IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 4, December 2023, pp. 1793~1802
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i4.pp1793-1802

Classification and visualization: Twitter sentiment analysis of Malaysia's private hospitals

Khyrina Airin Fariza Abu Samah[1], Nur Maisarah Nor Azharludin[1], Lala Septem Riza[2], Mohd Nor Hajar Hasrol Jono[1], Nor Aiza Moketar[3]
[1]College of Computing, Informatics and Media, Universiti Teknologi MARA Melaka Branch, Jasin Campus, Melaka, Malaysia
[2]Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3]Faculty of Information Technology and Communication, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

Article Info

Article history:

Received Nov 7, 2022
Revised Jan 30, 2023
Accepted Mar 10, 2023

Keywords:

Classification
Data visualization
Naïve Bayes
Twitter sentiment analysis

ABSTRACT

Malaysia has many private's hospitals. Thus, feedback is important to improve service quality, becoming reviews for other patients. Reviews use the channel service provided on social media, such as Twitter. Nevertheless, online reviews are unstructured and enormous in volume, which leads to difficulties in comparing private hospitals. In addition, no single websites compare private hospitals based on users' interests, bilingual reviews, and less time-consuming. Due to that, this study aims to classify and visualize the Twitter sentiment analysis of private hospitals in Malaysia. The scope focuses on five factors: 1) administrative procedure, 2) cost, 3) communication, 4) expertise, and 5) service. Term frequency-inverse document frequency is used for text mining, information retrieval techniques, and the Naïve Bayes, a machine learning algorithm for the classification. The user can visualize the specified state's private hospitals and compare them with any selected state. The system's functionality and usability have been tested to ensure it meets the objectives. Functionality testing proved that the private hospital's Twitter sentiment could be predicted based on the training and testing data as intended, with 77.13% and 77.96% accuracy for English and Bahasa Melayu, respectively, while the system usability scale based on the usability testing resulted in an average final score of 95.42%.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Khyrina Airin Fariza Abu Samah
College of Computing, Informatics and Media, Universiti Teknologi MARA Melaka Branch
Jasin Campus, Merlimau, 77300, Melaka, Malaysia
Email: khyrina783@uitm.edu.my

1. INTRODUCTION

Private hospitals in Malaysia contribute high-quality services and patient satisfaction due to the industry's intense competition. Nonetheless, very few studies quantify the service quality of private hospitals [1]. As the number of private hospitals increases annually, they must compete to provide the finest care to their patients, enhancing their hospital's reputation [2]. Social media reviews are one of the most efficient ways to collect data that can serve as an indicator of the service quality improvement of private hospitals. It may help all stakeholders in healthcare [3]. Nonetheless, because social media reviews are unstructured and abundant, they may lead to a fairly erroneous result [4].

Choosing the best private hospital for treatment is vital for every consumer, as no website or application can compare users' preferences, resulting in unhappiness with the selection process. The centre website is a voice for national resilience and the strengthening of centrist thought in Malaysia. It has only compared the service prices for public and private hospitals in Malaysia [5]. There was no mention of a specific hospital providing the service. Plan-do-check-act (PDCA), 5S, Kaizen, Control charts, and root cause analysis (RCA) are a few of the quality methodologies utilized by hospitals to achieve high-quality performance and boost patient satisfaction in recent years [6]. Zakaria and Wahab [7] conducted a study using descriptive and inferential analysis, and a questionnaire was utilized to analyse consumer perceptions, satisfaction, and behavioural intentions. Because it does not directly compare private hospitals, it is impossible to choose based on preferences, and it is a time-consuming comparison, the generalizability of these results is limited. For example, a study by [8] only focuses on hospital performance in Pakistan, and a study by [9] focuses on the India Institute Medical with no specific algorithm used.

Social media has become one of the essential venues for global communication and information gathering in the modern global economy. According to Dixon [10], worldwide social media users reached 4.2 billion in January 2021. Moreover, social media provides a platform where individuals may search for information, exchange ideas, and even virtually display their personal and professional lives [11]. Twitter is a popular social networking platform among active Internet users, particularly young people aged 25 to 34. Tweets allow users to communicate their thoughts and ideas with others. Kemp [12] stated that in 2021, Twitter would have approximately 397 million monetizable active usage and 187 million daily users. This large number of users suggests that Twitter is also a platform where users receive and share information.

Numerous sectors actively utilize online reviews because they influence consumer decisions. However, online reviews are limited because they are predominantly displayed in English [13]. Because Bahasa Malaysia is the most widely used language in Malaysia, it might have a negative impact on the outcome of decisions. In another significant study, Antonio *et al.* [14] discovered that analysis based on many languages produced more accurate results. It presents an opportunity for private hospitals to attract more patients.

Therefore, this study entails the development of a web-based dashboard to visualize the performance of private hospitals in Malaysia based on Twitter sentiment analysis (SA) from January 2021 to December 2021. The retrieved tweets only address the public's perception of Malaysian private hospitals regarding the administrative procedure, communication, cost, expertise, and service. The scope of the study includes 146 private hospitals in all 14 Malaysian states. The collected tweets reflect public sentiment regarding reviews of private hospitals in Malaysia based on the following five factors: administrative procedure, communication, cost, expertise, and service.

We have compared the existing machine learning algorithm such as artificial neural network, random forest, support vector machine, Naïve Bayes and K-nearest neighbour in order to identify the best technique. We chose and applied Naïve Bayes (NB), a straightforward learning technique based on Bayes' rule and the strong assumption that the attributes of a class are conditionally independent [15]. NB is among the most successful and efficient inductive learning algorithms for machine learning and data mining [16]. The NB classifier was used to evaluate the model using an algorithm to classify the dataset. The model applies the training set's labeled data to the dataset to classify it.

Data visualization is a crucial instrument for getting valuable information. It should depict data with charts and graphs and convey them intuitively [17]. This study utilized four visualization techniques: a line chart, a bar chart, a pie chart, and word clouds, since the extracted data from Twitter is more effective in displaying. Consequently, it is easier to identify large data sets' trends, patterns, and outliers. This study would consolidate all data into a more comprehensible visual format to facilitate user comprehension. The data is visualized using Plotly, Python's open-source interactive graphics tool. The model is created utilizing English and Bahasa Malaysia datasets to analyse sentiment in both languages. The outcomes can be used to increase client satisfaction and retain them. Thus, consumers will pursue prospective new markets and resolve customer issues more effectively. The paper structure is as follows: The first section is an introduction followed by the research methods in section 2. Section 3 focuses on the findings and discusses their accuracy, functionality, and usability. Finally, section 4 concludes the analysis by quickly noting possible future improvements.

2. RESEARCH METHOD

2.1. System design

System design is defined as implementing a system's product development concepts. Developing design diagrams facilitates the design process. It included the use case diagram, flowchart, and user interface.

2.2. Back-end development

The research design depicted in Figure 1 is the overall web-based dashboard development. The method of the study was divided into 4 sections for elaboration. During system development, the back end, often known as server-side development code, is the data access layer. The system's back end is written in Python, from data preparation to model deployment. Important back-end tasks for training and testing data include data collection, pre-processing, NB classification model development, and model deployment.

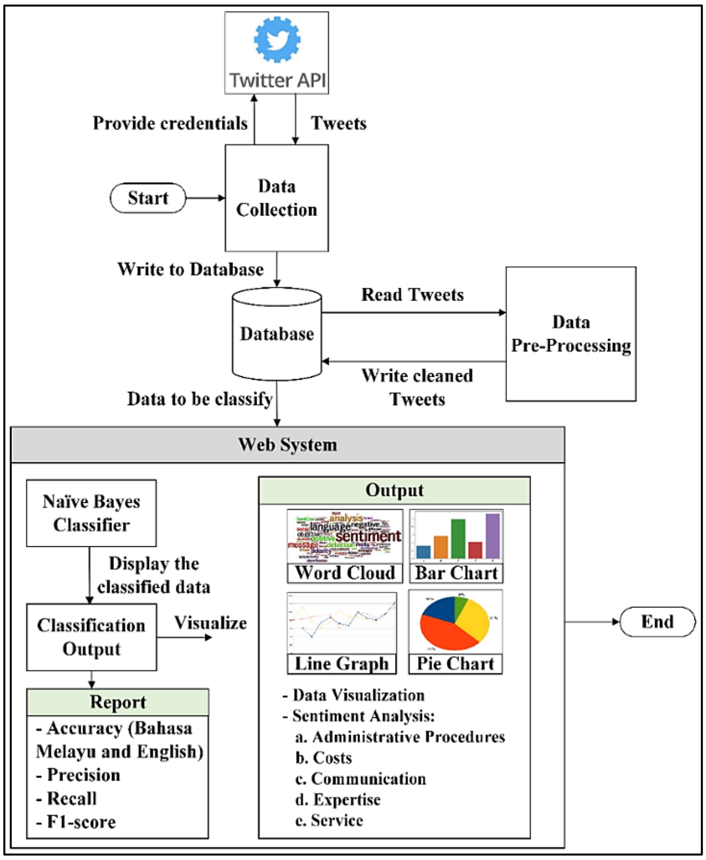


Figure 1. Flow diagram of research design

2.2.1. Data collection

Text classification for both Malay and English are performed using machine learning algorithms. The data source for the English model is taken from the website [18]. It contains 800,000 positive and negative data points. Meanwhile, data sources to train the Malay model are taken from [19]. The gathered neutral data for the English model is Malay conversion using onlinedoctranslator.com, which translates using Google Translate containing the additional neutral data for the Malay model. The English model has 1,614,640 data for training and testing, whereas the Malay model contains 531,679 data. The English and Malay datasets are utilized to determine whether the sentiment data is in English or Malay.

The data for the 14 states, Johor, Perlis, Terengganu, Malacca, Kelantan, Pahang, Sabah, Sarawak, Negeri Sembilan, Kedah, Perak, Penang, Wilayah Persekutuan, and Selangor are taken using real-world data from Twitter. The tweets were scrapped between January 1 and December 31, 2021 using Twint, where there is no case sensitivity for terms. The scraped tweets are then saved as comma separated values (CSV) files. The gathered data is manually analysed from the scraped data to eliminate empty cells from the tweet column. Table 1 shows the comparison of results for the private hospitals in Malaysia with 4,689 raw data in English and Malay collected, with 3,717 total positive mentions, 43 neutral mentions and 926 negative mentions. The raw data includes 36 variables, such as the tweet ID, username, tweet date, tweet content, language, and tweet link. The CSV file is read using the Pandas package.

2.2.2. Data pre-processing

Before encoding, text pre-processing is carried out to clean up the data [20]. Preparing data for use by eliminating and discarding extraneous text that does not add value to the model and instead decreases its quality is a technique known as text pre-processing [21], [22]. Natural language toolkit (NLTK) and 're' are the two Python packages used for text pre-processing. Only three columns are available for the final dataset: data, username, and tweet. We remove the unneeded columns. The dataset's text is cleaned by changing all characters to lowercase to prevent case-sensitive issues during pre-processing. Then, characters such as emojis, punctuation, and excessive whitespace were eliminated. The elimination of keywords such as links, hashtags, and mentions. In addition, duplicate tweets and null values from the dataset were removed to further minimise the data's dimensionality.

Table 1. Comparison of results for state's private hospitals for combination of both English and Malay

States	Total Mentions	Total Positive Mentions	Total Neutral Mentions	Total Negative Mentions
Johor	47	20	17	10
Kedah	16	14	0	2
Kelantan	2	2	0	0
Malacca	44	39	0	5
Negeri Sembilan	76	60	1	15
Pahang	31	27	0	4
Penang	356	311	0	45
Perak	257	249	0	8
Perlis	28	20	0	8
Sabah	55	43	0	12
Sarawak	74	64	0	10
Selangor	2,201	1,685	18	498
Terengganu	3	3	0	0
Wilayah Persekutuan	1,496	1,180	7	309
Total	**4,689**	**3,717**	**43**	**926**

However, this dataset is still of high dimension. Stop words are removed from the data to reduce dimensionality because they add no value. In English, stop words include "the", "and", "of", and "on". English stop words can be found in the NLTK library's pre-built function. Malay stop words are manually imported from [23] for stop word removal in the Malay model. The data was tokenized to form a bag of words, which is the process of extracting the words from the remainder of the text. After tokenization, raw text is transformed into collections of tokens, each of which is often a single word. In addition, the stem process known as lemmatizing was carried out. It is an approach to text normalization that eliminates suffixes. It decreases the number of words to diminish the text's dimension further. After pre-processing, the completed dataset is saved to the working directory.

2.2.3. Naïve Bayes classification model

Naive Bayes (NB) classifier evaluates the model, which categorises the dataset using an algorithm [24]. The model takes the pre-labeled data from the training set and applies it to the dataset for classification. The probability of an event is determined by the NB theorem utilising the probabilistic joint distribution of previous occurrences [25]. In this research, the model is helped to understand the context of positive, neutral, and negative phrases using a pre-labelled training dataset.

The text representation is a structured representation of a collection of expressions and words that counts how many times the phrase "Bag of Words" appears (BOW) [26]. It entailed extracting features from the tokens of words obtained and transforming them into a vector that a machine learning model could learn. This technique includes counting the term frequency, inverse document frequency, and normalising the vectors to unit length, where all steps from the bag of words (BOW). Term frequency-inverse document frequency (TF-IDF) is a statistical measure that determines how essential a word is in a document when the first two phases of BOW are combined [27], [28]. The TF-IDF weight is the weight used in information retrieval and text mining. Term frequency (TF) assessed the frequency of phrase occurrence in a single document was using, while the significance level was determined using inverse document frequency (IDF).

Cross-validation utilises the training data to guarantee that the model does not overfit the data [29], [30]. Several hyperparameter configurations are investigated to divide the model into pieces randomly. The model includes eight evaluated parameter configurations and 10 KFold validations. As a result, the model was trained and evaluated 80 times. The data is separated into training and testing datasets with an 80:20 split for both English and Malay models. Implementation in the real world is the next step in evaluating the model's performance. It is evaluated using the test holdout dataset. The evaluation yielded a classification report and a confusion matrix as performance measures. Examining the accuracy measure, confusion matrix, and classification report data. The data is delivered through the Twitter application programming interface (API) for sentiment predictions before the data visualisation process begins, and the model's effectiveness is assessed.

2.2.4. Model deployment

Model deployment is the process of deploying a machine-learning model for practical usage. Frequently, the phrase refers to making a model accessible via real-time APIs so that information can be retrieved in real-time. At the stage of model deployment, the predicted categorized tweets are generated with sentiment labels of "0", "2", and "4", which represent negative, neutral, and positive attitudes, respectively. Once settled, constructing the prediction of the sentiment using the model classifier on the gathered data and evaluating its efficiency, the data is shown using Plotly. Plotly is an open-source Python library for interactive graphics. The approach begins with loading the data into Python Pandas data frames. Jupyter Notebook is then used to code the data from the Excel file. Then, charts will be made utilizing the chart studio in online Plotly, together with the entered data. Consequently, an interactive visualization tool is designed to depict real-world data processing results using the outcome. The suggested method will visualize the text data for private hospitals using word cloud visualization. The words will be shown in various colors, with the size of each word emphasizing their frequency in the text data. The terminology utilized by private hospital businesses will be shown in a cloud for simple viewing.

2.3. Testing development

We perform the test after completing the dashboard to ensure the system functions well and can be reliable for the users to view. Functionality testing is necessary to guarantee that all system features work correctly and that any unusual behaviour is swiftly detected and corrected [31]. Functional testing aims to test each system function to ensure that the functional criteria outlined in earlier chapters are met. This test is based on constructing test cases drawn from system requirements. Usability testing is performed on a system by a group of representative users to determine how accessible it is to use [32]. Users are prompted to evaluate the system functionalities while being observed to determine whether users encounter any issues when using the system. Some recommendations are provided to help users with usability difficulties.

2.4. Front-end development

Front-end web development, also known as client-side development, translates data into a graphical interface using HyperText Markup Language, Cascading Style Sheets, and JavaScript to construct a website that enables users to view and interact with the data. The Python web application environment consists of data visualization tools for generating charts and graphs of sentiment data. The developments involve three modules: the dashboard page, each state dashboard and the comparison among the states.

3. RESULTS AND DISCUSSION

3.1. Accuracy testing

The simple Python code is used to evaluate the accuracy of the Naïve Bayes classification model. Figure 2 depicts the accuracy testing results for the English model of the training dataset. The score for accuracy is 77% when expressed as a percentage. This score indicates that the model correctly categorized seven out of ten correct responses as “positive”, “neutral”, or “negative”. In the confusion matrix, the “negative” class is represented by 0, the “neutral” class by 2, and the “positive” class by 4.

The accuracy score for the Malay model of the training dataset is depicted in Figure 3. The confusion matrix’s accuracy score is expressed in percentage form. It is 77% similar to the English model, indicating that the sentiment result is 77% accurate, with the algorithm correctly categorizing seven out of ten right outcomes as “positive”, “neutral”, or “negative”. The low accuracy score of 77% for both models result from the small amount of neutral sentiment data compared to a large amount of negative and positive sentiment data.

```
accuracy score:  0.7712864787197146


confusion matrix:
 [[128566     12  32784]
 [   345     40    242]
 [ 40471      4 120464]]


              precision    recall  f1-score   support

           0       0.76      0.80      0.78    161362
           2       0.71      0.06      0.12       627
           4       0.78      0.75      0.77    160939

    accuracy                           0.77    322928
   macro avg       0.75      0.54      0.55    322928
weighted avg       0.77      0.77      0.77    322928
```

Figure 2. Result of accuracy testing for English model

```
accuracy score:  0.779604284439155


confusion matrix:
 [[30294    49 12600]
 [  356  1231   310]
 [10368    21 52323]]


              precision    recall  f1-score   support

           0       0.74      0.71      0.72     42943
           2       0.95      0.65      0.77      1897
           4       0.80      0.83      0.82     62712

    accuracy                           0.78    107552
   macro avg       0.83      0.73      0.77    107552
weighted avg       0.78      0.78      0.78    107552
```

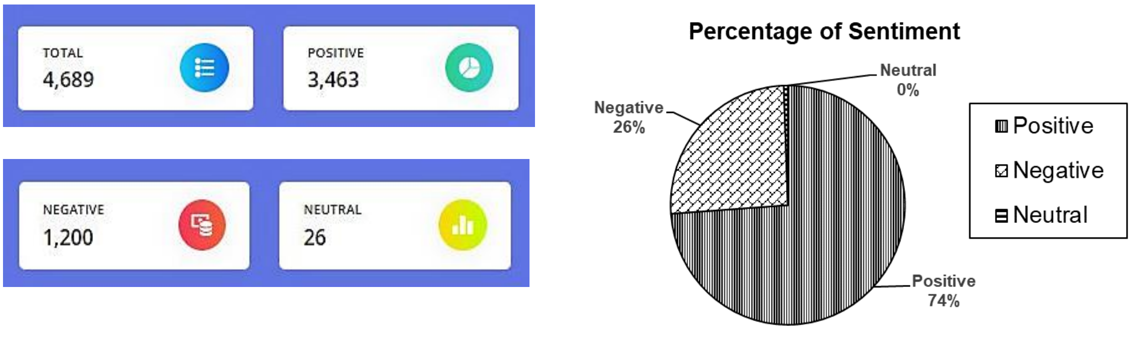
Figure 3. Result of accuracy testing for Malay model

3.2. Overview dashboard visualization

The web-based “Dashboard” page included visualization of the bar charts for the overall sentiment based on factors, pie charts and word clouds for positive, negative, and neutral sentiments. Each of the states has the same dashboard visualization with the details of the hospital name. Also, the comparison between states can be visualized.

3.2.1. Overall sentiment analysis

The system's dashboard plots and displays the complete data analysis. There were visualized using data visualization techniques such as pie charts, bar charts and word cloud for better visions. Figure 4(a) shows the overall sentiments in the selected hospitals from the 14 states. Based on the total of 4,689 sentiments, it was distributed to 3,463 positive, 1,200 negatives and 26 neutral sentiments. The user may immediately compare positive, negative, and neutral sentiments using the pie chart's total sentiment level and the color differences for each, as in Figure 4(b). The dashboard on specific 5 factors classification sentiments is visualized in the bar graph in Figure 5. Users may also visualize the total mentions of each private hospital in each state for each month.



(a) (b)

Figure 4. Dashboard for (a) overall sentiments for 14 states and (b) pie chart of the sentiment's distribution

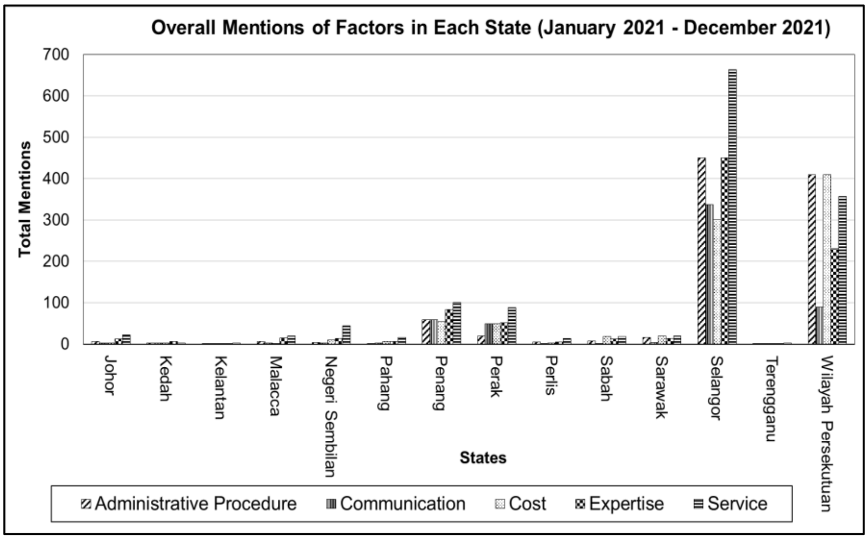


Figure 5. Dashboard for the specific classification sentiments on the identified 5 factors

3.2.2. Visualization sentiment analysis

Figure 6 visualizes the overall classified sentiments using word cloud visualization. The green color word cloud in Figure 6(a) represents the positive sentiment text data such as *sedap selalu*, *dekat* and *tip top*. Next, Figure 6(b) denotes the word cloud of negative sentiment data such as Johor, Medina and Gleneagles, as frequently mentioned on Twitter. The last Figure 6(c) represents neutral sentiment in blue colors such as *dekat*, *swasta* and *vaksin*. Depending on word size, the dataset's word frequency varies. With increasing dataset size, it appears more frequently.

(a) (b) (c)

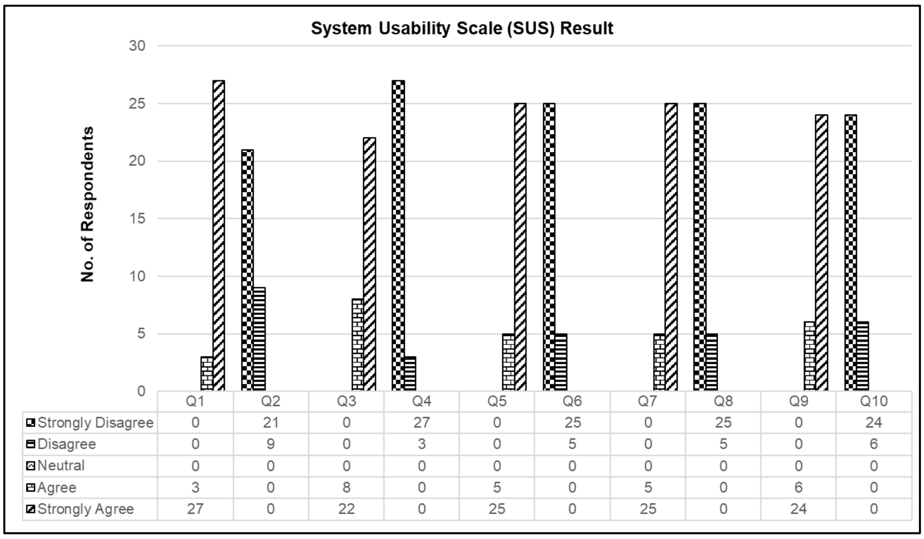
Figure 6. Dashboard for overall classified sentiments using word cloud visualization
(a) positive sentiments, (b) negative sentiments, and (c) neutral sentiments

3.3. Functionality testing

To verify that every system feature is functional, testing is necessary to ensure it functions correctly and that any abnormal system behavior is quickly recognized and solved. Functional testing aims to verify that each system function meets the functional criteria stated in earlier chapters. It is conducted by developing test cases based on system requirements. The findings demonstrate that the system performed as intended without prompted failures. The visualization provides a readily accessible reporting tool, allowing the user to view and comprehend trends and patterns immediately. We succeeded in finishing the dashboard and passing the functionality test.

3.4. Usability testing

A system is subjected to usability testing by a group of representative users to determine its usability. Users are required to evaluate the system's capabilities while being observed to discover whether they encounter any problems. Some suggestions are made to assist users with usability issues. The system usability scale (SUS) consists of 10 user-response questions. Figure 7 shows the ten SUS statements' scores displayed in a bar graph, which depicts the scale of the SUS statements based on user rankings. The graph illustrates that most users selected items with odd numbers, which are positive claims. It shows that customers are satisfied with the system and do not need any technical support to use all of its functions. Users are generally delighted with the system.



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Strongly Disagree	0	21	0	27	0	25	0	25	0	24
Disagree	0	9	0	3	0	5	0	5	0	6
Neutral	0	0	0	0	0	0	0	0	0	0
Agree	3	0	8	0	5	0	5	0	6	0
Strongly Agree	27	0	22	0	25	0	25	0	24	0

Figure 7. Bar chart of SUS result

The SUS scores' histogram is depicted in Figure 8. The frequency of users who responded to the SUS is shown on the histogram's y-axis. In addition, the x-axis displays the percentage of the SUS score range. According to the histogram, the graph demonstrates a normal distribution with a 90% to 100% range and a 2% interval. 11 respondents fall within the peak range between 94% and 96%. Seven respondents are below the median value, and twelve are above the median. The 30 responders to the SUS questionnaire had an average SUS score of 95.42%. If the SUS score is greater than 85, the system is highly usable; between 70 and 85, it is rated good to outstanding; between 50 and 70, it is competent but has some usability concerns that need to be addressed; and below 50, it is termed impractical and inappropriate [33]. With a score above 85%, this web-based application has been verified to be useful. Most respondents gave positive feedback and said they would recommend the product to their friends.

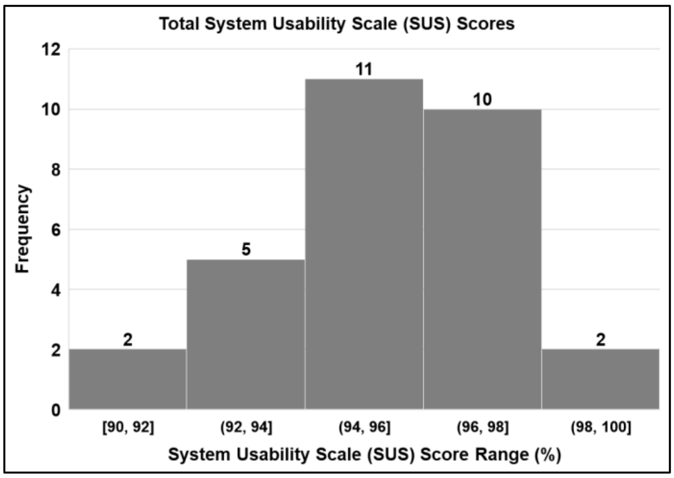


Figure 8. Histogram of SUS result

4. CONCLUSION

Classification and visualization of Malaysia's Private Hospitals based on Twitter sentiment analysis is a web application designed to analyze Twitter users' perceptions and visualize the SA of private hospitals in 14 states in Malaysia. The Nave Bayes classification model developed for this project may be used by the user on any textual data because it is embedded in the system application. The developed platform and application data were able to help anyone to evaluate private hospitals' performance to make decisions in the future. Positive, neutral, and negative classifications were used based on five factors: administrative procedure, communication, cost, expertise, and service. Multiple visualizations within the system application make it simple for customers to comprehend private hospitals in each Malaysian state. The functionality that enables users to watch tweets in real-time from the official Twitter account of private hospitals enables consumers to remain current on the most recent information from private hospitals. In order to interpret slang, abbreviations, and sarcastic words into meaningful values that help determine the sentiment, future studies need to define these terms in dictionaries for different languages.

ACKNOWLEDGEMENTS

The research was funded by Universiti Teknologi MARA Cawangan Melaka under the TEJA Grant 2022 (GDT 2022/1-19).

REFERENCES

[1] S. Swain, "Do patients really perceive better quality of service in private hospitals than public hospitals in India?," *Benchmarking*, vol. 26, no. 2, pp. 590–613, 2019, doi: 10.1108/BIJ-03-2018-0055.
[2] S. N. S. Annuar and N. S. N. Jaffery, "Determining the effect of service quality towards patients' satisfaction in private hospitals in Kota Kinabalu, Sabah Malaysia," *Academic Journal of Business and Social Sciences*, vol. 2, no. 1, pp. 1–15, 2018.
[3] A. I. A. Rahim, M. I. Ibrahim, K. I. Musa, S. L. Chua, and N. M. Yaacob, "Assessing patient-perceived hospital service quality and sentiment in malaysian public hospitals using machine learning and facebook reviews," *International Journal of Environmental Research and Public Health*, vol. 18, no. 18, 2021, doi: 10.3390/ijerph18189912.
[4] I. S. Ahmad, A. A. Bakar, and M. R. Yaakub, "Movie revenue prediction based on purchase intention mining using YouTube trailer reviews," *Information Processing and Management*, vol. 57, no. 5, 2020, doi: 10.1016/j.ipm.2020.102278.
[5] S. Muniapan and D. I. Zakaria, "Healthcare costs in Malaysia," *The Center*, no. October, 2019, [Online]. Available: https://www.centre.my/post/healthcare-costs-in-malaysia.
[6] S. Ahmed, N. H. Abd Manaf, and R. Islam, "Measuring quality performance between public and private hospitals in Malaysia," *International Journal of Quality and Service Sciences*, vol. 9, no. 2, pp. 218–228, 2017, doi: 10.1108/IJQSS-02-2017-0015.
[7] N. Zakaria and A. Wahab, "Customer perception, satisfaction and behavioural intentions towards hospital meal services in government and private hospitals in Alor Setar, Kedah," *Universiti Malaysia Terengganu Journal of Undergraduate Research*, vol. 3, no. 3, pp. 195–206, 2021, doi: 10.46754/umtjur.v3i3.231.

[8] M. B. Khan, "Automatic assessment of performance of hospitals using subjective opinions for sentiment classification," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 420–427, 2020.
[9] V. S. Gupta and S. Kohli, "Twitter sentiment analysis in healthcare using Hadoop and R," in *International Conference on Computing for Sustainable Global Development*, 2016, pp. 3766–3772.
[10] A. S. Huang, A. A. N. Abdullah, K. Chen, and D. Zhu, "Ophthalmology and social media: An in-depth investigation of ophthalmologic content on Instagram," *Clinical Ophthalmology*, vol. 16, pp. 685–694, 2022, doi: 10.2147/OPTH.S353417.
[11] P. Oltulu, A. A. S. R. Mannan, and J. M. Gardner, "Effective use of Twitter and Facebook in pathology practice," *Human Pathology*, vol. 73, pp. 128–143, 2018, doi: 10.1016/j.humpath.2017.12.017.
[12] S. Aldera, A. Emam, M. Al-Qurishi, M. Alrubaian, and A. Alothaim, "Exploratory data analysis and classification of a new Arabic online extremism dataset," *IEEE Access*, vol. 9, pp. 161613–161626, 2021, doi: 10.1109/ACCESS.2021.3132651.
[13] M. M. Mariani, M. Borghi, and U. Gretzel, "Online reviews: Differences by submission device," *Tourism Management*, vol. 70, pp. 295–298, 2019, doi: 10.1016/j.tourman.2018.08.022.
[14] N. Antonio, A. de Almeida, L. Nunes, F. Batista, and R. Ribeiro, "Hotel online reviews: different languages, different opinions," *Information Technology and Tourism*, vol. 18, no. 1–4, pp. 157–185, 2018, doi: 10.1007/s40558-018-0107-x.
[15] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Variable selection for Naïve Bayes classification," *Computers and Operations Research*, vol. 135, 2021, doi: 10.1016/j.cor.2021.105456.
[16] F. Ofori, E. Maina, and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review," *Journal of Information and Technology*, vol. 4, no. 1, pp. 2616–3573, 2020.
[17] K. A. F. A. Samah, N. F. A. Misdan, M. N. H. H. Jono, and L. S. Riza, "The best Malaysian airline companies visualization through bilingual Twitter sentiment analysis: A machine learning classification," *International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 130–137, 2022, doi: 10.30630/joiv.6.1.879.
[18] M. Kazanova, "Sentiment140 dataset with 1.6 million tweets | Kaggle," *Kaggle.Com*, 2017, [Online]. Available: https://www.kaggle.com/kazanova/sentiment140%0Ahttps://www.kaggle.com/kazanova/sentiment140%0Ahttps://www.kaggle.com/kazanova/sentiment140/data#training.1600000.processed.noemoticon.csv.
[19] H. Zolkepli, "Malay-dataset," *Github*, 2022, [Online]. Available: https://github.com/huseinzol05/malay-dataset.
[20] E. Nugraheni, "Indonesian Twitter data pre-processing for the Emotion Recognition," *2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019*, pp. 58–63, 2019, doi: 10.1109/ISRITI48646.2019.9034653.
[21] P. R. Gupte *et al.*, "A guide to pre-processing high-throughput animal tracking data," *Journal of Animal Ecology*, vol. 91, no. 2, pp. 287–307, 2022, doi: 10.1111/1365-2656.13610.
[22] K. Shanmugavadivel *et al.*, "An analysis of machine learning models for sentiment analysis of Tamil code-mixed data," *Computer Speech and Language*, vol. 76, 2022, doi: 10.1016/j.csl.2022.101407.
[23] G. Diaz, "Stopwords-iso," *Github*, 2022, [Online]. Available: https://github.com/stopwords-iso/stopwords-ms/blob/master/stopwords-ms.txt.
[24] H. Zhang *et al.*, "Developing novel computational prediction models for assessing chemical-induced neurotoxicity using naïve Bayes classifier technique," *Food and Chemical Toxicology*, vol. 143, 2020, doi: 10.1016/j.fct.2020.111513.
[25] M. Mu *et al.*, "Prediction of algal bloom occurrence based on the naive Bayesian model considering satellite image pixel differences," *Ecological Indicators*, vol. 124, 2021, doi: 10.1016/j.ecolind.2021.107416.
[26] D. Yan, K. Li, S. Gu, and L. Yang, "Network-based bag-of-words model for text classification," *IEEE Access*, vol. 8, pp. 82641–82652, 2020, doi: 10.1109/ACCESS.2020.2991074.
[27] A. Thakkar and K. Chaudhari, "Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks," *Applied Soft Computing Journal*, vol. 96, 2020, doi: 10.1016/j.asoc.2020.106684.
[28] W. Aljedaani *et al.*, "Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry," *Knowledge-Based Systems*, vol. 255, 2022, doi: 10.1016/j.knosys.2022.109780.
[29] V. Steve Arrul, P. Subramanian, and R. Mafas, "Predicting the football players' market value using neural network model: a data-driven approach," *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2022*, 2022, doi: 10.1109/ICDCECE53908.2022.9792681.
[30] A. Koul, C. Becchio, and A. Cavallo, "Cross-validation approaches for replicability in psychology," *Frontiers in Psychology*, vol. 9, no. JUL, 2018, doi: 10.3389/fpsyg.2018.01117.
[31] A. Nikiforova and J. Bicevskis, "Towards a business process model-based testing of information systems functionality," *ICEIS 2020 - Proceedings of the 22nd International Conference on Enterprise Information Systems*, vol. 2, pp. 322–329, 2020, doi: 10.5220/0009459703220329.
[32] B. Cemellini, P. van Oosterom, R. Thompson, and M. de Vries, "Design, development and usability testing of an LADM compliant 3D Cadastral prototype system," *Land Use Policy*, vol. 98, 2020, doi: 10.1016/j.landusepol.2019.104418.
[33] A. A. Aziz, I. N. Hafawati Adam, J. Jasmis, S. J. Elias, and S. Mansor, "N-gain and system usability scale analysis on game based learning for adult learners," *2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2021*, 2022, doi: 10.1109/ICRAIE52900.2021.9704013.

BIOGRAPHIES OF AUTHORS

Khyrina Airin Fariza Abu Samah is a senior lecturer at the College of Computing, Informatics and Media in Universiti Teknologi MARA (UiTM), Melaka Jasin Campus. Before joining UiTM, she has 13 years of working experience in the semiconductor industry. She has a Diploma, Bachelor's Degree and Master's Degree in Computer Science and PhD in Information Technology. Her research interest is in Artificial Intelligent, Algorithm Analysis, Clustering and Optimization, Evacuation Algorithms, Machine Learning. She can be contacted at email: khyrina783@uitm.edu.my.

Nur Maisarah Nor Azharludin is a Bachelor's student in Computer Science currently undergoing her internship programme. Her research interests are machine learning algorithms, data analysis and visualization. She is actively involved in the research group at UiTM. Currently, she is under an internship at Universiti Teknologi Malaysia. She can be contacted at email: 2019294714@student.uitm.edu.my.

Lala Septem Riza received PhD in Computer Science from Universidad de Granada, Spain in 2015. He works in the Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia. He teaches machine learning, big data platform, and statistical data science. His research interests are in machine learning, data science, and education. He can be contacted at email: lala.s.riza@upi.edu.

Mohd Nor Hajar Hasrol Jono is a Senior Lecturer at the College of Computing, Informatics and Media, Universiti Teknologi MARA (UiTM), Melaka Jasin Campus. Graduated with a PhD in Multimedia Education, Universiti Pendidikan Sultan Idris. MSc Information Technology, UiTM and Bachelor's Degree in Computer Science, Universiti Teknologi Malaysia. Once served as a software developer at a leading company in Malaysia. As an academic, research is one of the areas of interest for developing ideas, sharing skills, and applying the latest innovations according to the latest technology and education trends. He can be contacted via email: hasrol@uitm.edu.my.

Nor Aiza Moketar is a Senior Lecturer in the Faculty of Information Technology and Communication, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia. She holds a PhD and MSc in Software Engineering and Intelligence from Universiti Teknikal Malaysia Melaka and BSc in Information Technology from University Malaya. Before commencing as an academician, she worked as a software developer in a multi-national company. Her research interests are Software Engineering, requirements, testing and Intelligence. She can be contacted at email: nor.aiza@utem.edu.my.