Sentiment Analysis for Comparing the Performance of Random Forest and Decision Tree Algorithms in Predicting Presidential Election Results in Indonesia

1st Kevin Husodo
Computer Science
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11530
kevin.husodo@binus.ac.id

2nd Putri Febiyani
Computer Science
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11530
putri.febiyani@binus.ac.id

3rd Muhammad Fikri Hasani
Computer Science
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11530
muhammad.fikri003@binus.ac.id

4th Pandu Wicaksono
Computer Science
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11530
pandu.wicaksono005@binus.ac.id

***Abstract*—Indonesia's presidential election is important for its democracy, with social media playing a massive role in shaping public opinion amid the challenge of disinformation. Sentiment analysis using Decision Tree and Random Forest algorithms enables expected results and enhances political strategies. Random Forests are better at handling overfitting and complexity, while Decision Tree is easier to interpret but liable to overfitting, providing insights to improve political participation and decision-making. This research uses Python to predict the effects of the Indonesian presidential election and utilizes data mining techniques, specifically the Random Forest and Decision Tree algorithms. Twitter data collected from Mendeley Data, including 30,000 tweets about Indonesian presidential candidates, was preprocessed through steps such as cleaning, normalization, and tokenization. Model training and evaluation use accuracy metrics and confusion matrices to assess performance. Random Forest shows better recall for the positive class, while Decision Tree excels in precision. Understanding Decision Tree is slightly more accurate (83.93%) than Random Forest (83.22%), and this helps to improve political strategies by forecasting public opinion. This study compares the use of random forests and decision trees for sentiment analysis on Indonesian presidential candidate's tweets. Every model has nearly equal precision, recall and F1-score, but Decision Tree only increases accuracy a little bit. For the positive class Random Forest performs well in recall. These results will enable one to choose between models depending on the classification requirements as well as interpret findings from sentiment analysis activities.**

Keywords—Random Forest, Decision Tree, Sentiment Analysis, Presidential Election, Twitter.

I. Introduction

The presidential election in Indonesia is a pivotal moment in the country's democratic journey. Since the democratic transition in 1998, Indonesia has conducted eight direct presidential elections, demonstrating increasing public political participation [1]. As the world's fourth most populous country with rich cultural diversity, Indonesia's presidential election significantly impacts political stability and the direction of national development [2].

In this digital era, social media has emerged as the primary platform for individuals to express their opinions and sentiments on various issues, including politics [3]. Indonesia boasts a large social media user population alongside a high degree of political polarization [4]. Disinformation and misinformation also pose serious challenges during elections. However, political participation remains relatively low [5]. In this context, sentiment analysis becomes a valuable tool for understanding and analyzing public opinion, promoting political participation, strengthening democracy, and ensuring that people's voices are heard and considered in the decision-making process [6].

Sentiment data analysis and prediction are crucial tools in supporting presidential elections. The increasing availability of data creates opportunities to employ machine learning algorithms for predicting election outcomes with high accuracy [7]. This enables stakeholders such as political parties, candidates, and political analysts to formulate strategies and make more effective decisions[8].

Popular machine learning algorithms in the realm of political elections include Decision Tree and Random Forest [9]. Decision Tree provides an easy-to-understand and interpret approach but is prone to overfitting. On the other hand, Random Forest, an extension of Decision Tree, mitigates overfitting and enhances prediction accuracy by aggregating multiple decision trees [10].

This study aims to compare the performance of Decision Tree and Random Forest in predicting the outcomes of the presidential election in Indonesia. Using sentiment data from previous presidential elections, this study will analyze the advantages and disadvantages of both algorithms and identify the factors contributing to prediction accuracy.

This study aims to compare the performance of Decision Tree and Random Forest in predicting the outcomes of the presidential election in Indonesia. Using sentiment data from previous presidential elections, this research will assess the advantages and disadvantages of each algorithm and explore the factors contributing to prediction accuracy. The anticipated outcomes of this study are to contribute to the scholarly literature on data analysis and prediction in political elections. This research is expected to assist stakeholders in understanding the potential and limitations of machine learning algorithms in predicting the results of the presidential election in Indonesia and provide insights for developing more effective strategies and decision-making processes.

II. Literature Review

A. Sentiment Analysis and Election Projections

Previous studies on Indonesian general elections, especially the 2019 elections for President, Vice President, DPR (House of Representatives), DPD (Regional Representative Council), and DPRD (Regional House of Representatives) members, utilized in-depth statistical and qualitative methods. Findings underscore the elections' crucial role in promoting effective governance, justice, and social welfare through transparent and integrity-driven democratic mechanisms [11].

Studies on first-time voter participation emphasize diverse forms of democratic engagement, including voting, participation in political campaigns, discussions, and media consumption. These studies highlight the importance of these activities in fostering a democratic society [12]. Another study examines the evolution of Indonesia's political survey institutions, emphasizing the importance of adhering to professional standards and ethical conduct for their credibility and contribution to political democracy [13]. Sentiment analysis, crucial in election projections, faces challenges in processing informal social media text. However, lexicon-based techniques, along with machine learning methods like Naïve Bayes, effectively address these issues, offering insights for accurate election forecasts [14], [15].

Moreover, recurrent model-based approaches such as LSTM address the complexities of multi-emotion expressions in text data, enhancing sentiment analysis for political election projections [16]. Overall, sentiment analysis, drawing from diverse data sources, plays a critical role in predicting political election outcomes, providing valuable insights into voter preferences and current political dynamics. Methods like Random Forest or Decision Tree further refine sentiment analysis for election projections.

B. Social Media

The evolution of social media has significantly impacted people's perceptions, social interactions, and learning methods over time. Studies spanning from 1992 to 2023 highlight profound shifts in human social behaviors and self-perceptions attributed to social media engagement. This phenomenon underscores the critical need to foster digital literacy within educational settings to harness the advantages of social media while mitigating potential negative outcomes. [17].

A quantitative study utilizing a cross-sectional design investigates the relationship between social media exposure and mental health among individuals during the COVID-19 pandemic. Findings from a survey of 430 individuals highlight a significant correlation between high social media exposure and increased levels of depression and anxiety disorders. These insights underscore the urgent need for increased attention to mental health among the population, emphasizing the role of educational institutions and mental health services in providing adequate support [18].

C. Random Forest

Random Forest is a critical ensemble learning approach in contemporary statistical analysis, offering numerous benefits and extensive applications. However, it also presents certain limitations that need to be considered. Studies examining the use of Random Forest (RF) in medical image classification highlight its advantages in overcoming overfitting in complex datasets, thereby producing robust and reliable models for identifying disease states. Nevertheless, it is noted for its lower interpretability compared to linear models. Our meta-analysis, based on a database of 251 papers, where 42% and 68% apply RF and SVM respectively, aims to statistically compare these classifiers in terms of frequency and accuracy [19].

Recent research evaluating the application of Random Forest in network security analysis shows the effectiveness of this technique in detecting anomalous patterns in complex network traffic. By combining multiple Decision Trees, Random Forest can identify cyber-attacks and intrusions with a high degree of accuracy, often achieving accuracy scores as high as 93.01% on datasets like UNSW-NB15 and 99.90% on the TON_IoT network with appropriate feature selection. However, challenges related to parameter settings and proper data processing remain a major concern in implementation [20].

D. Decision Tree

Segmentation responsibilities, consisting of disposing of blood vessels for Optic Disc (OD) segmentation with an accuracy of 99.61%. numerous optimization strategies can further beautify the performance of DT, in particular in datasets like UCI ML and CICIDS2017, ensuing inside the highest accuracy in type and segmentation responsibilities [21].

Furthermore, several research highlight ongoing efforts to enhance the overall performance of ID3 and C4.5 algorithms in records category. those efforts recognition on optimizing the characteristic choice process, including the usage of the IGR criteria in C4.5, decreasing features using fraction department strategies, or using optimization algorithms like PSO. these findings display advanced accuracy and performance in data type, mainly in medical domains which include coronary heart disorder and breast cancer analysis. but, in addition research is needed to use those findings in actual-global scenarios and measure adoption with the aid of researchers and industry[21].

III. Methodology

This study uses data mining techniques with the Python programming language to predict the results of presidential candidate elections. The methods used include Random Forest and Decision Tree, which are generally used in election data analysis.

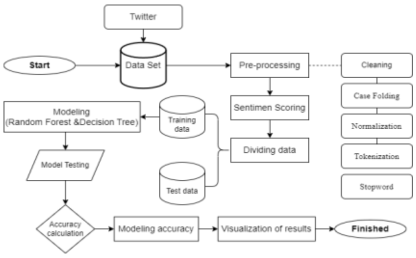

Fig.1. Flowchart

A. Data Collection

The information series calls for the usage of social media networks to reap associated records. Programs consisting of Twitter are used to accumulate information to analyze sentiment, which is used to expect election effects in Indonesia. The social account of the president in question is where this record is obtained. Since it serves as a foundation for next ranges in sentiment analysis, including record processing, feature extraction, model training, and model evaluation, the record collection method is critical.

B. Data Preprocessing

Sentiment analysis requires preparing information should, especially Twitter records. Preparing text records for machine learning model requires cleaning and organizing it. The model's performance can be significantly influenced by effective preprocessing. Data preprocessing includes:

- Cleaning: Exclude non-valuable text such as URLs, mentions, hashtags, numbers and special characters.
- Case Folding: Keep text in lower case to maintain consistency
- Normalization: Enhancing an unstructured textual content which is essential to improve the quality of casual text like tweets.
- Tokenization: Division of a piece of writing into single words or tokens
- Stop Words removal: Get rid of words that are excessive such as "or", "and" and "the"

These techniques do this by enhancing the data on the texts for sentiment analysis thus enabling the model to identify underlying emotions better. However, different training techniques may be used depending on the data set and task at hand [22].

C. Data Labeling

Crucial in the analysis of data for predicting election outcomes in Indonesia is the creation of labels. This starts by identification of a target variable. After that, each data entry can be assigned one among two labels: positive or negative. Consistency in labeling is important as it will contribute to accurate predictions from the developing model. Besides this, methods should be applied for treating missing values and incomplete information to ensure correct model specification. Therefore, label creation is an important stage in sorting out data before it can be analyzed and modeled.

D. Model Training

The use of preprocessed information, system getting to know strategies like Random forest and decision Tree are educated to create accurate models that expect sentiment. The performance of those models is classified using measures which include accuracy.

- Random forest: This ensemble studying approach builds a couple of selection bushes on unique information subsets to decorate precision and resilience. With the aid of aggregating the predictions of those timber, Random Forest improves generalization and decreases overfitting.
- Decision Tree: This technique constructs a tree-like shape where internal nodes constitute selections based on functions, branches represent choice effects, and leaf nodes represent final predictions. Choice bushes are accurate, smooth to interpret, and visualize by splitting statistics at nodes based at the maximum giant capabilities [23].

During training, preprocessed data is fed into the algorithms, and model performance is evaluated on a separate validation set. Iterative training and hyperparameter tuning are used to optimize model performance, enabling the model to predict the sentiment of new, unseen data.

E. Model Evaluation

comparing a sentiment analysis version entails determining how nicely it plays. To assure the correctness and efficacy of the version, overall performance evaluation is critical. the subsequent are some common measures for assessing sentiment evaluation fashions:

- Confusion Matrix: A confusion matrix shows the performance of the version by evaluating the predicted and actual emotion labels. It provides correct negatives, accurate positives, and fake negatives.

IV. Result And Analysis

A. Data Collection

The facts turned into accumulated through Mendeley information, a web platform that offers free get entry to to numerous datasets. The dataset used on this look at turned into posted on December 13, 2023 and consists of 30,000 entries. The decision to apply Mendeley statistics become primarily based on its reputation as a source of high-quality, relevant, applicable and reliable datasets. each entry in the dataset is organized in a established manner and stored in CSV format, making it clean to get admission to and examine the information. extra facts approximately this dataset may be determined at: https://data.mendeley.com/datasets/7w5zvr8jgp/5.

	Date	Text	Username
0	Wed Dec 13 17:59:27 +0000 2023	Poling Diadakan Oleh Kader PSI Dan Yang Menang...	msw_andi
1	Wed Dec 13 17:48:49 +0000 2023	Unggul Telak Dalam Debat Capres Anies Basweda...	msw_andi
2	Wed Dec 13 17:32:11 +0000 2023	Sihlakan Retweet bagi yang dukung @aniesbaswed...	NafisahKH2022
3	Wed Dec 13 16:50:05 +0000 2023	@DPP_PKB @aniesbaswedan @cakimiNOW good job pa...	pikiranlugu
4	Mon Dec 18 01:40:53 +0000 2023	@Fahrihamzah Wakanda No More ? Indonesia Forev...	RakhaBilly6

Fig 2. Dataset

The dataset used in this research includes the text of the tweet uploaded, the username of the tweet creator, and the date the tweet became uploaded. After that, the information goes through a preprocessing degree, that's the guidance of statistics earlier than it's miles carried out to device studying models or algorithms. the stairs in preprocessing include information cleansing, facts transformation, reproduction removal, normalization, encoding, categorization, dimension discount, and elimination of beside the point features. Preprocessing objectives to refine and adjust the layout of uncooked facts to make it extra suitable for modeling evaluation. by performing this technique, the statistics may be equipped to be analyzed using device getting to know techniques, making sure accurate and meaningful analysis consequences.

Table 1. Preprocessing

No	Before Preprocessing	After Preprocessing	candidate
1	I think Prabowo is the best choice for our country! 😊	prabowo is the best choice for our country	Prabowo Subianto
2	Ganjar's policy is truly impressive, but can he realize it? 🤔	ganjar's policy is truly impressive, but can he realize it	Ganjar Pranowo
3	Anies has done a lot for our city, he is a true leader!	anies has done a lot for our city, he is a true leader	Anies Baswedan

B. Data Labeling

Creating labels is a key step in data analysis to predict the results of General Elections in Indonesia. This process begins by identifying the target variable. Each data entry is then labeled based on positive and negative labels. It is important to ensure consistency in labeling so that the developed model can provide accurate predictions.

	Final_text	label
0	anies president info	Positive
1	gerindra party politician sandiaga uno answers...	Positive
2	mr anies continued guard becomes president	Positive
3	may allah swt save nation state republic indon...	Positive
4	poor chotimah uncle anies family decided ele...	Positive
...	...	...
7831	populi center survey ganjar pranowo presidenti...	Positive
7832	pa ganjar pranowo nex president indonesia	Negative
7833	meaning reward pranowo president	Positive
7834	people ready support ganjar run president repu...	Positive
7835	happy national aviation day october fly high ...	Positive

Fig 3. Labeling

C. Classification Report

In the classification modeling stage, data is divided with a ratio of 75% for training data and 25% for testing data using the random forest and decision tree methods. The model created uses a random state value of 42.

Fig 4. Classification Report RF

In the graph, three categories are evaluated using Precision, Recall, and F1-Score, represented by blue, green, and red, respectively.

- Precision: 0.78 for the negative class and 0.85 for the positive class.
- Recall: 0.54 for the negative class and 0.94 for the positive class.
- F1-Score: 0.64 for the negative class and 0.89 for the positive class.
- Macro Averages: These provide an overall performance measure, indicating good precision, higher recall, and a balanced F1-Score across classes.

Overall, the model shows better performance in identifying the 'positive' class than the 'negative' class, as shown by higher recall and F1-score values for the 'positive' class. However, there is a tendency for the model to more often incorrectly predict the 'negative' class as 'positive', which is reflected in the high false positive rate.

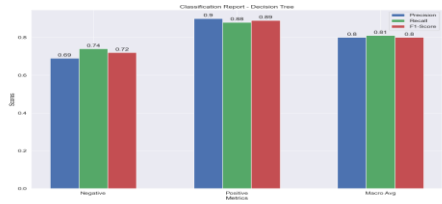

Fig 5. Classification Report DT

In the graph, three categories are evaluated using Precision, Recall, and F1-Score, represented by blue, green, and red, respectively.

- Precision: The model's precision is 0.78 for the negative class and 0.85 for the positive class.
- Recall: The recall is 0.54 for the negative class and 0.94 for the positive class.
- F1-Score: The F1-score is 0.64 for the negative class and 0.89 for the positive class.
- Macro Average: The Decision Tree model shows good overall performance with balanced precision, recall, and F1-Score, indicating effective classification of the data.

Based on test results with decision trees, the model has good performance in classifying positive classes with high precision, recall and F1-score. However, there are challenges in classifying the negative class, as shown by the low recall. Nevertheless, the model can well recognize the positive class.

Model Comparative Evaluation

Best	Positif	Negatif
Precision	Decision Tree	Random Forest
Recall	Random Forest	Decision Tree
F1-Score	Both	Decision Tree

Fig 6. Comparative Evaluation

The table (Fig. 6) reflects the predictive ability of each model on test or validation data. Decision Tree stands out in precision for positive classes, while Random Forest has a higher recall for positive classes. This shows the difference in the two models' approach to data and prediction. Random Forest combines many decision trees, resulting in different variations in the data, which can increase recall for positive cases. Meanwhile, Decision Tree focuses on direct decisions based on data features, which can increase precision for positive cases.

D. Data Modelling

This research applies topic modeling to identify patterns and trends in large and complex text data. By grouping documents based on content similarity, topic modeling facilitates efficient exploration and understanding of text data. This technique also identifies keywords or main topics within each group of documents.

The word cloud method is used to visualize the main topics that often appear related to elections in Indonesia. The positive and negative word cloud of the Random Forest and Decision Tree algorithms are attached to show the topics identified by each algorithm, giving an idea of how these two algorithms process sentiment in predicting election results.

Fig 7. Positive RF

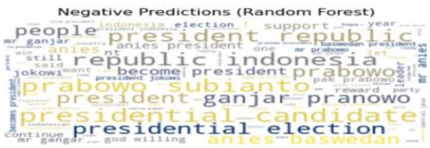

Fig 8. Negative RF

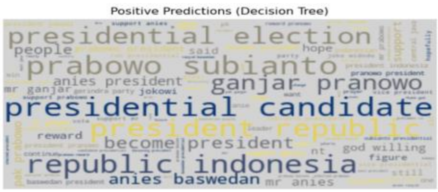

Fig 9. Positive DT

Fig 10. Negative DT

Two word clouds depict sentiment prediction results for Indonesian presidential candidates using the Random Forest and Decision Tree algorithms. The Random Forest word clouds show positive sentiments with words like "presidential candidate," "republic indonesia," "Prabowo Subianto," and "Anies Baswedan" (Fig 7. Positive RF), and negative sentiments with terms like “problem,” "issue," and candidates' names in a negative context (Fig 8. Negative RF). The Decision Tree word clouds reveal positive sentiments with "Prabowo Subianto" (Fig 9. Positive DT) and negative sentiments with "president" and "Anies" (Fig 10. Negative DT). Both algorithms indicate the presence of positive and negative sentiments surrounding the candidates, with Prabowo Subianto featured prominently in both contexts.

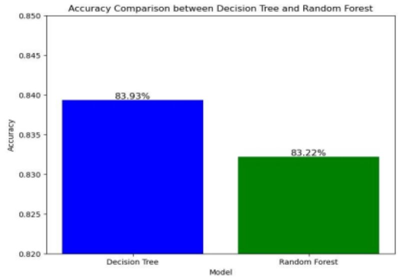

Fig 11. Accuracy Algorithm

Bar graph results from measuring accuracy between two machine learning models, namely Decision Tree and Random Forest. This graph visualizes the performance comparison of both models in terms of accuracy. These results indicate that, in general, the Decision Tree model is slightly superior in terms of accuracy compared to the Random Forest model for the dataset used in this experiment.

V. Conclusion

Consequently, based on the data, using Random Forest and Decision Tree techniques for prediction purposes, several deductions and suggestions can be made.:

- In this research, Random Forest and Decision Tree are the two sentiment analysis techniques employed. Random Forest is one of the best classification tools because it uses a machine learning algorithm while Decision Tree can also be used on sentiment analysis tasks to determine which path will lead to a better solution.
- Based on the data collected, Decision Tree had an accuracy of 83.93% while random forest had an accuracy of 83.21%. This tiny difference implies that Decision Tree is only slightly more precise than Random Forest in terms of accuracy. In other words, for this specific situation, when compared to Random Forest, Decision Tree is slightly more accurate.
- However, even though the Random Forest model was found to have some merits in several aspects, there was a significant disadvantage due to recall being so low as 54%. Having this little recall indicates that the model has difficulty capturing all authentic positives. For instance, this may result in most positive sentiments being left out by the model while they should have been classified
- This paper can provide a quick contrast between two unique techniques of studying systems, Random Forest and Decision Tree, in the area sentiment analysis around presidential aspirants. Which may be used as a guide to help choose the best approach for similar sentiment analysis tasks.

References

[1] M. Kitanova, “Youth political participation in the EU: evidence from a cross-national analysis,” *J*

Youth Stud, vol. 23, no. 7, pp. 819–836, Aug. 2020, doi: 10.1080/13676261.2019.1636951.

[2] M. Lane, “Understanding Indonesia’s 2024 Presidential Elections: A New Polarisation Evolving”.

[3] P. Rita, N. António, and A. P. Afonso, “Social media discourse and voting decisions influence: sentiment analysis in tweets during an electoral period,” *Soc Netw Anal Min*, vol. 13, no. 1, Dec. 2023, doi: 10.1007/s13278-023-01048-1.

[4] P. Anggraheni, N. T. Setyowati, and H. Harry, “SOCIAL MEDIA AND POLITICAL PARTICIPATION IN INDONESIA: RESTRICTIONS ACCESS AT ANNOUNCEMENT RESULTS OF 2019 PRESIDENTIAL ELECTION,” *ASPIRATION Journal*, vol. 2, no. 1, pp. 85–123, Jul. 2021, doi: 10.56353/aspiration.v2i1.23.

[5] M. Cantarella, N. Fraccaroli, and R. Volpe, “Does fake news affect voting behaviour?,” *Res Policy*, vol. 52, no. 1, Jan. 2023, doi: 10.1016/j.respol.2022.104628.

[6] K. Schwoerer, “An exploratory study of social media’s role in facilitating public participation in e-rulemaking using computational text analysis tools,” *Policy Internet*, vol. 15, no. 2, pp. 178–203, Jun. 2023, doi: 10.1002/poi3.332.

[7] P. Chauhan, N. Sharma, and G. Sikka, “The emergence of social media data and sentiment analysis in election prediction,” *J Ambient Intell Humaniz Comput*, vol. 12, no. 2, pp. 2601–2627, Feb. 2021, doi: 10.1007/s12652-020-02423-y.

[8] C. Revanth, C. V. Yaswanth, D. B, S. S. R. K, and V. M, “Personal Career Recommendation System,” *International Research Journal on Advanced Science Hub*, vol. 5, no. Issue 05S, pp. 67–71, May 2023, doi: 10.47392/irjash.2023.s009.

[9] R. Jehad and S. A.Yousif, “Fake News Classification Using Random Forest and Decision Tree (J48),” *Al-Nahrain Journal of Science*, vol. 23, no. 4, pp. 49–55, Dec. 2020, doi: 10.22401/ANJS.23.4.09.

[10] I. D. Mienye, Y. Sun, and Z. Wang, “Prediction performance of improved decision tree-based algorithms: A review,” in *Procedia Manufacturing*, Elsevier B.V., 2019, pp. 698–703. doi: 10.1016/j.promfg.2019.06.011.

[11] M. Syafei and M. R. Darajati, “Design of General Election in Indonesia.”

[12] A. Sholichah and I. Nurchotimah, “The Forms of Political Participation in General Elections by Beginner Voters,” 2020.

[13] M. Murod, L. Andriyani, D. Gunanto, Usni, and N. A. Hamid, “The Political Role of the Survey Agency in Indonesia Democracy,” *Journal of Local Government Issues*, vol. 5, no. 2, pp. 176–192, Sep. 2022, doi: 10.22219/logos.v5i2.20232.

[14] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Social Network Analysis and Mining*, vol. 11, no. 1. Springer, Dec. 01, 2021. doi: 10.1007/s13278-021-00776-6.

[15] L. Abualigah, H. E. Alfar, M. Shehab, and A. M. A. Hussein, “Sentiment Analysis in Healthcare: A Brief Review,” in *Studies in Computational Intelligence*, vol. 874, Springer, 2020, pp. 129–141. doi: 10.1007/978-3-030-34614-0_7.

[16] N. K. Singh, D. S. Tomar, and A. K. Sangaiah, “Sentiment analysis: a review and comparative analysis over social media,” *J Ambient Intell Humaniz Comput*, vol. 11, no. 1, pp. 97–117, Jan. 2020, doi: 10.1007/s12652-018-0862-8.

[17] Y. Purnama and A. Asdlori, “The Role of Social Media in Students’ Social Perception and Interaction: Implications for Learning and Education,” *Technology and Society Perspectives (TACIT)*, vol. 1, no. 2, pp. 45–55, Oct. 2023, doi: 10.61100/tacit.v1i2.50.

[18] K. E. Anthony, S. J. Venette, A. S. Pyle, B. C. Boatwright, and C. E. Reif-Stice, “The role of social media in enhancing risk communication and The role of social media in enhancing risk communication and promoting community resilience in the midst of a disaster promoting community resilience in the midst of a disaster The role of social media in enhancing risk communication and promoting community resilience in the midst of a disaster,” 2019. [Online]. Available: https://tigerprints.clemson.edu/communication_pubs

[19] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, “Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13. Institute of Electrical and Electronics Engineers Inc., pp. 6308–6325, 2020. doi: 10.1109/JSTARS.2020.3026724.

[20] R. A. Disha and S. Waheed, “Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique,” *Cybersecurity*, vol. 5, no. 1, Dec. 2022, doi: 10.1186/s42400-021-00103-8.

[21] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[22] S. Pandya and P. Mehta, “A Review On Sentiment Analysis Methodologies, Practices And Applications,” *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, p. 2, 2020, [Online]. Available: www.ijstr.org

[23] I. H. Sarker, A. Colman, J. Han, A. I. Khan, Y. B. Abushark, and K. Salah, “BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model,” Dec. 2019, doi: 10.1007/s11036-019-01443-z.