The Best Malaysian Airline Companies Visualization through Bilingual Twitter Sentiment Analysis: A Machine Learning Classification

Khyrina Airin Fariza Abu Samah [a,*], Nur Farhanah Amirah Misdan [a], Mohd Nor Hajar Hasrol Jono [a], Lala Septem Riza [b]

[a] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka Kampus Jasin, Melaka, Malaysia
[b] Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia
Corresponding author: *khyrina783@uitm.edu.my

Abstract—Online reviews are crucial for business growth and customer satisfaction. There is no exception for the airlines' company, which places third as the biggest contributor to Malaysia's Gross Domestic Product. Customer opinions play an important role in maintaining the reputation and improving the quality of service of the airlines. However, there is no specific platform for online review. Most online ratings obtain English, leading to inaccurate results as not all reviews regarding different languages are considered. Airlines currently have no specific platform for online reviews despite being critical for business growth, performance, and customer experience improvement. Hence, this paper proposed implementing a web-based dashboard to visualize the best Malaysian airline companies. The airline companies involved are AirAsia, Malaysia Airlines, and Malindo Air. We designed and developed the proposed study through the bilingual analysis of Twitter sentiment using the Naïve Bayes algorithm. Naïve Bayes algorithm is a machine learning approach to do classification. The tweets extracted were analyzed as metrics that advance airline companies' online presence. Testing phases have shown that the classifier successfully classified tweets' sentiment with 93% accuracy for English and 91% for Bahasa. Every feature in the web-based dashboard functions correctly and visualizes a detailed analysis of sentiment. We applied the System Usability Scale to test the study's usability and managed to get a score of 94.7%. The acceptability score 'acceptable' result concluded that the study reflects a good solution and can assist anyone in understanding the public views on airline companies in Malaysia.

Keywords— Bilingual model; classification; Naïve Bayes; Twitter sentiment analysis; web-based dashboard.

I. INTRODUCTION

An airline is a business that specializes in transporting people and cargo by air. As a result of codeshare agreements, which allow one airline to run the same flight under the code, airlines can provide various services to their customers. An air operating certificate or license granted by a governmental aviation organization is commonly recognized by airlines. The most significant Malaysia's Gross Domestic Product (GDP) contributor is tourism, a third contributor besides the manufacturing and commodities industry [1]. A Gross value-added to tourism industries (GVATI) recorded a contribution of 15.9% is to the GDP compared to the year before with 15% [2]. The category passenger transport shared 3.9%, which came sixth after retail trade, food and beverage, services, accommodation, and recreational. It seems that in the tourism

industry, transport contributes a vital part. In Malaysia, the airline industry dominates the two biggest airline companies, Malaysia Airlines and AirAsia [3]. AirAsia, Malaysia Airlines, and Malindo Air are the top three airlines in Malaysia with the highest annual passenger traffic from 2012 to 2018, excluding the foreign airlines [4]. Skytrax's World Airline Star Rating (WASR) program claims to be a global benchmark for airline standards worldwide, including Malaysia, and utilizes a quality scale from 1 to 5 stars [5]. Besides WASR, there are also Online Travel Agents (OTA), companies in the airline and travel markets such as Kayak, Kiwi.com, Skyscanner, and Skiplagged.

There is a need to focus on the passenger's experience and satisfaction due to an intense battle in the aviation industry [6]. Airlines currently have no specific platform for online reviews despite it being critical for business growth, performance, and customer experience improvement.

Moreover, issues on electronic word of mouth (eWOM) are considered reliable, fast, and widespread because they can easily be transmitted among customers [7].

Even though WASR and OTA can compare the airlines, both have limitations, such as not seeing the direct comparison and only showing ticket prices information. At the same time, it is meaningful to derive the reviews and opinions expressed by an experienced customer to gain insights predict the present direction to move forward and the future [8]. The applications only offer a price comparison for OTA without including other factors. Most consumers also prefer utilizing online travel agencies and meta-search engines to acquire market coverage rather than an exhaustive direct search on individual airline websites [9]. Hence, the outcome achieved might not be satisfying and could lead to a bad airline experience.

Most online ratings obtained their result from the online platform using the English language [9]. This could lead to inaccurate results as not all reviews regarding different languages are considered. Neimann and Montgomery [10] stated that when research teams conduct systematic reviews, most neglected happened to other published studies that use language than English. Researchers had argued about the possibility of bias when excluding non-English studies.

Achieving growth in the airline market requires understanding the factors influencing business travelers to select an airline [11]. Quality passenger services are now the key to airline profitability and growth, especially in this highly competitive environment [12]. The service provided by airlines to their passengers performs a significant part in determining customer's satisfaction. As mentioned on the Skytrax website, there are various services such as inflight meals, inflight entertainment, seat comfortableness, staff services, and ticket pricing. Other than that, there are also other factors, such as flight delays. The flight delays could affect passengers' sentiment as they are more emotionally negative [13], [14].

In the new era of globalization, social media has become one of the most significant communication factors worldwide. Saibaba [15] concluded that the number of social network platform users almost tripled from 2010 to 2020. 4.32 billion mobile internet users, 4.66 billion active internet users, 4.15 billion mobile social media users, and 4.2 billion social media users. Also, Internet Users Survey 2020 conducted by the Malaysian Communications and Multimedia Commission (MCMC)'s stated 28.7 million Internet users in 2020 [16]. It is known that social media is where people can express their opinions, views, and findings without barriers. With 37.1%, Twitter is Malaysia's fourth most popular social networking site. Hence, Twitter sentiment analysis could be one way to analyze the Malaysian airline companies' reputations through data visualization.

Natural language processing, text analysis, computational linguistics, and biometrics are used in sentiment analysis (SA) to systematically analyze, extract, measure, and detect affective states and personal information. Traditionally, SA is about the polarity of the opinion [17]. If people have positive, neutral, or negative opinions about things and a service product whose review is widely expressed on the internet has been the main object of SA.

Thus, SA can be interpreted as one of the best methods to analyze business reputations among customer reviews, expressed in various online platforms regardless of language, age, and social status worldwide. The increasing interest in e-commerce is also a popular source of voicing and analyzing thoughts, feelings, judgments, and feedback. This is also why e-commerce website consumers rely mainly on current customers' feedback. In turn, suppliers and service providers evaluate consumer perceptions to enhance the quality and standards of their goods and services [18].

Therefore, this study involves developing a web-based dashboard in visualizing the performance of the best airline companies in Malaysia through Twitter sentiment analysis. The tweets extracted only cover public opinion regarding Malaysian airline companies' reputations on various factors. The model is built based on English and Bahasa Malaysia datasets to conduct bilingual sentiment analysis. The visualized results can be used to maximize customers' satisfaction and ensure retention. In this way, the consumers will proactively target new potential markets and resolve customer issues more effectively.

II. MATERIAL AND METHOD

The implementation is divided into four parts for the research method: system design, back-end development, front-end development, and testing development.

A. System Design

The study starts with designing the overall web-based dashboard flow, the use case diagram, and the user interface. The design process helps ensure that the research aims can be achieved and realistic.

B. Back-End Development

Fig. 1 shows the research design for back-end development. The process involved a complete flow started with data collection, data pre-processing, NB classifier, evaluation performance metrics, and model deployment of the web-based dashboard.

1) Data Collection: Machine learning models for Malay and English languages are built to classify the text. The dataset for the English model is collected from https://www.kaggle.com/kazanova/sentiment140, which contains 800,000 negative data and 800,000 positive data. In contrast, the Malay dataset is gathered from https://github.com/huseinzol05/malay-dataset/ tree/master/ sentiment/translate/twitter-sentiment for the training and testing of the Malay model. From the dataset, 344,733 negative data and 312,985 positive data are obtained. However, both datasets collected for the English and Malay model only contain binary classification data, positive and negative, as in Table 1.
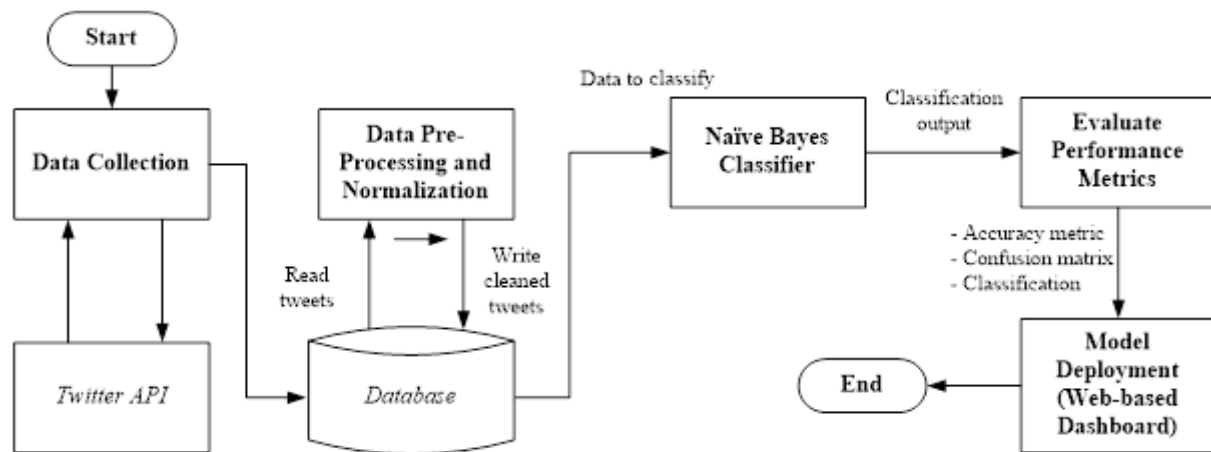
Fig. 1 Research Design for Back-End Development

Model	Data Description		
English	Source	https://www.kaggle.com/kazanova/sentiment140 https://www.kaggle.com/kritanjalijain/twitter-sentiment-analysis-lstm/#data	
	Number of Data	Positive	800,000
		Negative	800,000
		Neutral	75,046
		Total	1,675,046
Malay	Source	https://github.com/huseinzol05/Malay-Dataset	
	Number of Data	Positive	312,985
		Negative	344,733
		Neutral	-
		Total	657,718

Therefore, the additional dataset is collected from https://www.kaggle.com/kritanjalijain/twitter-sentiment-analysis-lstm/#data to add neutral data for training and testing data for both models. However, only 75,046 data were discovered in the dataset. For the additional neutral data to the Malay model, the neutral data gathered for the English model is translated into Malay using https://www.onlinedoctranslator.com/, which uses Google Translate to translate the document. The English model's total training and testing data are 1,675,046 data and 657,718 data for the Malay model.

Twitter has many Malaysian users, most of whom post and write their tweets to share their opinions using the Malay language. The data for the three companies, AirAsia, Malaysia Airlines, and Malindo Air, are extracted from Twitter to obtain real-world data. The scrapped tweets start from January 1, 2019, until December 31, 2019 using the keywords 'AirAsia', 'Malaysia Airlines', and 'Malindo Air' refer to AirAsia, Malaysia Airlines, and Malindo Air companies, respectively. The date chosen is before the COVID-19 pandemic to get accurate data. For more data obtained, hashtags of the companies' names were also used to scrap the tweet: '#AirAsia', '#MalaysiaAirlines', and '#MalindoAir'. The tweets are scrapped using Twint in Spyder software, where the keywords typed are not case sensitive. The scrapped tweets are saved as JSON (JavaScript Object Notation) file type. Table II shows the total collected data is 7,410 data for AirAsia, 1,829 data for Malaysia Airlines, and 400 data for Malindo Air.

Month	Air Asia			Malaysia Airlines			Malindo Air		
	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
Jan	250	197	180	52	23	29	12	6	14
Feb	252	184	192	75	21	25	17	6	5
Mar	293	252	253	67	58	73	19	9	10
Apr	235	163	194	51	31	34	13	4	14
May	207	169	177	63	23	26	14	2	10
Jun	265	169	192	65	99	81	14	9	10
Jul	323	172	218	80	57	47	14	7	10
Aug	289	162	197	68	20	21	9	3	11
Sep	245	135	174	91	50	66	12	31	17
Oct	213	137	208	84	40	53	35	21	7
Nov	226	151	210	53	53	40	13	2	6
Dec	227	115	184	58	23	29	10	4	14
Total	3025	2006	2379	807	498	524	182	104	114

2) Data Pre-Processing: Text data need to be transformed into numerical form to do Machine Learning. The real-world extracted data were encoded numerically using encoding techniques such as Bi-gram, n-gram, TF-IDF, BagOfWord, and Word2Vec. Data cleaning is done before the encoding process is called text pre-processing. Text pre-

processing is the process of cleaning and preparing the data by removing and eliminating unnecessary features in the data. There is no added valuable meaning but would only interrupt the quality of the developed model.

The two Python libraries used in the text pre-processing task are NLTK and re. The final dataset consists of only four columns: data, username, tweet, and language, where the unwanted columns are removed from the dataset. Thus, dataset text cleaning is done by converting all characters to lowercase. It is to avoid case-sensitive issues while pre-processing. Next, the step continued by removing unwanted characters such as emojis, punctuation marks, and extra whitespaces. The terms such as links, hashtags, and mentions in the tweets are also removed. Besides, the duplicate tweets and null values in the dataset have been dropped to reduce the data's dimensionality further.

However, this dataset is still high-dimensional data. To reduce the dimensionality, stop words are removed from the data as no additional valuable meaning. The words such as 'the,' 'and, 'of, and 'on' are considered stop words in English. Stop words in English can be obtained in the pre-built function in the NLTK library. However, stop words from the Malay language are manually imported from https://github.com/stopwords-iso/stopwords-ms/blob/master/stopwords-ms.txt for removal of stop words in the Malay model.

3) Naïve Bayes Classifier: Machine learning (ML) uses computers to simulate human understanding [19]. The model evaluation is conducted using the Naïve Bayes (NB) classifier, which is used to classify the dataset. After learning the pre-labeled data from the training set, the model classifies the dataset. The NB theorem calculates an event's probability based on the probabilistic joint distribution of other events [20]. NB is chosen as the ML algorithm to build the model. It is the most effective algorithm that works quickly, saves a lot of time, and is suitable for multi-class prediction problems [21]. In this project, the pre-labeled training dataset fed into the model assists the model in learning the context for positive, neutral, and negative sentences.

The text representation represents the collection of phrases and words structured and calculates the number of occurrences of the phrase called Bag of Words (BOW). It extracts the feature by turning the tokenized words into a vector that a machine learning model can learn. The three steps from the BOW are used to count the term frequency, count the inverse document frequency, and normalize the vectors to unit length. The first two steps of BOW combined are known as Term frequency-inverse document frequency (TF-IDF). It is a statistical measure to determine the word's importance in the document [22]. In information retrieval and text mining, we use the TF-IDF weight. TF measured the frequency appearance of a term in a document; meanwhile, IDF measured the significance level. The formula for calculating TF and IDF is (1) and (2), respectively. The formula for calculating TF and IDF is (1) and (2), respectively.

$$TF(t) = \frac{(Number\ of\ times\ term\ appears\ in\ document)}{(Total\ number\ of\ terms\ in\ the\ document)} \quad (1)$$

$$IDF(t) = log_e \frac{Total\ number\ of\ documents)}{(Number\ of\ documents\ with\ term\ t\ in\ it)} \quad (2)$$

We use the training data for the cross-validation process to ensure whether the model is overfitting the data. To partition the model into random sections, we evaluate different hyperparameter setups. There are eight tested parameters configurations and 10 KFold validations for the model. Thus, the model was trained and tested 80 times.

4) Evaluation Performance Metrics: In the next step, we evaluate the model's performance on the test holdout dataset in the real-world implementation. From the evaluation, the performance metrics obtained are a classification report and a confusion matrix. The accuracy metric, confusion matrix, and classification report results are observed. Lastly, before the data visualization process, the data acquired are run with the model through Twitter Application Programming Interface (API) for sentiment predictions, and its performance is evaluated.

5) Model Deployment of Web-Based Dashboard: Model deployment is the task of exposing an ML model to actual use. The term is often used synonymously to make a model available via real-time APIs. The predicted classified tweets are produced at the model deployment stage with '0', '2', and '4', which indicate negative, neutral, and positive sentiments.

After sentiment predictions using the model classifier on the data obtained are made and followed by evaluating its performance. We use Plotly, an open-source interactive graphics toolkit for Python, to visualize it. The step started by importing the data obtained into the Pandas data frames in Python. The charts are generated along with the details entered through some coding. The result obtained from the analysis is then used to create an interactive visualization tool to illustrate the outcomes of real-world data analysis.

Data visualization using word cloud visualizes the text data for the airline companies of the system. The words used different colors, and the word's size emphasizes the frequency of the words that appear in the text data. In addition, the words are arranged in a cloud form, showing the terms used for airline companies most simply and easily. The following subsection discussed the front-end development of the system.

C. Front-End Development

Front-end web development, or alternatively referred to as client-side development, converts data to a graphical interface through HTML, CSS, and JavaScript as the language to build a website to view and interact with that data [23]. The web application environment for Python includes the data visualization tools to create the charts and graphs of the sentiment data. There are six modules involved in the developments: overview page, AirAsia dashboard, Malindo Air dashboard, Malaysia Airlines dashboard, real-time Twitter updates page and competitive analysis page.

D. Testing Development

Once the previous phases are complete, the system considers almost done and needs to perform the testing. We validate the development through two methods: functionality and usability of the web-based dashboard.

1) Functionality Test: This phase aims to verify the whole system's functionality and whether it complies with the

previous phase's requirements. The navigation menu is tested based on the six modules developed in front-end development. Functions included in the test cases are tested by entering the input and examining the output to check whether the functions are successful.

2) Usability Test: A usability test is conducted to discover the system workflow and determine whether the information is correctly formatted and delivered. Improvements to the system can improve user experience towards the system besides designing better system workflows. It is also important to recognize whether the user has found an alternative way to access the system abnormally caused improper development. Therefore, the System Usability Scale (SUS) determines its effectiveness from the user's perspective. The SUS is a reliable and practical tool to measure the usability of a product [24], [25]. It can be used to see if there is a problem with the design of digital products and services. The user needs to rate the system by answering a set of questions by rating each question on a scale of one (lowest) to five (highest). There were 30 random respondents interested in evaluating the airline companies' performance. They were instructed to complete specific tasks to evaluate each system's features using a Google Form questionnaire. The form included user information, including their name, the most recommended airlines, and the user's experience using the application.

III. RESULTS AND DISCUSSION

We divided the result into four parts: accuracy test, overview dashboard visualization, the functionality test, and usability of the system.

A. Accuracy Test

We obtained the Naïve Bayes Classifier Model's testing accuracy from the simple coding of Python. The result in Fig. 2 shows the testing accuracy for the English model with a score of 93% in percentage form. It indicates that the sentiment result is 93% correct, which means that the model correctly classified nine correct results into 'positive', 'neutral' or 'negative' out of 10 attempts. From the confusion matrix, 0 represents the 'negative' class, 2 represents the 'neutral' class, and the 'positive' class is 4.

```
accuracy score:  0.9320107241843641

confusion matrix:
[[150415     24    8095]
 [   351  14210    406]
 [ 13614     29  144070]]

              precision    recall  f1-score   support

           0       0.92      0.95      0.93    158534
           2       1.00      0.95      0.97     14967
           4       0.94      0.91      0.93    157713

    accuracy                           0.93    331214
   macro avg       0.95      0.94      0.94    331214
weighted avg       0.93      0.93      0.93    331214
```
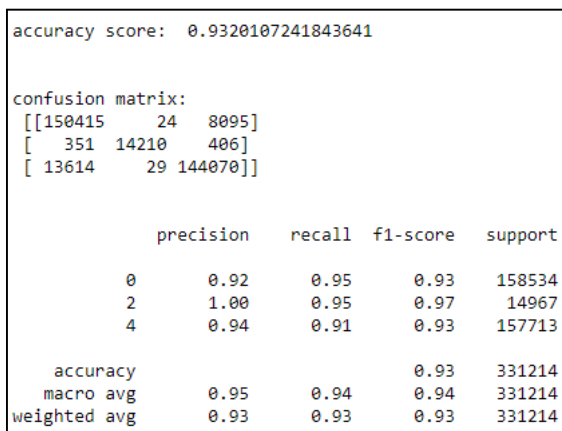
Fig. 2 Result of Accuracy Testing for English Model

Fig. 3 shows the accuracy score for the Malay model. In the confusion matrix, similar to the English model, the accuracy score is 91% in percentage form. It indicates that the result of the sentiment is 91% correct, where the model was able to correctly classify nine correct results into 'positive', 'neutral' or 'negative' out of 10 attempts.
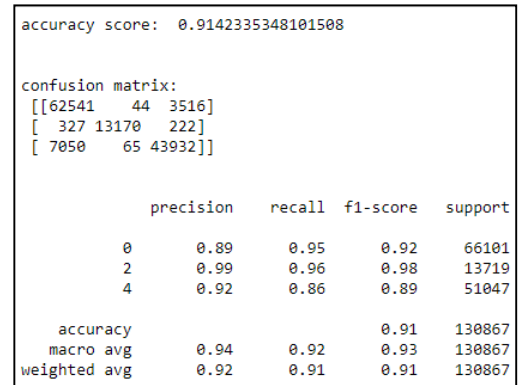
```
accuracy score:  0.9142335348101508

confusion matrix:
[[62541     44   3516]
 [  327  13170    222]
 [ 7050     65  43932]]

              precision    recall  f1-score   support

           0       0.89      0.95      0.92     66101
           2       0.99      0.96      0.98     13719
           4       0.92      0.86      0.89     51047

    accuracy                           0.91    130867
   macro avg       0.94      0.92      0.93    130867
weighted avg       0.92      0.91      0.91    130867
```

Fig. 3 Result of Accuracy Testing for Malay Model

B. Overview Dashboard Visualization

Real-world data analysis is plotted and illustrated on the system's dashboard. The data is displayed in different visualizations such as bar charts, pie charts, gauge charts, and word clouds for better insights. Table III shows the Brand Talkable Favorability (BTF) score for the airline companies, AirAsia, Malindo Air, and Malaysia Airlines. Despite the values being quite similar and not substantial, AirAsia has the worst BTF, 35.8% compared to 43.0% for Malindo Air and 42.7% for Malaysia Airlines.

TABLE III
CLASSIFICATION COMPARISON FOR AIRASIA, MALINDO AIR, AND MALAYSIA AIRLINES

Airline Company	Total Mentions	Total Positive Mention	Total Neutral Mention	Total Negative Mention	Brand Talkability Favorable
AirAsia	7,410	3,025	2,006	2,379	35.8%
Malindo Air	400	182	104	114	43.0%
Malaysia Airlines	1,829	807	498	524	42.7%

1) Overall Brand Talkable Favorability (BTF): The top dashboard reveals the BTF for each airline company in 2019 through Twitter mentions as in Fig. 4.

Fig. 4 Overall BTF for the Airline Companies

2) Overall Sentiment: All the airline companies have higher positive mentions than neutral and negative mentions, as depicted in Fig. 5, bar chart of the overall sentiment of companies in 2019 through Twitter mentions.
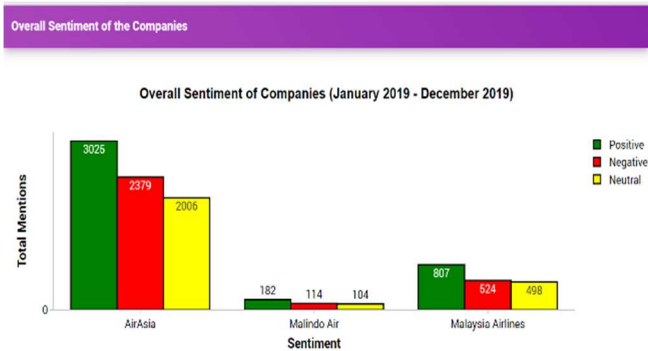
Fig. 5 Bar Graph: Overall Sentiments for the Airline Companies

3) Overall Total Mentions: Series plot of all airlines companies totals mentions on Twitter by month as in Fig. 6.
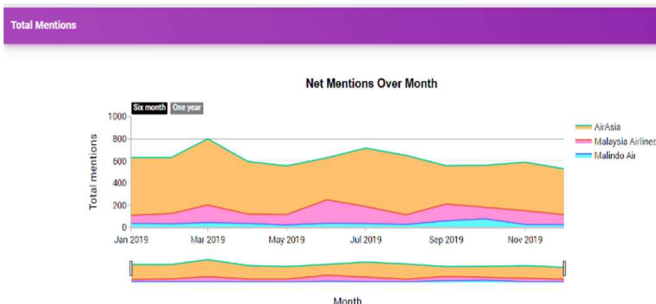

Fig. 6 Series Plot of All Airlines Companies

4) Overall Word Cloud: Fig. 7 shows the overall word cloud for the airline companies' classified sentiment analysis. The positive sentiment text data illustrates in graphical representation using word cloud using green color. The word's size differentiates the word frequency in the dataset. The larger the size, the more frequent it appears in the dataset. Some of the words related to the positive sentiment that appears in the word cloud are 'good', 'best', and 'great'. The red color represents the words related to the negative sentiment. The words 'delay', 'not', and 'no' appears in the word cloud. Lastly, the word cloud for neutral sentiment in yellow color represents the words related to the neutral mention.

Fig. 7 Word cloud: Overall classified sentiments for the airline companies

5) Overall Statistics: Each dashboard for airline companies consists of three visualizations: 1) net sentiment

by month using a bar chart or line chart, 2) overall sentiment level, and 3) sentiment breakdown. Fig. 8 shows the captured dashboard for AirAsia.

Fig. 8 Overall Statistics Dashboard Visualization for AirAsia

6) Word Cloud Specific Airline Companies: The word cloud obtained in Fig. 9 represents the word in AirAsia's company text corpus. The bar chart depicts the words that are related to positive and negative sentiments.
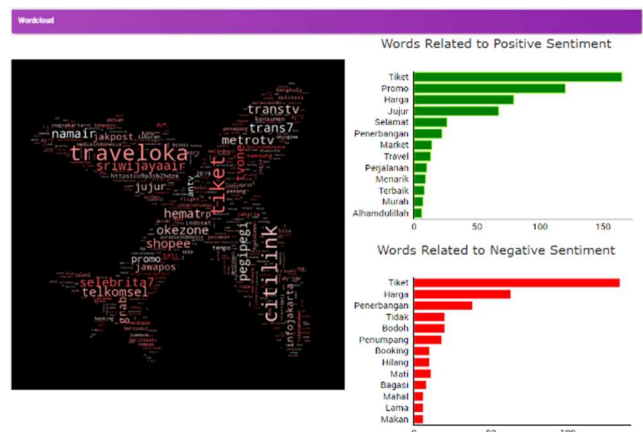

Fig. 9 Word cloud and bar chart of sentiment words for AirAsia

7) Compare Performance of Companies: The system displays the competitive analysis page to the user. The user can compare the performance between companies of their choice by clicking the number provided in Fig. 10. Each selection compares airline companies for total mentions, sentiment breakdown and line chart for 'positive', 'negative' and 'neutral'.

Fig. 10 Interface of competitive analysis

C. Functionality Test

Testing is important to ensure that all the system's features operate correctly and quickly identify and fix any odd behavior by the system. Also, to test each function of the system to determine the functional requirements in the research method, work accordingly. We managed to complete the dashboard and successfully passed the functionality test.

D. Usability Test

The ten SUS statements score is visualized in a bar chart, as in Fig. 11, where it shows the scale of the SUS statements according to the users given a ranking. The plotted graph shows that most users choose scale 5, representing "strongly agree" for the odd-numbered questions, the positive statements. In contrast, the even-numbered questions received mostly scale 1, showing that the users "strongly disagree" with the negative statements in the questions. It indicates that the users have a good experience with the system and do not require any technical assistance to use its features. In general, the users are all satisfied with the system.
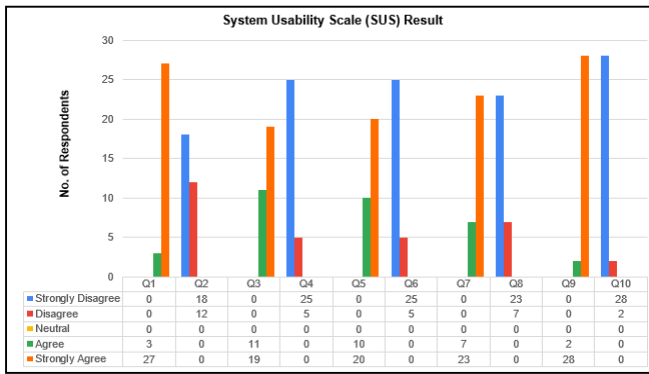

Fig. 12 Bar chart of SUS result

Fig. 12 illustrates the histogram of the SUS scores. The *y-axis* of the histogram represents the frequency of users that answered the SUS, while the *x-axis* shows the percentage of the SUS score range. Based on the histogram, the data spreads between 90% to 100%. The graph has a normal distribution with a range starting from 90% until 100% with a 2% interval. The histogram peak is at 94% to 96%, where it contains the highest number of respondents, which are 12 respondents. Ten respondents fall below the central value, and eight respondents fall above the central value.
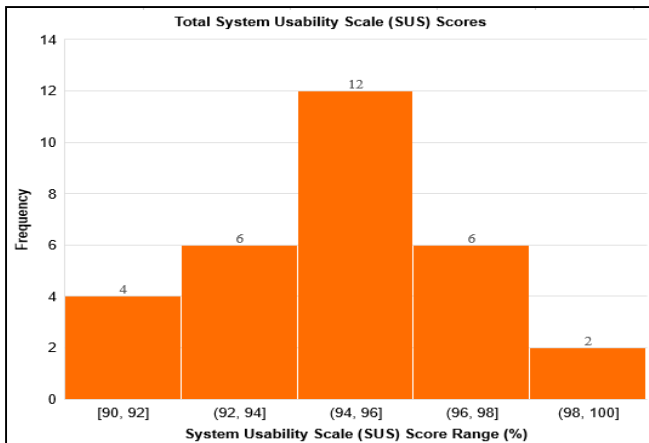

Fig. 12 Histogram of SUS scores

Overall, the SUS score of 94.7% is acquired from the 30 selected respondents in the SUS questionnaire. The SUS average score's baseline is 68%. A score below 68% is considered below average, and more than 68% is considered above average. The scores under baseline will need to issue with the design to be resolved and researched more.

IV. CONCLUSION

The web-based dashboard was developed to visualize the bilingual sentiment analysis of Twitter users' perception of Malaysia Airlines, AirAsia, and Malindo Air from January 1, 2019, to December 31, 2019. The user can utilize the Naïve Bayes Classifier model built for this project on any textual data because it is embedded in the system application. Three classifications were used, 'positive', 'neutral' and 'negative'. The information obtained from the system application can assist the user in determining the performance of the companies and help in decision-making for future use. The various visualizations provided in the system application make it easy for the user to gain better insights into each company. For future study, the dictionary for various languages such as slang, short forms, and sarcastic words should be defined to interpret them into meaningful values that could help determine the sentiment.

REFERENCES

[1] R. Hirschmann, "Travel and tourism in Malaysia – statistics and facts," 2020. https://www.statista.com/topics/5741/travel-and-tourism-in-malaysia/ (accessed Jul. 31, 2021).

[2] M. U. Mahidin, "Tourism satellite account 2019," 2020.

[3] D. R. Tobergte and S. Curtis, "Factors That Influence Customer Satisfaction and Behavioural Intentions in Airline Industry," *Journal of Chemical Information and Modeling*, vol. 53, no. 9, pp. 1689–1699, 2013.

[4] Y. Mamrina, D. M. H. Kee, A. Tursynbay, H. A. B. M. Athiyah, and R. Muhamach, "A case study on critical success factors of AirAsia," *International Journal of Tourism and Hospitality in Asia Pasific*, vol. 3, no. 1, pp. 77–88, 2020.

[5] N. Avogadro, P. Malighetti, R. Redondi, and A. Salanti, "A tale of airline competition: When full-service carriers undercut low-cost carriers fares," *Journal of Air Transport Management*, vol. 92, pp. 1–10, 2021.

[6] W. Messner, "Value or quality? Differences in airlines' customer satisfaction strategies across national markets," *Services Marketing Quarterly*, vol. 41, no. 3, pp. 205–235, 2020.

[7] H. J. Ban and H. S. Kim, "Understanding customer experience and satisfaction through airline passengers' online review," *Sustainability (Switzerland)*, vol. 11, no. 15, 2019.

[8] H. J. Kwon, H. J. Ban, J. K. Jun, and H. S. Kim, "Topic modeling and sentiment analysis of online review for airlines," *Information (Switzerland)*, vol. 12, no. 2, pp. 1–14, 2021.

[9] K. A. F. A. Samah, A. A. S. A. Sani, N. Sabri, S. Ibrahim, A. F. A. Fadzil, and L. S. Riza, "Visualizing communication of service providers reputation during COVID-19 pandemic: A conceptual model," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.4 Special Issue, pp. 558–568, 2020.

[10] L. Neimann Rasmussen and P. Montgomery, "The prevalence of and factors associated with inclusion of non-English language studies in Campbell systematic reviews: A survey and meta-epidemiological study," *Systematic Reviews*, vol. 7, no. 1, pp. 1–12, 2018.

[11] B. M. Zaharias, "Analyzing Customer Satisfaction," *Review of Indonesian Economic and Business Studies*, vol. 7, no. 1, pp. 1–12, 2016.

[12] N. Murugan, "Influence of socio demographic factors in the choice of airline between air asia & malaysia airlines," in *Asean Tourism Research Conference*, 2017, pp. 187–194.

[13] C. Song, J. Guo, and J. Zhuang, "Analyzing passengers' emotions following flight delays- a 2011–2019 case study on SKYTRAX comments," *Journal of Air Transport Management*, vol. 89. pp. 1–7, 2020.

[14] C. Raynes and K. W. H. Tsui, "Review of Airline-within-Airline strategy: Case studies of the Singapore Airlines Group and Qantas Group," *Case Studies on Transport Policy*, vol. 7, no. 1, pp. 150–165, 2019.

[15] S. Saibaba, "Bibliometric analysis and visualization of online consumer behavior – past, present and future," *Saudi Journal of Business and Management Studies*, vol. 6, no. 8, pp. 268–291, 2021.

[16] M. C. and M. Commission, "Internet user survey 2020," 2020.

[17] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018.

[18] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1495–1545, 2019.

[19] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Systems with Applications*, vol. 97, pp. 205–227, 2018.

[20] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings*, 2017, pp. 900–903.

[21] M. J. Sánchez-Franco, A. Navarro-García, and F. J. Rondán-Cataluña, "A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services," *Journal of Business Research*, vol. 101, pp. 499–506, 2019.

[22] S. Qaiser and R. Ali, "Text mining: Use of tf-idf to examine the relevance of words to documents," *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25–29, 2018.

[23] A. Holzer and J. Ondrus, "Mobile app development: Native or web?," *Tematics and Informatics*, vol. 23, no. 1, pp. 22–31, 2011.

[24] A. Kaya, R. Ozturk, and C. A. Gumussoy, "Usability neasurement of nobile applications with system usability scale (SUS)," in *Lecture Notes in Management and Industrial Engineering*, 2019, no. January, pp. 389–400.

[25] A. Smyk, "The system usability scale & how it's used in UX," 2020. https://xd.adobe.com/ideas/process/user-testing/sus-system-usability-scale-ux/ (accessed Sep. 02, 2021).