Sentiment Analysis towards Car Reviews With Data Visualization

Mohamad Hafiz Khairuddin*, Nurazian Binti Mior Dahalan, Albin Lemuel Kushan, Nur Farhana Binti Mohd Nasir

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Cawangan Melaka Kampus Jasin, 77300 Merlimau, Melaka

DOI: https://dx.doi.org/10.47772/IJRISS.2025.910000585

Received: 27 October 2025; Accepted: 02 November 2025; Published: 19 November 2025

ABSTRACT

Nowadays, there are too many car reviews on the internet, worldwide. Big manufacturing companies use user feedback to improve product quality by understanding the user perspective. Customers will read reviews on websites or on social media platforms before deciding which cars to buy, and may consider testing at a nearby showroom. So, reviews are very important to both the manufacturer and the customer. Nevertheless, it is hard to extract useful information from hundreds or thousands of reviews on websites or social media platforms. Sentiment analysis is applied across various areas, such as business and products, to analyse and learn from people's opinions. Fine-grained sentiment analysis is best for analysing the polarity of a sentence and determining its sentiment —positive, negative, or neutral. After preprocessing the reviews, extract features and use Naïve Bayes to classify sentiment. The results will be displayed in the dashboard visualisation so the user can read all the reviews properly. Functional testing is conducted to ensure the system runs smoothly, as it should. There is a need to improve this system, as some of these car models are not very common in Malaysia. Later, we can get data on Malaysia's standard car models and apply the system to them. The model's classification method accuracy could be improved by training and testing the system on a large number of reviews.

Keywords: Car Review, Sentiment Analysis, Naïve Bayes

INTRODUCTION

With the ongoing economic development, more and more families are considering buying a car. For ordinary families, buying a car is an important and relatively expensive thing. So, it is important to choose a car which has a suitable price and quality. Examples of top automobile companies are Audi, BMW, Honda, and Mitsubishi. Users have many choices when it comes to their favourite automobile companies, but most people prefer those that offer high-quality products and good service. With that, a high-quality product can be an important service. Quality is defined as meeting or exceeding the client's expectations.

TikTok, Facebook, Google, and other online platforms like Shopee are now among the most popular platforms for users to share their buying experiences through online reviews and ratings, or to share important information for users. (Adwan et al., 2020). On many of these platforms, buyers can share their feelings, opinions, or even suggestions about the product or the manufacturer. For example, users post their comments on online websites that are easy to access and usually include a star rating ranging from 1 to 5 (Alamanda et al., 2019). So, users no longer refer to others when buying a car, as they can read reviews on websites. In particular, while users read reviews, ratings do not always accurately reflect their review sentiment.

Furthermore, the available data in raw format is not an easy task to analyse in a limited time. The reviews could be a challenge for a user to extract important information, as they include too many types of information in a single sentence. Therefore, a study suggests that sentiment analysis is needed, as it can help classify reviews into positive and negative categories to improve service performance and product quality (Panchal & Deshmukh, 2020).

Sentiment analysis (also referred to as subjectivity analysis, opinion mining, or emotion artificial intelligence) is a natural language processing (NLP) technique that identifies patterns and features from a large text corpus

(Lamba, Manika, & Margam; Madhusudhan, 2022). Negative and slightly negative ratings frequently result in sales loss. In addition, machine learning, lexicon-based, and aspect-based sentiment analysis are practical approaches that can produce categorical sentiment (positive or negative).

A Naïve Bayes classifier can classify Twitter tweets into positive or negative (Al-Natour & Turetken, 2020). It also operates on the principles of probability and assumes independence among features, which is why it is called "naïve". For sentiment analysis, it can effectively handle textual data considering the presence of certain words or phrases (Deshmukh et al., 2023). The dashboard visualisation highlights sentiment analysis results to help the user better understand them.

Problem Statement

Nowadays, customers do not have to enter the shop to extract the information about the automobile they wish to purchase. They can get all kinds of information immediately by clicking a mouse and browsing social media platforms. However, with easy access to information, this creates a disequilibrium between demand and supply (Shamsher Singh & Ameet Sao, 2021). So, people just believe reviews on any social media platform, but not on the official website. Every car buyer in the country starts their search on the World Wide Web. Social media platforms have become common channels for businesses and organisations to market their products. Extracting meaningful information from consumer reviews, such as the most frequent words and their relationships, provides the company with insights to address and resolve issues quickly (Kim, E., & Chun, S., 2019).

So, consumers face challenges in choosing the right cars from a vast network due to the sheer volume of data, diverse types, and the low density of valuable information. Collecting consumer feedback through online surveys is both expensive and time-consuming for automobile companies (Panchal & Deshmukh, 2020). This is because reviews may not depict the quality of the company's products. Consequently, automobile companies find it challenging to deliver service or product quality that surpasses consumer expectations, leading to a direct loss of both potential customers and revenue.

Thus, Awais et al. (2020) suggest that companies should collect user experience data and perform sentiment analysis to assess the polarity of the text —whether it is a positive or negative review. Extracting insights from such feedback can contribute to knowledge. When the company analyses the polarity of reviews, it can gauge user opinions on the quality and effectiveness of its products.

Related Works

There are three types of related work similar to this project: Machine Learning Model for Sentimental Analysis of Amazon Reviews, Sentiment Analysis for social media using SVM Classifier, and Sentiment Analysis of YouTube Movie Trailer Comments using Naïve Bayes.

Machine Learning Model for Sentimental Analysis of Amazon Reviews

This research aims to deepen the understanding of online product reviews by examining a large Amazon dataset comprising numerous star ratings and comments (Umamageswari et al., 2024). The motivation for this project is that consumers currently use product reviews as a decision-making tool when buying. It represents the quality and dependability of the products. So, this research aims to ensure that ratings and reviews are correlated, not vice versa. The research used three models: Random Forest, Gradient Boosting, and a Hybrid (Random Forest & Gradient Boosting). The accuracy results for Random Forest are 73%, while for Random Forest are 76%. The hybrid model shows 92%. So, it is concluded that a combination of individual classifiers can outperform a single effective classifier.

Sentiment Analysis for social media using SVM Classifier of Machine Learning

This research shows the significance of doing sentiment analysis for businesses and organisations. Support Vector Machines (SVMs) are machine learning techniques used for sentiment analysis (Huang, 2023). The research found that sentiment analysis using SVM has been proven to be a practical approach for analysing social media data in business and organisations. The research is focusing on sentiment analysis of US-Airlines-related tweets. The precision, recall, and F1-score indicate that SVM is a promising approach for sentiment

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue X October 2025

analysis and text polarity classification. With these valuable findings, businesses and organisations can better manage social media sentiment and make more informed decisions. The accuracy for this project is 91.8%. The precision is 91.3%, and the recall is 82.3%. The F1-score is 86.9, and these are the results of analysing US-Airlines related tweets by the SVM algorithm.

Sentiment Analysis of YouTube Movie Trailer Comments using Naïve Bayes

This research analyses viewers' comments and opinions on YouTube about Money Heist, a Netflix TV series (Novendri et al., 2020). There are four seasons of Money Heist. However, many still comment neutrally or give positive feedback about the series. So, sentiment analysis is conducted to classify opinions using Naïve Bayes. It is chosen because a previous study showed satisfactory results. The Naïve Bayes algorithm has been used in text mining because it is simple yet can achieve high accuracy. The result is considered successful because the accuracy rate is 81%. The precision is 74.83%, and the recall is 75.22%.

METHODOLOGY

This phase involved tasks such as Project Framework, Use Case Diagram, and Flow Chart.

Project Framework

Figure 1 illustrates the phases of the project framework for system development. In here, it shows the sequence of each phase along with the activities, as well as the expected outcomes after the development is completed

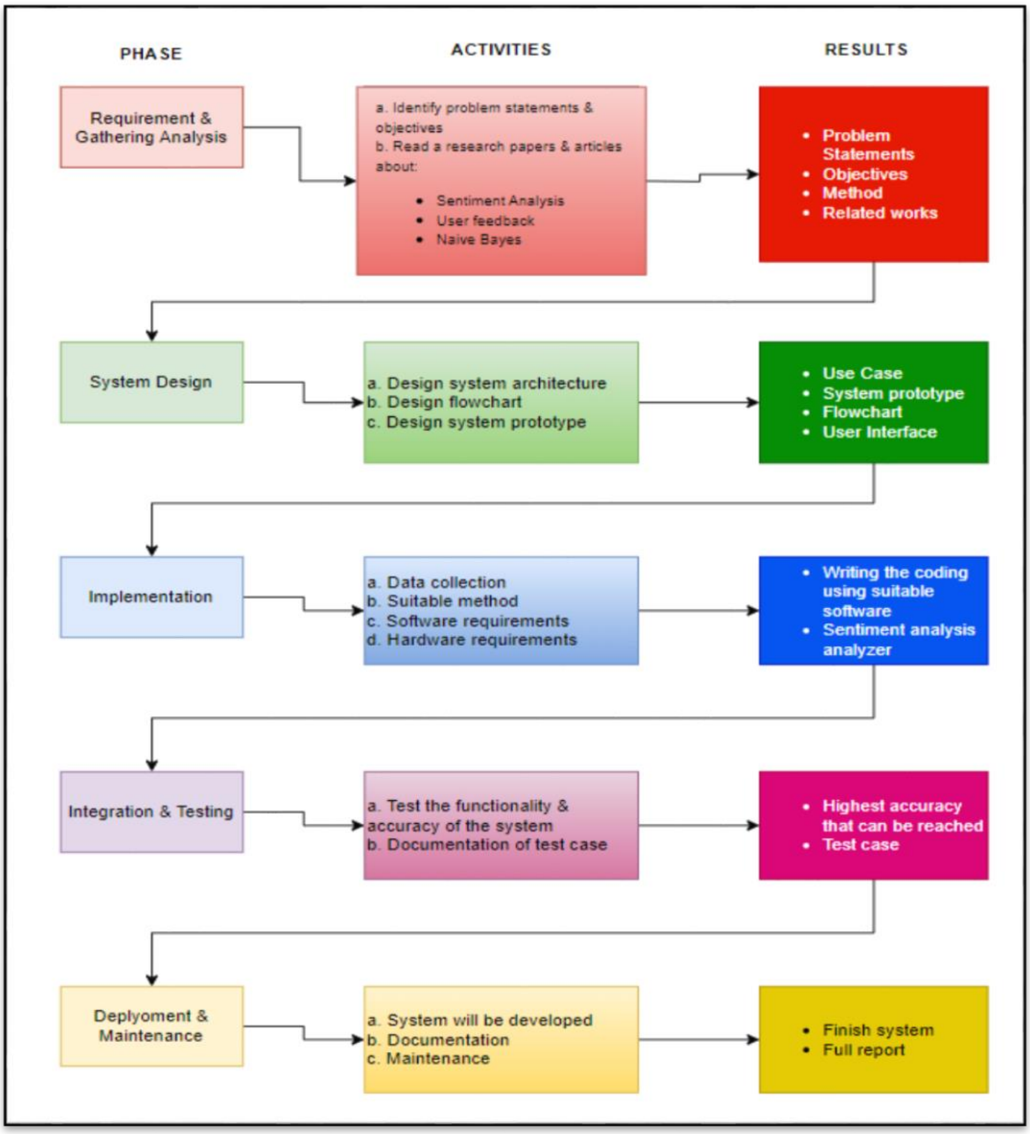

Fig 1. Project Framework

Use Case Diagram

A use case diagram is designed and created to meet the requirements in the design phase of the Modified Waterfall Methodology. A use case can show the interaction between the user and the system. It can also see how many tasks are needed to use the system. Figure 2 shows the use case diagram.

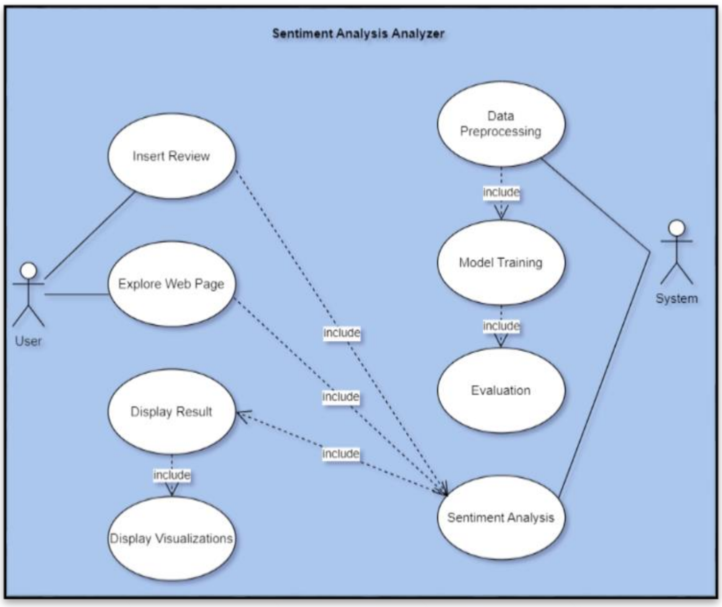

Fig 2. Use Case Diagram

Flowchart

A flowchart is a diagram that shows how data moves through an organisation. It gives a clear picture of the actions taken and the order in which they are carried out within a system. The system flowchart is shown in Figure 3.

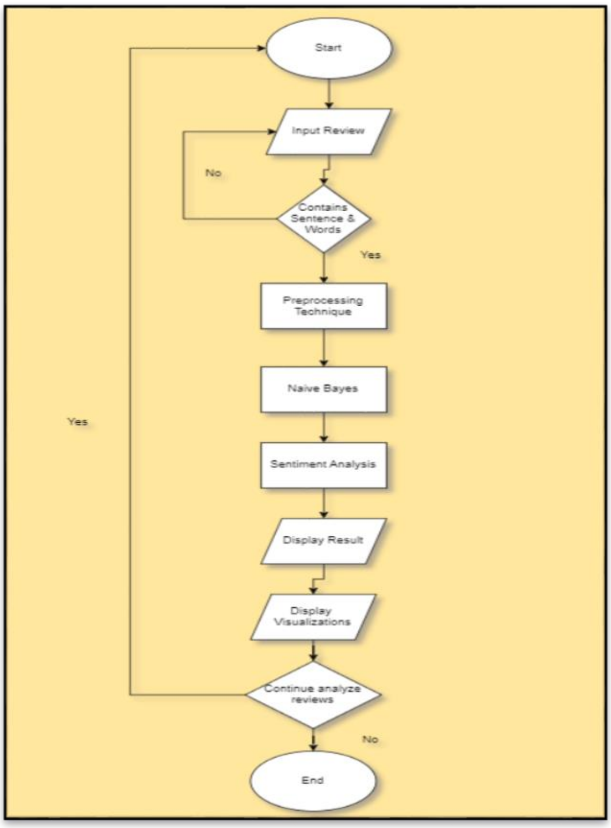

Fig 3. Flowchart

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue X October 2025

RESULTS

In an accuracy test, all the processed data has been split into 20:80, with 20% used for testing and the remaining 80% for training. This ratio is chosen because it yields the highest accuracy score among the other ratios and has already been tested. The technique used is called the Confusion Matrix to measure and enhance the performance of the developed model. Figure 4 shows the result of the Confusion Matrix by percentage.

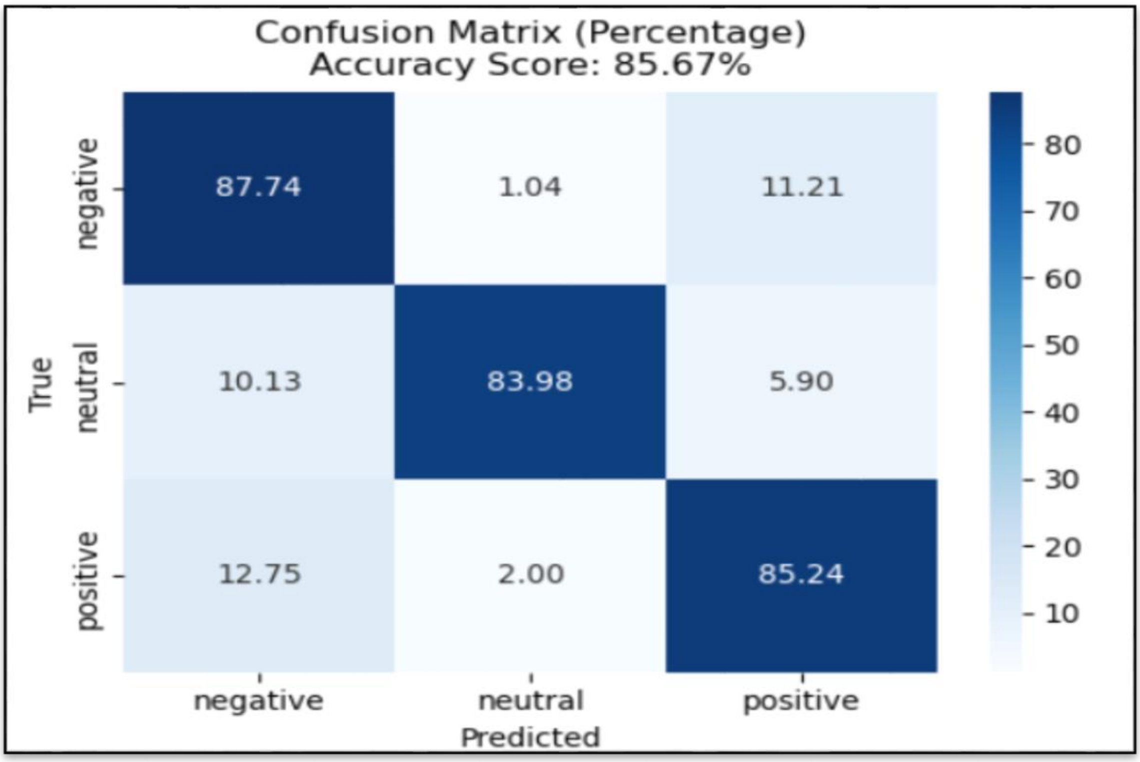

Fig 4. Confusion Matrix in Percentage

In Figure 5.1, the rows of the matrix correspond to accurate labels for positive, negative, and neutral sentiment. The columns for the predicted labels of positive, negative, and neutral. It is presented in Table I in 4 parts: True Positive, True Negative, False Positive, and False Negative.

Table I Classifier Parameters

True / Predicted	Negative	Neutral	Positive
Negative	87.74	1.04	11.21
Neutral	10.13	83.98	5.90
Positive	12.75	2.00	85.24

True Negative (TN)

The true class is Negative, and the classifier correctly predicted Negative with 87.74.

False Positive (FP)

The true class is Negative, but the classifier predicted either Neutral or Positive. It is 1.04 as (Neutral) and 11.21 as (Positive).

False Negative (FN)

The true class is Neutral or Positive, but the classifier predicted as Negative. It is 10.13 (true Neutral) and 12.75 (true Positive).

True Positive (TP)

The true class is either Neutral or Positive, and the classifier correctly predicted either Neutral or Positive. It is 83.98 (actual Neutral) + 5.90 (true Positive) + 85.24 (true Positive).

Figure 5 shows the evaluation results for the Naïve Bayes model.

```
Accuracy: 0.8566989808600547
Confusion Matrix:
 [[4788   57  612]
 [ 541 4486  315]
 [ 675  106 4512]]
Classification Report:
              precision    recall  f1-score   support

    negative       0.80      0.88      0.84      5457
     neutral       0.96      0.84      0.90      5342
    positive       0.83      0.85      0.84      5293

    accuracy                           0.86     16092
   macro avg       0.86      0.86      0.86     16092
weighted avg       0.86      0.86      0.86     16092
```

Fig 5. Naïve Bayes Model Evaluation Results

CONCLUSION

In conclusion, several functions have been recognised during project development. This system utilised the Naïve Bayes model to categorise the reviews into polarities. The analysis results are then displayed in a dashboard using Anvil Editor, an open-source app framework. It is also convenient for designing a simple interface. Several limitations have also been identified during the project's development and testing phases. The project relies on inadequately labelled data, where the abundance of positive reviews relative to negative and neutral ones could affect the classifier's performance. The system focuses only on English reviews because the dataset used in this system development comes from overseas users. So, it might not perform very well with other languages. The project takes around 25 seconds to load the visualisation graph when the user clicks the link because the image is quite large. Therefore, a few recommendations for future improvements are required. Utilising a properly labelled dataset that maintains a balance between positive and negative instances is recommended for this project. When the dataset is balanced, it can improve the classifier's performance. The system could be more useful if it recognised other languages, since the world is a diverse place with many people, races, and languages. The system can be upgraded by loading the graph image when the user clicks, or by loading all when the user clicks the visualisation link. Despite the limitation, it meets all the project objectives: to design a machine learning model and a dashboard visualisation for sentiment analysis of car reviews, to develop a web-based platform for sentiment analysis with data visualisation, and to test the effectiveness of the web-based system.

REFERENCES

1. Adwan, O. Y., Al-Tawil, M., Huneiti, A. M., Shahin, R. A., Abu Zayed, A. A., & Al- Dibsi, R. H. (2020). Twitter sentiment analysis approaches: A survey. International Journal of Emerging Technologies in Learning, 15(15), 79–93.
2. Alamanda, D. T., Ramdhani, A., Kania, I., Susilawati, W., & Hadi, E. S. (2019). Sentiment analysis using text mining of Indonesia Tourism reviews via social media. International Journal of Humanities, Arts and Social Sciences, 5(2), 72– 82.
3. Al-Natour, S., & Turetken, O. (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. International Journal of Information Management.

4. Awais, M., Batool, S., Mirza, A. M., Sajid, A., Khokhar, A. S., & Zafar, A. (2020). Patient's feedback platform for quality of services via "Free Text Analysis" in healthcare industry. EMITTER International Journal of Engineering Technology, 8(2), 316–325.
5. Deshmukh, A., Sonar, S. D. B., Ingole, R. V., Agrawal, R., Dhule, C., & Morris, N. C. (2023). *Satellite image segmentation for forest fire risk detection using Gaussian mixture models*. In *Proceedings of the 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 806–811). IEEE.
6. Huang, Q. (2023). Sentiment analysis for social media using SVM classifier of machine learning. Applied and Computational Engineering, 4(1), 86–90. https://doi.org/10.54254/2755-2721/4/20230354
7. Kim, E., & Chun, S. (2019). Analyzing online car reviews using text mining. Sustainability, 11(6), 1611. https://doi.org/10.3390/su11061611
8. Lamba, Manika & Margam, Madhusudhan. (2022). Sentiment Analysis. 10.1007/978- 3- 030-85085-2_7.
9. Novendri, R., Callista, A. S.., Pratama, D. N., & Puspita, C. E. (2020). Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes. Bulletin of Computer Science and Electrical Engineering, 1(1), 26–32. https://doi.org/10.25008/bcsee.v1i1.5
10. Panchal, D. S., Kawathekar, S. S., & Deshmukh, S. N. (2020). Sentiment analysis of healthcare quality. International Journal of Innovative Technology and Exploring Engineering, 9(3), 3369–3376
11. Singh, S., & Sao, A. (2021). Impact of social media marketing in consumer buying behavior in automobile industry: an empirical study in Delhi. Turkish Online Journal of Qualitative Inquiry, 12(7), 6278–6292. https://tojqi.net/index.php/journal/article/view/4832
12. Umamageswari, A., Pratishwaran, R. J., Reddy, M. P., & Raj, R. Y. S. (2024). Machine learning model for sentiment analysis of Amazon reviews. In Proceedings of the 2024 International Conference on Electronic Systems and Intelligent Computing (ICESIC) (pp. 139–144). IEEE.