

Multistage Sentiment Classification Model using Malaysia Political Ontology

Nur Farhana Ismail, Nur Atiqah Sia Abdullah*, Zainura Idrus

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia

Abstract—Now-a-days, people use social media platforms such as Facebook, Twitter, and Instagram to share their opinions on particular entities or services. The sentiment analysis can get the polarity of these opinions, especially in the political domain. However, in Malaysia, current sentiment analysis can be inaccurate when the netizen tempts to use the combination of Malay words in their comments. It is due to the insufficient Malay corpus and sentiment analysis tools. Therefore, this study aims to construct a multistage sentiment classification model based on Malaysia Political Ontology and Malay Political Corpus. The reviews are carried out in sentiment analysis, classification techniques, Malay sentiment analysis, and sentiment analysis on politics. It starts with the data preparation for Malay tweets to produce tokenized Malay words and then, the construction of corpus using corpus filtering, web search, and filtering using linguistic patterns before enhancing with political lexicons. The process continues with the classifier construction. It started with a generic ontology with Malaysia's political context. Lastly, twelve features are identified. Then the extracted features are tested using different classifiers. As a result, Linear Support Vector Machine yields an accuracy of 86.4% for the classification. It proved that the multistage sentiment classification model improved the Malay tweets classification in the political domain.

Keywords—Malay corpus; political ontology; sentiment analysis; sentiment classification; social media

I. INTRODUCTION

Social media is a common platform for internet users. Netizens can spread and viral issues quickly via social media like Facebook, Twitter, Blog, Instagram, and online platforms. Social media allows people to voice opinions freely on current matters. Their opinions are beneficial, especially for the business, marketing strategies, and policymakers include government. Sentiment analysis tools can analyze their comments into exploitable information.

The existing sentimental analysis classifiers manage to analyze different languages such as English, French, Indian, Arabic, and Chinese. However, it has yet insufficiently in analyzing the Malay language accurately. Each comment containing Malay words will be classified as neutral in most of the social media monitoring tools. It is one of the reasons for the Malay sentiment classifier to support the research on classifying the Malay language, which use lexicon and k-nearest neighbor [1], lexicon [2] and other classification

methods [3]. Besides, there is lacking Malay sentiment analysis that covers the political domain [4].

The author in [4] has conducted a study on the political inclination classifier model for Malay text in social media data. This study focuses on Malay sentiment analysis in the political domain. However, it needs to be improved to get a more accurate sentiment classification. Besides, some corpora are for the abbreviations [5] and hadith [6]. However, there is a need to create a corpus in the political domain.

Currently, most of the agencies use social media monitoring tools to extract comments for strategic planning in marketing, customer behavior, political inclination, and etc. However, the comments that contain Malay words are classified as neutral. This shortage motivates this research to improve the sentiment classification. It sets the interest to investigate the sentiment classification in the political domain due to the political scenario in Malaysia.

The main idea of this study is to propose a multistage sentiment classification model using Malaysia Political Ontology and Malay Political Corpus. This model aims to increase the sentiment classification by adding the entity classification to the existing process. Besides, new features are suggested based on the entity classification. By using this model, the analysis and monitoring process of social media can speed up. Besides, this research helps to expand the knowledge in this field to get better accuracy of sentiment. This paper continues with Section II explains the related works. Section III presents the methods and processes. Section IV and Section V contain the results and discussions. Finally, Section VI concludes the research.

II. RELATED WORK

A. Sentiment Analysis

Sentiment analysis is known as opinion mining where the purpose is to determine people's opinion towards certain entities such as an event, product, management, politician, and government issues [7]. Some studies on Malay sentiment analysis use different approaches to construct the Malay sentiment classifiers. Previous studies use the machine learning approaches for sentiment classification include machine learning [8-10], immune network [11], artificial immune network [12] and hybrid approaches [1, 6, 13].

*Corresponding Author

Sponsor by Ministry of High Education (MOHE) Malaysia, under the Fundamental Research Grant Scheme (FRGS/1/2018/ICT04/UITM/02/9).

TABLE I. MALAY SENTIMENT ANALYSIS

Author	Content of Article			
	Approach	Dataset	Politic	Ontology
[10]	Machine Learning	Newspaper	No	No
[11]	Machine Learning	Newspaper	No	No
[12]	Machine Learning	Newspaper	No	No
[8]	Machine Learning	Online Review	No	No
[9]	Machine Learning	Online Review	No	No
[2]	Lexicon based	Facebook Twitter	No	No
[1]	Hybrid	Online Review	No	No
[4]	Hybrid	Twitter	Yes	No
[13]	Hybrid	Article	No	Yes

Other than that, a lexicon-based approach is used to perform classification [2]. Although the machine learning approaches are used for Malay sentiment analysis, it does not mean this approach is the only one to perform Malay sentiment classification. Reference [13] proved that the hybrid approach gets the highest accuracy in the sentiment classification, which is better than previous studies.

Table I shows the Malay sentiment analysis in terms of approach, types of datasets, political domain, and application on ontology. Three major approaches are machine learning, lexicon-based, and hybrid approach used in Malay sentiment analysis.

From Table I, the study [13] improved the accuracy in sentiment analysis for the Malay language. It deals with informal language style and multilingualism that has become the norm of communication in social media. It used a hybrid approach, which is a combination of machine learning and a knowledge base. The ontology helps to get the more accurate sentiment with 94.34%. The polarities are positive, negative, neutral, and mixed. However, it is not in the political domain. Hence, it is a motivation for sentiment analysis on the political domain of other countries to find out the workflow in political sentiment classification.

B. Sentiment Analysis in Political Domain

Some approaches that classify the sentiment in the political domain are the machine learning approach, lexicon-based approach, ensemble approach, and multistage classification approach. Table II shows the sentiment analysis in the political domain in various countries use mostly tweets as the dataset.

Two studies [14-15] performed sentiment classification using the machine learning approach. The researchers use tweets as the dataset in Egypt and Indonesia. The lexicon-based approaches in [16-17] used online news and tweets that related to Indonesia and Turkey political context. The research [4, 18] performed sentiment classification using a hybrid approach. These two researches used tweets and corpus-based approach in Malaysia and India. Lastly, the research [19] used a multistage classification approach to get higher accuracy. The study was carried out in United States by using hybrid approach.

TABLE II. SENTIMENT ANALYSIS IN POLITICAL DOMAIN

Author	Content of Article			
	Approach	Dataset	Corpus	Country
[14]	Machine Learning	Tweets	No	Egypt
[15]	Machine Learning	Tweets	No	Indonesia
[16]	Lexicon based	Online News	Yes	Indonesia
[17]	Lexicon based	Tweets	Yes	Turkey
[4]	Hybrid	Tweets	Yes	Malaysia
[18]	Hybrid	Tweets	Yes	India
[19]	Hybrid	Tweets	No	United States

From Table II, the researchers [19] predicted the presidential election of the United States using Twitter sentiment analysis. The multistage classification approach classified the tweets of Donald Trump and Hillary Clinton. The accuracy of sentiment classification for Donald Trump and Hillary Clinton is 0.99% and 0.98%. From the reviews on the sentiment analysis in political domain, the studies [16, 19] become the anchors to construct a Malay sentiment classifier in this study.

C. Ontology

In [16], the ontology in a sentiment classifier helps to analyze social media content and gets an accurate sentiment. Ontology is a set of concepts related to entities, and the ontological hierarchy is constructed from the relations between concepts of entities.

There is still a lack of political ontology for the political domain in Malaysia. The construction of Malay political ontology (MPO) is adapted from [20]. Reference [20] constructed Australian politic ontology uses BBC politic ontology. There are four main concepts in BBC political ontology, which are person, place, organization, and event.

However, Australian politicians and parties are the main concepts in the ontology because it focuses on the election. These concepts with 53 instances for politicians, and 4 instances for parties. The Australian political structure is similar to Malaysian. Therefore, this study becomes another reference model to construct Malaysia Political Ontology (MPO).

D. Multistage Classification

The study [19] identified the winner from the election of the United States. The study proposed a multistage classification to classify the entity and sentiment of the tweets. The first stage of classification, the classifier called as entity classifier that classifies a general stream data into the respective entities. The classifier is trained with the entire dataset labelled by the entities. For the next stage of classification, the classifier called a sentiment classifier that classifies the sentiment of the tweets written refer to that particular candidate. Therefore, each candidate has a classifier associated with him or her. The classifier is trained with a dataset that pertaining to only its candidate.

Hybrid approach is used to perform the multistage classification [19]. The hybrid approach combines the machine learning with knowledge-based approach to improve the accuracy of sentiment analysis. The postings are classified into positive, negative or neutral. The result from [19] shows the accuracy of 94.34%, which is better than Naïve Bayes (NB), k-nearest neighbor (kNN) and Support Vector Machine (SVM).

III. METHODOLOGY

The dataset in this research is obtained from the Centre for Media and Information Warfare Studies (CMIWS) of Universiti Teknologi MARA (UiTM) Malaysia. There is a total of 1207 tweets in the political domain. There are six phases in this research, which are Preliminary Study, Data Pre-processing, Corpus Construction, Entity Construction, Multistage Classification Modelling, and lastly Evaluation.

A. Preliminary Study

In the preliminary study phase, literature reviews on articles to identify the research gap of sentiment analysis. It includes sentiment classification techniques, sentiment analysis in the Malay language, and political domain. Besides, it highlights the problem statement, research questions, objectives, scopes, and selected sentiment analysis techniques in the Malay language and political area.

B. Data Pre-Processing

This study focuses on Malay tweets. From the tweets collection, 752 Malay tweets are extracted. These tweets are the netizen comments related to political issues and selected politicians in Malaysia. The data pre-processing includes six processes to remove the noisy data. These processes begin with the removal of external links, symbols, and numbers. It continues with the lowercase conversion, abbreviation correction, stop words removal, word stemming, and then tokenization.

The external links, symbols, and numbers in a tweet are removed because these elements do not contain any meaning in classification. This process is similar to the previous studies [4, 15, 17-19]. Then, the remaining words are converted into lowercase to ease the word checking in the sentiment classification process. The abbreviations are converted into formal words as the abbreviations cannot be classified correctly [4].

Then, the process continues with removing the stop words as these words contain no meaning to analyze [14, 16]. The stemming process uses Fatimah stemmer [21] to get the seed words in the Malay language. The last process in the data pre-processing is word tokenization [4, 14] that split the words and store them in the database.

C. Corpus Construction

The corpus construction process from [22] is adapted in this research. It contains corpus filtering, web search, and filtering using linguistic patterns and domain-specific polarity lexicon. Our focus is building a Malay Political Corpus.

The corpus filtering contains two processes, which are word extraction and unique word selection. Firstly, 18 political words are extracted as the initial list. Next, the words that are

related to the election are selected. At the end of these processes, it produces a list of five political words that are related uniquely to the election.

After that, these election words are used together with a Linguistic pattern for web search to find more election-related words. There are two patterns in this searching process. In the first pattern, 'Pilihanraya' is used, which represents election in Malay. In the second pattern, it combines 'Pilihanraya' with 'seed words' related to the election.

- Pattern 1: Pilihanraya include(s)*.
- Pattern 2: Pilihanraya + "seed word".

These unique seed words include 'calon' (candidate), 'kempen' (campaign), 'manifesto' (manifesto), 'parti politik' (political party), and 'pengundi' (voter). Table III contains Malay seed words and sample lexicon related to election. There are 182 lexicons after this searching process.

TABLE III. LIST OF LEXICON

Seed Word	Sample of Lexicon	Total Lexicon
Calon	Wakil, Tanding, Kredibiliti, Khidmat, Pilih	65
Kempen	Strategi, Demokrasi, Lawan, Provokasi, Taktik	48
Manifesto	Subsidi, Janji, Sistematis, Tawar, Bersih, Hapus	15
Parti Politik	Agenda, Bangkang, Kuasa, Propaganda	36
Pengundi	Pangkah, Protes, Daftar, Undi, Sokong, Tolak	18

Then, the process continues with setting the polarity and score for these lexicons. The lexicon is classified into positive, negative, or neutral polarity. The positive word has a score of 1 to 5 based on the meaning of the lexicon. The negative word has a score of -5 to -1, while the neutral word gets 0. At the point, the political corpus is successfully constructed.

D. Entity Construction

In this process, the political parties are classified into government or opposition. In the entity construction, the generic ontology construction and instances enrichment from [20] are adapted.

A generic ontology has a set of concepts related to an entity. The relations between these concepts are organized into an ontological hierarchy. The ontology is constructed using Protégé 5.5.0 tool and using OntoGraf for visualizing the relationship in the ontology.

Four main concepts in the ontology include Person, Place, Organization, and Event. The person concept is the class of people in Malaysian political environments like voters and politicians. The place concept is related to the electoral areas such as state and constituency. The organization concept in Malaysian politics includes political parties, government, and council. The event concept relates to Election Day and campaign.

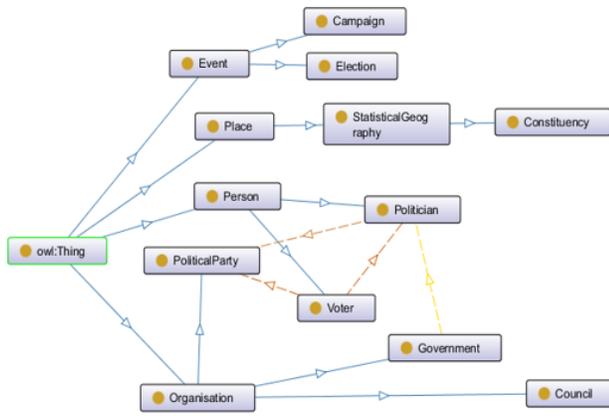


Fig. 1. Main Concept of Malaysian Politic Ontology.

This ontology is not sufficient to classify the entities in the tweets. Therefore, it needs to be enriched with instances before being fully utilized. The Parliament of Malaysia 2021 is referred to assign the instances for a political party and its candidates. We successfully construct the Malaysia Political Ontology (MPO) shown in Fig. 1.

The people and organization concept are used in this study to represent politicians and political parties. This ontology helps to classify the entity in the tweets. For example, with the instances of the political parties in Malaysia, the algorithm is developed to classify the entity into government or opposition. It helps to identify the sentiment towards the political parties in Malaysia. This information is valuable during the election.

E. Multistage Classification Modeling

The multistage classification from [19] is adapted in this model. There are two stages in multistage classification, which are entity classification and sentiment classification.

Entity classification is the classification based on the Malay Political Ontology (MPO). With MPO, the algorithm helps to classify the politicians and political parties into government or opposition. By this entity classification, it helps in the analysis of sentiment, especially during the pre-election. It speeds up the classification and sentiment analysis process based on the politicians and political parties.

TABLE IV. LIST OF ADAPTED FEATURES

Type	Description
F1	Presence of positive words
F2	Presence of negative words
F3	Presence of positive words in proportion to the presence of negative words
F4	Frequency of positive words in proportion to the frequency of negative words
F5	Weighted probabilities of a positive tweet
F6	Weighted probabilities of a negative tweet

Sentiment classification consists of the feature identification, feature extraction, training dataset, and testing dataset using Support Vector Machine (SVM) classifier. The trained dataset needs to be vectorized data. Therefore, it needs to vectorize data through feature extraction. There are twelve identified features in this research. Table IV shows six features adapted from [1].

From Table IV, the six features and descriptions are aimed to cater to the positive and negative words, proportion calculation, and weighted probabilities calculation.

There are six new features (see Table V) in this model that reflect the political domain in Malaysia. These features are used to cater the entity and words in political domain.

TABLE V. LIST OF NEW FEATURES

Type	Description
F7	Presence of domain specific positive words
F8	Presence of domain specific negative words
F9	Presence of domain specific positive words in proportion to the presence of domain specific negative words
F10	Frequency of domain specific positive words in proportion to the frequency of domain specific negative words
F11	Presence of a positive entity
F12	Presence of a negative entity

The feature extraction converts the words into vectors before training and testing. There are twelve sets of the formula, as stated in Table VI.

TABLE VI. FORMULA FOR FEATURES EXTRACTION

Type	Formula	Description
F1	$F1(T) = p$	T is referred to tweet; p is number of positive words in T
F2	$F2(T) = n$	n is number of negative words in T
F3	$F3(T) = p/n$	p is presence of positive words in T; n is presence of negative words in T
F4	$F4(T) = fp/fn$	fp is frequency of positive words in T; fn is frequency of negative words in T
F5	$F5(T) = p*(1-P+)*mp+$	p is number of positive words in T; P+ is probability of seeing positive words in a tweet; mp+ is manually assigned scores for positive score from lexicon
F6	$F6(T) = p*(1-N+)*mn$	n is number of negative words in T; N+ is probability of seeing negative words in a tweet; mn- is manually assigned scores for negative score from lexicon
F7	$F7(T) = dp$	dp is number of domain specific positive words in T
F8	$F8(T) = dn$	dn is number of domain specific positive words in T
F9	$F9(T) = dp/dn$	dp is presence of domain specific positive word in T; dn is presence of domain specific negative word in T
F10	$F10(T) = fdp/fdn$	fdp is frequency of domain specific positive word in T; fdn is frequency of domain specific negative word in T
F11	$F11(T) = pe$	pe is number of positive entities in T
F12	$F12(T) = pn$	pn is number of negative entities in T

After the extraction process, all data become vectorized data and assigned with sentiment polarity. The data is now ready to be trained and tested using MATLAB. 80% of vectorized data train using the Support Vector Machine classifier in MATLAB. The training dataset process runs eight times on different features for each technique in Support Vector Machine. This experiment aims to find the best features that can achieve high accuracy of sentiment.

After the experiments, the Linear Support Vector Machine classifier achieved high accuracy during the training process. A total of 20% of the remaining vectorized data test the Linear SVM classifier. The results are evaluated.

F. Evaluation

The result from the testing process was evaluated by the experts using the Delphi technique. Three experts in the political domain have cooperated in the evaluation phase. The experts are given the sample tweets to label the sentiment polarity separately in the first round of the evaluations. The anonymous responses are shared with the group after the first round. The experts are then allowed to adjust their answers in subsequent rounds. The final sentiments by the experts are collected to compare with the multistage sentiment classifier to measure the accuracy of the classifier.

IV. RESULTS

The main result for this study is the multistage sentiment classification model. It is a combination of entity and sentiment classification. At the first stage, the politicians and political parties are classified into government or opposition using Malaysia Political Ontology (MPO). This knowledge-based technique helps to identify the entity in the tweets. This stage is crucial in sentiment analysis, especially during the pre-election. The data analysts have to analyze the netizens' opinions

quickly to ensure election candidates decide on the pledge to win the election.

The MPO is constructed using the ontology concept and reflect the political entities in Malaysia, which the political parties are classified into government or opposition. With the enrichment of instances from the Parliament of Malaysia 2021, the MPO can be used to classify the political entities into government (positive), and opposition (negative).

At the second stage, the data from the political corpus are used for sentiment classification. This political corpus is specifically constructed based on the election-related words in the Malay language. It follows the processes of corpus filtering, web search, and filtering using linguistic patterns and domain-specific polarity lexicon. At the end of this stage, the lexicons are classified into positive, negative, or neutral polarity with setting of the score. In this entity classification and political corpus, twelve features are identified and extracted using specific formulas to get the vectorized data. Next, the polarity of the data is set before it is classified using SVM. At the end of the processes, the tweets are classified into positive, negative, and neutral.

Fig. 2 shows the multistage sentiment classification model for Malaysia Political Ontology. It illustrates the whole processes for the multistage sentiment classification. The political tweets need to be pre-processed accordingly to form a political corpus. At the same time, the entity classification uses the MPO to classify the polarity of the political parties. Then, it continues with the final sentiment classification using SVM.

The model is executed in a prototype to prove the concept. The experiments are also carried out using MATLAB. The same dataset is used to compare the results with the previous study.

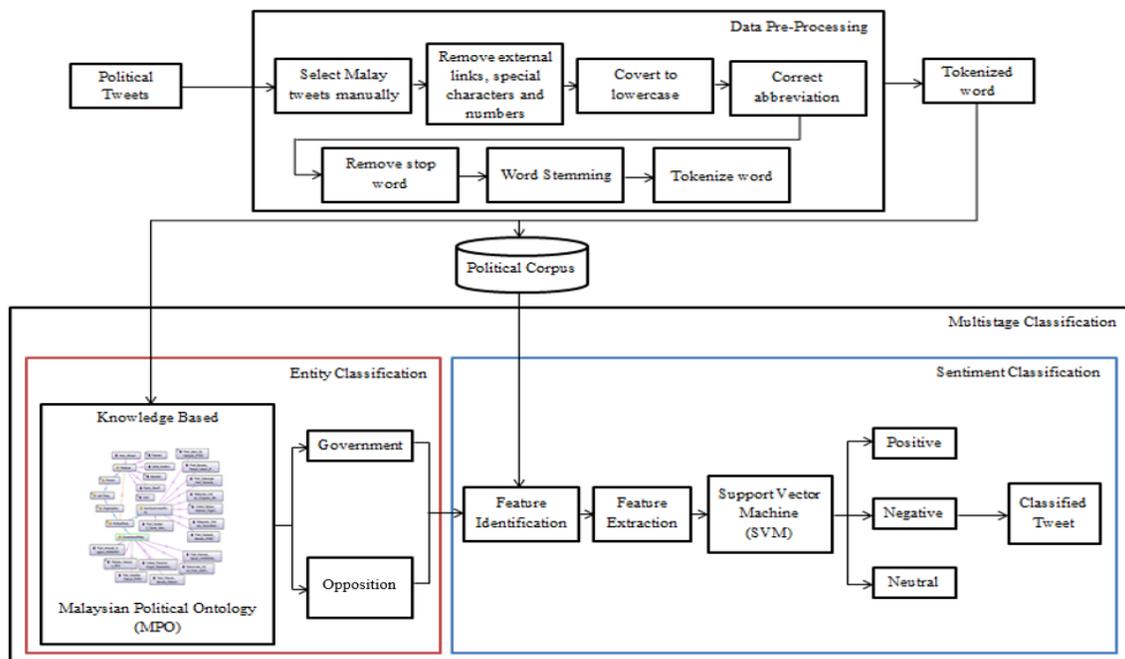


Fig. 2. Multistage Sentiment Classification Model using Malaysia Political Ontology and Malay Political Corpus.

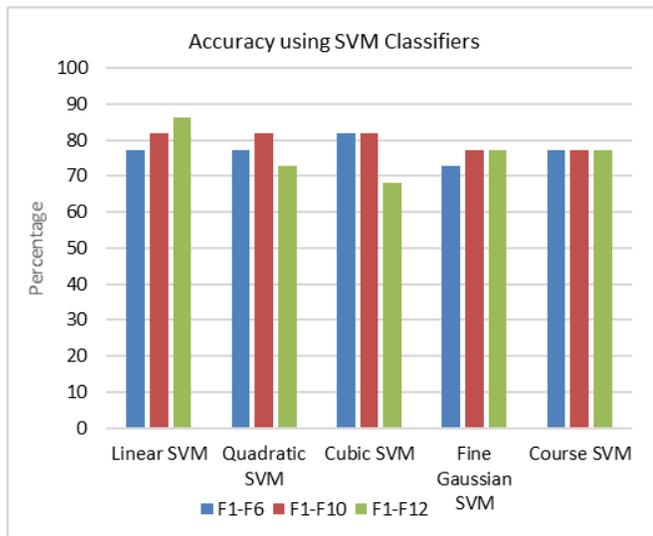


Fig. 3. Accuracy of Classification.

There are five experiments conducted to find the highest accuracy. The classifiers include Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, and Course Gaussian SVM (see Fig. 3).

From Fig. 3, F1-F6 are the features from the existing study. The best accuracy is 81.8% for the F1-F6. The additional features, F7-F12, have improved the accuracy to 86.4% for the classification. The evaluation results show that the 12 features improve the accuracy of the sentiment analysis.

The evaluation of the model is carried out using the Delphi technique. Then, the results between the political experts and the system are compared. We found that the expert results are similar to the results of the experiments.

V. DISCUSSION

There are few points to discuss from this study. There are available Malay corpora include abbreviation corpus [5] and hadith corpus [6]. However, these corpora are used for the specific domain and are not suitable for the politic. In the political domain, there is a relation between the verb and the candidate or current situation. To ease the classification process, therefore there is a need for specific corpus construction. This study fills the gap by constructing Malay Political Corpus. The terminologies related to the election are the main focus of this study.

The multistage sentiment classification model is successfully constructed in this study. This classification has two stages to classify tweets, which are entity classification and sentiment classification. At the first stage, the politician and party entities are classified as government or opposition using Malaysia Political Ontology (MPO). By classifying the entity, it sentiments the opinions of the netizen towards the politicians and political parties. In the second stage, the classified entity and the political corpus are used for further classification. All data are vectorized using twelve features and finally classifying the polarity of the tweets. The vectorized data are trained and tested using the Support Vector Machine technique. The result

shows that the accuracy is increased by 4.6% compared to the previous study.

However, a few important challenges need to be highlighted in the study. Firstly, the pre-processing process is not covered for negation handling. This negation handling is important to keep the meaning of words. For instance, the word 'tak menang' gives a meaning for 'not winning, with means 'kalah' or lost in English. Thus, the polarity of this word is negative. If there is no negation handling process, the word 'tak' will denote as negative and 'menang' will denote as positive, then the polarity of this word will be neutral. This process affects the accuracy of sentiment.

Secondly, the political corpus in this study is still lacking in lexicon and needs to enrich more political words in future studies. The Malaysia Political Ontology (MPO) only covers person and organization concepts of the political domain. The event and place concepts are out of scope in the study. Therefore, there is a need to covers four main concepts to improve the sentiment analysis in the political domain.

VI. CONCLUSION

Sentiment analysis tools help to analyze the social media comments into exploitable information for strategic planning in business, marketing, finance, entertainment and politic. The existing sentimental analysis classifiers manage to analyze different languages, but it has yet insufficiently in analyzing the Malay language accurately. Most of the comments that contain Malay words are classified as neutral. This shortage motivates this research to find a better solution for sentiment classification. This research sets the interest to investigate the sentiment classification in the political domain. It is one of the most popular areas due to the political scenario in Malaysia. From the literature, the existing Malay corpora such as abbreviation corpus and hadith corpus are not suitable for the political domain. In the political domain, there is a relation between the verb and the candidate or current politic situation. Therefore, this research constructs Malay Political Corpus, Malaysia Political Ontology, and proposes a multistage classification model. The experiments show the effect of twelve features on the performance of the sentiment classification. The multistage classification model proves that it improves the sentiment classification result by determining the entity in the political domain. The combination of entity classification and sentiment classification in the multistage classification can be a better solution in classifying the sentiments related to the political domain. Besides, the hybrid approach in multistage classification improves accuracy. With the political corpus and the linear SVM, the enhanced features selection increases the accuracy to 86.4%. In the future study, some improvements can be made to the pre-processing, enrich the list of words in the political corpus, and using a larger dataset.

ACKNOWLEDGMENT

The authors gratefully acknowledge the use of service and facilities of the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA. This study receives funding from the Ministry of High Education (MOHE)

Malaysia, under the Fundamental Research Grant Scheme (FRGS/1/2018/ICT04/UITM/02/9).

REFERENCES

- [1] A. Alsaffar and N. Omar, "Integrating a lexicon based approach and k nearest neighbour for MALAY sentiment analysis," *Journal of Computer Science*, vol. 11, no. 4, pp. 639–644, 2015.
- [2] M. H. Hijazi, L. Libin, R. Alfred, and F. Coenen, "Bias aware Lexicon-based sentiment analysis of Malay dialect on social media data: A study on the SABAH LANGUAGE," *2016 2nd International Conference on Science in Information Technology (ICSITech)*, 2016.
- [3] N. A. Abdullah and N. I. Rusli, "Multilingual sentiment analysis: A systematic literature review," *Pertanika Journal of Science and Technology*, vol. 29, no. 1, 2021.
- [4] N. I. Shaari, "Political Inclination Sentiment Analysis using Lexicon and Support Vector Machine Approaches for Malay Tweets," unpublished.
- [5] N. Omar, A. F. Hamsani, N. A. S. Abdullah, and S. Z. Z. Abidin, "Construction of Malay abbreviation corpus based on social media data," *Journal of Engineering and Applied Sciences*, vol. 12, no. 3, 468–474, 2017.
- [6] S. S. Sazali, N. A. Rahman, and Z. A. Bakar, "Characteristics of malay translated hadith corpus," *Journal of King Saud University - Computer and Information Sciences*, 2020.
- [7] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.
- [8] T. Al-Moslmi, S. Gaber, A. Al-Shabi, M. Albared, and N. Omar, "Feature selection methods effects on machine learning approaches in Malay sentiment analysis," In Proc. 1st International Conference on Innovation in Science and Technology, 2015, pp. 1-2.
- [9] A. Alsaffar, and N. Omar, "Study on feature selection and machine learning algorithms for Malay sentiment classification," In Proc. IEEE 6th International Conference on Information Technology and Multimedia, 2014, pp. 270-275.
- [10] H. Alshalabi, S. Tiun, N. Omar, and M. Albared, "Experiments on the use of feature selection and machine learning methods in automatic malay text categorization," *Procedia Technology*, vol. 11, pp. 748–754, 2013.
- [11] N. Isa, M. Puteh, and R. M. H. R. Kamarudin, "Sentiment classification of Malay newspaper using immune network," In Proc. of the World Congress on Engineering, 2013, vol. 3, pp. 3-5.
- [12] M. Puteh, N. Isa, S. Puteh, and N. A. Redzuan, "Sentiment mining of Malay newspaper (SAMNews) using artificial immune system," In Proc. of the World Congress on Engineering, 2013, vol. 3, pp. 1498-1503.
- [13] A. A. Sadanandan, A. Osman, H. Saifuddin, M. K. Ahamad, D. N. Pham, and H. Hoe, "Improving accuracy in sentiment analysis for Malay language," In Proc. 4th International Conference on Artificial Intelligence and Computer Science, 2016, pp. 28–29.
- [14] T. Elghazaly, A. Mahmoud, and H. A. Hefny, "Political sentiment analysis using twitter data," In Proc. of the International Conference on Internet of Things and Cloud Computing, 2016, pp.1-5.
- [15] H. T. Gemilang, A. Erwin, and K. I. Eng, "Indonesian President candidates 2014 sentiment analysis by using Twitter data," In Proc. IEEE 2014 International Conference on ICT for Smart Society, 2014, pp. 101–104.
- [16] M. A. F. Yatim, Y. Wardhana, A. Kamal, A. A. Soroinda, F. Rachim, and M. I. Wonggo, "A corpus-based lexicon building in Indonesian political context through Indonesian online news media," In Proc. IEEE 2016 International Conference on Advanced Computer Science and Information Systems, 2016, pp. 347–352.
- [17] E. Uysal, S. Yumusak, K. Oztoprak, and E. Dogdu, "Sentiment analysis for the social media: A case study for Turkish general elections," In Proc. of the South East Conference, 2017, pp. 215-218.
- [18] R. Jose, and V. S. Chooralil, "Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble approach," In Proc. IEEE 2016 International Conference on Data Mining and Advanced Computing, 2016, pp. 64–67.
- [19] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, "Election result prediction using Twitter sentiment analysis," In Proc. IEEE International conference on inventive computation technologies, 2016, vol. 1, pp. 1-5.
- [20] P. Wongthongtham and B. A. Salih, "Ontology-based approach for identifying the CREDIBILITY domain in social big data," *Journal of Organizational Computing and Electronic Commerce*, vol. 28, no. 4, pp. 354–377, 2018.
- [21] S. A. Fadzli, A. K. Norsalehen, I. A. Syarilla, H. Hasni, and M. S. S. Dhalila, "Simple rules malay stemmer," In Proc. International Conference on Informatics and Applications, 2012, pp. 28-35.
- [22] S. S.K.Rastogi, R. Singhal, and A. Kumar, "An improved sentiment classification using lexicon into svm," *International Journal of Computer Applications*, vol. 95, no. 1, pp. 37–42, 2014.