Analysis and Visualization of Political Sentiments on Twitter Using Machine Learning Algorithms

Adlina Najihah Mohd Pouzi[1], Shuzlina Abdul- Rahman[2], Siti Sakira Kamaruddin[3] and Norlina Mohd Sabri[4*]

[1] Tata Consultancy Services, Axiata Tower 9, Kuala Lumpur Sentral, Malaysia
[2] Research Initiative Group of Intelligent Systems, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Selangor, Malaysia
[3] School of Computing, Universiti Utara Malaysia, Sintok, Malaysia
[4] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Cawangan Terengganu, Malaysia
norli097@uitm.edu.my

Abstract. The Malaysian political system is a constitutional monarchy with a federal parliamentary democracy. In the era of technological advancement, social media has become a huge part of human life. The relationship between politicians and the public has progressed through the use of social media, where they connect with the citizens and build personal connections with them. However, the people's voice remains in the grey where demands and needs are not easily identified. It is critical for the government to understand public opinions to define strategies and make decisions. Hence, in this work, sentiment analysis has been implemented by focusing on the public sentiments toward the Malaysian government, specifically on the governance of the Perikatan Nasional political party during the Covid-19 pandemic. In this study, a corpus-based lexicon method was used to identify the positive, negative and neutral labels for the training data. These labels were used in the machine learning models, which were the Naive Bayes, Support Vector Machine and Logistic Regression. Furthermore, the models' performance has been compared and the results have shown that Naive Bayes has surpassed other models with the highest performance in accuracy, precision and F1- score. The main contribution of this study is that this research has successfully analyzed the relative performance of three machine learning algorithms for the political sentiments. The results of the sentiment analysis were visualized using Power BI for a clear view with a deeper understanding.

Keywords: Sentiment Analysis, Machine Learning, Politics.

1 Introduction

The internet and social networking sites have brought a huge progression in creating a whole new level of engagement between the government's public voting and political members [1]. Social networking services (SNS) such as Facebook and Twitter platforms allow individuals to learn information, share information, discuss, exchange opinions, and develop networks. Many political MPs around the world use Twitter as a

N. A. S. Abdullah et al. (eds.), *Proceedings of the International Conference on Innovation & Entrepreneurship in Computing, Engineering & Science Education (InvENT 2024)*, Advances in Computer Science Research 117,
https://doi.org/10.2991/978-94-6463-589-8_26

channel for interacting and influencing citizens to support the goals of their political party [2]. Sentiment analysis is an opinion-mining technique that analyses people's feelings, opinions, attitudes and emotions about things like products, services, organizations, individuals, issues, events, and topics [3]. Sentiment Analysis (SA) is one of the Natural Language Processing (NLP) techniques that machines use to comprehend human language to communicate with each other [4]. It is used to decide whether the data is positive, negative or neutral using the combination of Natural Language Processing (NLP) and machine learning techniques. It is used in the political field to keep track of political beliefs and to find consistency and inconsistency between government claims and actions which can also be used to forecast election outcomes. It is critical for the government to earn the general public's trust and approval to maintain political stability [5]. A loss of public confidence will severely harm the government's ability to function. Thus, the government must show competence by overcoming every shortcoming and flaw. The government was led by Prime Minister Muhyiddin Yassin from 1st March 2020 until August 2021. The government rule at the beginning of the COVID-19 pandemic in Malaysia and their response in reducing the COVID-19 impact were important for the people. The government has been compelled to take stricter measures such as a limited lock-down called the Movement Control Order (MCO) to manage the outbreak and prevent the healthcare system from collapsing [6].

Every action taken during the pandemic has a critical expectation from the people as a consequence of the pandemic, as many have lost their loved ones, lost their jobs, and it has caused countless business bankruptcies, etc. It is important to recognize the opinions on political events or topics in order to improve the policies, and positions [7]. Hence, the goal of this study is to analyze the public sentiment on political movements and the sentiment towards the initiatives taken by the government sparked by the recent Coronavirus pandemic. The sentiment analysis could provide insight for the government to see if the policies and initiatives win the people and can use the information to deploy new strategies. Three machine learning algorithms have been selected in this study, which are Support Vector Machine (SVM), Naive Bayes and Logistic Regression. These are the well known algorithms with good performace in sentiment classification problems. In this study, to further understand the results, the findings were visualized using Microsoft Power BI.

The paper is organized as follows: Section 2 briefly discusses the machine learning algorithms, whereas Section 3 details the methods used in the study. Section 4 summarizes the findings, and Section 5 wraps up the research with a conclusion and recommendations for further research.

2 Literature Review

2.1 Sentiment Analysis

Sentiment Analysis (SA) is one of the Natural Language Processing (NLP) techniques that machines use to comprehend human language to communicate with each other [4]. It is used to decide whether the data is positive, negative or neutral using the combination of Natural Language Processing (NLP) and machine learning techniques. By identifying expressions that are commonly used by people from texts, sentiment analysis can calculate the tonality and polarity to evaluate a person, entity or event [8]. Furthermore, SA also has different approaches namely the machine learning approach, lexicon-based approach and hybrid approach. There are several studies that have implemented SA to analyze political related views. A researcher has analyzed the reasons why some citizens vote while others do not, by comparing the voting behavior of casting and non-casting voters [9]. Another study has utilized the supervised learning algorithm such as Naïve Bayes and SVM in the study to predict which party will win in the election [10]. There was also a study to investigate people's acceptance of Pakatan Harapan as Malaysia's new government [11].

2.2 Machine Learning Algorithms

The three models namely, SVM, NB and LR were chosen to be compared in this study as they are frequently utilized in past studies due to their good performance in sentiment classifications. One of the supervised machine learning models that are commonly used in classification and regression is the Support Vector Machine (SVM). The purpose of the support vector machine algorithm is to locate a hyperplane in an N-dimensional space, where N denotes the number of features that separate data points in the case of sentiment analysis, SVM performs it by locating the hyperplane that separates the n-dimensional classes plotted. Through the work of Sachdeva, the researcher described SVM as easy to be tuned with fewer hyper-parameters [12]. Although it has its own limitations, SVM performs relatively well among other classification models [13].

The Bayesian classification approach is both a supervised learning and a statistical classification method. Because of its simplicity of use in both the training and classification stages, the Naive Bayes text classifier is widely used [14]. The algorithm is based on Bayes' Theorem. By determining probability, the theorem allows one to capture the model's uncertainty in a systematic way [15]. The Naive Bayes technique is very simple, easy to implement, and effective, but there are limitations attached to it such as incomplete training data, continuous variables and attribute independence.

Logistic Regression is utilized to determine the output or outcome when there are one or more independent variables. The output result might be in binary form, that is, 0 or 1. While a classifier predicts the likelihood of a categorical dependent variable which is the dependent being in two categories - yes or no known as logistic regression, multinomial logistic regression is used when there are more than two classes [16]. Sheela [17] proposed the use of LR, Naive Bayes and Bernoulli to classify sentiment and the result shows that LR and Naive Bayes has significant result in classifying correctly and it was found that LR performs well with a large sample size.

3 Methodology

The research framework contains data collection, data pre-processing, data modeling and dashboard design and development. The involved steps are discussed in the subsequent sub-sections.

3.1 Data Collection

The data were scraped from Twitter using Snscrape, which is a scraper for various SNS. It is a Python tool for scraping tweets via Twitter's API without any restrictions or requests limits and without needing a Twitter developer account. It scrapes information such as user profiles, hashtags, searches and returns the results, such as relevant postings from the chosen platform. The data were scraped from Twitter from the 1st January until the 1st August of 2021 for the purpose of researching the opinions on the subject of the general action execution and functioning of the government in assisting the country, especially after the third wave of COVID-19 hit in late of December 2020. In order to study the performance of the government, the keywords used to scrape the data were more towards the head of ministries of the government such as Muhyiddin, Annuar Musa, Azmin Ali, Ismail Sabri, Adham Baba and Perikatan Nasional. The data collected were totaled up to 83,555 tweets.

3.2 Data Preprocessing

There were 3 main phases in the data preprocessing. The phases were data cleaning, data transformation and data labeling.

Data Cleaning. The cleaning of the datasets was done using Python and RapidMiner. The first step is cleaning the raw data such as removing URLs, hashtags, retweets (RT) and tweets mentioned which was done using Python. RapidMiner was used to remove duplicated tweets, as well as symbols contained in the tweets. RapidMiner utilized three operators to clean the dataset. First is Read Excel which is used to read dataset that was stored in Excel format after the datasets had been successfully scraped using Snscrape. Then the second operator used was Remove Duplicates to delete unnecessary redundant tweets. The next operator used is called Replace operator. The scrapped tweets may contain punctuation or special characters which need to be removed to perform sentiment analysis. After the data was cleaned and the noise was removed, the total number of data have been reduced to 60928 tweets.

Data Transformation. The first step involved in this phase is lowering the case. This is due to the model might treat the word with capitalized character differently from one without a capitalized letter, that can result in declination in accuracy. Next is translating the tweets to English, spell checking to minimize the inaccuracy in the data, tokenization, pos-tagging, lemmatization, stopping words removal and generating n-gram using Orange Data Mining. To translate Malay tweets to English, the GoogleTranslate formula and the add-on feature called Translate My Sheet in Google Sheets were utilized.

Once the dataset is cleaned and translated, the next step is to apply tokenization because the data are in the form of text, which is not accepted by the classifier. Tokenization is the process of dividing a piece of text into tokens. Words, characters, and sub-words can all be considered tokens. The chosen approach for tokenization is using whitespace. Because the chosen lemmatization approach is Wordnet lemmatization, the POS tag will be applied first. POS tag or part-of-speech tag is a particular label assigned to each token (word) in a text corpus to denote the part of speech as well as other grammatical categories such as tense, number (plural/singular), case etc. Pos-tag followed by lemmatization is said to be essential to reduce a word to its root form, especially in the lexicon-based method. The lemmatizer will benefit from the information provided by the tagging. The next step is word normalization. For this study, WordNet lemmatization is implemented. The next step in data transformation before modeling is stop words removal. A stop word is a widely used word such as (”the,” ”a,” ”an,” or ”in”) that a search engine has been designed to reject both while indexing and retrieving results as the result of a search query. A sentence does not need them to provide meaning. Removing unnecessary words that the model must evaluate, will improve the model’s performance. The last step is generating an n-gram. This study uses the combination of unigram and bigram to generate the features to improve the classification result. The N-gram of range = [1, 2] was applied in this study. After applying the n-gram for the sentence ”thank you azmin ali”, the output will be [’thank’, ‘you’, ‘azmin’, ’ali’, ’thank you’, ‘you azmin’, ‘azmin ali’].

Data Labeling. Data labeling is required in preparing the data before moving on to the development phase. The pre-processed dataset is labeled to achieve high-accuracy results. It is labeled as either positive, negative or neutral based on the polarity score. The features computed from the extraction were in the form of a dictionary that included positive, negative, neutral and compound scores. Data were labeled using the VADER compound score as shown in Table 1.

Table 1. Range of compound score.

Sentiment	Compound Score
Positive	≥ 0.05
Neutral	-0.05 < Score < 0.05
Negative	≤ -0.05

3.3 Model Development

Three machine learning classifiers, which were Support Machine Vector (SVM), Naive Bayes (NB) and Logistic Regression (LR) were implemented and later compared based on their accuracy and performance. The labeled dataset was split into different ratios of training and test sets. The models were trained using the training set and tested on the testing set once the classifier had been trained.

The process started with retrieving the labeled dataset. Using the Data Sampler widget, the dataset was split into two, which were for the training set and the test set using the sampling technique. Different sampling techniques were used in this study for comparison purposes such as different k-folds of cross-validation (5-fold and 10-fold), Stratified sampling and split ratios of 80:20, 70:30, and 60:40. These are common methods in machine learning to sample datasets. After loading the labeled dataset, both training and test set data are pre-processed using the text pre-process widget. The pre-processing includes transforming data into lower cases, tokenization, POS-tagging, lemmatization, stop words removal and generating n-gram. This study uses a combination of unigram and bi-gram [1, 2] for the modeling. After pre-processing, a bag of word widgets is used to identify the term frequencies of the corpus. TF-IDF is then applied to convert words into numerical vectors for term weighting. A statistical measure of how essential a word is to a document in a collection or corpus is the TF-IDF [18]. After that, the Select Column widget was used for attribute and feature selection where the important and relevant feature was selected and the target label was chosen. The sample data was then used to train the models. Then, the three classifiers were applied. The remaining data act as test data which was used to evaluate the performance. The remaining data was also connected to the prediction module to see how each classifier classifies on the test data. In the test and score module, after training the data, the model was evaluated by testing it on the test data. Finally, for the performance evaluations of the models, the confusion matrix widget was used. This widget returns a list of values for classification task performance.

3.4 Dashboard Development

The dashboard visualizes the result of the sentiments. It provides a clear view of the public opinion on the government performance outcome based on the analysis of the amount of positive sentiment, neutral and negative sentiment. The dashboard aims to give a graphical view of the sentiment results. This phase of the study goes through the visualization that is plotted using the result of the sentiment analysis as well as the designing of the dashboard interface using Microsoft Power BI.

4 Result and Discussion

4.1 Sentiment Extraction

Polarity detection of textual data has been performed using the methods of VADER. The result obtained was in the form of a dictionary that included positive, negative, neutral and compound scores. Table 2 shows the result of VADER sentiment extraction. VADER labeled the total data set into 22272 positive tweets, 22088 neutral tweets and 16568 negative tweets.

Table 2. Results of Vader sentiment extraction.

Class	Total
Positive	22272
Neutral	22088
Negative	16568

4.2 Different Techniques and Performances

The dataset was evaluated and compared using k-fold cross-validation (10-fold). The evaluation was also done using stratified sampling with a 70% training set size. The k-fold cross validation is a popular method for evaluating classification algorithms because it reduces the bias in classifier performance estimation. The stratified sampling divides the population into subgroups with the right number of instances randomly sampled from each subgroup to ensure the test set is representative of the whole population. The labeled dataset is used for both training and testing in this approach. Table 3 shows the result of applying the two techniques in this study.

Table 3. Comparison results of two techniques for result evaluation.

Classifiers	Technique	Results				
		AUC	CA	F1	PR	RC
SVM (RBF)	k-fold Cross Validation	0.965	0.690	0.611	0.584	0.690
	Stratified Sampling	0.979	0.965	0.965	0.965	0.965
Naive Bayes	k-fold Cross Validation	**1.00**	**0.997**	**0.997**	**0.997**	**0.997**
	Stratified Sampling	**1.00**	**0.998**	**0.998**	**0.998**	**0.998**
Logistic Regression	k-fold Cross Validation	1.00	0.990	0.990	0.990	0.990
	Stratified Sampling	1.00	0.990	0.990	0.990	0.990

It can be seen that Naive Bayes' performance outperforms other models in both techniques since the results of AUC, accuracy (CA), F1-score (F1), precision (PR) and recall (RC) have achieved the highest compared to other two algorithms. Therefore, it can be concluded that Naive Bayes has proven to be the best model for this project. As mentioned in [15] and other studies, the VADER approach is sufficient to produce a good result as the evaluations show a significant rise in performance when used on a large dataset.

4.3 Different Split Ratio Comparison

In this project, three split ratios were used: 80:20, 70:30, and 60:40. Each split ratio results in a varied level of accuracy and the split ratio with the highest level of accuracy is highlighted in Table 4. The results of Support Vector Machine (SVM) show that it works best when used with a larger training size (80:20) in terms of accuracy. Meanwhile, both Naive Bayes and Logistic Regression results reveal that both models do not require a larger training size to achieve the best result as they both work the best in the 70:30 ratio. In conclusion, it was proven that the ideal training and testing ratio for ML models was 70:30, as reported in a work by a previous researcher [19].

Table 4. Results of different split ratio.

Classifiers	Ratio	Comparison Results				
		AUC	CA	F1	PR	RC
SVM	80:20	0.983	0.728	0.631	0.652	0.728
	70:30	**0.992**	**0.725**	**0.626**	**0.843**	**0.725**
	60:40	0.986	0.720	0.623	0.841	0.720
Naive Bayes	80:20	1.000	0.997	0.997	0.997	0.997
	70:30	**1.000**	**0.998**	**0.998**	**0.998**	**0.998**
	60:40	0.999	0.997	0.997	0.997	0.997
Logistic Regression	80:20	1.000	0.989	0.989	0.989	0.989
	70:30	**1.000**	**0.990**	**0.990**	**0.990**	**0.990**
	60:40	1.000	0.987	0.987	0.988	0.987

4.4 Dashboard Visualization

The dashboard of the public's perspectives of the government performance was done in three pages. The first page was created to provide readers with a clear understanding of each month's perspective improvement or deterioration with respect to the government governance during the period chosen. The second page provides information on the Ministers of the government namely Muhyiddin Yassin, Adham Baba, Anuar Musa, Azmin Ali, and Ismail Sabri. The third page is designed to give information or topics that are most talked about in each sentiment using a sentiment filter through a word cloud. Based on the results, the highest opinion towards Perikatan Nasional is negative followed by neutral and positive. Using the information, the government and the public can track the consistency and inconsistency of the government claims and actions. Fig. 1 shows the third page which provides the profound information on both datasets of Perikatan Nasional and the politicians. The word cloud summarizes the text and investigates the subjects covered by each dataset.

Fig. 1. Sentiment analysis results in Page 3 from Dashboard.

5 Conclusion

This study compares three classification algorithms to assess the performance in classifying sentiment text. The evaluation process involves several aspects, including sampling approach, data size and training size. The best classifier among other models in this study is Naive Bayes with five-fold cross-validation, while the best-split ratio between the training set and test set is 70:30. The finding of the study is that it has successfully analyzed the public's sentiment and acceptance towards the political or government situation during the Covid-19 pandemic. The period of study was considered critical as the country has to survive from the consequences of the pandemic. For future improvement, real-time sentiment analysis should be considered as it could track and analyze mentions of related keywords, whenever they occur. Aside from that, more diverse keywords involving the opposition party can be used in future to better understand public opinion regarding the country's political situation. Future work also might look at Malay corpus instead, as the Malay manner of speaking and expression can differ from English, possibly including appropriately identifying sarcasm. Sarcasm might appear to be positive even though the context recognized in Malay is negative.

Acknowledgments. The authors would like to thank the College of Computing, Informatics and Mathematics from UiTM Shah Alam and UiTM Cawangan Terengganu for the support given towards the completion of the research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alashri, S., Kandala, S. S., Bajaj, V., Ravi, R., Smith, K. L., Desouza, K. C.: An Analysis of Sentiments on Facebook during the 2016 US Presidential Election. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 795–802. IEEE, San Francisco (2016)
2. Britzolakis, A., Kondylakis, H., Papadakis, N.: A Review on Lexicon-based and Machine Learning Political Sentiment Analysis using Tweets. International Journal of Semantic Computing **14**(4), 517–563 (2020)
3. Budiharto, W., Meiliana, M.: Prediction and Analysis of Indonesia Presidential Election from Twitter using Sentiment Analysis. Journal of Big data **5**(1), 1–10 (2018)
4. Beppe, K.: Tweeting opinions: How does twitter data stack up against the polls and betting odds?. Linnaeus University, Sweden (2018)
5. Ndubai, R. E., Mbeche, I. M.: A Study of the Intervening Effect of Political Stability on the Relationship between Performance Contracting and Measurement, and Public Service Delivery in Kenya. Open Access Library Journal **4**(4), 1-14 (2017)
6. Altahir, A. A., Mathur, N., Thiruchelvam, L., Mustafa Abro, G. E., Radzi, S., Dass, S. C., Gill, B. S., Sebastian, P., Zulkifli, S. A., Asirvadam. V. S.: Modeling the Impact of Lockdown on Covid-19 Spread in Malaysia. bioRxiv, (2020)
7. Gangwar, A., Mehta, T.: Sentiment Analysis of Political Tweets for Israel using Machine Learning. arXiv preprint arXiv, (2022)

8. Haselmayer, M., Jenny, M.: Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding. Quality & Quantity **51**, 2623–2646 (2017)
9. Hossain, M. K. R., Aktar, M. M., Islam, M. S.: People's Voting Behavior In Local Election: A Study On Annadanagar Union Parishad , Pirgachha , Rangpur. IOSR Journal of Humanities and Social Science **22**(03), 1-14 (2017)
10. Khurana Batra, P., Saxena, A., Shruti, Goel, C.: Election result prediction using twitter sentiments analysis. In: 6th International Conference on Parallel, Distributed and Grid Computing, pp. 182–185. IEEE, India (2020)
11. Gobithaasan, R. U., Farhana, N. U. R., Che, S.: Sentiment Analysis of People ' S Acceptance Towards the New Malaysian Government Using Naïve Bayes Method. UMT Journal of Undergraduate Research **2**(3), 93–102 (2020)
12. Sachdeva, S. J., Abhishek, R.: Comparing Machine Learning Techniques for Sentiment Analysis. Ijarcce **8**(4), 67–71 (2019)
13. Rahmath, H., Ahmad, T.: Sentiment analysis techniques-A comparative study. International Journal of Computational Engineering & Management **17**(4), 25–29 (2014)
14. Tripathy, A., Agrawal, A., Rath. S. K.: Classification of sentimental reviews using machine learning techniques. Procedia Computer Science **57**, 821–829 (2015)
15. Ting, S., Ip, W., Tsang, A. H.: Is Naive Bayes a Good Classifier for Document Classification. International Journal of Software Engineering and Its Applications **5**(3), 37–46 (2011)
16. Wankhade, M., Rao, A. C. S., Dara, S., Kaushik, B.: A Sentiment Analysis of Food Review using Logistic Regression. Int J Sci Res Comput Sci Eng InformTechnol **2**(7), 251–260 (2017)
17. Sasikala, P., Sheela, L. M. I.: Sentiment analysis and prediction of online reviews with empty ratings. International Journal of Applied Engineering Research **13**(14), 11525–11531 (2018)
18. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval **2(**1-2), 1–135 (2008)
19. Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., Pham, B. T.: Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering, 1–15 (2021)