PCMJ

Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

CLASSIFICATION AND VISUALISATION OF MALAYSIAN COURIER SERVICES BASED ON X SENTIMENT ANALYSIS USING NAÏVE BAYES

NURUL AIN NAJWA MOHAMAD AMRAN
Universiti Teknologi Mara Cawangan Melaka Kampus Jasin
decaynra02@gmail.com

Article Info

Abstract

Users express their views on various topics through social media, including experiences with courier services. Getting insights from such a huge volume of unstructured textual data is scientific challenging. This paper discusses the perception of the Malaysian courier companies J&T and DHL. Sentiment analysis helps businesses determine customer's opinion and how to improve service quality, but personalized studies on Malaysian logistics are few. This research tackles the challenge of identifying sentiment in X caused by informal language and acronyms. A machine learning techniques Naive Bayes classification model was constructed to tackle this problem with the development process following the Waterfall model. The dataset that was collected underwent tokenization, lemmatization, stop-word removal, and then TF-IDF feature extraction before it was classified into positive neutral and negative sentiments. Performance was improved through hyperparameter tuning on stratified k-fold cross validation training and validation sets. After tuning, the model achieved 86% training accuracy and 78% testing accuracy, hence improving upon the classification performance. The confusion matrix along with precision, recall, and F1-score evaluated the performance of the model. For accuracy in real-time sentiment tracking, support vector machines SVM or KNN could be leveraged. A bigger dataset with aspect-based sentiment might reveal more about service problems. This study contributes to sentiment analysis in Malaysia and demonstrates its applicability in enhancing the logistics customer experience.

Received: March 2025
Accepted: September 2025
Available Online: November 2025

Keywords: Sentiment Analysis, Courier Services, Naïve Bayes, Machine Learning, Malaysia

INTRODUCTION

The rapid advancement of technology has led to the emergence of several social media platforms, enabling individuals to communicate with one another regardless of geographical barriers. X, formerly known as Twitter, is a widely used social media network. According to Karami et al. (2020), X has a total of 1.3 billion accounts, with 336 million active users who together publish 500 million tweets every day as of 2020. X is an essential application in today's day because to its real-time and constantly updated information sharing capabilities.

Given the wide range of issues addressed on X, it is natural that each user will express their own multitude of ideas. Courier services were also discussed. The exponential growth of e-commerce, fueled by the internet, together with the surge in demand for courier services due to online shopping, and the advancements in delivery methods have significantly pushed the development of the courier business (Lin et al., 2020). Hence, the selection of courier services has a major impact for customers, as it enables them to make an informed decision that aligns with their own satisfaction. The purpose of the project is to classifying courier services based on X by using sentiment analysis. As to the study conducted by (Dang & Maria N.Moreno-Garcia, 2022) sentiment analysis refers to the extraction of information about an entity and the automatic identification of any subjective aspects associated with the object being studied. The aim is to determine if the content produced by users expresses their positive, negative, or neutral viewpoints.

Project Significance

The primary objective of this project is to create a web-based system that helps customers of courier services in Malaysia make informed decisions specifically about DHL Express and J&T Express. In addition, this project may also aid in detecting the challenges encountered by customers via their experience with the provided service, enabling the analysis of customer satisfaction and service quality. Furthermore, it may assist courier services in determining their level of excellence and user experience with the service they provide.

LITERATURE REVIEW

Machine Learning Approach

Machine learning, frequently referred to as ML, is branch of Artificial Intelligence (AI) that utilizes previous data to train algorithms and make predictions or perform tasks.

Naïve Bayes

The classification method is referred to as Naïve Bayes, named after Thomas Bayes, a statistician who first formulated Bayes' theorem (Sinaga et al., 2020). The Naïve Bayes algorithm is used for sentiment analysis, specifically to assess whether is a supervised learning method is essential in machine learning for solving classification issues. The Naïve Bayes

Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

algorithm enables quicker model building and prediction generation. The following is Equation (1), which represent the theorem of Naïve Bayes (Sinaga et al., 2020):

$$PP(HH|EE) = \frac{DD(EE|HH)xxxx(HH)}{xx(EE)} \quad \mathbf{1}$$

Naïve Bayes (NB) is a probabilistic classifier that use Bayes' Theorem to estimate the probability of a specific label based on a given collection of features (Wankhade et al., 2022). In Bayesian models, particularly Naïve Bayes, the concept of strong independence across features implies that the presence or absence of one features in a dataset (Sinaga et al., 2020). Plus, it only requires a small amount of training data to estimate the parameters necessary for classification (Dey et al., 2016).

Data Visualisation

Plotly

Plotly is a powerful data visualization library known for creating interactive charts with features like zooming, panning, and data inspection (Jadidoleslam et al., 2020). It supports integration into web pages using Dash and works offline in Jupyter and IPython notebooks. Although its syntax may be challenging for beginners (Stancin & Jovic, 2019), Plotly's rich documentation, interactivity, and versatility make it a valuable tool for data analysis and visualization.

STATISTICAL REPORT TOOL

Bar Chart

A bar chart uses bars to visually compare categories of data with two axes: one for categories and the other for values. The height or length of the bars reflects the data's magnitude, making trends, patterns, and comparisons easily understandable (Parul Gandhi, 2020).

Pie Chart

A pie chart represents data as slices of a circle, with each slice proportional to its value. It is commonly used to compare proportions of components to a whole, providing a simple and clear way to visualize relationships (Parul Gandhi, 2020).

Word Clouds

Word clouds visually represent textual data by highlighting frequently used words, where size or color indicates importance. They are widely used in sentiment analysis for their accessibility, readability, and effectiveness in showcasing prominent terms (Parul Gandhi, 2020).

System Platform

Web

The web, or World Wide Web, enables access to text and multimedia content via internet browsers without requiring software installation. Web-based systems are cost-effective, accessible, and eliminate the need for hardware or software updates, as providers handle maintenance and updates. This reduces administrative burdens and ensures users always have updated features and security, making web-based solutions convenient and efficient for users and organizations.

Programming Language

Python

Sentiment analysis via Naïve Bayes model would work well with Python given its simplicity, ease of use, coupled with extensive library support (Das et al., 2021). Different libraries like TextBlob, Vader, Henry among many others is offered by Python for this purpose. These APIs are designed in such a way that they can be easily understood by any ordinary person thereby making the training process of building Naïve Bayes classifiers faster than one would anticipate. Additionally, Python can easily fit into various data processing applications without any problems.

PCMJ
Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

METHODOLOGY

Flowchart

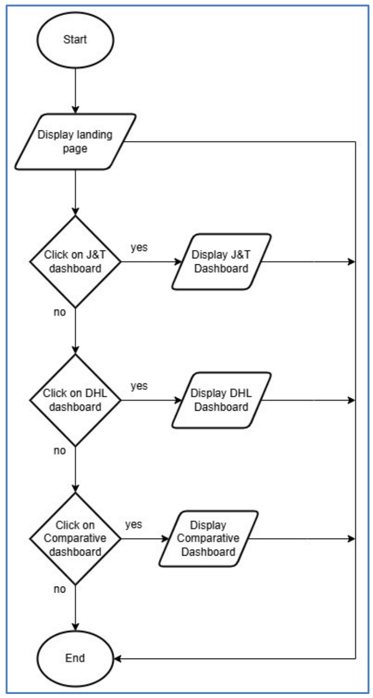


Figure 1 : Flowchart Design

The system flowchart begins from the point wherein a user is directed to the landing page. This page is the first encounter a user has. Upon accessing the landing page, the user encounters a navigation menu located at the top that lists options to access the J&T Express dashboard and DHL Express dashboard.

By clicking the dashboard option under J&T Express from the navigation menu, user will be sent to the J&T Express dashboard where user will be able to visually analyse data related to J&T Express services.

In a similar vein, if user select this option on the newly launched DHL Express dashboard, user will be taken directly to the DHL Express dashboard and will find there visually depicted data concerning DHL Express services.

Lastly, user have option to click on Comparative dashboard to view both DHL and J&T sentiment comparison with stacked and clustered chart.

Diagram Model

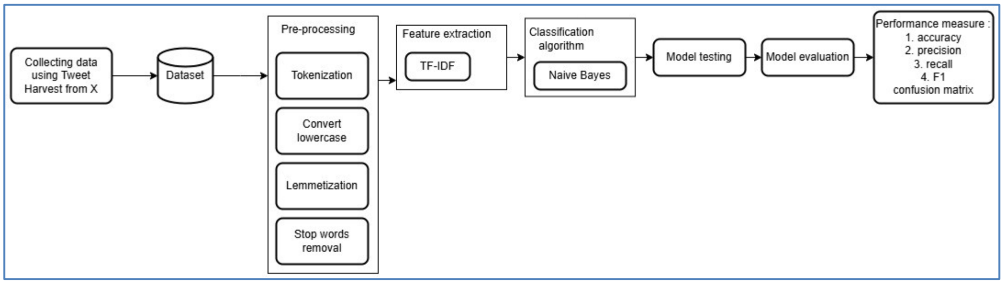


Figure 2 : Diagram Model

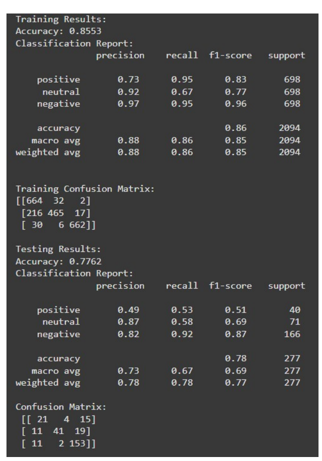
The model diagram illustrates the process of building the sentiment analysis system. Data collection uses Tweet Harvest to extract reviews about J&T Express and DHL Express from X (formerly Twitter). The pre-processing phase involves tokenization, converting text to lowercase, removing stop words, and applying lemmatization to standardize words.

Feature extraction is conducted using Term Frequency-Inverse Document Frequency (TF-IDF) to transform text into numerical data, enabling machine learning models to process it. The Multinomial Naïve Bayes algorithm, a probabilistic classifier suited for text classification, is used to categorize tweets into positive, negative, or neutral sentiments.

The datasets are split into training and testing sets for model learning and evaluation. Model performance is measured using accuracy, precision, recall, F1 score, and a confusion matrix, providing a comprehensive evaluation of classification effectiveness.

RESULT AND DISCUSSION

Accuracy of Model with Parameter Tuning



```
Training Results:
Accuracy: 0.8553
Classification Report:
              precision    recall  f1-score   support

    positive       0.73      0.95      0.83       698
     neutral       0.92      0.67      0.77       698
    negative       0.97      0.95      0.96       698

    accuracy                           0.86      2094
   macro avg       0.88      0.86      0.85      2094
weighted avg       0.88      0.86      0.85      2094


Training Confusion Matrix:
[[664  32   2]
 [216 465  17]
 [ 30   6 662]]

Testing Results:
Accuracy: 0.7762
Classification Report:
              precision    recall  f1-score   support

    positive       0.49      0.53      0.51        40
     neutral       0.87      0.58      0.69        71
    negative       0.82      0.92      0.87       166

    accuracy                           0.78       277
   macro avg       0.73      0.67      0.69       277
weighted avg       0.78      0.78      0.77       277

Confusion Matrix:
 [[ 21   4  15]
 [ 11  41  19]
 [ 11   2 153]]
```

Figure 3 : Result of Training and Testing

The tuned model achieved an accuracy of 85.53% on the training dataset. Positive sentiment recall increased to 95%, effectively capturing most positive reviews, while precision for neutral sentiment rose to 92%, although recall slightly dropped to 67%. The negative sentiment class showed strong performance with 97% precision and 95% recall, making it the most accurately classified category. The training confusion matrix highlights these improvements, with positive and negative sentiments being better distinguished, though challenges remain in differentiating neutral sentiments.

On the testing dataset, the tuned model achieved an accuracy of 77.62%, indicating better generalization for new data. Positive sentiment recall improved from 47% to 53%, though precision remains low at 49%, signaling continued misclassification. Neutral sentiment precision increased to 87%, but recall decreased to 58%, showing improved reliability in predictions but challenges in correctly identifying all neutral samples. The negative sentiment class maintained strong performance with 82% precision and 92% recall.

The testing confusion matrix shows that out of 40 positive samples, 21 were correctly classified, with others misclassified as neutral or negative. For the 71 neutral samples, 19 were misclassified as negative, while the negative class achieved the highest accuracy, with 153 out of 166 samples correctly identified. Overall, the tuned model shows better performance in classifying sentiments, particularly for positive and negative categories.

Data Visualisation of J&T Express

J&T's sentiment analysis reveals that 50.4% of client comments are positive (201 mentions), 42.1% are neutral (168 mentions), and only 7.52% are negative (30 mentions). The pie chart visually emphasizes this distribution, showing most feedback as positive or neutral. The bar chart complements this by using column heights to represent sentiment counts, with positive feedback being the highest, followed by neutral and negative sentiments. Both charts align, highlighting a strong positive perception of J&T among customers, with minimal negative sentiment.

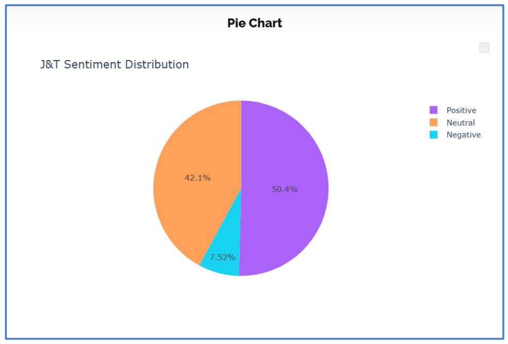


Figure 4 : Pie Chart of J&T

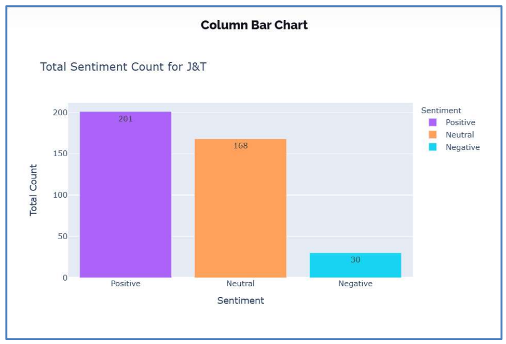


Figure 5 : Bar Chart of J&T

Figure 6 : Word Cloud of J&T

Based on user comments on X, the word clouds provide a graphic illustration of J&T Express's sentiment distribution. These visualisations help in the public's understanding of J&T Express across many sentiment levels.

PCMJ
Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

Data Visualisation of DHL Express

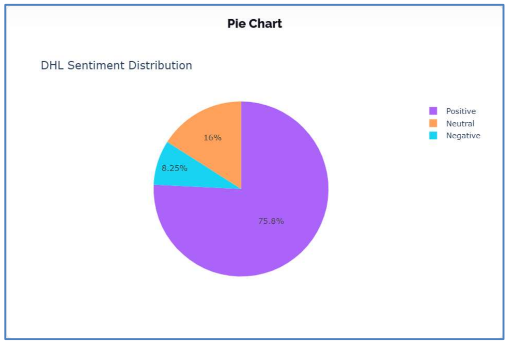


Figure 7 : DHL Pie Chart

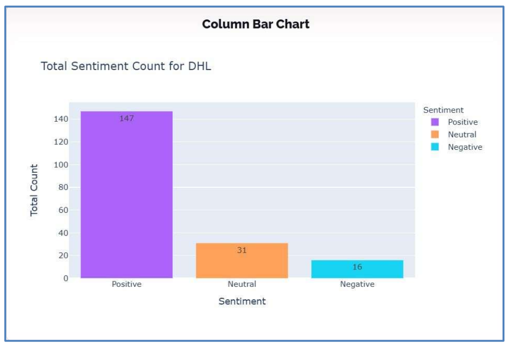


Figure 8 : DHL Bar Chart

DHL Express's sentiment analysis highlights 194 mentions, with 75.8% positive (147 mentions), 16% neutral (31 mentions), and 8.25% negative (18 mentions). The pie chart emphasizes the strong positive sentiment, reflecting high customer satisfaction, while the bar chart visually compares the sentiment distribution, showing positive feedback as the highest, followed by neutral and negative sentiments. Both charts confirm that DHL Express enjoys a solid reputation with minimal negative feedback.

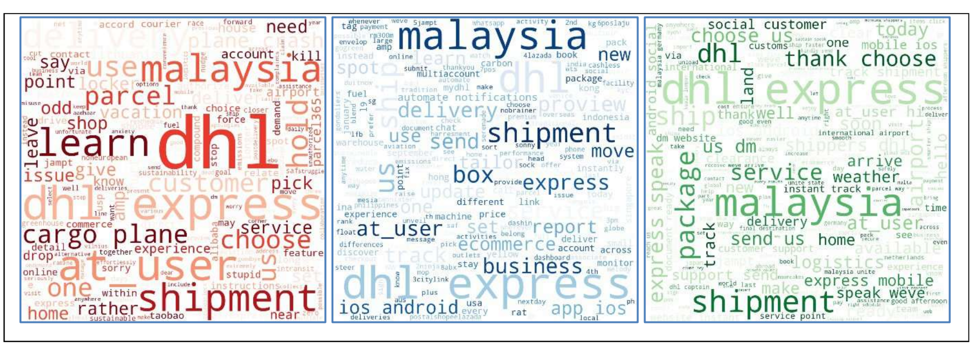


Figure 9 : DHL Word Cloud

PCMJ
Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

DHL word clouds reveal customer perceptions of their services. The red (negative) cloud highlights issues like "delay," "problem," and "complicating," reflecting dissatisfaction with logistics. The blue (neutral) cloud includes terms like "shipment," "business," and "update," indicating general discussions on tracking and ecommerce. The green (positive) cloud showcases satisfaction with words like "thank," "fast," and "support," highlighting DHL's efficiency and reliability.

Data Visualisation of Comparative Dashboard

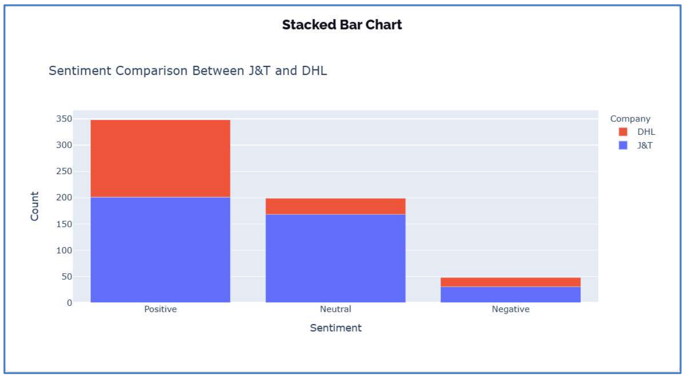


Figure 10 : Comparison Stacked Bar Chart

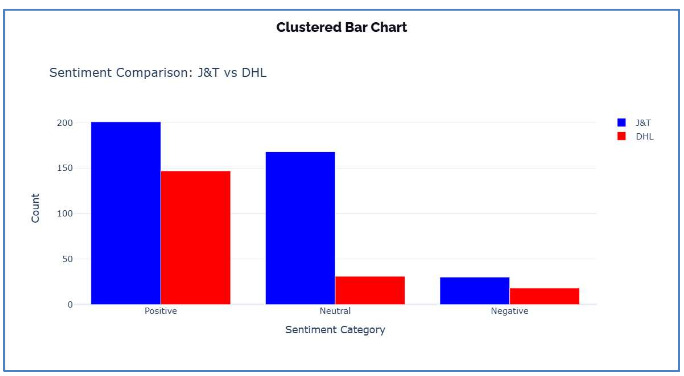


Figure 11 : Comparison Clustered Bar Chart

The stacked bar chart on Figure 10 compares the sentiment distribution for J&T Express and DHL Express. For each courier service, there is one bar divided into three color-coded segments that represent positive, neutral, and negative sentiments. Sentiment trends are easily comparable through this method, showing which company gets more positive feedback and which has more negative mentions. It visually represents the proportional difference in sentiment between the two services. The clustered bar chart on Figure 11 compares the sentiment distribution of J&T Express and DHL Express closely. Unlike stacking the sentiment categories in one bar, this layout groups each type of sentiment (positive, neutral, and negative)

Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

with separate bars for J&T Express and DHL Express. Such an arrangement helps to compare the exact values of sentiments between the two companies and also understand patterns; for instance, if one courier is consistently mentioned more positively or has more complaints than the other.

Functionality Testing

Table 1 : Functionality Testing

Test Case	Expected Result	Success/Failure
View Landing Page	The system begins by showing the landing page.	Success
View J&T Dashboard	The system displays the results of the sentiment analysis on DHL Express as well as visualization of the data.	Success
View DHL Dashboard	The system displays the results of the sentiment analysis on J&T Express as well as visualization of the data.	Success
View Comparative Dashboard	The system will display comparison visual both DHL and J&T data of sentiment.	Success

Discussion

A key challenge with the sentiment analysis model is the imbalance in data distribution between positive, neutral, and negative mentions, particularly for J&T and DHL Express. For example, DHL has 147 positive comments but only 18 negative ones, limiting the model's ability to learn from minority classes like negative sentiments. This imbalance may reduce the model’s accuracy in detecting negative feedback and skew overall customer satisfaction analysis. Misclassification is another issue, with overlapping features causing the model to incorrectly categorize sentiments, as seen in J&T's lower precision and recall for neutral sentiment. Additionally, potential overfitting is evident, with the model achieving 86% accuracy in training but dropping to 78% in testing, indicating difficulty generalizing to new datasets. These issues impact the model’s reliability for real-world applications.

CONCLUSION

This study successfully developed a web-based system for sentiment analysis of J&T Express and DHL Express using Naïve Bayes classification. The first objective was to classify X (formerly Twitter) feedback on courier services in Malaysia. This was achieved through preprocessing (tokenization, stopword filtering, stemming) and sentiment labeling with VADER, followed by training a Multinomial Naïve Bayes (MNB) model with TF-IDF vectorization, achieving 86% accuracy in training and 78% in testing. The second objective was to create a web app for classification and sentiment visualization. This was achieved using Flask and Plotly, providing interactive visualizations like bar charts, pie charts, and word clouds. The system enables effective sentiment analysis and comparison of trends for J&T Express and DHL Express. While the results are satisfactory, future improvements could include expanding the dataset, using advanced algorithms, and integrating real-time visualizations.

REFERENCES

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis Of Review Datasets Using Naïve Bayes‘ And K-NN Classifier. *International Journal Of Information Engineering And Electronic Business*, *8*(4), 54–62. Https://Doi.Org/10.5815/Ijieeb.2016.04.07

Parul Gandhi. (2020). Data Visualization Techniques: Traditional Data To Big Data. *Data Visualization: Trends And Challenges Toward Multidisciplinary Perception*, *June*, 1–179. Https://Doi.Org/10.1007/978-981-15-2282-6

Sinaga, L. M., Sawaluddin, & Suwilo, S. (2020). Analysis Of Classification And Naïve Bayes Algorithm K-Nearest Neighbor In Data Mining. *IOP Conference Series: Materials Science And Engineering*, *725*(1). Https://Doi.Org/10.1088/1757-899X/725/1/012106

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A Survey On Sentiment Analysis Methods, Applications, And Challenges. In *Artificial Intelligence Review* (Vol. 55, Issue 7). Springer Netherlands. Https://Doi.Org/10.1007/S10462-022-10144-1

Jadidoleslam, N., Goska, R., Mantilla, R., & Krajewski, W. F. (2020). Hydrovise: A non-proprietary open-source software for hydrologic model and data visualization and evaluation. *Environmental Modelling and Software*, *134*, 104853. https://doi.org/10.1016/j.envsoft.2020.104853

Stancin, I., & Jovic, A. (2019). An overview and comparison of free Python libraries for data mining and big data analysis. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, 977–982. https://doi.org/10.23919/MIPRO.2019.8757088

Das, N., Gupta, S., Das, S., Yadav, S., Subramanian, T., & Sarkar, N. (2021). A Comparative Study of Sentiment Analysis Tools. *Proceedings of the 2021 IEEE International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2021*, 1–7. https://doi.org/10.1109/ICSES52305.2021.9633905