

TravelVibe: Aspect-Based Sentiment Analysis of Hotel Booking Application Reviews Using Support Vector Machines

Mohamad Hafiz Khairuddin^{1*}, Nurfazrin Zulaikha Binti Zunaini², Nurazian Binti Mior Dahalan³,
Azlin Binti Dahlan⁴, Zamlina Binti Abdullah⁵

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

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.100900282>

Received: 17 September 2026; Accepted: 22 September 2026; Published: 08 October 2026

ABSTRACT

Online reviews strongly influence travel decisions, yet their volume and informal language make manual interpretation impractical. This paper presents TravelVibe, a web-based aspect-based sentiment analysis system for reviews of Tripadvisor, Booking.com, and Agoda mobile applications. A total of 30,000 English-language Google Play Store reviews posted from 2022 to 2025 were collected. Reviews were cleaned, spell-corrected, lemmatised, and represented using unigram and bigram term frequency-inverse document frequency features. Non-negative matrix factorisation supported the derivation of four aspect categories—Customer Support, Hotel Experience, Price and Deals, and Travel Planning—whereas VADER compound scores supplied initial positive, neutral, and negative sentiment labels. Two calibrated linear support vector machine classifiers were trained for sentiment and aspect prediction. The Synthetic Minority Over-sampling Technique was applied only to the training data to address unequal class distributions. The baseline sentiment classifier achieved 80.99% accuracy; after balancing, accuracy increased to 91.42%, neutral-class recall rose from 0.50 to 0.98, and neutral F1-score rose from 0.58 to 0.95. Aspect accuracy increased from 88.45% to 88.94%, with improved balance for underrepresented aspects. Fourteen functional test cases passed, covering review analysis, filters, platform dashboards, and downloadable visualisations. The results indicate that TF-IDF and linear SVM remain effective for interpretable review analytics, although the use of automatically generated labels and semantically overlapping aspects limits the strength of external validity claims.

Keywords: aspect-based sentiment analysis, hotel booking applications, Support Vector Machine, SMOTE, TF-IDF, TravelVibe

INTRODUCTION

Mobile hotel-booking applications have become central to accommodation search, comparison, and reservation. Their usefulness depends not only on inventories and prices but also on the perceived credibility of other users' reported experiences. Online reviews, therefore, operate as a large, continuously updated source of information about platform reliability, service quality, booking friction, pricing transparency, and the actual hotel experience. Chen et al. (2022) demonstrated that online review content can materially influence purchasing decisions, while Fan et al. (2021) showed that the structure of hotel-review content affects its perceived helpfulness. For platform administrators, these reviews also provide operational evidence that can guide service improvements.

The practical problem is scale. A popular application may accumulate thousands of reviews written in informal language, containing spelling errors, abbreviations, mixed opinions, and references to several service features in the same text. Reading the reviews individually is slow and inconsistent. Conventional document-level sentiment analysis reduces each review to a single polarity and may therefore obscure the reasons behind satisfaction or dissatisfaction. A positive statement about price can coexist with a negative statement about customer support. Aspect-based sentiment analysis (ABSA) addresses this limitation by identifying the topic under discussion and

associating sentiment with it. Recent reviews of ABSA research confirm its growing use across domains and the continuing importance of domain-specific aspect design (Hua et al., 2024; Birjali et al., 2021).

Machine-learning classifiers are attractive for this setting because review text becomes a high-dimensional, sparse representation after vectorisation. Support Vector Machines (SVMs) have consistently demonstrated strong performance in text classification, including on travel and service-review datasets (Dahlan et al., 2022; Mustakim & Priyanta, 2022; Maroof et al., 2024). Nevertheless, performance can be distorted by class imbalance. Neutral reviews and less frequently discussed aspects may be overwhelmed by dominant classes, producing apparently acceptable overall accuracy while yielding poor recall for important minority categories. The present work consequently combines linear SVM classification with the Synthetic Minority Over-sampling Technique (SMOTE).

This study converts that analytical pipeline into TravelVibe, a web application that analyses reviews from Tripadvisor, Booking.com, and Agoda. Its objectives are: (1) to design an ABSA model and suitable visualisations; (2) to develop sentiment and aspect classifiers using SVM; and (3) to evaluate model accuracy and system functionality. The contribution is a complete, interpretable workflow linking large-scale review collection, NMF-assisted aspect labelling, VADER-assisted polarity labelling, TF-IDF feature engineering, class balancing, calibrated SVM prediction, and interactive decision-support views.

LITERATURE REVIEW

Hotel-booking review dimensions

Hotel-booking applications mediate several stages of the travel journey. Four dimensions were selected because they recur in both the literature and the collected data. Customer Support covers responsiveness, refunds, cancellations, communication, and problem resolution. Trust in online booking is closely linked to the platform's ability to resolve transaction-specific problems (Baki, 2020). Travel Planning includes search, itinerary organisation, maps, booking convenience, and coordination of travel arrangements. Price and Deals captures affordability, discounts, fees, price transparency, and value for money. Hotel Experience represents the extent to which the property, room, location, amenities, cleanliness, and service correspond to expectations formed during booking.

These dimensions are related but not mutually exclusive. A refund complaint may refer to both price and support; the word 'booking' may appear in a travel-planning comment or a support complaint. This semantic overlap is a central difficulty for aspect classification and later appears in the model's error patterns. It also explains why aspect-level analysis offers more useful operational detail than a single overall polarity, while still requiring careful label design.

Aspect-based sentiment analysis

Sentiment analysis can be undertaken at the document, sentence, or aspect level. Document-level classification assigns a single label to the entire text, whereas sentence-level analysis provides finer granularity. ABSA goes further by identifying entities, attributes, or predefined categories and then determining the sentiment associated with them. Zhu et al. (2019) framed aspect-category sentiment analysis as the joint interpretation of category and polarity, and Akhtar et al. (2020) demonstrated the value of jointly learning aspect extraction and sentiment classification. In applied service settings, however, a pipeline of separate classifiers can remain useful because it is transparent, computationally moderate, and easy to deploy.

Prior studies on hotel and travel support the choice of SVM. Bachtiar et al. (2020) reported that SVM outperformed Naive Bayes across several aspects of hotel reviews. Godakandage and Thelijjagoda (2020) also compared SVM and Naive Bayes for hotel recommendation. Mustakim and Priyanta (2022) achieved their best result using a tuned SVM on mobile travel application reviews. More broadly, Maroof et al. (2024) found competitive SVM performance for service-industry aspects and sentiment classification. These findings motivate an interpretable SVM baseline before introducing more complex contextual language models.

Feature representation, topic modelling, and imbalance

TF-IDF remains a practical representation for sparse text classification because it increases the contribution of discriminating terms and reduces the effect of corpus-wide common words. Its simplicity supports reproducibility and efficient inference (Jain et al., 2024). Including bigrams captures local phrases such as 'not good', 'customer service', and 'hidden fee' that may be more informative than isolated tokens. Explicit negation handling further preserves sentiment direction.

NMF decomposes a non-negative document-term matrix into topic and term factors. It has been used effectively for short-text topic modelling, particularly when paired with TF-IDF (Shi et al., 2018; Egger & Yu, 2022). In this study, it supports label construction rather than serving as the final predictive classifier. VADER supplies preliminary polarity labels through lexicon and rule-based scoring. This weak-supervision strategy makes it possible to efficiently prepare a large labelled corpus, but it also means that the supervised SVM ultimately learns labels generated by other algorithms rather than a fully human-annotated ground truth.

Class imbalance is addressed using SMOTE, which synthesises minority-class training points by interpolating between neighbouring observations. The intention is not merely to raise accuracy but to reduce systematic neglect of neutral sentiment and less frequent aspect classes. To prevent leakage, oversampling is confined to the training partition after the train-test split.

METHODOLOGY

Research design and dataset

The project followed a modified waterfall development process encompassing requirements, design, implementation, and testing. Reviews were obtained using the Google Play Scraper Python library for the mobile applications of Tripadvisor, Booking.com, and Agoda. The collection period covered 2022-2025. Up to 10,000 valid reviews were retained for each platform, producing 30,000 records. Each record preserved review content, user rating, platform, and submission date. Identifiers and other fields not required for the analytical tasks were excluded.

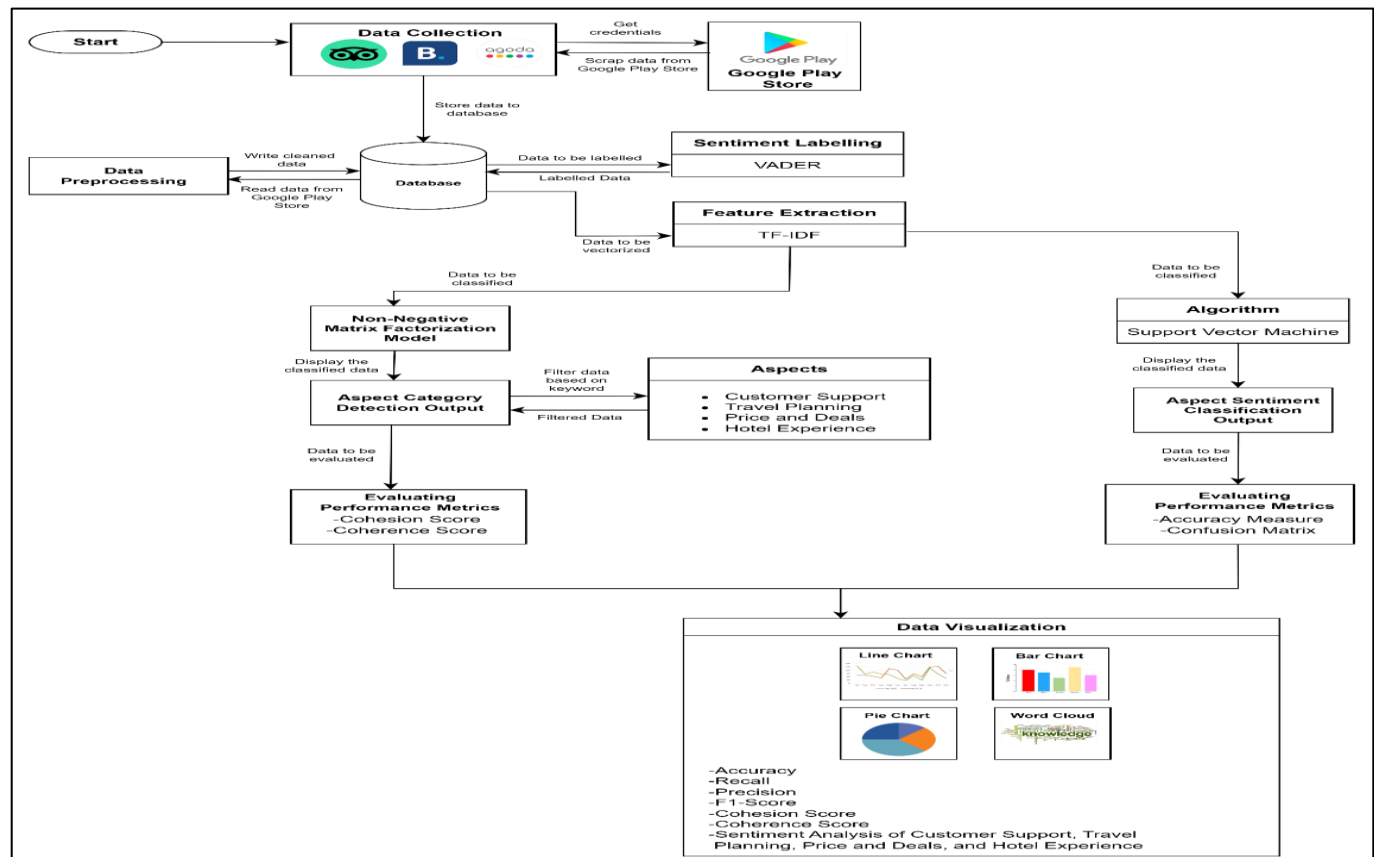


Figure 1. Research design of the TravelVibe project.

Text preprocessing

Raw reviews were standardised through a staged preprocessing pipeline. Emojis and punctuation were removed using regular expressions; whitespace was normalised; spelling correction was performed using SymSpell; tokens were lemmatised using NLTK; and common stopwords were removed. Reviews shorter than five words were excluded from NMF topic refinement, and consecutive duplicate words were reduced. For sentiment modelling, negation terms were joined to their following terms so that expressions such as 'not good' became 'not_good'. The processed review text was stored separately from the original text to preserve traceability.

Label construction

Aspect labels were constructed through NMF topic modelling. Cleaned reviews were represented with TF-IDF and initially decomposed into six latent topics. Generic platform names and travel terms identified during inspection were added to a custom stopword list, and the topic model was rerun. Topic keywords and their semantic meanings were examined, after which related topics were mapped into four operational categories: Customer Support, Hotel Experience, Price and Deals, and Travel Planning. The final four-topic solution achieved reported c_v coherence and normalised pointwise mutual information values of 0.6661, indicating internally consistent keyword groups.

Sentiment labels were generated with VADER compound scores. Scores greater than or equal to 0.05 were labelled positive, scores less than or equal to -0.05 were labelled negative, and scores between those thresholds were labelled neutral. Empty or invalid reviews defaulted to neutral during the automated routine. These sentiment and aspect labels became the targets for the two supervised classifiers.

Feature extraction and classifiers

For both tasks, reviews were divided into stratified training and test sets using an 80:20 ratio. TF-IDF vectorizers were fitted on the training text and then applied to the test text. Sentiment features included unigrams and bigrams with $\max_df=0.90$ and $\min_df=5$. The linear sentiment classifier used $C=1.0$. Aspect features also included unigrams and bigrams, with $\max_df=0.90$ and $\min_df=3$. The linear aspect classifier used $C=0.6$ and $\text{class_weight}='balanced'$.

SMOTE was applied only to the training matrix to create synthetic minority-class instances. Each LinearSVC model was wrapped with CalibratedClassifierCV using sigmoid calibration so that the web application could display class-confidence estimates. Evaluation used accuracy, class-specific precision, recall and F1-score, macro averages, classification reports, and confusion matrices. Baseline models without SMOTE were compared with their SMOTE-enhanced counterparts.

System implementation and testing

The trained classifiers and vectorizers were serialised as Python pickle files and integrated into a Flask web application. TravelVibe provides an overview dashboard, platform-specific pages, an aspect-analysis page with date filters, and an analyser for new text. Visual outputs include sentiment summary cards, word clouds, pie charts, bar charts, and monthly sentiment trends. Functional testing comprised 14 test cases covering page display, filters, chart controls, platform views, review input, and prediction output.

System Development

TravelVibe architecture

TravelVibe separates the analytical pipeline from the presentation layer. The Flask backend loads the saved vectorizers and calibrated classifiers, preprocesses a submitted review, and returns both the predicted sentiment and aspect. Aggregated CSV data support dashboard views for the entire corpus and each platform. This modular structure allows a classifier or vectorizer to be replaced without redesigning the front end.

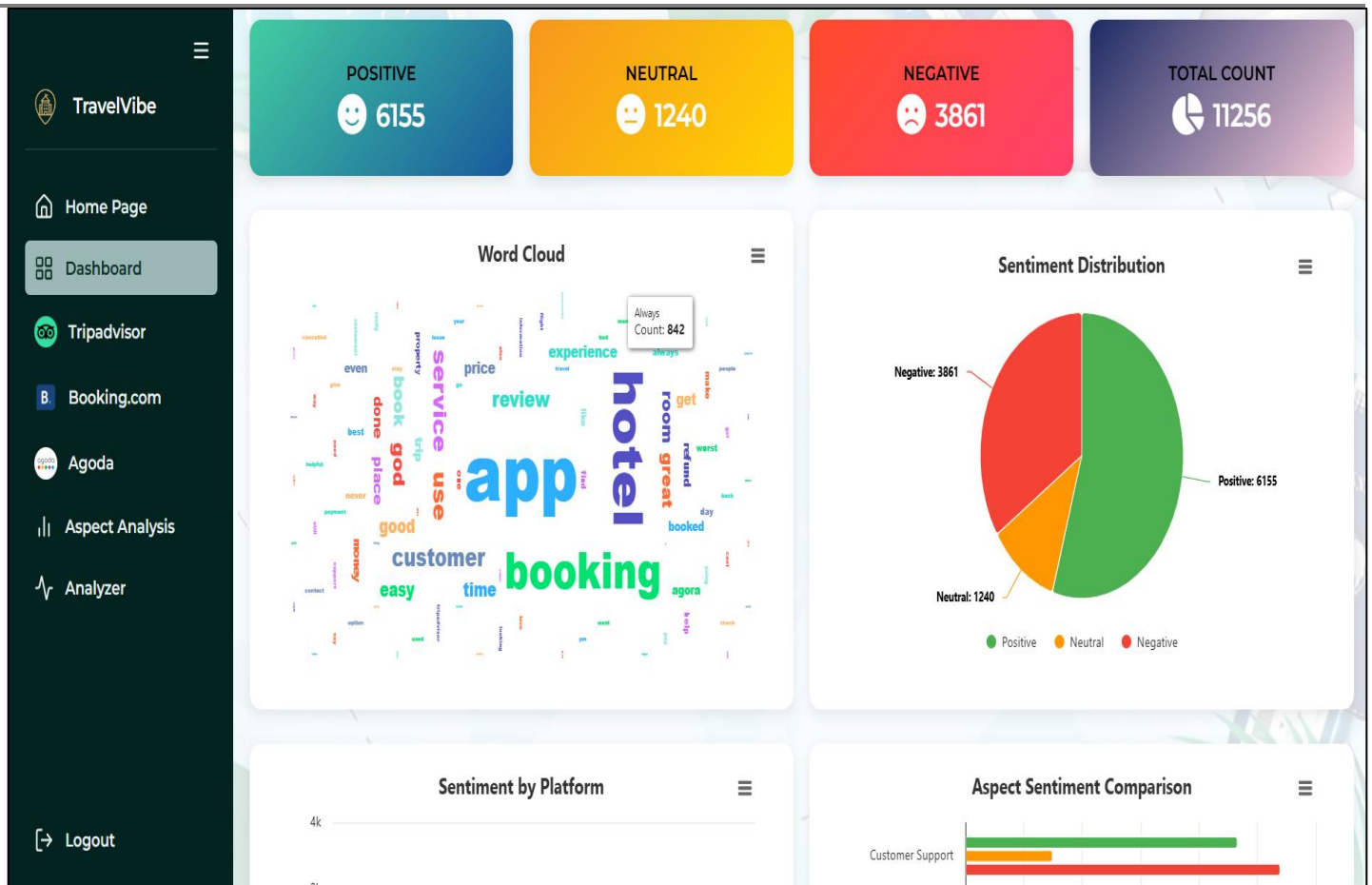


Figure 2. TravelVibe dashboard presenting aggregate sentiment and aspect visualisations.

Interactive analyzer

The analyzer gives a direct demonstration of the two-stage classification process. A user enters a review, which is passed through the same preprocessing and vectorisation logic used during model development. The sentiment and aspect models then generate a polarity label, an aspect category, and calibrated confidence estimates. This feature transforms the system from a static reporting dashboard into an interactive review and analysis tool.

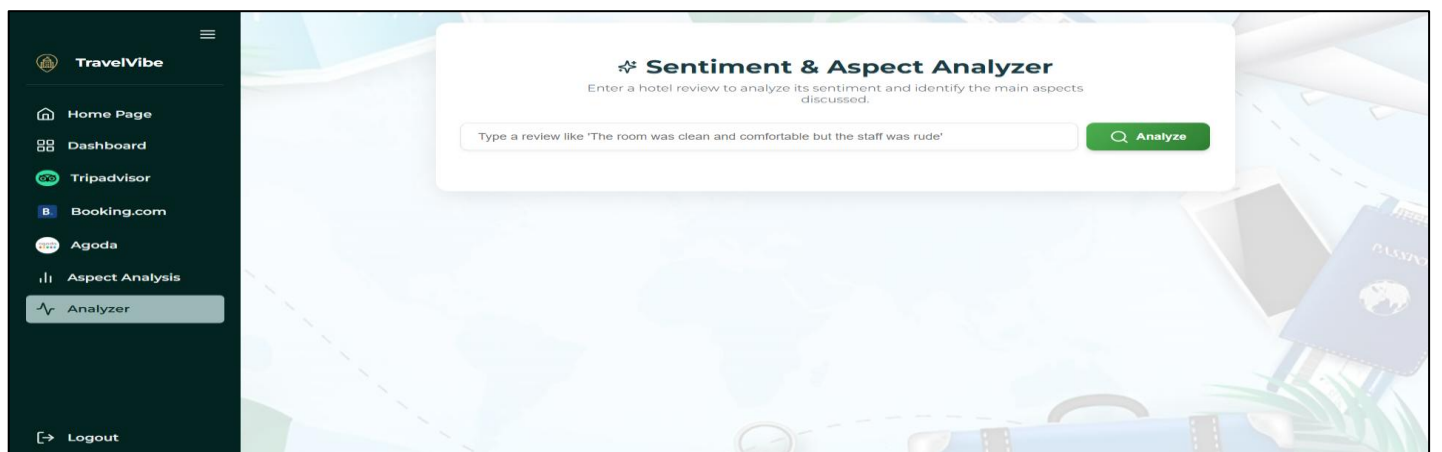


Figure 3. Analyser interface for real-time sentiment and aspect prediction.

Decision-support views

The overview dashboard summarises total positive, neutral, and negative reviews and provides cross-platform visual comparisons. Dedicated Tripadvisor, Booking.com, and Agoda pages retain a consistent layout so that

stakeholders can compare common terms and aspect-level sentiment. The aspect-analysis page provides a year-range filter for 2022-2025 and allows a selected category to be examined in greater detail. Chart actions support full-screen viewing, printing, and downloading in common image formats.

RESULTS AND DISCUSSION

NMF topic evaluation

The final NMF output produced four interpretable themes. One topic emphasised hotel, booking, room, refund, and price; another emphasised ease of use and navigation; a third captured trip- and platform-related feedback; and a fourth emphasised customer, service, support, contact, and dissatisfaction. Both reported coherence measures were 0.6661. The topic solution, therefore, provided a plausible mechanism for organising the corpus, though semantic inspection and manual mapping were still necessary to align latent topics with the four predefined application aspects.

Sentiment classification

The baseline linear SVM achieved 0.8099 accuracy. Its principal weakness was neutral sentiment: precision was 0.68, recall was 0.50, and F1-score was 0.58. This pattern indicates that many reviews labelled neutral were absorbed into the positive or negative classes. Following training-set oversampling, overall accuracy increased to 0.9142. Neutral precision reached 0.92, recall reached 0.98, and F1-score reached 0.95. Macro-average precision and recall both reached approximately 0.9133.

Table 1 Sentiment-Classification Performance Before And After Smote

Metric	SVM without SMOTE	SVM with SMOTE
Accuracy	0.8099	0.9142
Neutral precision	0.6800	0.9200
Neutral recall	0.5000	0.9800
Neutral F1-score	0.5800	0.9500
Macro-average precision	0.7700	0.9133
Macro-average recall	0.7300	0.9133

The 10.43 percentage-point gain in accuracy is substantial, but the class-specific changes are more important. The neutral-class recall gain of 0.48 indicates that balancing altered what the classifier learned about the middle category. Synthetic training observations expanded the minority region of feature space and reduced majority-class dominance. The result supports SMOTE as a useful intervention for this dataset, but does not prove that every newly recovered neutral prediction is semantically correct, since the original VADER labels were automatically assigned.

Aspect classification

The baseline aspect classifier achieved 0.8845 accuracy. Customer Support was the most distinctive category, with reported precision of 0.88, recall of 0.95, and F1-score of 0.91. Errors concentrated among semantically related categories. Travel Planning reviews were sometimes classified under Customer Support, while Hotel Experience could overlap with Price and Deals due to shared terms such as 'booking', 'room', 'location', and 'service'.

With SMOTE, aspect accuracy rose modestly to 0.8894. The Price and Deals category achieved a precision of 0.93, a recall of 0.82, and an F1-score of 0.87; Travel Planning achieved a recall of 0.83 and an F1-score of 0.87,

respectively. The macro-average precision improved from 0.89 to 0.91. These results suggest that balancing mainly improved consistency across categories rather than producing a dramatic increase in aggregate accuracy.

Table 2 Aspect-Classification Performance Before And After Smote

Metric	SVM without SMOTE	SVM with SMOTE
Accuracy	0.8845	0.8894
Price and Deals F1-score	0.8400	0.8700
Macro-average precision	0.8900	0.9100

Functional evaluation

All 14 specified functional test cases passed. The tests covered access to the landing, home, dashboard, Tripadvisor, Booking.com, Agoda, aspect-analysis, and analyser pages; chart-action controls; zoom reset; aspect selection; year filtering; review entry; and output generation. This confirms that the principal analytical and visualisation functions were connected correctly in the tested environment.

Table 3 Summary of Travelvibe Functional Testing

Functional area	Representative checks	Outcome
Navigation	Landing, home, dashboard, platform and analyser pages	Passed
Visualization controls	Action menu, reset zoom, aspect selection and filtering	Passed
Review analysis	Text input, Analyse action, sentiment and aspect output	Passed
Platform reporting	TripAdvisor, Booking.com and Agoda visualisations	Passed

DISCUSSION

The findings show that an interpretable classical machine-learning pipeline can deliver strong performance for a large review corpus. Sparse TF-IDF features align naturally with a linear maximum-margin classifier, and bigrams plus explicit negation handling preserve locally meaningful phrases. Probability calibration adds a useful interface-level capability even though LinearSVC does not natively output probabilities. The resulting design is computationally lighter and easier to inspect than many deep neural alternatives.

The sentiment experiment demonstrates why overall accuracy should not be reported alone. Before SMOTE, the model's 80.99% accuracy concealed a neutral recall of only 0.50. After balancing, neutral recall reached 0.98. The aspect experiment tells a different story: aggregate accuracy changed by less than half a percentage point, yet minority-aspect metrics became more balanced. In both cases, the value of SMOTE lies in class coverage and prediction fairness, not simply in maximising a single score.

Several cautions are necessary. First, the evaluation measures agreement with VADER- and NMF-derived labels rather than agreement with independent expert annotation. High accuracy partly indicates that SVM can reproduce the patterns encoded by those labelling procedures. Second, NMF topics required manual consolidation into four categories, introducing judgment into the target definition. Third, a random split of reviews from the same applications and time period may not reflect performance on a future platform, a different country, or genuinely new language patterns. Fourth, calibrated probability values should be interpreted as model confidence rather than proof of correctness.

Despite these constraints, TravelVibe has practical value as a prototype. It turns unstructured feedback into platform- and aspect-specific summaries, exposes trends through filters and visualisations, and permits immediate analysis of new reviews. For administrators, the aspect categories help direct attention toward support, pricing, planning, or stay-related concerns. For researchers, the study offers a reproducible baseline against which contextual transformer models and human-labelled datasets can be compared.

CONCLUSION AND FUTURE WORK

This paper presented TravelVibe, a web-based ABSA system for 30,000 Google Play Store reviews of Tripadvisor, Booking.com, and Agoda collected from 2022 to 2025. The system combines NMF-assisted aspect labelling, VADER-assisted sentiment labelling, TF-IDF unigram and bigram features, calibrated linear SVM classification, SMOTE, and interactive Flask visualisations. The SMOTE-enhanced sentiment classifier achieved 91.42% accuracy and greatly improved neutral recall and F1-score. The aspect classifier achieved 88.94% accuracy, with improved coverage of underrepresented categories. All 14 functional test cases passed.

Future research should replace or validate the automatically generated targets using a stratified, independently human-annotated test set and report inter-annotator agreement. Contextual models such as BERT, sentence transformers, or domain-adapted language models should be compared under identical splits. Multi-label or sentence-level aspect assignment could better capture reviews that address multiple concerns. Temporal holdout evaluation, cross-platform transfer tests, multilingual reviews, explainability methods, and real-time data ingestion would further establish robustness and practical scalability.

REFERENCES

1. Akhtar, M. S., Garg, T., & Ekbal, A. (2020). Multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing*, 398, 247-256. <https://doi.org/10.1016/j.neucom.2020.02.093>
2. Bachtiar, F. A., Paulina, W., & Rusydi, A. N. (2020). Text mining for aspect based sentiment analysis on customer review: A case study in the hotel industry. *IICST2020*, 105-112.
3. Baki, R. (2020). Analysis of factors affecting customer trust in online hotel booking website usage. *European Journal of Tourism, Hospitality and Recreation*, 10(2), 106-117. <https://doi.org/10.2478/ejthr-2020-0009>
4. Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. <https://doi.org/10.1016/j.knosys.2021.107134>
5. Chen, T., Samaranayake, P., Cen, X. Y., Qi, M., & Lan, Y. C. (2022). The impact of online reviews on consumers' purchasing decisions: Evidence from an eye-tracking study. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.865702>
6. Dahlan, A., Gunawan, A., & Wibowo, F. W. (2022). Sentiment analysis of airline ticket and hotel booking of Traveloka using Support Vector Machine. *2022 5th International Seminar on Research of Information Technology and Intelligent Systems*, 537-542. <https://doi.org/10.1109/ISRITI56927.2022.10052839>
7. Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology*, 7, 1-16. <https://doi.org/10.3389/fsoc.2022.886498>
8. Fan, W., Liu, Y., Li, H., Tuunainen, V. K., & Lin, Y. (2021). Quantifying the effects of online review content structures on hotel review helpfulness. *Internet Research*, 32(7), 202-227. <https://doi.org/10.1108/INTR-11-2019-0452>
9. Godakandage, M., & Thelijjagoda, S. (2020). Aspect based sentiment oriented hotel recommendation model exploiting user preference learning. *2020 IEEE 15th International Conference on Industrial and Information Systems*, 409-414. <https://doi.org/10.1109/ICIIS51140.2020.9342744>
10. Hua, Y. C., Denny, P., Wicker, J., & Taskova, K. (2024). A systematic review of aspect-based sentiment analysis: Domains, methods, and trends. *Artificial Intelligence Review*, 57(11). <https://doi.org/10.1007/s10462-024-10906-z>

11. Jain, S., Jain, S. K., & Vasal, S. (2024). An effective TF-IDF model to improve text classification performance. 2024 IEEE 13th International Conference on Communication Systems and Network Technologies, 1-4. <https://doi.org/10.1109/CSNT60213.2024.10545818>
12. Maroof, A., Wasi, S., Jami, S. I., & Siddiqui, M. S. (2024). Aspect-based sentiment analysis for service industry. IEEE Access, 12, 109702-109713. <https://doi.org/10.1109/ACCESS.2024.3440357>
13. Mustakim, H., & Priyanta, S. (2022). Aspect-based sentiment analysis of KAI Access reviews using NBC and SVM. IJCCS, 16(2), 113-124. <https://doi.org/10.22146/ijccs.68903>
14. Shi, T., Kang, K., Choo, J., & Reddy, C. K. (2018). Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. The Web Conference 2018, 1105-1114. <https://doi.org/10.1145/3178876.3186009>
15. Zhu, P., Chen, Z., Zheng, H., & Qian, T. (2019). Aspect aware learning for aspect category sentiment analysis. ACM Transactions on Knowledge Discovery from Data, 13(6). <https://doi.org/10.1145/3350487>