Toxicity and Sentiment Analysis About Digital Bounty on Social Media

1st Fazjar Sekti Aji
Faculty of Information Technology, Satya Wacana Christian University
Salatiga, Indonesia
972020020@student.uksw.edu

2nd Ade Iriani
Faculty of Information Technology, Satya Wacana Christian University
Salatiga, Indonesia
ade.iriani@uksw.edu

3rd Hendry
Faculty of Information Technology, Satya Wacana Christian University
Salatiga, Indonesia
hendry@uksw.edu

2023 12th International Conference on Awareness Science and Technology (iCAST) | 979-8-3503-2469-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/iCAST57874.2023.10359318

***Abstract*—*The presence of social media has a positive impact but social media is also a means for negative things to happen, such as cyberbullying, data theft, and sexual crimes. Digital bounty or competition to find perpetrators digitally is one of the activities of cyberbullying and digital vigilantism (doxing). Digital bounties have negative impacts such as psychological and socio-economic impacts on perpetrators (bounty targets) and encourage reduced public trust in applicable legal institutions. The number of opinions on social media requires sentiment classification and toxicity measurements. To get the best results, a classification method is searched for the appropriate dataset by comparison, using k-fold validation (KNN, NBC, SVM, DT), in the classification process using the selected method. Give positive opinion results, (supporting digital bounty posts of 1540 and negative of 439, and in toxicity, analysis using Communalytic, most comments in the dataset scored 0 – 0.4*.

***Keywords*— sentiment analysis, toxicity analysis, doxing.**

I. Introduction

In this modern era, the world is being hit by developments in technology. Developments have brought changes in all layers of life. One of the consequences of the development of technology is the internet. Along with the times, the internet is growing rapidly, and one of its benefits is as a medium that connects people around the world, better known as social media [1]. Social media as the software allows various individuals to communicate, share, gather, and play. Social media itself in Indonesia is one of the biggest destinations for internet users in Indonesia, social media that are often used in Indonesia are Facebook, Instagram, Twitter, etc. [2]. The presence of social media has a positive impact, but social media is also a means for negative things to happen, such as cyberbullying, data theft, and sexual crimes.

Vigilantism or Persecution is a situation when people take on the role of law enforcement without being given legal authority, against a person or several citizens, and without considering whether their actions are based on justice or not. Digital Vigilantism is a form of Vigilantism or Persecution along with the development of the internet and the formation of social media, for example, in the form of naming, shaming, or doxing [3]. When vigilantism is expressed through digital media, where the perpetrator seeks and circulates the target's personal information, this action aims to share information about the details of the target's fault or violation and penalize the kind of visibility that is intentionally imposed on the target. Digital Vigilantism is a global phenomenon, such as in North America, where online responses to the 2011 Vancouver riots and the 2013 Boston bombing hunt severely affected the lives of those who were (wrongly) targeted [4].

Digital bounty or competition to find perpetrators digitally is one form of cyberbullying and digital vigilantism(doxing) activities. Digital bounty activities during the COVID-19 pandemic increased along with the increase in crime rates and internet users [5], [6]. Digital bounty has negative impacts such as psychological and socio-economic impacts on perpetrators (target bounties) and encourages reducing public trust in legal institutions. "(1) Unless otherwise stated by laws and regulations, the use of any information through electronic media concerning a person's data must be carried out with the consent of the person concerned. (2) Every person whose rights are violated as referred to in paragraph (1) can file a lawsuit for the losses incurred", adhering to Article 26, paragraphs 1 and 2, Law No. 11 of 2008 [7], [8]. Toxicity is also one of the activities of cyberbullying and digital vigilantism (naming & shaming). Toxicity is defined as "rude, disrespectful, or unreasonable comments that might lead you to leave the discussion". Comments containing toxicity are common in various social media posts, including Digital bounty.

Facebook and Twitter are social media that are often used by Indonesians, with users reaching 65.8% and 20.6% of the total internet users in Indonesia in 2019-2020 based on the APJII survey [2]. Both social media are often misused as a means of digital bounty. Digital bounties on Facebook are often found in regional information groups and regional buying and selling groups, Digital bounties on Twitter can be found in popular tab searches with certain keywords. Sentiment analysis provides useful indicators for various purposes, which can be found in the comments.

Several studies on sentiment analysis in several cases have been carried out. One of them is research on sentiment analysis with a comparison of classification algorithms which is also carried out using the SVM and NBC methods. This study uses 9637 data from the 2018 West Java governor election tweet. The results show that the Support Vector Machine Algorithm produces an average accuracy of 92.61% with an AUC of 0.950, and the Naïve Bayes algorithm produces an average accuracy of 93.29% with an AUC of 0.525 [9]. In addition to the governor election, research on sentiment analysis using NBC has previously been carried out to classify various opinions from social media such as Jakarta government's lockdown policy [10], Hate Speech on Online News Portals [11], Infrastructure Development [12], car market sentiment [13], and film opinion [14].

Other studies also conducted sentiment analysis by comparing the Naive Bayes Method classification algorithms, Decision Tree, and SVM regarding public opinion on the COVID-19 booster vaccination. This study uses 3000 tweet data. The results show that the Support Vector Machine Algorithm produces an average accuracy of 79%, a precision of 79.97%, and an AUC of 75.40%, the Naïve Bayes algorithm produces an average accuracy of 70%, a precision of 83.81%, and AUC 71, 06%, the Decision Tree Algorithm produces an average accuracy of 83.33%, precision 81.06%, and AUC 74.15% [15].

Related research in toxicity analysis on tweeters about comments or sentiments regarding the use of face masks during the COVID-19 pandemic. The study compared the results of the Toxicity between tweets with hashtags pro-mask and anti-mask. Research suggests toxicity as a form of verbal aggression and as an expression of anger creates an additional and powerful barrier for individuals to be exposed to on social media and is as dangerous as misinformation and disinformation. Specifically, the toxicity surrounding the wearing of masks: indicates a public outcry, which in turn threatens the success of risk communication efforts (creating a calm society, and understanding what they must do for their immediate environment) around an issue, alienating individuals from the discussion, leading to increased public disengagement with the matter; raises doubts, effectively eroding trust in Twitter's authorities and public health units [16]. Several previous studies on toxicity have also been carried out in various media such as YouTube [17], Reddit [18], and Facebook [19].

Several previous studies can prove that to find out public opinion on a problem that occurs and becomes a topic of conversation on social media, sentiment analysis can be done using machine learning. Based on this and the background of the problem, this study aims to understand public sentiment about Digital bounty classification and toxicity analysis methods using classification and analyze the relationship between the value of toxicity with sentiments that support the existence of Digital bounty.

II. Research Method

This research uses a quantitative approach. This study aims to determine public sentiment about Digital bounty using classification and toxicity analysis. The RapidMiner application performs pre-processing, training, testing, and classification algorithm comparison.

A. Data Collection(Scraping)

Scraping is data collection on a website using methods and tools to be saved to a database [20]. Scraping or data collection in this study consists of 2 parts, namely the Twitter section and the Facebook section, sentiment scraping on Digital bounty using an application called Communalytic with the Twitter API, Digital Bounty Sentiment Data on Facebook is only taken from the Regional Facebook Group which is public and the process scraping is done manually, This is because after the CA scandal exploded [21], Facebook decided to close the API service (the only legitimate tool that allows third parties to access and download Facebook user information), although there are other ways such as screen scraping that are ethically acceptable but may still expose researchers to legal risk due to TOS (Term Of Service) violations, so that Facebook's publishable datasets must not contain personal information about users, and must delete all potentially identifying information that is not automatically collected [22][23].

B. Preprocessing

Preprocessing Data is the extraction and cleaning of data that will be used at the Data Data Training [24], Preprocessing data is the most important part of sentiment analysis which covers more than 50% of the sentiment analysis process [25]. Before entering the preprocessing stage, Facebook sentiment data is translated from regional languages to Indonesian before being combined into one with Twitter sentiment data, the preprocessing data is carried out using RapidMiner or manual (labelling), stemming process using API py Sastrawi and stop word process using the stop word dictionary from NLTK Corpus. There are several parts of Preprocessing as follows:

- Normalization Features: username, URL, special characters, punctuation marks, and other characters will be replaced with spaces, then remove excess spaces
- Transform Case: words containing uppercase letters will all be converted into letters in small or lowercase.
- Tokenize: used to divide the text in the document into sequence Token for term weighting (TD-IDF)
- Stemming: used to make suffix words into essential words
- Filter Token (By Length): Words shorter than three letters and longer than 25 letters will be removed
- Stopword Removal: Removes irrelevant words or conjunctions.

C. Comparison

Sentiment analysis is a method for obtaining information from various platforms. The purpose of sentiment analysis is to process and understand textual data automatically to obtain sentiment information contained in an opinion sentence, whether it is a positive or negative opinion [26]. Sentiment analysis analyzes old cases to find patterns from the data using the method, following method chosen for comparison in this study

- Naive Bayes Classifier (NBC), NBC is a Supervised Learning that uses probability and statistics that predict opportunities based on previous experience (training). [27][13][14][28].
- Support Vector Machine (SVM), SVM method is a relatively new method for making predictions on regression or classification cases. SVM classification tries to separate the data space using nonlinear or linear classification between different classes. The concept of SVM classification is as a hyperplane that acts as a separator of two data classes [29].
- Decision Tree (DT), A decision tree is a tree-like collection of nodes intended to make decisions about the affiliation of values to classes or estimates of numerical target values. Each node represents a split rule for one particular Attribute. For classification, this rule separates the values belonging to different classes, for regression it separates them to reduce the error optimally for the selected parameter criteria [15], [30], [31].
- K-Nearest Neighbors (KNN), The K-Nearest Neighbor method groups data into predetermined classes based on the level of similarity between the data and the testing data. Later the data will be grouped into a class by looking at many "k" values for the closest distance to the testing data [32].

The training process is carried out to build a model from training and testing data the classification model is generated from the training process using new data or what is called data testing. Comparison in the selection of classification algorithms is carried out with three conditions: accuracy, precision, and AUC. A comparison is made with K-fold cross-validation, K-fold cross-validation is a method used to assess the performance of a predictive model by dividing the data into sections, then training in one section and then testing in another in an iterative approach to ensure all data can be used. Used for testing [33]. After a comparison is made for the selection of a classification algorithm, a classification algorithm is implemented on the dataset to determine the sentiment in the dataset.

D. Toxicity Analysis

Toxicity is defined as "rude, disrespectful, or unreasonable comments that might lead you to leave the discussion." Toxicity measurement in this study will use Communalytic tools with the help of the Perspective API, the Perspective API is an API developed by Jigsaw (a team from Google), Perspective is trained to recognize various attributes (whether comments are toxic, threatening, insulting, off-topic, etc.) using millions of samples collected from several online platforms and reviewed by human annotators [34]. Toxicity analysis in this study is used to measure the toxicity score of identity attacks, insults, and threats. Communalytic is a web-based research tool that can collect and analyze publicly available data [35]. Communalytic's Toxicity analysis on a scale of 0 – 1 (the closer to 1 the higher the toxicity), as well as scores for a certain message, attributes such as assault, insults, and profanity, which are set using the Perspective API [36].

III. Result And Discussion

The following is an explanation of the research conducted, in which this study carried out two different analyses. The first analysis used was sentiment analysis, while the second was toxicity analysis.

A. Sentiment Analysis

Facebook and Twitter data collection regarding Digital Bounty With the target of cases in 2021-2022, 2043 comments/tweets (889 Facebook & 1154 Twitter) were collected in Indonesian. Digital bounty is a very open social phenomenon. This is evidenced by several cases that have been scrapped from social media (Facebook & Twitter), totalling 2043 cases; although some people have realized the negative impact of Digital bounties, such as making regulations on groups or communities on Facebook, there are still many people who use social media as a means of vigilantism. Social media companies (Facebook & Twitter) also do not support the existence of a Digital bounty, and this can be seen in Digital bounty, which disappears within 2x24-7x24 hours. There are some limitations in this study. As we have explained, where Digital bounty is easily lost or deleted, and we don't have access to many deleted Twitter and Facebook posts, making the data available is not much so the cause of Digital bounty is very varied, making direct and fair comparison impossible in many cases. These limitations highlight the need for cross-platform, privacy-sensitive protocols for sharing data with researchers [37].

Facebook and Twitter data collection regarding Digital Bounty With the target of cases in 2021-2022, 2043 comments/tweets (889 Facebook & 1154 Twitter) were collected in Indonesian. Digital bounty is a very open social phenomenon. This is evidenced by several cases that have been scrapped from social media (Facebook & Twitter), totalling 2043 cases; although some people have recognized the negative impact of digital bounties, such as imposing regulations on Facebook groups or communities, there are still many people who use social media as a means of vigilantism. Social media companies (Facebook & Twitter) also do not support the existence of a Digital bounty, and this can be seen in Digital bounty, which disappears within 2x24 - 7x24 hours.

There are some limitations in this study. As we have explained, where Digital bounty is easily lost or deleted, and we don't have access to many deleted Twitter and Facebook posts, making the data available is not much so the cause of Digital bounty is very varied, making direct and fair comparison impossible in many cases. These limitations highlight the need for cross-platform, privacy-sensitive protocols for sharing data with researchers [37].In this study, Rapidminer was used to perform pre-processing (Feature Normalization, Transform case, Tokenizing), Stemming (API Literature), and Stopword (NLTP). This study looks for a classification method that is suitable for the dataset by comparing the classification method (NBC, DT, K-NN, and SVM) with k-fold cross-validation.

The dataset is divided into two, namely training data (labelling) and testing data, in this study the data is divided in a ratio of 25%:75%. Data labelling in this study refers to indications of doxing, shaming, and efforts to digital viral bounty. After going through pre-processing and labelling, it is followed by a comparison of classification algorithms according to the dataset. Training data is compared using K-fold and SMOTE implementation (as a class balancer). SVM produces an accuracy of 75.01%, and an AUC of 0.909, DT produces an accuracy of 72.09%, and an AUC of 0.626, K-NN produces an accuracy of 71.86%, and an AUC of 0.831, NBC produces an accuracy of 82.77% and AUC of 0.556. The greater the AUC value (closer to 1), the better, and an AUC with a score of 0.5 and below is almost the same as the model that performs random guesses, to see whether AUC scores good or bad, it can be classified as follows:

- 0.5 = No discrimination
- 0.5-0.7 = Poor discrimination
- 0.7- 0.8 = Acceptable discrimination
- 0.8-0.9 = Excellent discrimination
- >0.9 = Outstanding discrimination

This standard follows the rules of Hosmer and Lemeshow [38]. The results of the classification algorithm performance can be seen in Tables 2 to 5.

Table 1. The results of k-fold SVM

75.01%/0.909	true negative	true positive	class precision
pred. negative	115	12	90.55%
pred. positive	91	194	68.07%
class recall	55.83%	94.17%	

Table 2. Result of k-fold DT

72.09%/0.626	true negative	true positive	class precision
pred. negative	157	66	70.40%
pred. positive	49	140	74.07%
class recall	76.21%	67.96%	

Table 3. K-NN k-fold results

71.86%/0.831	true negative	true positive	class precision
pred. negative	182	92	66.42%
pred. positive	24	114	82.61%
class recall	88.35%	55.34%	

Table 4. NBC k-fold results

82.77%/0.556	true negative	true positive	class precision
pred. negative	162	27	85.71%
pred. positive	44	179	80.27%
class recall	78.64%	86.89%	

Sentiment analysis Digital bounty can be done using the NBC, DT, K-NN, and SVM classification methods. The comparison of classification methods results shows that SVM is very compatible with the processed dataset with test results using 10 k-fold cross-validations (accuracy and AUC). SVM was also chosen because of its ability to minimize errors in the training data and is influenced by dimensions (Risk Minimization Structure) which can be realized by selecting the hyperplane with the largest margin [39]. The parameter value selected by the SVM operator is the standard value given by Rapidminer. Changes were made to the penalty factor parameter C, which is the misclassification tolerance value of SVM. In this study, the input value for C is 1. This number provides much better accuracy than the standard value given by the system (0). After the comparison is complete, proceed with testing data classification using SVM. The classification process with SVM produces 1540 positive sentiments (sentiments in favor of Digital bounty) and 439 negative sentiments (not supporting Digital bounty).

The parameter value selected by the SVM operator is the standard value given by Rapidminer. Changes were made to the penalty factor parameter C, which is the misclassification tolerance value of SVM. In this study, the input value for C is 1. This number provides much better accuracy than the standard value given by the system (0). After the comparison is complete, proceed with testing data classification using SVM. The classification results using SVM show that the supporting opinion is more dominant, at 78%, and the opposing opinion, at 22%. The classification process with SVM produces 1540 positive sentiments (sentiment in favor of Digital bounty) and 439 negatives. After processing the data, it produces accuracy and AUC in the use of SVM. the accuracy produced by SVM is quite high if compared to another study whose accuracy produced below 80% [39], the figure shows an accuracy of 96.11% because there are 16 datasets with new labels that are different from the previous dataset labels (training data.) produced below, however, in the k-fold SVM test result (75%), it is still very low when compared to other studies with the results of the k-fold testing of other studies reaching 95% [29].

This study uses a word cloud to see words that often appear in the dataset. A Word cloud is a visual representation of text data and provides clarity in identifying trends and patterns [40]. The following is a word cloud generated from a dataset of social media user sentiment on the digital bounty.

- Positive

  Orang, up, ketemu, uang, anjing were the most dominant words.

- Negative

  Cek, palsu, polisi, langsung, tanda were the most dominant words.

B. *Toxicity Analysis*

Sentiment analysis Digital bounty can be done using the NBC, DT, K-NN, and SVM classification methods. The comparison of classification methods results shows that SVM is very compatible After knowing the sentiment data, followed by toxicity analysis using Communalytic, from 1979 the data that went through preprocessing (Normalization of features), 1600 can be done toxicity analysis. Table 5 shows examples of comments with toxicity scores, and Figure 4 shows the percentage of toxicity scores obtained from Communalytic.

Table 5. Examples of comments and toxicity scores

No	Score	Comments
1	0.003748	Dilacak saja lewat bantuan polisi nomer telefon kan masih aktif itu
2	0.141940	laporin polisi aja udah biar ga tuman jangan kasih ampun
3	0.224223	Maaf kak itu teman nya kenal sama dia darimana Sudah pernah ketemu apa belum Kalo belum kemungkinan dia ada di dalam jeruji besi nyari uang dengan cara menipu kaya begitu

No	Score	Comments
4	0.388749	wkwkw iya mana di indonesia hukum yg mengatasi kaya ginian tuh lemah banget ujung2nya disuruh ikhlasin gimana ga makin merajalela penipu penipu model begini gampang bener jadi penipu di indo modal ga punya malu aja ?? ga bakal masuk penjara juga
5	0.450452	Serang saja jadi biarkan dia mengalami gangguan mental
6	0.583504	Idih lu klepto dan tukang bokis ya Makanya hidup jangan kemakan gengsi Betina giling
7	0.67902285	Mesti bocil malu maluin jateng bangsat bangsat
8	0.785242	iya dasar muka badak sudah mukanya jelek suka mencuri pula sudah komplit tinggal di buang saja
9	0.8546526	kak kak cantik cantik kriminal jadi pelacur aja kak
10	0.935843	seharusnya pukuli saja sampai mati

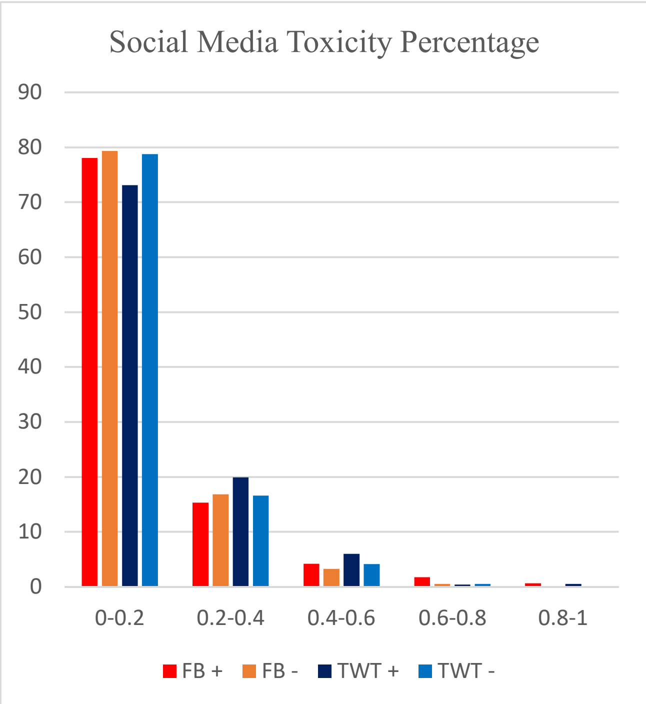

Figure 4 *FB&TWT Toxicity analysis*

The results of the toxicity analysis show that most of the toxicity of the majority dataset scored 0 – 0.2, some with a score of 0.3 – 0.6, and very few with a score of 0.7 – 1, the results of the toxicity analysis were carried out by Join (data merging) with the results of sentiment analysis to compare the Toxicity from social media (Facebook & Twitter) and sentiment (Positive & Negative). Based on the joint results obtained, it can be concluded that toxicity scores with a range of 0.4-1.0 are more often found in sentiments with positive labels. The toxicity score 0-0.4 is more or less the same for sentiment with source (Facebook & Twitter) and label (positive & negative). The difference between sentiment with Facebook & Twitter sources is in the toxicity score of 0.6-0.8 where Facebook sentiment, especially with a positive label, is higher than Twitter sentiment.

IV. Conclusion

This study produces a framework combining two analyses: sentiment and toxicity. The result is a word cloud, chart, and tabular. Based on the results found, it can be concluded that the analysis of Digital bounty sentiment can be done using the classification method; at the comparative stage of the classification method, four methods are compared, namely KNN, DT, SVM, and NBC. Of the four methods compared, SVM was selected with an accuracy of 75.01% and an AUC of 0.909, although the accuracy of SVM is smaller than NBC, but with NBC's AUC of only 0.550 and close to 0.5 [37]. The results of the SVM classification on public sentiment towards Digital bounty are positive, with positive opinion results (supporting digital bounty on the SVM classification method of 1540 and negative of 439. Factors that cause positive sentiment (supports Digital bounty post) include indications of doxing, shaming, and efforts to viral digital bounty. The results of the toxicity analysis show that most score toxicity of the 90% dataset is 0 – 0.4, this proves that the toxicity on digital bounty posts is not high. The relationship between toxicity and digital bounty posts can be seen in comments with a toxicity score of 0.4 – 1 where the majority of comments come from positive labeled sentiments. This study has not compared with the other methods such as implementing SMOTE, PSO, GA, etc. Therefore, this research could still be developed in the future to increase accuracy.

REFERENCES

[1] J. A. Pratama, Y. Suprijadi, and Z. Zulhanif, “The Analisis Sentimen Sosial Media Twitter Dengan Algoritma Machine Learning Menggunakan Software R,” *Jurnal Fourier*, vol. 6, no. 2, p. 85, Oct. 2017, doi: 10.14421/fourier.2017.62.85-89.

[2] APJII, “LAPORAN SURVEI INTERNET APJII 2019-2020,” 2020.

[3] M. Isnaini, S. Sarwoprasodjo, R. A. Kinseng, and K. Kholil, “Praktik vigilantisme digital di media sosial dalam konflik antarkelompok,” *Jurnal Studi Komunikasi (Indonesian Journal of Communications Studies)*, vol. 4, no. 3, p. 749, Nov. 2020, doi: 10.25139/jsk.v4i3.2468.

[4] D. Trottier, “Digital Vigilantism as Weaponisation of Visibility,” *Philos Technol*, vol. 30, no. 1, pp. 55–72, Mar. 2017, doi: 10.1007/s13347-016-0216-4.

[5] M. I. Bustomi, “Polda Metro Jaya: Angka Kriminalitas Naik Selama Pandemi Covid-19,” *Kompas*, Oct. 05, 2021.

[6] A. A. Triana and A. M. Fauzi, “Dampak Pandemi Corona Virus Diserse 19 Terhadap Meningkatnya Kriminalitas Pencurian Sepeda Motor Di Surabaya,” *Syiah Kuala Law Journal*, vol. 4, no. 3, pp. 302–309, Dec. 2020, doi: 10.24815/sklj.v4i3.18742.

[7] *Undang Undang No.19 Tahun 2016*. Indonesia, 2016.

[8] *Undang Undang No.11 Tahun 2008*. Indonesia, 2008.

[9] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar, and S. Alfarizi, “Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023”, doi: 10.31294/jtk.v4i2.

[10] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, “SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM,” *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: https://t.co/NfhnfMjtXw

[11] A. N. Ulfah and M. K. Anam, “Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM),” vol. 7, no. 1, pp. 1–10, 2020, [Online]. Available: http://jurnal.mdp.ac.id

[12] A. Fitri Niasita, P. P. Adikara, and S. Adinugroho, “Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan

Automated Lexicon Word2Vec dan Naive-Bayes," 2019. [Online]. Available: http://j-ptiik.ub.ac.id
[13] D. Rustiana Program Studi Sistem Komputer Perguruan Tinggi Raharja and N. Rahayu Magister Teknologi Informatika Perguruan Tinggi Raharja, "ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES," *Jurnal SIMETRIS*, vol. 8, 2017.
[14] R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," vol. 3, no. 1.
[15] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Technology and Science (BITS)*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
[16] N. Alperstein, D. J. Barnett, and P. Pascual-ferra, "Toxicity and verbal aggression on social media : Polarized discourse on wearing face masks during the COVID-19 pandemic," 2021, doi: 10.1177/20539517211023533.
[17] A. Obadimu, E. Mead, M. Maleki, and N. Agarwal, "Developing an Epidemiological Model to Study Spread of Toxicity on YouTube," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 266–276. doi: 10.1007/978-3-030-61255-9_26.
[18] H. Almerekhi, S. B. B. J. Jansen, and C. S. B. H. Kwak, "Investigating Toxicity Across Multiple Reddit Communities, Users, and Moderators," in *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, Association for Computing Machinery, Apr. 2020, pp. 294–298. doi: 10.1145/3366424.3382091.
[19] R. P. Sidiq, B. A. Dermawan, and Y. Umaidah, "Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 3, p. 356, Sep. 2020, doi: 10.32493/informatika.v5i3.6571.
[20] Vojtech Draxl, "Web Scraping Data Extraction from websites," 2018.
[21] "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far - The New York Times." https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html (accessed May 06, 2022).
[22] M. Mancosu and F. Vegetti, "What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data," *Social Media and Society*, vol. 6, no. 3, Jul. 2020, doi: 10.1177/2056305120940703.
[23] "Facebook TOS." https://web.facebook.com/terms.php?_rdc=1&_rdr (accessed May 06, 2022).
[24] R. Arief and K. Imanuel, "ANALISIS SENTIMEN TOPIK VIRAL DESA PENARI PADA MEDIA SOSIAL TWITTER DENGAN METODE LEXICON BASED Universitas Gunadarma 1, 2 Jalan Margonda Raya No 100 Depok Jawa Barat 16424 Sur-el : rifiana@staff.gunadarma.ac.id 1 , karel4404@gmail.com 2," *Jurnal Ilmiah MATRIK*, vol. 21, no. 3, 2019.
[25] "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says." https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=12c0eb4a6f63 (accessed May 06, 2022).
[26] O. Dwiraswati and K. Nazaruddin Siregar, "ANALISIS SENTIMEN PADA TWITTER TERHADAP PENGGUNAAN ANTIBIOTIK DI INDONESIA DENGAN NAIVE BAYES CLASSIFIER SENTIMENT ANALYSIS ON TWITTER ABOUT THE USE OF ANTIBIOTICS IN INDONESIA WITH NAIVE BAYES CLASSIFIER," 2019. [Online]. Available: www.search.twitter.com
[27] M. Winda, P. Sistem, I. Sekolah, T. Manajemen Informatika, D. Komputer, and N. Mandiri, "INTI NUSA MANDIRI ANALISIS SENTIMEN OPINI PUBLIK MENGENAI SARANA DAN TRANSPORTASI MUDIK TAHUN 2019 PADA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES, NEURAL NETWORK, KNN DAN SVM", [Online]. Available: http://www.nusamandiri.ac.id
[28] F. S. Pamungkas, B. D. Prasetya, and I. Kharisudin, "Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 3, pp. 689–694, 2019, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/
[29] V. Kevin, S. Que, : Analisis, S. Transportasi, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)," 2020. [Online]. Available: www.tripadvisor.com,
[30] N. Tri Romadloni, I. Santoso, S. Budilaksono, and M. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL COMMUTER LINE."
[31] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.
[32] A. Deviyanto, M. R. Didik Wahyudi, and T. Informatika UIN Sunan Kalijaga Yogyakarta Jl Marsda Adi Sucipto No, "PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *Jurnal Informatika Sunan Kalijaga)*, vol. 3, no. 1, pp. 1–13, 2018, [Online]. Available: https://twitter.com/search?l=id&q=AHY%20since%3A2017-01-01%20until%3A2017-01-
[33] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)," *Renew Energy*, vol. 115, pp. 411–422, 2018, doi: 10.1016/j.renene.2017.08.061.
[34] "About the API - FAQs." https://support.perspectiveapi.com/s/about-the-api-faqs (accessed May 01, 2022).
[35] A. Gruzd, P. Mai, and Z. Vahedi, "Studying Anti-Social Behaviour on Reddit with Communalytic."
[36] "Tutorial: Toxicity Analysis – Communalytic." https://communalytic.com/video-tutorials/tutorial-4-toxicity-analysis/ (accessed May 06, 2022).
[37] I. Pasquetto, B. Swire-Thompson, and M. A. Amazeen, "Tackling misinformation: What researchers could do with social media data," *Harvard Kennedy School Misinformation Review*, Dec. 2020, doi: 10.37016/mr-2020-49.
[38] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Wiley, 2013. doi: 10.1002/9781118548387.
[39] A. N. Ulfah and M. K. Anam, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," vol. 7, no. 1, pp. 1–10, 2020, [Online]. Available: http://jurnal.mdp.ac.id
[40] R. Kusumaningrum, S. Adhy, and Suryono, "WCLOUDVIZ: Word cloud visualization of Indonesian news articles classification based on Latent dirichlet allocation," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 16, no. 4, pp. 1752–1759, Aug. 2018, doi: 10.12928/TELKOMNIKA.v16i4.8194.