text, and extract meaningful features. Data was collected between April 6, 2025, and May 13, 2025, following the film's official release on March 31, 2025, coinciding with its peak public discussion window. A semi-supervised learning strategy was applied, where 532 tweets were manually labelled into positive, neutral, or negative sentiment categories, mitigating the extensive need for labelled data. To address the class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. The labelled data were then used to train a Random Forest classifier, achieving an accuracy of 90% and balanced F1 scores across all classes. The model was subsequently applied to classify the remaining unlabeled tweets, which revealed a dominant proportion of positive sentiments toward the film. These results obtain strong public approval of "JUMBO" and demonstrate the effectiveness of combining machine learning with semi-supervised techniques for sentiment analysis, particularly in the context of local cultural products. This research can be an initial stage in a broader roadmap for analyzing the success factors of Indonesian animated films through AI-driven approaches.

Keywords:
Sentiment analysis, Twitter, Random Forest, Semi-supervised learning, Film review classification.

1. Introduction

Social media's rapid development has transformed how people express opinions and share experiences, including reactions to films and entertainment content [1]. Among social media platforms, Twitter stands out as a rich source of real-time public opinion due to its concise and spontaneous nature [2]. In Indonesia, the animated film "JUMBO" has gained significant attention and set a new record as the most-watched local animated film in history. The film's viral popularity presents a valuable opportunity to understand public sentiment through computational methods.

Sentiment analysis, or opinion mining, has become an essential technique in natural language processing for classifying user-generated content based on emotional tone [3]. Traditional sentiment classification methods require a substantial amount of labelled data, which can be time-consuming and resource-intensive. To address this limitation, semi-supervised learning has emerged as a practical approach that combines a small set of labelled data with a larger pool of unlabeled data to enhance classification performance [4].

This research applies a sentiment analysis approach to categorize public opinions about "JUMBO" into three classes: positive, neutral, and negative. Two thousand fifty tweets were

Corresponding Author: Eko Rahmat Slamet Hidayat Saputra (erachmat@amikom.ac.id)
1 Eko Rahmat Slamet Hidayat Saputra, Universitas Amikom Yogyakarta, erachmat@amikom.ac.id
2 Arvin Claudy Frobenius, Universitas Amikom Yogyakarta, arvinclaudy@amikom.ac.id

collected using keyword-based scraping, with 532 manually labelled and the rest predicted using an RF classifier trained under a semi-supervised learning framework. To address the class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied [5, 6]. The objective of this study is to evaluate public sentiment toward "JUMBO", assess the performance of RF in this context [7], and explore the potential of semi-supervised learning in opinion classification [4].

The novelty of this work lies in integrating traditional machine learning with semi-supervised learning and resampling techniques to analyze the success of a local cultural product [8]. While previous studies have extensively explored sentiment analysis in various domains, including general movie reviews, there remains a notable gap in dedicated research focusing specifically on public sentiment towards unique local animated films, especially in the Indonesian context. This study contributes to the early stages of building a roadmap for evaluating animated film success using AI-based analysis [9][10].

2. Related Works

Sentiment analysis has been widely applied in various domains, including product reviews, political opinion tracking, and entertainment media analysis [1, 11, 12]. Zamzami et al. [13] applied a Modified Balanced RF combined with Mutual Information for feature selection in the context of movie reviews. They achieved notable improvements in sentiment classification accuracy for film reviews. Their work emphasized the importance of handling imbalanced datasets in textual classification tasks.

Another study by Jihad and Sulistyaningsih [11] utilized an RF classifier to analyze sentiment in movie reviews from IMDb. The authors demonstrated that RF performed reliably on medium-sized datasets and could generalize well across sentiment classes [7]. This supports the suitability of RF for sentiment classification in social media contexts [7].

Semi-supervised learning approaches have also been increasingly adopted to reduce reliance on large-scale labelled datasets. Guellil et al. [14] proposed a semi-supervised sentiment analysis framework using RF and deep learning models to classify Arabic social media posts. Their results showed that semi-supervised learning could yield competitive performance while significantly reducing annotation effort [4]. In the Indonesian context, Khomsah and Aribowo [1] implemented a semi-supervised learning model using logistic regression for sentiment annotation. They showed that it reduced the need for expert labelling without degrading performance. Similarly, Ayuningtyas et al. [15] compared LSTM and GRU for semi-supervised sentiment labelling and found that the GRU model produced better accuracy in handling imbalanced textual data.

To address the class imbalance, several studies applied the Synthetic Minority Over-sampling Technique (SMOTE) [5, 6]. Syafutra and Kusrini [16] reported that using SMOTE in combination with machine learning algorithms significantly improved classification results in Twitter-based sentiment analysis, particularly in political discourse. These prior works collectively highlight the effectiveness of RF [7], semi-supervised learning [4], and SMOTE [5, 6]. Limited research, however, has focused on analyzing public sentiment toward local cultural products such as Indonesian animated films [8]. This study builds upon these foundational works by integrating the three techniques to analyze sentiment about the film “JUMBO” [9, 10], contributing a novel perspective to the field of social media-based film analytics.

3. Proposed Method

This study employs a sentiment classification using RF as a powerful ensemble learning method primarily used for classification and regression tasks. In the context of sentiment analysis, RF can effectively classify text data (e.g., tweets, reviews, comments) into categories such as positive, negative, or neutral sentiment. RF is widely used for classification tasks, including sentiment analysis, due to its robustness and ability to handle high-dimensional data. In essence, RF constructs multiple decision trees during training and outputs the mode (for classification) of the classes predicted by individual trees. It combines bagging (bootstrap aggregating) with the random selection of features at each node to ensure low correlation between trees, which improves accuracy and reduces overfitting.

In this study, we define the dataset as $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where x_i is a feature vector extracted from text (e.g., TF-IDF, word embeddings), and $y_i \in \{positive, negative\}$ is the sentiment label. RF generates B decision trees $\{T_1, T_2, \ldots, T_B\}$, each trained on a bootstrap sample $D_b \subset D$ and a random subset of features F_b.

The prediction for a new input x is made by majority vote in equation 1:

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \ldots, T_B(x)) \quad (1)$$

Each tree T_b is built by recursively selecting the best split based on a criterion like Gini impurity in equation 2:

$$Gini(D) = 1 - \sum_{i=1}^{C} p_i^2 \quad (2)$$

where p_i is the proportion of class iii in dataset D, and $C$is the number of classes.

In sentiment analysis, especially with imbalanced data (e.g., more positive than negative reviews), RF may be biased toward the majority class. To overcome this, SMOTE (Synthetic Minority Over-sampling Technique) is applied during preprocessing. SMOTE generates synthetic samples for the minority class by interpolating between existing samples and their nearest neighbors in equation 3:

$$x_{\text{new}} = x_i + \delta \cdot (x_{\text{nn}} - x_i) \quad (3)$$

where x_i is a minority class instance, x_{nn} is one of its k-nearest neighbors, and $\delta \sim U(0,1)$ is a random value. After balancing the dataset using SMOTE, RF is trained on the augmented dataset, which improves its ability to learn minority class patterns and increases generalization performance.

The final model prediction is computed from the ensemble of all decision trees. The strength of RF lies in its variance reduction and low bias, while SMOTE ensures class balance, both contributing to improved performance in sentiment classification. Performance is typically measured using metrics like:

- **Accuracy**: $\frac{TP+TN}{TP+FP+TN+FN}$
- **Precision**: $\frac{TP}{TP+FP}$
- **Recall**: $\frac{TP}{TP+FN}$
- **F1-Score**: $2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$

These metrics validate the classifier's ability to correctly predict both positive and negative sentiments.

4. Experimental Setup

This section outlines the tools, libraries, procedures, and configurations for implementing the proposed sentiment analysis method.

1. Data Collection

This study employed the Python library to efficiently collect Indonesian-language tweets related to the animated film “JUMBO” without requiring Twitter API authentication. Using keywords such as “JUMBO,” “film JUMBO,” and “nonton JUMBO,” the scraping process retrieved 2,050 tweets during the peak discussion period. The collected data captured a diverse range of public opinions and naturally expressed sentiments.

2. Preprocessing and Cleaning

Text preprocessing was carried out using regular expressions and the Sastrawi library, tailored for the Indonesian language. The process involved removing URLs, mentions, hashtags, numbers, emojis, and punctuation to eliminate noise from the raw text. Case folding converted all characters to lowercase, followed by tokenization. Subsequently, stopwords were removed using Sastrawi’s built-in list, and stemming was applied to reduce words to their root form. These steps ensured that only semantically relevant and linguistically normalized tokens were retained, resulting in a clean dataset (`tweet_clean`) suitable for feature extraction and sentiment classification tasks.

3. Labeling and Balancing

This study gathered a total of 2,050 tweets, with 532 tweets manually labeled into three sentiment classes: positive (295), neutral (191), and negative (46). These labeled tweets served as the training dataset for sentiment classification. The remaining 1,518 tweets remained unlabeled and were processed using a semi-supervised learning approach to leverage the larger dataset while minimizing manual labeling effort. Due to the evident class imbalance, particularly the underrepresentation of negative sentiment, we adopt the SMOTE approach. SMOTE generated synthetic examples of the minority class to balance the dataset and mitigate bias during training. This preprocessing step ensured that the classifier learned equally from all sentiment classes, thereby improving the robustness and fairness of the sentiment analysis model.

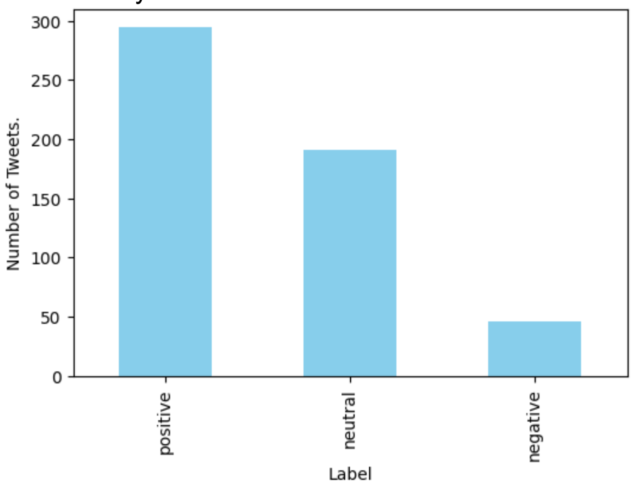

Fig. 2. Label Distribution Before SMOTE

This study fixes an imbalanced dataset using SMOTE (Synthetic Minority Over-sampling Technique), which synthetically generates new samples for the minority classes to balance the dataset. After applying SMOTE, each sentiment class had 295 instances, resulting in a balanced training set of 885 samples.

4. Feature Extraction and Model Configuration

We utilize the vectorization technique using the TF-IDF method to convert textual data into numerical features. This approach captured the relevance of each term relative to the tweet corpus, enhancing the model's ability to distinguish key sentiment-bearing terms. The feature vectors were then input into a RF classifier implemented via scikit-learn. The RF model was configured with 100 decision trees (`n_estimators=100`) and a fixed `random_state=42` to ensure reproducibility. The model was trained on 80% of the labeled dataset and tested on the remaining 20%. Evaluation metrics—accuracy, precision, recall, and F1-score—were employed to assess classification performance, providing a comprehensive view of the model's ability to generalize across sentiment categories.

5. Prediction and Visualization

This study trained the RF model to classify the remaining 1,518 unlabeled tweets into sentiment categories and analyzed them to gauge overall public opinion toward the film. This analysis included visual representations: pie charts illustrated the proportion of each sentiment class, while word clouds highlighted the most frequent terms associated with each sentiment category. The entire experimental process was conducted using Python in a Google Colab environment.

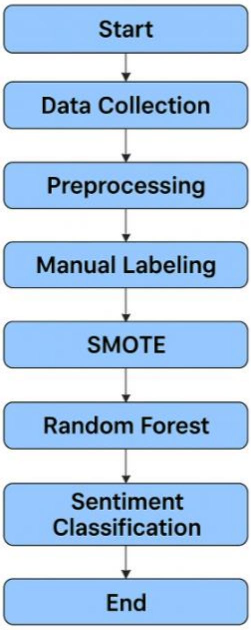

Fig. 1. Research Flow

5. Results and Analysis

This section evaluates the sentiment classification model and analyzes the predicted results on unlabeled tweets regarding the animated film "JUMBO". After training on the balanced labelled dataset (via SMOTE), the RF classifier achieved the following performance on the 20% test set (n = 177 tweets), as Table 1.

Table 1. Classification Report

Class	Precision	Recall	F1-score
Positive	0.83	0.93	0.87
Neutral	0.90	0.79	0.84
Negative	1.00	0.98	0.99
Macro Avg	0.91	0.90	0.90

The classifier achieved an overall accuracy of 90%, with a strong F1-score balance across all sentiment classes. The results demonstrate the effectiveness of combining SMOTE and RF for sentiment classification on imbalanced data. To provide clearer insight into the classification, Table 2 presents examples of tweets categorized by their predicted sentiment:

Table 2. Examples of Classified Tweets

Sentiment Category	Example Tweet
Positive	“Walaupun ada celah di satu dua hal tapi jujur bangga sekali sama Jumbo. Akhirnya Indonesia punya film lokal berkualitas yang 'anak-anak' banget. Karakter animasi soundtrack semuanya berkesan. Congrats Jumbo.”
Neutral	“Jumbo film terlaris dalam sejarah masih tayang di bioskop cek jadwalnya selasa 13 mei 2025”
Negative	“Jumbo menurut gw jelek alurnya kecepeten karakternya nyebelin”

According to the experimental result, the sentiment prediction on the remaining 1,518 unlabeled tweets revealed a strong dominance of positive sentiment, with 1,171 tweets (77.1%) classified as positive, followed by 340 neutral tweets (22.4%) and only 7 negative tweets (0.5%). This distribution reflects an overwhelmingly favorable public response to the film *JUMBO* on Twitter. The high prevalence of positive sentiment aligns with the film’s commercial success and suggests strong audience approval. However, the notably low proportion of negative sentiment may also reflect biases in user expression, such as social desirability or selective sharing behavior, where users are more inclined to post positive opinions or avoid negative critiques in public discourse.

To enhance interpretability, word clouds were generated for each sentiment class. Figure 3 shows the word cloud for positive opinions, prominently featuring words such as “bagus”, “anak”, and “suka”, suggesting that the film resonated well with families and children. Figure 4 illustrates the word cloud for neutral opinions, where tweets often contain promotional or informational content (e.g., “tiket”, “tonton”). Conversely, Figure 5 displays the word cloud for negative opinions, highlighting dissatisfaction with animation quality or storyline in the few negative tweets.

Fig. 3. Word Cloud for Positive Opinion

Fig. 4. Word Cloud for Neutral Opinion

Fig. 5. Word Cloud for Neutral Opinion

A pie chart further illustrated the sentiment composition, affirming the overwhelming positivity in public opinion. Figure 6 presents this distribution, clearly showing the large proportion of positive sentiment, moderate neutral sentiment, and very low negative sentiment, consistent with the numerical breakdown provided above.

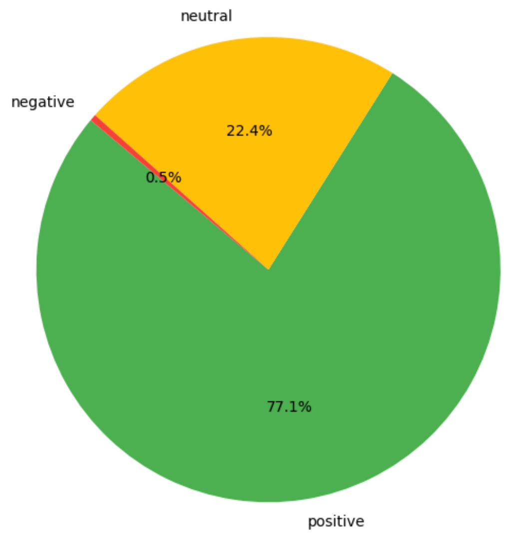

Fig. 6. Distribution of Opinions

The experimental results validate the proposed semi-supervised learning strategy for effective sentiment classification with minimal manually labelled data. The model exhibited high accuracy and robustness, even with only 532 manually labelled tweets. Additionally, SMOTE was critical in balancing class distributions and improving model generalization. The findings support that social media sentiment can be a reliable proxy for measuring public perception of cultural products. Moreover, this approach can be extended to comparative studies involving other local films to extract success indicators. Therefore, this approach demonstrates an efficient sentiment classification pipeline in semi-supervised learning contexts, maintaining robustness despite limited labeled data and ensuring balanced prediction outcomes.

6. Conclusion

This study investigated public sentiment toward the Indonesian animated film "JUMBO" using a machine learning-based sentiment classification framework. The study successfully classified 2,050 tweets with high accuracy using only 532 manually labelled samples by applying a semi-supervised learning approach with RF and SMOTE. The trained model achieved an accuracy of 90%, and the predicted distribution of sentiments revealed that public response to the film was overwhelmingly positive. Integrating SMOTE proved essential in balancing the imbalanced dataset and improving the model's performance, particularly for the minority class (negative sentiment). TF-IDF for feature extraction and RF for classification demonstrated effectiveness in a multilingual, informal-text setting like Twitter. This research contributes to the growing field of social media analytics in the cultural domain by providing a scalable method to measure public perception. It also sets a precedent for using AI to evaluate the success factors of local creative industries.

However, this study has several limitations. The data collection was limited to tweets from a specific time window, which might not capture long-term sentiment shifts or initial reactions outside that period. Furthermore, reliance on keyword-based scraping might miss relevant discussions that do not explicitly use the chosen keywords. The manual labelling process, while rigorous, is inherently subjective to some extent, and the focus on three sentiment classes (positive, neutral, negative) might oversimplify nuanced emotional expressions. Finally, while effective, the choice of RF might be surpassed by more complex deep learning models for specific linguistic nuances or large datasets.

Future studies may expand this research by incorporating multimodal sentiment analysis (e.g., images, emojis), analyzing temporal sentiment shifts, or comparing sentiments across different platforms (e.g., YouTube, TikTok). Additionally, implementing deep learning models such as BERT or LSTM may yield improved performance for longer and more nuanced texts.

Acknowledgement

The author would like to sincerely thank Universitas Amikom Yogyakarta, particularly the Faculty of Computer Science, for their academic support and research environment that made this study possible. Appreciation is also extended to colleagues and reviewers who provided valuable feedback during the development of this work. Special thanks to the annotators who assisted in the manual labelling process of the tweet dataset.

References

[1] S. Khomsah and A. S. Aribowo, "Semi-supervised learning model using logistic regression for sentiment annotation," *LEDGER: Journal Informatics and Information Technology*, vol. 1, no. 4, pp. 171–174, 2022.

[2] Albladi, M. Islam, and C. D. Seals, "Sentiment analysis of Twitter data using NLP models: A comprehensive review," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2025, doi: 10.1109/ACCESS.2025.3541494.

[3] K. Cortis and B. Davis, "Over a decade of social opinion mining: a systematic review," *Artificial Intelligence Review*, vol. 54, no. 7, pp. 4873–4965, 2021, doi: 10.1007/s10462-021-10030-2.

[4] M. Jafarlou and M. M. Kubek, "Reducing labeling costs in sentiment analysis via semi-supervised learning," *ACM Reference Format*, vol. 1, no. 1, 12 pages, 2024, arXiv:2410.11355v1 [cs.LG].

[5] G. Husain, D. Nasef, R. Jose, J. Mayer, M. Bekbolatova, T. Devine, and M. Toma, "SMOTE vs. SMOTEENN: A study on the performance of resampling algorithms for addressing class imbalance in regression models," *Algorithms*, vol. 18, no. 1, pp. 37, 2025, doi: https://doi.org/10.3390/a18010037.

[6] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, pp. 106, 2013, doi: 10.1186/1471-2105-14-106.

[7] G. Venkateshwarlu, S. Akhila, V. Kavyasree, S. Vishnu, and V. S. Prasad, "Enhanced text classification using random forest: Comparative analysis and insights on performance and efficiency," *International Journal of Computer Engineering in Research Trends (IJCERT)*, vol. 11, no. 1s, pp. 1–8, 2024, doi: https://doi.org/10.22362/ijcert/2024/v11/i1/v11i1s01.

[8] Indrajit, "Shaping global perceptions: A sentiment analysis of Indonesian made boots," *Management Studies and Entrepreneurship Journal*, vol. 6, no. 3, pp. 2758–2770, 2025.

[9] L. Li, Z. Guan, and F. Xing, "Text sentiment analysis of film reviews using Word2Vec-LSTM," in *Third International Conference on Electronic Information Engineering, Big Data, and Computer Technology (EIBDCT 2024)*, vol. 13181, p. 131810T, SPIE, 2024.

[10] S. Onalaja, E. Romero, B. Yun, and F. Javed, "Aspect-based sentiment analysis of movie reviews," *SMU Data Science Review*, vol. 5, no. 3, pp. 10, 2021.

[11] M. A. A. Jihad and E. Sulistyaningsih, "Sentiment analysis of film reviews using the Random Forest algorithm," *Jurnal e-Proceeding Teknik Informatika*, vol. 8, no. 1, pp. 98–105, 2021.

[12] T. Velmurugan, M. Archana, and A. S. Nongmaithem, "Analysing effectiveness of sentiments in social media data using machine learning techniques," *Journal of Computer and Communications*, vol. 13, no. 1, pp. 136–151, 2025, doi: 10.4236/jcc.2025.131010.

[13] F. N. Zamzami, A. Prasetyo, and Y. Nugroho, "Sentiment analysis of film reviews using Modified Balanced Random Forest and Mutual Information," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 15, no. 2, pp. 63–72, 2021.

[14] Guellil, F. Azouaou, and A. Adeel, “A semi-supervised approach for sentiment analysis of Arabic messages,” *Complex & Intelligent Systems*, vol. 7, pp. 2657–2670, 2021, doi: 10.1007/s42979-021-00510-1.
[15] P. Ayuningtyas, S. Khomsah, and Sudianto, “Sentiment labeling based on semi-supervised learning using LSTM and GRU algorithms,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 217–229, 2024.
[16] D. Syafutra and Kusrini, “Sentiment analysis of the Omnibus Law on Twitter using machine learning and resampling techniques,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 12, no. 1, pp. 55–64, 2025.