

Gigalyze: Aspect-Based Sentiment Analysis of Gig Economy Platform Services Using Support Vector Machines

Nurazian Binti Mior Dahalan¹, Nurul Nabila Binti Mohd Rajil², Mohamad Hafiz Khairuddin^{3*}, Azlin Binti Dahlan⁴, Zamalina 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.2026.100900234>

Received: 16 September 2026; Accepted: 21 September 2026; Published: 07 October 2026

ABSTRACT

The expansion of digital freelancing has increased the volume of user feedback associated with gig economy platforms, making manual interpretation slow and inconsistent. This paper presents Gigalyze, a web-based aspect-based sentiment analysis system for Google Play Store reviews of Fiverr, Freelancer, and Upwork. The dataset contains 30,000 English-language reviews, comprising 10,000 reviews per platform, collected for the Malaysian market between 2022 and 2025. Reviews were cleaned, spell-corrected, normalised, lemmatised, and represented using term frequency-inverse document frequency features. Non-negative matrix factorisation supported the derivation of four platform-service aspects: User Experience, App Performance and Functionality, Account and Support Issues, and Payment and Earnings. VADER compound scores supplied initial positive, neutral, and negative sentiment labels. Separate linear Support Vector Machine classifiers were developed for sentiment and aspect prediction, while the Synthetic Minority Over-sampling Technique was applied to the training data to reduce class imbalance. After balancing, sentiment accuracy reached 83.74%, neutral recall increased from 0.26 to 0.49, and macro F1-score increased from 0.68 to 0.73. Aspect accuracy reached 88.52%, with User Experience recall improving from 0.86 to 0.88. All 14 functional test cases passed. These results demonstrate the feasibility of interpretable machine-learning analytics for gig-platform reviews while also highlighting limitations stemming from automatically generated labels, overlapping aspect vocabularies, and subtle emotional expressions.

Keywords: aspect-based sentiment analysis, gig economy, Support Vector Machine, SMOTE, TF-IDF, Fiverr, Freelancer, Upwork, Gigalyze

INTRODUCTION

The gig economy is reshaping work by connecting independent workers and clients through digital platforms. In Malaysia, platform-mediated work has become an important source of additional income and flexible employment. Harun et al. (2020) demonstrated the growing significance of gig work in Malaysian labour statistics, while Muhyi et al. (2023) identified economic and flexibility-related forces that encourage workers to enter the sector. Digital freelancing platforms extend this model beyond location-based services, allowing workers to offer professional services to clients worldwide.

Fiverr, Freelancer, and Upwork are prominent examples. Fiverr organises services as purchasable gigs, Freelancer emphasises project bidding and crowdsourcing, and Upwork supports recruitment, contracting, communication, payment processing, and dispute resolution. These platforms create opportunities, but the quality of participation depends on interface usability, application reliability, customer support, account administration, payment security, and workers' ability to receive earnings. Platform design and information asymmetry can materially affect worker outcomes and autonomy (Dedema & Rosenbaum, 2024; Pilatti et al., 2024).

Google Play Store reviews contain direct evidence of these experiences. Their quantity and informal nature, however, make manual analysis impractical. A single review may mention application crashes, account restrictions, delayed payments, and poor customer support. Conventional sentiment analysis produces an overall polarity but often fails to explain the specific service feature responsible for it. Aspect-based sentiment analysis (ABSA) addresses this limitation by linking opinions to specific product or service categories (Brauwers & Frasincar, 2022; Hua et al., 2024).

This study develops Gigalyze to analyse gig-platform reviews at both sentiment and aspect levels. Its objectives are to design an ABSA model with suitable visualisations, implement sentiment and aspect classifiers using Support Vector Machines (SVMs), and evaluate the predictive performance and web system functionality. The principal contribution is an end-to-end, interpretable pipeline combining Malaysian-market review collection, NMF- and VADER-assisted labelling, TF-IDF feature representation, SVM classification, SMOTE balancing, calibrated confidence scores, and interactive platform comparison.

LITERATURE REVIEW

Gig Economy Platforms and Service Aspects

Digital labour platforms operate as market intermediaries, governing how workers present skills, find projects, communicate with clients, receive payments, and resolve disputes. Case studies of Fiverr, Freelancer, and Upwork show that the platforms differ in service organisation and contracting mechanisms while sharing a dependence on reputation, profiles, and platform-managed transactions (Green et al., 2018a; Green et al., 2018b; Green et al., 2021). These socio-technical arrangements make application quality a meaningful part of the working environment rather than a peripheral convenience.

Four aspects were selected from the review corpus. User Experience represents ease of use, navigation, accessibility, and overall interaction quality. App Performance and Functionality covers speed, stability, crashes, updates, and the correct operation of features. Account and Support Issues include verification, suspension, access, complaint handling, and responsiveness. Payment and Earnings represent fees, withdrawals, delayed payments, pricing, jobs, and income. These aspects provide operationally actionable categories for both platform companies and workers.

Aspect-Based Sentiment Analysis

ABSA normally includes aspect extraction or category detection followed by polarity classification. Aspect-category sentiment analysis is especially suitable when a study defines a manageable set of service categories. It allows feedback to be aggregated by operational concern even when users express similar sentiments using different terms. Ali et al. (2020) applied ABSA to ride-sharing platforms, and Indrawan et al. (2020) examined users' preferences for gig economy platforms. Their work supports using review analytics to identify platform-specific weaknesses that general star ratings cannot reveal.

SVM is well-suited to high-dimensional, sparse text. It identifies a separating hyperplane with a large margin and often performs strongly with TF-IDF features. Surveys continue to report its reliability for classification problems, although performance depends on feature design, class distribution, and hyperparameters (Cervantes et al., 2020). Previous application-review studies have also reported effective SVM-based ABSA results (Mustakim & Priyanta, 2022; Nurul Hidayati et al., 2023).

Topic Modelling, Weak Labelling, and Imbalance

NMF factorises a non-negative document-term matrix into document-topic and topic-term representations, providing an interpretable way to discover recurring themes (Lee & Seung, 2001; Abdelrazek et al., 2023). In Gigalyze, NMF is used to support aspect label construction rather than final prediction. VADER supplies initial sentiment labels through lexicon and rule-based compound scores. This weakly supervised strategy

"platform". Gensim phrase modelling combined meaningful bigrams and trigrams such as customer_support and payment_issues. Topics were manually interpreted as Payment and Earnings, User Experience, Account and Support Issues, and App Performance and Functionality.

VADER calculated a compound sentiment score between -1 and +1. Reviews scoring at least 0.05 were labelled positive, those scoring at most -0.05 were labelled negative, and intermediate scores were labelled neutral. The resulting generated sentiment and aspect labels served as target classes for supervised learning.

Model Development and Evaluation

The data was split into 80% for training and 20% for testing. The aspect split used stratification to preserve category proportions. TF-IDF was fitted to the training text and used to transform the test text, preventing vocabulary leakage. Aspect features excluded terms that occurred in fewer than 5 documents or in more than 90% of documents, and included both unigrams and bigrams. LinearSVC classifiers were trained separately for sentiment and aspect prediction.

SMOTE was applied only to the training features. Baseline and balanced classifiers were compared using accuracy, precision, recall, F1-score, macro F1, classification reports, and confusion matrices. Because LinearSVC does not natively return probabilities, CalibratedClassifierCV with Platt scaling converted decision values into class-confidence estimates for the interactive analyser.

System Development

Gigalyze Web Application

Gigalyze was implemented with Python and Flask. The back end loads the saved vectorizers, trained models, and review dataset, while HTML and CSS provide the presentation layer. The dashboard aggregates positive, neutral, and negative totals and compares results across Fiverr, Freelancer, and Upwork. Additional pages provide platform-specific summaries, aspect analysis, interactive word clouds, frequent-word exploration, and filtered source reviews.

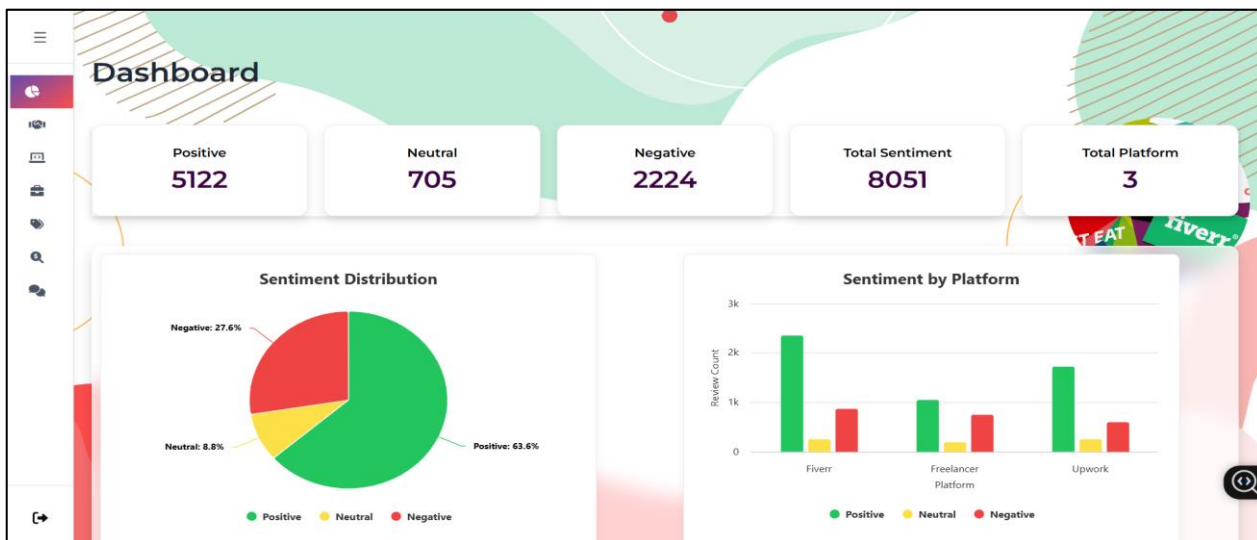


Figure 2. Gigalyze dashboard for aggregate and cross-platform review analysis.

Interactive Analyser

The analyser accepts a gig-related review and returns the predicted sentiment, aspect, and calibrated confidence score. Input controls reject blank submissions and reviews shorter than three words. Definitions of the four aspects are displayed alongside the result to help users interpret the classification. The feature demonstrates how the trained models can be used beyond the historical dashboard data.

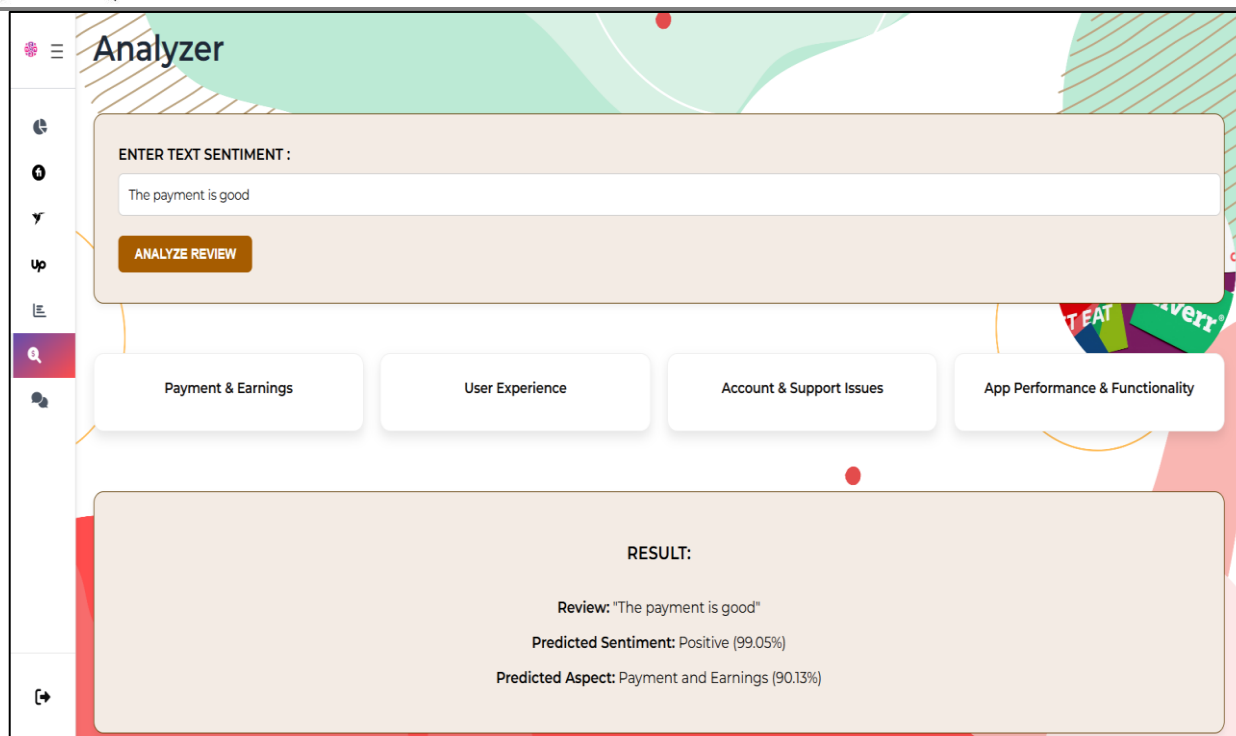


Figure 3. Gigalyze analyser displaying aspect, sentiment, and calibrated confidence.

RESULTS AND DISCUSSION

NMF Evaluation

The four-topic NMF solution produced coherent groups corresponding to the selected aspects. Topic keywords included terms related to freelancing and experience, application operation, support and account access, and jobs, pay, and money. The reported coherence score was 0.6783. Word cloud inspection supported the interpretation that each group contained semantically related terms, although manual judgment was still necessary to name topics and assign them to operational categories.

Sentiment Classification

The baseline sentiment model achieved 83.36% accuracy. Positive sentiment performed strongly, with a precision of 0.87, a recall of 0.93, and an F1-score of 0.90. Negative sentiment achieved approximately 0.78 across precision, recall, and F1. Neutral sentiment was the principal weakness: recall was 0.26 and F1-score was 0.36. Of 134 neutral test reviews, only 35 were correctly classified, while 60 were predicted as positive and 39 as negative. After SMOTE, sentiment accuracy rose slightly to 83.74%, but minority-class performance improved materially. Neutral recall increased to 0.49 and F1-score to 0.51; 66 neutral reviews were correctly predicted. Macro F1 increased from 0.68 to 0.73. The main benefit of balancing was therefore improved coverage of neutral sentiment, rather than a large increase in headline accuracy.

TABLE 1 Sentiment Model Performance Before And After Smote

Metric	Without SMOTE	With SMOTE
Accuracy	83.36%	83.74%
Neutral recall	0.26	0.49
Neutral F1-score	0.36	0.51
Macro F1-score	0.68	0.73

Aspect Classification

The baseline aspect classifier achieved 88.45% accuracy. User Experience obtained a precision of 0.94, a recall of 0.86, and an F1-score of 0.90. App Performance and Functionality achieved an F1-score of 0.88; Account and Support Issues achieved 0.87; and Payment and Earnings achieved 0.89. Some Account and Support reviews were confused about functionality or payment because terms describing access, jobs, and transactions can span categories.

With SMOTE, accuracy reached 88.52%. User Experience recall improved to 0.88 and F1-score to 0.91, while Account and Support Issues recall improved from 0.85 to 0.86. The macro F1 remained 0.89. These small changes indicate that the initial aspect model was already strong and that balancing mainly reduced disparities rather than transforming overall performance.

TABLE 2 Aspect Model Performance Before And After Smote.

Metric	Without SMOTE	With SMOTE
Accuracy	88.45%	88.52%
User Experience recall	0.86	0.88
User Experience F1-score	0.90	0.91
Macro F1-score	0.89	0.89

Functional and Validation Testing

All 14 functional test cases passed. They covered navigation to the landing page, dashboard, platform pages, analyser, and review word-cloud page; aspect and sentiment selection; valid and invalid analyser input; and frequent-word and filtered-review functions. Additional validation used unseen, real-world review examples and compared sentiment and aspect outputs with human-judged expectations.

TABLE 3 Summary Of Gigalyze Functional Testing.

Area	Checks	Outcome
Navigation	Landing, dashboard, Fiverr, Freelancer, Upwork and analyser pages	Passed
Filters and visualisation	Aspect, sentiment, word cloud and frequent-word controls	Passed
Input validation	Valid text, blank input and a minimum three-word requirement	Passed
Prediction	Sentiment, aspect and calibrated confidence output	Passed

DISCUSSION

The results confirm that TF-IDF and linear SVM provide a credible and interpretable baseline for gig-platform review analytics. Overall accuracy alone would have understated the original model's difficulty with neutral reviews. The increase in neutral recall from 0.26 to 0.49 and macro F1 from 0.68 to 0.73 provides stronger evidence of SMOTE's value than the 0.38 percentage-point accuracy gain. For aspect classification, balancing produced only incremental improvement because all four classes already performed comparatively well.

The aspect findings are operationally meaningful. Payment and Earnings feedback can reveal withdrawal delays, fees, job availability, and transaction problems. Account and Support Issues can expose verification or access barriers. App Performance and Functionality highlights crashes, speed, and feature failures, while User Experience captures usability and navigation. Separating these categories helps platforms prioritise improvements and helps workers compare services beyond aggregate ratings.

The evidence must nevertheless be interpreted cautiously. Both target variables were generated algorithmically: VADER provided sentiment labels, and NMF, along with manual topic interpretation, provided aspect labels. Test accuracy, therefore, measures how effectively the SVM reproduces those labels, not definitive agreement with independently annotated human judgements. Reviews from the same platforms and period may also exhibit language patterns that overlap across the training and test sets. Sarcasm, mixed sentiment, indirect criticism, and multi-aspect reviews remain difficult for TF-IDF representations to capture.

CONCLUSION AND FUTURE WORK

This paper presented Gigalyze, an ABSA web application for 30,000 Malaysian-market Google Play reviews of Fiverr, Freelancer, and Upwork. The system combines structured preprocessing, NMF-assisted aspect labelling, VADER-assisted sentiment labelling, TF-IDF, linear SVM, SMOTE, probability calibration, and interactive visualisation. The final sentiment and aspect classifiers achieved 83.74% and 88.52% accuracy, respectively. SMOTE substantially improved neutral recall and produced a smaller improvement for the underrepresented User Experience aspect. All functional tests passed.

Future work should evaluate the models on a stratified, independently human-annotated test set and report inter-annotator agreement. Multi-label aspect classification would better represent reviews that discuss several platform issues. Contextual models such as BERT or sentence transformers, alongside dependency parsing and targeted feature engineering, should be compared under identical splits. Temporal holdout tests, multilingual Malaysian reviews, broader platform coverage, and real-time ingestion would provide stronger evidence of generalizability and deployment readiness.

REFERENCES

1. Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A. (2023). Topic modeling algorithms and applications: A survey. *Information Systems*, 112, 102131. <https://doi.org/10.1016/j.is.2022.102131>
2. Ali, S., Wang, G., & Riaz, S. (2020). Aspect based sentiment analysis of ridesharing platform reviews for Kansei engineering. *IEEE Access*, 8, 173186-173196. <https://doi.org/10.1109/ACCESS.2020.3025823>
3. Brauwiers, G., & Frasincar, F. (2022). A survey on aspect-based sentiment classification. *ACM Computing Surveys*, 55(4). <https://doi.org/10.1145/3503044>
4. Budaya, I. G. B. A., & Suniantara, I. K. P. (2024). Comparison of sentiment analysis algorithms with SMOTE oversampling and TF-IDF implementation on Google Reviews for public health centers. *MALCOM*, 4(3), 1077-1086. <https://doi.org/10.57152/malcom.v4i3.1459>
5. Cervantes, J., Garcia-Lamont, F., Rodriguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215. <https://doi.org/10.1016/j.neucom.2019.10.118>
6. Dedema, M., & Rosenbaum, H. (2024). Socio-technical issues in the platform-mediated gig economy: A systematic literature review. *Journal of the Association for Information Science and Technology*, 75(3), 344-374. <https://doi.org/10.1002/asi.24868>
7. Green, D. D., McCann, J., Vu, T., Lopez, N., & Ouattara, S. (2018a). Gig economy and the future of work: A Fiverr.com case study. *Management and Economics Research Journal*, 4(2), 281. <https://doi.org/10.18639/MERJ.2018.04.734348>
8. Green, D. D., Walker, C., Alabulththim, A., Smith, D., & Phillips, M. (2018b). Fueling the gig economy: A case study evaluation of Upwork.com. *Management and Economics Research Journal*, 4, 104. <https://doi.org/10.18639/MERJ.2018.04.523634>
9. Green, D. D., Polk, X., O'Donnell, H., Doughty, K., Carr, M., & Costa-Cargill, D. (2021). The gig economy: A case study analysis of Freelancer.com. *Management and Economics Research Journal*, 7(2), 1-6. <https://doi.org/10.18639/merj.2021.1413412>
10. Harun, N., Ali, N. M., & Khan, N. L. M. A. (2020). An experimental measure of Malaysia's gig workers using labour force survey. *Statistical Journal of the IAOS*, 36(4), 969-977. <https://doi.org/10.3233/SJI-200749>

11. 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>
12. Indrawan, N. A., Sucahyo, Y. G., Ruldeviyani, Y., & Gandhi, A. (2020). What users want for gig economy platforms: Sentiment analysis approach. 2020 6th International Conference on Science in Information Technology, 68-73. <https://doi.org/10.1109/ICSITech49800.2020.9392060>
13. Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 13.
14. Muhyi, S. N. A., Omar, S. I., & Adnan, S. F. (2023). The drivers force the gig workers into gig economy: The case of Malaysia. *Journal of Business and Social Development*, 11(1), 1-12. <https://doi.org/10.46754/jbsd.2023.03.001>
15. Mustakim, H., & Priyanta, S. (2022). Aspect-based sentiment analysis of KAI Access reviews using NBC and SVM. *IJCCS*, 16(2), 113. <https://doi.org/10.22146/ijccs.68903>
16. Nurul Hidayati, S., Hamami, F., & Fa'rifah, R. Y. (2023). Aspect-based sentiment analysis on FLIP application reviews using Support Vector Machine. *Journal of Informatics and Telecommunication Engineering*, 7(1), 183-197. <https://doi.org/10.31289/jite.v7i1.9768>
17. Pilatti, G. R., Pinheiro, F. L., & Montini, A. A. (2024). Systematic literature review on gig economy: Power dynamics, worker autonomy, and the role of social networks. *Administrative Sciences*, 14(10), 267. <https://doi.org/10.3390/admsci14100267>