JOIN (Jurnal Online Informatika)
Volume 7 No.1 | June 2022: 132-141
DOI: 10.15575/join.v7i1.900

p-ISSN: 2528-1682
e-ISSN: 2527-9165

Sentiment Analysis for the Brazilian Anesthesiologist Using Multi-Layer Perceptron Classifier and Random Forest Methods

Jelita Asian[1], Moneyta Dholah Rosita[2], Teddy Mantoro[3]
[1,2]School of Computer Science, Nusa Putra University, Indonesia
[3]Media-Tech Lab, Department of Computer Science, Sampoerna University, Indonesia

Article Info

Article history:

Received August 13, 2022
Revised September 06, 2022
Accepted September 06, 2022
Published September 07, 2022

Keywords:

Brazilian Anesthesiologist
MLP Classifier
Random Forest
Sentiment Analysis

ABSTRACT

Sexual harassment is defined as giving sexual attention both verbally, either in speech or writing, and physically to victims who are predominantly women, On July 13, 2022, there was a tweet featuring a video of sexual harassment that made it trend in various countries. The video irritated Twitter users and made various comments resulting in various sentiments that can be analyzed using sentiment analysis. The purpose of this study is to see what the public thinks about the sexual harassment case of Brazilian anesthesiologist. Besides the sentiment analysis, another aim of this study is to see how objective are those sentiments based on their polarity. This study uses a comparison of two methods in sentiment analysis, namely Multi-Layer Perceptron Classifier and Random Forest, and labeling automatically using TextBlob. This results in 94.44% accuracy, 94.44% precision, 92% recall and 93% f1_score. For MLP Classifier and accuracy 96.42%, precision 94.44%, recall 96.66% and f1_score 95.56% for Random Forest. Sentiment polarity score from the TextBlob is -0.5 and subjectivity is 0.4 which indicates that most statements are negative and subjective score is 0.4, which means those sentiments are subjective in nature.

Corresponding Author:

Jelita Asian
School of Computer Science,
Nusa Putra University,
Jl. Raya Cibolang No. 21, Sukabumi, West Java, Indonesia
Email: jelitayang@gmail.com

1. INTRODUCTION

Sexual harassment is defined as giving sexual attention both verbally, either in speech or in writing, and physically to a victim, who are predominantly women, even though it is beyond the wishes of the victims [1]. Victims often accepted sexual harassment as if they were 'natural' and hardly reported the acts because they thought they might get retribution from the offenders especially if the offenders are in a position of power and they might not get the justice they deserved. Sexual harassment is the most common act of sexual crimes against women in society.

With the popularity of social media including Twitter, almost anything can be uploaded by users. On July 13, 2022, there was a viral tweet on Twitter featuring a video of sexual harassment that made the offender trending in various countries, The video footage shows an anesthesiologist at a Brazilian hospital, who performed sexual harassment on a woman who at that time underwent a caesarean section. The pregnant woman's head was visible. The perpetrator went under the woman's cover, sexually abused his patient while she was lying down and another doctor was still performing the surgery. After he finished performing the action, he then took a tissue and wiped the victim's mouth to hide the evidence of his crime. The video irritated Twitter users. They threw scathing words at the doctor. These various comments generated various sentiments that could be analyzed using sentiment analysis.

Sentiment Analysis is one of the applications of Natural Language Processing (NLP) which is used to judge somebody's opinion and conveyed emotion to the reader in a form of a document or a speech [2]. There are two types of data categorized by sentiment analysis: facts and subjective opinions. An example of a fact is "I just bought a new Samsung phone" and an example of an opinion is "I love my new Samsung". These two could also overlap for example "I love the clarity of the camera in my new Samsung." Sentiment analysis is supposed to separate facts and opinions. The main task of sentiment analysis is to identify whether there is a sentiment in the source text and categorize the sentiment as positive, negative, or neutral.

In this paper, sentiment analysis is used to analyze public opinions from the Web in the form of tweets. Sentiment analysis for tweets is inherently difficult because the nature of tweet with its limited length and it has slangs, misspellings, and emoticons [3]. The purpose of this study is to see what the public thinks about the sexual harassment case of Brazilian anesthesiologist. Besides the sentiment analysis, another aim of this study is to see how objective are those sentiments based on their polarity.

As compared to Naïve Bayes and Support Vector Machine algorithms, Multi-Layer Perceptron (MLP) perform the best in term of accuracy for sentiment analysis [4] (MLP) Classifier is a feed forward artificial neural network (ANN) with one or more hidden layers [3]. Its purpose is to minimize prediction error by finding the most optimum combination of connection of weights and biases. In general, MLP consists of an input layer that is a collection of neurons for data input, at least one hidden layer as a computational neuron and one output layer as a storage neuron for computational results. In MLP, there are two important parameters, namely the activation function and the optimization function [5].

Research by Dashtipour [6] used two deep learning models (deep auto encoders and deep convolutional neural networks (CNNs)) to review Persian movies. The proposed deep learning models are analyzed and compared with the state-of-the-art shallow multilayer perceptron (MLP) based machine learning model. Simulation results demonstrate the enhanced performance of deep learning over state-of-the-art MLP. Neural Network resembles the biological neurons in which it takes input from several input signals into one output signal. The type of output is determined by the activation function used [7]. The activation functions and their derivatives determine the computational complexity of the network. Optimization function of Neural Network is used to find the best solution out of all possible solutions by analyzing current solution and achieve one or more of the objectives of ANN [8].

In addition to using the MLP Classifier, this research was also carried out using the Random Forest method. Random Forest produced the highest accuracy as compared to K-NN, Bernoulli, Decision Tree, Support Vector Classifier, and XG-Boost in order to detect anxiety based on a study by Saifullah et al [9]. Random Forest method is popular because it can deal with various predictions and only a few parameters need to be adjusted [10]. It can be accurate to predict data of small size with high dimensional feature. Random Forest, is a method developed from the CART (Classification and Regression Trees) method, which is also a method or algorithm of the decision tree technique. Random Forest is a combination of each existing decision tree technique, then merged and combined into a model [11]. In this paper, we have compared Support Vector Machine, Random Forest and Random Forest Support Vector Machine algorithms (RFSVM) which are very suitable for generating rules in classification technique [12]. From the experimental results, it is concluded that Random Forest Support Vector Machine algorithm seems better than the other algorithms for product reviews dataset offered by Amazon.

2. MATERIAL & METHOD

The research is done in several stages as shown in the research flow below. It includes data collection, preprocessing data, labeling data, data visualization, then finally the sentiment analysis using MLP Classifier and Random Forest. The steps of this sentiment analysis research are shown in Figure 1.

For evaluation, the metric of accuracy, recall, precision and F1 score are used [13]. Accuracy indicates the true prediction (true positive and true negative) out of all predictions. Precision shows how many true predictions (true positives) are correct out of all predictions the system deem correct (all positives). Recall indicates how many correct predictions retrieved by the system out of all the correct predictions. F1-score is used to combine recall and precision and its purpose is to balance the two metrics especially if either the recall or precision value is too low.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

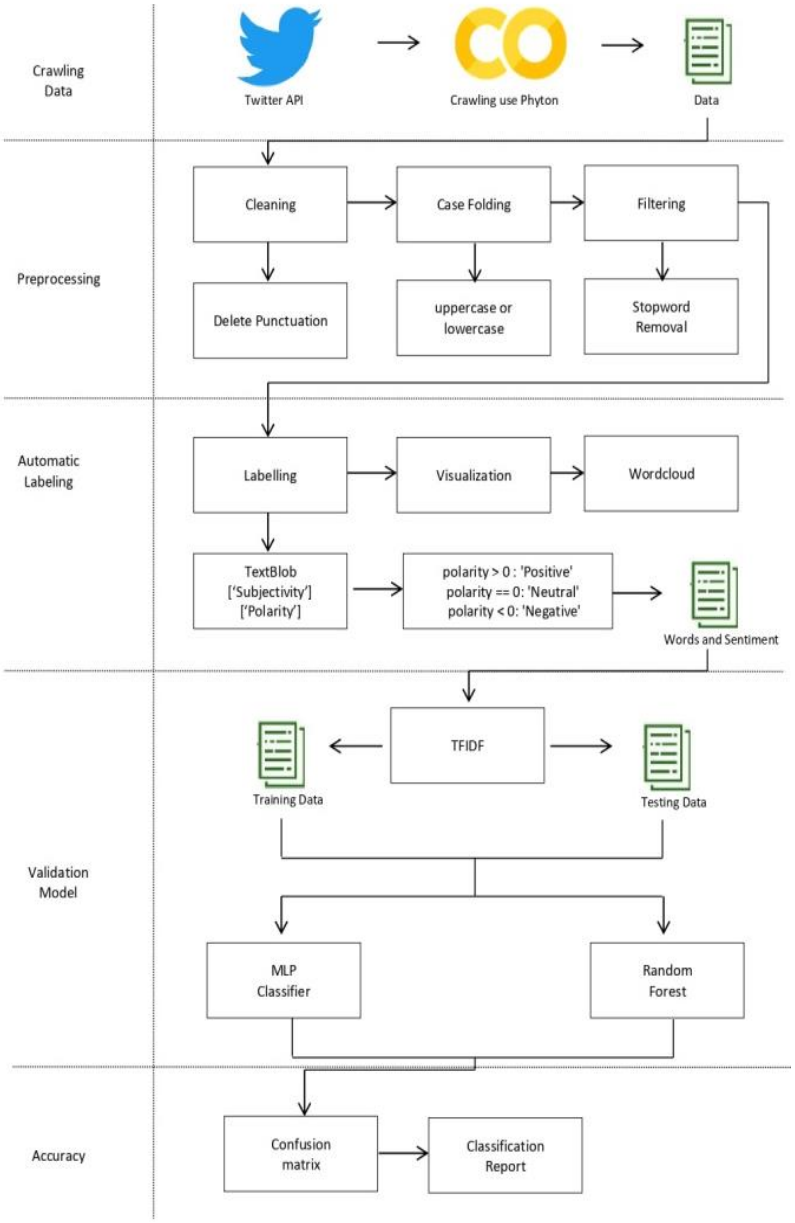

Figure 1. Sentiment Analysis Methodology

2.1. Data Collection

This study crawled on twitter through the Twitter API using the Tweepy module provided by the Python programming language. Test data were gathered for each keyword by using Tweepy library in Python. This library can retrieve some posts related to the specific twitter hashtag [14]. Tweepy is well developed and documented and is commonly used to collect tweets for sentiment analysis [15]. The tweets were scrawled using the keyword "brazilian anesthesiologist". After that the raw tweet dataset is stored in the form of excel file .Xlsx which will be further processed later. Data crawling took place on July 20, 2022. The data collected contains 839 tweets.

2.2. Data Pre-Processing

After the process of crawling the data and saving it, the next step is preprocessing the data, because the dataset is not yet structured. The main task of data preprocessing is to eliminate and overcome data noises

for optimal calculation results. The stages in data preprocessing include data cleaning, case folding, and filtering [16].

2.2.1. Data Cleaning

In the first stage, it is done by deleting special characters in the dataset, while the special characters that are deleted are punctuation marks such as (full stops (.), commas (,), question marks(?), exclamation marks(!), numerical numbers (0-9), and other non-alphabetical characters ($, %, *). This process also removes words that do not match the parsing results, such as usernames beginning with the "@" symbol, the hashtag "#" symbols, Uniform Resource Locator (URL), and emoticons. These signs or symbols and numbers are omitted because they do not have much effect on the determination of labels.

2.2.2. Case Folding

After cleaning, the next step is the folding case. Case folding is the uniformization of all forms of letters, so change all letters either to uppercase or lowercase [17]. This is used to minimize mistakes caused by different capitalizations done by different writers. For example: Jakarta, JAKARTA, and jakarta mean the same thing but might be considered as different words if they were not case folded.

In social media, especially Twitter, writing tweets, there must be differences in the shape of letters, case-folding stages are a changing process of the shape to lowercase letters (lower case) or can also be called uniformity of letters [18].

2.2.3. Filtering

Filtering is a step to eliminate words that have no meaning which are called stopwords. While stopword is a valid word, it does not necessarily convey the meaning of the sentence. Examples of such words in English are “the”, “a”, “an”, “the”, “is”, and “are”. Stopwords removal is the removal of common and often used words but does not have a significant effect in sentences [19]. Stopwords are generally used in information retrieval such as, the google search engine. Index subtraction in the text with the omission of some verbs, adjectives and other adverbs can be included in the stopword list [20].

2.3. Data Labeling

After preprocesing, the next most important thing is to do the labeling. Automatic labeling with Python TextBlob was carried out. Python TextBlob is a Python library (versions 2 and 3) used to process textual data. TextBlob is often used along with Tweepy for sentiment analysis [15]. TextBlob provides APIs that can be used for natural language processing (NLP) such as noun phrase extraction, sentiment analysis, classification, and translation [19]. The resulting TextBlob object is used to process natural language learning and its library can only recognize the English language [20].

2.4. TF-IDF

TF-IDF is an algorithm for calculating the weight of a term is important when it has a high frequency in a single document, but the low frequency in the collection of all other documents. TF-IDF used to improve the efficiency of raw data, while the voting mechanism is used to extract more stable clustering result [22]. As defined, TF is the term frequency in a single document. Terms can be words, or phrases. For documents, the frequency for each term may vary greatly. Therefore, frequency is an important attribute of the term to discriminate itself from other terms. Sometimes, term frequency is directly used as the value of TF. That is, the TF value of term i is

$$TF_i = tf_{ik} \quad (1)$$

Where tf_i denotes the frequency of term i in document j. Since the number of term frequency may be very large, the following formula is also often used to calculate TF value [18].

$$TF_i = log_2(tf_{ij}) \quad (2)$$

As for IDF, various formulas have been proposed. A basic formula was given by Robertson. A later discussion between Spärck Jones [18] and Robertson resulted in the following formula of IDF:

p-ISSN: 2528-1682
e-ISSN: 2527-9165

$$IDF_i = log_2\left(\frac{N}{n_j}\right) + 1 = log_2(N) - log_2(n_j) + 1 \quad (3)$$

where N is the total number of documents in the collection and nj is the number of documents that contain at least one occurrence of the term i.

2.5. Data Visualization

Word cloud is used to determine the dominant topic in each cluster. Word cloud is a form of text data visualization [23]. Word Cloud is also used to show which words or tags are most prevalent in this sentiment analysis. It shows intuitive overview about the sentiment by looking at the majority of words used to describe the feelings of the users [24].

2.6. Multi-layer Perceptron classifier (MLP)

The MLP algorithm is an algorithm that adopts the way neural networks work in living things. This algorithm is famous for being reliable because of the learning process that can be carried out directionally. Learning this algorithm is carried out by updating the back propagation. The determination of the optimal weight will lead to the right prediction results [25]. In MLP, a standard Sigmoid function is used where the weighting amount of a number of inputs and biases are entered into the activation level through the transfer function to produce output, and the units are arranged in a feed-forward topology layer called feed forward neural network (FFNN) [26]. When there is more than one hidden layer, the output of the hidden layer is inserted into the next hidden layer and a separate weight is used for summation to each subsequent layer. MLP consists of an input layer, one or more hidden layers, and an output layer.

2.7. Random Forest

Random Forest, is a method developed from the CART (Classification and Regression Trees) method, which is also a method or algorithm of the decision tree technique [27]. What distinguishes the random forest method from the CART method is that Random Forest applies the bootstrap aggregating (bagging) method and also random feature selection or can be called random feature selection [28].
Random Forest is a combination of each existing decision tree technique, then combined and combined into a model. There are three main points in the Random Forest method, the three main points are (1) bootstrapping sampling to build a prediction tree; (2) each decision tree predicts with a random predictor; (3) then Random Forest makes predictions by combining the results of each decision tree by means of majority votes for classification or average for regression [29].

3. RESULTS AND DISCUSSION

3.1. Data Collecting

The following table shows the samples of 839 tweets obtained by entering the keyword "brazilian anesthesiologist".

Table 1. Crawled Data Result

Tweeted	Username	Tweet
1549148616568310000	Theepretttyone	@TKJanaeee Yess look up Brazilian anesthesiologist
1549136803575330000	leolove_3	brazilian anesthesiologist video, A Brazilian doctor, Giovanni QuintellaBezerra, was caught on camera as he sexually assaults a sedated woman undergoing a C-section. #brazilian #anesthesiologist #Video ðŸ‘‡ðŸ‘‡ https://t.co/4iEnq2WjRl
1549108900896570000	Barbieriana_	@MnHopes A Brazilian anesthesiologist was raping women during their C Sectionsâ€¦.do you know how hazardously horny you have to be??? Straight to prison

3.2. Data Pre-processing

Here are the results of preprocessing by taking one sample text. Starting from the raw data from the results of the crawling process, then preprocessing.

Table 2. Before Pre-processing

Original tweet:
@noelcastanza: That brazilian anesthesiologist rape video almost made me quit social media Atrocities have always happend but it can't…

Table 3. After Pre-processing

Clean tweet:
that brazilian anesthesiologist rape video almost made me quit social media atrocities have always happend but it can t

3.3. Data Labeling

For automatic labeling is done using Python TextBlob. Here's a python script to do auto-labeling.

```
Algorithm 1. TextBlob For Automatic Label
Input
from textblob import TextBlob
Output
defanalyze_sentiment(tweet):
analysis = TextBlob(tweet)
if analysis.sentiment.polarity> 0:
return 'Positive'
elifanalysis.sentiment.polarity == 0:
return 'Neutral'
 else:
return 'Negative'
defprepCloud(Topic_text,Topic):
Topic = str(Topic).lower()
```

Then we call the function with df. The results sample results are shown below.

Table 4. Labeling Result

Clean Tweet	Result
brazilian anesthesiologist videoabrazilian doctor giovanniquintellabezerra was caught on camera as he sexually assaults a sedated woman undergoing a csectionbrazilian anesthesiologist	Negative
a brazilian anesthesiologist was raping women during their c sectionsdo you know how hazardously horny you have to be straight to prison	Negative

The probability of positive and negative statement is shown below:

• text=TextBlob ("That brazilian anesthesiologist rape video ")

• print(format(text.sentiment))

• Sentiment (polarity= -0.5, subjectivity=0.4

The result is available in two categories: polarity and subjectivity. Polarity takes value in between -1 to +1. The value 0 indicates neutral, <0 indicates highly negative and >0 indicates highly positive statement. On the other hand, subjectivity also takes the value between 0 to +1. Here 0 indicates statement is highly objective and >0 indicates highly subjective. The above result shows the sentiment polarity score of -0.5 and subjectivity of 0.4. Polarity of -0.5 clearly states that smost sentiments are negative and subjective score of 0.4 indicates that those sentiments are subjective in nature. That makes sense because this case is not a fact like the features of a phone but a sexual harassment case that causes different emotions to different people depending on their own experience. Table 5 shows the result after data labelling.

Table 5. Sentiment Analysis Using TextBlob

Total Positive Tweets are	108

p-ISSN: 2528-1682
e-ISSN: 2527-9165

Total Negative Tweets are	516
Total Neutral Tweets are	215

After autolabeling using Python, TextBlob label there are 108 tweets with positive sentiment, 215 neutral, and 516 negatives. The calculation of the percentage of sentiment classification is obtained as follows:

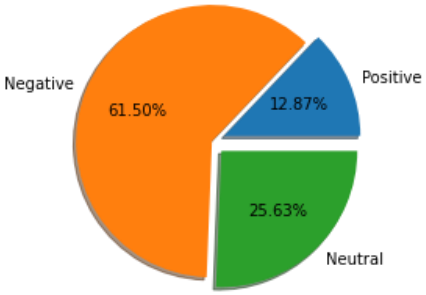

Figure 2. Sentiment Analysis Percentage

3.4. Visualization

Below is the data Visualization using Word Cloud:

Figure 3. Word Cloud Visualization

The most common words used include “discretion” and “sick” which showed that most tweets showed that the viewers were feeling sick with this sexual harassment. It was also up to viewer’s discretion about what to do with the video.

3.5. Validation Model

After crawling, preprocessing and automatic labeling with TextBlob, the next step is to measure the validation model using the Sklearn library in the Python programming language with testing data of 0.3 and training data 0.7. as for the script, it can be seen below:

a. Multi-layer Perceptron classifier (MLP)
Measure accuracy, precision, recall and F1_Score using the Multi-layer Perceptron Classifier (MLP) Method. The script can be seen below.

```
Algorithm 2. Validation Model with Random Forest
Input
from sklearn.neural_network import MLPClassifier
NN = MLPClassifier()
```

```
NN.fit(x_train, y_train)
y_NN = NN.predict(x_test)

Output
print('Akurasi = ', accuracy_score(y_test, y_NN))
print(confusion_matrix(y_test, y_NN))
print(classification_report(y_test, y_NN))
```

With the results as provided in the Tabel 6.

Table 6. MLP Classification Report

	Precision	Recall	F1-score	Support
Negative	0.95	0.97	0.96	158
Neutral	0.93	0.89	0.91	63
Positive	0.93	0.90	0.92	31
Accuracy			0.94	252
Macro avg	0.94	0.92	0.93	252
Weighted avg	0.94	0.94	0.94	252

In the Table 6, the accuracy result is 94.44%, precision is 94.44%, recall is 92% and f1_score is 93%.

b. Random Forest

Measure accuracy, precision, recall, and F1_Score using the Random Forest Method. As for the script, it can be seen below:

```
Algorithm3. Validation Model with Random Forest
Input
from sklearn.ensemble import RandomForestClassifier
RandomForest = RandomForestClassifier()
RandomForest.fit(x_train, y_train)

Output
y_RF = RandomForest.predict(x_test)
print('Akurasi = ', accuracy_score(y_test, y_RF))
print(confusion_matrix(y_test, y_RF))
print(classification_report(y_test, y_RF))
```

With the following results as in Table 7.

Table 7. Random Forest Classification Report

	Precision	Recall	F1-score	Support
Negative	1.00	0.97	0.98	158
Neutral	0.91	0.97	0.94	63
Positive	0.91	0.94	0.92	31
Accuracy			0.96	252
macro avg	0.94	0.92	0.93	252
weighted avg	0.94	0.96	0.95	252

In the Table 7, Random Forest produces accuracy of 96.42%, precision of 94.44%, recall of 96.66% and f1_score of 95,56%. This shows that Random Forest gives better results in term of accuracy, precision, and f1_score as compared to MLP. Random Forest retrieves and identifies correctly the sentiments of tweets. It is possible because the tweet sample is relatively small in term of volume and the size of each tweet due to Twitter word limitation, and each user uses different words to express their sentiment hence the huge dimensional size. TextBlob shows that the sexual harassment case gets more negative sentiment which amount to 61.50%, while positive sentiment got 12.8%, and neutral 25%. The Word Cloud also shows that majority of the viewers felt sick of the video of sexual harassment with the word “sick” was shown as the biggest cloud.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

4. CONCLUSION

In this research it can be concluded Multi-Layer Perceptron Classifier after labelling using TextBlob produces 94.44% accuracy, 94.44% precision, 92% recall and 93% f1_score. On the other hand, Random Forest produces accuracy of 96.42%, precision 94.44%, recall 96.66% and f1_score 95.56%. Sexual harassment cases get more negative sentiment in social media twitter. In general, Random Forest performs better than MLP perhaps due to high dimensionality of the tweets and small number of texts in the tweets and sample sizes. Sentiment polarity score from the TextBlob is -0.5 and subjectivity is 0.4. Polarity of -0.5 indicates that most statements are negative and subjective score is 0.4, which means the sentiments are subjective in nature. That makes sense because this is about sexual harassment case could cause different emotions for different people depending on their experience.

In the current study, the sentiment produced is the result of labelling using TextBlob. For future study, subjective sentiment analysis could also be used as a gauge about how well the sentiment analysis works. These Random Forest and MLP classifier could be tested on tweets of more positive natures and see whether they perform as well as the negative ones. Besides tweets, the system could also be tested on other documents such as news articles or user reviews on Amazon.

5. REFERENCES

[1] L. A. Rudman, E. Borgida & B. A. Robertson. Suffering in silence: Procedural justice versus gender socialization issues in university sexual harassment grievance procedures. *Basic and Applied Social Psychology*, 17(4), 519–541. 1995. https://doi.org/10.1207/s15324834basp1704_6

[2] T. Luo, S. Chen, G. Xu, J. Zhou, Sentiment Analysis. In: Trust-based Collective View Prediction. *Springer, New York, NY*. 2013,https://doi.org/10.1007/978-1-4614-7202-5_4

[3] D. A. Alboaneen, H. Tianfield and Y. Zhang, 2017, Sentiment analysis via multi-layer perceptron trained by meta-heuristic optimisation *2017 IEEE International Conference on Big Data (Big Data*), pp. 4630-4635, doi: 10.1109/BigData.2017.8258507

[4] Jyoti Budhwar, Sukhdip Singh, 2021, Sentiment Analysis based Method for Amazon Product Reviews, *International Journal Of Engineering Research & Technology (Ijert) Icact* – 2021 (Volume 09 – Issue 08)

[5] J. Kusuma, B. Herawan, W. Wanayumini and R. Rosnelly "Komparasi Metode Multi Layer Perceptron (MLP) dan Support Vector Machine (SVM) untuk Klasifikasi Kanker Payudara Komparasi Metode Multi Layer Perceptron (MLP) dan Support Vector Machine (SVM) untuk Klasifikasi Kanker Payudara" 2022 *in Multimedia Artificial Intelligence Networking Database* 7:1, 51-60, June 2022.MIND, vol.7 no 1 pp 51 – 60, Juni 2022.

[6] Dashtipour, K., Gogate, M., Adeel, A., Ieracitano, C., Larijani, H., Hussain, A. (2018). Exploiting Deep Learning for Persian Sentiment Analysis. In: , et al. Advances in Brain Inspired Cognitive Systems. BICS 2018. *Lecture Notes in Computer Science, vol 10989. Springer, Cham*. https://doi.org/10.1007/978-3-030-00563-4_58.

[7] D.R. Vivas, M. Javier, B. Victor, O. L. G'omez and J. H. Reina. "Neural-Network Quantum States: A Systematic Review." 2022.

[8] A. Maher & Hussain, Suhail & Ustun, T. Selim & Sarker, Mahidur & Hannan, M. A. & Mohamed, Ramizi & Ali, Jamal & Mekhilef, Saad & Milad, Abdalrhman. Artificial Neural Networks Based Optimization Techniques: A Review. Electronics. 10. 2689. 10.3390/electronics10212689. 2021

[9] S. Saifullah, Y. Fauziah, and A. S. Aribowo (2021), "Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety Based on Social Media Data," DOI: 10.26555/jifo.v15i1.a20111.

[10] G. Biau and E. Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016.

[11] K. Budiman, N. Zaatsiyah, U. Niswah, and F. F*aizi, "*Analysis of Sexual Harassment Tweet Sentiment on Twitter in Indonesia using Naive Bayes Method through National Institute of Standard and Technology Digital Forensic Acquisition Approach", *JAIST*, vol. 2, no. 2, pp. 21-30, Oct. 2020.

[12] Y. Al Amrani, M. Lazaar, K.E. El Kadiri, Random forest and support vector machine based hybrid approach to sentiment analysis *Procedia Comput. Sci*., 127 (2018), pp. 511-520

[13] Mantoro, M. Merdianti and M. A. Ayu, "Sentiment Analysis of the Papuan Movement on Twitter Using Naïve Bayes Algorithm," *2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED), Sukabumi, Indonesia*, 2021, pp. 1-5. doi: 10.1109/ICCED53389.2021.9664868.

[14] T. Mantoro, M. A. Ayu, R. T. Handayanto. Machine Learning Approach for Sentiment Analysis in Crime Information Retrieval, *3rd International Conference of Computer And Informatics Engineering IC2IE 2020*, Jakarta, 15-16 Sept 2020.

[15] B. Balachander, P. S. J. Sentimental Analysis of Twitter Data Using Tweepy and Textblob. *International Journal of Advanced Science and Technology*, 2020 *29*(3), 6537 - 6544. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/7243

[16] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Comput. Sci.*, vol.161, pp. 707–714, 2019, doi: 10.1016/j.procs.2019.11.174.
[17] C.D. Manning, P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008. Retrieved from https://nlp.stanford.edu/IR-book/information-retrieval-book.html.
[18] Bahrawi. Sentiment Analysis Using Random Forest Algorithm Online Social Media Based. *Journal Of Information Technology and Its Utilization*, Volume 2, Issue 2, December-2019: 29-332. h.29-33..
[19] V. A. Fitri, R. Andreswari and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of AntiLGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Computer Science*, vol. 161, pp. 765-772, 2019.
[20] R. Sari and R. Y. Hayuningtyas, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website" *Indones. J. Softw. Eng*., vol. 5, no. 2, pp. 51–60, 2019, doi: 10.31294/ijse.v5i2.6957.
[21] I. G. S. Mas Diyasa, N. M. I. Marini Mandenni, M. I. Fachrurrozi, S. I. Pradika, K. R. Nur Manab and N. R. Sasmita, "Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob", *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, pp. 012034, 2021.
[22] M.A. Ayu, S.S. Wijaya, T. Mantoro, An automatic lexicon generation for Indonesian news sentiment analysis: A case on governor elections in Indonesia, *Indonesian Journal of Electrical Engineering and Computer Science*, 16 (3), pp. 1555-1561, 2019.
[23] R. Kurniasari, R. Santoso, and A. Prahutama, "Analisis Kecenderungan Laporan Masyarakat Pada "LAPORGUB..!" Provinsi Jawa Tengah Menggunakan Text Mining Dengan Fuzzy C-Means Clustering," *Jurnal Gaussian*, vol. 10, no. 4, pp. 544-553, Dec. 2021. https://doi.org/10.14710/j.gauss.v10i4.33101.
[24] S. S. Chai, W. L. Cheah, K. L. Goh, Y. H. R. Chang, K. Y. Sim, K. O. Chin, "A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in Sarawak, Malaysia", *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 2794888, 11 pages, 2021. https://doi.org/10.1155/2021/2794888.
[25] F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word Cloud Explorer: Text Analytics Based on Word Clouds," *2014 47th Hawaii International Conference on System Sciences*, 2014, pp. 1833-1842, doi: 10.1109/HICSS.2014.231.
[26] K. Khadijah and R. Kusumaningrum. "Ensemble Classifier untuk Klasifikasi Kanker Payudara." IT JOURNAL RESEARCH AND DEVELOPMENT, 2019, https://doi.org/10.25299/itjrd.2019.vol4(1).3540.
[27] A. Setiawan., I.G. S. M. Diyasa, M. Hatta, and E. Y. Puspaningrum. "Mixture gaussian V2 based microscopic movement detection of human spermatozoa" *International Journal of Advances in Intelligent Informatics*, Vol. 6 (2), pp. 210-222, 2020.
[28] I. Sunni dan D. H. Widyantoro, "Analisis Sentimen Dan Ekstraksi Topik Penentu Sentimen Pada Opini Terhadap Tokoh Publik" *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, vol. 1, no. 2, pp. 200-206, Jul. 2022.
[29] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 1, no. 1, pp. 27-31, 2018.