

# A Case Study on Social Media Analytics for Malaysia Budget

Ahmad Taufiq Mohamad, Nur Atiqah Sia Abdullah\*

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

**Abstract**—Malaysia citizen always looks forward to the budget announcement, which is presented by the government each year. Due to the direct effect on the economy, the citizens' opinions are crucial in understanding what they want and whether the budget satisfies them or not. Social media analytics can gather netizens' opinions on Twitter and conduct sentiment analysis. Most of the corpora in previous sentiment analysis research use English-based corpus. However, the current scenario of tweets in Malaysia uses a combination of English-Malay words. Therefore, this study uses a hybrid of the corpus-based and support vector machine approach. Semantic corpus-based combines the Malay and English words. Then, the domain-specific corpus on Malaysia Budget is constructed, which is budget corpus. Two separate analyses include category classification and sentiment analysis. Overall, most netizens have a positive sentiment about Malaysia's Budget with 56.28% of the tweets being positive sentiments. The majority of the netizens focus on social welfare and education that have the highest tweets. The discussion highlights the suggestion to improve the accuracy of this study.

**Keywords**—Malaysia budget; twitter; social media analytics; sentiment analysis; category classification; budget corpus

## I. INTRODUCTION

Social media platforms allow people to share content quickly, efficiently, and in real-time. There are numerous social networking services available to be utilized. Netizen intends to discuss current issues include politics, budget, products, and others using these social networking services. Budget is one of the important issues for a government. A government has annually presented the national budget plan for the next part of the year. The allocation and utilization of public funds in an efficient and prudent manner have always been a key concern in most of the budgets, including the Malaysia Budget. For this case study, the focus is on the Malaysia Budget 2020, which has been announced on 11 October 2019.

The budget consists of all aspects such as education, national defense, agriculture, transportation, and many more. This concern caused the government to reform its budget and finance to gain greater rationality and effectiveness in public financial management [1]. Therefore, the budget announcement is crucial for the ruling government.

Netizen shares their opinions through social media platforms about the budget before and after the budget presentation in Parliament. The government can use these

opinions as an evaluation of their budget allocation and citizens' satisfaction. Sentiment analysis can help in opinion mining. It makes use of natural language processing and text analytics to identify and quantify subjective information systematically. After the budget announcement, there is a need to conduct a social media analytic on these opinions to evaluate the peoples' sentiments towards the budget. It can be the guidelines for improving future budget analysis.

This paper aims to conduct a case study on social media analytics for the Malaysia Budget. The processes in the case study can be a guideline for future budget analysis. The paper continues with Section II that focuses on a literature review on social media analytics, sentiment analysis, and data visualization. Then, Section III shows the methodology for the study. It follows by the results and discussion in Section IV. It ends with the conclusion in Section V.

## II. RELATED WORK

### A. Twitter

Twitter has gained popularity among scholars, students, leaders, politicians, and the general public. It is one of the ideal public platforms for the rapid and comprehensive dissemination of political information and opinions [2], with an average of 330 million monthly active users and 500 million tweets. Retweeting is also a big part of Twitter, where people retweet or share other tweets with everyone else. The activity on Twitter involves the use of hashtags to aggregate tweets about the same subject.

### B. Social Media Analytics

The use of social media and the web creates a source of data that can be mined for new insights into how people communicate and behave, what they think and feel, and how they connect to each other [3]. Typically, retailers use social media in a few ways to promote their products. However, with people not only sharing content but also sharing their opinions, retailers or organizations find these opinions useful.

Gathering netizens' opinions about a topic or product or something in general, organizations can analyze their opinions and find ways to capitalize on them. If the netizens complain about a product of a rival company, organizations can use those opinions in creating a new product that will satisfy the netizens' needs. That is what social media analytics is all about; a practice of gathering data from social media websites and analyzing that data using social media analytical tools to make better decisions.

\*Corresponding Author.

### C. Data Preparation

Data preparation is crucial in social media analytics. It includes data scraping, data cleaning, data pre-processing, and stemming.

1) *Data scraping*: One of the ways to gather data from Twitter is through data scraping. The author in [4] describes scraping as getting the online data collection from social media and other websites in the form of unstructured data. Scraping has shown its capabilities in social media analytics, allowing new ways to collect and analyze social data [5].

2) *Data cleaning*: This process removes repeated data that are unrelated to the topic, removing typographical errors [4]. If there is incorrect or inconsistent data, it can lead to false conclusions, thus misdirecting the solutions. For example, this study analyzes the netizens' opinions if the data are incorrect, and the result could show netizens agrees with the budget allocation. However, the actual result could be the other way around.

3) *Data pre-processing*: Multiple steps of pre-processing [6] include removing stop words and stemming. However, it depends on the kind of analysis and expected output. Stop words are the most frequent words like articles (a, an, the), auxiliary verbs (be, am, is, are), prepositions (in, on, of, at), conjunctions (and, or, nor, when, while) that does not provide any information to the analysis. Therefore, stop words can be removed [7].

4) *Stemming*: In pre-processing, stemming is defined as coding multiple forms of a linguistic object into a 'rudimentary' shape with the same meaning [8] or obtaining the root, the stem of derived words [9]. Another part of data pre-processing is the categorization of data. It aims at classifying documents into a variety of pre-defined categories [10]. It is a process of assigning tags or categories to text according to its contents.

### D. Sentiment Analysis

Sentiment analysis is evolving rapidly as an automated linguistic relation and context review process [11] that involves a process of extracting attitudes, emotions, and feelings [4]. Social sentiment indicates how a person states his opinion and attitude towards an object [12]. The strength of sentiment or opinion is associated with the intensity of some emotions [13] such as joy and anger.

The evaluations can be based on consumer behavior research: rational evaluations and emotional evaluations. Rational evaluations are tangible beliefs and utilitarian attitudes, for example, "This house is worth the price". While emotional evaluations are based on non-tangible and emotional responses to events that go deep into the state of mind of individuals. For example, "This is the best house in the neighborhood". After evaluation, the polarity of the statements needs to be defined. Classification of polarity can be binary, ternary, or ordinal [11] depending on the aim of the sentiment analysis.

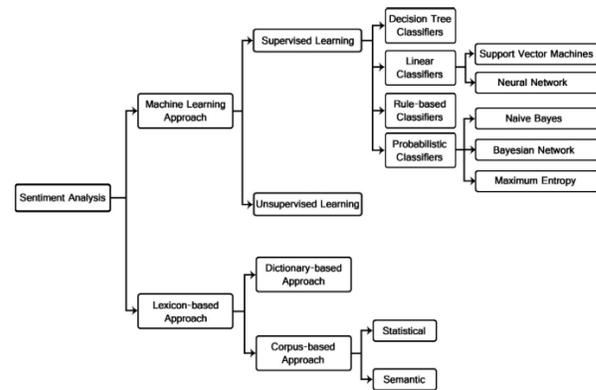


Fig. 1. Sentiment Analysis Techniques [15].

Fig. 1 shows the multiple techniques used for sentiment analysis. Sentiment detection approaches can be divided into the lexicon-based approach and the machine-learning approach [14]. The lexicon-based approach is divided into dictionary and corpus approaches. The machine learning approach (ML) uses popular algorithms such as Support Vector Machine (SVM) and Naïve Bayes (NB), which use linguistic features. ML approaches for sentiment analysis can be unsupervised or supervised machine learning.

Multiple articles on sentiment analysis are studied to gain knowledge on the performance of the classifiers mentioned. Table I shows numerous papers on sentiment analysis. The purpose of this comparison is to find the classifier that performs the best among others. From the review, SVM is the best as it performs better than other classifiers in terms of sentiment classification. There is a Malay Opinion Corpus [16] used as a data source. It is similar to this study, where the extracted tweets are Malay words with a combination of English words.

After a literature review of the sentiment analysis techniques, this study uses a hybrid of semantic corpus-based and machine learning SVM. The semantic corpus-based was chosen because it gives the sentiment values directly and suitable for domain/context-specific data. The SVM was chosen as the classifier as it is the best among others.

### E. Data Visualization

Data visualization can express data in a visual form that finds blind spots [27], which helps users acquire knowledge about the data. Data can be observed from different perspectives and used more in-depth observation and analysis. Because there are different degrees of data, the zoom feature should be implemented in data visualization [28].

Many researchers agree that data visualization can improve decision-making [29-31]. It helps an organization to view where they are and the process carried by an organization. An organization views and analyses the visualized data and can identify the problems for adjustments. Therefore, the organizations improve their decision-making through systematic data analysis to make changes to their process flow.

TABLE I. COMPARISON OF SENTIMENT ANALYSIS TECHNIQUES

Study	Technique	Feature Selection	Data Source/ Domain	Performance
[16]	SVM, Naïve Bayes, KNN	Information Gain (IG), Gini index, Chi-squared	Malay Opinion Corpus	IG-based performed best, SVM performed best (85.33%)
[17]	Naïve Bayes, KNN, and Decision Tree	Word tokens	Roman-Urdu opinions	Naïve Bayes highest accuracy (97.50%)
[18]	SVM, Naïve Bayes, Rocchio, C4.5, KNN	Word tokens	Reuters-21578, Ohsumed	SVM outperforms other classifiers
[19]	SVM, Artificial NN	Words, TF-IDF, IG	Amazon.com	ANN produce better or at least comparable results to SVM
[20]	Convolutional NN	Uses filters to produce features	Movie reviews, SST-1 SST-2, Subj, TREC, CR, MPQA	CNN achieve the highest accuracy for 4 out of 7 tests
[21]	Rule-based	Lexical features	Social media text	Performed as well as other techniques
[22]	Hybrid of Rule-based with supervised learning	N-gram, POS Tags, @USER, Hashtag, URL, Discourse	Twitter	Hybrid improve SVM Predictions
[23]	Bayesian Network	Chi-square	Twitter	Precision: 55.57%
[24]	KNN, Random Forest, Naïve Bayes, Bayesian Network	Words	Twitter	KNN highest accuracy for with stop words (99.65%) and without stop words (96.64%)
[25]	Maximum Entropy	Words	Chinese online product reviews	Accuracy: 80.87%
[26]	Unsupervised learning	POS tagging	Korean language	F1 score: 71.8%

#### F. Types of Data Visualization

There are five data visualization categories, namely temporal, network, geospatial, hierarchical, and multidimensional.

1) *Temporal visualization*: It is one of the simplest and quickest ways to represent important time-series data. Temporary datasets usually include location and time datasets. Sometimes, these datasets may contain different characteristics [32], depending on the data sources. Temporal data have items that have a start and finish time with possibilities of data overlapping with each other.

2) *Network visualization*: A network dataset comprises an arrangement of a set of known connections among entities [33]. Network visualization shows complex relationships between several elements. A network visualization displays undirected and directed graph structures. This kind of visualization sheds light on the relationships between entities. Round nodes represent entities, while lines indicate their relationships. The vivid display of network nodes can reveal non-trivial data discrepancies.

3) *Geospatial visualization*: techniques supporting the analysis of geospatial data using interactive visualization. One of the earliest forms of information visualization is geospatial visualization. There exist a substantial number of applications these days, in which it is crucial to analyze relationships that include geographic location [34]. Geospatial visualization takes place in several real-world situations such as wildland fire fighting, forestry, archaeology, environmental studies, and urban planning that call for decision-making and processes for information formation.

4) *Hierarchical visualization*: Hierarchical visualization is suitable for numerous data types that are automatically hierarchical or ideal for a recursive grouping [35]. Hierarchical data are organized in a tree structure in which each data element identifies a node in the tree. At the same time, each

node can have child nodes. Hierarchical data visualization allows the user to drill down through multiple levels.

5) *Multidimensional visualization*: It manages datasets with several variables to correspond to the visual structure of one-dimension, two-dimension, or higher dimensions [36]. Usually, this technique can represent data depending on one or two variables [37].

### III. METHODOLOGY

There are several phases in the methodology include business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The data preparation phase covers all activities to construct the final dataset from the initial raw data. After the data extraction, the data preparation phase focuses on cleaning the data before the modeling phase. The result of this phase produces a new dataset for the modeling phase.

#### A. Data Collection

RapidMiner software is used for data collection as it is easy to use and without coding. Using the “search Twitter” node in RapidMiner, an access token is needed from Twitter to gain access for extraction. After getting the access token, it sets the keywords for extracting tweets. The related keywords are Budget2020 and “Belanjawan 2020”, which obtain through Google Trends. The process of data extraction continues in 22 days (2 to 23 October 2019). The extraction happens at around 11.30 pm every single day and saves in separate excel files. At the end of data extraction, there are a collection of 9,638 tweets with 12 columns.

#### B. Data Cleaning

Dataset needs to be cleaned from unwanted data such as repetition of data and not related to the topic. Each data is scanned to remove noise, which is unrelated data. For the repetitive data, using the excel function of removing duplicates proves to be enough for removing repetitive data, keeping only the unique texts.

### C. Data Classification

As the budget covers various topics, netizens' tweets also diversify according to the categories in the budget. Therefore, there are nine budget categories with different keywords. The tweets are processed according to the words in the nine categories. Table II shows the sample of related terminologies for nine categories. The nine categories are selected because they are the most talked-about topic by netizens. After analyzing the tweets, the keywords used by the netizens are captured.

After thoroughly going through all the tweets and identifying the keywords, a corpus is created for category classification. The corpus is saved as a CSV file. Each row contains the words and the category. The related words are as follows: 26 words for agriculture, 151 words for the economy, 46 words for education, 41 words for general, 22 words for health, 45 words for others, 19 words for public services, 52 words for social welfare, and 36 related words for transportation. There are 438 words in the budget corpus.

Besides categorizing the words in the budget, the polarity of each tweet is assigned. There are three polarities in this study, positive, negative, and neutral. There are different words according to the polarity. Due to the multilingual comments by the netizen, the words are in Malay and English. Table III shows the related words for the three different polarities. The polarity represents the emotion and feeling of the netizen towards the budget categories. Positive words represent happiness and satisfaction in the budget category. The negative words show the dissatisfaction with the budget.

The processes of classification for the polarity are the same as the category classification. All the sentiment words are used to train the model to identify the sentiment of each tweet. After the classification processes, there are 114 words for the positive polarity, 90 words for neutral polarity, and 92 words for negative polarity. The corpus is used to train the model to classify the sentiment for each tweet.

### D. Modeling

For this modeling phase, there are a few steps to be taken. The first step is selecting a suitable model. For this study, the model focused on sentiment analysis. There are two approaches for sentiment analysis, which are machine learning and lexicon-based models. Machine learning models belong to supervised classification. Two sets of documents, training, and testing, are needed for classification purposes. The training set is used in an automatic classifier to differentiate the characteristics of tweets. The testing set is used to check the performance of the classifier.

For lexicon-based models, it employs dictionaries of words annotated with their semantic polarity and sentiment strength. It uses a corpus to help in the sentiment classification process. Then, the corpus is used to calculate a score for the polarity of the document or dataset. After comparing the techniques for sentiment analysis models, a hybrid of corpus-based and Support Vector Machine (SVM) is selected as the technique in this study. Fig. 2 shows the designed process of sentiment analysis for this study.

TABLE II. WORDS IN SELECTED CATEGORIES

Category	Related Words
Agriculture	Pertanian, sawit, getah, padi, peladang, pesawah, farmer
Economy	M40, B40, gst, tax, cukai, pekerjaan, gaji, ekonomi, epf
Education	Pendidikan, education, TVET, pendidikan tertiary, biasiswa
General	Menteri, Belanjawan, Budget, government, kerajaan, Malaysia
Health	Perubatan, patients, vaccines, medical, kesihatan, disease
Public Services	MINDEF, polis, tentera, kebombaam, pertahanan, TLDM, defence
Social Welfare	Bumiputra, disabled, OKU, wanita, children, orang asli
Transportation	Pengangkutan awam, tol, kereta, minyak, BSH, fuel, Petrol
Others	Persiaran, hiburan, seni, pelancongan, budaya, belia, sukan

TABLE III. SAMPLE WORDS IN THREE POLARITIES

Polarity	Related Words
Positive	best, sokong, satisfying, love, agree, baik, happy, excited, glad, uplifting
Negative	#pakatanhancing, malap, garbage, rugi, bodoh, worse, keburukan
Neutral	not sure, entah, unfortunately, why, but, cadangan, diharapkan

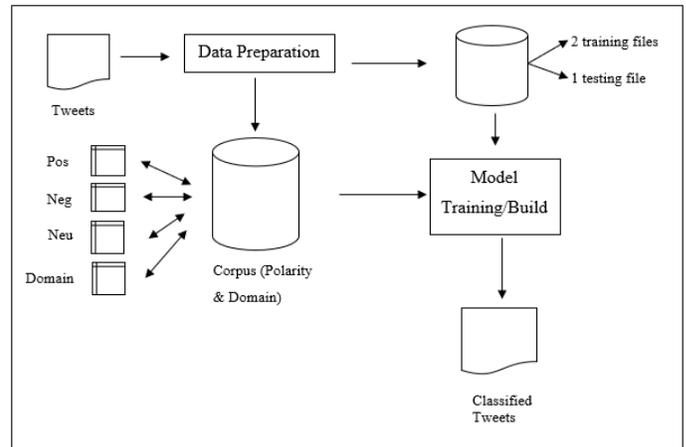


Fig. 2. Sentiment Analysis Techniques [15].

After pre-processing and the preparation of the data, the results of the classifications are set into a corpus. The corpus is used for modeling, where the extracted dataset is trained by two training files and tested using a single file using Support Vector Machine (SVM). The two training files are files containing the category keywords and sentiment keywords. The result of the testing is then compared with the benchmark sentiment dataset to check the accuracy.

The result of both analyses is visualized in a dashboard form. The visualization is presented using PowerBI software based on a dashboard approach where multiple charts are in a single view. As the visualizations have different techniques, multidimensional visualization is more suitable for this study due to the data types. Table IV shows the types of visualizations and data representations for the dashboard.

TABLE IV. DASHBOARD VISUALIZATION

Chart	Data Representation
Pie chart	Percentage and number of polarity
Column chart	Total of tweets in each category
Column chart	Total of tweets in each category with sentiment

The pie chart is used to show the percentage and amount of polarity. The column chart is used to show the total tweets in each category, and sentiments. As SVM has many parameters to experiment with, this study finds the best parameter that produces the best score. GridSearch function is used to find the best value for each parameter. To summarize, the best parameters for this SVM model are:

- Best C: 1000.
- Best Kernel: rbf.
- Best Gamma: 0.001.

These parameters are used for the category and sentiment classification as training files used in the GridSearch function yield the same result.

#### IV. RESULTS AND DISCUSSION

This study explains the processes of gathering and creating a budget corpus for analysis. A hybrid of SVM and corpus-based approaches is proposed for category classification and sentiment analysis. The modeling uses two training files and one testing file for both analyses. Based on the results, social welfare and education are the most popular category in Malaysia Budget 2020 (see Table V).

The netizen talked about the welfare of disabled people, natives, senior citizens, and people from different religions and races. On the topic of education, netizen comment about education issues ranging from kindergarten to university. The other categories in the tweets are agriculture, economy, health, and others. Table V shows the actual number of tweets in each category in the budget.

For sentiment classification, the result of the model training using the parameters achieved a score of 46.59%. The result of the testing model achieved an accuracy of 31.04%. Table VI shows the classification report of the sentiment classification.

Based on Table VI, the accuracy for the sentiment classification is 31.04%. Among the sentiments, neutral sentiment has the highest accuracy of 39.30%, f1-score of 32.54%, precision of 75.37% but the lowest recall score of 20.75%. For the negative sentiment, the model achieved 28.57% accuracy, 28.30% for f1-score, 19.79% in precision, and 49.60% for recall. Lastly, for positive sentiment, it achieved an accuracy of 27.53%, f1-score of 30.31%, precision of 19.48%, and recall of 68.21%.

Fig. 3 shows the noticeable gap between the actual and result of sentiment classification.

There are 843 actual positive tweets, but 2,951 tweets are predicted positive. For the negative sentiment, it is about 502 tweets, but 1,258 tweets are predicted negative. Lastly, neutral has the most sentiment tweets (n=3,952), but the model only

predicted 1,088 neutral tweets. The difference between the classification results is because of the lacking of words in the budget corpus.

With category and sentiment classification, the result is visualized in a dashboard. It helps the analysts in making a more in-depth critical analysis of the study.

Fig. 4 shows a dashboard for the social media analytics in Malaysia Budget 2020. There are three charts in the dashboard. A pie chart displays the tweets based on the sentiment. The column charts illustrate the tweets by time and category. A slicer function is also added to the dashboard to filter the visualization based on a specific category.

The dashboard is evaluated using convenience sampling. From the survey, the dashboard is understandable for the users. The majority of the users agreed that the dashboard is simple and useful for quick insight into the budget.

TABLE V. ACCURACY OF CATEGORY CLASSIFICATION

Category	Actual Number	Result
economy	856	746
education	261	3263
transportation	129	0
health	63	111
agriculture	89	1521
public services	532	0
social welfare	138	3284
general	1361	0
others	2206	94

TABLE VI. REPORT OF SENTIMENT CLASSIFICATION

Item	Accuracy (%)	f1-score (%)	Precision (%)	Recall (%)
negative	28.57	28.30	19.79	49.60
neutral	39.90	32.54	75.37	20.75
positive	27.53	30.31	19.48	68.21
accuracy		31.04	31.04	31.04
macro avg		30.38	38.22	46.19
weighted avg		31.78	61.21	31.04

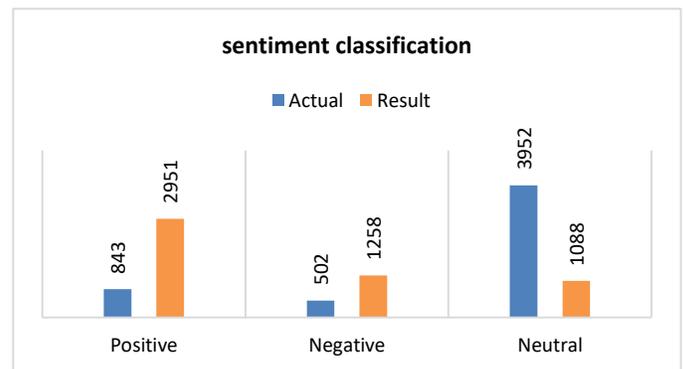


Fig. 3. Bar Chart of Sentiment Classification.

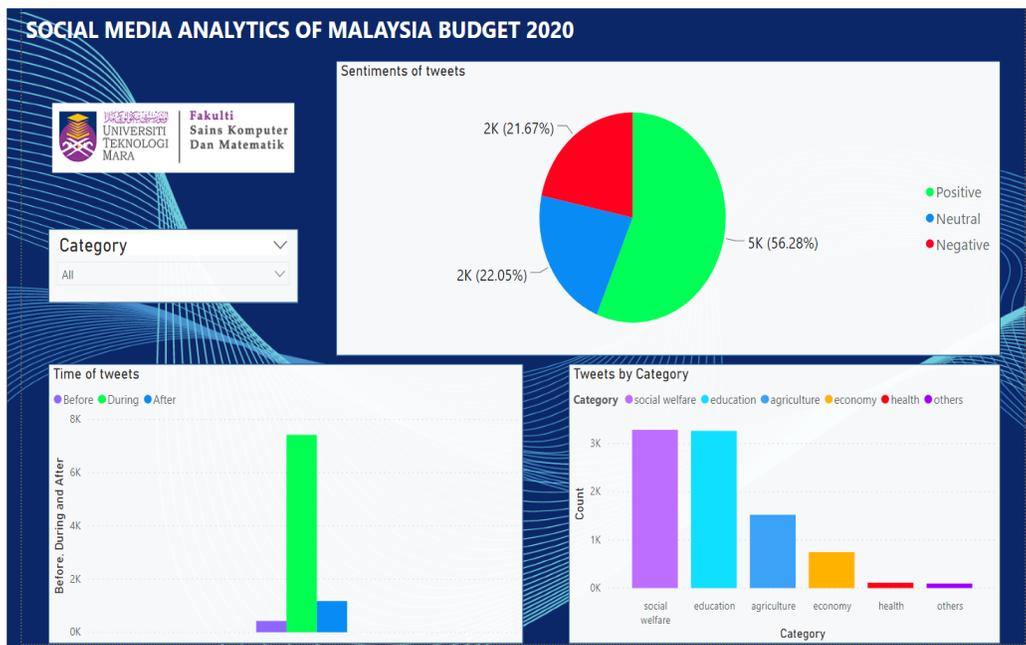


Fig. 4. Social Media Analytics Dashboard for Malaysia Budget 2020.

With only 5,297 tweets used out of 9,638 raw tweets, the data collection for the analysis is insufficient. As the study focuses on Twitter for data collection, perhaps using other social media platforms like Facebook can increase the data collection related to the budget. Thus, it will lead to higher accuracy of category and sentiment classification of tweets. Furthermore, the study only explored a narrow range of possible parameters values. A recommendation is to explore the SVM parameters such as kernels, shrinking, tolerance for stopping criterion, and class weight. Moreover, a more complex model for analysis would be an opportunity to be explored.

## V. CONCLUSION

The budget presentation has a direct impact on the economy of the country. Therefore, the citizens' opinions are crucial in understanding the actual needs and their satisfaction. Social media analytics can process opinions with sentiment analysis. This study chose a corpus-based approach to extract the Malay and English words that focus on Malaysia's Budget. A hybrid of SVM and corpus-based approaches is used for category classification and sentiment analysis. Overall, the netizens are positive about Malaysia's Budget with 56.28% of the overall tweets. The netizens are more concerned about the social welfare and education aspect of the budget as both categories have the highest tweets. Further exploration of the SVM parameters and complex model for analysis is the potential area to be studied.

## ACKNOWLEDGMENT

The authors express their gratitude to the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia for supporting this study.

## REFERENCES

- [1] N. A. Siddiquee, "Managing for results: Lessons from public management reform in Malaysia," *Int. J. Public Sect. Manag.*, vol. 23, no. 1, pp. 38–53, 2010.
- [2] S. Stieglitz, L. Dang-Xuan, A. Bruns, and C. Neuberger, "Socialmedia analytics," *Bus. Inf. Syst. Eng.*, vol. 6, no. 2, pp. 89–96, 2014.
- [3] G. Moss, H. Kennedy, S. Moshonas, and C. Birchall, "Knowing your publics: The use of social media analytics in local government," *Inf. Polity*, vol. 20, no. 4, pp. 287–298, 2015.
- [4] B. Batrinca and P. C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms," *AI Soc.*, vol. 30, no. 1, pp. 89–116, 2014.
- [5] N. Marres and E. Weltevrede, "SCRAPING THE SOCIAL?: Issues in live social research," *J. Cult. Econ.*, vol. 6, no. 3, pp. 313–335, 2013.
- [6] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6266–6282, 2013.
- [7] A. Fahrmi and M. Klenner, "University of Zurich Old wine or warm beer : target-specific sentiment analysis of adjectives Old Wine or Warm Beer : Target-Specific Sentiment Analysis of Adjectives," no. April, pp. 60–63, 2008.
- [8] Z. Xiang, Z. Schwartz, J. H. Gerdes, and M. Uysal, "What can big data and text analytics tell us about hotel guest experience and satisfaction?," *Int. J. Hosp. Manag.*, vol. 44, pp. 120–130, 2015.
- [9] N. Zainuddin and A. Selamat, "Sentiment analysis using Support Vector Machine," *I4CT 2014 - 1st Int. Conf. Comput. Commun. Control Technol. Proc.*, no. I4ct, pp. 333–337, 2014.
- [10] Y. Ko and J. Seo, "Automatic text categorization by unsupervised learning," 2000, pp. 453–459.
- [11] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, 2019.
- [12] Y. Mejova, "Sentiment Analysis : An Overview Comprehensive Exam Paper," *Science (80-. )*, pp. 1–34, 2009.
- [13] Y. Lin, X. Wang, and A. Zhou, "Opinion spam detection," *Opin. Anal. Online Rev.*, no. May, pp. 79–94, 2016.
- [14] D. Maynard and A. Funk, "Automatic detection of political opinions in tweets," *CEUR Workshop Proc.*, vol. 718, pp. 81–92, 2011.

- [15] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [16] 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," *1st ICRIL-International Conf. Innov. Sci. Technol. (IICIST 2015)*, no. April, pp. 2–5, 2015.
- [17] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 28, no. 3, pp. 330–344, 2016.
- [18] T. Joachims, "Text Categorization with SVM: Learning with Many Relevant Features," pp. 2–7.
- [19] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, 2013.
- [20] Y. Kim, "Convolutional neural networks for sentence classification," *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1746–1751, 2014.
- [21] J. Wilson and C. Hernández-Hall, "Physics laboratory experiments," *Eighth Int. AAAI Conf. Weblogs Soc. Media*, p. 18, 2014.
- [22] P. Chikersal, S. Poria, and E. Cambria, "SeNTU: Sentiment Analysis of Tweets by Combining a Rule-based Classifier with Supervised Learning," no. *SemEval*, pp. 647–651, 2015.
- [23] M. S. Asriadi, M. S. Mubarak, and Adiwijaya, "Classifying emotion in Twitter using Bayesian network," *J. Phys. Conf. Ser.*, vol. 971, no. 1, pp. 0–14, 2018.
- [24] A. P. Jain and V. D. Katkar, "Sentiments analysis of Twitter data using data mining," *Proc. - IEEE Int. Conf. Inf. Process. ICIIP 2015*, pp. 807–810, 2016.
- [25] H. Wu, J. Li, and J. Xie, "Maximum entropy-based sentiment analysis of online product reviews in Chinese," no. *May*, pp. 559–562, 2017.
- [26] Y. Ko and J. Seo, "Automatic text categorization by unsupervised learning," pp. 453–459, 2000.
- [27] Q. Fu, W. Liu, T. Xue, H. Gu, S. Zhang, and C. Wang, "a Big Data Processing Methods for Visualization," pp. 571–575, 2013.
- [28] F. Bao and J. Chen, "Visual framework for big data in d3.js," *Proc. - 2014 IEEE Work. Electron. Comput. Appl. IWCA 2014*, pp. 47–50, 2014.
- [29] Y. Zheng, W. Wu, Y. Chen, H. Qu, and L. M. Ni, "Visual Analytics in Urban Computing: An Overview," *IEEE Trans. Big Data*, vol. 2, no. 3, pp. 276–296, 2016.
- [30] H. Yan, J. Wang, and C. Xia, "Research and Application of the Test Data Visualization," *Proc. - 2017 IEEE 2nd Int. Conf. Data Sci. Cyberspace, DSC 2017*, pp. 661–665, 2017.
- [31] A. Lodde, "Network Visualisation," *Network*, 2009.
- [32] A. Shrestha, "Visualizing Spatio-Temporal data," 2014.
- [33] J. Madadhain, D. Fisher, P. Smyth, S. White, and Y. Boey, "Analysis and visualization of network data using JUNG," *J. Stat. Softw.*, vol. 10, no. 2, pp. 1–35, 2005.
- [34] D. A. Keim, "Information Visualization: Scope, Techniques and Opportunities for Geovisualization," in *Exploring Geovisualization*, 2005, pp. 21–52.
- [35] P. Neumann, S. Schlechtweg, and S. Carpendale, "Arctrees: Visualizing Relations in Hierarchical Data," *Proceeding EUROVIS'05 Proc. Seventh Jt. Eurographics / IEEE VGTC Conf. Vis.*, pp. 53–60, 2005.
- [36] R. M. Pillat, E. R. A. Valiati, and C. M. D. S. Freitas, "Experimental study on evaluation of multidimensional information visualization techniques," *ACM Int. Conf. Proceeding Ser.*, vol. 124, no. April 2014, pp. 20–30, 2005.
- [37] H. Pohlheim, "Visualization of Evolutionary Algorithms - Set of Standard Techniques and Multidimensional Visualization," *GECCO'99 - Proc. Genet. Evol. Comput. Conf. San Fr. CA*, pp. 533–540, 1999.