UNIVERSITI TEKNOLOGI MARA

X SENTIMENT ANALYSIS: CLASSIFICATION AND VISUALIZATION OF COURIER SERVICES IN MALAYSIA USING NAIVE BAYES AND PLOTLY

SITI KHADIJAH BINTI ZAZALI

BACHELOR OF COMPUTER SCIENCE (HONS.)

JULY 2024

Universiti Teknologi MARA

X Sentiment Analysis: Classification and Visualization of Courier Services in Malaysia Using Naive Bayes and Plotly

Siti Khadijah Binti Zazali

Thesis submitted in fulfilment of the requirements for Bachelor of Computer Science (Hons.) College of Computing, Informatics and Mathematics

July 2024

SUPERVISOR APPROVAL

X SENTIMENT ANALYSIS: CLASSIFICATION AND VISUALIZATION OF COURIER SERVICES IN MALAYSIA USING NAIVE BAYES AND PLOTLY

By

SITI KHADIJAH BINTI ZAZALI
2021610118

This thesis was prepared under the supervision of the project supervisor, Khairul Nizam Abd Halim. It was submitted to the College of Computing, Informatics and Mathematics and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons.)

Approved by

………………………………
Khairul Nizam Abd Halim
Project Supervisor

JULY 8, 2024

STUDENT DECLARATION

I certify that this thesis and the project to which it prefers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

……………………………………………
SITI KHADIJAH BINTI ZAZALI
2021610118

JULY 8, 2024

ACKNOWLEDGEMENT

Alhamdulillah, all praises to Allah S.W.T. for His Almighty and His utmost blessings, I was able to complete this research within the time duration given. First and foremost, I want to express my gratitude to my supervisor, Mr Khairul Nizam bin Abd Halim, for his endless support and insightful guidance in completing this project. He played the largest part in helping me to be able to deliver excellent work through this project.

I also want to thank my lecturer, Dr Noor Hasimah binti Ibrahim Teo, for her continuous support throughout the entire project. Her assistance significantly shaped the precursor of this project as well as kept the project progress on track.

A token of appreciation also goes to Ts. Dr. Khyrina Airin Fariza Abu Samah, my examiner, for guiding the completion of this project. Her attention to detail has helped me tremendously in perfecting this report.

My token of appreciation also goes to my supportive friends, Nurul Bahirah binti Mohd Nasir and Erin Hanna Yeo, for their encouragement and help in completing this project. Struggling with our projects together has empowered me to finish this project.

ABSTRACT

Online reviews are one of the most valuable assets for a company, including courier service businesses. Since there is no central platform to analyze all reviews of courier services in Malaysia, businesses and customers resort to social media for online opinions. Twitter, now known as X, is a popular open-forum site, making it a great platform for analyzing sentiment about a chosen topic. However, manual analysis approach is time-consuming and inefficient. Some of the reviews can even be biased and unreliable. Not only that, most online sentiment analysis only considers English language. Due to these issues, this project aims to design a classification and visualization system for X sentiment analysis of courier services in Malaysia for both English and Malay language. This project uses Naive Bayes algorithm for classification and Plotly library to perform the visualization. The methodology used is the Modified Waterfall model which is made up of five phases: requirement analysis, design, implementation, testing and documentation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABSA	Aspect-Based Sentiment Analysis
CSS3	Cascading Style Sheet 3
DHL	Dalsey, Hillblom, and Lynn
JNE	Jalur Nugraha Ekakurir
J&T	Jet and Tony
KNN	K-Nearest Neighbour
NB	Naïve Bayes
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
SDLC	Software Development Life Cycle
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
UML	Unified Modeling Language

CHAPTER 1

INTRODUCTION

This chapter discusses the background of research, problem statements, objectives, scope and significance of this project.

1.1 Background of Research

Twitter, now known as X, is a popular microblogging platform. It is a social networking platform for users to interact online through tweets or short texts (Hetler, 2023). X simplicity enables users to quickly read through a certain topic, thanks to its 280-character limit. It also offers interactive features such as likes and retweets, which promptly facilitate information sharing to audiences all over the world. The use of hashtags is also a popular practice on X. Hashtags are handles written in tweets with the intention to categorize the tweet under a similar subject. This element further eases information acquisition for any particular issue. All of these convenient features combined enables X to become a popular open-forum site, making it a suitable platform for analyzing sentiment about any topic. Since there is no central platform specifically for reviews of all courier services in Malaysia, X is chosen as the most suited alternative.

Courier service industry in Malaysia has been flourishing since the COVID-19 pandemic due to the rise of the e-commerce industry during the same time. This conjunction happens since the services offered by courier businesses are essential for moving the e-commerce wheels to ensure the demands of smooth flow of goods are fulfilled. In order to improve their product or service quality, businesses often rely on feedback (Needle, 2023). Online reviews are one of the most valuable assets for a company. It provides insights into actual experiences, giving customers an idea of the product or service's quality and value (Willas, 2023). However, the complicated and huge amount of data can make manual analysis method inefficient to extract insights from the data.

Sentiment analysis is a natural language processing (NLP) process that analyzes text to figure out the emotional tone behind the message, either positive, negative, or neutral (Gupta, 2018). This is a popular way for businesses to determine the public opinions about a product, service or the business itself. In this era of digitalization, sentiment analysis is more crucial than ever for businesses since it automatically analyzes the sentiment of the large volume of online reviews, allowing these companies to drive meaningful information from the analysis. Not only that, by categorizing the sentiments into meaningful aspects and visualizing the data, businesses can benefit from the insights, ultimately improving the business quality and performance.

1.2 Problem Statement

In an ideal world, courier service providers and customers have access to a reliable and insightful system for classifying and visualizing sentiment on X about courier services in Malaysia. The system would empower companies to enhance their services and address customer concerns, ultimately improving customer satisfaction. At the same time, the system would enable customers to make data-driven decisions when deciding the best courier services according to their preferences.

A survey was carried out towards 92 respondents to dive into of Malaysians' perspectives about courier services in Malaysia. The findings were striking, with an overwhelming majority (96.7%) expressing challenges in locating reviews for courier services online. Furthermore, an astonishing 97.8% of respondents voiced a strong desire for a centralized platform dedicated to analyze these reviews. These statistics underscore the urgent need for the development of a comprehensive system to analyze reviews of Malaysian courier services.

Due to the absence of a system to automatically classify sentiment regarding courier services, users rely on manual analysis to review the feedback. The manual approach is time-consuming and inefficient, which is why an automated sentiment analysis is needed to facilitate the process (Al-Jarrah et al., 2023). This problem is further supported by the large amount of data on the Internet, which makes it difficult to check the digital word-of-mouth communications between the consumers (Mindoro et al., 2022). Some of the feedback might even be biased, confusing, or meaningless (Mahardika et al., 2022). This issue of varying quality of reviews imposes a challenge to the businesses and customers in choosing reliable opinions.

Furthermore, the majority of online rating platforms only consider reviews in the English language, which can cause misleading results since it does not take into account other languages (Samah et al., 2022). This is a crucial aspect since Malaysians mainly use Malay language in their daily interactions. A statistics report by Internet World Stats (Top Ten internet languages in the world - internet statistics, 2020) says that only 25.9% of internet users communicate in English, which underlines the importance of sentiment analysis in other languages (Scheibengraf, 2023).

The lack of visualization in sentiment analysis systems also becomes an issue. Data visualization is critical to aid the understanding of data analysis (Liang, 2020). Data analysis results alone can be difficult to be understood by general public if no visualization is provided. It also hinders the decision-making process since data interpretation takes a longer time in the absence of data visualization. This issue is validated by majority of the survey respondents (98.9%) since they emphasized the importance of incorporating visualization features into the platform dedicated for analyzing courier service reviews. Hence, data analysis and visualization must go hand-in-hand to deliver the most beneficial insights to the users.

To address these problems, there is a need for a classifying and visualizing system for X sentiment analysis of courier services. The data will be collected from X, where relevant tweets will be extracted. By using Natural Language Processing (NLP) techniques and machine learning algorithms, this system will process textual data in two languages - English and Malay. The system will then categorize posts into different sentiment categories as well as visualize the sentiments results for better understanding.

1.3 Objectives

Outlining objectives can bring clarity to the project's intended outcomes and act as a roadmap for the development of the system. The objectives proposed as the solution to the problems above are:

i. To design a web-based classification and visualization system for X sentiment analysis of courier services in Malaysia using Naïve Bayes.
ii. To develop the designed system using Plotly.
iii. To test the functionality and accuracy of the system.

1.4 Scope

The scope of this project is defined into four segments. Firstly, the people involved are courier service providers, namely J&T Express, SPX Xpress and DHL Express, and the respective customers. Secondly, the area covered is standard courier services in Malaysia. Next, the functionality involved is the analysis of X sentiment and its classification and visualization. Lastly, the data used is taken from X within the date range of 1st April 2023 until 31st March 2024.

1.5 Significance

The main significance of this project is it could benefit the courier service providers by providing clear insight into their business performance. By visualizing the customers feedback towards various aspects of their service, the companies could take insightful actions in order to improve their business performance. Additionally, it could also benefit the courier service customers in making a data-driven decision when choosing the best courier service for their preference. Both parties could benefit from the system since it automates the manual process of analysing the reviews online. This significantly reduces the time taken for decision-making. Not only that, the system also analyses sentiment in both English and Malay language, which is useful for Malaysian businesses and customers who utilize both languages in their daily interactions.

1.6 Summary

This chapter discusses the background of research, problem statement, objective, scope, and significance of the project. The goal of this project is to develop a classification and visualization system for X sentiment analysis of courier services in Malaysia using Naive Bayes to classify the sentiment analysis and Plotly to visualize the data.

CHAPTER 2

LITERATURE REVIEW

This chapter discusses the overview of courier services, courier service performance, sentiment analysis, visualization, development approach and related works previously done in recent years. Figure 2.1 depicts a visual overview of the literature review that will be discussed in depth in this chapter.

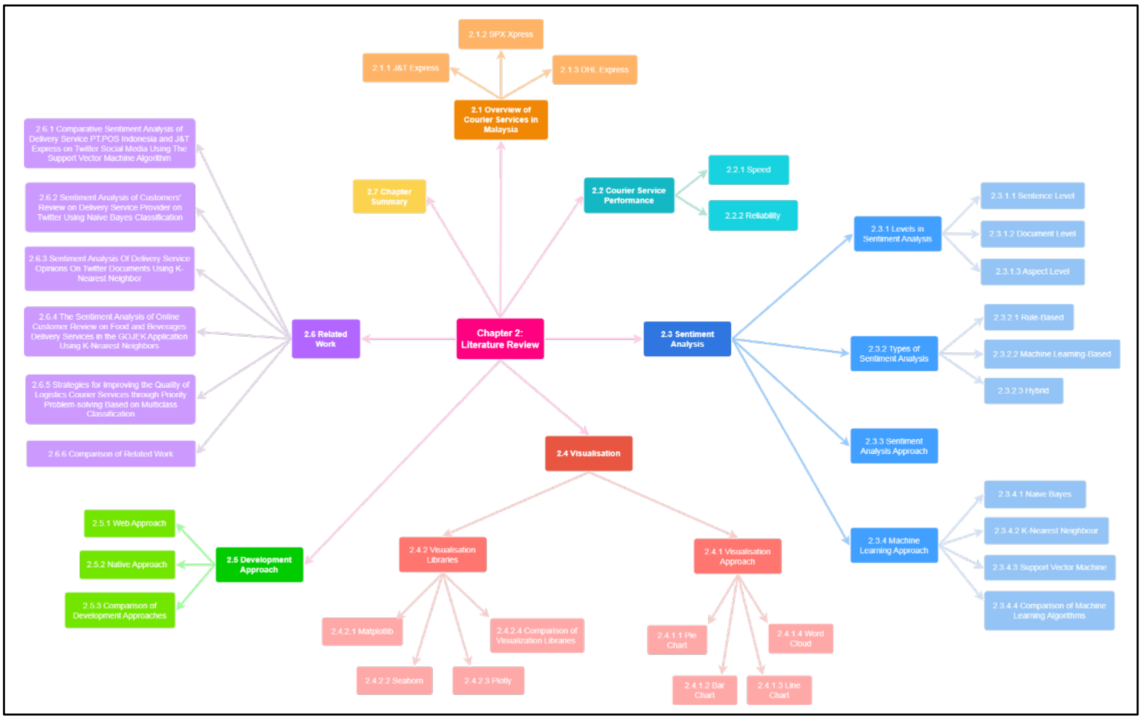


Figure 2.1 Literature Review Overview

2.1 Overview of Courier Services in Malaysia

Ever since the global pandemic, e-commerce has spurted its growth all over the world, including Malaysia. This has led to the rise of courier service businesses, which is essential in fulfilling the demanding job of parcel delivery from one place to another. A courier service refers to fast postal delivery that includes tracking and tracing capabilities (Royal Malaysian Customs Department, 2018). There are different types of courier services that offer different services according to various needs of businesses and individuals, mainly express, international, overnight and standard courier services. In this project, the focus will be on standard courier services since it is most used for e-commerce business in Malaysia. Standard courier services are commonly used for delivering parcels locally. Based on the conducted survey, the top three most popular standard courier services used by consumers in Malaysia are J&T Express (42.4%), SPX Xpress (31.5%) and DHL Express (22.8%).

2.1.1 J&T Express

Jet and Tony (J&T) Express is an Indonesian logistics and courier service company that operates in various countries in Southeast Asia, including Malaysia. The company has positioned itself as a key player in the e-commerce logistics sector, partnering with online businesses and platforms to handle the last-mile delivery of parcels. Figure 2.2 shows the logo of J&T Express.

Figure 2.2 Logo of J&T Express

2.1.2 SPX Xpress

SPX Xpress, or better known as Shopee Express, is a logistics service provided by Shopee, one of the leading e-commerce platforms in Southeast Asia. As Shopee grew its user base and expanded its reach, it invested in building a robust logistics infrastructure to support the increasing volume of e-commerce transactions. SPX Xpress helps streamline the process for sellers and enhances the shopping experience for buyers by ensuring timely and secure delivery of orders. Figure 2.3 shows the logo of SPX Xpress.

Figure 2.3 Logo of SPX Xpress

2.1.3 DHL Express

Dalsey, Hillblom, and Lynn (DHL) is a global logistics and courier company, and its operations in Malaysia reflect its international presence and comprehensive range of services. Given the growth of e-commerce, DHL has developed specific solutions to meet the logistics needs of online businesses. This includes services tailored for e-commerce retailers, such as efficient last-mile delivery and cross-border shipping. Figure 2.4 shows the logo of DHL.

Figure 2.4 Logo of DHL

After understanding the landscape of courier services in Malaysia, it becomes imperative to assess their performance to make informed decisions. Evaluating the performance of these services involves considering various factors. Let's delve into how we can effectively evaluate their performance

2.2 Courier Service Performance

Courier service performance is crucial for businesses and individuals relying on the timely and secure delivery of packages. The overall effectiveness of a courier service can be evaluated based on many key factors, such as delivery speed, reliability, traceability, customer service and cost. According to Performance Standards for Postal Services (Malaysian Communications and Multimedia Commission, 2011), the service performance standard for domestic parcel service is based on its speed and reliability. Hence, this project will be using these two criteria in evaluating the courier service performance. The factors are discussed further in the next subsection.

2.2.1 Speed

Speed, in context of delivery, is a value-added service indicating the duration from when an item is ordered to its delivery at the customer's address (Tandon, 2022). For businesses, the promptness of delivery holds significant weight, particularly in industries where timely distribution of goods is essential for maintaining customer satisfaction and competitive edge. Therefore, a dependable courier service is expected to not only offer short delivery times but also consider factors such as distance and urgency to ensure seamless and timely delivery of shipments.

2.2.2 Reliability

Reliability in courier services is defined as the consistency in meeting promised delivery times and maintaining the quality of service provided (Mentzer et al., 2001). A reliable courier service is crucial for the secure and timely transportation of goods. It goes beyond meeting deadlines, encompassing careful handling, accurate tracking, transparent communication, and responsive customer support. Being a reliable courier service ensures the customers' trust in handling their items. Ultimately, reliability is the cornerstone of a courier service's reputation, influencing customer loyalty and the success of businesses relying on efficient logistics.

In summary, the evaluation of courier service performance hinges on crucial factors such as speed and reliability. Speed is pivotal for ensuring prompt delivery, while reliability underscores the trustworthiness and efficiency of courier services. With these criteria in mind, we now turn our focus to sentiment analysis, aiming to delve deeper into customer perceptions and experiences with courier services.

2.3 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is the process of categorizing texts into positive, negative, or neutral sentiment (Gupta, 2018). It is a natural language processing (NLP) technique that is most commonly used for text classification. Sentiment analysis aims to examine individuals' opinions in order to assist businesses in their growth efforts. Sentiment analysis has a wide range of applications, including social media monitoring, customer feedback analysis, and market research.

2.3.1 Levels in Sentiment Analysis

Sentiment analysis can be performed at three levels, which are sentence level, document level, and aspect level (Nandwani & Verma, 2021). Each level offers unique insights into the sentiment, ranging from individual sentences to the entire document or specific aspects mentioned within the text.

2.3.1.1 Sentence Level

At sentence level sentiment analysis, it determines whether a particular sentence expresses a positive, negative, or neutral sentiment (Pandey & Vishwakarma, 2024). The goal is to determine the sentiment polarity of each sentence independently, without considering the sentiment of the entire document or specific aspects within the text. This level of analysis is useful for understanding the sentiment conveyed within shorter textual segments, such as tweets, reviews, or comments.

2.3.1.2 Document Level

In contrast to the sentence level, document-level sentiment analysis considers the entire document as a whole to determine its overall polarity (Mehta & Pandya, 2020). The objective is to determine the overall sentiment or emotional tone conveyed by the document. This level of analysis provides a holistic view of the sentiment present in the text and is commonly used in applications such as movie reviews, product feedback analysis, or sentiment analysis of news articles.

2.3.1.3 Aspect Level

Aspect-level sentiment analysis examines specific aspects or features of sentiment within textual data, providing insights into the sentiment expressed towards particular entities or attributes mentioned in the text (Nuha & Lin, 2023). Unlike document or sentence-level analysis, this approach dives deeper into understanding sentiments expressed towards specific entities, attributes, or topics within the text. It is especially valuable in domains like product reviews, where users express opinions about various features within a single review.

2.3.2 Types of Sentiment Analysis

Sentiment analysis can be categorized into three types (Li, 2024), which are the rule-based approach, the machine learning-based approach, and a hybrid approach that combines elements of both rule-based and machine learning-based approaches.

2.3.2.1 Rule-Based

Rule-based sentiment analysis is a technique employed for identifying and delineating particular aspects or characteristics within a provided text (Prabhune et al., 2023). These rules often involve the use of lexicons, dictionaries, or linguistic rules to assign sentiment scores to words or phrases. The sentiment of a text is then determined based on the aggregation of these scores. While rule-based approaches can be straightforward to implement and interpret, they may struggle with ambiguity and may not capture nuanced sentiments effectively.

2.3.2.2 Machine Learning-Based

Machine learning approach is employed for sentiment polarity classification, which involves categorizing sentiments as negative, positive, or neutral using both training and testing datasets (Birjali et al., 2021). These algorithms can then classify new texts based on the patterns learned during training. Machine learning approaches are often more flexible and capable of capturing complex relationships within text data. However, they require large amounts of labeled data for training and may be computationally intensive.

2.3.2.3 Hybrid

Hybrid sentiment analysis combines elements of both rule-based and machine learning-based approaches. This method first uses predefined rules or lexicons for initial sentiment classification, requiring minimal labeled data, and then employs machine learning to improve results (Li, 2024). This combination leverages the strengths of both approaches, allowing for more accurate and robust sentiment analysis results. Hybrid approaches can provide a balance between interpretability and performance, making them suitable for various sentiment analysis tasks.

In summary, the machine learning approach is preferred for sentiment analysis because it offers better scalability and adaptability, handling diverse language nuances more effectively than the rule-based approach. It is also simpler to implement and maintain compared to the hybrid approach, providing higher accuracy and robust performance.

2.3.3 Sentiment Analysis Approach

Feature engineering involves a process of identifying and eliminating redundant and irrelevant attributes from the feature list, consequently bolstering the accuracy of sentiment classification (Wankhade et al., 2022). The aim of this task is to extract valuable insights that describe essential characteristics found within the text (Birjali et al., 2021). When implementing feature engineering techniques, one fundamental aspect to consider is term frequency. Term frequency-inverse document frequency (TF-IDF), is a technique utilized to express text in matrix format, with each numerical value indicating the significance of these terms within a specific document (Nandwani & Verma, 2021). To implement TF-IDF, data preprocessing is done to remove noise and irrelevant information, then tokenize the text into individual words, and calculate the frequency of each term across the dataset.

Moreover, incorporating N-gram features can capture the contextual information and nuances in language usage that contribute to sentiment expression. In an N-gram vector representation, the text is depicted as a combination of distinct groups of N adjacent terms or words, known as N-gram means (Nandwani & Verma, 2021). Implementing N-gram features involves tokenizing the text into N-grams of varying lengths using sliding windows and representing each document as a feature vector containing the frequencies of different N-grams. By considering the context in which words appear, N-gram features provide valuable insights into the sentiment expressed in the text and enhance the effectiveness of sentiment analysis models.

2.3.4 Machine Learning Approach

Machine learning is a dynamic and rapidly evolving field within artificial intelligence that focuses on developing algorithms and models capable of learning from data to make predictions or decisions without being explicitly programmed. Its algorithms can learn from data, improving their performance over time and enabling the creation of sentiment analysis models that accurately predict the sentiment conveyed in textual data (Li, 2024). This approach aims to enable computers to recognize patterns, extract insights, and make informed decisions based on large datasets.

Machine learning techniques can be broadly categorized into supervised learning, unsupervised learning and semi-supervised learning (Rawat et al., 2022). In supervised learning, algorithms are trained using labelled data, where each input has a corresponding output label. Unsupervised learning works with unlabelled data to discover patterns or structures. Semi-supervised learning combines both approaches by using a small amount of labelled data alongside a larger set of unlabelled data. This project employs supervised learning, relying on labelled data to train the model.

In sentiment analysis, the most commonly used algorithms include Naive Bayes (NB), K-Nearest Neighbours (KNN), and Support Vector Machine (SVM). These algorithms are favoured due to their effectiveness in classifying text data and extracting sentiments from various textual sources (Joshi et al., 2021). The following subsections discuss each algorithm in more detail.

2.3.4.1 Naive Bayes

NB is a probabilistic classifier that employs Bayes' theorem to estimate the likelihood of a given set of features belonging to a specific label (Wankhade et al., 2022). It is widely used for classification tasks, particularly in natural language processing and sentiment analysis. The "naive" aspect stems from the assumption that features are conditionally independent given the class label, simplifying the computation of probabilities. Despite this simplification, NB often performs well in practice, particularly with large and high-dimensional datasets. In sentiment analysis, NB calculates the probability of a text belonging to each sentiment class based on word frequencies and assigns the class with the highest probability. Its efficiency, simplicity, and effectiveness in handling textual data make it a popular choice, although its performance relies on the independence assumption, which may not hold true in all real-world scenarios.

2.3.4.2 K-Nearest Neighbour

KNN is a machine learning algorithm used for both classification and regression tasks. KNN algorithm categorizes objects by identifying the closest resemblance to the object among the learning data (Pravina et al., 2022). The algorithm does not involve a training phase, making it simple and intuitive. However, it can be computationally expensive for large datasets and is sensitive to irrelevant features. Key considerations include the choice of the distance metric and the number of neighbors (k). Despite its limitations, KNN finds applications in various domains, such as image recognition and recommendation systems, where the emphasis is on local patterns in the data.

2.3.4.3 Support Vector Machine

SVM is a powerful supervised machine learning algorithm utilized for classification purposes, with its primary objective being to ensure it serves as the optimal linear separator for classification (Mehta and Pandya, 2020). It achieves this by maximizing the margin, which is the distance between the hyperplane and the nearest data points of each class. SVM is effective in handling both linear and non-linear relationships in data, thanks to the use of kernel functions that implicitly map input data into higher-dimensional spaces. This algorithm is particularly useful in scenarios with complex decision boundaries and is less prone to overfitting. SVM has found applications in diverse fields, including image classification, text categorization, and bioinformatics, owing to its ability to handle high-dimensional data and provide robust generalization performance.

2.3.4.4 Comparison of Machine Learning Algorithms

NB, KNN and SVM algorithms are compared in Table 2.1.

Table 2.1 Machine Learning Algorithms Comparison

Algorithm	Advantages	Disadvantages
NB	- Computationally efficient - Can be trained on relatively small datasets	- Poor performance if strong dependence occurs between features
KNN	- Does not require prior assumptions about data distribution	- Struggle with high-dimensional feature spaces - Computationally expensive for large dataset
SVM	- Effectively operates in high-dimensional spaces with limited training samples - Less prone to overfitting	- Sensitive to kernel selection and its hyperparameters - Computationally demanding for large dataset

(Source: Rajath et al., 2023)

When comparing NB, KNN and SVM algorithms, each exhibits distinct advantages and disadvantages. NB stands out for its computational efficiency and suitability for relatively small datasets, making it an attractive option in scenarios where computational resources are limited. However, it may falter if strong feature dependencies are present. KNN, on the other hand, does not require assumptions about data distribution, but it struggles with high-dimensional feature spaces and can become computationally expensive with large datasets. SVM is effective in high-dimensional spaces with limited training samples and is less prone to overfitting, yet it demands careful kernel selection and is computationally demanding for large datasets. Despite its simplicity, NB is chosen as the most suitable choice due to its computational speed, as well as its suitability for small datasets, like the one in this project. It simplifies the model selection process, and the algorithm can perform admirably even with the naive assumption.

2.4 Visualization

Data visualization typically seeks to aid individuals in grasping the significance of data by representing it visually (Khun & Thant, 2019). Effective data visualization plays a crucial role in making complex data more accessible, interpretable, and actionable. There are various approaches and tools available for data visualization, and the choice often depends on the nature of the data and the goals of the analysis.

2.4.1 Visualization Approaches

Visualization approaches refer to methods and techniques used to represent and communicate information visually. These approaches are commonly employed in various fields to help individuals better understand complex data, patterns, and relationships. Some common visualization approaches are discussed below.

2.4.1.1 Pie Chart

A pie chart is a popular way to show basic statistics (Kozak et al., 2015). It consists of a circular chart divided into slices, with each slice representing a proportion of the total. The size of each slice corresponds to the percentage it represents in relation to the entire dataset. Pie charts are commonly employed in scenarios where you want to illustrate the relative contribution of different categories to the total, such as in business reports or survey results. Figure 2.5 shows a pie chart example of population by age group.

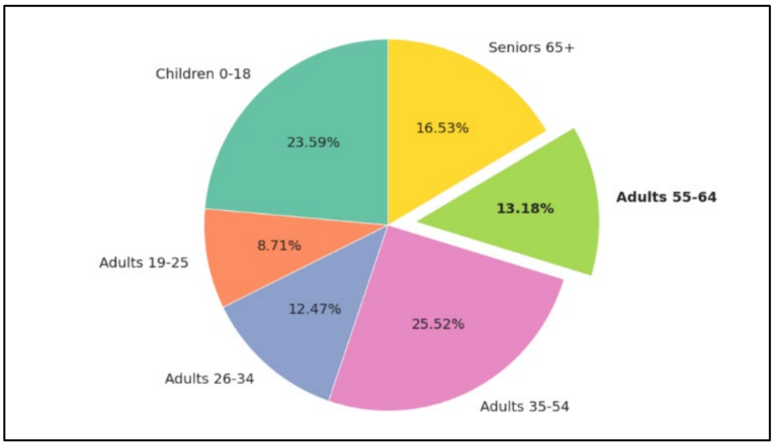


Figure 2.5 Pie Chart of Population by Age Group
(Source: Singh, 2022)

2.4.1.2 Bar Chart

Bar charts are commonly used for presenting categorical data (Starbuck, 2023). They consist of rectangular bars, either arranged horizontally or vertically, with the length of each bar proportional to the value it represents. Bar charts provide a clear visual representation of data distribution, making them widely used in market research, sales reports, and situations where a visual comparison of quantities is necessary. Figure 2.6 depicts a bar chart example of total sales of product.

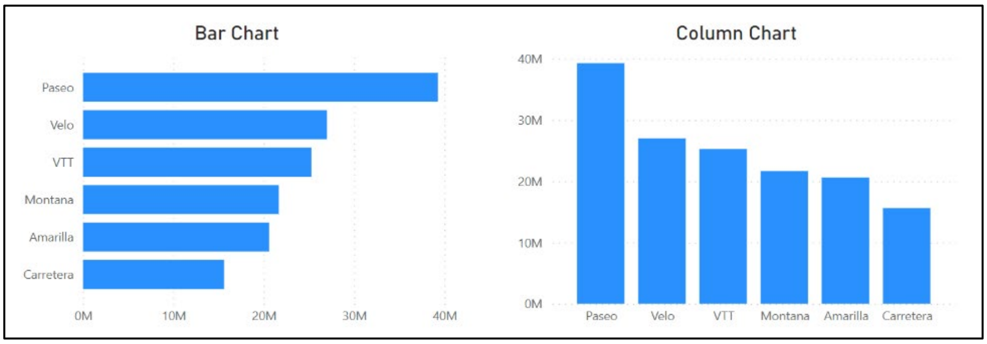


Figure 2.6 Bar Chart of Total Sales of Product
(Source: Gamble, 2022)

2.4.1.3 Line chart

Line charts visualize continuous data trends over time (Starbuck, 2023). This makes them ideal for illustrating continuous data sets like stock prices or sales figures. With axes typically depicting time or categories horizontally and numerical values vertically, they provide a clear snapshot of patterns and outliers, aiding decision-making in fields such as finance and scientific research. Figure 2.7 displays a line chart example of school enrolment by year.

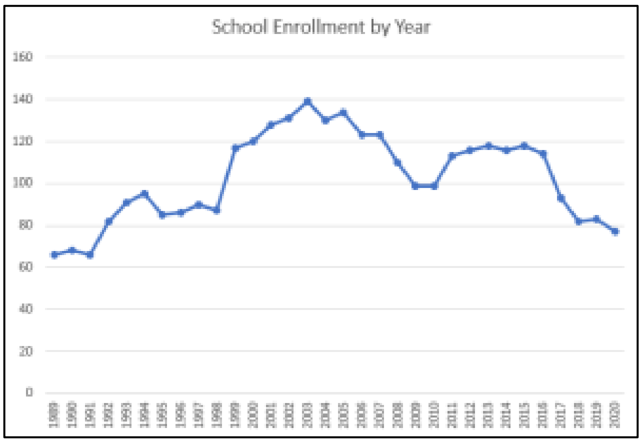


Figure 2.7 Line chart of School Enrolment by Year

(Source: Frost, 2021)

2.4.1.4 Word Cloud

Word clouds are a visual tool that display text by showing the most frequently used words in a visually appealing format (Heimerl et al., 2014). Words are arranged in a random or aesthetically pleasing manner, with the size of each word proportional to its frequency in the text. More frequent words appear larger and bolder, while less frequent words are smaller and lighter. Word clouds are often used in text analysis, social media sentiment analysis, and presentations to highlight key terms or topics within a body of text. Figure 2.8 shows a word cloud example of a product feedback analysis.

Figure 2.8 Word Cloud of Product Feedback Analysis

(Source: Visualizing Words: Inspiring Word Cloud Examples, 2024)

2.4.2 Visualization Libraries

Visualization libraries are software tools that provide pre-built functions and components for creating various types of visualizations. These libraries are designed to simplify the process of generating charts, graphs, and other visual representations of data, making it easier for developers and analysts to convey insights. Some popular visualization libraries are Matplotlib, Seaborn and Plotly.

2.4.2.1 Matplotlib

Matplotlib is a user-friendly data visualization library, constructed upon NumPy arrays with a low-level approach (Addepalli et al., 2023). It provides a wide range of static, animated, and interactive visualizations, making it a fundamental tool for data scientists and researchers. Its strength lies in its high level of customization and versatility, making it suitable for creating a wide range of static visualizations. Matplotlib offers extensive community support and documentation, making it a reliable choice for data scientists and researchers. However, its default styles may appear basic, necessitating additional customization for more modern aesthetics. Despite this, its seamless integration with Jupyter Notebooks and compatibility with various backends for rendering contribute to its enduring popularity. Figure 2.9 shows the logo of Matplotlib.

Figure 2.9 Logo of Matplotlib

2.4.2.2 Seaborn

Seaborn, a Python library, is designed for crafting statistical graphics, seamlessly integrating with panda's data structures, and offering a sophisticated interface to matplotlib (Addepalli et al., 2023). It excels in enhancing aesthetics and simplifying the syntax for creating attractive statistical visualizations. It comes with integrated color palettes and high-level functions that make it particularly suitable for users who prioritize ease of use and desire visually appealing plots. While it offers an enhancement in terms of aesthetics and ease of use, Seaborn has limitations in customization compared to Matplotlib, and its range of available plot types is somewhat more restricted. Figure 2.10 shows the logo of Seaborn.

Figure 2.10 Logo of Seaborn

2.4.2.3 Plotly

Plotly is a Python library designed to generate interactive visualizations accessible via the web (Addepalli et al., 2023). It supports multiple programming languages, including Python, R, and Julia. It distinguishes itself by offering interactive and web-based visualizations, making it an excellent choice for those who prioritize user engagement and collaboration. Its support for a wide range of chart types and animations further enhances its capabilities, catering to diverse data visualization needs. However, the library is relatively heavier, which can result in longer loading times, and it may not be the optimal choice for users seeking simple, static plots. Figure 2.11 shows the logo of Plotly.

Figure 2.11 Logo of Plotly

2.4.2.4 Comparison of Visualization Libraries

The visualization libraries are compared in Table 2.2.

Table 2.2 Visualization Libraries Comparison

Visualization Libraries	Advantages	Disadvantages
Matplotlib	- Offers many ways to customize visualizations. - Allows detailed control over how plots look. - Supports various types of plots and charts.	- Needs more code for some visualizations. - Takes longer to learn because of its complex syntax.
Seaborn	- Comes with built-in themes and color sets for appealing visuals. - Simplifies the creation of advanced statistical plots - Supports visualizing relationships between multiple variables.	- May have limitations for highly customized visualizations.
Plotly	- Offers interactive visuals that respond to user input. - Supports various chart types, including 3D plots and maps. - Enables the creation of interactive dashboards and web apps.	- May need more resources due to interactivity. - Learning to use interactive features effectively might take time.

(Source: Shaikh, 2023)

When comparing visualization libraries, Matplotlib offers extensive customization options and supports various plot types but requires more code and has a steep learning curve. Seaborn simplifies the creation of advanced statistical plots and comes with built-in themes, though it may have limitations for highly customized visuals. Plotly stands out with its interactive features, support for diverse chart types, and the ability to create interactive dashboards and web apps, making it the best choice for projects requiring dynamic and engaging visualizations. While it may require more resources, it is outweighed by the benefits of creating visually compelling and interactive visualizations.

2.5 Development Approach

In software development, the development approach refers to how applications are created and deployed. It involves choosing methods, tools, and platforms for building software. The development approach impacts how the software is designed and functions. There are two main approaches, which are the web and native approach. Each has its own strengths and weaknesses, so it is important to choose the right one based on the project's needs and resources. The following subsections analyses each approach in detail.

2.5.1 Web Approach

Web approach involves developing applications that are accessible via a web page, making them compatible with any device capable of running a web browser (Andrae & Kothuri, 2023). They use common web technologies such as JavaScript and Cascading Style Sheets 3 (CSS3). Web applications are compatible with any web browser, cost-effective, and can be easily deployed on multiple platforms. However, web apps have restricted access to device hardware and system functionalities and rely on constant server connections. The web approach is suitable for projects requiring broad accessibility and cost-effectiveness, such as informational websites, e-commerce platforms, and web-based tools.

2.5.2 Native Approach

Native approach involves creating applications tailored to a single platform (Andrae & Kothuri, 2023). Native apps leverage device features and optimized code, have direct access to device hardware, and able to utilize platform-specific APIs and libraries. However, developing native apps typically requires more resources, and may face limitations imposed by platform-specific guidelines and restrictions. The native approach is suitable for projects prioritizing performance, hardware integration, and user experience, such as mobile games, productivity apps, and multimedia-rich applications.

2.5.3 Comparison of Development Approaches

The development approaches are compared in Table 2.3.

Table 2.3 Development Approach Comparison

Development Approach	Advantages	Disadvantages
Web Approach	- Run on any web browser, cost-effective for development - Developed using common web technologies like JavaScript - Cost-effective for deploying on multiple platforms	- Limited access to device hardware and system functionalities - Less secure due to reliance on constant server connections
Native Approach	- Use device features for better performance and offline use - Access to device hardware like GPS and accelerometers - Supports advanced features like notifications	- Require more resources for development and may need redesign - Platform-specific restrictions may affect development

(Source: Selvarajah et al., 2013)

In choosing the development approach for this project, the benefits and drawbacks of both Web and Native approaches are weighed. Native approach provides better performance and access to device features like GPS, but it requires more resources and face platform-specific restrictions. Conversely, Web approach is cost-effective and accessible across multiple platforms, but it has limited hardware access and potential security risks. Considering the project's needs for broad accessibility and cost-effectiveness, the Web approach is preferred. It offers easy deployment on various platforms without significant resource investment.

2.6 Related Work

In order to gain clearer understanding of the project, previous related work done in the recent years are studied. By studying past works, it helps to form the project with a clearer vision as the works act as references to compare and contrast which methods are best to apply in this project.

2.6.1 Comparative Sentiment Analysis of Delivery Service PT.POS Indonesia and J&T Express on Twitter Social Media Using The Support Vector Machine Algorithm

The paper done by Euis Nur Fitriani Dewi, Aldy Putra Aldya, Andi Nur Rachman and Ara Ramdani is focused on the courier logistics domain. The study aims to carry out sentiment analysis on the opinions of users of Pos Indonesia and J&T Express delivery services on Twitter and to measure the performance of the SVM algorithm in classifying data. The data of PT.POS Indonesia and J&T Express reviews are taken from Twitter. The language taken for the reviews are in Indonesian. The algorithm used is SVM, where it achieved 80.14% accuracy using 70% training data and 30% test data. No visualization is done in this project.

2.6.2 Sentiment Analysis of Customers' Review on Delivery Service Provider on Twitter Using Naive Bayes Classification

The paper done by Ari Basuki is focused on the courier logistics domain. The study aims to evaluate the performance of the NB method for classifying customer feedback on courier delivery services obtained via Twitter. The data of J&T reviews are taken from Twitter. The language taken for the reviews are in Indonesian. The algorithm used is Naive Bayes, where it achieved 50.6% accuracy. Word cloud is the visualization done in this project.

2.6.3 Sentiment Analysis Of Delivery Service Opinions On Twitter Documents Using K-Nearest Neighbor

The paper done by Arsya Monica Pravina, Kretawiweka Nuraga Sani, Hendy Dwi Harfianto, Tesar Akram Pratama, Ari Fahrina and Yova Ruldeviyani is focused on the courier logistics domain. The study aims to analyze opinion sentiment of delivery service (JNE, J&T, TIKI, Pos Indonesia, and DHL Indonesia) and determine the aspect of timeliness and quality. The data of JNE, J&T, Pos Indonesia, TIKI and DHL Indonesia reviews are taken from Twitter. The language taken for the reviews are in Indonesian. The algorithm used is KNN, where it achieved 94.56% accuracy. The limitation of this paper is the unbalanced amount of data between delivery services was not considered. The number of tweets made about each delivery service can be different, as it follows how much that delivery service is being discussed by users on Twitter. Bar chart is the visualization done in this project.

2.6.4 The Sentiment Analysis of Online Customer Review on Food and Beverages Delivery Services in the GOJEK Application Using K-Nearest Neighbors

The paper done by Aal Fathrizqy Putra Mahardika, Aisyah Larasati and Abdul Muid is focused on the food logistics domain. The study aims to further analyze the reviews based on a machine learning approach and sentiment analysis to label each review as positive or negative. The data of Gojek reviews are taken from Google Play Store. The language taken for the reviews are in Indonesian. The algorithm used is KNN, where it achieved 83% accuracy. The limitation of this paper is the lack of information since not much review is collected from Google Play Store. Bar chart and ROC curve are the visualizations done in this project.

2.6.5 Strategies for Improving the Quality of Logistics Courier Services through Priority Problem-solving Based on Multiclass Classification

The paper done by Ratih Hendayani and Muhammad Cahyo Dharmawan is focused on the courier logistics domain. The study aims to review and determine which classification model is the most appropriate for the dataset, to find out the rating of sentiments and dimensions in order to measure the quality of logistics company services and to identify what problems need to be prioritized. The data of JNE reviews are taken from Twitter. The language taken for the reviews are in English. The algorithms used are NB and SVM, where it achieved 96.45% and 98.24% accuracy for NB and SVM respectively. No visualization is done in this project.

2.6.6 Comparison of Related Work

The related works are depicted in Table 2.4 to provide a clearer view to compare each aspect from the related works.

Table 2.4 Related Work Comparison

Reference	**Dewi et al. 2023**	**Basuki 2023**	**Pravina et al. 2022**	**Mahardika et al. 2022**	**Hendayani and Dharmawan 2020**	**Proposed system**
Title	Comparative Sentiment Analysis of Delivery Service PT.POS Indonesia and J&T Express on Twitter Social Media Using Support Vector Machine Algorithm	Sentiment Analysis of Customers' Review on Delivery Service Provider on Twitter Using Naive Bayes Classification	Sentiment Analysis Of Delivery Service Opinions On Twitter Documents Using K-Nearest Neighbor	The Sentiment Analysis of Online Customer Review on Food and Beverages Delivery Services in the GOJEK Application Using K-Nearest Neighbors	Strategies for Improving the Quality of Logistics Courier Services through Priority Problem-solving Based on Multiclass Classification	X Sentiment Analysis: Classification and Visualization of Courier Services in Malaysia
Domain	Courier logistics	Courier logistics	Courier logistics	Food logistics	Courier logistics	Courier logistics
Dataset	Twitter	Twitter	Twitter	Google Play Store	Twitter	X (Twitter)
Language	Indonesian	Indonesian	Indonesian	Indonesian	English	English, Malay
Algorithm	SVM	NB	KNN	KNN	NB, SVM	NB
Findings	Achieved 80.14% accuracy	Achieved 50.6% accuracy	Achieved 94.56% accuracy	Achieved 83% accuracy	Achieved 96.45% and 98.24% accuracy for NB and SVM	No findings yet.
Visualization	None	Word cloud	Bar chart	Bar chart, ROC curve	None	Bar chart, line chart, word cloud

Table 2.4 summarizes several studies on sentiment analysis, highlighting the methodologies, datasets, and results. Dewi et al. (2023) used SVM algorithm to analyze Twitter data for courier services, achieving an accuracy of 80.14%. Basuki (2023) employed NB classification on Twitter reviews, reaching 50.6% accuracy. Pravina et al. (2022) utilized KNN on Twitter data and achieved 94.56% accuracy. Mahardika et al. (2022) analyzed Google Play Store reviews for food delivery services using KNN, obtaining 83% accuracy. Hendayani and Dharmawan (2020) combined NB and SVM for courier services, achieving high accuracy rates of 96.45% and 98.24%, respectively.

In conclusion, while other algorithms such as KNN and SVM have demonstrated higher accuracy in similar studies, NB was chosen for this project due to its advantages in handling large volumes of text data efficiently and its robustness in various text classification tasks. NB offers simplicity, ease of implementation, and good performance with smaller datasets or noisy data, making it a practical choice for the sentiment analysis of courier services in Malaysia. Additionally, its speed in processing large-scale data align with the project's requirements.

2.7 Summary

This chapter discusses the overview of courier services, courier service performance, sentiment analysis, visualization, development approach and related works. The courier services that are chosen for this project are J&T Express, SPX Xpress and DHL Express and the performance that is measured from each courier are speed and reliability. For the sentiment analysis, machine learning approach is implemented. NB algorithm is chosen for this project due to its simplicity, computational speed, as well as its suitability for small datasets. Visualization approaches like bar chart, pie chart, line chart and word cloud will be used in this project. After evaluating a few visualisation libraries, Plotly is chosen for this project due to its interactive visualizations, support of numerous chart types and animations, and support of multiple programming languages. After comparing native and web development approaches, the web approach is picked for this project. Its convenient deployment on multiple platforms without the need of large resource investment makes it more appealing for this project. This chapter establishes a strong base for this project based on the study of the previous related works in recent years.

CHAPTER 3

METHODOLOGY

This chapter discusses the project methodology which acts as a guideline in completing the project. The main issue discussed includes the project model and all the phases involved in the model. The discussion obtained from this chapter will offer a clearer view into the processes required for successful completion of the project.

3.1 Introduction

Software Development Life Cycle (SDLC) is a methodology that outlines each stage of the software development process thoroughly (Gupta, 2023). It encompasses a series of well-defined phases, typically including requirements analysis, system design, implementation, testing, deployment, and maintenance. The SDLC aims to ensure the development of high-quality software that meets user requirements and industry standards while managing resources effectively. Various methodologies, such as the waterfall model, agile, and spiral model, can be employed within the SDLC framework to facilitate the development process, with each phase contributing to the overall success and reliability of the software product.

3.1.1 Waterfall Model

The waterfall methodology is a linear and sequential approach to software development that follows a structured, step-by-step process (Lutkevich, 2022). As shown in Figure 3.1, the model is divided into distinct phases, including requirement analysis, design, implementation, testing, deployment, and maintenance (Khan, 2023). Each phase must be completed before moving on to the next stages. The waterfall model emphasises thorough documentation and planning upfront, making it suitable for projects with well-defined and stable requirements.

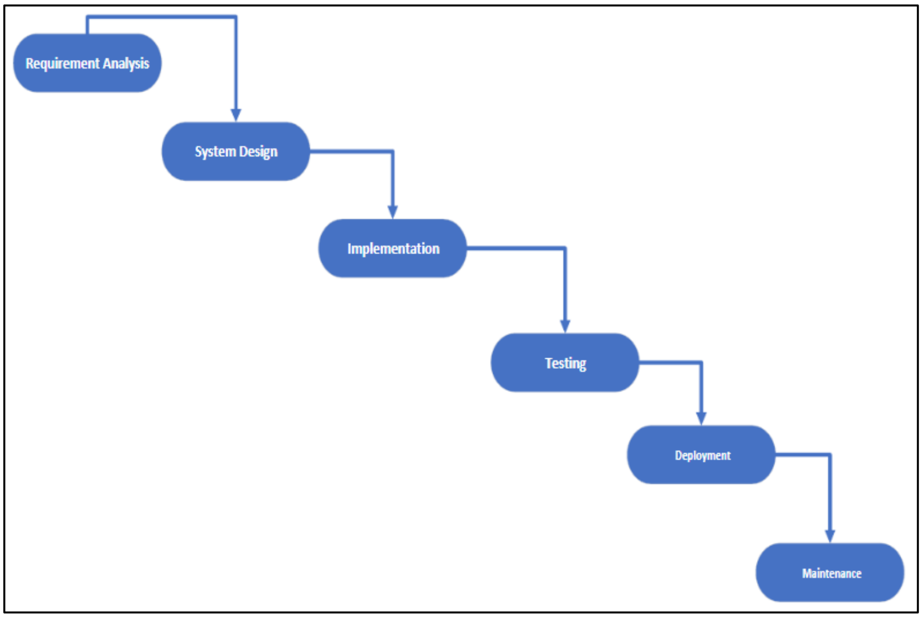


Figure 3.1 Waterfall Model

(Source: Khan, 2023)

However, the waterfall model has its limitations which include difficulties in managing project progress within stages and its inability to accommodate requirement changes (Khan, 2023). Managing progress within stages can be challenging because each phase must be completed before moving on to the next, making it hard to address issues that arise in earlier stages without disrupting the entire process. Additionally, the model's rigid structure means that once the requirements are defined at the beginning, it is difficult to incorporate changes later on. This inflexibility can lead to problems if new requirements emerge or if there are changes in project scope, resulting in a final product that may not fully meet the users' needs.

Despite its structured approach, the waterfall model limitations make it less suited for dynamic and unpredictable development environments. Hence, this calls for a more flexible version of the model, the modified waterfall model, in order to overcome the original model constraints.

3.1.2 Modified Waterfall Model

The rigidity of the waterfall model has led to the emergence of a more flexible and iterative development methodology. The modified waterfall model is a variation of the traditional waterfall methodology that introduces some flexibility to accommodate certain drawbacks of the rigid sequential approach (Petersen et al., 2009). Unlike the classic waterfall model, limited iteration is allowed between phases, allowing for feedback and adjustments. This modification aims to address issues related to late-stage changes and stakeholder feedback that may arise in the development process. While the modified model maintains a structured and systematic approach, it provides a degree of adaptability to changes without compromising the overall integrity of the development process. Figure 3.2 shows the modified waterfall model.

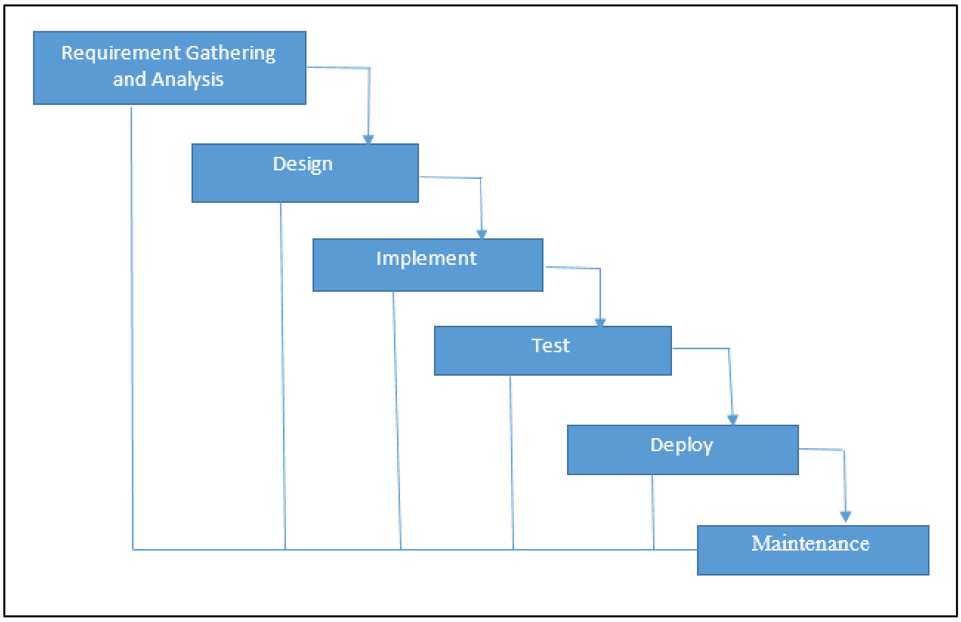


Figure 3.2 Modified Waterfall Model

(Source: Melinmani, 2014)

As seen in Figure 3.2, there are six stages in the modified waterfall model which are requirement analysis, design, implementation, testing, deployment, and maintenance (Melinmani, 2014). Due to the degree of this project, the deployment and maintenance phases are discarded. This left us with four stages, which are requirement analysis, design, implementation, and testing. Documentation stage is added to the end of the methodology to combine all progress and complete the project. The overall phases of this project methodology are shown in Table 3.1.

Table 3.1 Overview of Project Methodology

<table>
<tr><th>Phase</th><th>Activities</th><th>Deliverable</th></tr>
<tr><td rowspan="2">Requirement Analysis</td><td rowspan="2">• Identify research area
• Identify problem statement
• Define objectives and scope
• Study previous related works
• Perform extensive research work</td><td><u>Chapter 1</u>
• Research topic found
• Problem statement found
• Objectives defined
• Scope defined
• Significance defined</td></tr>
<tr><td><u>Chapter 2</u>
• Literature review performed</td></tr>
<tr><td>Design</td><td>• Determine project methodology
• Define flow of the system
• Define system interface
• Define algorithm design
• Define Gantt chart of the project</td><td><u>Chapter 3</u>
• Methodology defined
• Flowchart defined
• Use case diagram defined
• User interface design defined
• Algorithm design defined
• Gantt chart defined</td></tr>
<tr><td>Implementation</td><td>• Perform data collection
• Perform data pre-processing
• Perform model building
• Perform model evaluation
• Perform data visualization</td><td><u>Chapter 4</u>
• Research design defined</td></tr>
<tr><td>Testing</td><td>• Perform unit testing
• Perform integration testing
• Perform system testing</td><td><u>Chapter 5</u>
• Results and findings discussed
• Functionality test case completed</td></tr>
<tr><td>Documentation</td><td>• Combine all chapters</td><td>• Final report completed</td></tr>
</table>

3.2 Requirement Analysis Phase

Requirement analysis phase is the first phase of the modified waterfall model which involves identifying the research area, problem statement, defining objectives and scope, and performing literature review. Firstly, defining a research area entail defining the specific domain of study that will be the focus of the research. This phase involves a thorough exploration of potential topics, ensuring that the chosen research area is relevant, feasible, and aligned with the researchers' expertise and interests.

Secondly, defining a problem statement is a crucial element in research, emerging from the process of defining the research area. The problem statement outlines the challenges, gaps, or issues within this chosen area, providing a clear, concise statement of the problem the research aims to address. It establishes the study's significance, relevance, and direction, ensuring alignment with researchers' expertise and interests for an effective investigation.

Next, defining objectives and scope. It is important to set what is desired to be achieved in the study, defining the specific outcomes and boundaries within which the research will be conducted. This process involves setting clear and achievable objectives that guide stages of the research, ensuring a focused and purposeful investigation.

Lastly, literature review is a critical aspect of the requirement analysis phase. Existing scholarly works, academic papers, and relevant literature are studied to understand the current state of knowledge in the chosen research area. The literature review serves multiple purposes, including identifying gaps in existing knowledge, informing the development of research questions, and establishing a theoretical framework for the study. It also helps avoid duplication of works and build upon existing ones.

In summary, the requirement analysis phase serves as the foundational step in the modified waterfall model, marking the project's inception. By addressing the key elements early on, the modified waterfall model establishes a solid groundwork for subsequent phases, laying the groundwork for a systematic and well-informed approach to the overall research process.

3.3 Design Phase

Design phase is the second phase in the modified waterfall model which transforms the conceptual ideas and requirements outlined in the previous requirement analysis phase into a structured system design. The primary objectives include defining logical system design, defining flow of the system, and defining system interface. The logical system design involves creating a blueprint that details the overall structure and system components, their relationships, and the data flow between them. This phase yields several outcomes which are flowchart, use case diagram, user interface design, and algorithm design.

3.3.1 Flowchart

The flow of the system is defined to picture how the system is going to work and how the variables interact with each other. The process can be visualized using a flowchart. A flowchart is a visual representation of a process or system using standardized symbols to illustrate the sequence of steps, decision points, and actions involved. It serves as a graphical tool to convey the logical flow and structure of a system, making complex processes more understandable. The flow of the system is depicted in Figure 3.3.

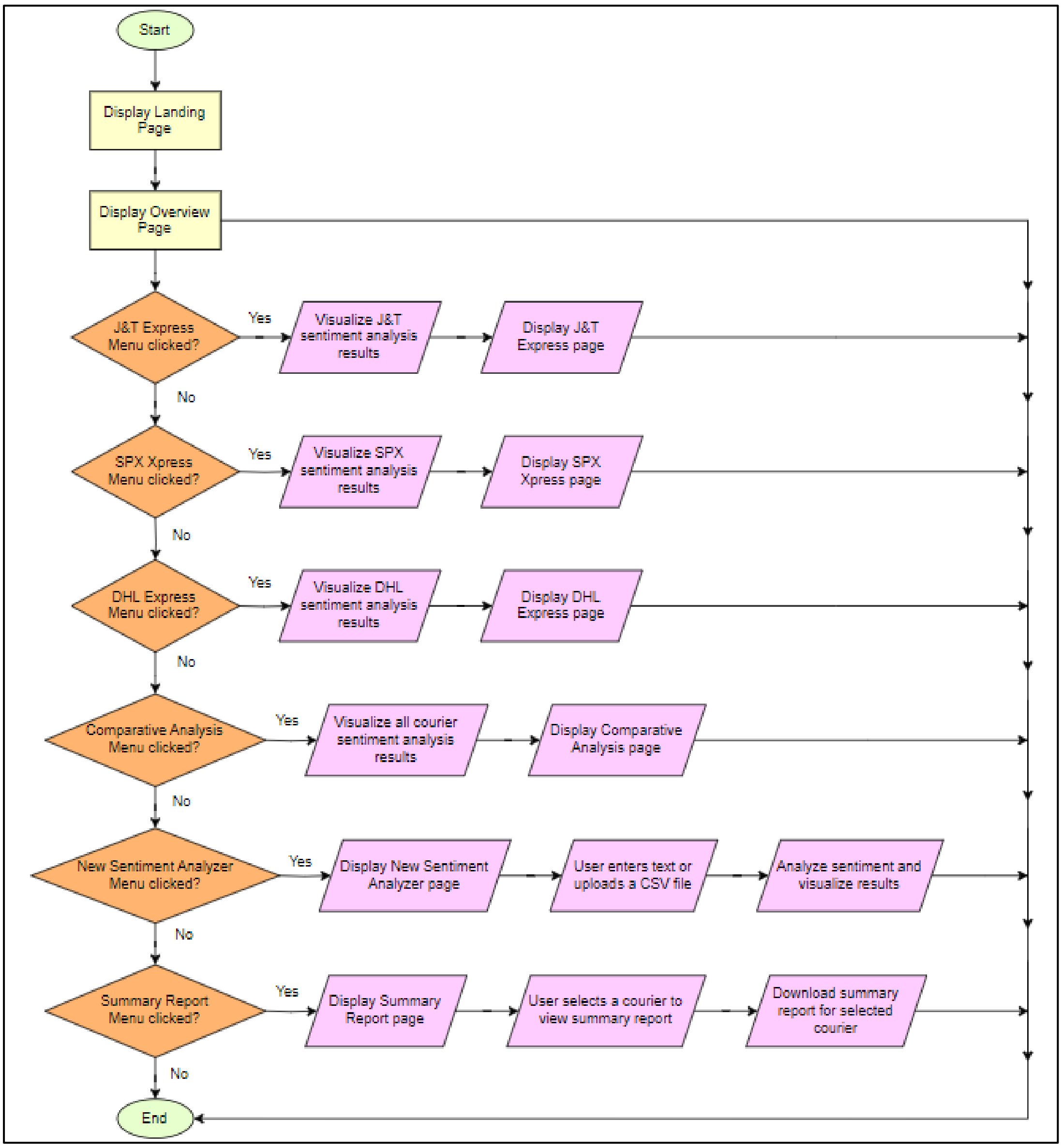


Figure 3.3 Flowchart of System

The flow of the system starts with the landing page. When user clicks on the 'Start' button, they will be redirected to the overview page. Here, the summary of sentiment analysis of courier services in Malaysia is shown. On the left side of the page, the navigation bar is located. User can find multiple buttons that can take them to different pages, such as 'J&T Express', 'SPX Xpress', 'DHL Express', 'Comparative Analysis', 'New Sentiment Analyzer' and 'Summary Report'.

When user clicks on the ‘J&T Express’ button, they will be redirected to the J&T Express page. Here, they can find the visualizations of J&T Express reviews. It is the same case for ‘SPX Xpress’ and ‘DHL Express’ buttons, where user will be redirected to the SPX Xpress and DHL Express pages respectively. When user clicks on the ‘Comparative Analysis’ button, the courier services performance is compared. The reviews comparison of the courier services will be visualized.

When user clicks on the ‘New Sentiment Analyzer’ button, they will be redirected to a new page where they can enter a single text or upload a CSV file to analyze its sentiment. The result will be displayed on the same page. When user clicks ‘Summary Report’ menu, they will be redirected to Summary Report page. Here, they can download a PDF file consisting of a summary of the courier performance report.

3.3.2 Use Case Diagram

Use case diagram is a visual representation in the Unified Modelling Language (UML) that illustrates the various ways a system interacts with users or other systems to achieve specific functionalities. It provides a high-level overview of the system's behaviour by capturing different use cases, each representing an interaction scenario between an actor and the system. This project has two users, guest and courier admin, both has different access to the system. The use case diagram of this project is shown in Figure 3.4.

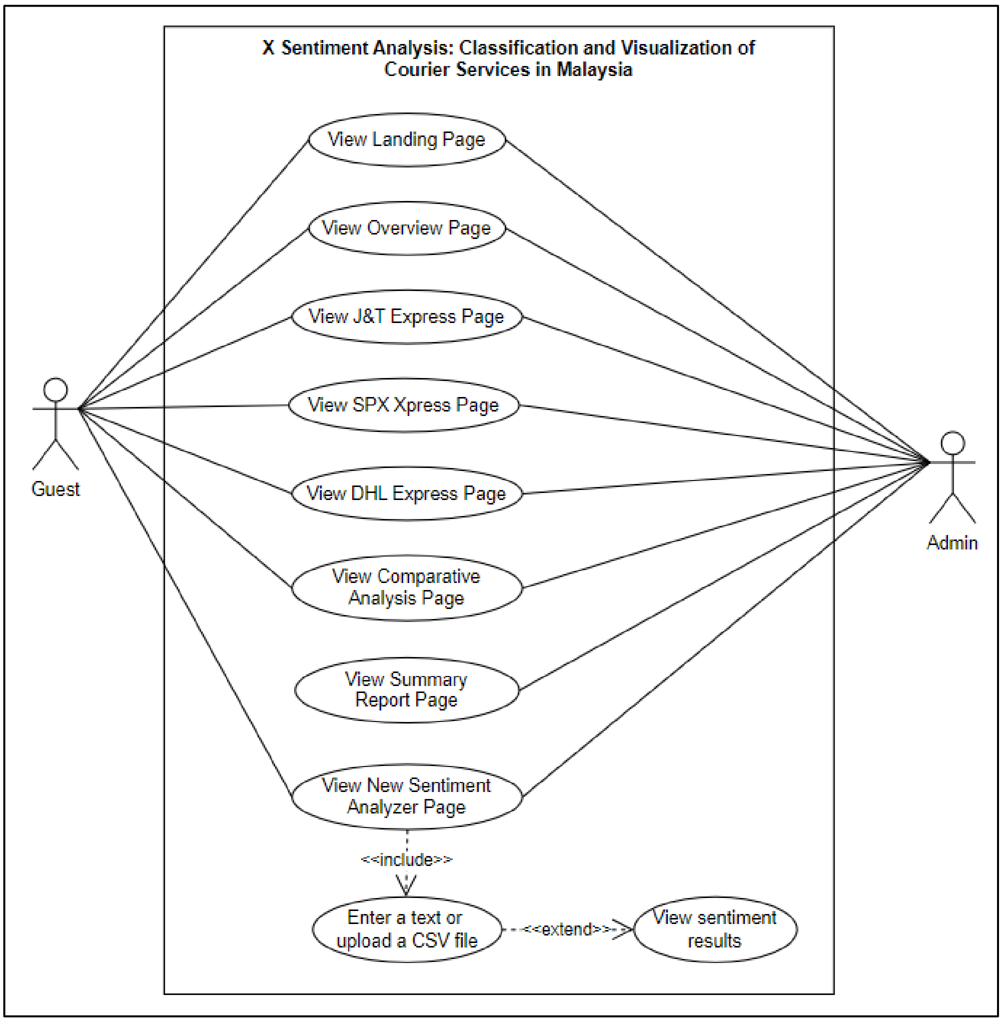


Figure 3.4 Use Case Diagram of System

The use cases are described in Table 3.2 to show how the system will be used to achieve specific tasks. Each use case represents a specific interaction between the user and the system.

Table 3.2 Use Case Description

Use Case	Description
View Landing page	User will view the landing page as soon as the system runs.
View Overview page	User will view the overview page which displays the data analysis summary of J&T Express, SPX Xpress and DHL Express.
View J&T Express page	User will view the sentiment analysis results of J&T Express data and its visualizations in the form of charts.
View SPX Xpress page	User will view the sentiment analysis results of SPX Xpress data and its visualizations in the form of charts.

View DHL Express page	User will view the sentiment analysis results of DHL Express data and its visualizations in the form of charts.
View Comparative Analysis page	User will view the visualization of the sentiment comparisons between the courier services.
View Summary Report Page	User will select to view a summary report of selected courier service. A summary report file will be downloaded.
View New Sentiment Analyzer page	User will view the analyzer page which allow user to analyze a text or a CSV file.
Enter Text or Upload CSV File	User will enter a single text or upload a CSV file into the sentiment analyzer.
View Sentiment Results	User will view the results displayed on the analyzer page.

3.3.3 User Interface Design

User interface design is the process of creating the visual and interactive elements of a software system to enhance user experience. It involves designing the graphical layout, interactive components, and overall aesthetics of the user interface, ensuring that it is intuitive, visually appealing, and user-friendly. The design should facilitate efficient and enjoyable interactions between users and the system. The proposed interface design for the system is shown in the figures below.

Figure 3.5 shows the Landing page, which is the first page that user sees when they enter the system There is a ‘Start’ button which will redirect the user to the Overview page.

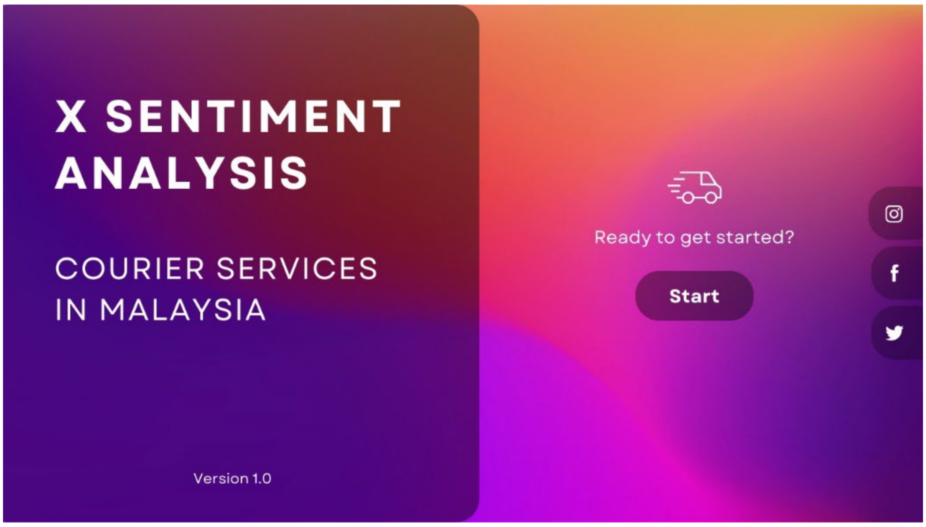


Figure 3.5 Landing Page

Figure 3.6 shows the Overview page, which displays the visualization of all courier services involved in this project. Several visualization figures are used such as line chart and bar chart to briefly summarize the review analysis of the courier services.

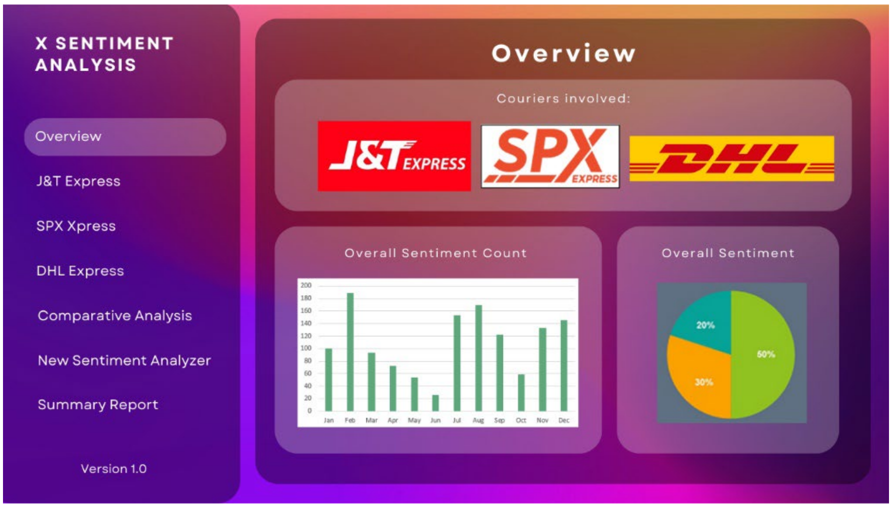


Figure 3.6 Overview Page

Figure 3.7 shows the J&T Express page, which depicts several data visualization techniques for the analysis results of J&T Express reviews

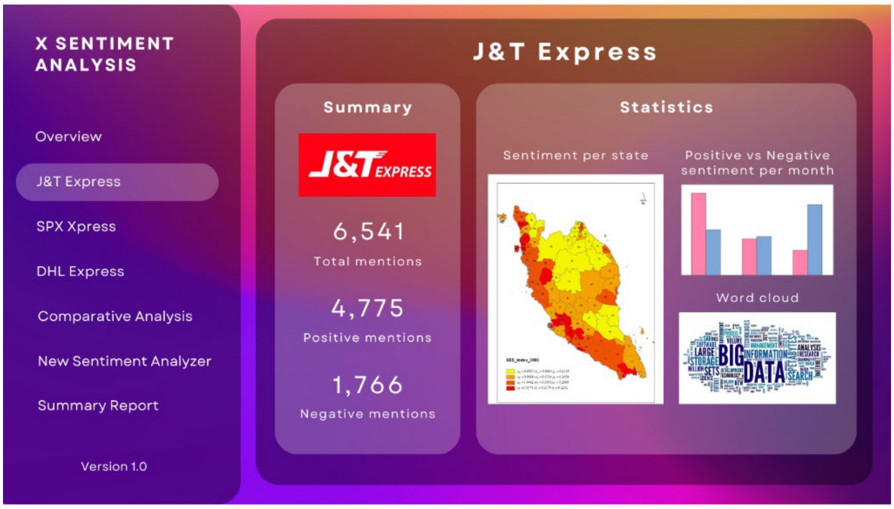


Figure 3.7 J&T Express Page

Figure 3.8 shows the SPX Xpress page, which depicts several data visualization techniques for the analysis results of SPX Xpress reviews

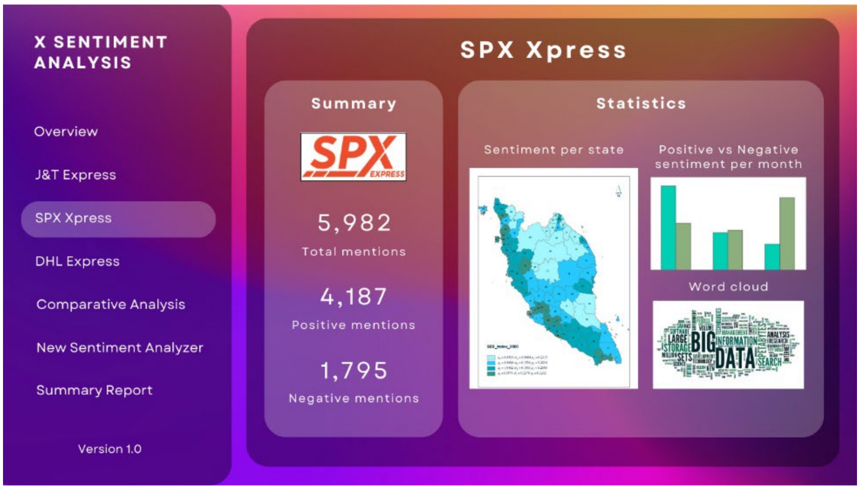


Figure 3.8 SPX Xpress Page

Figure 3.9 shows the DHL Express page, which depicts several data visualization techniques for the analysis results of DHL Express reviews.

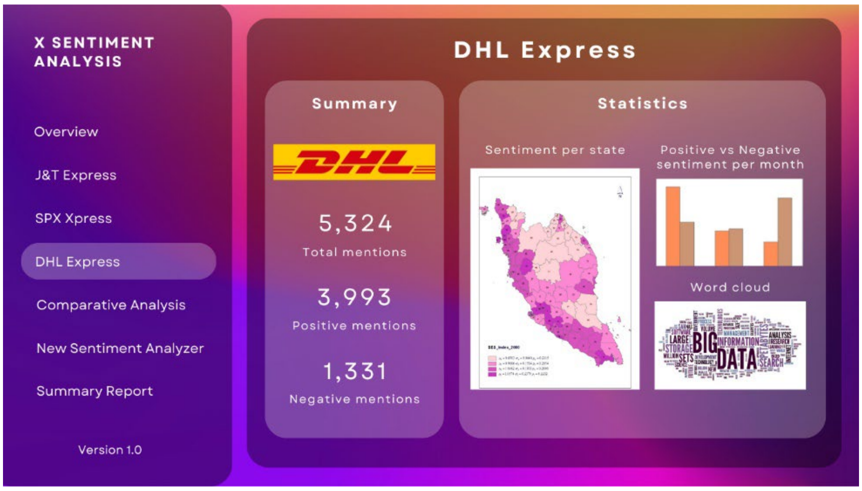


Figure 3.9 DHL Express Page

Figure 3.10 shows the Comparative Analysis page, which shows the comparative analysis results of the selected courier services.

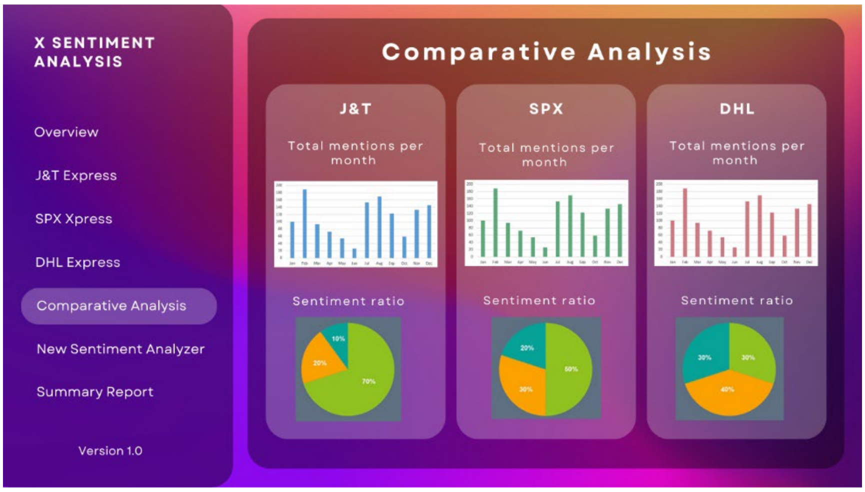


Figure 3.10 Comparative Analysis Page

Figure 3.11 shows the New Sentiment Analyzer page, which allows user to enter a text or upload a CSV file into the sentiment analyzer. The analysis results will be shown on the page.

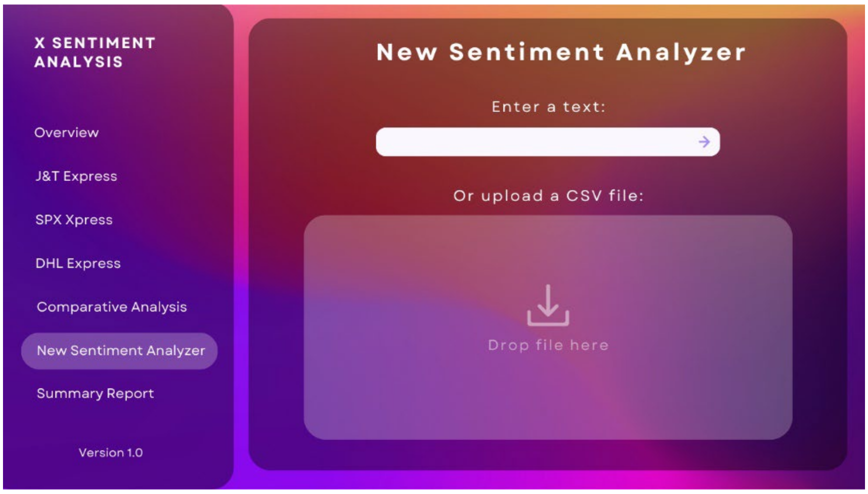


Figure 3.11 New Sentiment Analyzer Page

Figure 3.12 shows the Summary Report page, which allows user to download a summary report file of the courier performance.

Figure 3.12 Summary Report Page

3.3.4 Algorithm Design

Algorithm design is the process of defining a systematic set of instructions to efficiently solve a computational problem. This is where the step-by-step solution is defined to build the model which will be used to classify the sentiments. Firstly, labeled datasets are collected from Github. Next, the data is pre-processed where several processes are performed, including data cleaning, tokenization, stopword removal, lemmanization and feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF).

The pre-processed datasets are then split into training and testing datasets. The training dataset is used for model development. In this step, NB algorithm is applied to calculate the probability of each sentiment class (positive, negative, neutral) for each individual sentiment. Then, the sentiment is classified as sentiment class with the highest probability. After training the model, it is evaluated using testing data. This step yields confusion matrix and performance metrics which measures the model performance. Accomplishing all these steps is necessary to generate the developed NB model. The algorithm design for this project is depicted in Figure 3.13.

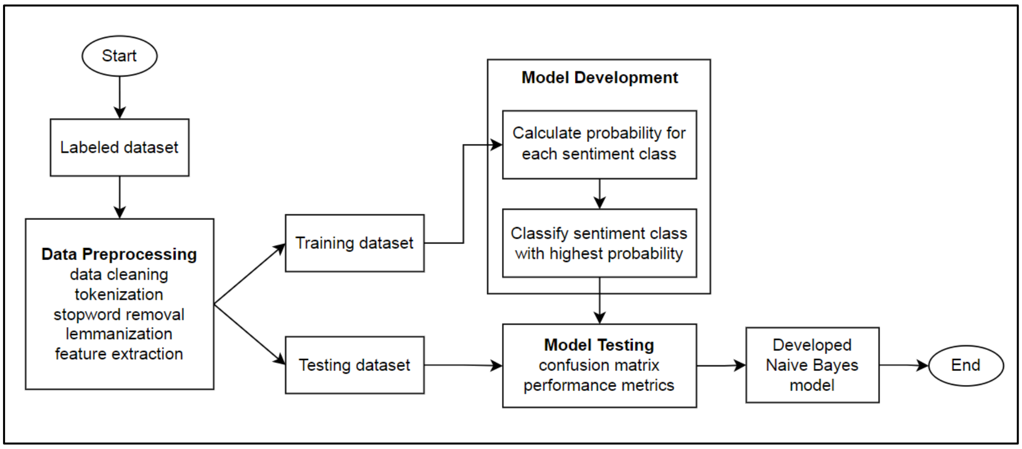


Figure 3.13 Algorithm Design of System

3.4 Implementation Phase

Implementation phase is a phase in the modified waterfall where the developed model is applied on real-world data. To provide a better overview, a research design is used to visualize the process. Research design refers to the strategy that is used to systematically investigate a research problem. It encompasses the overall framework that guides the collection, analysis, and interpretation of data in a study. The research design for this project is shown in Figure 3.14.

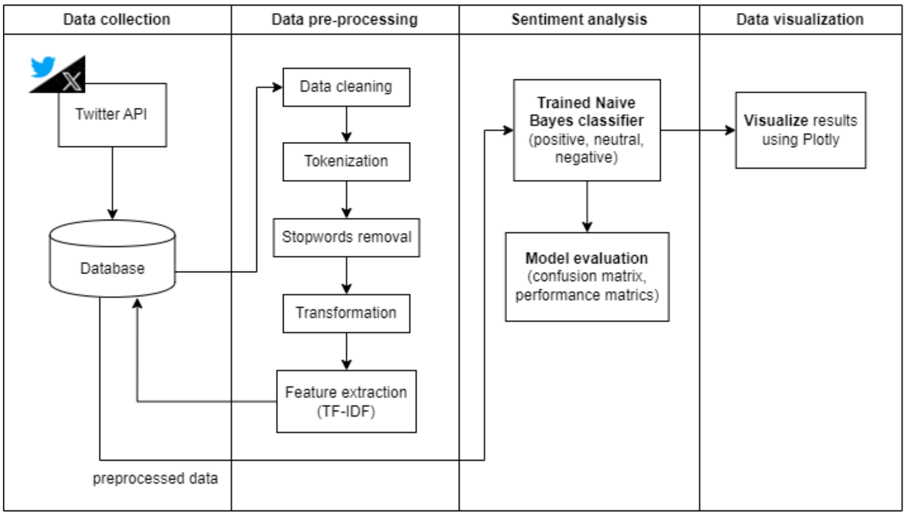


Figure 3.14 Research Design of System

Firstly, data collection is carried out by web-scraping tweets from Twitter from 1st April 2023 until 31st March 2024. The data is stored in the database. Next, the collected data undergo pre-processing, where several processes are carried out. The processes include data cleaning, tokenization, stopword removal, lemmatization and feature extraction, where TF-IDF is applied. The pre-processed data is written into the database. Then, the data undergoes sentiment analysis. This is where the previously developed NB classifier is used to classify each sentiment as either positive, negative or neutral. Then, the model is evaluated to measure its performance. Confusion matrix and performance metrics such as accuracy, precision, recall and F1 score are evaluated during this stage. Lastly, the sentiment results are visualize using Plotly library in Python to provide the users a better understanding of the results.

3.5 Testing Phase

Testing phase is a critical stage where the developed system undergoes inspection to ensure that it meets the specified requirements and functions correctly. Since the modified waterfall model allows for limited iteration, any issues identified during testing can be addressed before proceeding to the next stages.

Functionality testing focuses on verifying whether the software behaves as intended. This process begins with unit testing, where individual components are examined in isolation to ensure their correct functionality. Integration testing follows, validating the interaction between different components. System testing assesses the complete system, ensuring all parts work together seamlessly. The functional test case of this project is depicted in Table 3.3.

Table 3.3 Functional Test Case

Test scenario	Expected
View Main Page	System displays main page
View Overview Page	System displays overall courier services performance using visualization
View J&T Express Page	System displays J&T Express sentiment analysis results in several visualization forms
View SPX Xpress Page	System displays SPX Xpress sentiment analysis results in several visualization forms
View DHL Express Page	System displays DHL Express sentiment analysis results in several visualization forms
View Comparison Page	System displays comparison of sentiment analysis between different courier services
View Summary Report page	System displays the summary report of courier performance.
View New Sentiment Analyzer Page	System displays page to input new sentiment to be analyzed.
Enter Text or Upload CSV File	System allows user to enter a single text or upload a CSV file into the sentiment analyzer.
View Sentiment Results	System displays the sentiment analysis results on the analyzer page.

3.6 Hardware and Software Requirements

The suitable software and hardware components are essential when creating a project system. This is to ensure the success and efficiency of the project development. In this project, it will be developed as a web-based platform to allow accessibility from any device with internet connection and a web browser. It is imperative that the hardware and software can efficiently run web system development tools without encountering any performance issues. Table 3.4 and 3.5 depicts the hardware and software requirements used for developing this project.

Table 3.4 Hardware Requirements

Hardware	Description
Laptop	• Model: HP EliteBook 840 G9 • Processor: Intel(R) Core(TM) i5-8350U • System type: 64-bit operating system, x64-based processor • Memory: 16.0 GB • Storage: 165 GB

Table 3.5 Software Requirements

Software	Description
Windows 11 Pro	Device operating system.
Microsoft Word	Platform to write project report.
Microsoft Excel	Database to store scraped data.
Mendeley	Platform for references citation.
Diagrams.net	Platform to create LR overview, use case diagram, flowchart, algorithm design and research design.
Jupyter Notebook	IDE to perform data pre-processing and develop Naive Bayes model.
Visual Studio Code	IDE to develop system.

3.7 Project Timeline

The project timeline unfolds through distinct phases, starting with requirement analysis where research area, objectives, scope and literature review are defined. Subsequently, the design phase is carried out to define methodology, logical system design, and research design. The implementation phase involves data collection, data pre-processing, data analysis and data modelling. The testing phase is then performed to carry out unit testing, integration testing, system testing, and acceptance testing. Simultaneously, the ongoing documentation phase combines all chapters in the correct format to produce a complete final report. This approach is visualized using Gantt chart to give a clearer overview of the project timeline to ensure a systematic project development process. The Gantt chart can be found in Appendix C.

3.8 Summary

This chapter discusses the modified waterfall model and its phases. Each phase is described in detail, which helps to provide a clear understanding of how the project is going to be executed. Diagrams are used to clearly depict the overall overview of the system.

CHAPTER 4

SYSTEM DEVELOPMENT

This chapter delves into the development of the system. The phase is done in parts, which are front-end and back-end stages. During back-end development, tasks such as data collection, data pre-processing, and NB classifier model development is done. While in front-end phase, the user interface and system's functionalities are developed. The chapter meticulously examines the procedure, offering a comprehensive breakdown.

4.1 Back End Development

The back-end of a web application, encompasses server-side operations such as data processing and model creation to ensure seamless functionality of the front end. In this project, the back-end encompasses tasks such as collecting data, pre-processing data, developing a NB model, and conducting accuracy tests on the model. The subsequent subsection delves deeper into these processes.

4.1.1 Data Collection

There are two types data needed to be collected for this project, which are labelled data and real-world data. For labelled data, it is collected from Github, Kaggle, mycourier (https://www.mycourier.my/) and lookp (https://www.lookp.com/) websites. The Malay dataset was collected from a GitHub repository named malaysian-dataset by huseinzol05. It contains 62,385 negative data, 62,890 positive data and 909 neutral data. The English dataset is obtained from Kaggle, containing 7,781 negative data, 8,582 positive data and

11,118 neutral data. Figure 4.1 and 4.2 display some of the training data for Malay and English models.

Out[2]:

	text	sentiment
0	"saya sayang awak adalah yang terbaik! !"	positive
1	"dia mempunyai kesan ke atas semua orang"	positive
2	"Terima kasih untuk maklum balas anda . saya s...	positive
3	NaN	positive
4	selamatkan untuk akhirnya bergabung dengan t...	positive
5	"Saya menjawab kucing bodoh membantu saya mena...	positive
6	"im bertemu dengan salah satu besties saya mal...	positive
7	"sakit adalah benar-benar murah apabila sakit ...	positive
8	teman-teman anda membuat anda sup"	positive
9	"saya sangat cemburu	positive

Figure 4.1 Training Data for Malay Model

Out[3]:

	text	sentiment
0	I`d have responded, if I were going	neutral
1	Sooo SAD	negative
2	bullying me	negative
3	leave me alone	negative
4	Sons of ****,	negative
5	http://www.dothebouncy.com/smf - some shameles...	neutral
6	fun	positive
7	Soooo high	neutral
8	Both of you	neutral
9	Wow... u just became cooler.	positive

Figure 4.2 Training Data for English Model

Both Malay and English datasets are based from a general domain, not specified to courier domain. This can cause inaccuracy in developing the model, since this project's domain is related to courier services. Hence, an additional 'Courier' dataset is manually created for this project. The reviews of multiple couriers in Malaysia are extracted from mycourier and lookp websites, which are two platforms used to rate company services in Malaysia. These reviews are classified based on its star-rating, where those with star-rating 0 to 2 is classified as negative, while those with 3 stars are rated as neutral, and those rated with 4 to 5 stars are considered positive. A snippet of coding for this process is shown in Figure 4.3.

```python
# Function to classify reviews based on star ratings
def classify_review(stars):
    if stars >= 0 and stars <= 2:
        return 'negative'
    elif stars == 3:
        return 'neutral'
    elif stars >= 4 and stars <= 5:
        return 'positive'
    else:
        return 'unknown'  # Handle unexpected values

# Apply the classification to the DataFrame
df['sentiment'] = df['stars'].apply(classify_review)
```

Figure 4.3 Snippet of Coding for Classify Reviews

This tailored dataset ensures the model is equipped with specific insights, improving the sentiment analysis accuracy. Figure 4.4 shows a portion of the training dataset for Courier model.

	text	sentiment
0	very bad services, won't buy with buyer if cou...	negative
1	This one particular delivery man was rude and ...	negative
2	Curi barang cust , tktau jnt mana yang buat ta...	negative
3	Kurier paling teruk. Penghantaran selangor ke ...	negative
4	J&T LABUAN LEMBAB ! status parcel out for deli...	negative
5	??jnt NOT GOOD AT ALL DELIVER ALSO SLOW!?? no ...	negative
6	Worst experience - Purchases made on 2 Feb 202...	negative
7	This is the worst courier service I ever exper...	negative
8	barang delay lama ..lpstu chat last minit ckp s...	negative
9	JNT JOHOR BAHRU NOOB jangan kerja kalau layan ...	negative

Figure 4.4 Training Data for Courier Model

For real-world data, web scraping was performed on X to collect data which will be used as testing data. Sentiments about courier services were scraped using relevant keywords such as 'j&t express', 'spx xpress' and 'dhl express'. The date range for the tweets was from 1st April 2023 to 1st April 2024. The web scraping process was done using Tweet-Harvest tool and the scraped data was stored in CSV files. Fifteen columns were scraped, which were 'conversation_id_str', 'created_at', 'favorite_count', 'full_text', 'id_str', 'image_url', 'in_reply_to_screen_name', 'lang', 'location', 'quote_count', ‘reply_count’, ‘retweet_count‘, ‘tweet_url’, ‘user_id_str’ and ‘username’. A snippet of scraped data for J&T Express can be seen in Figure 4.5.

1	conversati	created_a	favorite_c	full_text	id_str	image_url	in_reply_t	lang	location	quote_cou	reply_cou	retweet_c	tweet_url	user_id_st	username
2	1.65E+18	Tue Apr 1	0	@jntexpre	1.65E+18	https://pb	jntexpress	en	Kuala Lum	0	0	0	https://tw	2.2E+09	lilmintchoc
3	1.69E+18	Sun Jul 30	0	Hi @jntex	1.69E+18			en	Anywhere	1	0	0	https://tw	29926964	fadira
4	1.65E+18	Fri Apr 14	0	From Sere	1.65E+18	https://pbs.twimg.cc		en	Kuala Lum	0	0	0	https://tw	57538197	ariefamron
5	1.66E+18	Sat May 2	0	@jntexpre	1.66E+18	https://pb	jntexpress	en		0	0	0	https://tw	1.37E+18	DrHebaElM
6	1.67E+18	Sun Jun 1	0	Hi @jntex	1.67E+18			en	Anywhere	0	0	0	https://tw	29926964	fadira
7	1.73E+18	Thu Dec 0	1	J&t e	1.73E+18	https://pbs.twimg.cc		en	Ipoh, Mal	0	1	0	https://tw	2E+08	pennyteh
8	1.73E+18	Mon Nov	0	Waited 11	1.73E+18	https://pbs.twimg.cc		en		0	0	0	https://tw	43691177	nay_bella
9	1.69E+18	Tue Aug 1	0	@jntexpre	1.69E+18	https://pb	jntexpress	en		0	0	0	https://tw	7.22E+17	alyahlim
10	1.73E+18	Wed Nov	1	Hi @jntex	1.73E+18	https://pbs.twimg.cc		en	Seoul, Rep	0	0	0	https://tw	94350310	qyymqamar
11	1.66E+18	Thu May 1	0	Hi @jntex	1.66E+18	https://pbs.twimg.cc		en	Kuala Lum	0	0	0	https://tw	32419164	__aimiaziza
12	1.65E+18	Tue Apr 2	0	@Shopeel	1.65E+18	https://pb	ShopeeMY	en	Kuala Lum	0	1	0	https://tw	2.26E+08	Ladyaf3
13	1.74E+18	Tue Jan 02	0	Becareful	1.74E+18	https://pbs.twimg.cc		en	Kuala Lum	0	1	0	https://tw	41548106	boysyamros
14	1.67E+18	Fri Jun 16	0	Around 1.	1.67E+18	https://pb	_teijashriy	en		0	1	0	https://tw	1.49E+18	_teijashriy1
15	1.67E+18	Fri Jun 16	0	The staff t	1.67E+18	https://pb	_teijashriy	en		0	0	0	https://tw	1.49E+18	_teijashriy1
16	1.70E+18	Fri Sep 01	0	Hello @jnt	1.70E+18			en		0	0	0	https://tw	1.05E+18	ntsyzmi
17	1.69E+18	Wed Aug	2	maybe yo	1.69E+18			en	Kuantan, P	0	0	2	https://tw	1.52E+18	HoOne96_G
18	1.65E+18	Mon Apr 1	0	klaw statu	1.65E+18	https://pbs.twimg.cc		en	somewhe	0	0	0	https://tw	1.3E+09	yaya_TVXQ
19	1.66E+18	Wed May	0	@jntexpre	1.66E+18	https://pb	jntexpress	en	Cyberjaya,	0	0	0	https://tw	1.63E+18	rnambiar24
20	1.65E+18	Thu Apr 2	6	@jntexpre	1.65E+18	https://pb	jntexpress	en	#NAME?	0	2	7	https://tw	3.12E+08	KogiEXO

Figure 4.5 Snippet of Scraped Data for J&T Express

4.1.2 Data Pre-Processing

Throughout the data pre-processing phase of the project, numerous measures were implemented to refine and organize the dataset for subsequent analysis. The initial stage involved data cleaning, entailing the elimination of redundant or extraneous information to enhance data quality and facilitate model development in subsequent phases. To ensure applicability in real-world scenarios, it became imperative to streamline the dataset by removing any superfluous columns. Consequently, extraneous columns were expunged, leaving only the ‘created_at’, ‘full_text’ and ‘lang’ columns intact. Figure 4.6 depicts the data with relevant columns.

	created_at	full_text	lang
0	Thu Mar 28 22:11:07 +0000 2024	5 green months in a row for the SPX. Congrats ...	en
1	Thu Mar 28 21:15:52 +0000 2024	Q1 2024 DXY up Gold up Rates up Dow up SPX up ...	en
2	Fri Mar 29 02:07:56 +0000 2024	Tucker Carlson Elon Musk Andrew Tate Alex Jone...	en
3	Thu Mar 28 21:30:43 +0000 2024	Look at the SPX Gold and Btc monthly charts. T...	en
4	Thu Mar 28 21:26:03 +0000 2024	PCE - hot powell - hawkish We ll open below $s...	en

Figure 4.6 Snippet of Data with Relevant Column

Next, missing data are reported and removed, if any. Subsequently, various cleaning operations are applied to the 'full_text' column. This includes removing URLs, mentions, hashtags, emojis, special characters, retweets, and duplicates. The text is then converted to lowercase, and extra whitespace is removed. Additionally, the 'created_at' column is converted to datetime format and reformatted to display only the date. Finally, the 'created_at' column is renamed to 'date'. These operations collectively ensure the dataset is cleaned and formatted appropriately for further analysis or model development. Figure 4.7 shows a snippet of the code for the process.

```
# Remove Hashtags
df['cleaned_text'] = df['full_text'].apply(lambda x: re.sub(r'#+', '', x))

# Remove Emojis
df['cleaned_text'] = df['full_text'].apply(lambda x: x.encode('ascii', 'ignore').decode('ascii'))

# Remove Special Characters
df['cleaned_text'] = df['full_text'].apply(lambda x: re.sub(r'[^\w\s]', '', x))

# Remove Retweets
df = df[~df['cleaned_text'].str.startswith('RT')]

# Remove duplicates
df.drop_duplicates(subset='cleaned_text', inplace=True)

# Convert to Lowercase
df['cleaned_text'] = df['cleaned_text'].str.lower()

# Remove Extra Whitespace
df['cleaned_text'] = df['cleaned_text'].apply(lambda x: re.sub('\s+', ' ', x).strip())

# Convert 'created_at' column to datetime
df['created_at'] = pd.to_datetime(df['created_at'], format='%a %b %d %H:%M:%S %z %Y')

# Reformat the date without the timezone part
df['created_at'] = df['created_at'].dt.strftime('%Y-%m-%d')

# Rename 'created_at' to 'date'
df.rename(columns={'created_at': 'date'}, inplace=True)
```

Figure 4.7 Snippet of Coding for Data Cleaning

Moving on, tokenization is performed to split text into individual tokens. This function is applied to the 'cleaned_text' column to generate tokenized representations. The NLTK library is used for English text, while Sastrawi library is used for Malay text. Figure 4.8 depicts a snippet of the code for the process.

```python
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Tokenization function
def tokenize_text(text):
    return word_tokenize(text)

# Tokenize 'cleaned_text' column
df['tokens'] = df['cleaned_text'].apply(tokenize_text)

# Create a Stemmer object
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# Function to apply stemming to tokens
def stem_tokens(tokens):
    stemmed_tokens = [stemmer.stem(token) for token in tokens]
    return stemmed_tokens

# Apply stemming to 'tokens' column
df['tokens'] = df['tokens'].apply(stem_tokens)
```

Figure 4.8 Snippet of Coding for Tokenization

Following tokenization, stopwords removal is done on the dataset. The NLTK library's English and Indonesian stopwords list are used. Since Malay and Indonesian languages often share similar phrases, Indonesian stopwords list are used for the process. To further clean the data, Malay stopwords are read from a text file and removed from the tokenized text data. Furthermore, a custom stopwords list comprising terms specific to the dataset, such as ‘jnt’, ‘dhl’ and ‘courier’, is defined and employed. Figure 4.9 shows the snippet of coding for the described process.

```python
# Get Indo stopwords from NLTK
stopwords_list = set(stopwords.words('indonesian'))

# Function to remove Indo stopwords
def remove_stopwords_indo(tokens):
    filtered_tokens = [token for token in tokens if token.lower() not in stopwords_list]
    return filtered_tokens

# Apply remove_stopwords_indo function to cleaned text
df['tokens'] = df['tokens'].apply(remove_stopwords_indo)

# Read Malay stopwords from a text file
with open('stopwords-malay.txt', 'r', encoding='utf-8') as file:
    stopwords_list = set(file.read().splitlines())

# Function to remove Malay stopwords
def remove_stopwords_malay(tokens):
    filtered_tokens = [token for token in tokens if token.lower() not in stopwords_list]
    return filtered_tokens
```

Figure 4.9 Snippet of Coding for Stopwords Removal

After that, the Natural Language Toolkit (NLTK) library's WordNet Lemmatizer is used during lemmatization to reduce tokens to their base or dictionary form. Moreover, a LabelEncoder object is instantiated to transform the 'lang' column into numerical labels (0 for English and 1 for Indonesian). The dataset is then partitioned into English and Indonesian subsets and saved as separate CSV files. Figure 4.10 depicts a snippet of coding for the process.

```
# Initialize WordNet Lemmatizer
lemmatizer = WordNetLemmatizer()

# Function to lemmatize tokens
def lemmatize_tokens(tokens):
    lemmatized_tokens = [lemmatizer.lemmatize(token) for token in tokens
    return lemmatized_tokens

# Apply lemmatization to 'tokens' column
df['tokens'] = df['tokens'].apply(lemmatize_tokens)
from sklearn.preprocessing import LabelEncoder

# Create a label encoder object
label_encoder = LabelEncoder()

# Transform 'lang' into numerical labels (EN-0 IN-1)
df['lang'] = label_encoder.fit_transform(df['lang'])
# Divide into EN and IN dataset

#ENGLISH
df_en = df[df['lang'] == 0]
df_en.to_csv('spx_en.csv', index=False)

#INDONESIAN
df_in = df[df['lang'] == 1]
df_in.to_csv('spx_in.csv', index=False)
```

Figure 4.10 Snippet of Coding for Lemmatization and Dataset Partition

4.1.3 Naive Bayes Algorithm

The NB classifier assumes a crucial role in model evaluation by employing probabilistic computations to classify the dataset. However, before applying the NB classifier, it is imperative to preprocess the data using the TF-IDF approach. These preprocessing steps are indispensable for converting text-based inputs into a numerical representation, either in matrix or vector format, which can be readily interpreted by machine learning algorithms. The TF (Term Frequency) and IDF (Inverse Document Frequency) are fundamental components of the TF-IDF transformation process. TF measures the frequency of a term, i within a document, j, while IDF calculates the inverse document frequency of term, i. This transformation results in vectors that are normalized to unit length.

The integration of TF-IDF with the NB classifier proves advantageous in identifying pertinent characteristics or keywords in the data indicative of the target class. This is attributed to the NB assumption of feature independence given the class labels. Leveraging the TF-IDF representation, the NB classifier can more accurately classify and predict based on the transformed data, effectively capturing these independent features.

After the data has been transformed, the NB algorithm is applied by initially training it on a labeled dataset. This training process involves computing the prior probabilities of each class, determined by the frequency of each class in the training data, and the likelihood of each feature given each class, based on the feature frequency in the training data for each class. These prior probabilities and likelihoods are subsequently combined using Bayes' theorem to compute the posterior probabilities of each class for the input data. The predicted class is then determined by selecting the class with the highest posterior probability.

To enhance the accuracy of the NB model, the technique of cross-validation was employed. This involved dividing the training data into multiple folds and sequentially using each fold as a test set. This approach helps prevent overfitting and promotes better generalization to new data. Specifically, the data was split into five folds, and GridSearchCV was utilized to explore different alpha values for the NB classifier. The alpha parameter controls the model's smoothing, a critical factor in text classification tasks. A parameter grid consisting of alpha values ranging from 0.1 to 10.0 was defined, resulting in 25 separate model trainings within each fold, totaling 125 model trainings. Figure 4.11 shows the snippet of coding to train the model.

```python
# Create TF-IDF vectorizer
tfidf_vectorizer = TfidfVectorizer()

# Fit and transform the text data
x = tfidf_vectorizer.fit_transform(x)

# Initialize KFold with the number of splits
kfold = KFold(n_splits=5, shuffle=True, random_state=0)

# Define the parameter grid for alpha values
param_grid = {'alpha': [0.1, 0.5, 1.0, 2.0, 5.0, 10.0]}

# Initialize Multinomial Naive Bayes classifier
classifier = MultinomialNB()

# Initialize GridSearchCV with the classifier, parameter grid, and cross-validation strategy
grid_search = GridSearchCV(classifier, param_grid, cv=kfold, scoring='accuracy')

# Initialize lists to store accuracy scores for each fold
accuracy_scores = []

# Iterate over each fold
for fold, (train_index, test_index) in enumerate(kfold.split(x)):
    print(f"Fold {fold + 1}:")

    # Split data into train and test sets
    x_train, x_test = x[train_index], x[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    # Perform GridSearchCV on the training data
    grid_search.fit(x_train, y_train)

    # Get the best estimator from the grid search
    best_classifier = grid_search.best_estimator_

    # Make predictions on the test set
    y_pred_test = best_classifier.predict(x_test)

    # Calculate accuracy for this fold
    fold_accuracy = accuracy_score(y_test, y_pred_test)
    accuracy_scores.append(fold_accuracy)
# Print the best hyperparameters
print("Best hyperparameters:", grid_search.best_params_)

# Get the best estimator from the grid search
best_classifier = grid_search.best_estimator_

# Evaluate classifier
conf_matrix = confusion_matrix(y, best_classifier.predict(x))
accuracy = accuracy_score(y, best_classifier.predict(x))
classification_rep = classification_report(y, best_classifier.predict(x))
```

Figure 4.11 Snippet of Coding for Model Training

The best classifier obtained from the grid search was then evaluated on the entire dataset. Figure 4.12, 4.13 and 4.14 depict the parameters used for the Malay, English and Courier model.

```
Best hyperparameters: {'alpha': 0.1}
```

Figure 4.12 Parameters for Malay model

```
Best hyperparameters: {'alpha': 2.0}
```

Figure 4.13 Parameters for English model

```
Best hyperparameters: {'alpha': 0.5}
```

Figure 4.14 Parameters for Courier model

4.1.4 Model Evaluation

The model evaluation involves assessing its performance through various metrics, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These metrics are utilized to compute classification accuracy, precision, recall, F1 score, and the confusion matrix, providing insights into the model's effectiveness. The confusion matrix visually represents the comparison between actual target values and predicted values generated by the machine learning model. These metrics are meticulously analyzed to gauge the model's performance, particularly when evaluated on the testing dataset. Figure 4.15 provides an overview of the confusion matrix.

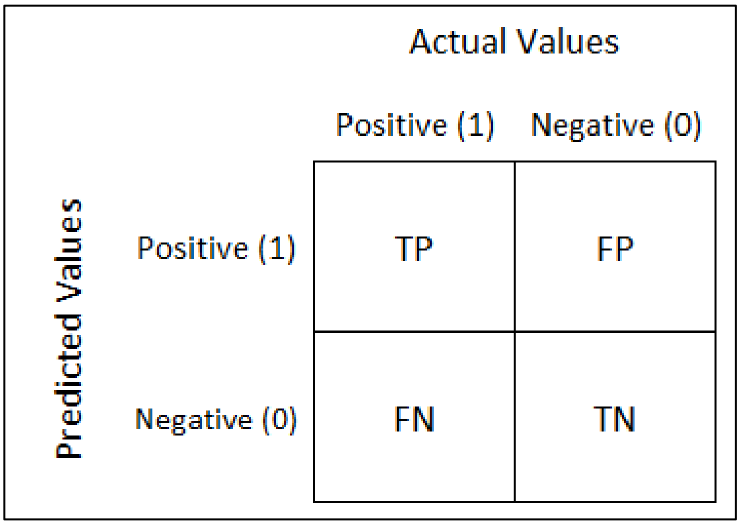


Figure 4.15 Overview of Confusion Matrix

(Source: Narkhede, 2018)

4.1.5 Model Deployment

Upon completion of the classifier model development, it was applied to the real-world data previously gathered from X. Employing NB for sentiment analysis, the model assigned labels to each text, categorizing them as positive, negative, or neutral. These categories are represented as '0' for negative, '1' for neutral, and '2' for positive sentiment by Scikit-learn.

Following sentiment analysis and model performance evaluation, the collected data underwent visualization using Plotly, a Python library known for its interactive graphics. These visualizations enable users to actively interact with the data by hovering over specific elements, triggering pop-ups displaying additional information about the charts. To facilitate this process, the data was initially organized into a Pandas DataFrame, and various interactive charts were generated using suitable coding techniques. Serving as valuable tools for exploring and comprehending the analysis results obtained from real-world data, these interactive visualizations offer users a deeper understanding of patterns and trends. Chapter 5 discusses the detailed explanations of the visualizations featured in the system.

The classification process yields two types of outputs. Firstly, a comprehensive report incorporating a confusion matrix and performance metrics such as accuracy, precision, recall, and F1-Score provides insights into the classifier's performance and its efficacy in predicting sentiment accurately. Secondly, data visualization using the Plotly library presents the analyzed data in an interactive and user-friendly format. By exploring these visualizations, users can gain a deeper understanding of patterns and trends in the sentiment analysis results derived from real-world data.

4.2 Front End Development

Front-end development is where the data is visualized and displayed for users on a website. The phase is done using programming languages such as HTML and CSS, as well as the Flask web framework. Flask is utilized for this project due to its user-friendly nature in web application environment and the various libraries available in building an attractive interface. Not only that, it also supports Python's data visualizations, simplifying the process of creating figures for displaying the sentiment analysis results. A more detailed explanation about the system interface is discussed in the following subsections.

4.2.1 Landing Page

Figure 4.16 shows the "Landing" page, the first page user sees when they enter the system. Here, the user can choose to use the system as a registered user or as a guest. "Sign In" button will take the users to the "Login" page, while the "Continue as Guest" hyperlink will direct them to the "Overview" page, where the overall performance of the courier services is displayed. This page also includes a remark which states that the data used for this system is dated from 1st April 2023 until 31st March 2024.

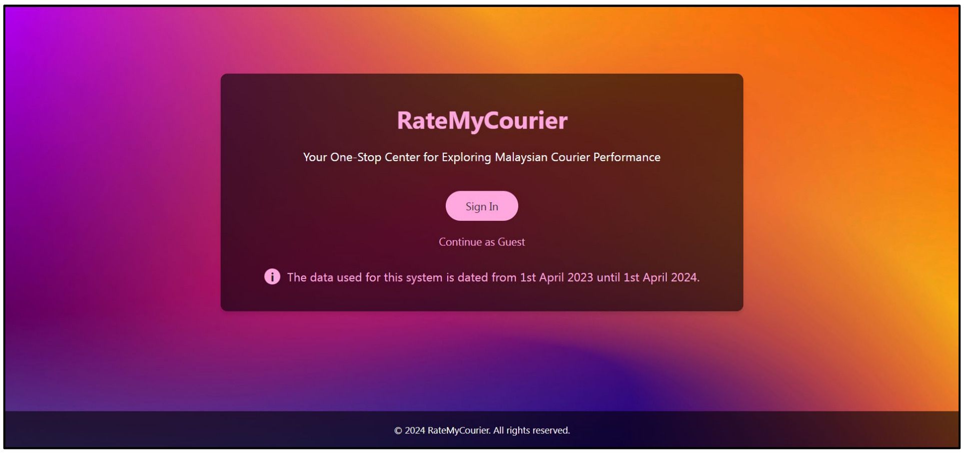


Figure 4.16 Landing Page

4.2.2 Login Page

Figure 4.17 presents the “Login” page, where users can sign into an existing account to use the system. They will be prompted to enter their email and password, then click the “Submit” button to log in. If the user wishes to create a new account, they can click the “Register” hyperlink which will take them to the “Register” page.

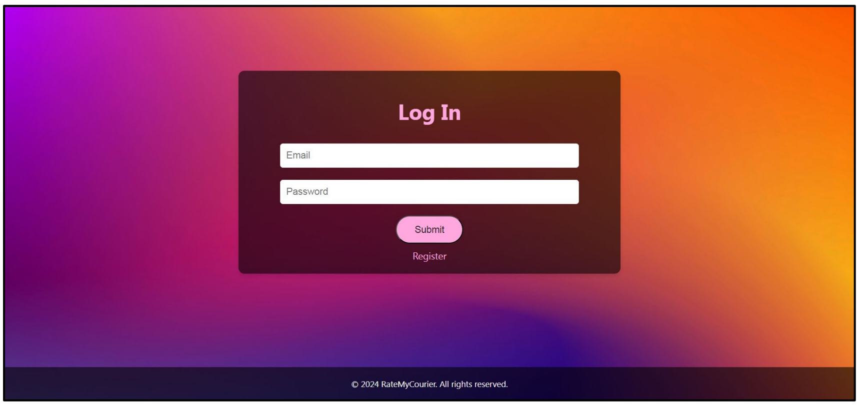


Figure 4.17 Login Page

4.2.3 Register Page

Figure 4.18 shows the “Register” page, where users can create a new account to enter the system. Details such as name, email and password will be filled by the users. Then, they can click the “Submit” button to register the account.

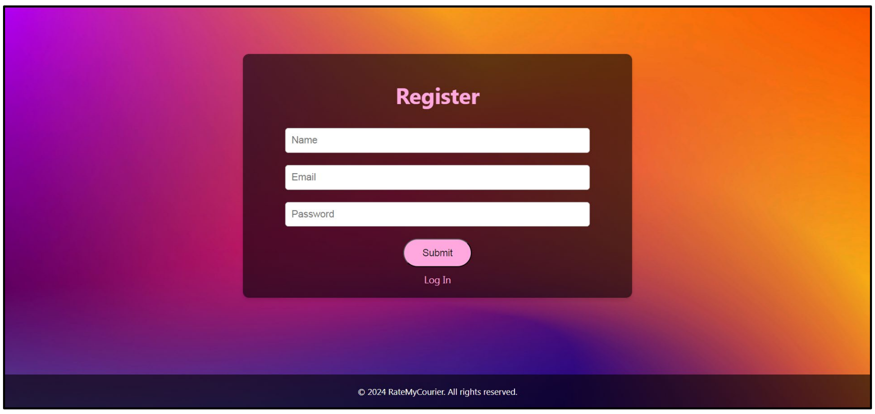


Figure 4.18 Register Page

4.2.4 Overview Page

Figure 4.19 presents the “Overview” page, which summarizes the overall performance metrics for all courier services engaged in this project. The courier services involved are J&T Express, SPX Xpress, and DHL Express. The sidebar menu facilitates easier navigation for the users, showing the main menus available in the system.

'Overall Sentiment Mentions' features a stacked bar chart that illustrates the total number of positive, negative, and neutral mentions for each courier. This visual representation provides a clear and immediate comparison of sentiment across all service providers. It helps in highlighting areas of strength and potential improvement.

Two critical aspects are evaluated for each courier: speed and reliability. 'Speed Sentiment Mentions' showcases a pie chart detailing the total mentions related to the speed of each courier. This chart enables users to understand how customers perceive the promptness and efficiency of each service. Similarly, 'Reliability Sentiment Mentions' presents a pie chart that depicts the total mentions concerning the reliability of each courier. This visualization allows users to assess how consistently and dependably each service provider meets customer expectations.

Together, these visual tools offer a comprehensive view of customer sentiment, enabling users to thoroughly assess the performance of each courier service. By analyzing both speed and reliability, stakeholders can gain valuable insights into the operational effectiveness of each courier, informing strategic decisions and identifying opportunities for enhancement.

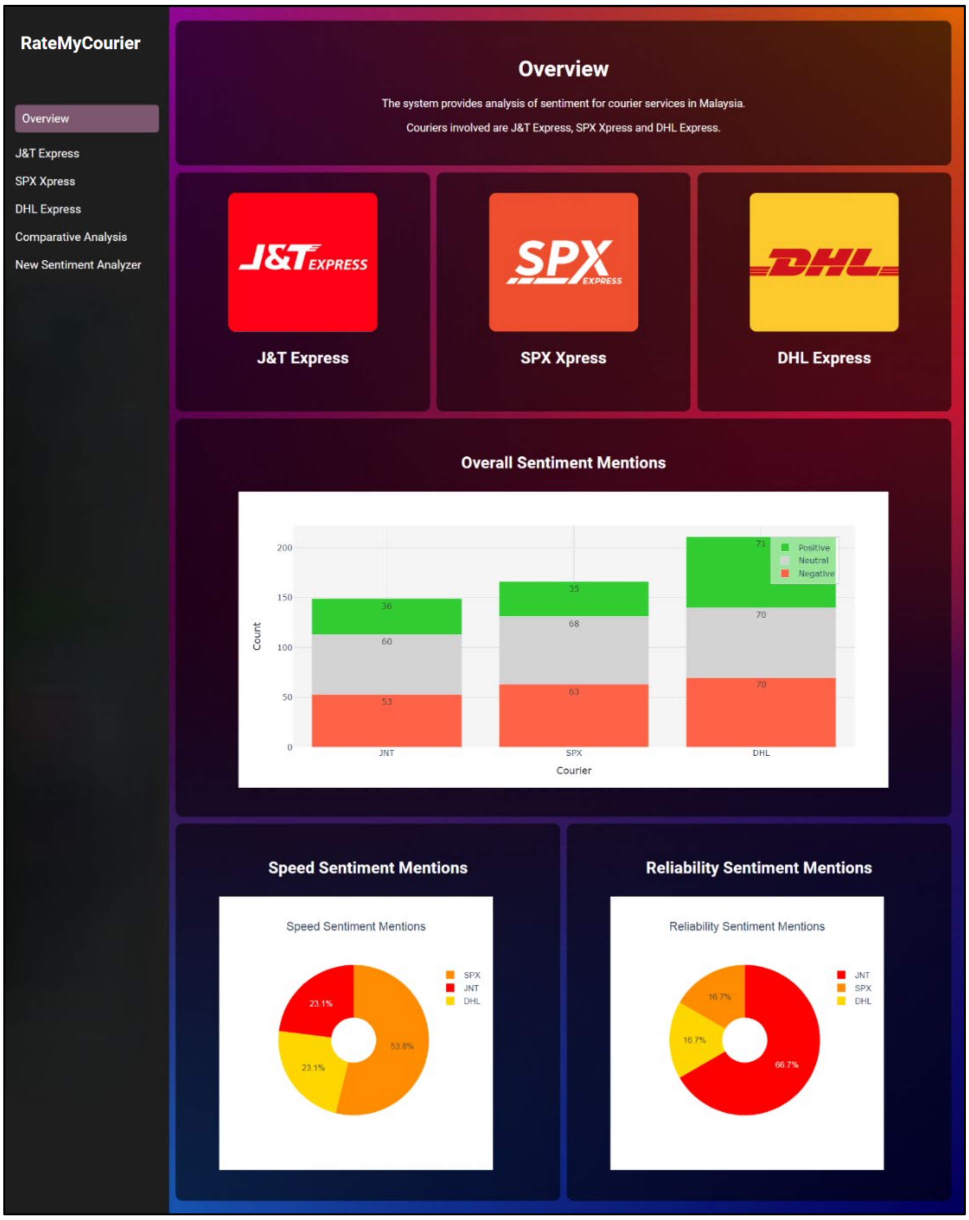


Figure 4.19 Overview Page

4.2.5 J&T Express Page

Figure 4.20 illustrates the "J&T Express" page, providing an in-depth analysis of sentiments regarding J&T Express. Below the company logo, their website link and contact number are displayed. 'Sentiment Count' features a bar chart that quantifies the total number of negative, neutral, and positive sentiments. Complementing this, the 'Sentiment Percentage' section represents the same data as a pie chart, highlighting the percentage of each sentiment category. 'Sentiment Over Time' provides a timeline chart, showing the distribution of negative, neutral, and positive sentiment mentions from 1st April 2023 to 31st March 2024. This analysis offers insights into sentiment trends over the past year.

Language-specific sentiment analyses are presented in the 'English Sentiment' and 'Malay Sentiment' sections, each depicted as pie charts illustrating the distribution of sentiments in the respective languages. Word clouds are also included, with the 'Negative Sentiment Word Cloud' and 'Positive Sentiment Word Cloud' visually representing the most frequently mentioned terms associated with negative and positive sentiments, respectively.

In addition, 'Speed Aspect' features a pie chart indicating the percentage of negative, neutral, and positive sentiments related to speed. Accompanying this is 'Speed Keywords', which displays a horizontal bar chart of the most frequently mentioned keywords related to speed. Similarly, 'Reliability Aspect' presents a pie chart showing the sentiment distribution concerning reliability, alongside 'Reliability Keywords', which provides a horizontal bar chart of keyword frequency related to reliability.

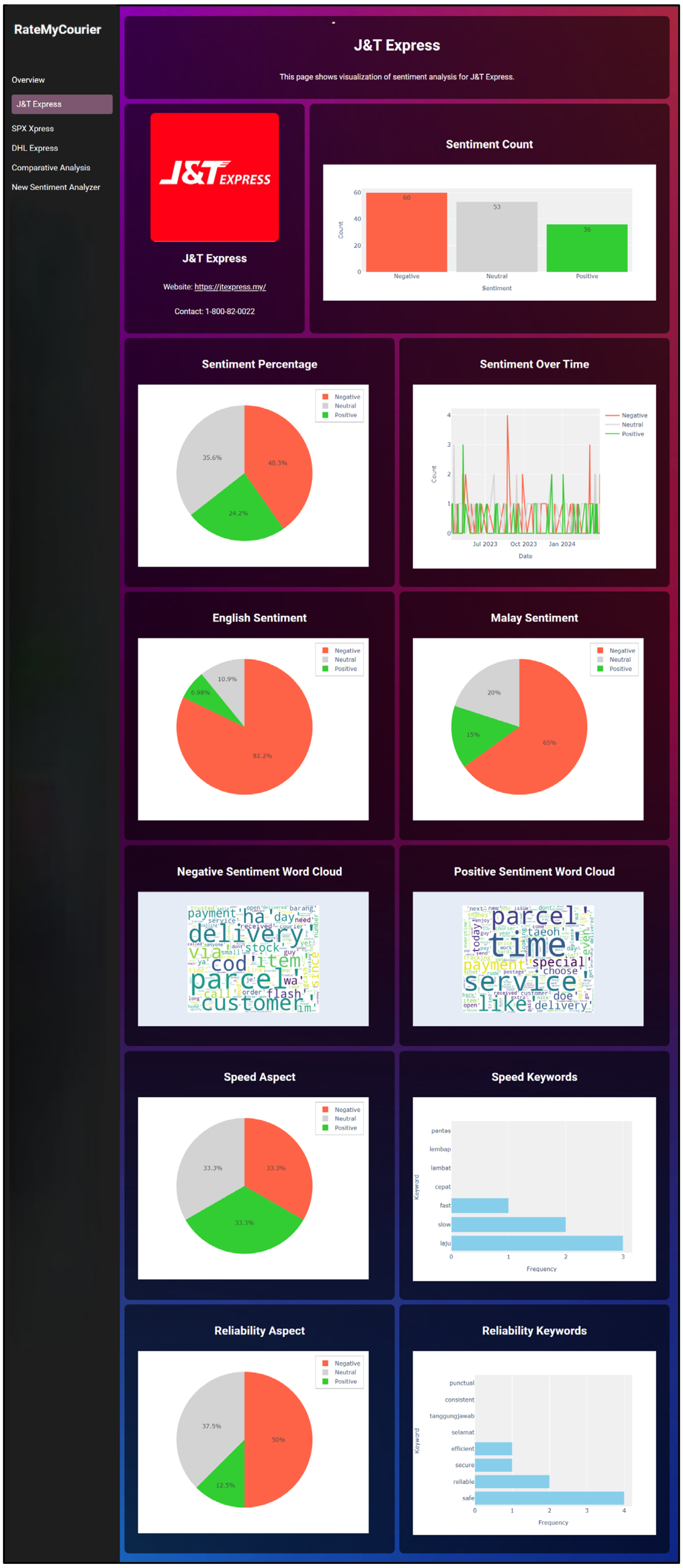


Figure 4.20 J&T Express Page

4.2.6 SPX Xpress Page

Figure 4.21 illustrates the “SPX Xpress” page, offering an in-depth analysis of sentiments regarding SPX Xpress. Below the company logo, their website link and contact number are displayed for easy access.

'Sentiment Count' features a bar chart that quantifies the total number of negative, neutral, and positive sentiments. Complementing this, 'Sentiment Percentage' presents the same data in a pie chart format, highlighting the percentage of each sentiment category. 'Sentiment Over Time' provides a timeline chart showing the distribution of negative, neutral, and positive sentiment mentions from 1st April 2023 to 31st March 2024, offering insights into sentiment trends over the past year.

Language-specific sentiment analyses are included in 'English Sentiment' and 'Malay Sentiment' sections, with each depicted as pie charts illustrating the distribution of sentiments in the respective languages. Word clouds are also presented, with the 'Negative Sentiment Word Cloud' and 'Positive Sentiment Word Cloud' visually representing the most frequently mentioned terms associated with negative and positive sentiments, respectively.

Additionally, 'Speed Aspect' features a pie chart indicating the percentage of negative, neutral, and positive sentiments related to speed. Accompanying this is 'Speed Keywords,' which displays a horizontal bar chart of the most frequently mentioned keywords related to speed. Similarly, the 'Reliability Aspect' presents a pie chart showing the sentiment distribution concerning reliability, alongside 'Reliability Keywords,' which provides a horizontal bar chart of keyword frequency related to reliability.

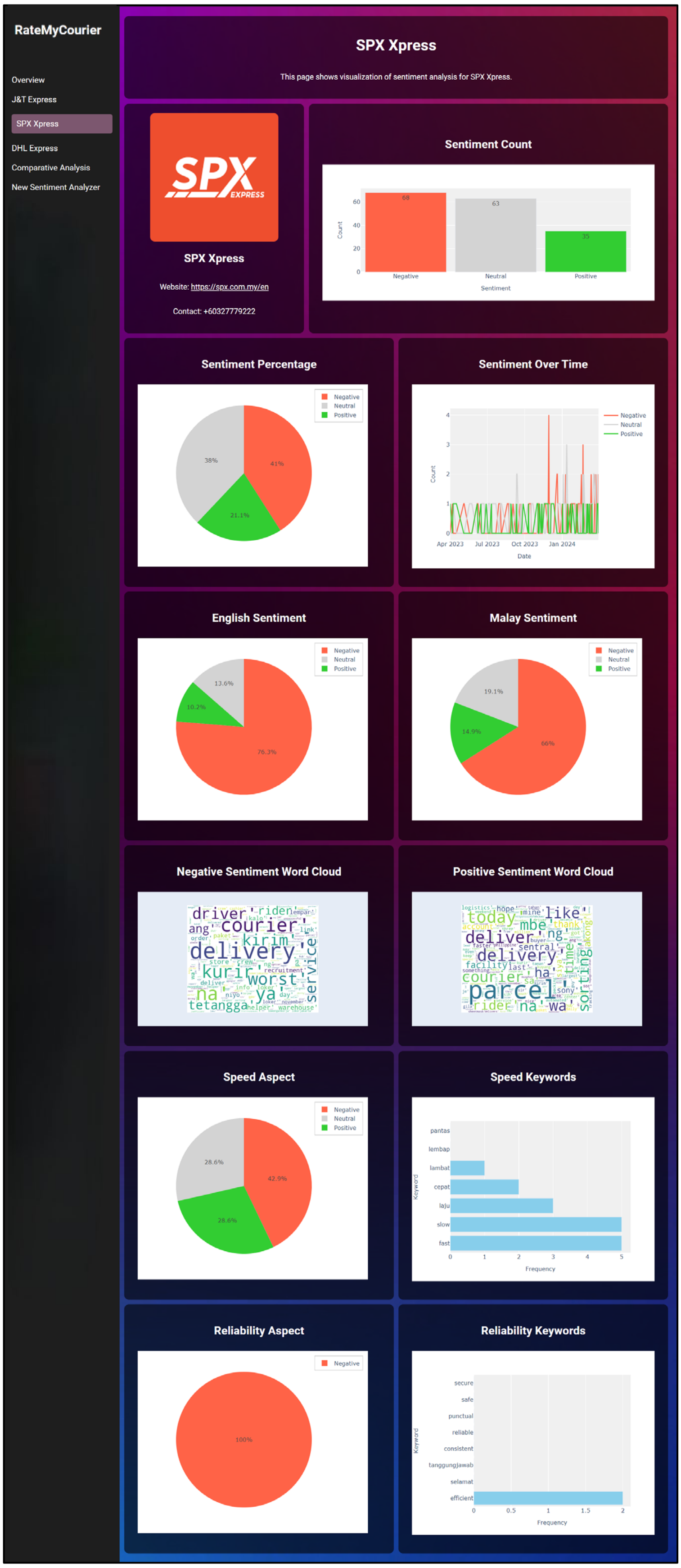


Figure 4.21 SPX Page

4.2.7 DHL Express Page

Figure 4.22 illustrates the “DHL Express” page, offering an in-depth analysis of sentiments regarding DHL Express. Below the company logo, their website link and contact number are displayed for easy access.

'Sentiment Count' features a bar chart that quantifies the total number of negative, neutral, and positive sentiments. Complementing this, 'Sentiment Percentage' presents the same data in a pie chart format, highlighting the percentage of each sentiment category. 'Sentiment Over Time' provides a timeline chart showing the distribution of negative, neutral, and positive sentiment mentions from 1st April 2023 to 31st March 2024, offering insights into sentiment trends over the past year.

Language-specific sentiment analyses are included in 'English Sentiment' and 'Malay Sentiment' sections, with each depicted as pie charts illustrating the distribution of sentiments in the respective languages. Word clouds are also presented, with the 'Negative Sentiment Word Cloud' and 'Positive Sentiment Word Cloud' visually representing the most frequently mentioned terms associated with negative and positive sentiments, respectively.

Additionally, 'Speed Aspect' features a pie chart indicating the percentage of negative, neutral, and positive sentiments related to speed. Accompanying this is 'Speed Keywords,' which displays a horizontal bar chart of the most frequently mentioned keywords related to speed. Similarly, 'Reliability Aspect' presents a pie chart showing the sentiment distribution concerning reliability, alongside 'Reliability Keywords,' which provides a horizontal bar chart of keyword frequency related to reliability.

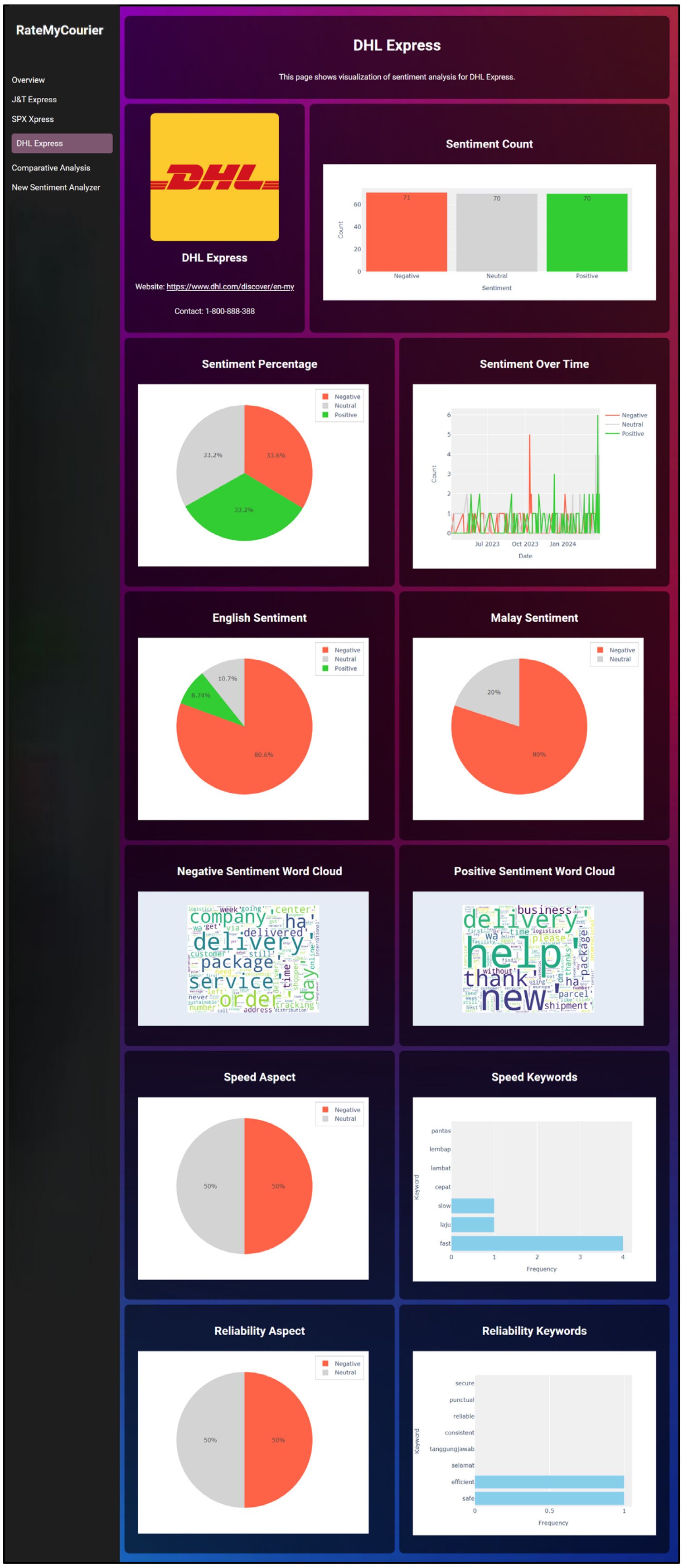


Figure 4.22 DHL Express Page

4.2.8 Comparative Analysis Page

Figure 4.23 presents the “Comparative Analysis” page, which offers an in-depth comparison of the sentiment analysis results for the courier services involved in this study. This page is designed to provide a clear and comprehensive view of how each courier service is perceived across various sentiment metrics. To facilitate easier comparison, the sentiment percentages for each courier are displayed in side-by-side pie charts. This visual representation allows users to quickly grasp the proportion of positive, neutral, and negative sentiments for each service provider, enabling straightforward comparative analysis.

'Sentiment Mentions' features a grouped bar chart that quantifies the count of negative, neutral, and positive sentiments for each courier. This chart provides a direct comparison of the volume of sentiment mentions. It helps to highlight the areas where each courier excels or needs improvement.

'Speed Aspect' showcases a stacked bar graph that illustrates the distribution of speed-related sentiments—negative, neutral, and positive—for each courier. This allows for a detailed examination of how each service performs in terms of delivery speed, a critical factor in customer satisfaction. Similarly, 'Reliability Aspect' presents a stacked bar graph depicting the distribution of reliability-related sentiments for each courier. This visualization offers insights into the perceived dependability of each service, an essential component of overall performance.

Figure 4.23 Comparative Analysis Page

4.2.9 New Sentiment Analyzer Page

Figure 4.24 presents the “New Sentiment Analyzer” page, a user-friendly interface designed to facilitate sentiment analysis. This page allows users to either enter text directly or upload a CSV file for analysis. Upon submission, the sentiment analyzer processes the input and displays the analysis results on the same page.

This feature is particularly useful for businesses and individuals seeking to analyze customer feedback, social media comments, or any text-based data. By providing both direct text input and CSV upload options, the “New Sentiment Analyzer” page offers flexibility and convenience to accommodate various user needs.

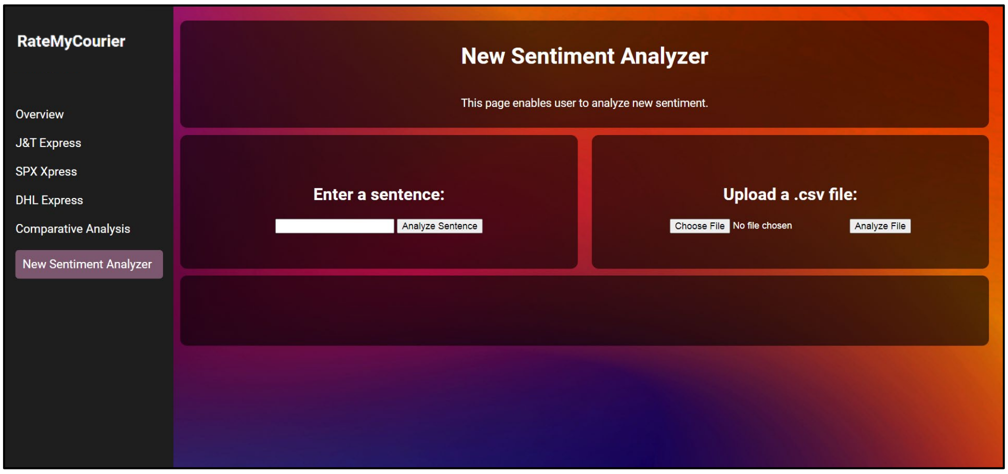


Figure 4.24 New Sentiment Analyzer Page

4.2.10 Summary Report Page

Figure 4.25 shows the “Summary Report” page. This page is exclusively accessible to authenticated users, providing personalized experience. It showcases a curated selection of courier logos, meticulously arranged to enhance visual clarity and user experience.

Central to the functionality of the Summary Report page are the 'Generate Report' buttons, positioned below each courier logo. These buttons empower users with the ability to effortlessly compile detailed reports tailored to their specific requirements. It's noteworthy that access to this functionality is meticulously governed to uphold user privacy and data integrity. Users are only permitted to generate reports for couriers affiliated with their respective organizations, thereby fostering a sense of ownership and confidentiality.

In essence, the Summary Report page represents a harmonious fusion of advanced analytics and user-centric design principles. By providing a platform for data-driven decision-making, it plays a pivotal role in empowering our users to make informed strategic choices, driving efficiency and maximizing operational effectiveness.

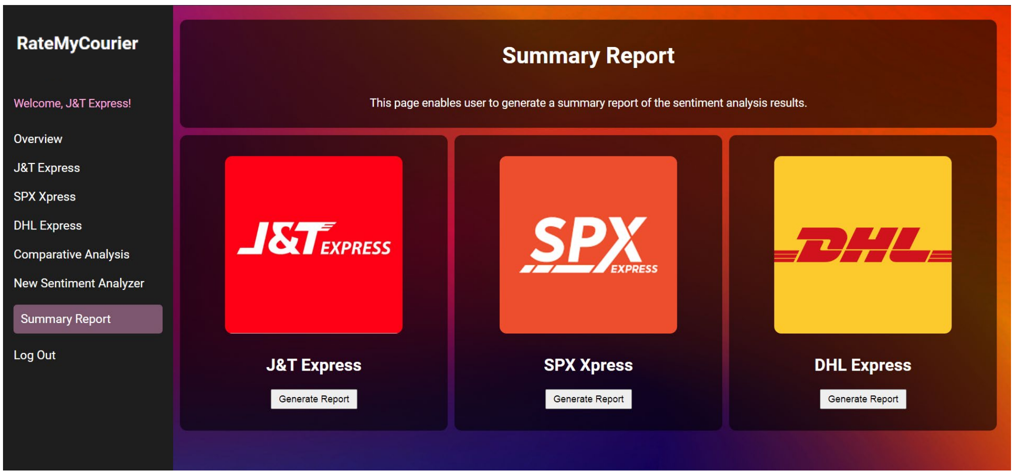


Figure 4.25 Summary Report Page

4.3 Summary

This chapter details the system development process, divided into front-end and back-end stages. The back-end development covers data collection, data pre-processing, and Naive Bayes classifier model development. The front-end phase involves creating the user interface and implementing system functionalities. The chapter provides a thorough breakdown of each step in the procedure.

CHAPTER 5

RESULTS AND DISCUSSIONS

This chapter discusses the accuracy of the Naïve Bayes classifier model, presenting analysis results based on real-world data. It also discusses the testing phase, where functionality testing of the system is carried out to ensure proper operation.

5.1 Accuracy Testing

The accuracy of the Naive Bayes classifier model for the English language is evaluated using a basic Python program. The resulting accuracy metrics for both the training and testing datasets are presented in Figure 5.1.

```
Train Accuracy:  0.864403987148859

Confusion Matrix:
 [[3384  456  226]
 [ 255 3562  222]
 [ 172  315 3547]]

Classification Report:
               precision    recall  f1-score   support

           0       0.87      0.75      0.81      5058
           1       0.76      0.86      0.80      5058
           2       0.84      0.84      0.84      5058

    accuracy                           0.82     15174
   macro avg       0.82      0.82      0.82     15174
weighted avg       0.82      0.82      0.82     15174

Test Accuracy:  0.8177804138658231

Confusion Matrix:
 [[3817  817  424]
 [ 328 4335  395]
 [ 225  576 4257]]

Classification Report:
               precision    recall  f1-score   support

           0       0.87      0.75      0.81      5058
           1       0.76      0.86      0.80      5058
           2       0.84      0.84      0.84      5058

    accuracy                           0.82     15174
   macro avg       0.82      0.82      0.82     15174
weighted avg       0.82      0.82      0.82     15174
```

Figure 5.1 Accuracy Test for English Model

The English classifier model achieved an accuracy of 86.44% on the training data, correctly predicting sentiment labels about 86% of the time during training. This means the model identified approximately 8 to 9 out of 10 sentiment labels correctly as "positive," "neutral," or "negative." The training data's confusion matrix reveals that the model made accurate predictions for 3,384 "negative" labels, 3,562 "neutral" labels, and 3,547 "positive" labels.

For the testing data, the model's accuracy was 81.78%. This indicates that it correctly predicted sentiment labels about 82% of the time on unseen data. The testing data's confusion matrix shows the model accurately predicted 3,817 "negative" labels, 4,335 "neutral" labels, and 4,257 "positive" labels, although there were some misclassifications.

Precision scores for each sentiment class are provided as well. On the training data, the precision scores are 0.87 for "negative", 0.76 for "neutral", and 0.84 for "positive", reflecting the proportion of accurate predictions for each class. On the testing data, the precision scores are 0.87 for "negative", 0.76 for "neutral", and 0.84 for "positive".

The accuracy of the Malay classifier model is presented in Figure 5.2. It showcases the performance metrics and accuracy testing of training and testing data specifically for the Malay language.

```
Train Accuracy:  0.9110371075166508

Confusion Matrix:
 [[632   8  54]
 [ 27 670  20]
 [ 70   8 613]]

Classification Report:
              precision    recall  f1-score   support

           0       0.76      0.88      0.81       876
           1       0.97      0.85      0.91       876
           2       0.83      0.79      0.81       876

    accuracy                           0.84      2628
   macro avg       0.85      0.84      0.84      2628
weighted avg       0.85      0.84      0.84      2628

Test Accuracy:  0.8405631659056316

Confusion Matrix:
 [[769  12  95]
 [ 78 747  51]
 [170  13 693]]

Classification Report:
              precision    recall  f1-score   support

           0       0.76      0.88      0.81       876
           1       0.97      0.85      0.91       876
           2       0.83      0.79      0.81       876

    accuracy                           0.84      2628
   macro avg       0.85      0.84      0.84      2628
weighted avg       0.85      0.84      0.84      2628
```

Figure 5.2 Accuracy Test for Malay Model

The Malay classifier model showed an accuracy of 91.10% on the training data. It correctly predicts sentiment labels approximately 91% of the time during the training phase. The confusion matrix for the training data indicates the model accurately predicted 632 "negative" labels, 670 "neutral" labels, and 613 "positive" labels.

For the testing data, the model achieved an accuracy of 84.06%. This means it correctly predicted sentiment labels around 84% of the time on new, unseen data. The confusion matrix for the testing data reveals the model made correct predictions for 769 "negative" labels, 747 "neutral" labels, and 693 "positive" labels, though some errors occurred.

The precision scores for each sentiment class provide further details on the model's performance. On the training data, the precision scores were 0.76 for "negative," 0.97 for "neutral," and 0.83 for "positive," indicating the proportion of accurate predictions for each class. On the testing data, the precision scores were 0.76 for "negative," 0.97 for "neutral," and 0.83 for "positive."

The accuracy of the Courier classifier model is presented in Figure 5.3. It showcases the performance metrics and accuracy testing of training and testing data specifically from the courier domain.

```
Train Accuracy:  0.9826086956521739

Confusion Matrix:
 [[77  0  0]
 [ 2 75  1]
 [ 0  1 74]]

Classification Report:
               precision    recall  f1-score   support

           0       0.98      0.84      0.90        95
           1       0.88      0.97      0.92        99
           2       0.96      0.99      0.97        94

    accuracy                           0.93       288
   macro avg       0.94      0.93      0.93       288
weighted avg       0.94      0.93      0.93       288

Test Accuracy:  0.9340277777777778

Confusion Matrix:
 [[80 12  3]
 [ 2 96  1]
 [ 0  1 93]]

Classification Report:
               precision    recall  f1-score   support

           0       0.98      0.84      0.90        95
           1       0.88      0.97      0.92        99
           2       0.96      0.99      0.97        94

    accuracy                           0.93       288
   macro avg       0.94      0.93      0.93       288
weighted avg       0.94      0.93      0.93       288
```

Figure 5.3 Accuracy Test for Courier Model

The Courier classifier model achieved an accuracy of 98.26% on the training data. It correctly predicts sentiment labels nearly 98% of the time during training. The confusion matrix for the training data shows that the model accurately predicted 77 "negative" labels, 75 "neutral" labels, and 74 "positive" labels.

For the testing data, the model attained an accuracy of 93.40%. It correctly predicts sentiment labels approximately 93% of the time on new, unseen data.

The confusion matrix for the testing data reveals the model made correct predictions for 80 "negative" labels, 96 "neutral" labels, and 93 "positive" labels, with some misclassifications.

Precision scores for each sentiment class further illustrate the model's performance. On the training data, the precision scores were 0.98 for "negative," 0.88 for "neutral," and 0.96 for "positive," indicating the proportion of correct predictions for each class. On the testing data, the precision scores were 0.98 for "negative," 0.88 for "neutral," and 0.96 for "positive."

5.2 Real-World Data Analysis Result

This project showcases real-world data analysis of courier services through various visualization methods, such as pie charts, bar charts, line charts, and word clouds. These visualizations offer a comprehensive and insightful view of the data, helping users better understand the analyzed information. The next subsection provides a detailed explanation of the data visualization techniques used in the system.

5.2.1 Data Visualization of Courier Services Overview

The data collection process gathered 4,365 tweets from the platform X between 1st April 2023 and 1st April 2024, focusing on three major courier services: J&T Express, SPX Xpress, and DHL Express. The dataset revealed that J&T Express was mentioned in 1,401 tweets, SPX Xpress in 1,514 tweets, and DHL Express in 1,450 tweets. Figure 5.4 illustrates this data distribution in a bar chart, showing SPX Xpress with the highest number of mentions, followed by DHL Express and J&T Express.

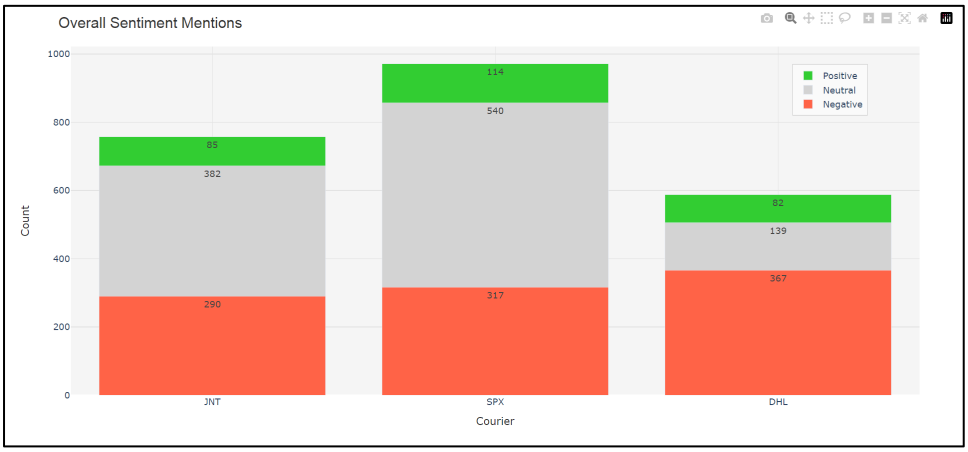


Figure 5.4 Overall Sentiment Mentions

For a more in-depth analysis of the reviews, the specific sentiments related to various aspects of courier services are also analyzed. The identified aspects in the collected data are speed and reliability. The keywords used for speed aspect includes 'lembap', 'slow', 'fast', 'lambat', 'laju', and 'cepat', while the keywords used for reliability aspect are 'efficient', 'secure', 'reliable', 'safe', and 'selamat'. These keywords are used to filter the general sentiments and determine total mentions for each aspect. Figure 5.5 and 5.6 show the pie charts of distribution of speed and reliability mentions for all couriers.

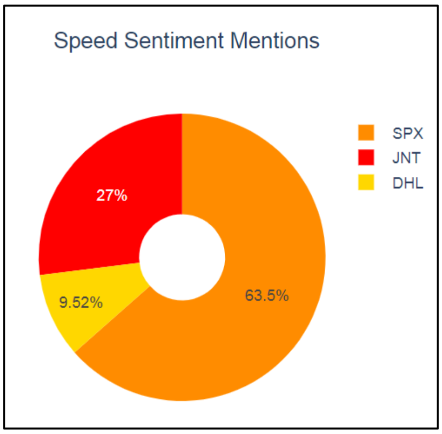


Figure 5.5 Speed Sentiment Mentions

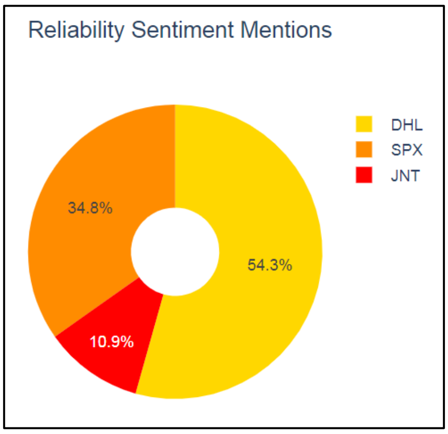


Figure 5.6 Reliability Sentiment Mentions

5.2.2 Data Visualization of J&T Express

Figure 5.7 showcases a bar chart displaying the total counts across each sentiment category for J&T Express. Complementing this, Figure 5.8 features a pie chart detailing the percentage distribution of sentiments by category. Negative sentiment is visually represented in red, neutral sentiment in grey, and positive sentiment in green. According to the bar chart, the largest count belongs to negative sentiment, totaling 382 mentions, followed by 290 mentions for neutral sentiment and 85 mentions for positive sentiment. The pie chart provides a percentage breakdown, illustrating that negative sentiment constitutes 50.5% of the total, while neutral sentiment accounts for 38.3%, and positive sentiment represents 11.2%.

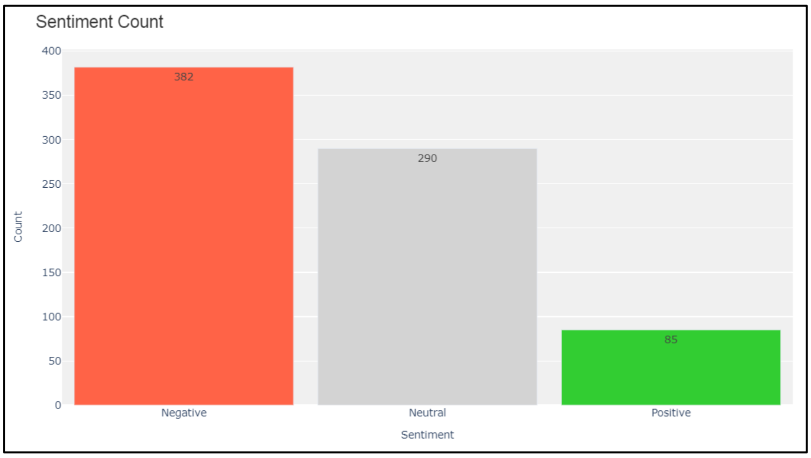


Figure 5.7 J&T Express Sentiment Count

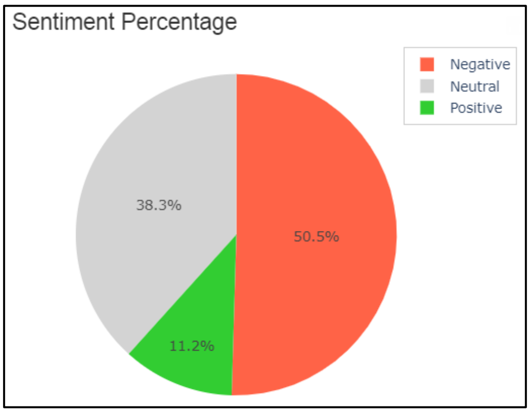


Figure 5.8 J&T Express Sentiment Percentage

Figure 5.9 illustrates a line chart depicting the evolution of each sentiment category from April 2023 to April 2024. Notably, peaks in negative mentions were observed between August 2023 and December 2023, with another notable spike in April 2024. Conversely, positive mentions reached their peak in January 2024 and maintained a relatively steady level throughout the year. The data reveals a gradual increase in X mentions starting from September 2023. This visualization offers valuable insights into the dynamics of sentiment trends and the overall distribution of sentiments across the analyzed timeframe.

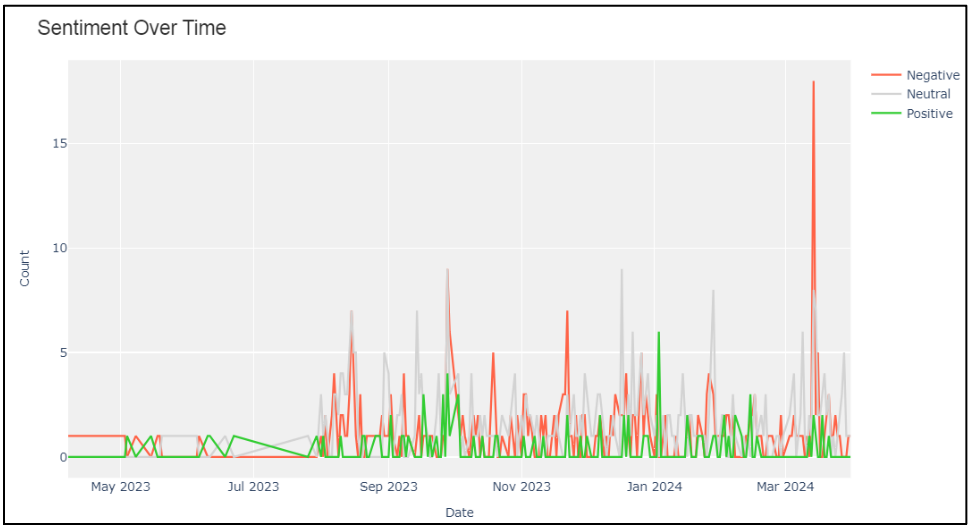


Figure 5.9 J&T Express Sentiment Over Time

Figure 5.10 displays a pie chart illustrating the sentiment distribution in English, with negative sentiment comprising 75.4%, neutral sentiment at 13.1%, and positive sentiment at 11.4%. Meanwhile, Figure 5.11 depicts a similar pie chart for Malay sentiment, where negative sentiment accounts for 63.9%, neutral sentiment for 21.5%, and positive sentiment for 14.6%. These charts provide a clear visual representation of how sentiments are categorized across both languages.

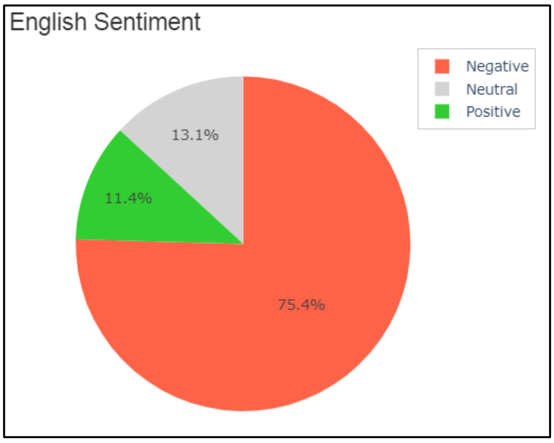


Figure 5.10 J&T Express English Sentiment

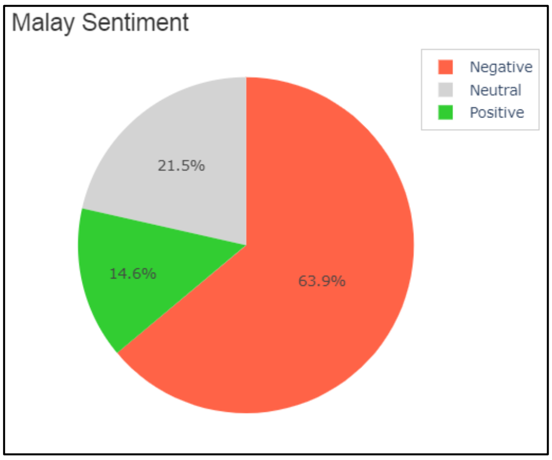


Figure 5.11 J&T Express Malay Sentiment

Figures 5.12 and 5.13 feature word clouds showcasing positive and negative mentions. These visual representations aim to highlight common phrases used by users when discussing J&T Express. Notably, terms such as 'barang' and 'parcel' prominently appear in the negative and positive word clouds respectively, indicating frequent discussions about parcel-related issues among users. Additionally, the word clouds reveal other significant phrases associated with J&T Express, such as 'delivery' and 'kirim'. They serve as a graphical summary of key themes and topics frequently raised in discussions about J&T Express on platform X.

Figure 5.12 J&T Express Negative Word Cloud

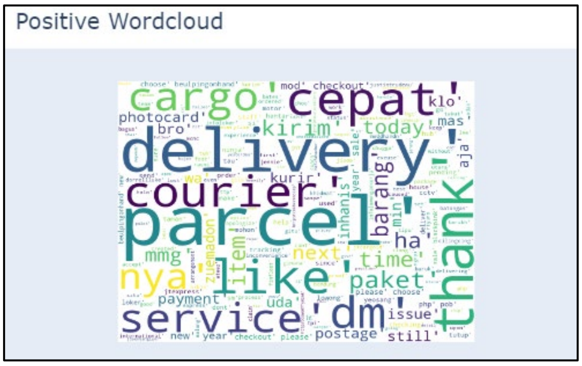


Figure 5.13 J&T Express Positive Word Cloud

Figure 5.14 presents a pie chart depicting sentiments related to speed, with negative sentiments accounting for 41.2%, neutral sentiments for 32.4%, and positive sentiments for 26.5%. In Figure 5.15, the frequency of keywords associated with the speed aspect is illustrated in a horizontal bar chart. The keyword 'cepat' emerges as the most frequently mentioned, underscoring its prominence in discussions, while 'slow' is notably the least mentioned keyword in the dataset.

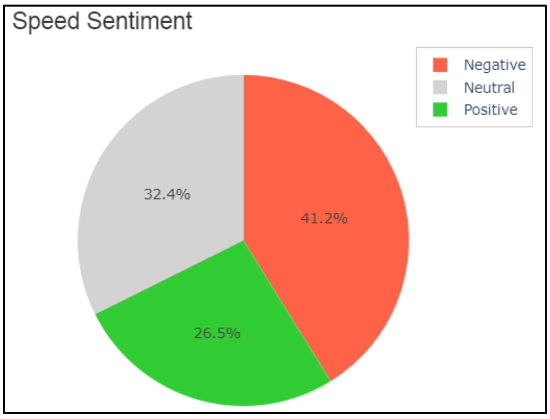


Figure 5.14 J&T Express Speed Sentiment

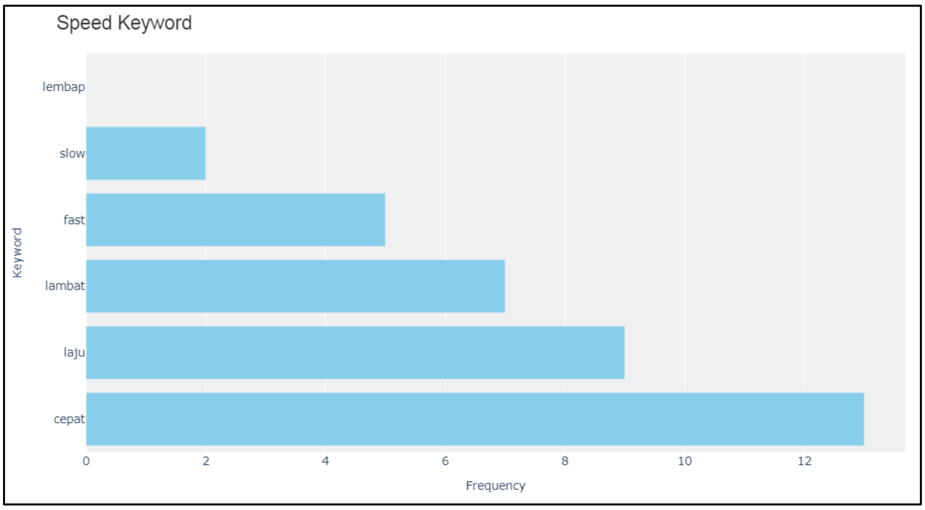


Figure 5.15 J&T Express Speed Keyword

Figure 5.16 showcases a pie chart that portrays sentiments regarding reliability, revealing that negative sentiments constitute 80% of the discussions, while neutral sentiments make up the remaining 20%. Interestingly, there were no positive mentions related to the reliability aspect for J&T Express. In Figure 5.17, the frequency of keywords linked to reliability is depicted in a horizontal bar chart format. The keyword 'selamat' stands out as the most frequently mentioned, highlighting its significant presence in discussions, whereas 'safe' appears as the least frequently referenced keyword in the dataset.

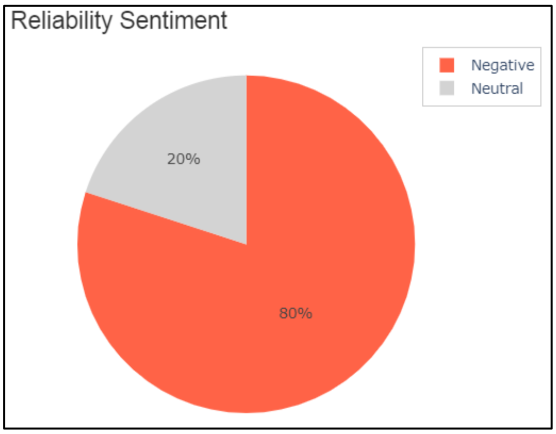


Figure 5.16 J&T Express Reliability Sentiment

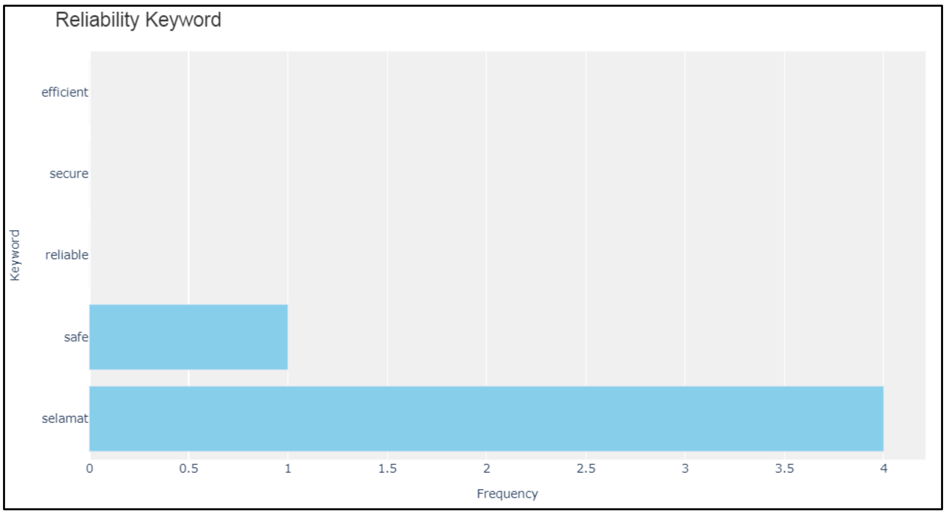


Figure 5.17 J&T Express Reliability Keyword

5.2.3 Data Visualization of SPX Xpress

Figure 5.18 showcases a bar chart displaying the total counts across each sentiment category for SPX Xpress. Complementing this, Figure 5.19 features a pie chart detailing the percentage distribution of sentiments by category. Negative sentiment is visually represented in red, neutral sentiment in grey, and positive sentiment in green. According to the bar chart, the largest count belongs to negative sentiment, totaling 540 mentions, followed by 317 mentions for neutral sentiment and 114 mentions for positive sentiment. The pie chart provides a percentage breakdown, illustrating that negative sentiment constitutes 55.6% of the total, while neutral sentiment accounts for 32.6%, and positive sentiment represents 11.7%.

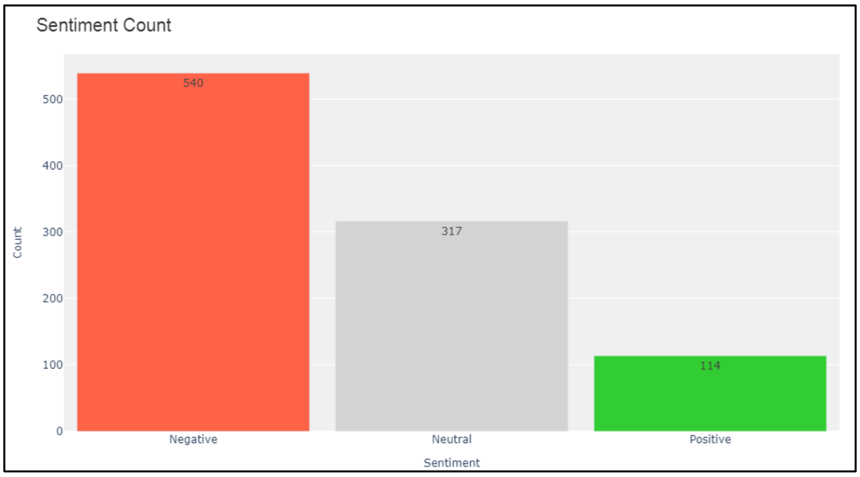


Figure 5.18 SPX Xpress Sentiment Count

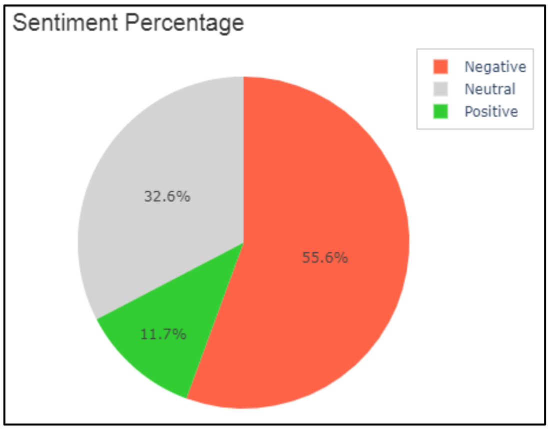


Figure 5.19 SPX Xpress Sentiment Percentage

Figure 5.20 presents a line chart that tracks the trends of each sentiment category from April 2023 to April 2024. A significant observation is the spike in negative mentions, particularly between August 2023 and September 2023, with a sharp increase from January 2024 continuing into April 2024. In contrast, positive mentions maintain a steady rate throughout the year. Interestingly, all sentiment categories—positive, neutral, and negative—show a gradual rise beginning around January 2024. This chart highlights notable fluctuations and trends in customer sentiment over the specified period.

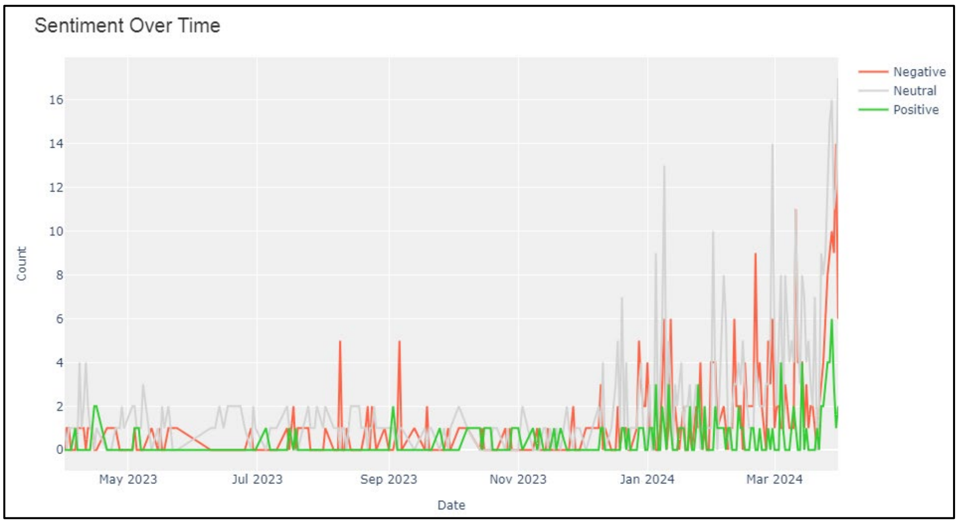


Figure 5.20 SPX Xpress Sentiment Over Time

Figure 5.21 presents a pie chart showing the distribution of sentiments in English, with negative sentiment making up 69.6%, neutral sentiment at 15.7%, and positive sentiment at 14.7%. Similarly, Figure 5.22 depicts a pie chart for Malay sentiments, where negative sentiment accounts for 68.3%, neutral sentiment for 19.6%, and positive sentiment for 12.1%. These charts offer a clear visual comparison of sentiment categorization in both languages.

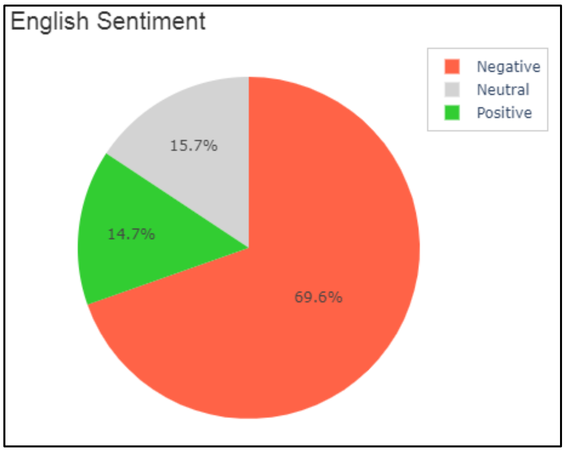


Figure 5.21 SPX Xpress English Sentiment

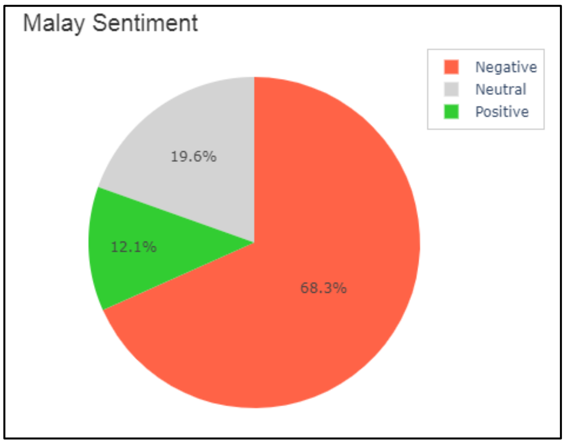


Figure 5.22 SPX Xpress Malay Sentiment

Figures 5.23 and 5.24 showcase word clouds that represent positive and negative mentions of SPX Xpress. These visuals emphasize the common phrases used by users in their discussions. The term ‘paket’ is notably prevalent in both word clouds, suggesting frequent conversation about parcel-related issues. Other prominent terms include ‘delivery’ and ‘kurir’ which highlight important themes and topics often mentioned in relation to SPX Xpress. These word clouds provide a concise graphical summary of user sentiment and discussion points.

Figure 5.23 SPX Xpress Negative Word Cloud

Figure 5.24 SPX Xpress Positive Word Cloud

Figure 5.25 features a pie chart illustrating sentiments related to speed: negative sentiments make up 52.5%, neutral sentiments comprise 36.3%, and positive sentiments represent 11.3%. Meanwhile, Figure 5.26 uses a horizontal bar chart to show the frequency of keywords associated with speed. The keyword ‘cepat’ appears most frequently, highlighting its importance in discussions, whereas ‘lembap’ is the least mentioned keyword in the dataset.

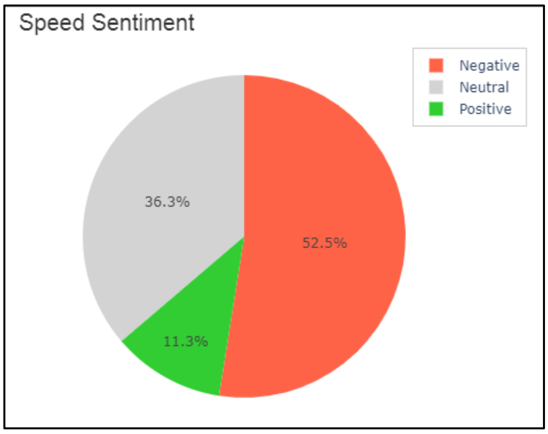


Figure 5.25 SPX Xpress Speed Sentiment

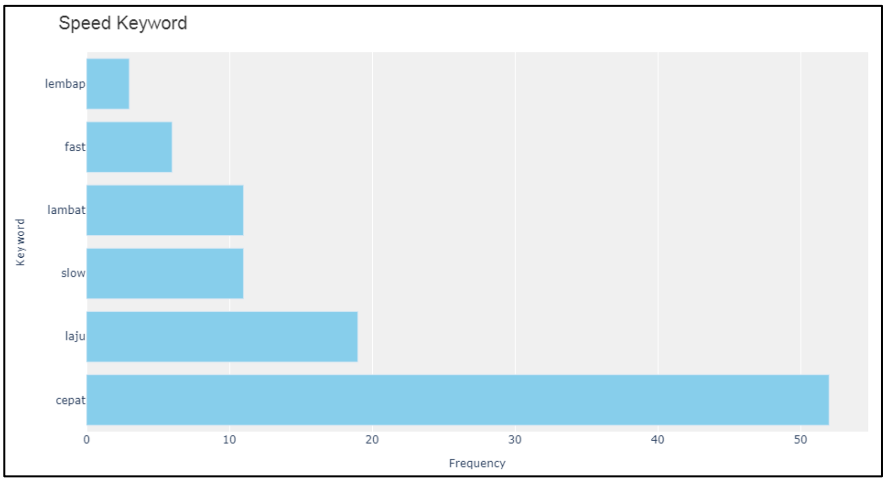


Figure 5.26 SPX Xpress Speed Keyword

Figure 5.27 displays a pie chart illustrating sentiments related to reliability. The chart reveals that negative sentiments dominate the discussion at 56.3%, followed by neutral sentiments at 31.3%, and positive sentiments at 12.5%. In contrast, Figure 5.28 presents a horizontal bar chart that shows the frequency of keywords associated with reliability. The keyword 'selamat' is the most frequently mentioned, underscoring its prominence in user discussions. On the other hand, 'efficient' and 'secure' are the least frequently referenced keywords in the dataset.

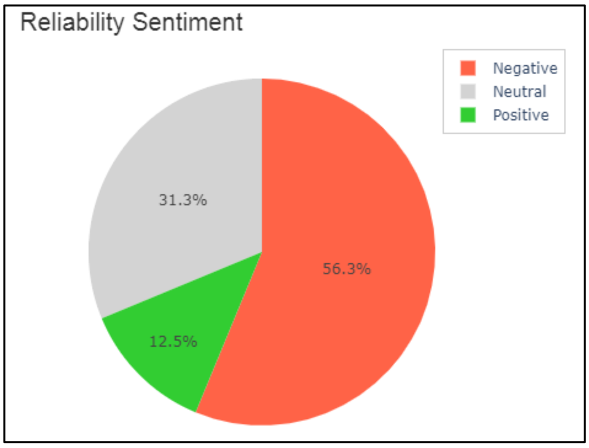


Figure 5.27 SPX Xpress Reliability Sentiment

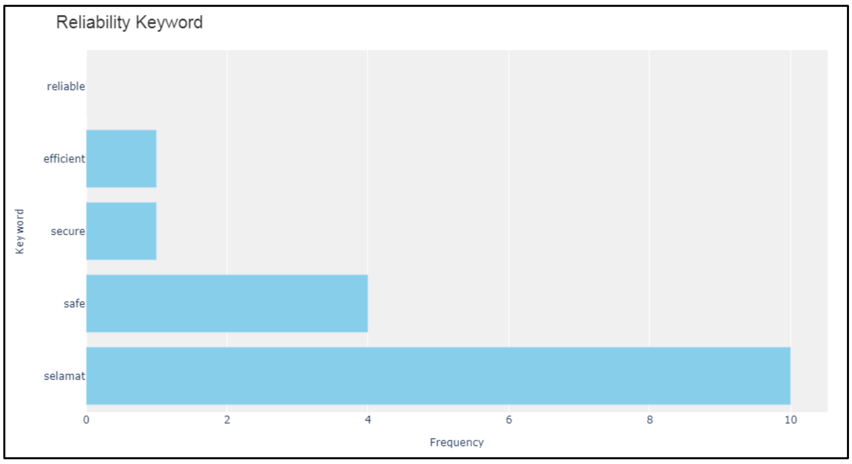


Figure 5.28 SPX Xpress Reliability Keyword

5.2.4 Data Visualization of DHL Express

Figure 5.29 showcases a bar chart displaying the total counts across each sentiment category for DHL Express. Complementing this, Figure 5.30 features a pie chart detailing the percentage distribution of sentiments by category. According to the bar chart, the largest count belongs to negative sentiment, totaling 367 mentions, followed by 139 mentions for neutral sentiment and 82 mentions for positive sentiment. The pie chart provides a percentage breakdown, illustrating that negative sentiment constitutes 62.4% of the total, while neutral sentiment accounts for 23.6%, and positive sentiment represents 13.9%.

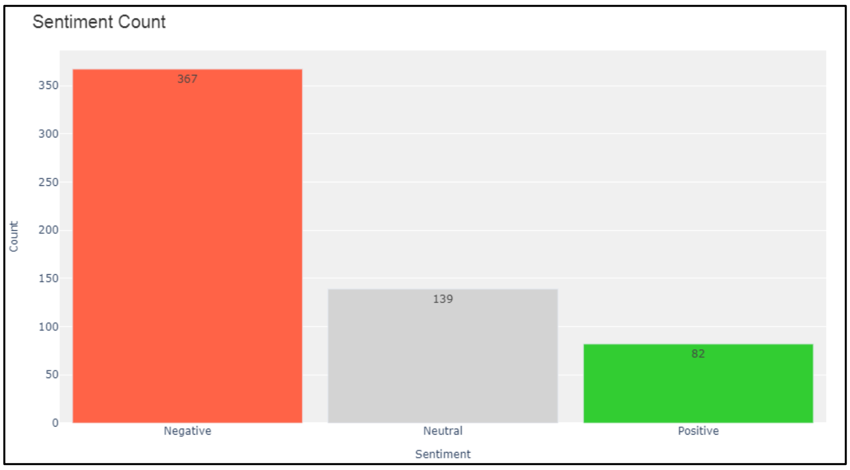


Figure 5.29 DHL Express Sentiment Count

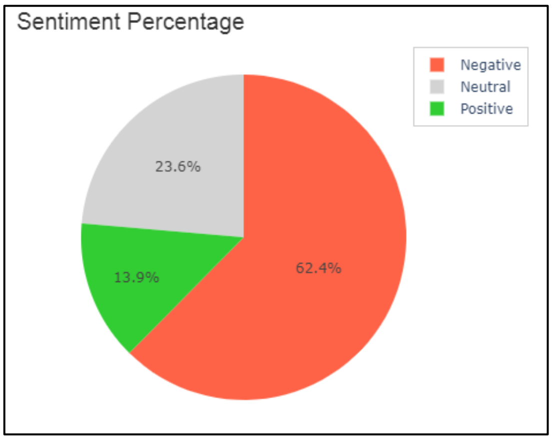


Figure 5.30 DHL Express Sentiment Percentage

Figure 5.31 presents a line chart tracking the trends in sentiment categories from April 2023 to April 2024. Negative mentions surge significantly as April 2024 approaches, while positive mentions peak in January 2024 and remain relatively stable thereafter. The chart indicates a gradual rise in the volume of mentions beginning in January 2024. This visualization provides crucial insights into sentiment dynamics and the distribution of sentiments over the analyzed period.

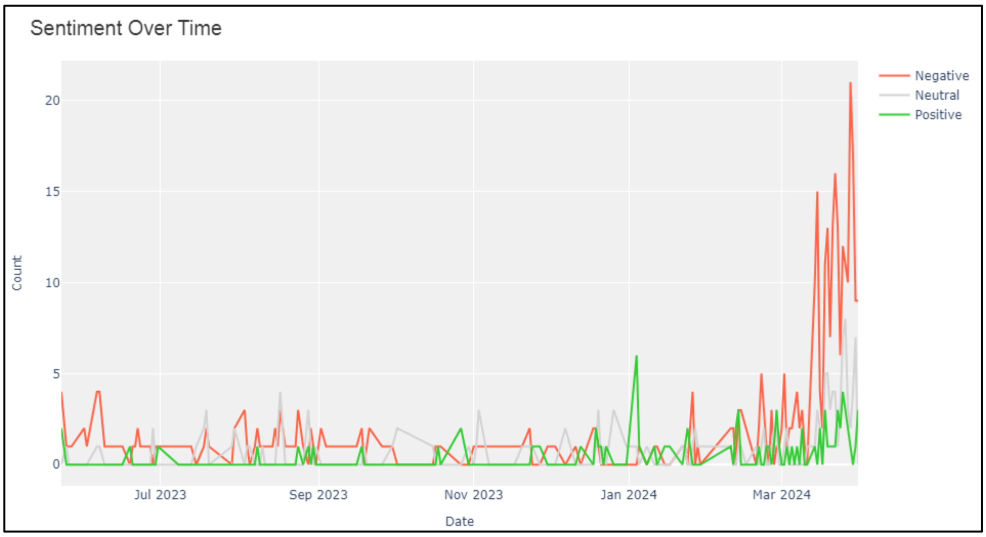


Figure 5.31 DHL Express Sentiment Over Time

Figure 5.32 shows a pie chart representing sentiment distribution in English: 76.4% negative, 19.9% neutral, and 3.65% positive. In contrast, Figure 5.33 presents a similar pie chart for Malay sentiment, with 62.1% negative, 28.8%

neutral, and 9.09% positive. These charts clearly visualize the sentiment categorization in both languages.

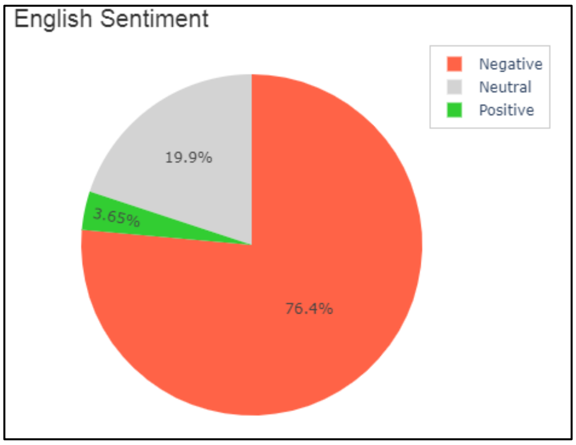


Figure 5.32 DHL Express English Sentiment

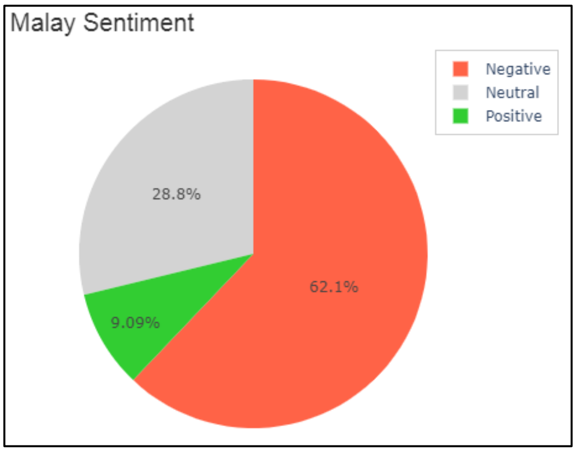


Figure 5.33 DHL Express Malay Sentiment

Figures 5.34 and 5.35 display word clouds that illustrate positive and negative mentions related to DHL Express. These visuals are designed to emphasize common phrases used by users when discussing the company. Notably, the term ‘shipment’ prominently appears in the negative word cloud, while ‘kastam’ is featured in the positive word cloud, indicating frequent discussions about shipment-related and customs-related issues, respectively. Additionally, terms such as 'delivery' and ‘hantar’ are also highlighted, providing a graphical summary of the key themes and topics often raised in conversations about DHL Express.

Figure 5.34 DHL Express Negative Word Cloud

Figure 5.35 DHL Express Positive Word Cloud

Figure 5.36 displays a pie chart illustrating sentiments regarding speed: 50% negative, 33.3% neutral, and 16.7% positive. In Figure 5.37, a horizontal bar chart shows the frequency of keywords related to speed. The keyword ‘fast’ is the most frequently mentioned, highlighting its significance in discussions, while 'slow' is the least mentioned keyword in the dataset.

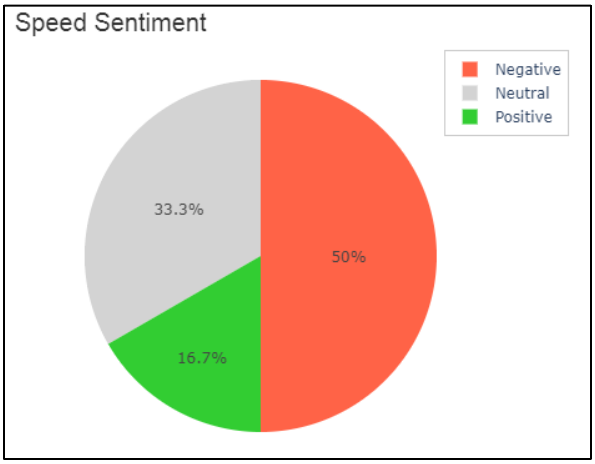


Figure 5.36 DHL Express Speed Sentiment

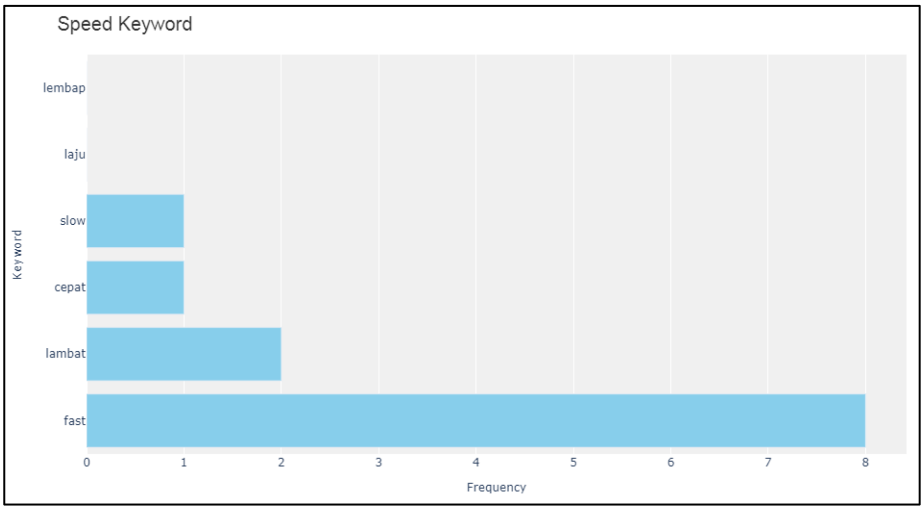


Figure 5.37 DHL Express Speed Keyword

Figure 5.38 presents a pie chart illustrating sentiments concerning reliability, with negative sentiments comprising 52%, neutral sentiments 28%, and positive sentiments 20% of the discussions. Figure 5.39 features a horizontal bar chart detailing the frequency of keywords associated with reliability. The keyword ‘safe’ is prominently featured as the most frequently mentioned term, indicating its significant presence in discussions, while ‘reliable’ is the least frequently referenced keyword in the dataset.

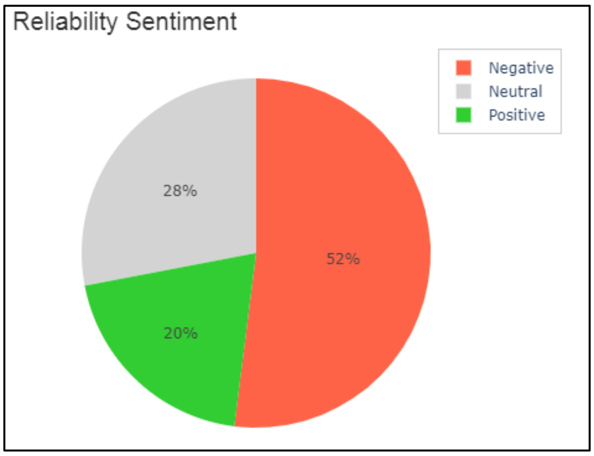


Figure 5.38 DHL Express Reliability Sentiment

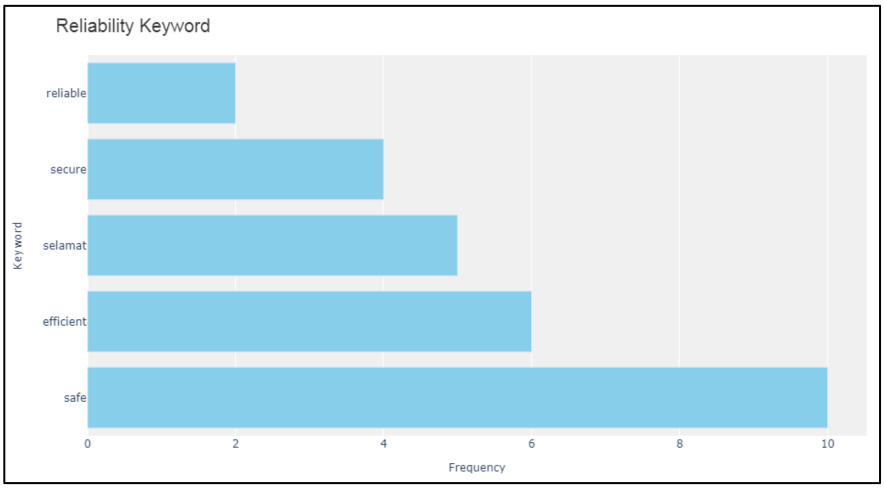


Figure 5.39 DHL Express Reliability Keyword

5.2.5 Data Visualization of Comparative Analysis

In this comparative analysis, the performance of J&T Express, SPX Xpress, and DHL Express is evaluated. Figure 5.40 displays a grouped bar chart that shows the total mentions for each courier. SPX Xpress leads in neutral and positive mentions, whereas DHL Express has the fewest. Conversely, DHL Express has the highest number of negative mentions, with J&T Express having the least. Notably, DHL Express is the only courier with more negative mentions than both neutral and positive ones. However, it is crucial to account for the imbalanced data quantity for each courier when interpreting these results.

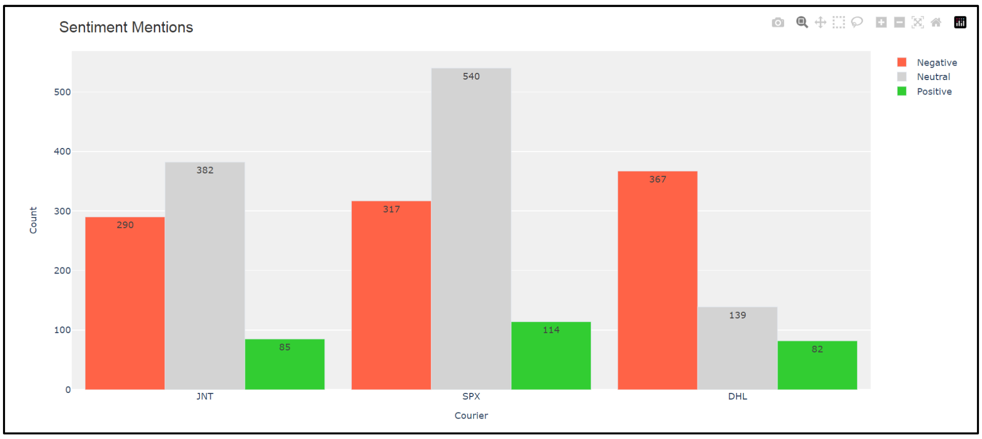


Figure 5.40 Comparison of Sentiment Mentions

The performance of each courier is also evaluated based on two key aspects: speed and reliability. Figure 5.41 presents the comparative analysis of speed sentiment for J&T Express, SPX Xpress, and DHL Express, indicating that SPX Xpress has the highest number of mentions, followed by J&T Express and DHL Express. Figure 5.42 depicts the comparison of reliability sentiment among the three couriers, showing that DHL Express leads in the number of mentions, followed by SPX Xpress, and J&T Express. Notably, J&T Express has no positive mentions in the context of reliability.

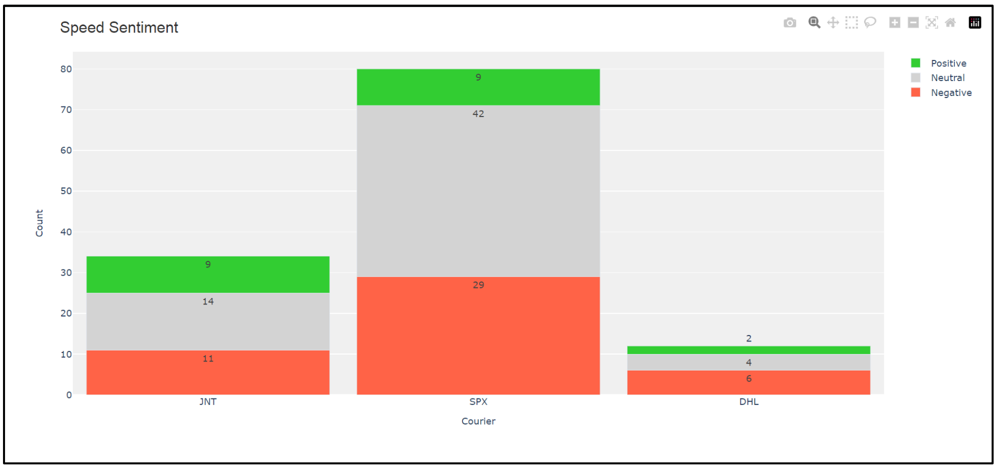


Figure 5.41 Comparison of Speed Sentiment

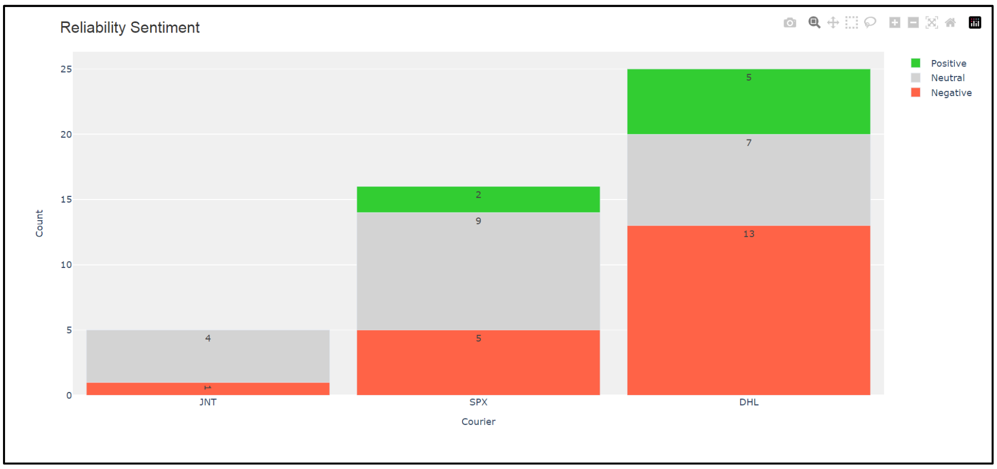


Figure 5.42 Comparison of Reliability Sentiment

5.3 Functionality Testing

Functionality testing is a type of software testing that ensures the system meets its functional requirements. It involves testing the system's functionalities using test cases to confirm that it performs as expected and that all functions operate correctly without errors or issues. The results of the functions and test cases that were evaluated are shown in Table 5.1.

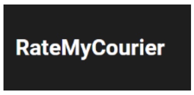
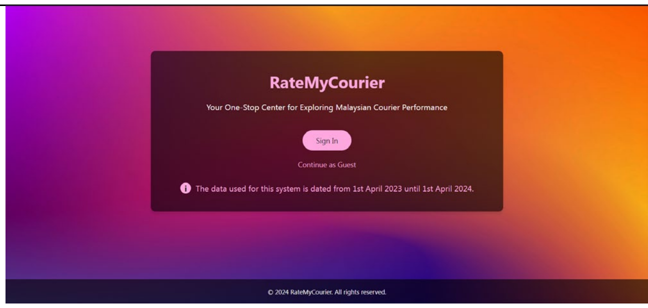
Table 5.1 Functionality Testing

Test Case	Expected Result	Actual result
View Landing Page	System displays landing page.	Success
View Login Page	System displays login page.	Success
View Register Page	System displays register page.	Success
View Overview Page	System displays visualizations of overall courier services performance.	Success
View J&T Express Page	System displays J&T Express sentiment analysis results in several visualization forms.	Success
View SPX Xpress Page	System displays SPX Xpress sentiment analysis results in several visualization forms.	Success
View DHL Express Page	System displays DHL Express sentiment analysis results in several visualization forms	Success
View Comparative Analysis Page	System displays comparison of sentiment analysis between different courier services.	Success
View New Sentiment Analyzer Page	System displays page to input new sentiment to be analyzed.	Success
View Summary Report Page	System displays the summary report page.	Success

5.3.1 View Landing Page

Table 5.2 discusses the test results of “View Landing Page”.

Table 5.2 Functionality Testing for Landing Page

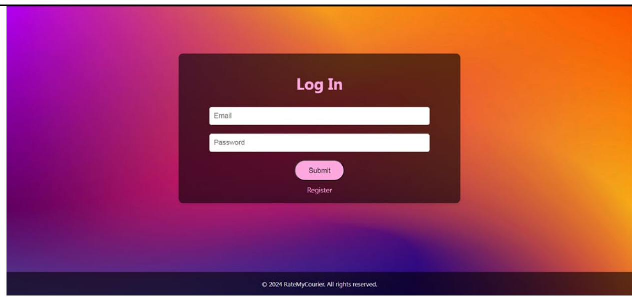
Test Objective	To verify the system displays the landing page correctly.
Potential Test Input	A. User clicks the “RateMyCourier” text  B. User clicks the “Log Out” menu on the sidebar menu 
Expected Test Output	Landing page loads successfully. “Sign In” button and “Continue as Guest” hyperlink are functional.
Test Procedures	A. User clicks the “X Sentiment Analysis” text B. User clicks the “Log Out” menu on the sidebar menu
Actual Test Results	

5.3.2 View Login Page

Table 5.3 discusses the test results of “View Login Page”.

Table 5.3 Functionality Testing for Login Page

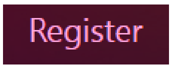
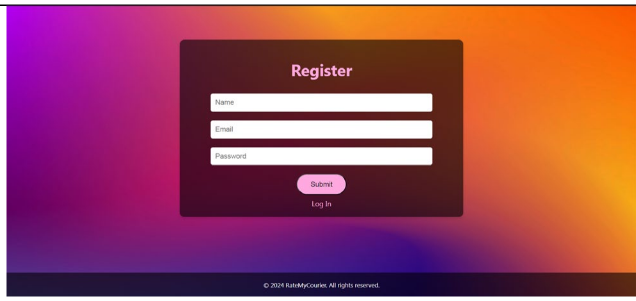
Test Objective	To verify the system displays the login page correctly.
Potential Test Input	A. User clicks the “Sign In” button on the landing page. 

	B. User clicks the “Login” hyperlink on the register page. 
Expected Test Output	Login page loads successfully. “Submit” button and “Register” hyperlink are functional.
Test Procedures	A. User clicks the “Sign In” button on the landing page. B. User clicks the “Login” hyperlink on the register page.
Actual Test Results	

5.3.3 View Register Page

Table 5.4 discusses the test results of “View Register Page”.

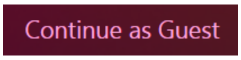
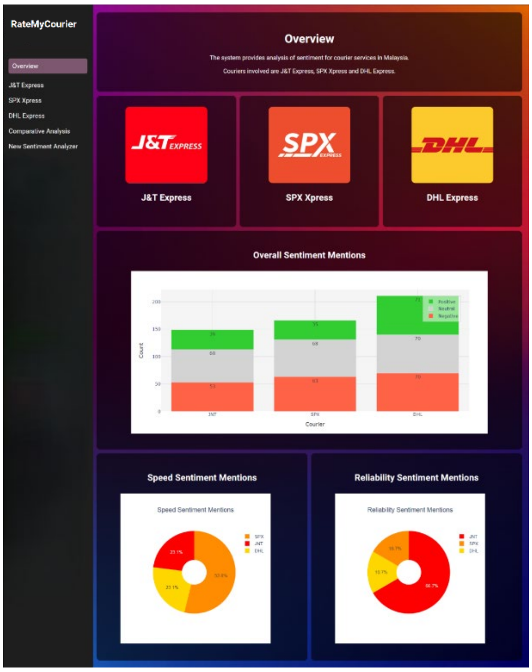
Table 5.4 Functionality Testing for Register Page

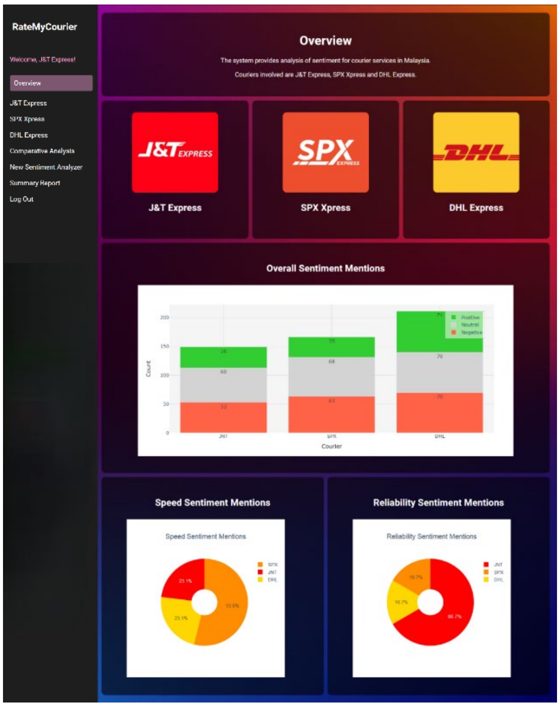
Test Objective	To verify the system displays the register page correctly.
Potential Test Input	User clicks the “Register” hyperlink. 
Expected Test Output	Register page loads successfully. “Submit” button and “Login” hyperlink are functional.
Test Procedures	User clicks the “Register” hyperlink on the login page.
Actual Test Results	

5.3.4 View Overview Page

Table 5.5 discusses the test results of “View Overview Page”.

Table 5.5 Functionality Testing for Overview Page

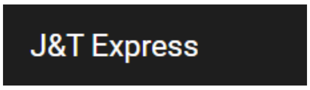
Test Objective	To verify the system displays the overview page correctly.
Potential Test Input	A. User clicks “Continue as Guest” hyperlink on landing page.  B. User clicks “Submit” button on login page after entering their registered email and password. 
Expected Test Output	Overview page loads successfully. Sidebar menus are functional. All visualizations on the page are displayed correctly and interactive.
Test Procedures	A. User clicks “Continue as Guest” hyperlink on landing page. B. User clicks “Submit” button on login page after entering their registered email and password.
Actual Test Results	A. When viewed as guest, the number of menus on the sidebar menu are limited to only six menus. 

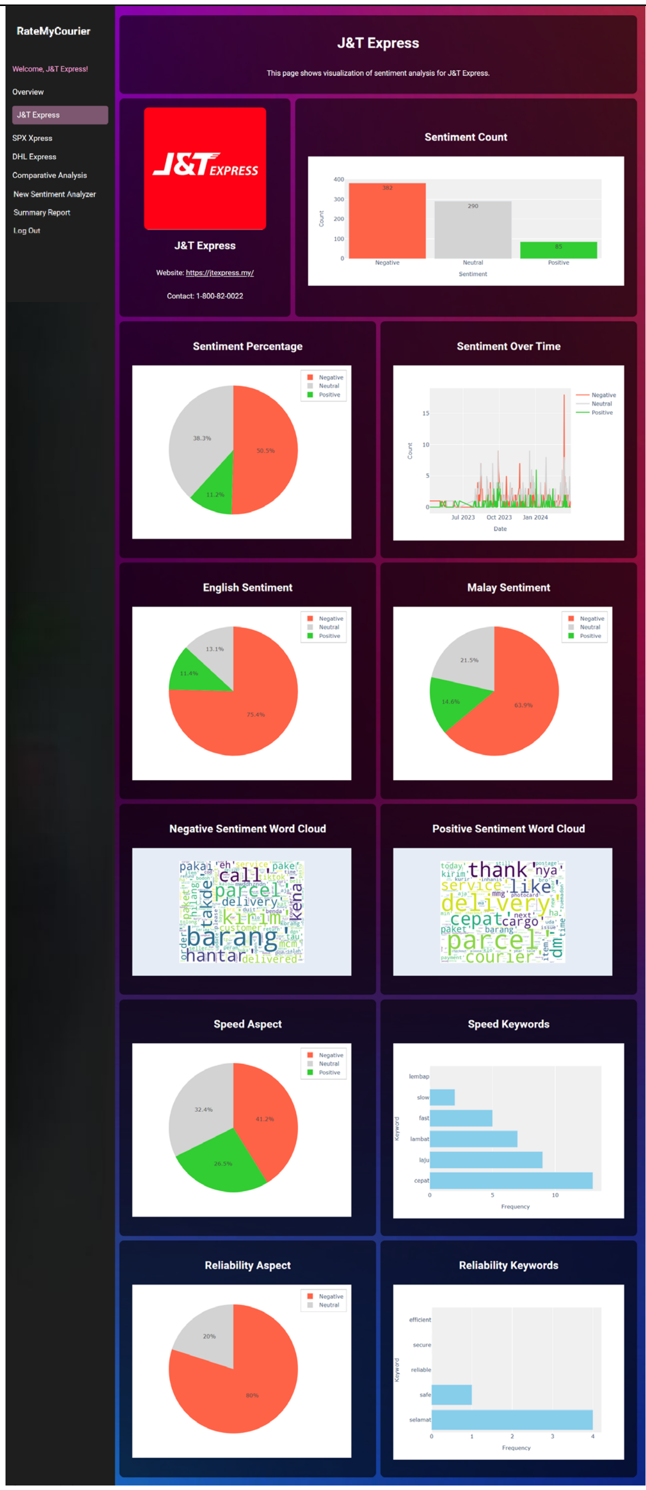
	B. When viewed as registered user, there are eight menus available on the sidebar menu. 

5.3.5 View J&T Express Page

Table 5.6 discusses the test results of “View J&T Express Page”.

Table 5.6 Functionality Testing for J&T Express Page

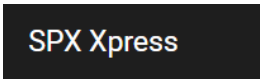
Test Objective	To verify the system displays the J&T Express page correctly.
Potential Test Input	User clicks “J&T Express” menu on the sidebar menu. 
Expected Test Output	J&T Express page loads successfully. All visualizations on the page are displayed correctly and interactive.
Test Procedures	User clicks “J&T Express” menu on the sidebar menu.

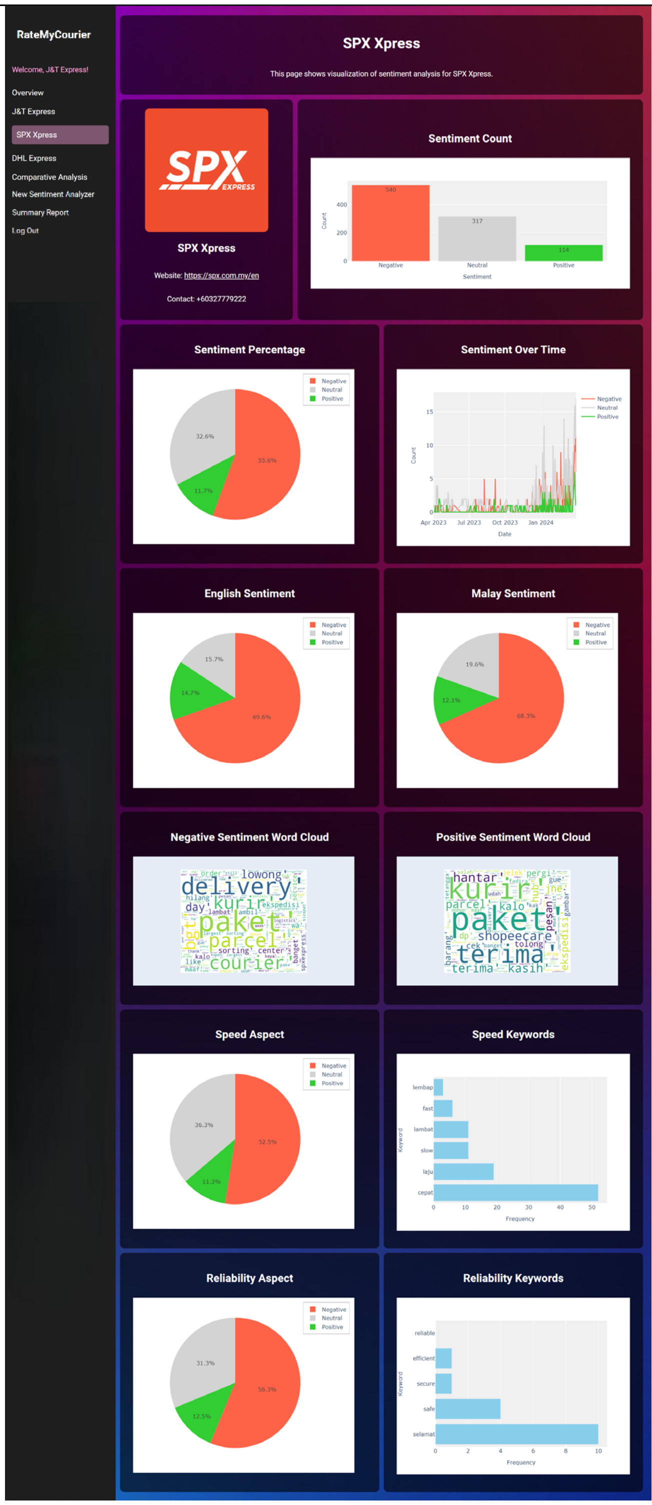
<table>
<tr>
<td>Actual Test Results</td>
<td>
</td>
</tr>
</table>

5.3.6 View SPX Xpress Page

Table 5.7 discusses the test results of “View SPX Xpress Page”.

Table 5.7 Functionality Testing for SPX Xpress Page

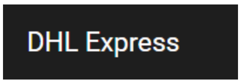
Test Objective	To verify the system displays the SPX Xpress page correctly.
Potential Test Input	User clicks “SPX Xpress” menu on the sidebar menu. 
Expected Test Output	SPX Xpress page loads successfully. All visualizations on the page are displayed correctly and interactive.
Test Procedures	User clicks “SPX Xpress” menu on the sidebar menu.

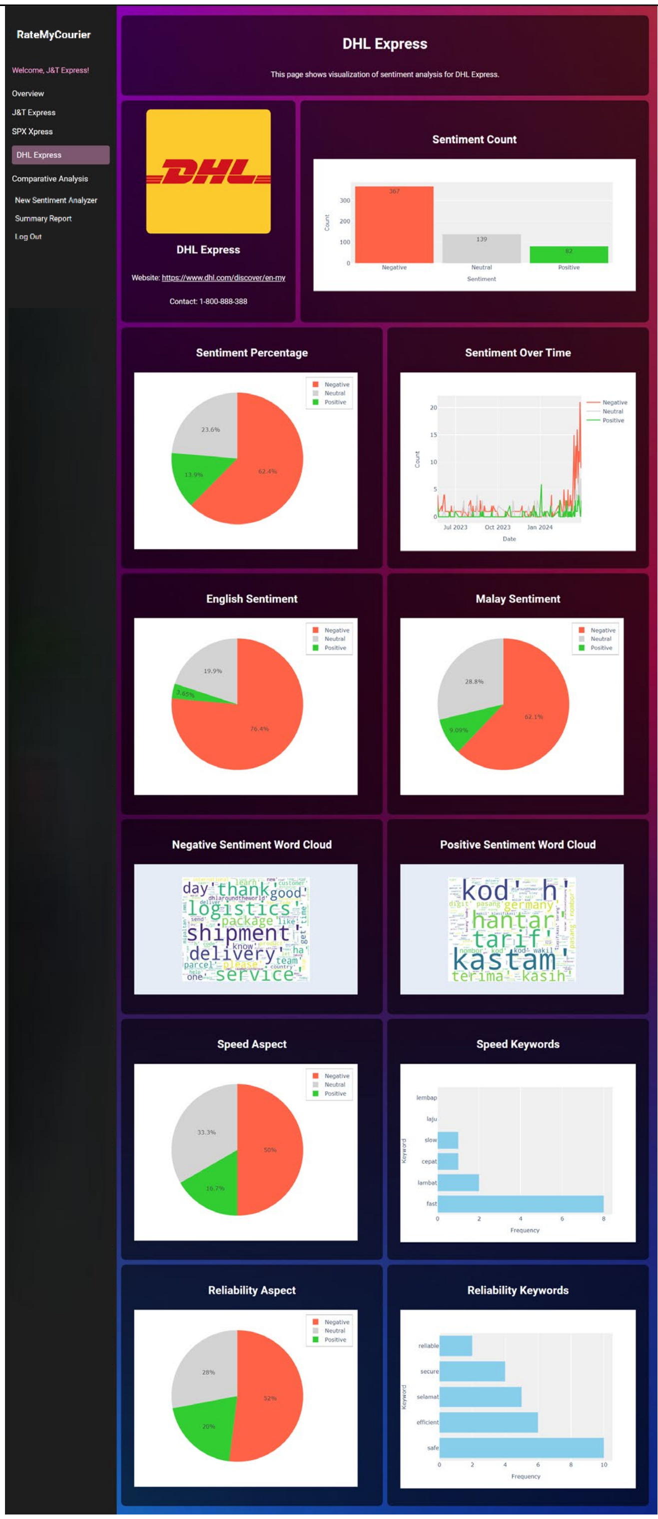
Actual Test Results	

5.3.7 View DHL Express Page

Table 5.8 discusses the test results of "View DHL Express Page".

Table 5.8 Functionality Testing for DHL Express Page

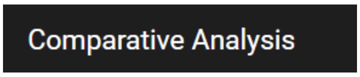
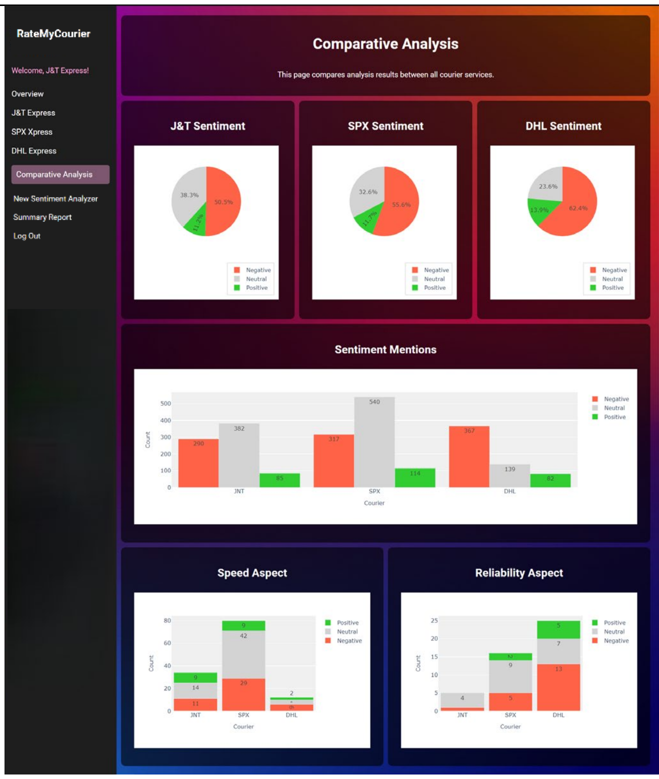
Test Objective	To verify the system displays the DHL Express page correctly.
Potential Test Input	User clicks "DHL Express" menu on the sidebar menu. 
Expected Test Output	DHL Express page loads successfully. All visualizations on the page are displayed correctly and interactive.
Test Procedures	User clicks "DHL Express" menu on the sidebar menu.

Actual Test Results	

5.3.8 View Comparative Analysis Page

Table 5.9 discusses the test results of “View Comparative Analysis Page”.

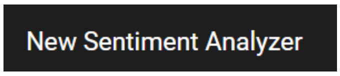
Table 5.9 Functionality Testing for Comparative Analysis Page

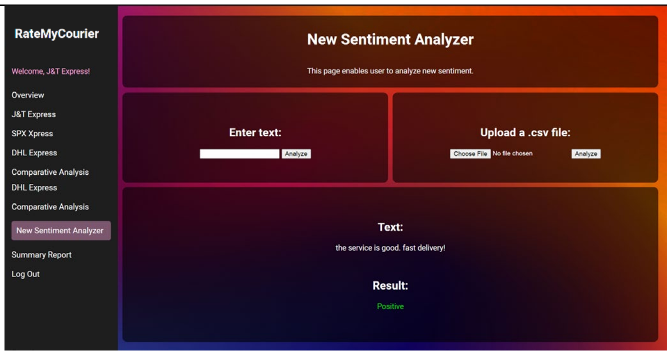
Test Objective	To verify the system displays the Comparative Analysis page correctly.
Potential Test Input	User clicks “Comparative Analysis” menu on the sidebar menu. 
Expected Test Output	Comparative Analysis page loads successfully. All visualizations on the page are displayed correctly and interactive.
Test Procedures	User clicks “Comparative Analysis” menu on the sidebar menu.
Actual Test Results	

5.3.9 View New Sentiment Analyzer Page

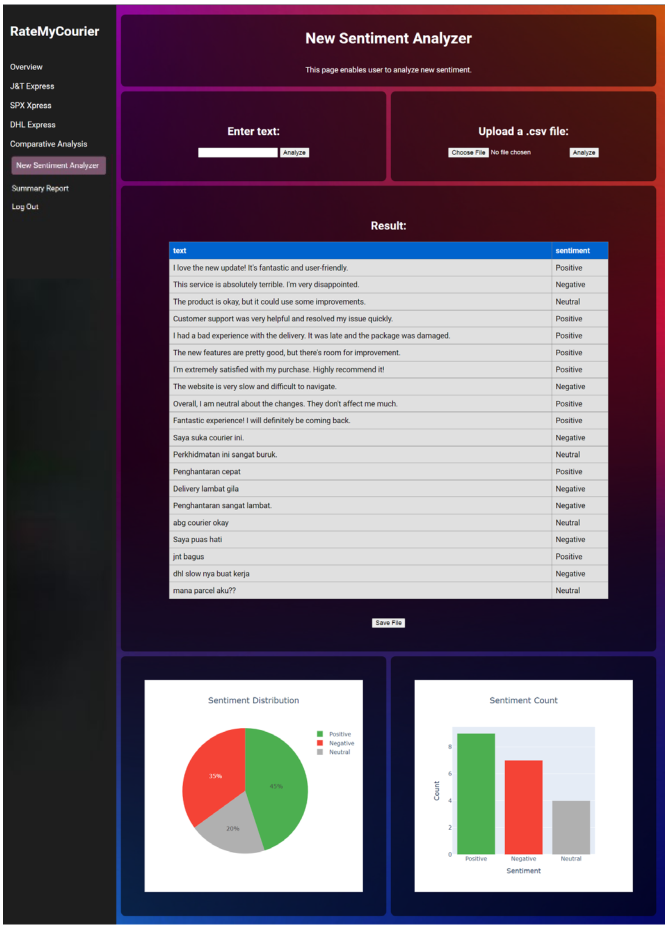
Table 5.10 discusses the test results of “View New Sentiment Analyzer Page”.

Table 5.10 Functionality Testing for New Sentiment Analyzer Page

Test Objective	To verify the system displays the New Sentiment Analyzer page correctly, and to ensure “Analyze” buttons work correctly for both text and file input, and to assure the output is successfully displayed.
Potential Test Input	User clicks “New Sentiment Analyzer” menu on the sidebar menu. 
Expected Test Output	New Sentiment Analyzer page loads successfully. The output for text and file input from user are displayed successfully.
Test Procedures	A. Text Input 1. User clicks “New Sentiment Analyzer” menu on the sidebar menu. 2. User enters text that they want to analyze. 3. User clicks “Analyze” button. 4. Sentiment result is displayed on the page. B. File Input 1. User clicks “New Sentiment Analyzer” menu on the sidebar menu. 2. User clicks “Choose File” button and uploads the .csv file that they wish to analyze. 3. User clicks “Analyze” button. 4. Sentiment results are displayed and visualised. 5. The results file is downloaded when the “Save File” button is clicked.
Actual Test Results	A. For text input, user enters a simple text in the input box and clicks the “Analyze” button. The entered text and sentiment result are displayed on the page.



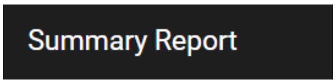
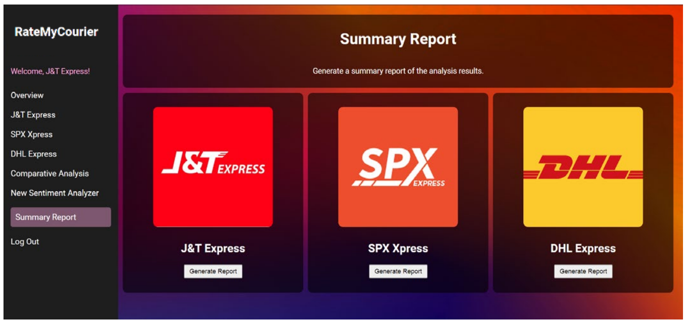
B. For file input, user uploads a .csv file that they wish to analyze and clicks the "Analyze" button. The sentiment results are displayed in a table. A "Save File" button is shown below the table, allowing user to download the results. The results are also visualized in pie chart and bar chart at the bottom of the page.



5.3.10 View Summary Report Page

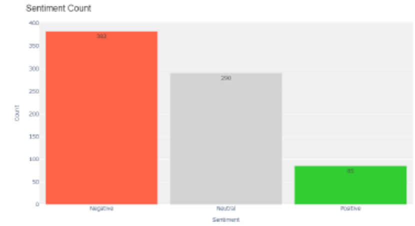
Table 5.11 discusses the test results of "View Summary Report Page".

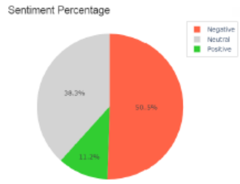
Table 5.11 Functionality Testing for Summary Report Page

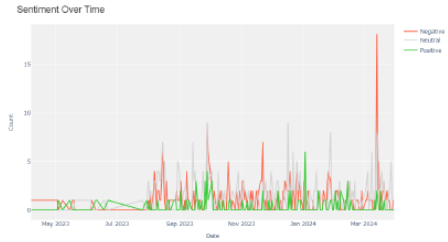
Test Objective	To verify the system displays the Summary Report page correctly, and to ensure the summary report file is produced successfully.
Potential Test Input	User clicks "Summary Report" menu on the sidebar menu. 
Expected Test Output	Summary Report page loads successfully. The summary report file is downloaded successfully when "Generate Report" button is clicked.
Test Procedures	User clicks "Summary Report" menu on the sidebar menu. User clicks "Generate Report" button for their own company. The summary report file is downloaded.
Actual Test Results	For this test, user is logged in as "J&T Express". User generates report for J&T Express.  Summary report for J&T Express is downloaded as PDF file.

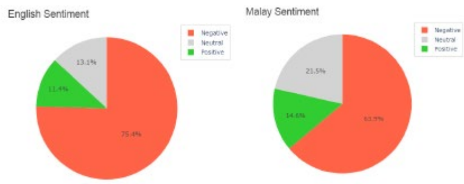
Summary Report

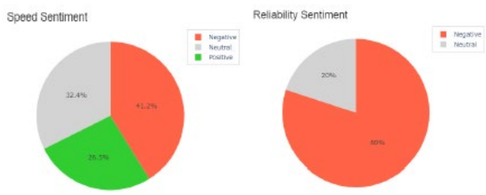
This report provides an overview of sentiment analysis for J&T Express.











5.4 Summary

A prediction model was created and assessed, achieving an accuracy of 81.78% for the English model, 84.06% for the Malay model, and 93.40% for the Courier model. This chapter provides an in-depth analysis of real-world data, illustrated through various charts that offer detailed insights. Additionally, functionality testing was conducted to evaluate the performance of each feature comprehensively.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

This final chapter concludes the findings that have been discovered throughout this project. The project limitations and recommendations are also discussed, with hope that the project could be improved in the future.

6.1 Project Conclusion

RateMyCourier is a web-based system that visualizes the result of sentiment analysis performed on tweets from X, about three courier services—J&T Express, SPX Xpress, and DHL Express—from 1st April 2023 to 1st April 2024. To analyse the sentiments, a Naïve Bayes classifier model was developed and integrated into the system. This aids the performance evaluation of the courier services and provides users with valuable insights for informed decision-making.

This project employs a variety of interactive charts to visualize the data, including bar charts, pie charts, line charts, and word clouds. These visualizations help users comprehend the data more thoroughly and extract significant insights. Additionally, the system provides visualizations from multiple aspects, allowing users to stay updated on various facets of courier performance.

6.1.1 Objective I

The first objective of this project is to design a web-based classification and visualization system for X sentiment analysis of courier services in Malaysia. In-depth research was done to set up the study's background, problem statements, objectives, significance, and scope. During the design phase, a flowchart, use case diagram, and algorithm were made to show the system's features. The project chose the top three courier services in Malaysia: J&T Express, SPX Xpress, and DHL Express. In system development phase, Flask is implemented to establish the web framework of the system. The Naïve Bayes classifier model was used to analyse sentiments expressed on X. By using Plotly, interactive visuals like bar charts, pie charts, line charts, and word clouds, were created to depict the sentiment analysis results. Finishing the design of the web system met the project's first objective.

6.1.2 Objective II

The second objective of this project is to develop the designed system using Naive Bayes and Plotly. The Naïve Bayes classifier model is an essential part of the project. Data is gathered using a web scraping tool called Tweet Harvest and then classified into positive, negative, and neutral sentiments with the Naïve Bayes classifier model. The sentiment analysis results are effectively displayed in the web application using the Plotly library. With the successful completion of these steps, it can be concluded that the project's second objective has been completed.

6.1.3 Objective III

The third objective of this project is to test the functionality of the system. The testing process includes functionality tests to validate the system's performance against predefined criteria. During functionality testing, the system has successfully demonstrated its ability to operate effectively and meet the intended requirements. Essential output metrics, such as precision, recall, F1-score, and accuracy rates for the classifier models, were calculated. For the English model, the accuracy during training is 86.44%, while during testing it is 81.78%. Similarly, for the Malay model, the accuracy during training is

91.10%, and during testing, it is 84.06%. For the Courier model, the accuracy during training is 98.26%, while during testing it is 93.40%. After undergoing numerous tests, the system performed as expected, leading to the conclusion that the third objective has been achieved.

6.2 Limitations

The system faces several challenges that hinder its full potential. One major issue is the lack of dedicated libraries for processing Malay data; the only available option is PySastrawi, which is tailored for Indonesian. Additionally, the NLTK library does not provide stop words for the Malay language. These constraints can result in inaccuracies when analysing and interpreting tweets in Malay.

Moreover, data extracted from Twitter includes many slangs, short forms, and abbreviations commonly used by users. These can lead to phrases being misinterpreted, causing them to lose their true meaning. Not only that, sarcastic expressions can be problematic for the classifier model, as it may fail to capture the intended meaning of tweets. Sarcasm and sarcastic statements present challenges in accurately detecting sentiment based solely on text, as they often involve saying the opposite of what is meant for humorous or mocking purposes. This nuance is difficult to identify, negatively impacting categorization accuracy.

Additionally, X imposes a daily tweet limit depending on the user's account type (normal or premium) to prevent excessive web scraping activities on the platform. While these security measures are crucial for user privacy and data protection, they can create obstacles in collecting enough data for thorough analysis and accurate sentiment classification.

6.3 Recommendations

Looking ahead, expanding the scope of data collection beyond X could offer a more nuanced understanding of customer sentiments. Exploring other social media platforms, such as TikTok, could enrich the dataset with diverse perspectives and insights. By incorporating reviews from a variety of sources, a more well-rounded view of user opinions can be gathered.

To further refine the sentiment analysis process, developing comprehensive dictionaries that cover diverse language elements like slang, abbreviations, and sarcastic comments would be highly advantageous. These specialized dictionaries would help in accurately interpreting the subtleties of language and improve the precision of sentiment classification.

In addition, experimenting with different machine learning algorithms, such as Support Vector Machines (SVM) and K-Nearest Neighbours (KNN), could prove beneficial. Testing these alternative methods may reveal new opportunities for enhancing the sentiment analysis model, leading to more advanced and flexible classification techniques.

REFERENCES

Addepalli, L., Lokhande, G., Sakinam, S., Hussain, S., Mookerjee, J., Vamsi, U., Waqas, A., & Vidya Sagar, S. D. (2023). Assessing the Performance of Python Data Visualization Libraries: A Review. International Journal of Computer Engineering in Research Trends, 10(1), 04. https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104

Al-Jarrah, I., Mustafa, A., & Najadat, H. (2023). Aspect-Based Sentiment Analysis for Arabic Food Delivery Reviews. ACM Transactions on Asian Low-Resource Language Information Processing, 22(7), 200. https://doi.org/10.1145/3605146

Andrae, F., & Kothuri, R. S. (2023). Guidelines on choosing between native and cross-platform development: A comparative study on the efficiency of native and cross-platform mobile development. KTH Royal Institute of Technology. Retrieved on December 15, 2023, from https://kth.diva-portal.org/smash/get/diva2:1801198/FULLTEXT01.pdf

Basuki, Ari. (2023). Sentiment Analysis of Customers' Review on Delivery Service Provider on Twitter Using Naive Bayes Classification. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 9(2), 420-428. https://doi.org/10.26555/jiteki.v9i2.26327

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134

Brooke, J. (1995, November). SUS: A quick and dirty usability scale. Retrieved on March 5, 2024, from https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale

Dewi, E., Aldya, A., Rachman, A., & Ramdani, A. (2023). Comparative Sentiment Analysis of Delivery Service PT.POS Indonesia and J&T Express on Twitter Social Media Using The Support Vector Machine Algorithm. IJISTECH (International Journal of Information System and Technology), 6(5), 664-676. https://doi.org/10.30645/ijistech.v6i5.284

Frost, J. (2021). Line Charts: Using, Examples, and Interpreting. Statistics By Jim. Retrieved on June 10, 2024, from https://statisticsbyjim.com/graphs/line-charts/

Gamble, Y. (2022, November 4). Bar and Column Charts in Power BI. Retrieved on January 25, 2024, from https://www.fe.training/free-resources/power-bi-data-visualization/bar-and-column-charts-in-power-bi/

Gupta, D. (2023, December 5). Software Development Life Cycle (SDLC). Retrieved on February 10, 2024, from https://www.geeksforgeeks.org/software-development-life-cycle-sdlc/

Gupta, S. (2018, January 19). Sentiment analysis: Concept, analysis and applications. Retrieved on March 15, 2024, from https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. In 2014 47th Hawaii International Conference on System Sciences (pp. 1833-1842). IEEE. https://doi.org/10.1109/HICSS.2014.231

Hendayani, R., & Dharmawan, M. (2020). Strategies for Improving the Quality of Logistics Courier Services Through Priority Problem-solving Based on Multiclass Classification. IOP Conference Series: Materials Science and Engineering, 879(1). https://doi.org/10.1088/1757-899X/879/1/012051

Hetler, A. (2023, July). What is Twitter?. Retrieved on April 2, 2024, from https://www.techtarget.com/whatis/definition/Twitter

Joshi, S., Dubey, R., Tiwari, A., & Jindal, P. (2021). Sentiment Analysis Algorithms: Classifiers and Their Comparison. In B. Singh, C. A. Coello Coello, P. Jindal, & P. Verma (Eds.), Intelligent Computing and Communication Systems. Algorithms for Intelligent Systems. Springer, Singapore.

Khan, S. M. (2023). Waterfall model used in software development reference: Software requirements engineering waterfall model. https://doi.org/10.13140/RG.2.2.29580.69764

Khun, N. H., & Thant, H. A. (2019). Visualization of Twitter Sentiment during the Period of US Banned Huawei. In 2019 International Conference on Advanced Information Technologies (ICAIT) (pp. 274-279). Yangon, Myanmar. https://doi.org/10.1109/AITC.2019.8921014

Kozak, M., Hartley, J., Wnuk, A., & Tartanus, M. (2015). Multiple pie charts: Unreadable, inefficient, and over-used. Journal of Scholarly Publishing, 46(3), 282-289. https://doi.org/10.3138/jsp.46.3.05

Li, J. (Ed.). (2024). Advances in Sentiment Analysis - Techniques, Applications, and Challenges. IntechOpen. https://doi.org/10.5772/intechopen.111293

Liang, T., Lu, S., & Liu, Q. (2020). Data Visualization System Based on Big Data Analysis. 2020 International Conference on Robots & Intelligent System (ICRIS), 76-79. https://doi.org/10.1109/ICRIS52159.2020.00027

Lutkevich, B. (2022, November). What is the waterfall model?. Retrieved on March 25, 2024, from https://www.techtarget.com/searchsoftwarequality/definition/waterfall-model

Mahardika, A. F. P., Larasati, A., & Muid, A. (2022). The Sentiment Analysis of Online Customer Review on Food and Beverages Delivery Services in the GOJEK Application Using K-Nearest Neighbors. 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management. https://doi.org/10.46254/AP03.20220167

MALAYSIAN COMMUNICATIONS AND MULTIMEDIA COMMISSION. (2012, January 31). Performance Standards for Postal Services (Domestic Parcel Service). Retrieved on April 5, 2024, from https://www.mcmc.gov.my/skmmgovmy/files/attachments/Domestic_Parcel_Service.pdf

MALAYSIAN COMMUNICATIONS AND MULTIMEDIA COMMISSION. (2024, February 9). LIST OF LICENSEES UNDER THE POSTAL SERVICES ACT 2012 (LIST OF NON-UNIVERSAL SERVICE LICENSEES). Retrieved on February 12, 2024, from https://www.mcmc.gov.my/skmmgovmy/media/General/pdf2/List-of-Non-Universal-Service-Licensees.pdf

Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices, and applications. International Journal of Scientific & Technology Research, 9(2), 601.

Melinmani, S. (2014, December 15). Waterfall Model pros and cons. Retrieved on January 10, 2024, from https://www.experts-exchange.com/articles/17441/Waterfall-Model-pros-and-cons.html

Mentzer, J. T., Flint, D. J., & Hult, G. T. M. (2001). Logistics service quality as a segment-customized process. Journal of Marketing, 65(4), 82-104. https://doi.org/10.1509/jmkg.65.4.82.18390

Mindoro, J., Susa, J., Malbog, M., Acoba, A., Nipas, M., & Gulmatico, J. (2022). Sentiment Analysis in Customer Experience in Philippine Courier Delivery Services using VADER Algorithm Thru Chatbot Interviews. Second International Conference on Power, Control and Computing Technologies (ICPC2T), 978(1). https://doi.org/10.1109/ICPC2T53885.2022.9777007

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1), 81. https://doi.org/10.1007/s13278-021-00776-6

Narkhede, S. (2018, May 9). Understanding confusion matrix. Towards Data Science. Retrieved on January 15, 2024, from https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Needle, F. (2023, August 24). The benefits of customer feedback, according to experts. HubSpot. Retrieved on February 20, 2024, from https://blog.hubspot.com/service/benefits-of-customer-feedback

Nuha, U., & Lin, C.-H. (2023). Aspect-based sentiment analysis with semi-supervised approach on Taiwan social distancing app user reviews. In 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 444-447). Bali, Indonesia. https://doi.org/10.1109/ICAIIC57133.2023.10067048

Pandey, A., & Vishwakarma, D. K. (2024). Progress, achievements, and challenges in multimodal sentiment analysis using deep learning: A survey. Applied Soft Computing, 152, 111206. https://doi.org/10.1016/j.asoc.2023.111206

Petersen, K., Wohlin, C., & Baca, D. (2009). The waterfall model in large-scale development. In F. Bomarius, M. Oivo, P. Jaring, & P. Abrahamsson (Eds.), Product-focused software process improvement. PROFES 2009. Lecture Notes in Business Information Processing (Vol. 32, pp. 245-258). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02152-7_29

Prabhune, A. G., Shri Hari, V. R., & Sethiya, N. K. (2023). Adaptation of a BERT model to the India restaurant data using rule-based approach for aspect-based sentiment analysis. In 2023 Second International Conference on Smart Technologies for Smart Nation (SmartTechCon) (pp. 457-461). Singapore, Singapore. https://doi.org/10.1109/SmartTechCon57526.2023.10391510

Pravina, A. M., Sani, K. N., Harfianto, H. D., Pratama, T. A., Fahrina, A., & Ruldeviyani, Y. (2022). Sentiment analysis of delivery service opinions on Twitter documents using K-nearest neighbor. JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 9(2), 996-1012. https://doi.org/10.35957/jatisi.v9i2.1899

Rajath, N., Kushal, V., Kodipalli, A., Rao, T., Pushpalatha, V., & Rohini, B. R. (2023). Sentiment analysis of Twitter using machine learning algorithms. In 2023 International Conference on Computational Intelligence for Information, Security and Communication Applications (CIISCA) (pp. 250-254). Bengaluru, India. https://doi.org/10.1109/CIISCA59740.2023.00056

Rawat, A., Maheshwari, H., Khanduja, M., Kumar, R., Memoria, M., & Kumar, S. (2022). Sentiment analysis of COVID-19 vaccines tweets using NLP and machine learning classifiers. In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) (pp. 225-230). Faridabad, India. https://doi.org/10.1109/COM-IT-CON54601.2022.9850629

Royal Malaysian Customs Department. (2018, August 21). Guide on courier services: Service tax 2018. Internal Tax Division. Retrieved on March 10, 2024, from https://mysst.customs.gov.my/assets/document/Industry%20Guides/GI/Guide%20on%20Courier%20Services.pdf

Sahinbas, K., & Avci, A. (2022). Sentiment analysis of customer review in online food delivery industry. International Journal of Engineering and Innovative Research, 4(3), 196-207. https://doi.org/10.47933/ijeir.1174377

Samah, K. A. F. A., Misdan, N. F. A., Jono, M. N. H. H., & Riza, L. S. (2022). The best Malaysian airline companies visualization through bilingual Twitter sentiment analysis: A machine learning classification. International Journal on Informatics Visualization (JOIV), 6(1), 130-137. https://doi.org/10.30630/joiv.6.1.879

Sankari, S., & Ebenezer, P. (2022). A sentiment analysis on tweets regarding food delivery companies via lexicon-based technique. International Research Journal of Modernization in Engineering Technology and Science (IRJMETS), 4(12). https://doi.org/10.56726/IRJMETS32057

Scheibengraf, M. (2023, September). Multilingual sentiment analysis. Crisol Translations. Retrieved on April 5, 2024, from https://www.crisoltranslations.com/our-blog/multilingual-sentiment-analysis/

SDLC Models: Agile, Waterfall, V-Shaped, Iterative, Spiral. (2017, August 4). Existek. Retrieved on May 2, 2024, from https://existek.com/blog/sdlc-models/

Selvarajah, K., Craven, M., Massey, A., Crowe, J., Vedhara, K., & Raine-Fenning, N. (2013). Native apps versus web apps: Which is best for healthcare applications? Lecture Notes in Computer Science, 8005, 189-196. https://doi.org/10.1007/978-3-642-39262-7_22

Sentiment Analysis: A definitive guide. (2023). MonkeyLearn. Retrieved on May 10, 2024, from https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20algorithms%20fall%20into,rule%2Dbased%20and%20automatic%20approaches

Sentiment analysis: Types, tools, and use cases. (2023, September). AltexSoft. Retrieved on June 15, 2024, from https://www.altexsoft.com/blog/sentiment-analysis-types-tools-and-use-cases/

Shaikh, M. (2023, August 16). Matplotlib vs. Seaborn vs. Plotly: A comparative guide. Medium. Retrieved on January 25, 2024, from https://medium.com/@mohsin.shaikh324/matplotlib-vs-seaborn-vs-plotly-a-comparative-guide-c99a0059c09f#:~:text=If%20you%20prefer%20precise%20control,Plotly%20offers%20a%20powerful%20solution

Singh, Y. (2022, July 24). How to customize pie charts using matplotlib. Proclus Academy. Retrieved on February 15, 2024, from https://proclusacademy.com/blog/customize_matplotlib_piechart/

Starbuck, C. (2023). Data visualization. In The Fundamentals of People Analytics. Springer, Cham. https://doi.org/10.1007/978-3-031-28674-2_15

Tandon, U. (2022). Shipment tracking, delivery speed, and product presentation as antecedents of repurchase intention: Predictors of online shopping repurchase intention. In S. Rahman (Ed.), Handbook of research on the platform economy and the evolution of e-commerce (pp. 20). Chitkara University, India.

Visualizing words: Inspiring word cloud examples. (2024). PPCexpo. Retrieved on March 25, 2024, from https://ppcexpo.com/blog/word-cloud-examples

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731-5780. https://doi.org/10.1007/s10462-022-10144-1

Waskom, M. L. (2021). seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021. https://doi.org/10.21105/joss.03021

What is SDLC Waterfall Model? (2023, June 23). Software Testing Help. Retrieved on April 20, 2024, from https://www.softwaretestinghelp.com/what-is-sdlc-waterfall-model/

Willas, S. (2023, December 12). 7 reasons online reviews are essential for your brand. Mention. Retrieved on June 30, 2024, from https://mention.com/en/blog/online-reviews/#:~:text=Online%20reviews%20represent%20real%2Dlife,can%20significantly%20influence%20purchasing%20decisions

APPENDICES

APPENDIX A: SURVEY QUESTIONS

Survey of Courier Services in Malaysia

This survey is carried out for a Final Year Project to understand the usage of courier services in Malaysia. All information is used strictly for research purposes only.

2021610118@student.uitm.edu.my Switch account

Not shared

* Indicates required question

What is your age? *

Your answer

What is your gender? *

- Male
- Female

Which state do you live in? *

Choose

How often do you use courier services? *

- Very often
- Often
- Rarely
- Very rarely

Which courier service do you use the most? *

- J&T Express
- DHL Express
- Ninja Logistics
- ABX Express
- City-Link Express
- Flash Malaysia
- GD Express
- Pgeon Express
- Skynet Express
- SPX Xpress
- Lazada Express
- Other:

Do you find it difficult to analyze online reviews about courier services? *

- Yes
- No

Do you think there is a need for a central platform that analyzes reviews of courier services in Malaysia? *

- Yes
- No

Do you think visualization for such platform is needed? *

- ◯ Yes
- ◯ No

Submit

Clear form

Never submit passwords through Google Forms.

This form was created inside of Universiti Teknologi MARA. Report Abuse

Google Forms

APPENDIX B: SURVEY RESULTS

Survey of Courier Services in Malaysia

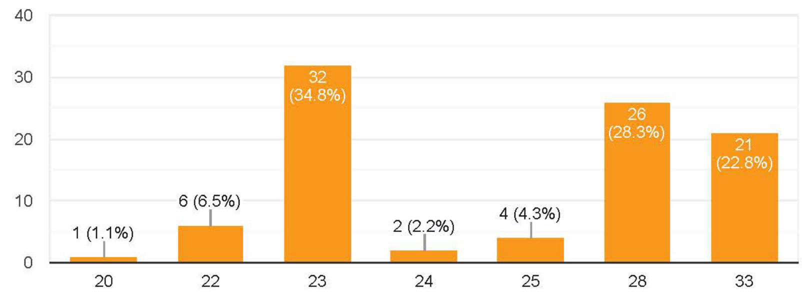
92 responses

Publish analytics

What is your age?

Copy

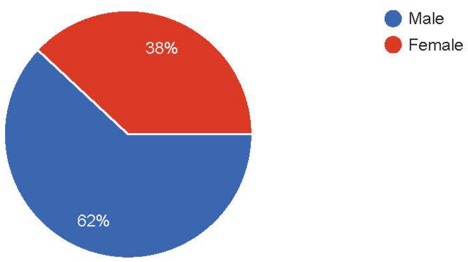
92 responses



What is your gender?

Copy

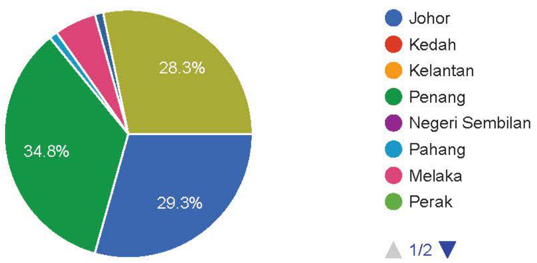
92 responses



Which state do you live in?

Copy

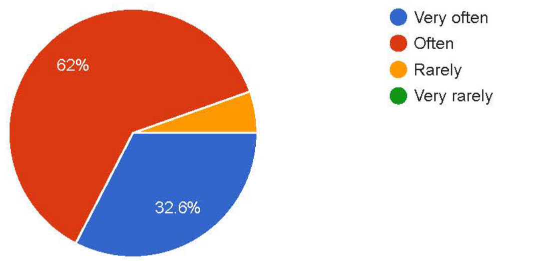
92 responses



How often do you use courier services?

Copy

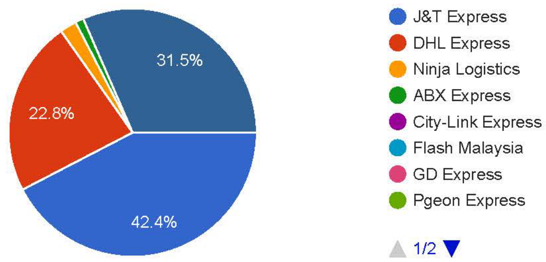
92 responses



Which courier service do you use the most?

Copy

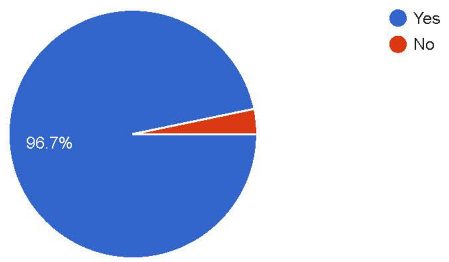
92 responses



Do you find it difficult to analyze online reviews about courier services?

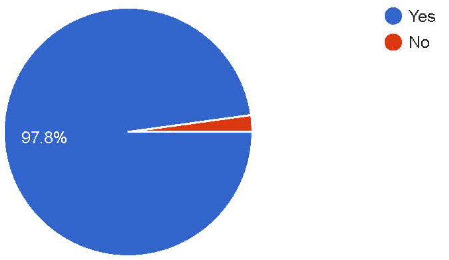
Copy

92 responses



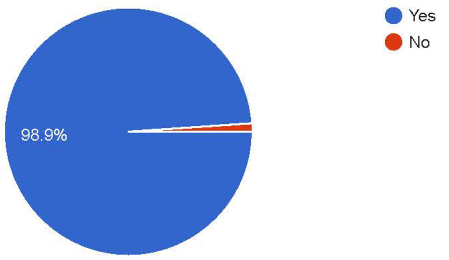
Do you think there is a need for a central platform that analyzes reviews of courier services in Malaysia?

92 responses



Do you think visualization for such platform is needed?

92 responses



This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

APPENDIX C: GANTT CHART

	2023			2024							
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug
Requirement analysis											
Design											
Implementation											
Testing											
Documentation											

APPENDIX D: PLAGIARISM REPORT

Similarity Report

16% Overall Similarity

Top sources found in the following databases:

- 8% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 14% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Universiti Teknologi MARA on 2021-08-01 Submitted works	1%
2	researchgate.net Internet	<1%
3	Universiti Teknologi MARA on 2017-07-24 Submitted works	<1%
4	ijistech.org Internet	<1%
5	ses.library.usyd.edu.au Internet	<1%
6	ieomsociety.org Internet	<1%
7	fastercapital.com Internet	<1%
8	Universiti Teknologi MARA on 2015-12-13 Submitted works	<1%

Sources overview

Universiti Teknologi MARA
KM 26 Jalan Lendu
78000 Alor Gajah
Melaka Bandaraya Bersejarah
Tel: +606 558 2000
Faks: +606 558 2001

Our reference : 600CM(PJI/RMU.5/5/12)
Date : 2nd April 2024

SITI KHADIJAH BINTI ZAZALI
College Of Computing, Informatics And Mathematics,
Universiti Teknologi MARA, Cawangan Melaka,
Kampus Jasin,
77300 Merlimau, Melaka

Dear Madam,

APPROVAL LETTER - UiTM MELAKA RESEARCH ETHICS

The Branch Ethics Review Committee (BERC) UiTM Cawangan Melaka has considered and approved your Research Ethics application.

Details of the approval are as follows:

Referral No.	BERC/MLK/120/2024
Proposal Title	X Sentiment Analysis: Classification And Visualisation Of Courier Services In Malaysia Using Naive Bayes And Plotly
Approval Period	1st October 2023 - 1st July 2024
Authorised Personnel	Sir Khairul Nizam Bin Abd Halim

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- The submission of the final report must be submitted to the Ethics Office on or before the anniversary of approval and completion of the subject.
- You must report as soon as practicable anything that might warrant a review of ethical approval of the project, including:
    - Serious or unexpected adverse events (which should be reported within 72 hours)
    - Unforeseen events that might affect the ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate immediate risk to participants)

Yours sincerely,

DR. KHAIRUNNISA ABD SAMAD
Chairperson
UiTM Branch Research Ethics Committee

c.c.: Rector, UiTM Melaka Branch