

VISUALIZING THE BEST MALAYSIAN AIRLINE COMPANIES THROUGH TWITTER SENTIMENT ANALYSIS USING NAÏVE BAYES

*Khyrina Airin Fariza Abu Samah¹, Nur Farhanah Amirah Misdan², Noor Afni Deraman³, Siti Nor Amalina Johari⁴, Nor Aiza Moketar⁵, Mohd Nor Hajar Hasrol Jono⁶

¹²³⁵⁶Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka
Kampus Jasin, Melaka, Malaysia

*Corresponding author's email: khyrina783@uitm.edu.my

ABSTRACT

In Malaysia, 37.1% of Internet users owned Twitter accounts in 2020. Besides that, the tourism industry is the third biggest contributor to Malaysia, putting the aviation and travel industry as part of the category. However, there is no specific platform for direct comparison for the online reviews among companies despite it is critical for business growth, performance and improvement of customer experience. Other than that, most online ratings obtained their result from the online platform using the English language only. Thus, this study aims to visualize the best Malaysian airline companies through Twitter sentiment analysis using Naïve Bayes (NB). The source of the data for this project is Twitter, where the tweets are extracted using dates and keywords. The data was pre-processed, and the model is run on real-world data. The model evaluation is conducted using the NB classifier. Two machine learning models for English and Bahasa Malaysia have been built for classification purposes based on the multi-class text classification. The results obtained are visualized in a dashboard. High accuracy score is achieved during testing and the project objectives are achieved. The future work that can be put into this project is to include other social media platforms for a wide reach to the companies.

Keywords: Naïve Bayes Classifier, Twitter sentiment analysis, Web-based visualization

1. INTRODUCTION

An airline is a company that provides air transport services for traveling passengers and freight. Airlines utilize aircraft to supply these services and may form partnerships or alliances with other airlines for codeshare agreements, in which they both offer and operate the same flight. Tourism placing the third as the most significant contributor to Malaysia's Gross Domestic Product (GDP) after the manufacturing and commodities industry (Hirschmann, 2020). According to the Department of Statistics, Malaysia (2020), 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%. According to the Malaysian Communications and Multimedia Commission (MCMC)'s Internet Users Survey 2020, there are 28.7 million Internet users in 2020. Twitter is in fourth place as the most widely used social media site in Malaysia with 37.1%. Social media application platforms have become the main ground for their users to comment because of their openness. Therefore, if certain products or services do not meet their customers' expectations, there are high chances that the customers' will express their personal opinion on the social platform.

Although there are WASR and OTA that compared airlines, WASR consists of many airlines, and users cannot see a direct comparison among airlines. Moreover, it can be time-consuming for the user to check each airline to make a comparison. Besides that, most online ratings obtained their result from the online platform using the English language. As time passes, this could lead to inaccurate results as not all reviews regarding different languages are considered. Neimann & Montgomery (2018) stated that when research teams conduct systematic reviews, studies published in a language other than English are mostly neglected. Researchers had argued about the possibility of bias when excluding non-English studies.

Thus, this study aims to adapt the Naïve Bayes (NB) algorithm to visualize the performance of the best airline companies in Malaysia through Twitter sentiment analysis. The targeted users are three Malaysian airline companies which are Malaysia Airlines, AirAsia, and Malindo Air. The system is a web-based application written in Python and the data is obtained from Twitter related to the three airlines dated from January 1st, 2019 to December 31st, 2019. The visualized results could be used in maximizing customers' satisfaction and ensuring retention. In this way, the consumers will be able to proactively target new potential markets and resolve customer issues more effectively.

2. MATERIALS AND METHODS

The back end is the data access layer of the system during the development, also known as the server-side development code. Figure 1 shows the algorithm design for the project development.

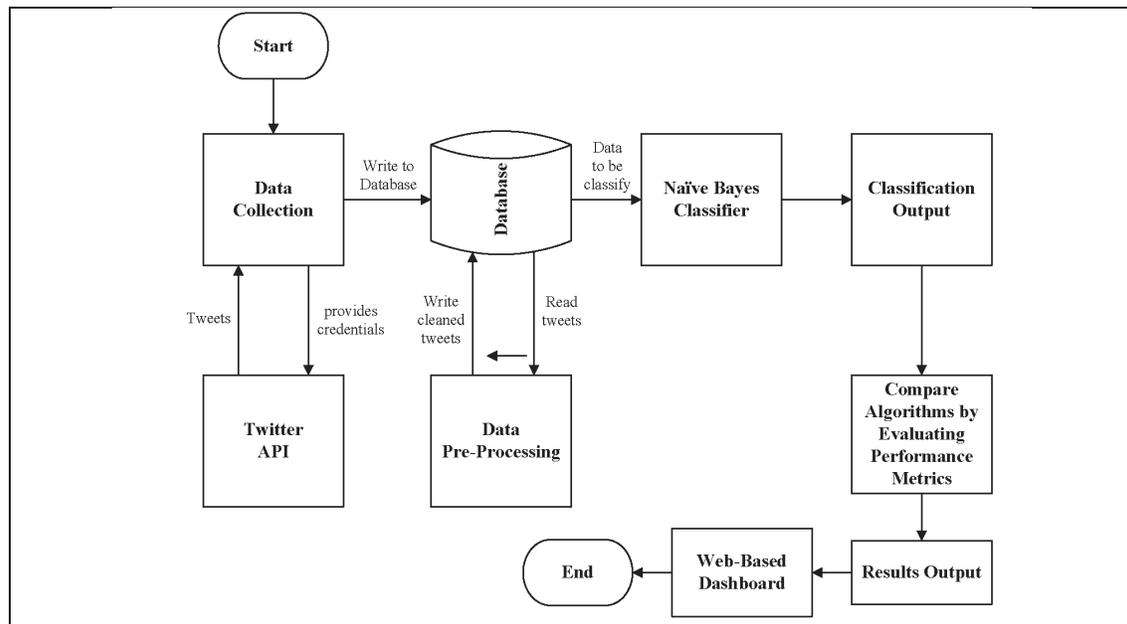


Figure 1. Back-End Development Phases for The Study

3. RESULTS AND DISCUSSION

3.1: Sentiment Analysis Trend Comparison

The user can display the competitive analysis page in which the user can compare performance between companies of their choice. Figure 2 shows the positive total mentioned according to the time-based collected data, where MAS and Malindo Air have a similar trend, where the total mentions is constant throughout the year. Meanwhile, AirAsia shows an inconstant trend where the total mentions reach its highest and July 2019 and the lowest in May 2019 and October 2019.

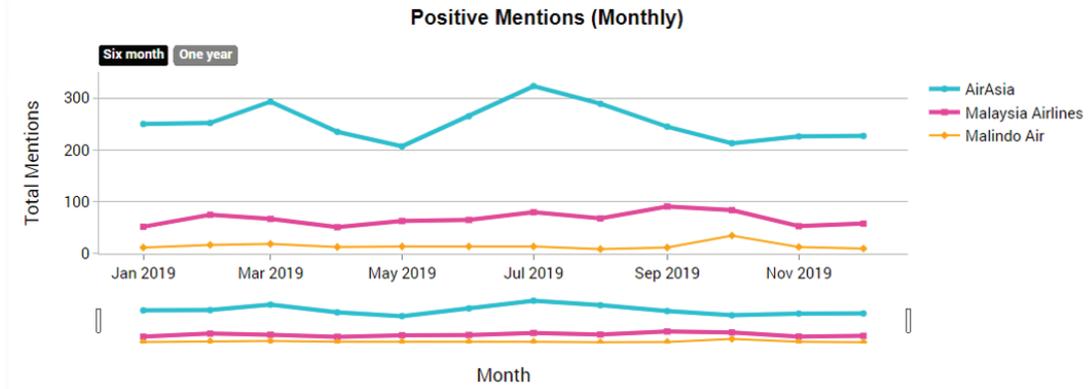


Figure 2. Monthly-based Trend for Positive Mentions Sentiment Analysis

Figure 3 shows the opposite trend for the negative total mentioned, where Malindo Air shows a constant trend, while AirAsia and MAS depict that the total mention increase in early and mid-year.

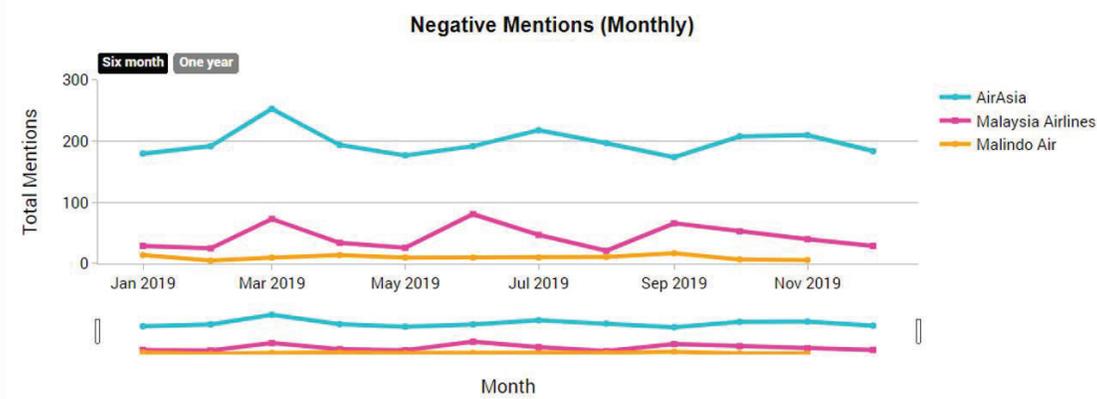


Figure 3. Monthly-based Trend for Negative Mentions Sentiment Analysis

Figure 4 shows the trends for the neutral total mentioned, where the three companies show a similar trend in March 2019, where the total mentions increase but drop in April 2019. However, as MAS and Malindo Air show an increment in June 2019 and September 2019, AirAsia face a constant drop until the end of the year.

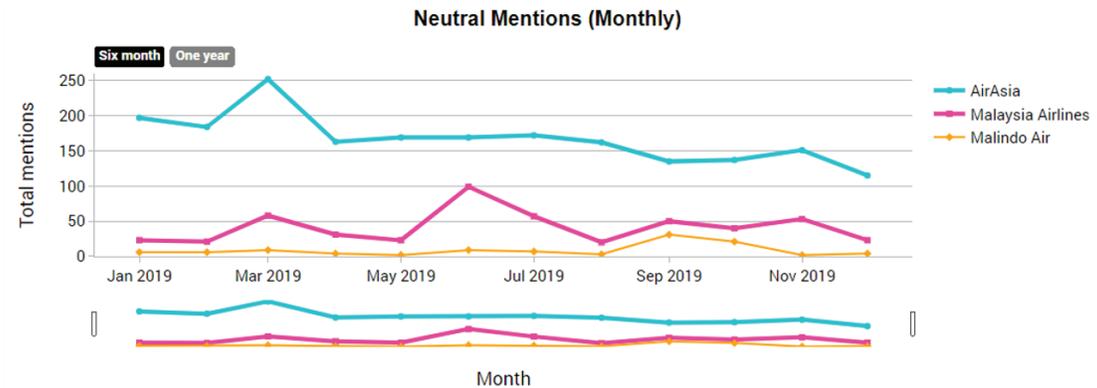


Figure 5. Monthly-based Trend for Neutral Mentions Sentiment Analysis

4. CONTRIBUTION AND USEFULNESS/COMMERCIALISATION

- i. This system can help Malaysian airline companies visualize the reputation of their company compared to other companies.
- ii. This system can help Malaysian airline consumers to see a direct comparison among the three airlines.
- iii. This project also contributes to the study's domain in the research area, as it involves creating a bilingual sentiment analyzer that includes Bahasa Malaysia and English.

5. CONCLUSION

This study's main purpose was to develop a web application that provides interactive visualization application that would make the analysis results readable and understandable for the stakeholders on the public sentiment of the policies provided by Malaysian airline companies. The application's testing phases show that the classifier successfully classified tweets' sentiment with higher than 90% accuracy. It is observed in the functionality testing that every feature in this application functions properly. In addition, this study also includes the real-time Twitter updates for each company to provide the user with the latest information from the companies' official Twitter account. For future recommendation, it is suggested that the reviews from other social media platforms are included in the study for better insight of customers' opinions towards the companies.

REFERENCES

- Hirschmann, R. (2020). Travel and tourism in Malaysia – statistics and facts. Retrieved July 31, 2021, from <https://www.statista.com/topics/5741/travel-and-tourism-in-malaysia/>.
- Department of Statistics, Malaysia (2020). Tourism satellite account 2019. *Department of Statistics Malaysia, Putrajaya*.
- Malaysian Communications and Multimedia Commission, M. (2018). Internet user survey 2018: Statistical brief numbered twenty-three. *Cyberjaya, Selangor*.
- Neimann Rasmussen, L., & Montgomery, P. (2018). The prevalence of and factors associated with the inclusion of non-English language studies in Campbell systematic reviews: A survey and meta-epidemiological study. *Systematic Reviews*, 7(1), 1–12.