Mathematical Sciences and Informatics Journal
Vol. 4, No. 2, Nov. 2023, pp. 66-75
http://www.mijuitm.com.my DOI: 10.24191/mij.v4i2.23847

Classification of Google Play Application Using Decision Tree Algorithm on Sentiment Analysis of Text Reviews

Aqil Khairy Hamsani
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Terengganu Branch Kuala Terengganu Campus, Terengganu, Malaysia
2021856342@student.uitm.edu.my

Ummu Fatihah Mohd Bahrin
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Terengganu Branch Kuala Terengganu Campus, Terengganu, Malaysia
ummufatihah@uitm.edu.my

Wan Dorishah Wan Abdul Manan
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Terengganu Branch Kuala Terengganu Campus, Terengganu, Malaysia
wand@staf.uitm.edu.my

Article Info
Article history:

Received May 01, 2023
Revised July 09, 2023
Accepted Aug 22, 2023

Keywords:

Google Play Application
Review
Classification
Sentiment Analysis
Decision Tree

ABSTRACT

The study begins with a comprehensive background, examining the significance of classifying Google Play Store applications using user reviews. The problem statement revolves around the need for an efficient approach to classify application reviews to their sentiments, such as the user's downloaded application not working as intended, and it can take time for users to read every single review. The project's objectives are to study the classifying approach of Google Play Store application reviews using the Decision Tree algorithm, develop a prototype of a classifying application program, and evaluate the accuracy model of the review classification program. To achieve these objectives, the methods employed involve data preprocessing and implementing the Decision Tree (DT) algorithm for classification. The classification model is trained and tested using various split ratios, and the optimal depth for the DT is determined through parameter tuning to achieve the best accuracy. Key results indicate that the developed prototype effectively classifies Google Play Store application reviews with an overall accuracy of 84.88%. This study has successfully achieved its objectives in creating a working Google Play application classification program. The classifier's accuracy and user-friendly interface make it a valuable tool for developers and users.

Corresponding Author:
Ummu Fatihah Mohd Bahrin
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Terengganu Branch Kuala Terengganu Campus, Terengganu, Malaysia
email: ummufatihah@uitm.edu.my

1. Introduction

In this current technological era, the development of mobile applications, in Google Play Store specifically, has risen rapidly. Recent studies show that the number of applications in the Google Play Store is around 3.5 million [1]. This number was enough for users to decide which application was good and which was not. Furthermore, mobile devices nowadays have become a requirement for everyone worldwide.

Before downloading any application, users can investigate the application using many approaches, such as looking through descriptions, photos, videos, and reviews of the application that the Google Play Store provides. It is challenging for Google Play Store to handle that large number of applications since it always has a trade-off between the scrutiny that has been put into checking the application and giving developers a quick time marketplace[2]. Other than reviews, the

PENERBIT PRESS
UNIVERSITI TEKNOLOGI MARA

information was submitted by the developer of the application itself to provide good news for the application. However, text reviews from other users can offer better information based on their application experiences.

Obtaining meaningful insights from text reviews presents several challenges. The sheer volume of reviews for popular apps can be overwhelming, making it difficult for users to get through the information efficiently [3]. Additionally, text reviews often vary in length and quality, with some being brief and lacking specific details, while others are more detailed and provide a richer perspective. This inconsistency can make it challenging to gauge the overall sentiment and extract actionable insights. Additionally, users' preferences and expectations differ, so what may be considered a drawback by one user could be a non-issue for another, making it essential to discern the context of each review.

Therefore, this project will use Sentiment Analysis on the application reviews to determine whether the application is good. The algorithm that is proposed for this project is Decision Tree (DT). DT algorithm is a classifier that can be used to train data and test it to produce the desired outcome. It can be used to classify sentiment analysis on text inputs [4].

The study makes notable contributions to the field of app review analysis by addressing the formidable challenges associated with extracting meaningful insights from a vast number of text reviews with machine learning technique. With the ever-increasing volume of app reviews for popular applications, the study recognizes the potential for information overload, emphasizing the need for more efficient means of data processing.

2. Literature Review

Sentiment is an emotion or attitude brought on by the customer's feelings. As consumer opinions are gathered and mined to determine an app's rating, sentiment analysis is also known as opinion mining. Sentiment analysis is a concept of machine learning [5]. Information is gathered and evaluated to determine whether there is a good or negative sentiment toward the tip. Sentiment analysis is a method of determining whether a selected text is positive, negative, or neutral. Sentiment analysis is used to obtain the polarity of the text from analyzing people's opinions. Sentiment analysis can retrieve various outputs, such as prediction and classification.

Sentiment analysis classification was used in information mining and disclosure. It is applied in different sectors and organizations, such as Amazon and eBay [6]. In variety, sentiment analysis aims to characterize the writer’s thoughts when writing a review or comment. Each word was analyzed to classify the overall emotion of the text as positive or negative. Research on Arabic sentiment analysis in 2022 uses sentiment analysis to classify datasets with sizes ranging from 147 to 5615943 sentences [7]. The classification varied from positive, negative, and neutral.

An application classification program using reviews was aimed to use a framework with different mappings from reviews to subjects of interest and a list of reviews for each issue that indicates how users feel about the issue. The framework was created using several approaches. In research in 2021, classification using the Naïve Bayes algorithm to distinguish between legit and fraudulent apps [8]. Figure 1 shows the framework of the fraudulent detection system used in the previous research. It analyzes data processing and sentimental comments as constructive or negative comments.

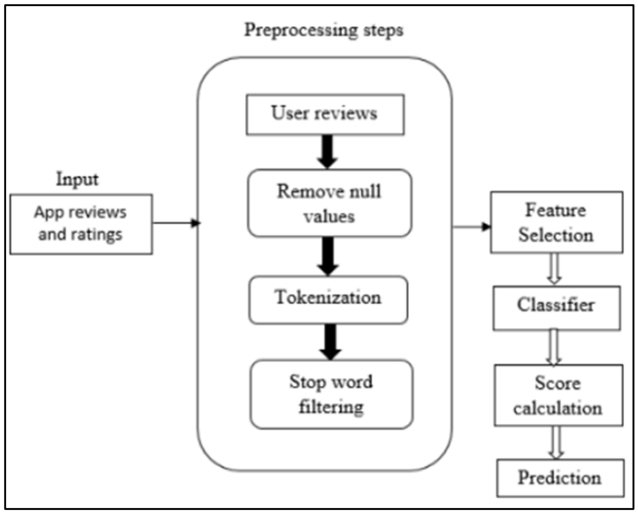


Figure 1. A framework of a fraudulent detection system.

DT Algorithm is a supervised approach that may be used for both classification and regression [9-11]. The DT's objective is to construct a model that predicts the value of a target variable using basic decision rules learned from diverse data sources. The DT may be trained by dividing the reference set into multiple subsets depending on feature value tests. The following technique is done recursively on every derived subset and is known as recursive partitioning. This recursion ends when the subset at a given node has the same value as the target variable or when splitting no longer adds value to the predictions [12].

3. Methodology

There are six (6) phases in the research framework: the preliminary phase, data collection phase, prototype design phase, implementation phase, testing and evaluation phase, and documentation phase. Each step has its tasks and activities. The preliminary stage consists of a literature review, such as reading articles and studying research papers. The data collection phase focused on finding a suitable dataset for the project. In this project, the dataset was reviewed from the Google Play Store. Next, the prototype design phase was to design the actual prototype of the project from the front end, like the user interface, to backend processes, like the classification model. After that comes the implementation phase, where the actual development of the project begins, in this phase, the coding and debugging will produce the desired outcome. Followed by that was the testing and evaluation phase. The project will be tested to meet the requirements of the output. Lastly, the documentation phase took place from the beginning until the end of the project development.

3.1 Training Dataset

The dataset used in the project is a set of application reviews that contain the information. For this project, the dataset used is titled “Google Play Store Apps Reviews” from Kaggle by Mehdi Slim. The dataset contains information such as the reviewer (string), content of the review (string), rating given (float), thumbs up from another reviewer (float) and date of the review (datetime). It has more than 110,000 reviews. The training will randomly take 90% of the total reviews, around 99,000 reviews from the dataset, and the remaining 10%, around 11,000 reviews, will be used for testing. The dataset was not required to be cleaned up because it does not contain the sentiment label for each review. Instead, the sentiment will be labelled automatically using the SentimentIntensityAnalyzer library during the development phase.

3.2 System Architecture

The system architecture consists of user and system parts. The user can send the input and retrieve the result output in the user part. In the system part, the system retrieves the input from the user, which is the application URL. Then, the system can fetch information using the URL the user gave from Google Play in the input. To fetch the data, the URL will be sent to Python’s Google Play

Scraper Library to process and scrape the reviews from the Google Play Store. The review data will then be sent back to the system, pre-processed data, and the classification by transmitting the processed data to the trained model. Finally, the result of the classification will be displayed by the system so that the user can see the final output of the system. Figure 2 shows the system architecture of the system. The system architecture consists of user and system parts.

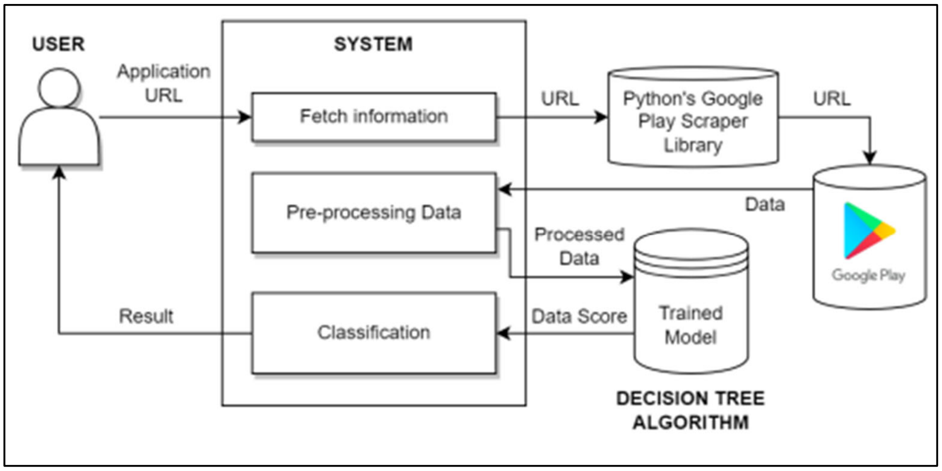


Figure 2. System architecture for the Google Play Application Classification project

3.3 Data Visualization

After the classification for each review of an application is performed, the process will calculate the amount of review for each sentiment category, which is positive, neutral, and negative. Other than in-text results, the result can be displayed on a bar and pie chart using ChartJS. ChartJS is a JavaScript library for visualizing data in various types of charts [13]. The use of ChartJS can help visualize the result for better understanding.

4. Results and Discussion

The results of the project will be discussed in this section. It will highlight how the classification model and the system are evaluated. The evaluation of the project was an important part of seeing how well the task was performed.

4.1 Confusion Matrix

Confusion matrix is a tabular representation of the performance of a classification model [14]. It shows the number of predicted classifications and the actual classification of the testing process. Figure 3 shows the confusion matrix for the trained model.

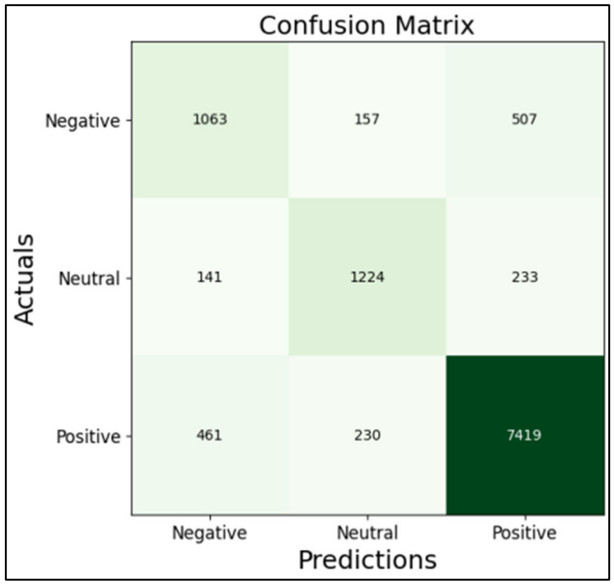


Figure 3. Confusion matrix produced from the classification model.

Evaluation matrices such as precision, recall, f1-score, and accuracy can be calculated based on the confusion matrix [15]. Table 1 shows the summary of the confusion matrix. Precision measures the accuracy of the positive predictions for a given class, recall (or sensitivity) quantifies how many of the true positives were correctly identified, and the F1-Score is a harmonic mean between precision and recall. Finally, accuracy represents the overall correctness of the model's predictions.

Table 1. Summary of the confusion matrix

	Precision	Recall	F1-Score	Accuracy	Data
Negative	0.6382	0.6154	0.6265	0.8947	1727
Neutral	0.7595	0.7665	0.7630	0.9354	1598
Positive	0.9096	0.9143	0.9120	0.8733	8110
				0.8488	11435

Based on experimental results, the DT model in this project has proven to be good and acceptable, with an overall accuracy of 84.88%. This shows that the model can most accurately classify the reviews in the dataset. It also aligns with the project objective to evaluate the accuracy of the Google Play Store application classification program. In this study, it appears that the Positive class has a significantly larger number of data points compared to the other two classes, which could impact the model's performance evaluation and its ability to generalize to different classes. In future work, further analysis to comprehensively address the imbalanced class distribution within the dataset will be a useful study, which could involve exploring advanced techniques for data rebalancing.

4.2. Interface Design

The interface of the project is what will be displayed to the user for input and output. This project contains several sections of the user interface. Figure 4 shows the input section for an application from Google Play Store, where the user can enter the application URL from Google Play Store.

Figure 4. Application URL input section.

The inserted application URL will be processed for the classification to fetch the information from the Google Play Store. Figure 5 shows the result page of application classification. The result page will contain information such as the application name, ID, icon, developer, version, URL, description, and percentage of all reviews classified.

Figure 5. Result page of application classification.

From the result page, the user can click on the ‘Show Scraped Reviews’ button to see the list of reviews of the application with its sentiment. Figure 6 shows the popup appearing when the user clicks the button. The scraped reviews and their sentiments will be displayed on the table. Users can choose options such as exporting to a CSV file, printing the reviews, or displaying the review list on a full page.

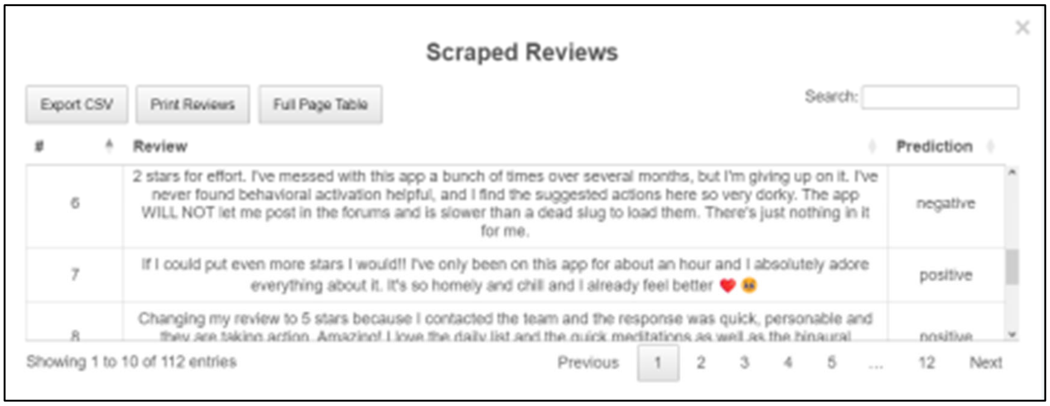


Figure 6. Scraped reviews list.

The last part of the result page is the statistics section. Users can click the 'Show Sentiments Statistic' button to see the statistics of the sentiments. This will pop up the graph in pie and bar charts and a table view of the number of sentiments for the application. Figures 7 and 8 show the sentiment statistics in multiple versions.

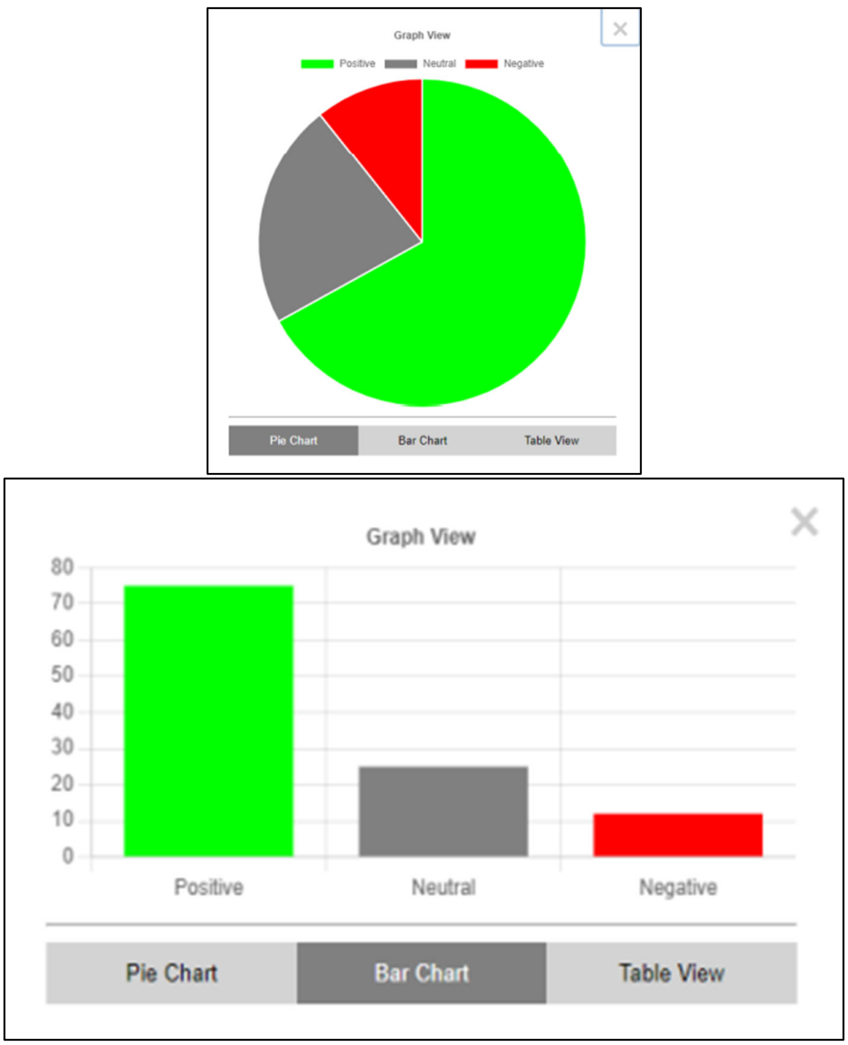
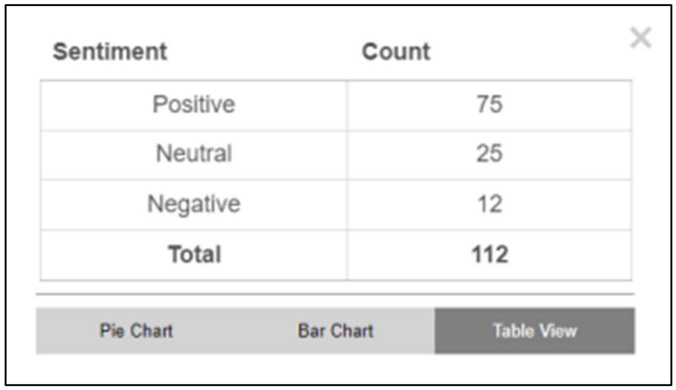


Figure 7. Graph view of sentiment statistics



Sentiment	Count
Positive	75
Neutral	25
Negative	12
Total	**112**

Figure 8. Table view of sentiment statistics

5. Conclusion

This project successfully achieved its main objectives, which were twofold. First, to measure performance of DT machine learning in classifying user sentiment from text review. Second, to develop an application for visualizing the semantic analysis of these text reviews. The study offers valuable contributions to the fields, including the automation of the review classification process for both end-users and application developers. Moreover, it provides invaluable insights into user sentiments, thereby assisting application developers in making informed decisions and aiding users in their choices.

However, it's crucial to underline the project's limitations. This system is presently confined to analyzing reviews exclusively from the Google Play Store, possibly missing out on valuable data from other platforms. Additionally, due to the computational constraints of processing extensive datasets, the application may take longer processing times when handling applications with a substantial volume of reviews.

To enhance the project's function and expand its scope, several recommendations can be proposed. One recommendation is the incorporation of an option to import and analyze reviews from CSV files, to offer more flexibility in data sources. Furthermore, an exciting feature for future development would be to extend the application's compatibility to include reviews from other application stores, thus broadening its classification capabilities and making it a more versatile tool for a wider audience.

Acknowledgments

The authors gratefully acknowledge the Universiti Teknologi MARA (UiTM), Terengganu branch.

Conflict of Interest

The authors declare no conflict of interest in this manuscript's subject matter or materials.

References

[1] N. Alfawzan, M. Christen, G. Spitale, N. Biller-Andorno, and others, “Privacy, data sharing, and data security policies of women’s mhealth apps: scoping review and content analysis,” JMIR mHealth uHealth, vol. 10, no. 5, p. e33735, 2022.

[2] S. Baek, W. Jung, and S. H. Han, “A critical review of text-based research in construction: Data source, analysis method, and implications,” Autom. Constr., vol. 132, p. 103915, 2021.

[3] N. Karunanayake, J. Rajasegaran, A. Gunathillake, S. Seneviratne, and G. Jourjon, "A multi-modal neural embeddings approach for detecting mobile counterfeit apps: A case study on Google Play store," *IEEE Transactions on Mobile Computing,* vol. 21, no. 1, pp. 16-30, 2020.

[4] K. Joshi, S. Kumar, J. Rawat, A. Kumari, A. Gupta, and N. Sharma, "Fraud App Detection of Google Play Store Apps Using Decision Tree," in *2022 2nd International Conference on*

Innovative Practices in Technology and Management (ICIPTM), 2022, vol. 2: IEEE, pp. 243-246.

[5] R. Safrin, K. Sharmila, T. S. Subangi, and E. Vimal, "Sentiment analysis on online product review," *International Research Journal of Engineering and Technology (IRJET),* vol. 4, no. 04, 2017.

[6] S. Choudhary and C. Chhabra, "Sentiment analysis of Amazon food review data," in *2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 2021: IEEE, pp. 116-120.

[7] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic sentiment analysis: A systematic literature review," *Applied Computational Intelligence and Soft Computing,* vol. 2020, pp. 1-21, 2020.

[8] S. A. Moeed, G. Ashmitha, and P. Niranjan, "Application of Machine Learning on Fraud App Detection," *Annals of the Romanian Society for Cell Biology,* vol. 25, no. 6, pp. 15556-15564, 2021.

[9] J. F. Courtney, “Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS,” Decis. Support Syst., vol. 31, no. 1, pp. 17–38, 2001.

[10] B. T. Jijo and A. M. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” Evaluation, vol. 6, p. 7, 2021.

[11] A. Anuar, N. H. Mohd Hussain, and H. Byrd, “Tree-based machine learning in classifying reverse migration,” Math. Sci. Informatics J., vol. 4, no. 1, pp. 49–56, 2023.

[12] H. Kaur and V. Mangat, "A survey of sentiment analysis techniques," in *2017 International conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC)*, 2017: IEEE, pp. 921-925.

[13] H. Da Rocha, Learn Chart. js: Create interactive visualizations for the web with chart. js 2. Packt Publishing Ltd, 2019.

[14] A. Luque, A. Carrasco, A. Martin, and A. de Las Heras, “The impact of class imbalance in classification performance metrics based on the binary confusion matrix,” Pattern Recognit., vol. 91, pp. 216–231, 2019.

[15] J. Xu, Y. Zhang, and D. Miao, “Three-way confusion matrix for classification: A measure driven view,” Inf. Sci. (Ny)., vol. 507, pp. 772–794, 2020.

Biography of all authors

Picture	Biography	Authorship contribution
	**Aqil Khairy Bin Hamsani** is a student of Universiti Teknologi MARA Terengganu Branch, Kuala Terengganu Campus. He is studying for a Bachelor's Degree in Computer Science. Research interests include machine learning and sentiment analysis. He can be connected at 2021856342@student.uitm.edu.my.	Design the research framework, perform data collection, system development and drafting articles.
	Ummu Fatihah Binti Mohd Bahrin is currently serving as a Lecturer at UiTM Kuala Terengganu Campus in the College of Computing, Informatics, and Mathematics. Her research focus centers on sentiment analysis and bio-inspired algorithms, bridging technology and human emotions. Starting her academic journey at UiTM in 2015, she has consistently showcased her dedication to the field. She can be contacted for further correspondence via email at ummufatihah@uitm.edu.my.	Data analysis, result interpretation and article validation.
	Wan Dorishah bt Wan Abdul Manan is a respected academic professional with the title of Associate Professor. She is associated with Universiti Teknologi MARA (UiTM) and is a member of the department of Computing, Informatics, and Mathematics. Wan Dorishah began her service at UiTM on June 16, 1986, and has since made significant contributions to the field of Information, Computer, and Communications Technology (ICT). Wan Dorishah's expertise extends to various areas within the field of ICT and her contributions to the academic community remain noteworthy. She can be reach via email at wand@uitm.edu.my.	Result analysis and article validation.