Sentiment Analysis of Malaysian Insurance Companies (SAMIC): A Visualization using Support Vector Machine Algorithm

Nur Farhana Binti Ahmad, Ts. Dr. Khyrina Airin Fariza Binti Abu Samah

Highlights

- This paper proposed the implementation of a web application to visualize Malaysia's best insurance companies to assist policyholders in understanding the public views of the policy offered by the top insurance companies in Malaysia, which are AIA, Prudential, and Great Eastern.
- This research aimed to design and develop a visualization application through Twitter sentiment analysis using the Support Vector Machine algorithm (SVM).
- Testing phases have shown that the classifier successfully classified tweets' sentiment with 90% accuracy, with every feature in the application functions properly and resulting in a 92.5% usability score.

Article Info

Received: 14/09/2021
Accepted: -

Keywords

Support Vector Machine algorithm, Web-based visualization, Twitter sentiment analysis.

Abstract

This study proposed the implementation of a web application to visualize Malaysia's best insurance companies to assist policyholders in understanding the public views of the policy offered by the top insurance companies in Malaysia, which are AIA, Prudential, and Great Eastern. The complexities of insurance purchasing issues include assessing financial needs and choosing an insurance policy from which established companies became confusing and challenging for potential policyholders as they will be entering into a long-term investment. They need to allocate adequate time to review each insurance company's offer to make a wise decision. Hence, this research aimed to design and develop a visualization application through Twitter sentiment analysis using the Support Vector Machine algorithm (SVM) as the classification model. The application acts as a medium to visualize tweets' sentiment analysis results that mentioned these insurance companies. Twitter was used as a source of data in this study. The tweets extracted using dates and keywords were analyzed as it is one of the metrics that will advance insurance companies' online presence. The classification model has been trained with over 1.2 million data and incorporated into the application, allowing the user to use the model on any English and Malay textual data reviews. Additionally, this application also includes real-time monitoring of tweets sentiment that uses the text pre-processing infrastructure and sentiment analyzer to evaluate public sentiment changes in response to user search keywords. Testing phases have shown that the classifier successfully classified tweets' sentiment with 90% accuracy, with every feature in the application functions properly and resulting in a 92.5% usability score.

1. INTRODUCTION

The insurance industry, including Islamic insurance, is a financial institution that offers risk management services, induces liquidity, diversifies financial losses, and promotes investment in the economy. According to the Malaysia Financial Sector Assessment Program [1] report, Malaysia's insurance sector accounted for about 6% of financial sector assets, generally divided into life insurance and general insurance. In general, insurance companies can be categorized as health and medical insurance, auto insurance, accidental insurance, life insurance, critical illness insurance, fire and home insurance, and other services. Many insurance companies are operating in Malaysia, offering a broad range of products and services tailored to Malaysians' needs from all walks of life. A total of 44 companies have been listed as the Licensed Insurance Companies and Takaful Operators list [2]. Choosing which insurance company to invest in can be challenging as the client will be entering into a long-term investment.

The survey conducted by YouGov in April 2019 finds that 45% of social media users in Malaysia use it as a platform to research services, review, or find new products to buy. Through online communities like the one that exists on Twitter, reviewers can directly impact customer decision-making. Therefore, insurance companies should value how social media can influence clients' perception that affects their purchasing intention. Companies that adopt a strategic approach to social media use will benefit from those who do not use social media [3]. Thus, this study aims to visualize the performance of leading insurance companies in Malaysia through Twitter Sentiment Analysis using the Support Vector Machine (SVM) algorithm. The visualized results could be used in maximizing customers' satisfaction and ensuring retention. In this way, insurers will be able to proactively target new potential markets and resolve customer issues more effectively.

*Correspondingauthor, e-mail:mbilen@mehmetakif.edu.tr

2. RELATED WORKS

This section explores an overview of Malaysia's insurance industry and the methods used to classified Twitter data.

2.1. Overview of the Insurance Industry in Malaysia

The insurance companies in Malaysia are under the supervision of Bank Negara Malaysia (BNM). BNM points out that Malaysia's insurance industry has undergone one of its most essential transformations since the Malaysian insurance regulatory system overhauled in the early 1990s [4]. The insurance industry was struggling with intense competition. Consequently, it has become vital for insurance companies, whether local or international, to be competitive and profitable in this field.

Any risk that can be quantified can potentially be insured. An insurance policy will specify in detail which of the policy covers risks. The insurance companies offered few insurance types, and each plan has a different package or premium. Nevertheless, the plan chosen depends on the policyholders' situation, age, lifestyle, and the type of insurance policy needed, such as retirement, health, or education purposes [5].

Life insurance, general insurance, and medical insurance are common types of insurance businesses in Malaysia. Life insurance helps to cover risks such as early death, sickness, and permanent injury. It pays out the insured or beneficiary a certain amount of money as compensation if anything unfortunate happens to the insured. General insurance comprises householder insurance, motor insurance, and personal accident insurance [6]. In the context of householder insurance, the policyholders will receive money if a natural disaster, fire destroy their personal property or house, and other cases included in the insurance plan and the protection of policyholders against claims made by third parties in certain related events [7]. If the policyholders' motor vehicle is involved in an incident, they may be claimed for damage or loss by means of motor insurance. According to the insurance plan purchased, personal accident insurance claims were obtainable when the policyholder became disabled due to an accident or passed away. Having health insurance is necessary to receive timely medical care and improve health and lives [8]. Besides, if the policyholder is admitted to the hospital, the insurance company will pay medical expenses, including hospital fees, rooms, and facilities.

According to the latest data released by YouGov BrandIndex [9] on their website, the leading insurance companies in Malaysia are Prudential, AIA, Great Eastern Life, Etiqa, and Allianz. All of the insurance companies in YouGov BrandIndex have been ranked based on the index score, which is the measurement of the overall company's performance, including the average impression, customer satisfaction, service quality, recommendation, and reputation. The index rankings chart displayed the highest average index scores for 12 months from July 1st, 2018 to June 30th, 2019. All scores are based on a representative sample of the general population over 18 years old. Prudential, AIA, and Great Eastern Life were the top 3 companies that achieved the highest index score, which is why these companies are chosen as the main scope of this study.

2.2. Sentiment Analysis

Sentiment analysis (SA) is an area of study within natural language processing (NLP) that recognizes and captures opinions within the text. Apart from evaluating opinions, it also evaluates the users' emotions, such as polarity, to identify if the opinion expressed is positive, negative, or neutral, and the holder of the opinion identifies as the person or group providing the opinions [10]. Public sentiment about a product, services, or other topics can be beneficial for different commercial applications. This unstructured information collected by SA will be converted automatically into structured data and divided into three major types, which are regular opinions, comparative opinions, and suggestive opinions that indicate a single entity or multiple entities. The regular opinions refer to a single entity and are used primarily to describe positive or negative perspectives for a particular text. Apart from that, comparative opinions compared or display a correlation between more than one entity and are primarily used for strategic intelligence because they help to elucidate the relationship between multiple entities [11].

2.2.1 Term Frequency Inverse Document Frequency (TF-IDF)

As part of the feature extraction technique, the term frequency-inverse document frequency (TF-IDF) is used and measured before it is passed to the machine learning model. TF-IDF is a numerical statistic widely used in text mining and information retrieval to shows the significance of keywords to specific documents or ranks the frequency of terms used to classify or categorize particular documents [12]. Term frequency calculates the number of times a word appears in a document, divided by the total number of words in that document. On the other hand, the logarithm of the number of documents in the corpus, divided by the number of documents in which the particular word appears, is computed as inverse document frequency [13]. Thus, TF provides a term's frequency, and IDF measures the importance of a term. The algorithm for conducting the TF-IDF is illustrated in Figure 1.

Algorithm

Input: Sentences free from stop words, and stemmed, lemmatized words.
Output: Aspects that retrieved using tf-idf.
Method: Load the input data file. Calculating the frequency of word in the given input file (number of occurrences of i in j).

TF(w) = Number of times term w appears in a document) / (Total number of terms in the document) Number of documents containing i.

IDF(w) = log_e(Total number of documents / Number of documents with term w in it)

***Figure 1**. TF-IDF Algorithm*

2.2.2 Support Vector Machine Algorithm

The classification technique to classify text into classes used in Machine Learning is including unsupervised learning and supervised learning [14]. The machine learning algorithms applicable to SA are primarily part of the supervised classification [15]. In machine learning techniques, data sets needed include a training set and test set. Training datasets are used to train the classifier model. Once a supervised classification technique has been selected, features in sentiment classification inspected will indicate how the documents are represented.

The support vector machine (SVM) is a supervised machine learning model that uses associated learning algorithms to analyze data for classification problems [16]. The algorithm produces the most efficient result in traditional text categorization by locating the best possible distinctions between positive and negative training samples [17]. SVM is chosen as the ML algorithm to build the model since it is the most effective algorithm that runs efficiently on large data sets and offers better results than other traditional classification algorithms [18]. Past research involving SA usage in social media data has shown that the SVM algorithm provides the most accurate text classification problems, achieved by constructing a hyperplane for the nearest trained examples with maximum Euclidean distance. Besides, the outstanding generalization capability of SVM, together with its optimal solution and discriminative power, makes it widely used in the area of text mining. The algorithm is also memory efficient due to its advantage of kernel mapping to high-dimensional feature spaces [19].

2.3. Web-based Visualization Application

A web application is a web-based program designed to operate on devices, including smartphones, tablets, and laptops, through a web browser [20]. Due to standardized technologies, mobile users can access the website on all devices in a similar manner by mobile browsers environment [21]. This approach relies on browser support for mobile platforms and the standardization of the web technology environment. Web technologies are considered most relevant to current trends in the software industry, whereby lifecycles are short, developments are rapid, and customer preferences are frequently evolving [22].

The web application approach was chosen for this study after analyzing several criteria and considering all aspects of constraints such as time constraints and the study's scope to be a fully functioning application. Many frameworks are available and easy to learn for web applications. Considering that websites are more practical for the target audience, rather than worry about the platform's restrictions that constrain native applications such as cross-platform interfaces and capability across various devices, it is more manageable to deal with cross-browser compatibility. The web application only needs to be developed once with cross-browser support, and it is useable on every device. Hence, in comparison to a native application, a web-based application is able to outperform in terms of development time and complexity.

3. RESEARCH METHOD

The three subsections below describe the research method used to conduct this study, including system use case, back-end, and front-end development.

3.1 System Use Case

The use case diagram is used to gather system requirements, including internal and external factors. Furthermore, interactions needed between the user and the system to accomplish a particular task can be represented and monitored using a use case diagram. The completed system has six main features, including an overview page, data dashboard, Twitter updates, sentiment analyzer, and competitive analysis. Figure 2 illustrates the flow of events from the user's perspective and the system's behavior responding to the requests.

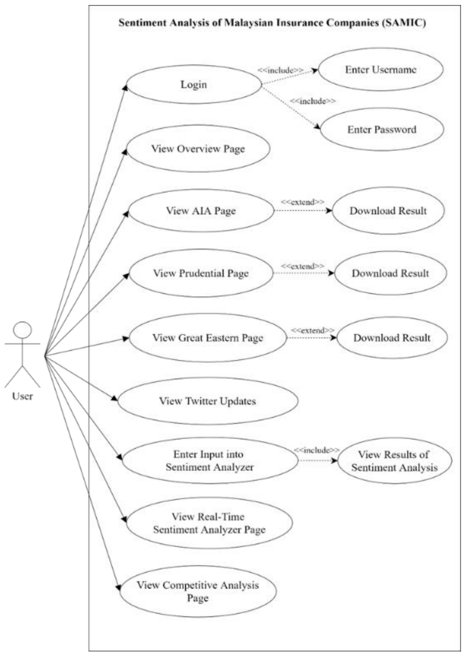


Figure 2. *Use Case Diagram of the Application*

3.2 Back End Development

This study involves full-stack web development, which implies that the back end and front end are two parts included in the web application. The next subsections explained all the steps involved in creating the web application that deploys the machine learning model, including data preparation, predictive modeling, and model deployment.

3.2.1 Data Preparation

Data preparation involves transforming raw data collected into a form that can be modeled using a machine learning algorithm. The raw data cannot fit into predictive modeling due to the algorithm's imposing requirements on the data to be numeric. Besides, statistical noise and errors in the data must also be corrected. As such, prior to being used to satisfy the ML model's requirements, the raw data must be prepared. Data preparation is a multi-step process that involves data collection, cleaning and pre-processing, feature engineering, and labeling.

Two machine learning models have been built for classification purposes, namely the Malay and the English multi-class text classification model. The difference between these two models is the way the data is pre-processed. The dataset from *https://malaya.readthedocs.io/en/latest/Dataset.html* has been collected from the Twitter-sentiment and polarity subfolder for training and testing the Malay model. This dataset is labeled to reflects its sentiment, whether it is positive or negative, using a semi-supervised model. All data collected in the text corpus is in the format of javascript object notation (JSON). The Malay model's total data is balanced for the positive and negative labeled data in which both of these classes comprised 637,760 sentences each.

On the other hand, 1,200000 positive and negative English sentences are collected from *http://help.sentiment140.com/for-students* to built a machine learning model based on English sentences. Both of these collected datasets are binary classification data, which only annotates negative and positive data. Therefore, to add neutral data to the training and testing dataset, the neutral data is gathered from *https://www.kaggle.com/crowdflower/twitter-airline-sentiment*. Next, the neutral dataset has been translated into the Malay language according to how Malaya-dataset is prepared, and wherein Google translator is used to translate the validated English dataset to a Bahasa Melayu dataset. However, only 12,547 neutral sentences were discovered, a small amount compared to positive and negative annotated data because a publicly accessible neutral dataset is difficult to obtain. The training and testing data that have been supplemented with neutral data for the Malay model are 1,287,977 and 1,212,547 for the English model.

Twitter has many Malaysian users, most of whom post and write their tweets to share their opinions using the Malay language. The data was scraped from Twitter to obtain real-world data. The historical data for these three insurance companies, AIA, Prudential, and Great Eastern, were collected, ranging from January 1st, 2017 until October 31st, 2020. The first approach was to retrieve tweets using general keywords that directly and indirectly, referring to these insurance companies, such as 'AIA', 'Prudential', and 'Great Eastern'. However, most tweets extracted using the first approach were not useful because they did not express any opinions required for a sentiment analysis application. Alternatively, it was more practical to use the keywords of the English term 'AIA Insurance' and Malay term 'AIA insurans' and analyze the hashtags these insurance companies mentioned. Trending hashtags of these insurance companies, including '#aia', '#prudential', and '#greateastern' have also been scrapped. The total number of collected data is 49,523 for AIA, 17,516 for Prudential, and 38,857 for Great Eastern.

Machine Learning requires text data in a numerical form. Encoding techniques such as BagOfWord, Bi-gram, n-gram, TF-IDF, and Word2Vec have been used to encode the real-world data collected from Twitter into a numeric vector. However, before encoding text data, it needs to be cleaned first. This process of preparing text data is called text pre-processing. Text pre-processing is a method of cleaning and preparing text data for use in a particular context. 'NLTK' and 're' are two of the Python libraries used in this study to perform text pre-processing tasks. Text cleaning has been done on the dataset, and it involves removing unwanted characters like emojis and properly formatting the text to remove any extra spaces. All characters were converted to a lowercase beforehand to avoid any case-sensitive operation. Terms such as mentions, hashtags, links, and usernames have been removed using patterns that can match the desired terms with a regular expression (Regex). Some tweets might also contain a Unicode character that cannot be read in an ASCII format. Mostly, these characters are used for emojis and non-ASCII characters. All of these characters have been removed as well. Duplicate tweets and null values have been dropped as they do not contribute useful information to the dataset and computationally inefficient.

3.2.2 Predictive Modeling

The model evaluation is conducted using the support vector machine (SVM) classifier, in which the algorithm is used to perform the dataset classification. SVM is a learning machine for two-group classification problems developed natively for binary classification problems and cannot be used directly with multi-class classification problems. Hence, One-vs-Rest (OvR) is incorporated into the classifier. OVR is a multi-class classification method that uses multiple binary SVM classifiers to obtain a multi-class prediction. It involves the splitting of the multi-class dataset into multiple binary classification problems. A binary classifier would then be trained on each binary classification, requiring each model to predict the probability of class membership or a probability-like score. The Argmax of these scores, which is the class index with the highest score, is used to predict the class label, and predictions are made using the most confident model.

The initial step in predictive modeling is to prepare text data to draw features for the machine learning (ML) algorithm to obtain an effective prediction. The bag-of-words concept is used as a weighting scheme to transform the whole corpus list of tokens into a matrix-vector that the model can understand and learn. A vocabulary of known words and a measure of known terms to represent all the words in model vocabulary as a list of tokens are required. The SciKit Learn’s CountVectorizer function converted the collection of tweets to a token count matrix. As the corpus of text contains millions of tweets, there would be several zero counts for every word in the corpus. The last step is to construct document vectors to score the word frequency in each document word with TF-IDF.

3.2.3 Model Deployment

Model deployment is one of the final steps in the machine learning life cycle. Predictive analytics based on trained data is carried out by using the model that has been developed. At the model development stage, the model took the data frames of tweets and returned the predicted classified sentiment label as ‘positive’, ‘neutral’, and ‘negative’ represented as ‘0’, ‘2’, and ‘4’, respectively. These data frames are sorted by date and afterward integrated into the production environment. In the implementation context, the model results are presented using the data visualization technique via a web service by displaying overall sentiment scores produced offline across the batch process. Prior to deployment, the model is tested to ensure that the test input set identified during development produces validated results by setting the production environment in accordance with the development environment. It was done by defining the environment specifics such as specific language versions and library dependencies for the model in Python requirements.txt file.

After sentiment predictions are made using the model classifier on the data obtained, and its performance is evaluated, the data is visualized using Plotly, an open-source interactive graphics library for Python. The data is first imported into the Pandas data frames in Python. Interactive charts are then generated. Additional information on the results of the analysis will be displayed through hover events. As in Table 1, the analysis results are used to create an interactive visualization tool to present real-world data analysis outcomes.

Table 1. *The Sentiment of Mentions by Months*

Month	AIA		Prudential		Great Eastern	
	Positive	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**
Oct 2020	919	353	170	80	794	111
Sep 2020	741	261	128	44	812	144
Aug 2020	795	269	233	119	775	95
Jul 2020	891	371	182	65	756	75
Jun 2020	720	301	166	64	663	69
May 2020	678	254	210	54	594	76
Apr 2020	669	223	235	71	530	66
Mar 2020	682	236	322	52	552	65
Feb 2020	1220	514	113	46	509	65

Jan 2020	865	294	144	43	566	55
Dec 2019	564	187	304	67	560	54
Nov 2019	849	363	136	48	568	65
Oct 2019	892	361	201	102	805	68
Sep 2019	677	262	191	79	597	76
Aug 2019	699	236	423	83	712	113
Jul 2019	826	277	1770	180	734	75
Jun 2019	848	179	1837	202	641	57
May 2019	947	359	486	52	729	75
Apr 2019	984	366	267	90	693	66
Mar 2019	788	269	183	51	768	72
Feb 2019	796	274	136	38	707	62
Jan 2019	741	283	160	64	899	107
Dec 2018	717	350	107	55	649	42
Nov 2018	787	239	224	47	777	93
Oct 2018	822	245	435	85	1031	72
Sep 2018	787	230	184	49	746	54
Aug 2018	782	294	190	60	720	50
Jul 2018	665	210	237	51	731	48
Jun 2018	961	230	179	82	648	47
May 2018	687	311	172	98	661	49
Apr 2018	917	382	209	96	800	52
Mar 2018	734	301	319	81	900	35
Feb 2018	590	239	148	73	671	53
Jan 2018	622	256	168	65	675	37
Dec 2017	611	223	117	42	554	39
Nov 2017	583	220	130	44	636	64
Oct 2017	736	309	163	64	850	52
Sep 2017	683	523	146	56	808	86
Aug 2017	508	252	170	64	683	55
Jul 2017	428	248	305	250	742	64
Jun 2017	455	194	148	46	616	62
May 2017	614	306	164	89	766	89
Apr 2017	787	282	161	90	863	76
Mar 2017	726	398	208	101	807	76
Feb 2017	733	319	177	68	752	51
Jan 2017	468	231	178	44	960	47

3.4 Front End Development

Front end development refers to the development of the ‘client-side’, where the emphasis is on what users see in their browser or application visually. HTML, CSS, Javascript, and Jquery are the front-end languages used to integrate Flask’s back-end framework. In a Python web application environment, the Flask web framework and the data visualization tools in Python are used to construct custom plots and charts after the data manipulation phases to leverage both front-end and back-end development power.

4. RESULT AND ANALYSIS

This segment presents and discusses the analysis results performed on real-world data and the testing carried out on accuracy, functionality, and usability testing.

4.1 Accuracy Testing

The classification report in Figure 3 provides the main classification metrics on every class of the data. In comparison, the heatmap represents the proportion in each class of correctly classified examples. The raw numbers of testing data are also stated within each cell. A simple metric, including classifier error rate and accuracy, has been used to measure the model's performance. In the confusion matrix, the 'negative' class is represented as 0, whereas the 'positive' class is denoted as 2, and 4 represented the 'neutral' class. The classifier has an accuracy of 90%. In simpler terms, out of 10 attempts, the model obtained approximately nine correct results based on the data provided from the testing data set, which were correctly classified as 'positive', 'neutral', and 'negative'. The precision, indicating the percentage of a class predicted labels that rightfully belonged to that class is 92% for the 'negative' class, while the 'positive' class is 88% and 31% for the 'neutral' class. The recall shows the percentage of labels that were correctly classified from the total of a class. For both the 'negative' and 'positive' class, the recall is 90%, while for the 'neutral' class, it is 19%. As compared to other classes, the 'negative' class is the most correctly classified, with only a few instances being misclassified. The outcomes of precision and recall values for each of the three classes confirm that the 'neutral' class suffers from substantially poorer results than other classes. Therefore, to evaluate the classifier approach, a baseline study has been specified and compared to compare the classifier performances.

```
accuracy score:  0.8956365537714208


confusion matrix:
 [[115219  12297    371]
 [  9897  96784    498]
 [   598   1089    399]]


              precision    recall  f1-score   support

           0       0.92      0.90      0.91    127887
           2       0.88      0.90      0.89    107179
           4       0.31      0.19      0.24      2086

    accuracy                           0.90    237152
   macro avg       0.70      0.67      0.68    237152
weighted avg       0.89      0.90      0.89    237152
```

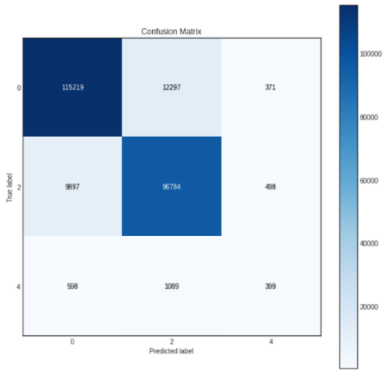


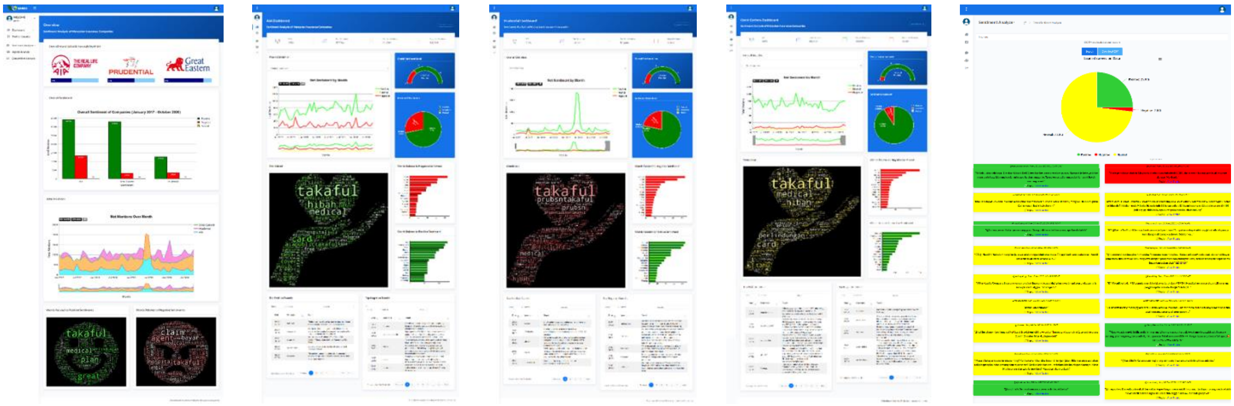
Figure 3. *Confusion Matrix of Malay Model*

Based on the same study conducted by Bouazizi and Ohtsuki [23] for multi-class sentiment analysis on Twitter, neutral tweets were also the hardest to classify, with a true positive rate equal to 28.5%. The overall accuracy of the validation testing data is equal to 76.3%. Additionally, the sentiment analysis model for Twitter data in the Polish language revealed that multi-class classification is still significantly more difficult to be modeled and only achieves an accuracy of 18.1% when used to classify tweets into three classes of positive, neutral, and negative. However, it produced nearly 71% accuracy and high precision of 87.77% when trained with the binary class [24]. Most studies on sentiment analysis overlooked 'neutral' examples, but the classification is important to distinguish between positive and negative examples. The Malay model's overall accuracy showed an excellent classification with an accuracy score of 90%. Thus, according to the overall accuracy, the model is acceptable to classify real-world data.

Over and above that, the relevant findings that can be concluded are that the classification model's performance is independent of the test set, and the quantification part can be carried out without overfitting issues for the classification of a real-world dataset. Since a balanced dataset for 'neutral' data is hard to collect a priori, it is believed that adding 'neutral' instances of training data, and more importantly, a balanced training dataset among all sentiment classes, could significantly improve the performance of the model.

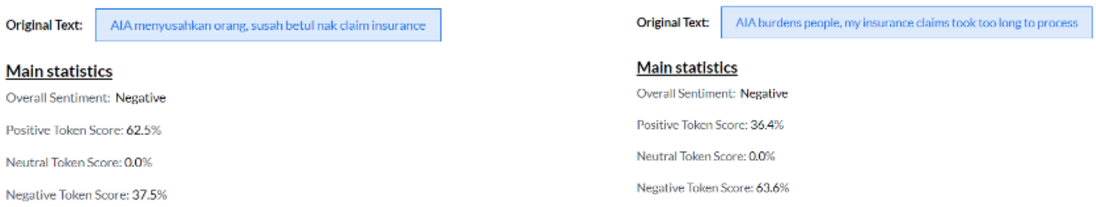
4.2 Functionality Testing

Functionality testing is essential to ensure that all application features operate correctly and that any detected error is fixed. The purpose of performing functionality testing is to test each function of the visualization application to determine how closely the specifications are matched by providing appropriate input and checking the output against the functional requirements. The test is conducted based on the development of test case scenarios derived from program specifications, with the list of functionalities in the use-case diagram was tested. Figure 4a represents the SAMIC overview page interface, which displays the summary analysis results, while Figures 4b, 4c, and 4d show the respective insurance companies' dashboards. Tweets are retrieved in real-time and sentiment analysis is conducted on the tweets as illustrated in Figure 4e. Conversely, the user can also manually input their review in Malay or English as in Figure 5a and Figure 5b to see whether the sentence is positive, negative, or neutral using the machine learning model built.



(a) ***(b)*** ***(c)*** ***(d)*** ***(e)***

Figure 4. *The interface of SAMIC Overview Page (a), AIA Dashboard (b), Prudential Dashboard (c), Great Eastern Dashboard (d), Real-time Sentiment Analyzer Page (e)*



(a) ***(b)***

Figure 5. *Result of Sentiment Analysis Performed on Malay Text (a), English Text (b)*

4.3 Usability Testing

The usability tests were conducted individually with the user using TeamViewer, where users gained remote access to the applications because of the Covid-19 situation, which prevented in-person usability testing from being conducted. Thirty users who have prior experience in purchasing insurance were selected to participate in the usability test and asked to complete some common tasks, such as evaluating each system's feature. The users responded to the ten questionnaires at the end of the evaluation. The questionnaire provided the following ten standard statements with five response options, indicating five Likert scales with anchors for strongly agree and strongly disagree. Users have ranked each question from 1 to 5 based on how much they agree with the statement they read, and it is compulsory to fill in all the questions. The statements in odd-numbered questions, including question 1, question 3, question 5, question 7, and question 9, are expressed positively. On the contrary, even-numbered questions such as question 2, question 4, question 6, question 8, and question 10 are phrased negatively.

Figure 6 shows the histogram of the SUS scores. The plotted graph has a normal distribution with a range starting at 88%, and followed by a new range for every increase of 2%. The highest frequency is 92% to 94%, of which nine respondents fall into that range. The histogram is centered on the same value with the highest frequency of 92% to 94%. Ten respondents fall in the category below the central value, and eleven respondents lie in the above central value. Overall, the SUS score of 92.5% was obtained from 30 users in the SUS questionnaire. The SUS average score's baseline is 68%, indicating that the system has average system usability. Whenever SUS exceeds 80%, the system has excellent usability [26]. To measure how the SUS scores are evaluated, 80.3 scores or higher is graded as an 'A' [27]. Hence, by receiving an excellent score of more than 80%, this application is proven to have good usability. It is also noted that most of the testers gave good reviews about the application and would recommend it to the potential insurance buyers they knew.

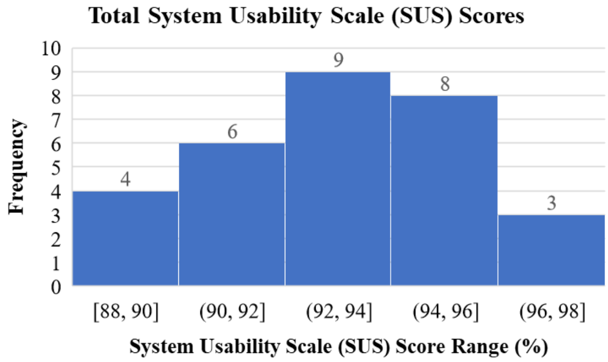


Figure 6. *SAMIC Histogram for Total SUS Scores*

5. CONCLUSION

This study's primary purpose was to develop an interactive visualization application that would make the analysis results readable and understandable for policyholders on the public sentiment of the policies provided by Malaysia's top insurance companies. Consequently, policyholders could make subtle and informed decisions to better understand outside information from real-world data. The application's testing phases show that the classifier successfully classified tweets' sentiment with 90% accuracy. It is observed in the functionality testing that every feature in this application functions properly. Furthermore, this application is proven to have good usability by receiving an excellent score of 92.5% in the SUS score. The support vector model used for the classification process has been incorporated into the application, allowing the user to use the model on any English and Malay textual data reviews. Additionally, this application also includes real-time monitoring of tweets sentiment that uses the data processing infrastructure and sentiment analyzer to evaluate public sentiment changes in response to user search keywords. For future recommendation, it is suggested that the quantity of neutral dataset be increased in multi-class classification, hoping that there will be a publicly available neutral dataset in Malay and English language so that more accurate results can be obtained to optimize the performance of the classifier.

ACKNOWLEDGMENTS

The authors would like to thank Research Management Centre under the grant of 600-IRMI/REI 5/3 (010/2018) of the University of Technology MARA.

REFERENCES

[1] "Malaysia : Publication of Financial Sector Assessment Program," 2018.

[2] N. J. Bao, R. Ramlan, F. Mohamad, and A. M. Yassin, “Performance of Malaysian insurance companies using data envelopment analysis,” Indones. J. Electr. Eng. Comput. Sci., vol. 11, no. 3, pp. 1147–1151, 2018.

[3] R. Sum and N. Nordin, “Decision-making biases in insurance purchasing,” J. Adv. Res. Soc. Behav. Sci., vol. 10, no. 1, pp. 165–179, 2018.

[4] “Bank negara annual report 2016”, Bank Negara Malaysia (BNM)., pp. 109-110, 2016.

[5] McMaken, L., “Types of insurance everyone needs”, 2019. [Online]. Available: https://www.investopedia.com/financial-edge/0212/4-types-of-insuranceeveryone-needs.aspx. [Accessed: 05-March-2021].

[6] Z. K. Zulkifly, Nur Ain Qamarina Zulkifly Kasim and J. B. Bidin, “Selection of personal medical and health insurance company by using fuzzy topsis,” J. Intelek, vol. 14, no. 1, 2019.

[7] M. S. Smith and C. R. Hayhoe, “Life insurance : Whole-life insurance,” Virginia Coop. Ext., vol. 354–145, pp. 1–3, 2009.

[8] R. R. Bovbjerg and J. Hadley, “Why health insurance is important,” Heal. Policy Briefs. Urban Institute. Washington, DC, no. 11, p. 3, 2007.

[9] YouGov-BrandIndex., “Malaysia top rankings index rankings by country”, 2019. [Online]. Available: https://www.brandindex.com/ranking/malaysia/2019-index/category/insurance. [Accessed: 05-March-2021].

[10] M. Ahmad, S. Aftab, S. Muhammad, and S. Ahmad, “Machine learning techniques for sentiment analysis: a review,” Int. J. Multidiscip. Sci. Eng., vol. 8, no. 3, pp. 27–32, 2017.

[11] S. Shayaa et al., “Sentiment analysis of big data: methods, applications, and open challenges,” Inst. Electr. Electron. Eng. Access, vol. 6, p. 21, 2018..

[12] S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to examine the relevance of words to documents,” *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018.

[13] K. Srividya and A. Mary Sowjanya, “Aspect based sentiment analysis using POS tagging and TFIDF,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 1960–1963, 2019

[14] S. S. Salman, M. U. K. Khan, and M. Z. I. Iqbal, “A systematic mapping study on testing of machine learning programs,” 2019.

[15] K. Sailunaz and R. Alhajj, “Emotion and sentiment analysis from twitter text,” J. Comput. Sci., vol. 36, p. 18, 2019.

[16] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, “Study of twitter sentiment analysis using machine learning algorithms on python,” Int. J. Comput. Appl., vol. 165, no. 9, pp. 29–34, 2017.

[17] A. Dey and S. Bandyopadhyay, “Automated glaucoma detection using support vector machine classification method,” Br. J. Med. Med. Res., vol. 11, no. 12, pp. 1–12, 2016.

[18] M. F. Luthfi and K. M. Lhaksamana, “Implementation of TF-IDF method and Support Vector Machine Algorithm for job applicants text classification,” *J. Media Inform. Budidarma,* vol. 4, pp. 1181–1186, 2020.

[19] C. Meng, S. Jin, L. Wang, F. Guo, and Q. Zou, “AOPs-SVM: A sequence-based classifier of antioxidant proteins using a Support Vector Machine,” *Front. Bioeng. Biotechnol.,* vol. 7, no. September, pp. 1–10, 2019.

[20] N. A. B. M. A. Ali and M. N. S. Sari, “The suitability of native application for university e-learning compared to web-based application,” Int. J. Sci. Res., vol. 4, no. 1, pp. 2045–2049, 2015.

[21] H. Heitkotter, Sebastian, S. Hanschke, and T. A. Majchrzak, “Evaluating cross-platform development,” pp. 120–121, 2013.

[22] J.-P. Erkkilä, “Web and native technologies in mobile application development,” Aalto University, 2013.

[23] M. Bouazizi and T. Ohtsuki, “A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter,” IEEE Access, vol. 5, no. May 2020, pp. 20617–20639, 2017.

[24] K. Chlasta, K. Wołk, and I. Krejtz, “Automated speech-based screening of depression using deep convolutional neural networks,” Procedia Computer Science, vol. 164, pp. 618–628, 2019.

[25] Wahome, R., “This is how Twitter sees the world: Sentiment analysis part one”, 2020. [Online]. Available: https://towardsdatascience.com/the-real-world-as-seen-on-twitter-sentiment-analysis-part-one-5ac2d06b63fb. [Accessed: 05-March-2021].

[26] Alathas, H., “How to measure product usability with the system usability scale (SUS) score”, 2018. [Online]. Available: https://uxplanet.org/how-to-measure-product-usability-with-the-system-usability-scale-sus-score-69f3875b858f. [Accessed: 05-March-2021].

[27] Wang, Y., “System usability scale : A quick and efficient user study methodology”, 2021. [Online]. Available: http://ixd.prattsi.org/2018/04/system-usability-scale-a-quick-and-efficient-user-study-methodology/r Science, 11(3), 1147–1151. [Accessed: 05-March-2021].