SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Aspect-Based Classification and Visualization of Twitter Sentiment Analysis Towards Online Food Delivery Services in Malaysia

Khyrina Airin Fariza Abu Samah[1,*], Nur Shahirah Jailani[1], Raseeda Hamzah[1], Raihah Aminuddin[1], Nor Afirdaus Zainal Abidin[1], Lala Septem Riza[2]

[1] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Melaka Branch, Jasin Campus, Melaka, Malaysia
[2] Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

ARTICLE INFO

Article history:
Received 22 June 2023
Received in revised form 6 September 2023
Accepted 29 October 2023
Available online 9 January 2024

Keywords:
Online Food Delivery; Twitter Sentiment Analysis; Aspect-Based; Naïve Bayes

ABSTRACT

Twitter has become a popular platform for the citizens of Malaysia. Twitter's ease of expressing opinions could be used to evaluate and review Malaysian Online Food Delivery (OFD) providers. Due to competition from other OFDs in Malaysia, companies need to know customer feedback. OFD reviews are unstructured and massive, making comparisons difficult. Next, some websites evaluate OFD yet only consider pricing, delivery time, and customer experience. Customers cannot visualize the comparison based on users' preferences, bilingual reviews, and it is less time-consuming to visualize OFD using the website. Thus, this study aims to design a web application system that uses Naïve Bayes to categorize Twitter sentiment analysis (SA) on Malaysia's best OFD. It is based on customer satisfaction, visualizing the results, developing the system, and evaluating its accuracy, functioning and usability. Users can read about specific OFD by viewing Twitter SA visualization or comparing them directly. Five aspect-based SA types were presented: affordable price, promotion and discount, review rating, delivery time and condition of food delivered. Functionality testing demonstrated the accomplishment of all objectives. The training and testing data could predict OFD's Twitter sentiment with 71.67% and 76.29% accuracy for English and Bahasa Melayu, respectively. The system's usability produced a 94.64% average score using System Usability Scale and was considered "excellent". Thus, it can be concluded that this study can solve the mentioned issues of OFD and ease the aspect-based comparison.

1. Introduction

Online food delivery (OFD) services offer the opportunity to order and have food made away from home delivered to the customer's door [1]. Customer service includes everything a firm does to ensure the satisfaction of its customers. They help to increase profits from sold products [2]. Customer satisfaction is a metric that assesses customers' satisfaction with the services they receive [3]. As a result, customer satisfaction and loyalty are the most important aspects in determining the market's success. A successful firm is built on customer happiness. Satisfying customers gives the

[*] *Corresponding author.*
E-mail address: khyrina783@uitm.edu.my

https://doi.org/10.37934/araset.37.1.139150

company a competitive advantage over competitors and helps the company make money [2]. Companies that seek to increase market share must supply customers with valuable and distinctive terms to defeat the competition. Customer satisfaction was found to be positively connected with financial performance [4].

Advanced information technology has revolutionized communication by allowing users to quickly obtain information and share their opinions on products and services on a big scale in real-time. Users can now express themselves through social media and review websites [3]. The effect of online reviews is highly dependent on their qualities. Negative reviews are more persuasive than good ones. Twitter is a popular method for Malaysians to voice their thoughts and views about online meal delivery. Twitter's ease of use in delivering opinions could be a chance to be used as an assessment and review tool for OFD services in Malaysia. Due to competition from other OFD services in Malaysia, they need to know what their customers think about their service. Based on this, companies may enhance their products, services, and marketing strategies, detect new trends, and measure the effectiveness of their advertising strategies [3].

In Malaysia, the OFD business is progressively growing in popularity. Before the COVID-19 outbreak, technological businesses such as GrabFood, Foodpanda and ShopeeFood helped to make OFD in Malaysia a growing trend in metropolitan areas [5]. The trend has increased using online food ordering apps, particularly considering the new COVID-19 pandemic standard of procedure. Food delivery services have proven to be a blessing to those who cannot travel or leave hectic schedules but would want food delivered in front of the door [6].

A website like *Vulcanpost.com* compared OFD services between ShopeeFood, GrabFood, Foodpanda, and AirasiaFood. Although it can visualize the comparison, only three comparisons are available. It uses characteristics such as price, time taken to arrive, and user experience. The user cannot visualize the comparison directly. OFDs such as Foodpanda, Bungkusit and GrabFood compete against each other in terms of features and services. They also have to compete against each other in terms of price, quality, and customer service [7]. Integrating online-to-offline (O2O) platforms introduces third-party factors that restaurants cannot control. A courier's attitude might influence customer perception and behavior. Most customers are drawn to O2O platforms because of their inexpensive pricing and ability to compare costs [8]. Han and Anderson [4] claimed that a customer satisfaction survey is a set of questions designed to help businesses learn how satisfied customers feel with various aspects of the company, including the quality of the products and services offered, the reputation of the brand, and the responsiveness of the customer service staff.

Many sectors actively engage with online reviews since they significantly influence customer decision-making. However, one disadvantage of online reviews is that they are usually available in English. Not all reviews in other languages are analyzed, which could lead to inaccurate results. When research teams perform systematic reviews, studies published in languages other than English are usually neglected [9-10]. Other than that, online reviews are naturally multilingual and multicultural. Hence, an analysis based on a single language risks not capturing the whole findings. In addition, significant obstacles may limit the full utilization of these data [11].

An eight-question survey was conducted using Google Form from 17th April 2022 until 24th April 2022. The survey was distributed to 113 respondents among friends and family members through WhatsApp and Twitter. It was designed to gather feedback from current and potential users of OFD services in Malaysia. Based on the survey of 113 respondents currently using OFD services, 91.2% agreed that it is time-consuming to manually compare Malaysia's OFD services. Other than that, 59.3% agreed that affordable price is the main aspect that users want to know regarding OFD services.

As a result, this research presents the development of a web-based application system to measure the effectiveness of Malaysian OFD using sentiment analysis (SA) from Twitter. The collected tweets discuss how the general public views Malaysian OFD based on five aspect-based sentiment analysis (ABSA): affordable price, promotion and discount, review rating, delivery time and condition of food delivered. In Malaysia, three OFD services are included in the study's scope.

We used Naive Bayes (NB), a simple learning technique based on Bayes' rule and the fundamental presumption that a class's characteristics are conditionally independent. For machine learning (ML) and data mining, NB is one of the best and most effective inductive learning algorithms [12]. The NB classifier assessed the model using an algorithm to categorize the dataset [13]. The model uses the labeled data from the training set on the dataset for classification. The benefit of NB is that it can estimate the classification-related parameters with minimal training data. The effectiveness of NB has been demonstrated across many domains [14].

For obtaining meaningful information, data visualization is an essential tool [15]. The author claimed that data visualization uses computer visuals to depict patterns, trends, and relationships between data pieces. As the data gathered from Twitter is more successful in displaying, this study used four different visualization techniques: a line chart, a bar chart, a pie chart, and word clouds. Large data sets make it simpler to spot trends, patterns, and outliers. The data is shown using Plotly after the model classifier makes sentiment predictions on the collected data and assesses its efficacy. A Python interactive graphics package called Plotly is free and open-source.

The model was developed using datasets in English and Bahasa Malaysia to analyze sentiment in both languages. The results can be used to boost customer satisfaction and keep them around. As a result, customers will seek out potential new markets and properly handle customer difficulties. The structure of this document is as follows: The first section is an introduction. The second section explains the research methods. Section 3 focuses on the result and discussion. Finally, section 4 concludes the analysis by quickly noting possible future improvements.

2. Methodology
2.1 System Design

In order to support the methodology, explanations and justifications for the research design and data collection techniques are provided. System design includes requirements, architecture, product design, modules, interfaces, and data. It included the system's flowchart, use case, and user interface.

2.2 Back-end Development

The general development of the web-based dashboard is shown in the research design in Figure 1. The study's methodology was separated into four parts for further explanation. Back-end development is the process of working on server-side software, which concentrates on website elements concealed from view throughout system development. The system's back end, which uses the NB algorithm, is implemented in Python and handles everything from data preparation through model deployment. Data collection, pre-processing, NB classification model building, and model deployment are significant back-end tasks for training and testing data.

2.2.1 Data Collection

The dataset for the three OFD services, Foodpanda, GrabFood, and ShopeeFood, was obtained from Twitter using a Python Twitter scraping application called Snscrape. It was done in order to get data from the actual world. The scope of the analyzed tweets is restricted to the general public's perspective on various aspects of OFD. The time range covered by the gathered tweets is from 1st January 2022 to 31st December 2022.

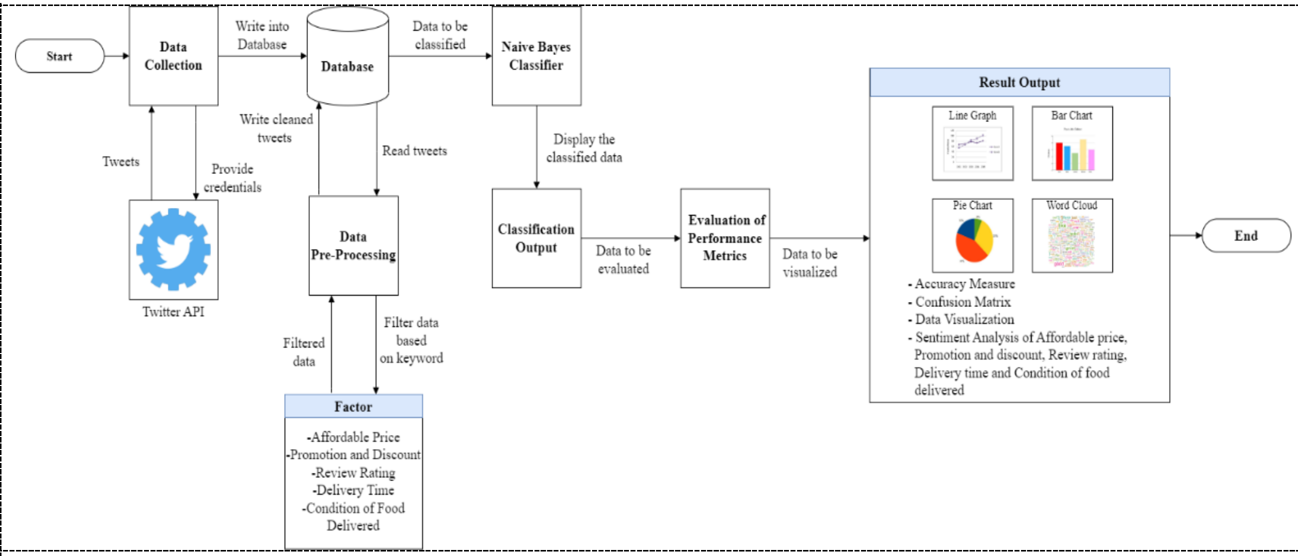


Fig. 1. Flow Diagram of Research Design

Following the scraping of data, the data that has been collected is then manually analyzed in order to get rid of empty cells in the tweet column. There are 4,677 raw data collected in English and Malay, 1,922 data collected for Foodpanda, 1,640 for GrabFood, and 1,115 for ShopeeFood. The data extracted for each OFD takes about three minutes to complete. The raw data includes six variables: the date the tweet was created, the username, the name, the tweet, the language, and the URL link. After scraping, the data is saved in a file type known as comma-separated values (CSV). When reading the CSV file, the Pandas library is used.

2.2.2 Data Pre-processing

Before beginning the encoding process, the text is cleaned up by a process known as pre-processing [16]. The data cleaning and preparation process during text pre-processing is called "pre-processing" [10]. Removing and deleting any features that are irrelevant to the analysis of the data. There is no use in adding any value to the model because doing so will lower its quality.

The work utilized two Python libraries called "NLTK" and "re" to perform text pre-processing. The completed dataset consists of only three columns: labeled data created, username, and tweet. Text cleaning is conducted on the dataset by changing all characters to lowercase. It is done to avoid any case-sensitive issues that may arise during pre-processing. After that, characters considered superfluous, such as emojis, punctuation marks, and excessive whitespace, were eliminated. The elimination process includes removing terms such as links, hashtags, and mentions. In addition, null values and duplicate tweets included in the dataset were deleted so that the dimensionality of the data could be decreased even further.

However, this dataset is still considered to be high-dimensional. Stop words are removed from the data to reduce dimensionality [10, 17]. Words such as "the", "and", "of", and "on" are examples of stop words in the English language. Words such as "ada", "akan" and "bukan" are examples of stop words in the Indonesian language. The pre-built function of the NLTK library includes a list of stop words in English and Indonesian. After that, the data were tokenized to produce a bag-of-words

(BOW), which separates the words from the remaining text. Following the tokenization process, the raw text is transformed into collections of tokens, each of which is typically a word. In addition to that, the stem process, also known as lemmatizing, was put into action. This method of normalizing text eliminates away with suffixes and is called stemming. It cuts down on the total number of words, making the text even less three-dimensional. The completed dataset consists of seven columns, which are as follows: data created, username, tweets, clean tweets, tokens, words, and lemmatized. A CSV file, which functions as a database, is used to save the data before any data visualization is performed on them. After completing the pre-processing step, the complete dataset is now prepared to be classified using the ML algorithm.

TextBlob is a simple API offered by a Python library [18]. It extracts subjectivity and polarity from Twitter texts accomplished through data pre-processing for improved data storage. It serves as the base of SA and is among its most crucial components. It establishes whether a particular tweet is supportive, critical, or neutral. The negative tweets count as -1, tweets that are positive count as +1, and tweets that are neutral count as 0. Subjective statements generally discuss a person's perception, emotion, or assessment. In contrast, knowledge founded on facts is referred to in objective words. Another float that falls between [0,1] is subjectivity. Figure 2 demonstrates the snippet of code used to retrieve tweets' SA, while Figure 3 displays the result retrieved after applying SA.

```
#get subjectivity and polarity of tweets with a function
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

#get polarity with a function
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

df['Subjectivity'] = df['CleanTweets'].apply(getSubjectivity)
df['Polarity'] = df['CleanTweets'].apply(getPolarity)
df
```

Fig. 2. Snippet Code for Retrieve Subjectivity and Polarity

```
#create a function to check negative, neutral and positive analysis
def getAnalysis(score):
    if score<0:
        return 'Negative'
    elif score ==0:
        return 'Neutral'
    else:
        return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)
df
```

Fig. 3. Snippet Code for Categorize in Three Analysis

Then, ABSA classification was categorized by looking at the relevance of tweets with particular keywords to the different factors. The keywords used to categorize the factors are displayed in a snippet code in Figure 4.

```
price = ['cheap','expensive','mahal','affordable','murah','tax','delivery charge']
promotion = ['voucher','promotion','celebration','discount','reward']
review= ['bagus','senang','worth','disappointing','susah','uninstall','bad service','cancel']
deliverytime = ['fast','slow','lambat','cepat','laju']
condition= ['spill','tumpah','salah','wrong']
```

Fig. 4. Snippet Code for Categorize in Five Factors

2.2.3 Naïve Bayes Classification Model

The dataset was classified using an algorithm tested using the NB classifier, which was used to evaluate the model [19]. The model uses the pre-labeled data found in the training set to the dataset before deciding its categorization [20]. The NB theorem is a method for computing an event's probability using the probabilistic joint distribution of previous events. In this study, using a pre-labeled training dataset helps the model learn the context for sentences that are either positive, neutral, or negative.

When the first and second phases of BOW are combined, a statistical metric known as the Term Frequency - Inverse Document Frequency (TF-IDF) is used to evaluate how important a given word is to the content of a given document [20]. The TF-IDF weight is utilized in both the process of information retrieval and text mining. The tool known as TF was utilized to determine the number of times a phrase appears in a given document. In contrast, IDF was utilized to ascertain the level of relevance [21].

The data is divided into training and testing in the proportion of 80:20, with training data representing 80 percent of the whole data set and testing data making up the remaining 20 percent. Regarding this undertaking, the test set ratio is 0.20. Because there is a possibility that there will not be enough data to train on, the test set should not be larger than the train set. The state of randomness determines the seed for the random number generator, which is number 45, representing this study's random state. It is the highest accuracy score among all other numbers.

After the splitting process, a prediction may be made on the response for the test dataset. The accuracy of the model's prediction for the test dataset was given as a percentage of correct predictions out of the total number of predictions produced by the model. The expression *"classifier = MultinomialNB()"* is used as the training model for the NB classification. The model is constructed with the help of the code *"classifier.fit(X_train, y_train)"*, which is passed the *X_train*, which contains the feature vector, and the *y_train*, which contains the desired variable. The procedure is referred to as the training process.

2.2.4 Model Deployment

Model deployment refers to putting ML models into actual use in the real world. In many contexts, this phrase refers to making a model accessible via real-time application programming interfaces (APIs), which enables information to be retrieved in real-time. During the stage where the model is deployed, the predicted categorized tweets are generated with sentiment labels of "0", "1", and "2", with the latter three having the corresponding meanings of negative, neutral, and positive sentiments.

After applying the model classifier to predict the sentiment based on the acquired data and evaluating how well it worked, the data is shown using Plotly. An open-source interactive graphics package for Python can be found under the name Plotly. Importing the data into Pandas data frames in Python was the initial step of the procedure. After that, the code for the Jupyter Notebook is generated using the data from the Excel file. After that, charts were produced by utilizing the chart studio in the online version of Plotly, together with the data entered. Consequently, the result is then utilized in constructing an interactive visualization tool that presents the results of real-world data analysis.

Word cloud data visualization was utilized to present the text data for OFD that the system will generate. The words shown in various colors and the size of each word emphasize how frequently that particular term appears in the text data. The terms employed for OFD firms will be immediately recognizable due to the placement of the words in a cloud.

2.3 Testing Development

After all previous phases are completed, the system is generally considered nearly finished and ready for testing. Running the test cases on the system is a step in the testing process. This step is to verify the overall system's functionality and ensure it complies with the standards outlined in the previous stage [22]. Functional testing is a technique for determining whether or not a piece of software meets pre-defined requirements. The primary purpose of functional testing is to ensure that a system performs as expected. The user will know whether the system fulfills the need based on the functionality testing results. A usability test is an essential step in designing a human-computer interface. It is the process of methodically gathering interface usability data and assessing and improving it. The primary goal of usability is to create products that are as simple to use as possible for people [23]. It is also important to determine if the user has discovered an additional method of accessing the system that is not caused by its poor design.

2.4 Front-end Development

Front-end web development, or client-side development, translates data to a graphical interface using the programming languages HTML, CSS, and JavaScript to create a website that enables users to view and interact with that data. Front-end developers accomplish this by utilizing a combination of design, technology, and programming to code the appearance of a website, in addition to taking care of any debugging that may be necessary. A Python online application environment has a set of data visualization tools, which may be used to create charts and graphs based on sentiment data. This study has eight interfaces. It includes a landing page, home page, dashboard page, individual OFD pages like Foodpanda, Grabood and ShopeeFood, a competitive analysis page, and a real-time Twitter updates page.

3. Result And Discussion

3.1 Accuracy Testing

The accuracy of the NB classification model can be evaluated using simple Python code. Table 1 summarizes the results of the accuracy testing performed on the English and Malay model of the training dataset. The overall score for English model accuracy is 71.67% when stated in percentage form. This score indicates that the sentiment result is 71.67% accurate, which indicates that the model correctly recognized seven out of ten correct answers as “positive”, “neutral”, or “negative”. In the confusion matrix, the class designated as “negative” is denoted by the number 0, the class designated as “neutral” by the number 1, and the class designated as “positive” by the number 2.

The accuracy score for the Malay model in the confusion matrix is 76.29% in percentage form, which indicates that the sentiment result is 76.29% correct, with the algorithm correctly classifying seven correct outcomes as “positive”, “neutral”, or “negative” out of a total of ten attempts. It indicates that the sentiment result is 76.29% accurate.

Table 1
Summary of Result Accuracy Testing for English and Malay Model

Model	Accuracy	Sentiment	Precision	Recall	F1-Score
English	71.67%	0 (negative)	0.48	0.65	0.55
		1 (neutral)	0.86	0.72	0.79
		2 (positive)	0.66	0.74	0.70
Malay	76.29%	0 (negative)	0.18	0.41	0.25
		1 (neutral)	0.92	0.81	0.86
		2 (positive)	0.40	0.58	0.47

3.2 Overview Dashboard Visualization

The web-based “Dashboard” page included a total sentiment dataset and word clouds representing positive, negative, and neutral sentiments. It provided an overview of the services offered by OFD. Next, the total SA for each OFD, a bar chart of the total sentiment of each ABSA based on OFD, a bar chart of the total sentiment of each OFD based on five ABSA: affordable price, promotion and discount, delivery time, review rating and condition of food delivered that user tweets on Twitter. For each ABSA, users can view the total of each sentiment.

3.2.1 Overall Sentiment Analysis

The system's dashboard plots and displays the complete data analysis. They were visualized using data visualization techniques such as pie charts, bar charts, line charts and word cloud for better visions. Figure 5 shows the main dashboard that indicates the overall sentiments for 3 OFD's dataset and each positive, negative, and neutral sentiment, as in Figure 5 (a) and a pie chart on the percentage of the sentiment, as in Figure 5 (b). Based on the total of 4,645 sentiments, it was distributed to 854 positive, 425 negative sentiments, and 3,366 neutral sentiments. The user may immediately compare positive, negative, and neutral sentiments using the pie chart's total sentiment level and color differences. The dashboard on five specific ABSA classification sentiments is visualized in the bar graph in Figure 6. Users may also visualize the total mentions of each OFD based on factors.

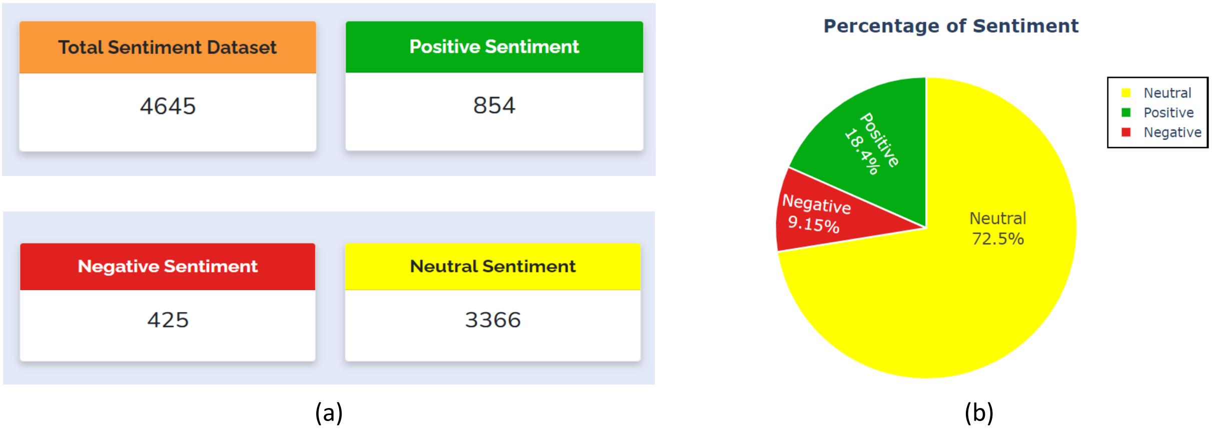


Fig. 5. Dashboard for (a) Overall Sentiments for 3 Online Food Delivery and (b) Pie Chart of the Sentiment's Distribution

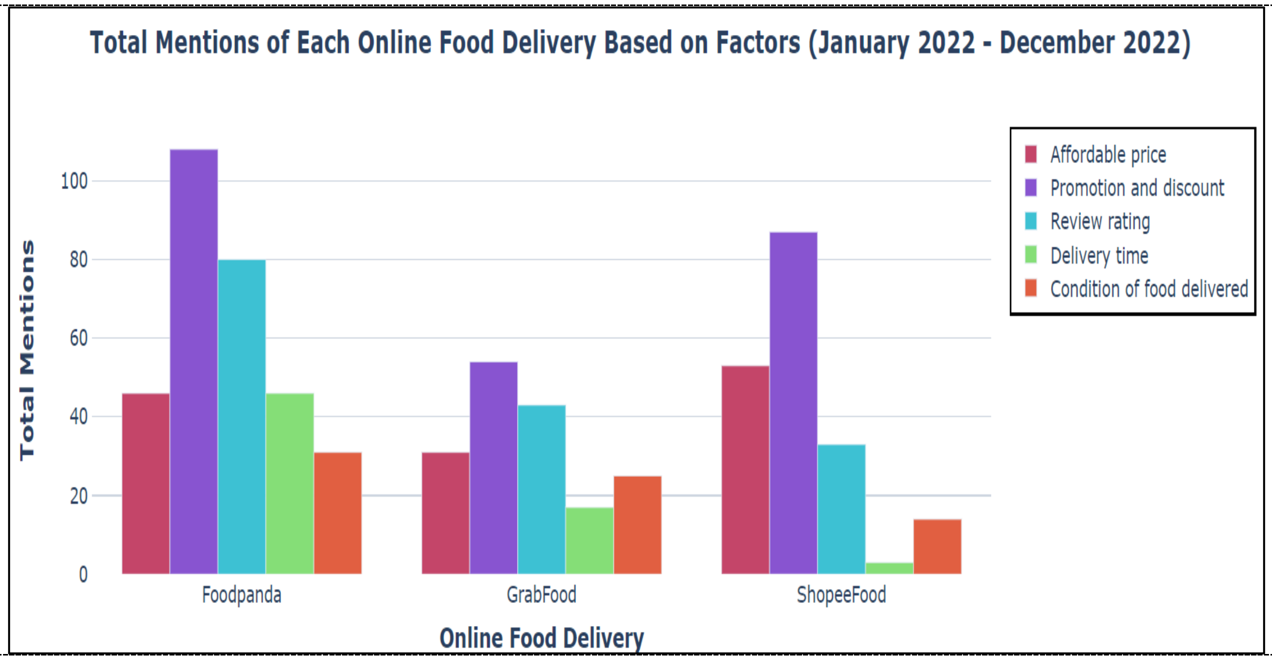


Fig. 6. Dashboard for Specific Classification Sentiments on 5 ABSA

3.2.2 Visualization Sentiment Analysis

Figure 7 shows three-word clouds for this study. Figure 7 (a) illustrates a word cloud representation of positive sentiment text data. The size of the term indicates the number of times it appears in the dataset. The larger the size, the greater the occurrence frequency in the dataset [10]. The word cloud contains words such as “great”, “enjoy”, “good” and “best” that are connected with a positive mention. A visualization of phrases that are connected with negative sentiment is shown in Figure 7 (b). Words and phrases such as “late”, “dirty”, “wrong”, “unpleasant” and “poor” are contained within the word cloud. The word cloud for the neutral emotion is displayed in Figure 7 (c). It shows the terms that are connected with the neutral mention, such as “order”, “makanan” and “promo”.

(a) (b) (c)

Fig. 7. Dashboard for (a) Positive Sentiments, (b) Negative Sentiments and (c) Neutral Sentiments

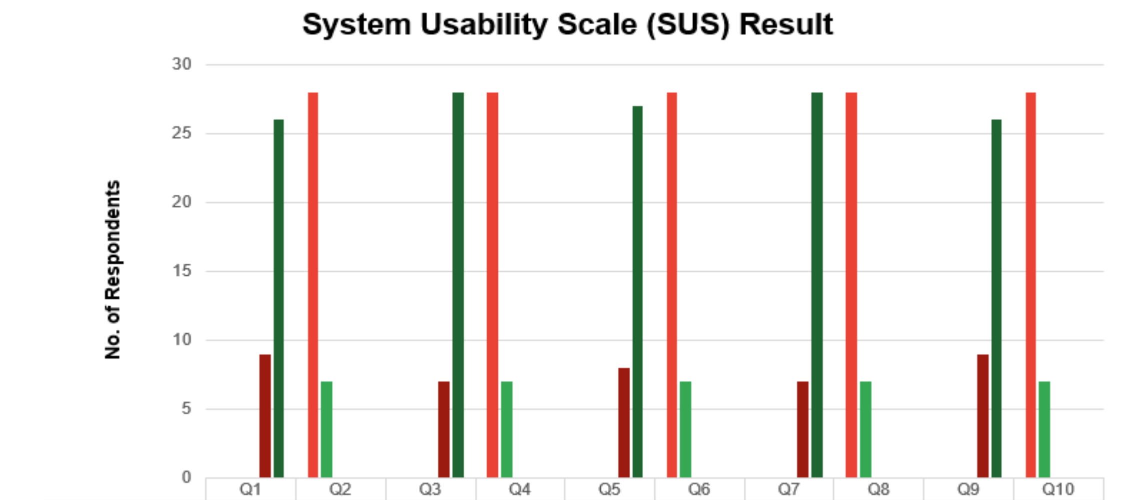
3.3 Functionality Testing

Testing is required to ensure that all of the elements of the system function as intended and that any unexpected behavior exhibited by the system is quickly identified and fixed. Functional testing aims to examine every aspect of a system to validate whether it satisfies the functional requirements. This test is carried out by constructing test cases based on the system requirements, which are then put through its tests. We succeeded in finishing the dashboard and passing the functionality test.

3.4 Usability Testing

Usability testing is the technique of evaluating the usability of a design with a group of representative users [10]. It often entails monitoring users while users attempt to perform activities and may be used for various design kinds. Usability testing can uncover design issues that would otherwise be overlooked. Observing the behavior of test users as they attempt to complete activities will provide crucial insights into the functionality of the design or product [24]. These insights can make modifications to the system. System usability is assessed using the System Usability Scale (SUS). SUS consists of 10 user-response questions. Figure 8 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 “strongly agree” on a scale of 5 for the odd-numbered items, which are positive assertions.

Conversely, the questions with even numbers had a majority of scale 1 responses, indicating that users “strongly disagree” with the negative allegations made in the questions. It indicates that the users have a good experience with the system and do not require any form of technical support in order to use all of the capabilities that it offers. The majority of the users express satisfaction with the system.



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Strongly Disagree	0	28	0	28	0	28	0	28	0	28
Disagree	0	7	0	7	0	7	0	7	0	7
Neutral	0	0	0	0	0	0	0	0	0	0
Agree	9	0	7	0	8	0	7	0	9	0
Strongly Agree	26	0	28	0	27	0	28	0	26	0

Fig. 8. Bar Chart of SUS Result

The SUS scores' histogram is depicted in Figure 9. The frequency of users who responded to the SUS is shown on the histogram's y-axis. In addition, the percentage of the SUS score range is shown on the x-axis. According to the histogram, the graph demonstrates a normal distribution with a 90% to 100% range and a 2% interval. A total of 12 respondents fall within the peak range between 94% and 96%; 13 are below the median value, and 10 are above the median. The 35 responders to the SUS questionnaire had an average SUS score of 94.64%. Any SUS score above 68 would be considered above average, while any lower than 68 would be considered below average. The procedure known as "normalizing" is the most effective approach to evaluating a score because it allows conversion to a percentile rank. If the SUS score exceeds 84, the percentage range is 96-100, and the adjective is "best imaginable". If the score is between 80 and 84, the percentage range is 90-95, and the adjective is "excellent". If the score is between 70 and 79, the percentage range is 60-89, and the adjective is "good". If the score is between 50 and 69, the percentage range is 15-59, and the adjective is "ok". If the score is less than 50, the adjective is "poor" and "worst imaginable" [25]. As a result, this study falls within a range that is considered acceptable, and the adjective is "excellent" with the achievement of the aims.

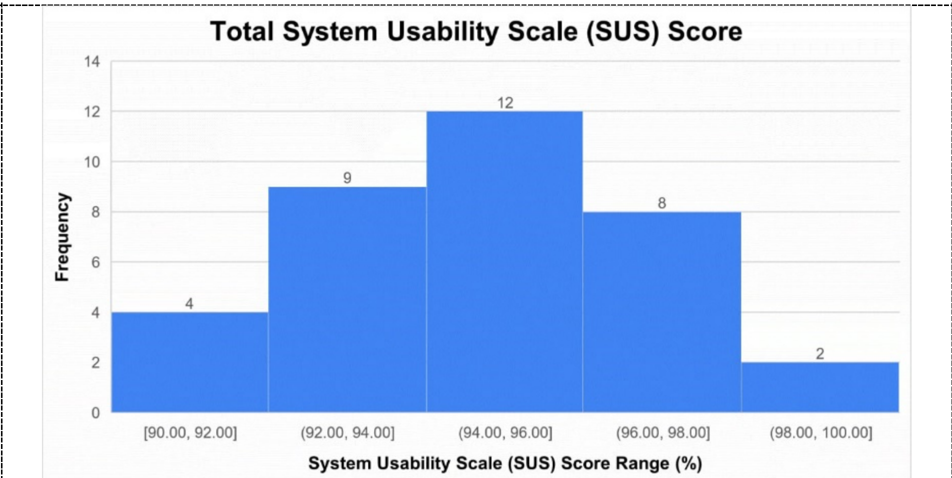


Fig. 9. Histogram of SUS Result

4. Conclusions

In this study, a web application system is designed and developed to categorize Malaysia's best OFD: Foodpanda, GrabFood and ShopeeFood based on Twitter SA. The visualization of SA was extracted from Twitter from 1st January 2022 to 31st December 2022. The NB classification model is embedded in the system application and helps the user determine the performance of OFD services and make future decisions. Five ABSA was selected as the main factor in classifying the best OFDs. It includes affordable price, promotion and discount, review rating, delivery time and condition of food delivered. Aside from that, the web application system includes numerous visualizations, making it easy for customers to understand the OFD services offered in Malaysia. Customers can save time and effort by using this visualization to compare each OFD, making it easier for the user to gain better and more precise insights. It is suggested that in the future, the reviews for Malay be read through one by one to ensure that no Indonesian words are added. Then, more information can be included by extending the time it takes to scrape data from Twitter. Since emojis show the user's intent in communicating their views, they should be a part of the classification process.

Acknowledgement
This research was funded by a grant from Universiti Teknologi MARA Cawangan Melaka (TEJA Grant 2023 GDT 2023/1-14).

References
[1] Keeble, Matthew, Jean Adams, Gary Sacks, Lana Vanderlee, Christine M. White, David Hammond, and Thomas Burgoine. "Use of online food delivery services to order food prepared away-from-home and associated sociodemographic characteristics: a cross-sectional, multi-country analysis." *International Journal Of Environmental Research And Public Health* 17, no. 14 (2020): 5190. https://doi.org/10.3390/ijerph17145190.

[2] Tan, Huikok, and Vincent Wee Eng Kim. "Examining the factors that Influence consumer satisfaction with online food delivery in Klang Valley, Malaysia." *The Journal of Management Theory and Practice (JMTP)* (2021): 88-95. https://doi.org/10.37231/jmtp.2021.2.2.115.

[3] Nguyen, Bang, Van-Ho Nguyen, and Thanh Ho. "Sentiment analysis of customer feedbacks in online food ordering services." *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy* 12, no. 2 (2021): 46-59. https://doi.org/10.2478/bsrj-2021-0018.

[4] Han, Saram, and Christopher K. Anderson. "The dynamic customer engagement behaviors in the customer satisfaction survey." *Decision Support Systems* 154 (2022): 113708. https://doi.org/10.1016/j.dss.2021.113708.

[5] Roslan, Nur Mira Alisa, and Nur Syazwani Mohd Nawi. "Food Delivery Service Application During Pandemic Covid-19." *Journal of Technology and Operations Management* 17, no. 1 (2022): 71-79. https://doi.org/10.32890/jtom2022.17.1.7.

[6] Prajapati, Bhavna, Arijit Goswami, and Archa Agrawal. "A Study on social media as a promotional tool for Food ordering companies and its impact on customers." *International Journal of Creative Research Thoughts* 8, no. 3 (2020): 1373-1386. https://doi.org/10.13140/RG.2.2.14869.27360.

[7] Singh, Ravindra Kumar, and Harsh Kumar Verma. "Influence of social media analytics on online food delivery systems." *International Journal of Information System Modeling and Design (IJISMD)* 11, no. 3 (2020): 1-21. https://doi.org/10.4018/IJISMD.2020070101.

[8] Cheong, Fan, and Rob Law. "Will Macau's restaurants survive or thrive after entering the O2O food delivery platform in the COVID-19 pandemic?." *International Journal of Environmental Research and Public Health* 19, no. 9 (2022): 5100. https://doi.org/10.3390/ijerph19095100.

[9] Neimann Rasmussen, Lauge, and Paul 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* 7, no. 1 (2018): 1-12. https://doi.org/10.1186/s13643-018-0786-6.

[10] Samah, Khyrina Airin Fariza Abu, Nur Farhanah Amirah Misdan, Mohd Nor Hajar Hasrol Jono, and Lala Septem Riza. "The best Malaysian airline companies visualization through bilingual twitter sentiment analysis: a machine learning classification." *JOIV: International Journal on Informatics Visualization* 6, no. 1 (2022): 130-137. http://dx.doi.org/10.30630/joiv.6.1.879.

[11] Agüero-Torales, Marvin M., José I. Abreu Salas, and Antonio G. López-Herrera. "Deep learning and multilingual sentiment analysis on social media data: An overview." *Applied Soft Computing* 107 (2021): 107373. https://doi.org/10.1016/j.asoc.2021.107373.

[12] Ofori, Francis, Elizaphan Maina, and Rhoda Gitonga. "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review." *Journal of Information and Technology* 4, no. 1 (2020): 33-55.

[13] Rahat, Abdul Mohaimin, Abdul Kahir, and Abu Kaisar Mohammad Masum. "Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset." In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pp. 266-270. IEEE, 2019. https://doi.org/10.1109/SMART46866.2019.9117512.

[14] Mahesh, Batta. "Machine learning algorithms-a review." *International Journal of Science and Research (IJSR).[Internet]* 9, no. 1 (2020): 381-386. https://doi.org/10.21275/ART20203995.

[15] Protopsaltis, Antonis, Panagiotis Sarigiannidis, Dimitrios Margounakis, and Anastasios Lytos. "Data visualization in internet of things: tools, methodologies, and challenges." In *Proceedings of the 15th International Conference on availability, reliability and security*, pp. 1-11. 2020. https://doi.org/10.1145/3407023.3409228.

[16] Nugraheni, Ekasari. "Indonesian twitter data pre-processing for the emotion recognition." In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 58-63. IEEE, 2019.

[17] Baweja, Vanshita R., Rajesh Bhatia, and Manish Kumar. "Support vector machine-based focused crawler." In *Inventive Communication and Computational Technologies: Proceedings of ICICCT 2019*, pp. 673-686. Springer Singapore, 2020. https://doi.org/10.1007/978-981-15-0146-3_63.

[18] Mahgoub, Abdulrahman, Hesham Atef, Abdulrahman Nasser, Mohamed Yasser, Walaa M. Medhat, M. Saeed Darweesh, and Passent M. El-Kafrawy. "Sentiment analysis: Amazon electronics reviews using BERT and Textblob." In *2022 20th International Conference on Language Engineering (ESOLEC)*, vol. 20, pp. 6-10. IEEE, 2022. https://doi.org/10.1109/esolec54569.2022.10009176.

[19] Hermansyah, Reza, and Riyanarto Sarno. "Sentiment analysis about product and service evaluation of pt telekomunikasi Indonesia tbk from tweets using textblob, naive bayes & K-NN method." In *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 511-516. IEEE, 2020.

[20] Ganiger, Shweta, and K. M. M. Rajashekharaiah. "Comparative study on keyword extraction algorithms for single extractive document." In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1284-1287. IEEE, 2018.

[21] Baruni, J. S., and J. G. R. Sathiaseelan. "Keyphrase extraction from document using RAKE and TextRank algorithms." *Int. J. Comput. Sci. Mob. Comput* 9 (2020): 83-93. https://doi.org/10.47760/ijcsmc.2020.v09i09.009

[22] Nikiforova, Anastasija, and Janis Bicevskis. "Towards a Business Process Model-based Testing of Information Systems Functionality." In *ICEIS (2)*, pp. 322-329. 2020. https://doi.org/10.5220/0009459703220329

[23] Nayan, Norshita Mat, and M. Hassan. "Customer satisfaction evaluation for online food service delivery system in Malaysia." *J. Inf. Syst. Technol. Manag* 5, no. 9 (2020): 123-136. https://doi.org/10.35631/jistm.5190010.

[24] Hook, Alan. "Exploring speculative methods: building artifacts to investigate interspecies intersubjective subjectivity." *Alphaville: Journal of Film and Screen Media* 17 (2019): 146-164. https://doi.org/10.33178/alpha.17.09

[25] Putri, Dwi Andini, Denda Rinaldi Hadinata, and Siti Nurwahyuni. "Website usability testing on the information and coordination center of COVID-19 in West Java using the system usability scale." *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)* 6, no. 1 (2020): 1-6. https://doi.org/10.33480/jitk.v6i1.1335.