Jurnal Teknologi Informasi dan Pendidikan
Volume 17, No. 1, March 2024
https://doi.org/10.24036/jtip.v17i1.807

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms

Sulistiawati Ahmad[1*]✉, Nur Insani[2], M Salim[3]
[1] Informatics, University Ichsan Gorontalo, Gorontalo, Indonesia
[2] Laws, University Ichsan Gorontalo, Gorontalo, Indonesia
[3] Informatic and Sains, University Ichsan Gorut, Gorontalo, Indonesia
**Corresponding Author: sulistiawatiahmad@gmail.com*

Article Information

Article history:

No. 807
Rec. October 26, 2023
Rev. January 08, 2024
Acc. January 08, 2024
Pub. January 11, 2024
Page. 75 – 86

Keywords:

Social media
Cyberbullying
Naïve Bayes
Random Forest
Support Vector Machine

ABSTRACT

Social media allows the public, especially the younger generation, to access information and knowledge or communicate with others online. Unfortunately, the phenomenon of bullying has evolved into cyberbullying, encompassing various forms of violence such as taunting, insults, intimidation, or harassment carried out by young individuals through digital technology or social media platforms. Therefore, considering the available data, there is a need for a method to classify text comments on social media, whether they fall into the category of cyberbullying or not. One of the methods used is the creation of a cyberbullying classification model using a Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes algorithms. This research aims to analyze cyberbullying in social media by comparing three different algorithms, namely Naïve Bayes, Random Forest, and SVM. The research results show that in the classification analysis, the Support Vector Machine (SVM) model performed the best, with an accuracy of 85%, precision of 79.93%, and recall of 94.29%. The Naive Bayes model also provided satisfactory results, with an accuracy of around 82.19%, precision of 81.29%, and recall of 85.10%. Meanwhile, the Random Forest (RF) model had a lower accuracy of approximately 73.15%, with a precision of 74.05% and a recall of 77.79%.

How to Cite:

Ahmad, S., Insani, N., & Salim, M. (2024). Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms. Jurnal Teknologi Informasi Dan Pendidikan, 17(1), 75-86. https://doi.org/10.24036/jtip.v17i1.807

Jurnal Teknologi Informasi dan Pendidikan
Volume 17, No. 1, March 2024
https://doi.org/10.24036/jtip.v17i1.807

1. INTRODUCTION

Cyberbullying can be carried out through SMS, text messages, applications, social media, forums, and even online games where others can participate and share content. [1] At this stage, cyberbullying typically involves sending, posting, and sharing negative, harmful, false, or malicious content towards others. This also includes sharing personal information that can lead to embarrassment or humiliation.[2]

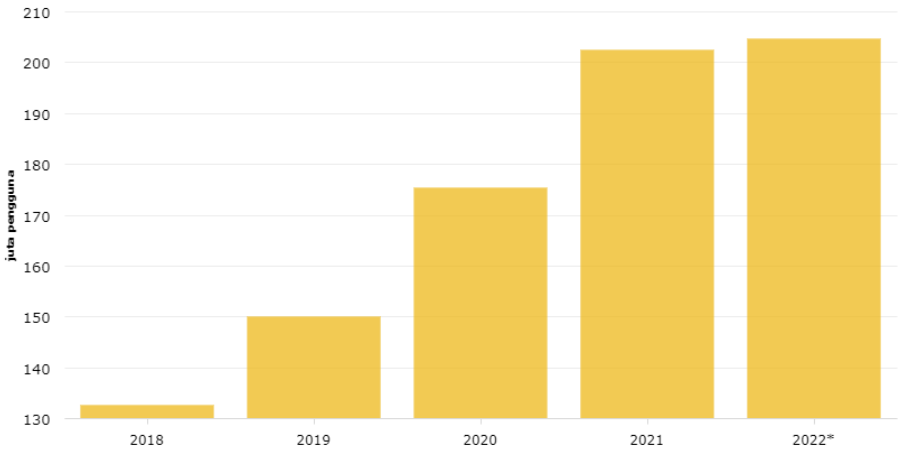

Figure 1. Number of Internet Users in Indonesia in 2018 – 2022[3]

The statistics regarding internet usage in Indonesia are quite staggering, with 212.9 million internet users at the beginning of this year, which is around 77% of the country's total population of 276.4 million. Unfortunately, cyberbullying is becoming increasingly prevalent, which is a form of repeated and continuous abusive behavior using electronic devices against a target who may find it difficult to defend themselves. [4], [5] Therefore, there is a need to classify text comments on social media as cyberbullying or not, and one of the methods used for this is creating a cyberbullying classification model using the Naive Bayes, Random Forest, and Support Vector Machine algorithms. [6]

One of the key innovations in this research is the comparison of three different algorithms, namely Naïve Bayes, Random Forest, and SVM, in conducting cyberbullying analysis on social media platforms. [7] This is an important step in identifying and classifying text comments on social media as either cyberbullying or not. [8] With the increasing prevalence of cyberbullying, this research provides valuable insights into the effectiveness of different algorithms in detecting and preventing these harmful behaviors on social media. [9] , [10] The study aims to conduct a sentiment analysis process on cyberbullying content from various social media platforms worldwide. The main objective of this analysis is to identify whether these texts contain emotional elements related to

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

cyberbullying or not.[11] Choosing to compare classification algorithms such as Support Vector Machine (SVM), Random Forest, and Naive Bayes stems from the desire to gain a comprehensive insight into their performance in various contexts. SVM, with its ability to address non-linear separation problems, Random Forest's robustness in handling complex features, and Naive Bayes' efficiency with datasets featuring independent features, represent a diversity of classification methods. Through this comparison, the aim is to assess the strengths and weaknesses of each algorithm, considering specific data characteristics, complexity levels, reliability on limited datasets, and contextual considerations in specific applications. Thus, the algorithm selection can be based on a comprehensive evaluation that takes into account the specific needs and constraints of the classification task at hand.

2. RESEARCH METHOD

2.1 Stages of the Research Process

This diagram will explain these processes in more detail:

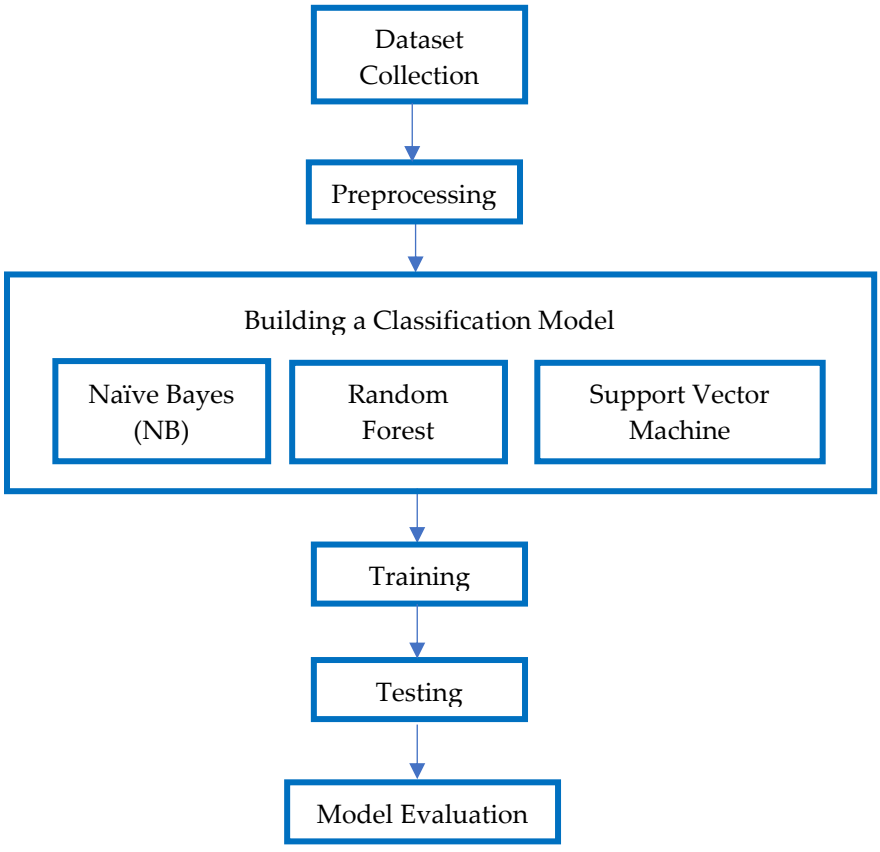

Figure 2. Stages of the Research Process

2.2 Data Collection

This stage involves classification algorithms built according to the initial research goals. Testing the dataset taken from the Indonesian Language Cyberbullying Kaggle data to classify bullying and non-bullying classes in this study is carried out using three classification algorithms, namely Random Forest, Naïve Bayes, and SVM, implemented with RapidMiner version 10.1.001.

The research method involved visiting the Instagram profiles of Indonesian celebrities, influencers, and public figures who have more than 500,000 followers, with posts made between August 2019 and April 2022. Photos and videos were selected for research purposes, and comments were copied and organized in Microsoft Excel. Manual labeling was carried out on the dataset, where comments labeled as bullying were assigned class 0, while comments labeled as non-bullying were assigned class 1. The dataset was then saved in .csv (Comma Separated Values) format.[12] , [13]

Table 1. Sample Comment Dataset

No	Comment	Class
1	"Kaka Tidur yaa, Udah pagi, gaboleh capek2"	1
2	" Makan Nasi padang aja begini badannya"	1
3	" Yang aku suka dari dia adalah selalu cukur jembut sebelum man..	0
4	" Hai Kak Isyana aku ngefans banget sama kak Isyana. Aku paling…	1

Table 2. Examples of Bullying sentences in the dataset

No	Comment
1	"@Ay.Kinantii Isyan skrg berubah ya☹ baju nya nakal"
2	" Makin jelek aja anaknya, padahal ibu ayahnya cakep2"
3	" Kok anaknya kayak udah tua gitu ya mukanya kk tasya"
4	" Muka anak nya ko tua banget yaa.. GK ngegemein GK ada lucunya"

The stages to be undertaken include reading the collected dataset that has been saved in .csv format. Subsequently, the dataset will go through the Text Preprocessing stage to clean the data so that it can be used in the next stage. [7] Furthermore, there is a FastText Embedding stage aimed at converting each word in the data into a vector form. The dataset will undergo the data splitting stage, where the data will be divided into training and testing data with a 70:30 ratio. The processed data will then go through a classification stage using three algorithms, and the results will be evaluated using a confusion matrix.[8] The final stage of research is comment detection.

2.3 Text Preprocessing

Before being processed and classified using a machine learning model, a preprocessing stage will be conducted to prepare the data for effective and efficient

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

processing by the machine learning model. [14] Through proper preprocessing, high-quality data will be generated, allowing the machine-learning model to produce more accurate and optimal predictions. [13], [15] In this study, the data will undergo four stages of text preprocessing, such as cleaning and converting to lowercase, tokenization and normalization, removal of common words, and word stemming. [16], Cleaning aims to purify the data by removing specific characters or symbols that may exist in the data to reduce noise [17] This involves eliminating URLs, non-ASCII characters, numbers, symbols, punctuation, mentions, hashtags, and converting uppercase letters to lowercase (case folding). An example of cleaning and converting to lowercase can be seen in Table 3 below.

Table 3. Cleaning and Case Folding

Before	Afther
"Kaka Tidur yaa, Udah pagi, gaboleh capek2"	"kaka tidur yaa udah pagi gaboleh capek2"
" Makan Nasi padang aja begini badannya"	" makan nasi padang aja begini badannya"

Next, the tokenization and normalization process is carried out, which aims to separate each word that forms a sentence into token pieces. [18] and normalize each word, including abbreviations or typos, so that they will be corrected to be more structured and can be further processed based on the prepared dictionary.

Table 4. Tokenization and Normalization

Before	After
"kaka tidur yaa udah pagi gaboleh capek2"	"kakak, tidur,ya, sudah, pagi, tidak, boleh lelah"
" makan nasi padang aja begini badannya"	" makan, nasi, padang, saja, seperti ini, badannya"

The subsequent preprocessing phase involves the removal of stopwords, to eradicate insignificant words that do not contribute to the classification process. [13]

	A
1	rt
2	ada
3	adalah
4	adanya
5	adapun
6	agak
7	agaknya
8	agar
9	akan
10	akankah
11	akhir
12	akhiri
13	akhirnya
14	aku
15	akulah
16	amat

Figure 3. Result Stopwords

The last preprocessing step, stemming, is intended to find the base form of words by stripping away their prefixes and suffixes. These removed affixes include prefixes such as "me," "ter," "ke," "ber," "di," and suffixes like "kan," "nya," "-i," and others.

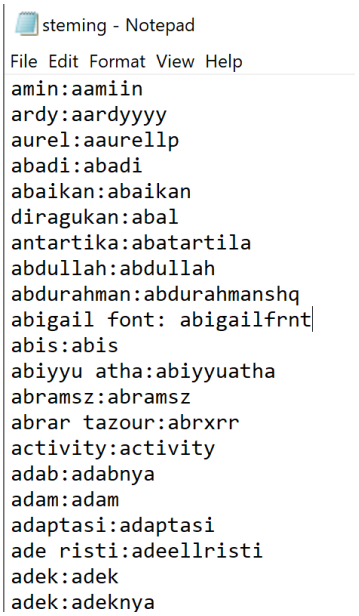

Figure 4. Result Steaming

3. RESULTS AND DISCUSSION

After the classification stage using the NB, RF, and SVM algorithms with parameter tuning, the best combination of hyperparameter values used in the three algorithm models will be determined. [19] Random Forest, Naïve Bayes, and SVM, implemented with the following tools:

.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

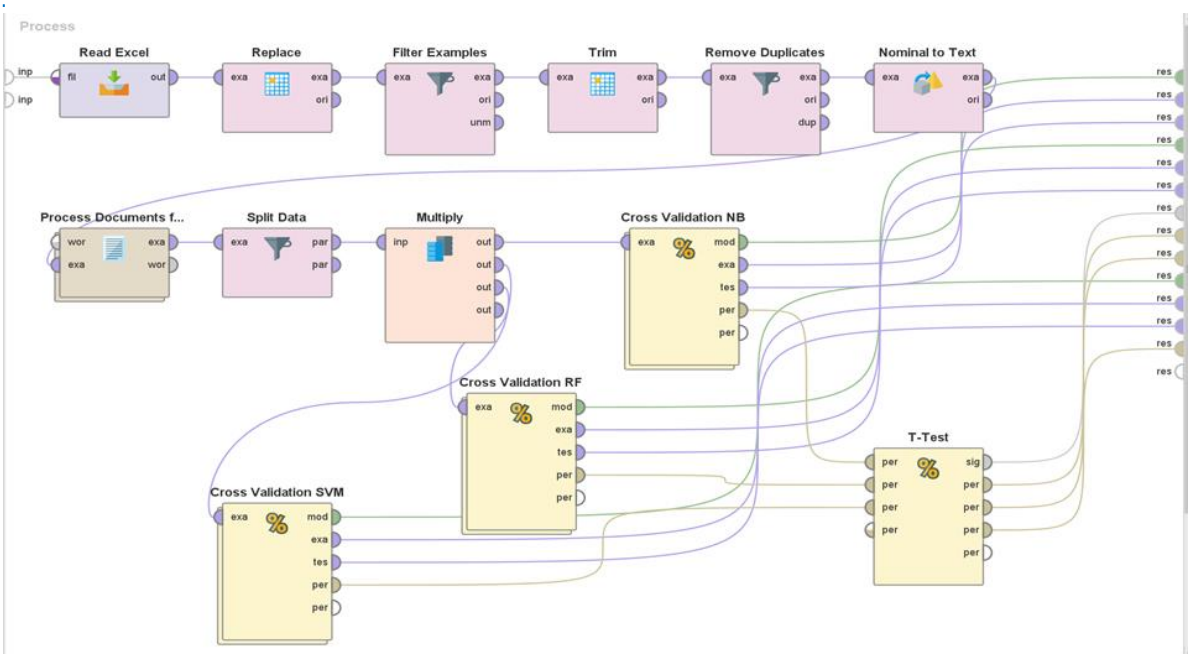

Figure 5. Random Forest Model Design, Naïve Bayes, and Support Vector Machine

4.1. Naïve Bayes, Random Forest dan Support Vector Machine

The research results demonstrate that three classification algorithms, namely Support Vector Machine (SVM), Naive Bayes, and Random Forest (RF), were evaluated in the classification analysis. In the tests, the Support Vector Machine (SVM) model exhibited the best performance with an accuracy rate of 85%, precision of around 79.93%, and recall of approximately 94.29%. This indicates that the SVM model possesses the ability to classify with a high degree of accuracy and effectively recognizes a significant portion of true positives (high recall).

Furthermore, the Naive Bayes model also yielded satisfactory results, with an accuracy rate of approximately 82.19%, precision of about 81.29%, and recall of roughly 85.10%. This suggests that the Naive Bayes model also performs well in classifying data, although slightly below the performance of the SVM model. On the other hand, the Random Forest (RF) model exhibited lower accuracy, around 73.15%, with a precision of approximately 74.05%, and recall of roughly 77.79%. This indicates that the RF model has lower performance in terms of accuracy, precision, and recall compared to the SVM and Naive Bayes models.

These research findings provide valuable insights into the performance of the three algorithms in classifying cyberbullying-related data. The SVM model stands out with excellent performance, while the Naive Bayes model delivers satisfactory results. Although

the RF model is still usable, it exhibits lower performance compared to the other two models. As a result, the choice of classification algorithm significantly impacts the results of cyberbullying analysis, and selecting SVM or Naive Bayes may be a better choice in this case.

AUC is useful because it provides a comprehensive picture of a model's performance without being limited to a specific threshold. In other words, AUC is not affected by a specific threshold, which can vary depending on the needs of the application. [20]

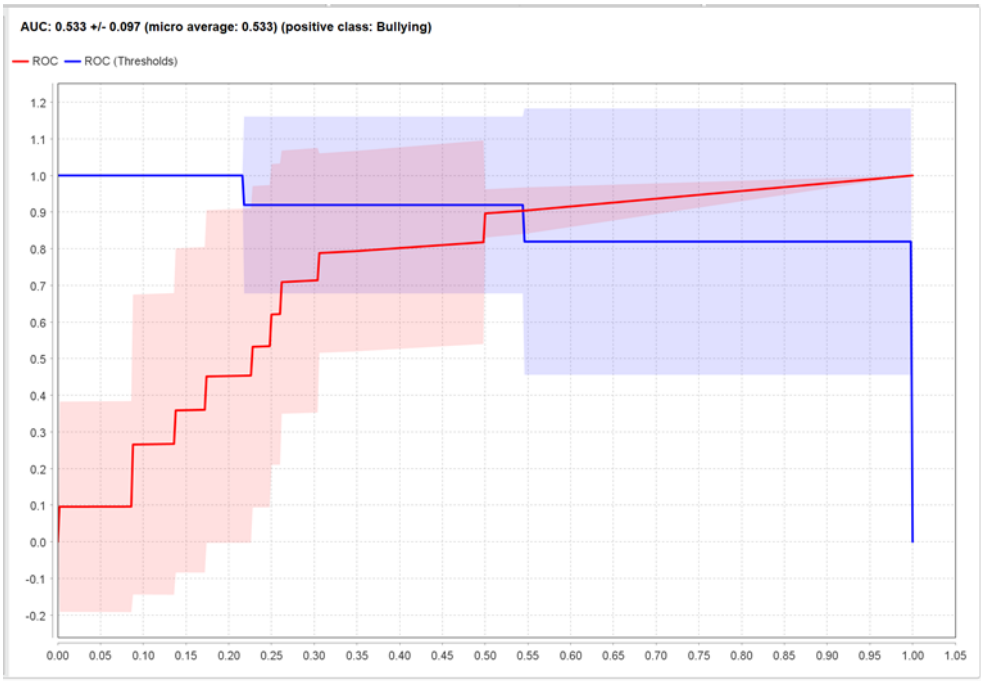

Figure 6. AUC Naïve Bayes

Figure 6 shows the presence of the Area Under the Curve (AUC) values on the Receiver Operating Characteristic (ROC) curve for Naive Bayes, specifically for the positive class of Bullying, providing an overview of how well the model can distinguish between Bullying and non-Bullying cases. If the AUC value approaches 1, it indicates that the model performs well in identifying instances of Bullying without generating too many false positives. The higher the AUC value, the better the model's ability to differentiate between these classes. Conversely, if the AUC value approaches 0.5, it signifies a model performance equivalent to random guessing. The interpretation of AUC values always depends on the application context and specific priorities related to sensitivity and specificity in correctly identifying cases of Bullying and avoiding false positives.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

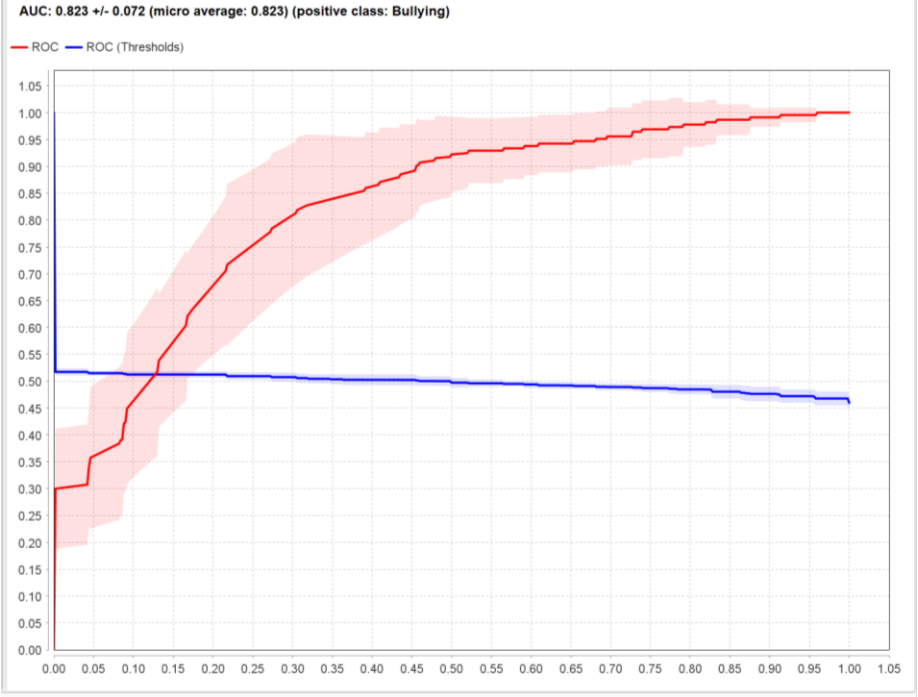

Figure 7. AUC Randon Forest

In Figure 7, the AUC (Area Under the Curve) value for Random Forest is 0.823, indicating the extent to which the Random Forest model can distinguish between positive and negative classes on the Receiver Operating Characteristic (ROC) curve. The higher the AUC value, the better the model's ability to differentiate between these classes. In this context, the value of 0.823 indicates good performance, approaching the maximum value of 1. A Random Forest model with this AUC is likely to have a good level of sensitivity in identifying positive cases and a high level of specificity in avoiding false positives. However, further interpretation may depend on the specific context of the application and the specific goals of the analysis.

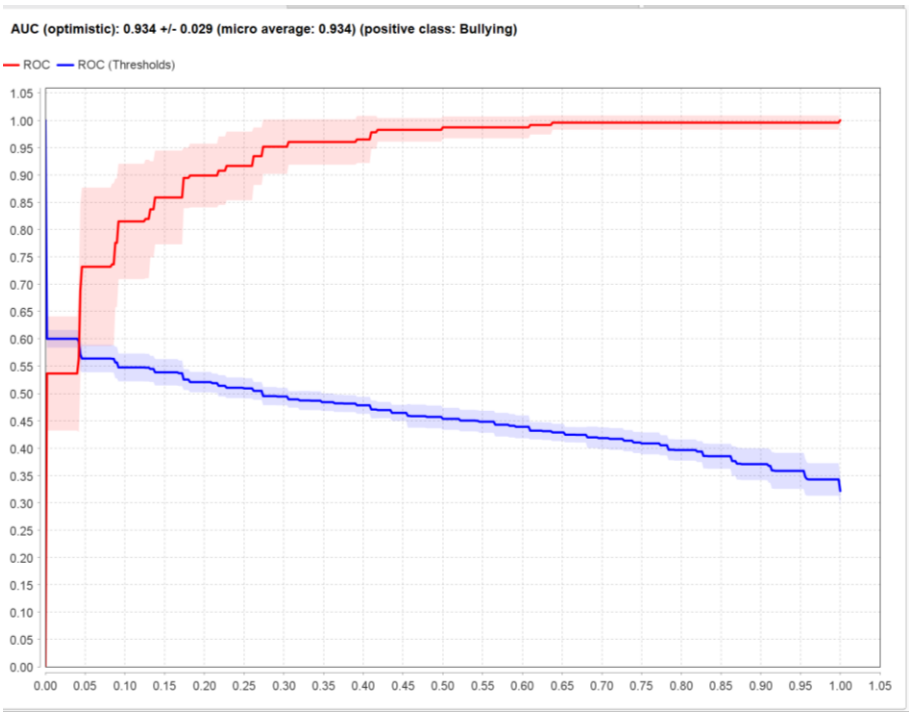

Figure 8. AUC SVM

Figure 8. The AUC (Area Under the Curve) score of 0.934 for the Support Vector Machine (SVM) indicates an excellent level of performance in distinguishing between positive and negative classes on the Receiver Operating Characteristic (ROC) curve. The higher the AUC score, the better the model's ability to differentiate between these classes. In this context, a score of 0.934 suggests that SVM has a high level of sensitivity in identifying positive cases and a high level of specificity in avoiding false positives. AUC approaching 1 indicates that SVM has a highly discriminative ability. The interpretation of the AUC score provides confidence in SVM's ability to distinguish between these classes, but it's important to note that this interpretation also depends on the specific context of the application and analysis goals.

4. CONCLUSION

This research has successfully made an analysis and comparison of three algorithms that classify Bullying and Not Bullying. The results showed that in the classification analysis, the Support Vector Machine (SVM) model had the best performance, with an accuracy of 85%, precision of 79.93%, and recall of 94.29%. The Naive Bayes model also gave satisfactory results, with an accuracy of about 82.19%, a precision of 81.29%, and a recall of 85.10%. Meanwhile, the Random Forest (RF) model has a lower accuracy, around 73.15%, with a precision of 74.05%, and recall of 77.79%. Therefore, for applications that require precise identification, SVM models may be the preferred choice, while Naive Bayes models

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

are also reliable. The Random Forest model, while still delivering good results, requires further performance improvements depending on the type of application used.

ACKNOWLEDGEMENTS

The author would like to thank various parties who have provided valuable support and contributions to this research. Thank you to family, peers, lecturers, institutions, respondents, and all who have supported this research. Contributions from various individuals and institutions have played an important role in the completion of this journal, as well as to Jurnal Teknologi Informasi dan Pendidikan which has allowed the Author.

REFERENCES

[1] H. Y. Al-aziz and S. Monalisa, "Comparison of Facebook and Instagram to Assess the Efectiveness of Advertising Channels in Customer Acquisition," *J. Teknol. Inf. dan Pendidik.*, vol. 15, no. 2, pp. 64–72, 2023, doi: 10.24036/jtip.v15i2.677.

[2] C. Destitus, W. Wella, and S. Suryasari, "Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 11, no. 2, pp. 107–111, 2020, doi: 10.31937/si.v11i2.1740.

[3] L. Septiani, "Jumlah Pengguna Internet di Indonesia 212,9 Juta - Teknologi Katadata.co.id," *Katadata.co.id*, Feb. 22, 2023. https://katadata.co.id/desysetyowati/digital/63f5d758a2919/jumlah-pengguna-internet-di-indonesia-212-9-juta (accessed Apr. 12, 2023).

[4] I. Saputra and D. Rosiyadi, "Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp," *Fakt. Exacta*, vol. 12, no. 2, p. 101, 2019, doi: 10.30998/faktorexacta.v12i2.4181.

[5] T. N. Lam, D. B. Jensen, J. D. Hovey, and M. E. Roley-roberts, "Heliyon College students and cyberbullying : how social media use affects social anxiety and social comparison," *Heliyon*, vol. 8, no. May 2022, p. e12556, 2023, doi: 10.1016/j.heliyon.2022.e12556.

[6] M. F. Naufal, T. Arifin, and H. Wirjawan, "Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial," *Jurasik (Jurnal Ris. Sist. …*, vol. 8, pp. 82–90, 2023, [Online]. Available: http://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/view/544%0Ahttp://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/download/544/522.

[7] A. Tariq *et al.*, "Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data," *Heliyon*, vol. 9, no. 2, p. e13212, 2023, doi: 10.1016/j.heliyon.2023.e13212.

[8] T. J. Melmambessy, "Analysis of the Opinion Students about The Online Learning System During the Pandemic Using The K-NN and Naïve Bayes Methods," *J. Teknol. Inf. dan Pendidik.*, vol. 16, no. 1, pp. 75–85, 2023, doi: 10.24036/jtip.v16i1.702.

[9] P. Yi and A. Zubiaga, "Session-based cyberbullying detection in social media: A survey," *Online Soc. Networks Media*, vol. 36, no. May, 2023, doi: 10.1016/j.osnem.2023.100250.

[10] M. Nabil, N. Pribadi, N. Chamidah, K. K. Cyberbullying, and R. Forest, "Klasifikasi Tweet Cyberbullying dengan Menggunakan Algoritma Random Forest," pp. 512–520, 2022.

[11] M. Karim, "Analisis Sentimen Pada Twitter Menggunakan Support Vector Machine Dengan Modifikasi Lexicon Inset Dan Sentis- Trength _ Id (Studi Kasus : Vaksinasi Covid-19)," 2022.
[12] W. W, "Comparison Of Clustering Levels Of The Learning Burnout Of Students Using The Fuzzy C-Means And K-Means Methods," *J. Teknol. Inf. dan Pendidik.*, vol. 16, no. 1, pp. 38–53, 2023, doi: 10.24036/jtip.v16i1.668.
[13] A. S. Syahab, "Comparison of Machine Learning Algorithms for Classification of Ultraviolet Index," vol. 15, no. 2, 2023.
[14] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," *Sustain. Oper. Comput.*, vol. 3, no. February, pp. 238–248, 2022, doi: 10.1016/j.susoc.2022.03.001.
[15] N. Lashkarashvili and M. Tsintsadze, "Toxicity detection in online Georgian discussions," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 1, p. 100062, 2022, doi: 10.1016/j.jjimei.2022.100062.
[16] S. R. Ahmad, "Website-Based E-Market (E-Patali) Application For Gorontalo City Central Market," *J. Teknol. Inf. dan Pendidik.*, vol. 16, no. 1, pp. 64–74, 2023, doi: 10.24036/jtip.v16i1.350.
[17] H. Santoso and R. A. Putri, "Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest Cyberbullying Comment Detection on Instagram Social Media Using Random Forest Algorithm," vol. 13, no. April, pp. 62–72, 2023.
[18] M. N. Huda, D. A. Fauzan, M. Raihan, S. Putra, and N. Sabila, "Jurnal KomtekInfo Optimalisasi Model Klasifikasi Sentimen Netizen Terhadap Merek," *J. KOmtekInfo*, vol. 10, no. 1, pp. 21–27, 2023, doi: 10.35134/komtekinfo.v10i1.360.
[19] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, and A. Hussain, "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms," *Healthc. Anal.*, vol. 3, no. April, p. 100185, 2023, doi: 10.1016/j.health.2023.100185.
[20] N. Nafi'iyah, "Svm Algorithm for Predicting Rice Yields," *J. Teknol. Inf. dan Pendidik.*, vol. 13, no. 2, pp. 50–54, 2020, doi: 10.24036/tip.v13i2.341.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id