Journal of Informatics and Web Engineering

Vol. 5 No.1 (February 2026) eISSN: 2821-370X

Hybrid Sentiment Analysis Model for Customer Feedback Interpretation Using Lexicon, Machine Learning and Deep Learning Techniques

Jou Jia Yi[1], Lew Sook Ling[2], Tan Li Tao[3*],

[1]Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.

[2]Centre for Advanced Analytics, CoE for Artificial Intelligence, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.

[3]Faculty of Law, The Sir David Williams Building, 10 West Road, Cambridge CB3 9DZ, United Kingdom

**corresponding author: (ltt28@cam.ac.uk; ORCiD: 0009-0004-7133-3045)*

Abstract - Customer feedback is pivotal in enhancing service quality and user satisfaction across digital platforms. However, traditional sentiment analysis methods often struggle with informal languages, contextual nuances, and aspect-specific opinions. In this paper, a hybrid sentiment analysis framework is proposed, utilizing lexicon-based (Valence Aware Dictionary and sEntiment Reasoner (VADER)), machine learning (Support Vector Machine and Random Forest), and deep learning (BERT) techniques to achieve improved sentiment classification accuracy and interpretability compared to previous studies. The framework incorporates advanced preprocessing techniques, such as emoji normalization, handling of negation, and detection of intensifiers, to better capture emotional information in user-generated content. The objectives of this study are to develop a robust sentiment analysis system that can accurately classify user sentiment and extract aspect-specific insights from customer feedback. Aspect-based Sentiment Analysis (ABSA) was also employed to provide detailed evaluations of specific service components, including driver behaviour, app performance, and pricing. In this study, experimental results using the Uber Customer Reviews Dataset demonstrate that the proposed hybrid model achieves 99% accuracy, significantly outperforms the individual model, and obtains a macro F1-score of 0.98. These findings confirm that integrating lexicon-based, machine learning, and deep learning approaches enhances sentiment classification effectiveness and supports data-driven decision making based on user experience.

Keywords—Aspect-Based Sentiment Analysis, Bidirectional Encoder Representations from Transformers, Customer Feedback, Natural Language Processing, Sentiment Analysis, Support Vector Machines, Uber, Valence Aware Dictionary and sEntiment Reasoner

Received: 17 May 2025; Accepted: 7 August 2025; Published: 16 February 2026

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.1.4

Published by MMU Press. URL: journals.mmupress.com/jiwe

1. INTRODUCTION

In today's data-driven service-oriented economy, customer feedback is extremely valuable to organizations that want to improve service quality, user satisfaction, and organizational efficiency. There are a multitude of available user opinions about Internet services in the form of online reviews, comments on app stores, and postings on social media. When these feedback, sources are analysed appropriately, they can yield insights that can be acted upon [1], [2]. However, traditional feedback analysis methods are often manual, time-consuming, and prone to subjective bias, rendering them unsuitable for large-scale or real-time applications [3].

Hybrid sentiment analysis is a multipronged approach that incorporates lexicon-based methods (VADER), machine learning models (Support Vector Machine and Random Forest) [4], and deep learning (BERT), leveraging the advantages of both machine learning and human effort. In addition, when producing the output, the Aspect-based Sentiment Analysis (ABSA) is utilized to can uncover sentiment about different components of service, such as driver behaviours, value of app(s), and pricing [5], [6]. To address these limitations, a hybrid sentiment analysis framework with a mix of methodologies is recommended to improve sentiment analysis in terms of veracity, interpretability, and granularity.

The proposed framework was evaluated using the Uber Customer Reviews Dataset 2024, which contains over 4,900 user-generated reviews. Accuracy, macro F1-score, and interpretability [7] were used to evaluate the models. The evaluations indicated that the hybrid model achieved better evaluation results than either approach alone. This study provides a more effective and pragmatic approach for interpreting customer feedback [8]. This study adds knowledge for building intelligent systems capable of supporting data-driven decision-making in customer experience management.

Previous studies have explored hybrid sentiment analysis approaches that combine lexicon-based techniques with machine learning or deep learning models to enhance performance and interpretability. Some techniques use graph structures to better learn complex sentence structures, whereas others use multi-task learning to tackle aspect-level sentiment detection. These techniques fuse both contextual and structural information, but they often rely on large amounts of labelled training data and considerable computational power.

Nevertheless, many existing hybrid models do not address the challenge of informal language normalization or target customer feedback in ride-hailing services where sentiment expressions are domain-related and unstructured. To address these gaps, the proposed framework leverages VADER sentiment scoring, Random Forest classification, and contextual embeddings produced from Bidirectional Encoder Representations from Transformers (BERT). This framework also incorporates unique preprocessing steps, such as emoji translation, negation, and intensifier consideration, to improve the robustness and accuracy when interpreting informal customer reviews.

2. LITERATURE REVIEW

Sentiment analysis has emerged as a fundamental methodology for understanding the user opinions expressed in online media. The development of Natural Language Processing (NLP) and artificial intelligence has fostered the creation and use of more varied approaches to automatically classify sentiment. These approaches generally fall into lexicon-based categories, which are machine learning-based categories.

2.1 Lexicon-Based Sentiment Analysis

Lexicon-Based Sentiment Analysis (LBSA) is performed using dictionaries of sentiments to determine the sentiment scores for words or phrases. Valence Aware Dictionary and sEntiment Reasoner (VADER) is a popular LBSA engine suitable for assessing short informal texts [9]. While lexicon-based approaches are intuitive and easy to compute, sarcasm, multivalued dimensions, negations, and terminology specific to domains can be troublesome [10]. To mitigate the limitations of lexicon-based approaches, a hybrid model combining lexicon-based scores with machine learning classifiers has been suggested. This study integrated VADER into a hybrid model with machine learning and deep learning models to enhance the sentiment classification efficiency of informal user-generated reviews.

In this study, the compound score and polarity probabilities with the VADER algorithm were incorporated as input features for machine learning classifiers such as Support Vector Machine (SVM) and Random Forest (RF). The compound score value was determined using the formula for normalization, as shown in Equation (1).

$$\boldsymbol{Compound} = \frac{\sum_{i=1}^{n} s_i}{\sqrt{\sum_{i=1}^{n} s_i^2 + \alpha}}\ , \boldsymbol{where}\ \alpha = 15 \tag{1}$$

where s_i represents the value of each token obtained from the VADER sentiment lexicon. These quantitative specifications, involving compound and polarity probabilities and other measures based on language, such as negation or intensifiers, were combined into a coherent feature vector and treated as an input feature for the machine learning classifiers. The hybrid approach improved the model's ability to detect more nuanced meanings of sentiments in informal user generated content.

To overcome the drawbacks of lexicon-based models, a hybrid model that uses lexicon-based scores and a machine-learning-based classifier has been proposed. For example, Muhammad et al. showed that a method that combined lexicon-based scores with SVMs and topic modelling (Latent Dirichlet Allocation (LDA)) effectively captured sentiment in informal user-generated content on Twitter [11].

2.2 Machine Learning-Based Sentiment Analysis

Sentiment Analysis with Machine Learning uses algorithms, including SVM and RF, to learn sentiment rules from labelled data. SVM is a good candidate for this type of high-dimensional feature space, and RF is less likely to overfit and provide better results in cases of unbalanced datasets. [12]. However, both models rely on feature engineering, and if the text being analysed is informal or noisy, models such as these may underperform. In this study, SVM and RF were employed alongside advanced preprocessing techniques to improve sentiment classification performance in informal Uber customer reviews.

The SVM model aims to construct a hyperplane that maximizes the margin between sentiment classes. The decision function is defined as Equation (2).

$$f(x) = w^T x + b \tag{2}$$

where x is the input feature vector, w is the weight vector, and b is bias. The model seeks to minimize the norm of the weight vector while ensuring correct classification and is formulated as Equation (3).

$$min_{w,b}\ \tfrac{1}{2}||w||^2\ \textit{subject to}\ y_i\ (w^T x_i\ +\text{b}) \geq 1 \tag{3}$$

where $y_i \in \{-1,1\}$ is the sentiment label of training sample x_i . This formulation ensures maximum separation between classes and enhances generalization.

The RF model, on the other hand, is an ensemble method consisting of multiple decision trees, each trained on random subsets of both data and features. The final prediction is obtained by majority voting across trees using Equation (4).

$$\hat{y} = mode(\{T_1(x), T_2(x), \dots, T_n(x)\}) \tag{4}$$

where is $T_i(x)$ is the prediction made using the *i-th* decision tree. This approach reduces variance and is particularly effective in handling heterogeneous and noisy data such as user-generated reviews.

2.3 Deep Learning-Based Sentiment Analysis

Deep-learning-based sentiment analysis, particularly with BERT, has significantly improved the ability to identify contextual semantics in text. Although BERT has been proven to perform the best in various sentiment classification tasks, it requires considerable computational capacity and fine-tuning to domain-based data [13]. BERT has recently proposed enhancements, such as auxiliary sentence approaches, hierarchical BERT, and domain-based fine-tuning, to work more effectively in ABSA tasks. In this study, BERT was fine-tuned on Uber customer review data to capture contextual and aspect-specific sentiments more accurately as part of the overall hybrid sentiment analysis framework.

In this study, BERT was fine-tuned on Uber customer review data to capture contextual and aspect-specific sentiments more accurately as part of the overall hybrid sentiment analysis framework. The fine-tuning process involves using the final hidden state of the special classification token [CLS] as input to a softmax classifier, which predicts the sentiment label. The classification is computed as Equation (5).

$$\hat{y} = softmax(W \cdot h_{[CLS]} + b) \tag{5}$$

where $h_{[CLS]}$ represents the contextualized embedding from the final BERT layer corresponding to the [CLS] token, W is the weight matrix, b is the bias term, and $\hat{y}$ denotes the predicted probability distribution over sentiment classes (e.g., positive, neutral, and negative). This formulation enabled BERT to effectively map deep semantic representations into discrete sentiment categories.

2.4 Aspect-based Sentiment Analysis (ABSA)

ABSA allows for fine-grained sentiment classification by determining the sentiment polarity towards specific aspects mentioned within a review, such as "driver behaviour" or "app performance," while there is more actionable data available for improving services, ABSA is more complex and needs to have accurate aspect term extraction [14]. Deep learning models with attention mechanisms and bidirectional encoders have demonstrated improved ABSA performance using ABSA methods, machine learning methods, deep learning models, and ABSA [15]. Experimenting with ABSA on the Uber review dataset will allow for the extraction of sentiments about service aspects and will produce these insights at a higher level to allow for actionable service improvement. Prior research has shown that LSTM-based ABSA models can achieve high accuracy when applied to informal text data such as tweets about extreme weather events [16].

The model has two stages for implementing ABSA in this case study. The first stage is aspect term extraction, and the second stage is sentiment polarity classification. Aspect term extraction is completed using a rule-based matching method applying an aspect lexicon. In the aspect lexicon, there are keywords, including "driver," "price" or "waiting time."

For sentiment classification, BERT is fine-tuned using a sentence-pair input format, where each input consists of an aspect term and the original sentence, structured as Equation (6).

$$\text{[CLS] Aspect Term [SEP] Sentence [SEP]} \tag{6}$$

The output embedding of the [CLS] token is passed to a softmax classifier to predict the sentiment polarity corresponding to a given aspect. The classification is computed as Equation (7).

$$= softmax(W \cdot h_{[CLS]} + b) \tag{7}$$

where $h_{[CLS]}$ is the contextual embedding generated by BERT, W and b are trainable parameters, and ŷ denotes the predicted sentiment label (e.g., positive, neutral, negative) with respect to a specific aspect. Although the mathematical classification function is identical to the standard sentence-level sentiment classification, the difference lies in the input structure and the learning objective. ABSA guides the model to focus specifically on the sentiment associated with a given aspect term, enabling fine-grained sentiment analysis rather than general sentence-level prediction.

2.5 Recent Advances in Transformer-Based Sentiment Analysis

Transformer-based models have continued to advance sentiment analysis tasks. Among them, RoBERTa and DistilBERT are two improved variants of BERT that offer improved speed, accuracy, and efficiency.

DistilBERT is a compressed version of BERT that retains most of its performance, while being faster and lighter. A study reported that Distil BERT has good performance on short texts, such as customer reviews, and is ideal for real-time applications [17].

RoBERTa is an improved method of BERT, by removing the "next sentence prediction" task, and training BERT on more data. A study in 2025 discussed how RoBERTa achieved high accuracy in classifying social media conversations about the UK's Central Bank Digital Currency (CBDC) within a few rounds of fine-tuning [18].

Researchers have also examined hybrid approaches that combine lexicon-based with deep learning techniques. In a paper published in 2025, we saw for instance, we saw how researchers integrated VADER sentiment scores with contextual embedding (Distil BERT) to improve real-time sentiment analysis of tweets, providing both interpretability and contextual information. [19].

An alternative method is RoBERTa-BiLSTM, which combines embeddings made using RoBERTa with a BiLSTM network and an attention mechanism. This method works particularly well on unstructured data such as social media and movie reviews because it can encapsulate long-term dependencies while disentangling influential sentiment phrases [20].

In a study published in 2024, BERT, Robertas, DistilBERT, and GPT-2 were evaluated simultaneously. The study demonstrated that RoBERTa had the best overall accuracy, but Distil BERT worked more efficiently in terms of computational resources and took better advantage of specific tasks [21].

3. RESEARCH METHODOLOGY

This section describes the methodology used to study customer feedback, using topics from sentiment analysis. First, customer review data were collected and pre-processed, and features were extracted to represent the input textual data in numerical form. Subsequently, the lexical method (VADER) [22], machine learning methods (SVM and RF), and deep learning (BERT) based on the Business Context Sentiment Analysis approach were applied. Part-of-speech tagging and ABSA permit the measure to be at the aspect level (e.g., driver and service). The overall model performance was compared and evaluated against standard evaluation metrics to identify effective models for the business domain. In this study, this multi-model approach is implemented on the Uber customer review dataset to identify the most suitable method for extracting actionable sentiment insights in the ride-sharing context.

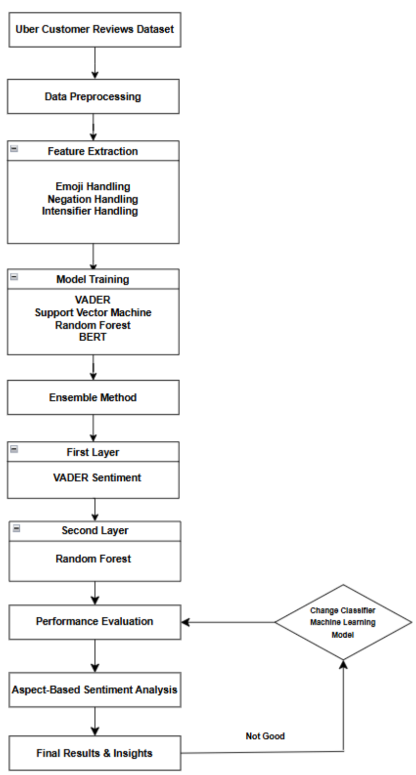

Figure 1. Proposed Framework

The overall framework, as shown in Figure 1, begins with data preprocessing including emoji normalization, negation handling, and intensifier detection. Features were then extracted using both lexicon-based (VADER) and semantic embeddings (Word2Vec, BERT). Multiple models—VADER, SVM, RF, and BERT— were trained, and a two-layer ensemble was applied: the first layer used VADER sentiment, while the second layer used RF for final prediction. If the performance was unsatisfactory, the classifiers were tuned. In the final stage, ABSA is applied to extract aspect-level sentiment insights.

To provide clearer insight into the operational flow of the framework, the following steps outline the complete pipeline.

Step 1: Uber Review Dataset
The process begins by collecting user comments from Uber's 2024 customer review dataset. These reviews reflect the real feedback from ride-sharing users.

Step 2: Data Preprocessing
First, the raw text was cleaned. This involves removing special symbols, correcting spelling errors, and standardizing the format. Common preprocessing tasks include tokenization (splitting sentences into words), removing stop words (like “the” or “is”), and lemmatization (changing words to their base forms).

Step 3: Feature Extraction
After cleaning, key features are extracted. These include:

- Emoji Interpretation: Emojis are converted into corresponding emotional meanings.
- Negation Handling: Words like “not” are handled carefully to avoid misjudging the sentiment.
- Intensity Adjustment: Words such as “very” or “slightly” are used to modify sentiment strength.

Step 4: Model Training
Several models are trained using the extracted features.

- VADER, a rule-based model, gives sentiment scores directly from words.
- SVM and RF, two machine learning models, learn patterns from labelled data.
- BERT, a deep learning model, was fine-tuned to better understand the context of reviews.

Step 5: Ensemble Approach
The study uses a two-level approach to improve the results.

- In the first layer, VADER is used to get the initial sentiment score.
- In the second layer, RF uses the VADER result along with other features to predict the final sentiment.

Step 6: Model Evaluation
Each model’s accuracy is measured using standard metrics. If the performance is poor, the classifier in the second layer is replaced or retrained, as indicated by the loop in the diagram.

Step 7: ABSA
To obtain more detailed insights, the study performs ABSA. It focuses on specific parts of the service like “driver”, “waiting time,” or “app function”. The keywords were first identified using rule-based methods. Then, BERT was used to classify the sentiment related to each keyword.

Step 8: Result Generation
After ABSA, the system produces clear sentiment outputs related to different service aspects. These results help us to better understand customer opinions.

Following preprocessing, the data were passed through four classifiers: VADER (a lexicon-based tool), SVM, RF, and BERT (a transformer-based deep learning model). These classifiers are integrated into a two-layer ensemble structure. In the first layer, VADER provides a sentiment baseline that is then refined by the RF classifier in the second layer. This layered ensemble allows improved accuracy and interpretability. If the performance is found to be unsatisfactory, the second-layer machine-learning model can be replaced or retrained, as indicated in the feedback loop of the framework.

To extract more specific insights, ABSA is applied, enabling the identification of sentiments related to targeted service aspects, such as driver behaviour, pricing, and app performance. This addresses the need for fine-grained sentiment understanding, which is crucial for operational decision making.

The results clearly demonstrated the effectiveness of the proposed method. The hybrid model achieved an accuracy of 99% and a macro F1-score of 0.98, significantly outperforming the individual classifiers. These results fulfil the first research objective of improving the classification accuracy. The successful incorporation of advanced preprocessing directly supports the second objective of enhancing the model's ability to detect nuanced emotional expressions. Finally, the application of the ABSA validates the third objective, enabling detailed aspect-level insights that are not possible with traditional methods. Thus, the framework not only meets but also exceeds the expectations set by the initial problem statement and the research goals.

3.1 Dataset Description

This study is based on the Uber Customer Reviews Dataset (2024) retrieved from Kaggle 1. The dataset contains over 12,000 text reviews left by users, which included the following information for each review: text review/text feedback, star ratings (1–5), when the review was submitted (timestamp), what version the app was on at the time (app version), and possibly a reply from the developer. Therefore, these reviews encapsulate actual experiences of customers using Uber to access the ride-hailing service; hence, the customer review dataset is well-suited for sentiment analysis and service quality evaluation.

Table 1 summarizes the key attributes derived from the Uber customer review dataset that are essential for sentiment classification and model evaluation.

Table 1. Key Attributes of Uber Customer Reviews Dataset

Attribute	Macro F1-Score
content	The main body of the user's review text
score	A numerical rating from 1 to 5
thumbsUpCount	Number of likes received by the review.
appVersion	The version of the Uber app used at the time of review.
replyContent	Developer's response to the review.

To maintain the quality and consistency of our data, we performed extensive preprocessing on the Uber customer review dataset. This involved multiple preprocessing steps, the first of which was normalizing the text. Normalization included tokenization, stop word removal, and lemmatization, all of which were performed using the Natural Language Toolkit (NLTK) [23]. Emojis were also normalized to preserve emotional cues, and informal phrases were transformed into descriptions. Linguistic modifiers were also recognized as negation or intensifiers and negotiations were adjusted to affect sentiment scores accordingly. In this study, rows with null values or unwanted entries were removed during the preprocessing stage to create a clean and reliable dataset.

In addition to the basic cleaning process, the dataset was augmented to include sentiment features provided by VADER (a lexicon-based sentiment analyser) and semantic embeddings created using Word2Vec and BERT. Sentiment features allow multiple sentiment classification models to be trained and tested, utilizing lexicon-based, machine learning, and deep learning approaches that take advantage of these additional features to provide an ABSA to assess the overall sentiment and the sentiment towards service aspects in the reviews provided by users (driver behaviour, app performance, pricing) [24]. In this study, the added features were used to improve model learning and enable comprehensive sentiment analysis on the dataset of Uber customer reviews.

This comprehensive cleaning and feature engineering pipeline provides an appropriate framework for the hybrid sentiment analysis framework suggested in this research that would allow effective model training, reliable evaluation, and a working approach to sentiment classification for customer experience management [25]. In this study, this framework supported the implementation of integrating the VADER model, SVM, RF, and BERT to analyse Uber customer feedback with accuracy and reliability.

3.2 Data Preprocessing

A data preprocessing pipeline was implemented to enhance the integrity of the sentiment analysis text data. The traffic generated from raw Uber customer reviews requires several stages of cleaning and transformation to make it suitable for extracting features and training models.

All reviews underwent cleaning to remove duplicates, null handling, and to ensure that irrelevant or empty information was filtered out. Attributes such as user image, reply content, and repliedAt were dropped because of their high null rates, whereas the key fields of content and score were preserved. In this study, the cleaned dataset served as the basis for training sentiment analysis models to evaluate customer feedback in the Uber review dataset.

Next, tokenization was conducted using NLTK [26], which breaks each review into word tokens. Following that, stop words were removed using a stop word list which eliminated certain common but semantically weak words (e.g., “is”, “the”, “and”) to reduce noise. The tokens were then cleaned and lemmatized using WordNetLemmatizer, which changed the words to their base form (ex. “running” → “run”), thus ensuring each word would be in the same representation.

Informal text normalization was applied to further enhance sentiment detection. This included:

- Emoji conversion: Emojis were translated into textual descriptions (e.g., emoji → “smiling face”) using the emoji Python package.
- Negation handling: Sentiment scores were adjusted when negation terms (e.g., “not”, “never”) were detected in the sentence.
- Intensifier detection: Words such as “very” or “extremely” were identified and used to amplify the sentiment scores accordingly.

Finally, the pre-processed text was saved in new columns (content no stop words and content lemmatized) to create a clean and structured dataset. This forms the basis of the dataset used for feature extraction and model training in the future. In this study, these pre-processed columns were used to generate sentiment scores and embeddings that supported the accurate sentiment classification of Uber customer reviews.

3.3 Feature Extraction

To conduct meaningful sentiment classification, this study applies several feature extraction methods to transform unstructured text data into structured numerical representations suitable for model training.

First, lexicon-based sentiment scoring was performed using the Valence Aware Dictionary and Sentiment Reasoner (VADER). VADER provides sentiment scores across four categories: positive, negative, neutral, and compounded. These scores were incorporated into both the individual model evaluation and the ensemble hybrid sentiment classification framework.

In addition to lexicon scores, semantic representations were captured using two word-embedding techniques: Word2Vec and BERT. Word2Vec embeddings (pre-trained on the Google News corpus) provided dense 300-dimensional vectors that captured the syntactic and semantic relationships between words. These embeddings were used to enrich traditional machine-learning classifiers (SVM and RF).

The BERT embeddings were used to extract deeper contextual meanings. The BERT model was fine-tuned on the Uber customer review dataset to better target domain-specific expressions and sentiment context.

In addition, linguistic features extracted during data preprocessing, including emoji normalization, negation handling, and intensifier detection (Section 3.2), were added to the final feature set. These informal features provide a more subtle understanding of customer sentiments, especially within short, expressive, and user-generated reviews.

A rich and comprehensive feature set was constructed by combining lexicon-based scores, semantic embeddings, and informal linguistic indicators. This enabled sentiment analysis models to achieve improved accuracy, robustness, and interpretability when classifying sentiment and extracting aspect-specific insights from Uber customer reviews.

3.4 Sentiment Analysis Models

This study examined and assessed three types of sentiment analysis models: lexicon-based, machine-learning-based, and deep-learning-based. All types of models were trained and tested using the pre-processed Uber customer review dataset and tested to determine how well sentiment was classified into negative, neutral, and positive categories.

The lexicon-based model used was VADER (Valence Aware Dictionary and Sentiment Reasoner), intended for social media and short informal text. VADER determines polarity scores based on its sentiment lexicon, which accounts for the rule-based modification of scores (punctuation, capitalization, negation, and intensifiers) [27]. The compound score from VADER was used to classify reviews into sentiment categories with a threshold of ±0.05. To capture this, additional modifications were made to capture the negation and intensifier terms that were indicated in the preprocessing step [28]. In this study, VADER rule-based scoring was enhanced with preprocessing techniques to better handle informal expressions in Uber customer reviews, contributing to more accurate sentiment categorization.

Two machine-learning-based models were implemented as classifiers: SVM and RF. The SVM model created a hyperplane from the TF-IDF vectors and Word2Vec embeddings that maximized the margin between the sentiment classes. In general, the SVM model demonstrated high accuracy in binary classification, particularly for clearly polarized sentiments. The RF model used an ensemble of decision trees to improve robustness and reduce overfitting. RF was effective in differentiating between polarity categories and provided feature importance scores that helped with interpretability regarding which text features influenced sentiment [29]. In this study, both SVM and RF were trained on Uber customer review features to evaluate their performance in classifying sentiments and identifying influential linguistic patterns in user feedback.

The deep learning-based model used in this study was BERT. Using the hugging-face transformers library, a pre-trained BERT model was fine-tuned on the Uber dataset. BERT was designed to perform sequence classification and was trained with labelled sentiment in mind. BERT is so named for its ability to leverage context better than what it did receive some positive performance; it still generated a limited amount of performance because of class imbalance, as well as its relative lack of training data with respect to context, and provides an understanding of the limitations and concerns when applying deep learning models to domain-specific textual sentiment classification. In this study, BERT was utilized to capture complex contextual meanings in Uber customer reviews, offering deeper sentiment insights than traditional approaches.

The models utilized standard evaluation measures: accuracy, precision, recall, and F1-score. The findings showed that VADER provided a high level of interpretability and efficiency, machine learning models achieved a better balance among classes, and BERT demonstrated reasonable potential for depth of context with appropriate tuning. Altogether, these models provide an initial structure to be utilized in the hybrid system proposed later. In this study, these evaluation metrics were applied to assess the performance of each model on the Uber review dataset to guide the selection of the most effective sentiment classification approach.

3.5 ABSA

In this study, ABSA was applied to granular sentiment classification based on the sentiment polarity of distinct service components referred to in user reviews. While traditional sentiment analysis assigns a single sentiment label to the review, ABSA divides and extracts sentiment associated with the distinct aspects, like “driver behaviours”, “app performance”, “prices”, etc. [30]. In this study, ABSA was used to identify and analyse sentiments towards key service aspects in Uber reviews, enabling more targeted insights for service improvement.

The ABSA process commenced with aspect-term extraction, which was performed through keyword-based matching as well as syntactic analysis. Reviews were first pre-processed through tokenization and lemmatization prior to the analysis of noun phrases and dependency relations using the spaCy NLP library, allowing relevant terms (normally nouns or noun phrases) to be identified that typically represent service features. In this paper, this method was applied to extract key aspects such as “driver”, “app”, and “pricing” from Uber reviews to support aspect-level sentiment classification.

After the aspect terms are extracted, sentiment polarity classification is performed for each aspect. A sentiment scoring function was applied to the context window surrounding the aspect term, and the polarity score was calculated on a scale of -1 (strongly negative) to +1 (strongly positive). Each aspect was labelled as positive, neutral, or negative based on a set of thresholds. This means that the model could account for the nuanced sentiments expressed in a single review, such as the user disliked pricing, but was satisfied with the driver's behaviour. In this study, this approach was implemented to analyse Uber reviews at a finer level, enabling the identification of mixed sentiments within individual feedback for more precise service evaluation.

To illustrate the findings, a sentiment distribution chart was created that summarizes the polarity of the top 100 most mentioned aspects. From the analysis, we see that Aspects like "driver", and "cleanliness" tended to have a positive association in sentiment, while Aspects like "app performance", and "pricing" had a larger percentage of negative feedback. In this study, aspect-level sentiment visualization was used to highlight key areas of user satisfaction and dissatisfaction with Uber services, supporting actionable insights for service improvement.

Overall, this study's ABSA component provides deeper insights into customer satisfaction by isolating sentiment at the aspect level. This approach enhances the interpretability of sentiment analysis results and supports targeted service improvements in customer experience management [10]. In this study, ABSA was applied to Uber customer reviews to uncover specific strengths and weaknesses within the service, offering valuable directions for enhancing the user experience.

3.6 Evaluation Metrics

To thoroughly evaluate the effectiveness of the sentiment analysis models presented in this work, standard evaluation metrics, including accuracy, precision, recall, and F1 score, were applied. These metrics should be considered together to provide a balanced assessment of model performance across sentiment categories.

Accuracy refers to the overall proportion of correctly identified instances across all predictions and provides a general measure of accuracy. However, accuracy does not rank the predictions, and in the case of imbalanced datasets, such as the Uber review dataset with a surplus of positive sentiments, sole reliance on accuracy can be misleading. Therefore, the addition of other metrics is beneficial for more nuanced evaluations.

Precision refers to the number of true positives divided by the number of instances predicted to be positive and is a measure of the model's ability to avoid false positives. Nevertheless, it is important to consider this metric in applications where false-positive cases would lead to poor decisions. Recall refers to the number of true positives divided by the number of actual instances that were positive and is a measure of the model's ability to capture relevant cases and avoid false negatives.

To balance the trade-off between precision and recall, the F1 score, which is the harmonic mean of these two metrics, was selected as a criterion for model performance. The macro-averaged and weighted average F1 scores were then computed. The macro F1-score treats all classes equally, making it the most appropriate metric for examining model performance on classes in an imbalanced dataset, whereas the weighted F1-score includes the support (number of instances) for each class and provides a better representation of the overall score.

In addition to the scalar metrics described above, confusion matrices were created for each model to visualize the distribution of true versus predicted labels across the sentiment classes. These confusion matrices indicate the researcher's specific areas where the model may have had difficulty, such as misclassifying neutral reviews as positive or negative.

Additionally, comparative bar charts were developed alongside set precision-recall curves to be exploratory visualizations of model performance using different techniques, including lexicon-based (VADER), machine learning (SVM, RF), deep learning (BERT), and the hybrid technique proposed in this paper. Exploratory visualizations help to obtain a more intuitive understanding of the strengths or weaknesses of each technique and allow for a more data-driven use of a strategy for clinically meaningful sentiment analysis.

In this study, these evaluation tools were applied to assess the effectiveness of each sentiment analysis model, supporting informed conclusions regarding their applicability to real-world Uber customer feedback.

3.7 A New Hybrid Method

The primary contribution of this study is a hybrid sentiment analysis model that encompasses lexicon-based sentiment scoring via VADER and a machine learning classifier (RF) for enhanced accuracy and interpretability. The hybrid model is implemented by separating the initial measurement (VADER sentiment score) from the final prediction (RF), based on the class of inputs provided. The architecture of the hybrid model is as follows: For any informal text, first, used to generate sentiment scores.

Next, the scores from VADER are used to predict class probabilities for each distinct class of input via RF (therefore using "hybrid" (lexicon-based approach and a machine learning approach). This simultaneous hybrid process creates a VADER sentiment score which has advantageous "rule- based" capability of accommodating informal text; and provides the ensemble learning strength of RF.

Although the hybrid model is the main innovation of this study, many supporting methods have been utilized to improve performance. The first supporting method was emoji normalization, in which emojis were transformed into descriptive text to preserve all aspects of emotion. The next supporting method was negation handling to capture the correct sentiment polarity of phrases that used negation words such as "not" or "never". The third supporting method is the discovery of strong sentiment expressions using intensifiers (e.g., very, extremely). The final supporting method was Aspect Based Sentiment Analysis (ABSA) to recognize more fine-grained sentiment towards standard aspects of services such as "driver behaviour", "app performance", and "pricing".

In this study, the combination of individual and hybrid methods resulted in a robust, accurate, and interpretable sentiment analysis pipeline that is well suited for analysing informal and domain-specific customer feedback from Uber reviews.

4. RESULTS AND DISCUSSIONS

This section provides the performance results of the sentiment analysis models applied to the Uber Review dataset. Four models were used for the evaluation: VADER, SVM, RF, and BERT. The performance of the models was analysed based on the accuracy, precision, recall, and F1-score.

4.1 Model Performance Comparison

The hybrid model, the most successful model tested, attained the highest accuracy rate of the models tested (99.00%) and the highest Macro F1-Score (0.98). This demonstrates the benefits of using multiple sentiment-analysis techniques. VADER, although a lexicon-based model, performed unexpectedly well on the Macro F1-Score (0.95), even though it had much lower accuracy (72.4%) than the hybrid model. This may mean that it tends to overgeneralize the polarity of the sentiment. BERT performed well for accuracy (33.33%) but had a significantly lower F1-Score (0.17) in this configuration, possibly owing to the underfitting of the model or challenges with the dataset. SVM and RF performed well but were similar and did not perform as well as the hybrid model. In this study, BERT's performance of BERT highlights the limitations of deep learning when applied to small or imbalanced datasets. In this study, SVM and RF were found to provide more balanced results, but the combination of models demonstrated superior performance overall, confirming the effectiveness of the hybrid approach for analysing informal customer reviews.

Table 2 presents a comparative analysis of the accuracy and macro F1-score achieved by various models including VADER, SVM, RF, BERT, and the proposed hybrid model**.** Among them, the hybrid model performed the best, with the highest accuracy (99.00%) and macro F1-score (0.98). This result reflects the advantages of combining multiple sentiment analysis techniques. While VADER, a lexicon-based method, achieved a surprisingly high macro F1-score of 0.95, its accuracy remained lower at 72.4%, possibly indicating a tendency to misjudge the overall sentiment direction despite recognizing emotional cues.

The SVM and RF models yielded relatively high accuracy levels of 83.2% and 81.5%, respectively, but their macro F1-scores were notably lower, at 0.60 and 0.59. This suggests that although these models learn well from labelled data, they might favour dominant classes and overlook subtler sentiments.

Despite being a powerful deep learning model, BERT showed the weakest performance, with only 33.33% accuracy and a macro F1-score of 0.17. This could be due to insufficient training data, model underfitting, or class imbalance in the dataset. Although BERT is generally known for its strong contextual understanding, the model performed poorly in this experiment owing to computational limitations. The model was trained for only three epochs without GPU support, and no hyperparameter tuning or data-balancing techniques were applied. These constraints are likely to lead to underfitting and reduced generalization. Future work should address these issues by leveraging GPU resources and implementing a systematic tuning.

Overall, these findings indicate that single models have certain limitations when applied to noisy or unbalanced customer feedback. By leveraging the strengths of each individual model, the hybrid approach produced significantly more reliable outcomes for sentiment classification in this context.

Table 2. Comparison of Model Performance Metrics

Model	Accuracy	Macro F1-Score
VADER	72.4%	0.95
SVM	83.2%	0.60
RF	81.5%	0.59
BERT	33.33%	0.17
Hybrid	99.00%	0.98

Multiple visualizations were used to show the sentiment distribution and model performance. A sentiment distribution histogram showed an imbalance in the dataset, with most reviews classified as positive. An aspect-based sentiment distribution chart showed that aspects, like “driver” and “cleanliness,” had greater sentiment likelihood to be classified as a positive sentiment. Other aspects like “app performance” and “pricing” had a larger distribution of negative sentiment. The bar chart for model performance indicates that the hybrid model produced the best performance in terms of the classifier’s accuracy and F1-score. These visualizations helped see the trends in sentiments and behaviours in models, and they also helped make data-driven insights into the service to determine how the abnormality and impact of performance may be improved.

Figure 2 illustrates the sentiment distribution across the top 100 most frequently mentioned aspects of Uber customer reviews. The chart shows that aspects such as “driver” and “cleanliness” are predominantly associated with positive sentiment, while “app performance” and “pricing” have a higher proportion of negative sentiment. This visualization supports the effectiveness of the ABSA in identifying specific service components that influence customer satisfaction and dissatisfaction.

Figure 3 presents a performance comparison of five sentiment classification models: VADER, SVM, RF, BERT, and the proposed hybrid model. The Hybrid model achieved the best overall results with an accuracy of 99.00% and a macro F1-score of 98.0%, demonstrating the effectiveness of integrating lexicon-based, machine learning, and deep learning methods. VADER performed relatively well at a macro F1-score of 95.0%, but its accuracy was lower at 72.4%, suggesting a tendency to overgeneralize the sentiment polarity. SVM and RF achieved comparable results, with accuracies of 83.2% and 81.5% and macro F1-scores of 60.0% and 59.0%, respectively, indicating moderate but consistent performance. In contrast, BERT recorded the lowest values, with an accuracy of 33.33% and a macro F1-score of 17.0%, which may be due to model underfitting, or limitations of the dataset used.

4.2 Contributions

This study addressed key challenges in the sentiment analysis of customer feedback by aligning each research contribution with specific research objectives and existing problems in the field.

a) Hybrid Sentiment Analysis Framework
In this study, a combined sentiment analysis framework is developed by integrating VADER, SVM, RF, and BERT. This addresses the objective of improving classification accuracy by leveraging the strengths of lexicon-based, machine learning, and deep learning approaches. The framework responds to the current problem in which

traditional sentiment models struggle with informal language, lack contextual understanding, and exhibit inconsistent performance across domains.

b) Handling Informal Text Features
To meet the objective of accurately analyse informal user-generated content, this study incorporated advanced preprocessing techniques such as emoji normalization, negation handling, and intensifier detection. These enhancements address the common issue that standard models often fail to properly interpret informal expressions, resulting in reduced accuracy.

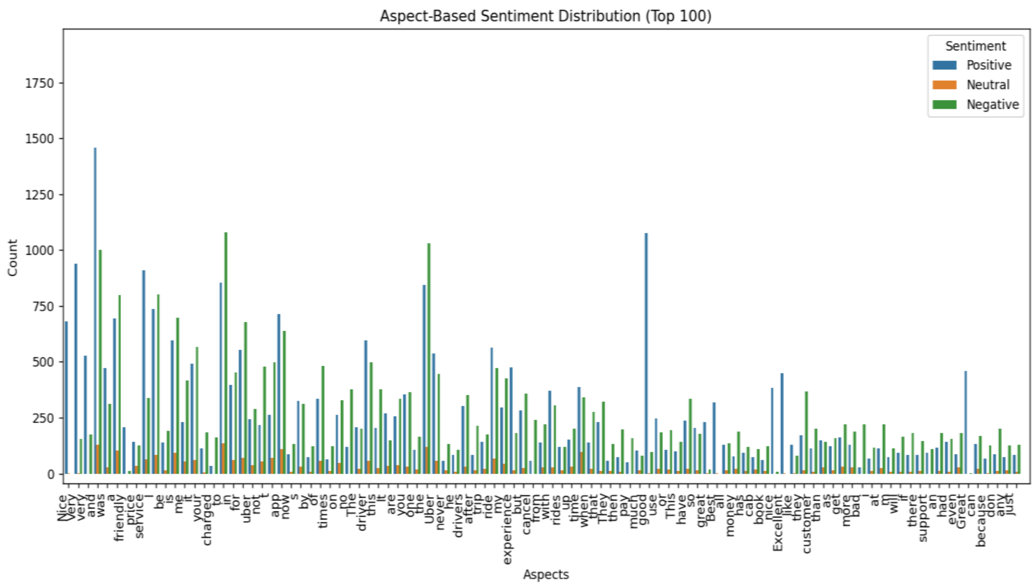

Figure 2. Sentiment Distribution of the Top 100 Aspects in Uber Reviews

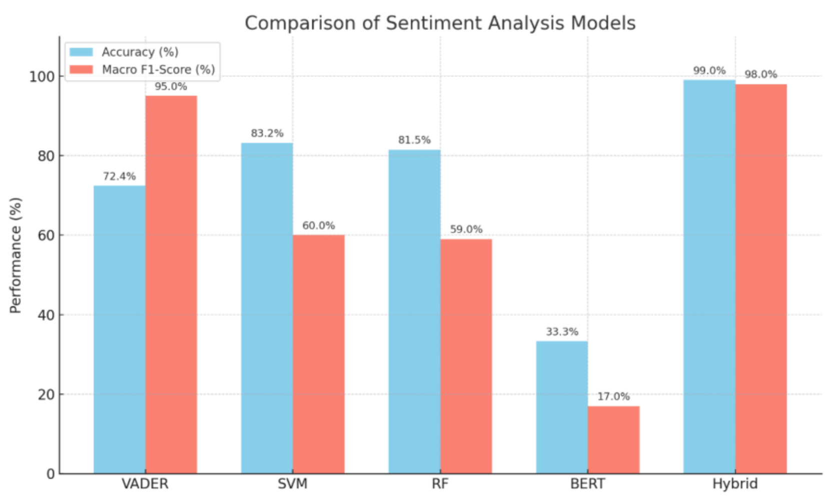

Figure 3. Comparison of Sentiment Analysis Models

c) ABSA
ABSA was implemented to achieve the objective of extract more detailed and actionable insights from reviews. By classifying sentiment towards individual service aspects (e.g., driver behaviour, app performance, pricing), this study resolved the limitations of traditional sentiment analysis, which tends to assign only a single label to an entire review and overlooks nuanced aspect-specific opinions.

d) High Performance Validated by Results
The objective of ensuring strong model performance was addressed by evaluating all models using accuracy, precision, recall, and F1-score. The hybrid model achieved 99.00% accuracy and a macro F1-score of 0.98, significantly outperforming individual models. This contribution tackles the problem of limited generalizability and imbalance in performance, as seen in many existing sentiment classifiers.

e) Real-World Application
The proposed framework was tested on over 12,000 real Uber customer reviews to achieve the objective of applying SA in practical settings. This confirms the practicality and robustness of the framework in handling actual customer feedback, addressing the issue that many academic models perform poorly on noisy datasets.

5. CONCLUSION

This study introduced a hybrid sentiment analysis framework that combines lexicon-based methods, machine learning models, and deep learning techniques. It was used to classify the sentiment in the Uber feedback. When combined with another coding method, it was better able to identify sentiments when using more casual, emotional, or shortened opinions. The classifier was based on word embeddings and aspect extracted to identify opinions applicable to specific features of the service (driver demeanour, car cleanliness, wait time, and fare for the ride).

The experiment was conducted using actual reviews of Uber users. The reviews contained familiar phrases such as "The driver was rude," "The car smelled bad," "He was 15 minutes late," "The cost was too high," and so on. The model could recognize the affective tone behind the comments and accurately associate them with appropriate service features. For example, the review "The driver kept using his phone while he was driving," was coded as a negative opinion and negatively related to driving safety. These tasks allow the platform to identify service failures using user comments.

This system can assist companies in several ways. First, it can sort large volumes of customer reviews and automatically flag negative ones. This reduces the time that customer support teams spend on manual review. Second, by grouping reviews based on service categories, businesses can measure the satisfaction levels in each area. For example, if reviews related to driver behaviour often contain negative words, the company knows where to improve. Third, by tracking repeated complaints in certain locations or time slots, the system can highlight operational issues. For instance, if many users mentioned delays in a specific area, the company could check the scheduling system or driver availability.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Jou Jia Yi: Data Preparation, Modelling, Validation, Writing – Original Draft Preparation;
Lew Sook Ling: Conceptualization, Supervision, Review – Editing.
Tan Li Tao: Conceptualization, Validation

CONFLICT OF INTERESTS

No conflicts of interest were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] K. Karoo, and M.V. Chitte, "Ethical Considerations in Sentiment Analysis: Navigating the Complex Landscape", *IRJMETS*, Dec. 2023, doi: 10.56726/IRJMETS46811.

[2] A. Clark, C. Fox, and S. Lappin, "The handbook of computational linguistics and natural language processing", in *Blackwell handbooks in linguistics*. Chichester, West Sussex; Malden, MA: Wiley-Blackwell, 2010.

[3] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis", *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, Jun. 2011, doi: 10.1162/COLI_a_00049.

[4] L. Xu, X. Pang, J. Wu, M. Cai, and J. Peng, "Learn from structural scope: Improving aspect-level sentiment analysis with hybrid graph convolutional networks", *Neurocomputing*, vol. 518, pp. 373–383, Jan. 2023, doi: 10.1016/j.neucom.2022.10.071.

[5] H. Jafarian, A. H. Taghavi, A. Javaheri, and R. Rawassizadeh, "Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language", in *2021 7th International Conference on Web Research (ICWR)*, Tehran, Iran: IEEE, May 2021, pp. 5–8. doi: 10.1109/ICWR51868.2021.9443131.

[6] Z. Wu, G. Cao, and W. Mo, "Multi-Tasking for Aspect-Based Sentiment Analysis via Constructing Auxiliary Self-Supervision ACOP Task", *IEEE Access*, vol. 11, pp. 82924–82932, 2023, doi: 10.1109/ACCESS.2023.3276320.

[7] Y. Chen, U. Liebau, S. M. Guruprasad, I. Trofimenko, and C. Minke, "Advancing Life Cycle Assessment of Sustainable Green Hydrogen Production Using Domain-Specific Fine-Tuning by Large Language Models Augmentation", *MAKE*, vol. 6, no. 4, pp. 2494–2514, Nov. 2024, doi: 10.3390/make6040122.

[8] K. Satyanarayana, "AI-Powered Customer Feedback Analysis and Sentiment Monitoring for Realtime Business Insights", *IJSREM*, vol. 09, no. 03, pp. 1–9, Mar. 2025, doi: 10.55041/IJSREM42513.

[9] D. Van Thin, D. N. Hao, and N. L.-T. Nguyen, "An Effective Contextual Language Ensemble Model for Vietnamese Aspect-based Sentiment Analysis", in *2022 9th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam: IEEE, Oct. 2022, pp. 30–34. doi: 10.1109/NICS56915.2022.10013429.

[10] Mamta and A. Ekbal, "Atmosphere kamaal ka tha (Was Wonderful): A Multilingual Joint Learning Framework for Aspect Category Detection and Sentiment Classification", *IEEE Transactions on Computational Social Systems*, vol. 11, no. 5, pp. 5892–5902, Oct. 2024, doi: 10.1109/TCSS.2024.3374450.

[11] H. Muhammad, F. A. Rafrastara, K. A. Setiadi, and A. Japardi, "Sentiment Analysis and Topic Modelling on Twitter Related to Mobile Legends: Bang Bang Game Using Lexicon-Based, LDA, and SVM", *Journal of Informatics and Web Engineering*, vol. 4, pp. 402–416, 2025.

[12] D. Nurjanah and H. Nurrahmi, "Cyberbullying Detection on Twitter using Support Vector Machine Classification Method", *bits*, vol. 3, no. 4, Mar. 2022, doi: 10.47065/bits.v3i4.1435.

[13] Q. T. Nguyen, T. L. Nguyen, N. H. Luong, and Q. H. Ngo, "Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews", in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam: IEEE, Nov. 2020, pp. 302–307. doi: 10.1109/NICS51282.2020.9335899.

[14] Z. Wu, G. Cao, and W. Mo, "Multi-Tasking for Aspect-Based Sentiment Analysis via Constructing Auxiliary Self-Supervision ACOP Task", *IEEE Access*, vol. 11, pp. 82924–82932, 2023, doi: 10.1109/ACCESS.2023.3276320.

[15] H. Jafarian, A. H. Taghavi, A. Javaheri, and R. Rawassizadeh, "Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language", in *2021 7th International Conference on Web Research (ICWR)*, Tehran, Iran: IEEE, May 2021, pp. 5–8. doi: 10.1109/ICWR51868.2021.9443131.

[16] T. Abdurahmonov, Muhammad Nabil Toby Abiyyu, Dzikru Nur Khaya, and M. Ary Heryanto, "Aspects-Based Sentiment Analysis of Extreme Weather on Twitter Using Long Short-Term Memory", *Journal of Informatics and Web Engineering*, vol. 4, no. no. 2, pp. 431–443, Jun. 2025.

[17] S. M. Tan, S. Roy, and A. Das et al., "Efficient Sentiment Classification using DistilBERT for Enhanced NLP Performance", *Journal of Quantum Computing*, vol. 4, no. 1, pp. 1–11, 2025, doi: 10.32604/jqc.2022.026658.

[18] Y. Aliyu, A. Sarlan, K. U. Danyaro, and A. S. B. A. Rahman, "Comparative Analysis of Transformer Models for Sentiment Analysis in Low-Resource Languages", *IJACSA*, vol. 15, no. 4, 2024, doi: 10.14569/IJACSA.2024.0150437.

[19] L. Nguyen, F. Wong, and Y. Zhao, "Real-Time Sentiment Insights from X Using VADER, DistilBERT, and Web-Scraped Data", *Real-Time Sentiment Insights from X Using VADER, DistilBERT, and Web-Scraped Data*, p. 100, Apr. 2025.

[20] M. M. Rahman, A. I. Shiplu, Y. Watanobe, and M. A. Alam, "RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis", *IEEE Transactions on Emerging Topics in Computational Intelligence,* vol. 1, no. 1, pp. 1-18, Jan. 2025, doi: 10.1109/TETCI.2025.3572150.

[21] A. Joshy and S. Sundar, "Analyzing the Performance of Sentiment Analysis using BERT, DistilBERT, and RoBERTa", in *2024 IEEE International Power and Renewable Energy Conference (IPRECON)*, Kollam, India: IEEE, Dec. 2024, pp. 1–6. doi: 10.1109/IPRECON55716.2022.10059542.

[22] D. Van Thin, D. N. Hao, and N. L.-T. Nguyen, "An Effective Contextual Language Ensemble Model for Vietnamese Aspect-based Sentiment Analysis", in *2022 9th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam: IEEE, Oct. 2022, pp. 30–34. doi: 10.1109/NICS56915.2022.10013429.

[23] D. Van Thin, D. N. Hao, and N. L.-T. Nguyen, "An Effective Contextual Language Ensemble Model for Vietnamese Aspect-based Sentiment Analysis", in *2022 9th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam: IEEE, Oct. 2022, pp. 30–34. doi: 10.1109/NICS56915.2022.10013429.

[24] H. Jafarian, A. H. Taghavi, A. Javaheri, and R. Rawassizadeh, "Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language", in *2021 7th International Conference on Web Research (ICWR)*, Tehran, Iran: IEEE, May 2021, pp. 5–8. doi: 10.1109/ICWR51868.2021.9443131.

[25] J. Dong, and Q. Qian, "A Density-Based Random Forest for Imbalanced Data Classification", *Future Internet*, vol. 14, no. 3, p. 90, Mar. 2022, doi: 10.3390/fi14030090.

[26] D. Van Thin, D. N. Hao, and N. L.-T. Nguyen, "An Effective Contextual Language Ensemble Model for Vietnamese Aspect-based Sentiment Analysis", in *2022 9th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam: IEEE, Oct. 2022, pp. 30–34. doi: 10.1109/NICS56915.2022.10013429.

[27] L. M. M. Campitelli et al., “Methodological and Ethical Considerations in the Use of Chordate Embryos in Biomedical Research”, *IJMS*, vol. 26, no. 6, p. 2624, Mar. 2025, doi: 10.3390/ijms26062624.

[28] D. Van Thin, D. N. Hao, and N. L.-T. Nguyen, “An Effective Contextual Language Ensemble Model for Vietnamese Aspect-based Sentiment Analysis”, in *2022 9th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam: IEEE, Oct. 2022, pp. 30–34. doi: 10.1109/NICS56915.2022.10013429.

[29] L. Xu, X. Pang, J. Wu, M. Cai, and J. Peng, “Learn from structural scope: Improving aspect-level sentiment analysis with hybrid graph convolutional networks”, *Neurocomputing*, vol. 518, pp. 373–383, Jan. 2023, doi: 10.1016/j.neucom.2022.10.071.

[30] H. Jafarian, A. H. Taghavi, A. Javaheri, and R. Rawassizadeh, “Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language”, in *2021 7th International Conference on Web Research (ICWR)*, Tehran, Iran: IEEE, May 2021, pp. 5–8. doi: 10.1109/ICWR51868.2021.9443131.

BIOGRAPHIES OF AUTHORS

	**Jou Jia Yi** received her bachelor’s degree in information technology, majoring in Business Intelligence and Analytics, from Multimedia University (MMU), Malaysia. She is currently pursuing her master’s degree in the same field at MMU. Her research interests include customer segmentation, clustering methods, and predictive analytics. She is passionate about data-driven solutions and aims to contribute to the advancement of intelligent business systems through applied research and innovation. She can be contacted at email: 1211306755@student.mmu.edu.my.
	**Lew Sook Ling** is an Associate Professor at the Faculty of Information Science and Technology, Multimedia University (MMU), Malaysia. She has been with MMU since 2001 and received her Ph.D. in 2013. She is also a Senior Member of IEEE. Her current research interests include educational technology, business analytics, image processing, and machine learning. She actively contributes to academic development through publications and collaborative research in emerging technologies. She can be contacted at email: sllew@mmu.edu.my.
	**Tan Li Tao** is pursuing a Bachelor of Arts (Honours) in Law at the University of Cambridge, United Kingdom. He is interested in examining the interaction of information technology and the social sciences. He declares that this work was conducted independently while enrolled at the University of Cambridge, and the views expressed are his own. He can be contacted at: ltt28@cam.ac.uk.