iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

HYBRID EXPLAINABLE SENTIMENT ANALYSIS FRAMEWORK

Vidyadhari S
MTech. Information Technology, Andhra University College of Engineering (A),
Andhra University, Visakhapatnam, India

Under the Guidance of: **Professor B. Prajna**

ABSTRACT
Sentiment analysis is widely used to identify opinions expressed in user-generated text, but conventional single-model approaches may provide limited robustness and limited interpretability. This paper presents a hybrid and explainable aspect-based sentiment analysis framework that integrates TF-IDF feature representation, a Linear Support Vector Machine (SVM), VADER sentiment analysis, DistilBERT-based sentiment inference, confidence scoring, SHAP-based explanation, and an interactive Streamlit dashboard. The system processes restaurant-domain aspect annotations from the SemEval-2014 Task 4 ABSA dataset. Each instance combines the target aspect and sentence context using an aspect-aware representation. TF-IDF with unigram and bigram features forms the primary supervised representation, while VADER and DistilBERT provide complementary sentiment signals. A weighted fusion mechanism gives the major contribution to the SVM prediction and smaller contributions to the other models. SHAP is used as a post-hoc explanation layer for the linear model, while the dashboard presents aspect-level sentiment, confidence information, analytics, word clouds, and downloadable results. The reported project evaluation obtained 63.69% for VADER and 93.28% for TF-IDF with SVM, with the hybrid configuration also reported at 93.28%. These figures are presented as reported project evaluation results rather than independent held-out test performance. The framework demonstrates an end-to-end practical approach for combining prediction, confidence information, explainability, and visualization in restaurant-review analysis.

Keywords:
Aspect-Based Sentiment Analysis, TF-IDF, SVM, VADER, SHAP, Explainable AI

1. INTRODUCTION
The rapid growth of online reviews has created a large volume of textual feedback that can be analyzed to understand customer opinions. Sentiment analysis attempts to determine whether a text expresses a positive, negative, or neutral opinion. Aspect-Based Sentiment Analysis (ABSA) extends conventional sentiment analysis by associating the sentiment with a particular aspect, such as food, service, price, ambience, or cleanliness.
This aspect-level view is important for restaurant reviews because one sentence can discuss several targets with different opinions. A review may praise the food while complaining about the service, for instance. Assigning one polarity to the entire sentence can hide this distinction.
Another challenge is interpretability. A prediction can be difficult to trust when the user cannot understand why the model produced it. Explainable Artificial Intelligence (XAI) techniques can provide information about the textual features contributing to a prediction. Confidence information can additionally communicate the relative strength of the selected class, although such a score should not automatically be interpreted as a calibrated probability.
This work proposes a hybrid explainable ABSA framework combining a supervised TF-IDF and Linear SVM classifier, a lexicon-based VADER model, and a pretrained DistilBERT sentiment model. Their outputs are combined using manually specified weighted voting. The system also generates a confidence score, applies SHAP to the linear SVM component, and provides an interactive Streamlit dashboard.

2. RELATED WORK
Traditional sentiment analysis commonly uses machine-learning representations such as bag-of-words and TF-IDF with classifiers including Naïve Bayes, logistic regression, and support vector machines. These approaches are computationally efficient and can provide strong performance when sufficient labeled training data are available.

Lexicon-based methods such as VADER use sentiment-related lexical information and rules to estimate polarity. They are lightweight and useful when rapid inference is required, but their general-purpose lexicons may not capture all domain-specific expressions.
Transformer-based language models have improved contextual text representation. DistilBERT provides a smaller transformer architecture suitable for practical inference compared with larger BERT variants. However, a pretrained sentiment model may not be specifically optimized for restaurant-domain three-class ABSA.
Hybrid approaches can combine complementary signals from different modeling paradigms. In the proposed framework, SVM supplies the dominant supervised signal, VADER contributes a lexical signal, and DistilBERT supplies contextual transformer information. SHAP is then used to add an explanation layer.

3. PROBLEM STATEMENT

A number of sentiment-analysis applications report one polarity for an entire review or sentence, which can overlook opinions attached to individual aspects. Classical machine-learning models may also have limited contextual representation, whereas transformer inference can be more resource intensive. A further issue is that a predicted label alone does not show which textual evidence influenced the decision. This work addresses these concerns by developing an aspect-aware hybrid framework that combines several sentiment signals and presents the output together with confidence-like information and SHAP explanations.

4. OBJECTIVES

- To preprocess restaurant-review sentences and retain aspect-level annotations.
- To construct aspect-aware text inputs by combining aspect terms with sentence context.
- To train a TF-IDF based Linear SVM sentiment classifier.
- To obtain complementary sentiment predictions using VADER and DistilBERT.
- To combine model outputs using a weighted fusion strategy.
- To compute a confidence score for the selected hybrid prediction.
- To generate SHAP explanations for the linear SVM component.
- To provide an interactive Streamlit dashboard for analysis and result download.

5. DATASET AND PREPROCESSING

The supervised training data use the restaurant portion of the SemEval-2014 Task 4 ABSA dataset, whose annotated XML records provide sentence text, aspect terms, and sentiment polarity. A separate Phase-B XML file supplied for this project contains unseen restaurant-review sentences, aspect information, and aspect categories, but does not contain gold sentiment polarity labels. The implementation parses both sources and organizes records around sentence, aspect, category, and sentiment fields when sentiment annotations are available.
During preprocessing, text is converted to lowercase and characters outside the selected alphanumeric and punctuation set are removed. The target is made explicit by constructing each input in the form [ASPECT] aspect_term [TEXT] sentence. This keeps the aspect identifiable while retaining its sentence context.
The implementation uses a labeled training XML input and an unseen Phase-B XML input for inference. TF-IDF is fitted only on the labeled training representation and then used to transform the unseen Phase-B representation. The vectorizer is configured for up to 10,000 features, unigram and bigram features, English stop-word removal, minimum document frequency of 2, and sublinear term-frequency scaling.

6. PROPOSED METHODOLOGY

6.1 Overall Architecture

The processing flow starts with a review or an annotated dataset record. After preprocessing, the aspect and its sentence context are joined into the aspect-aware input. This input is handled by VADER, TF-IDF with Linear SVM, and DistilBERT. Their predictions are then combined using the specified fusion weights. The selected class receives a confidence-like score, and SHAP is applied to the SVM component. The resulting information is displayed through Streamlit.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

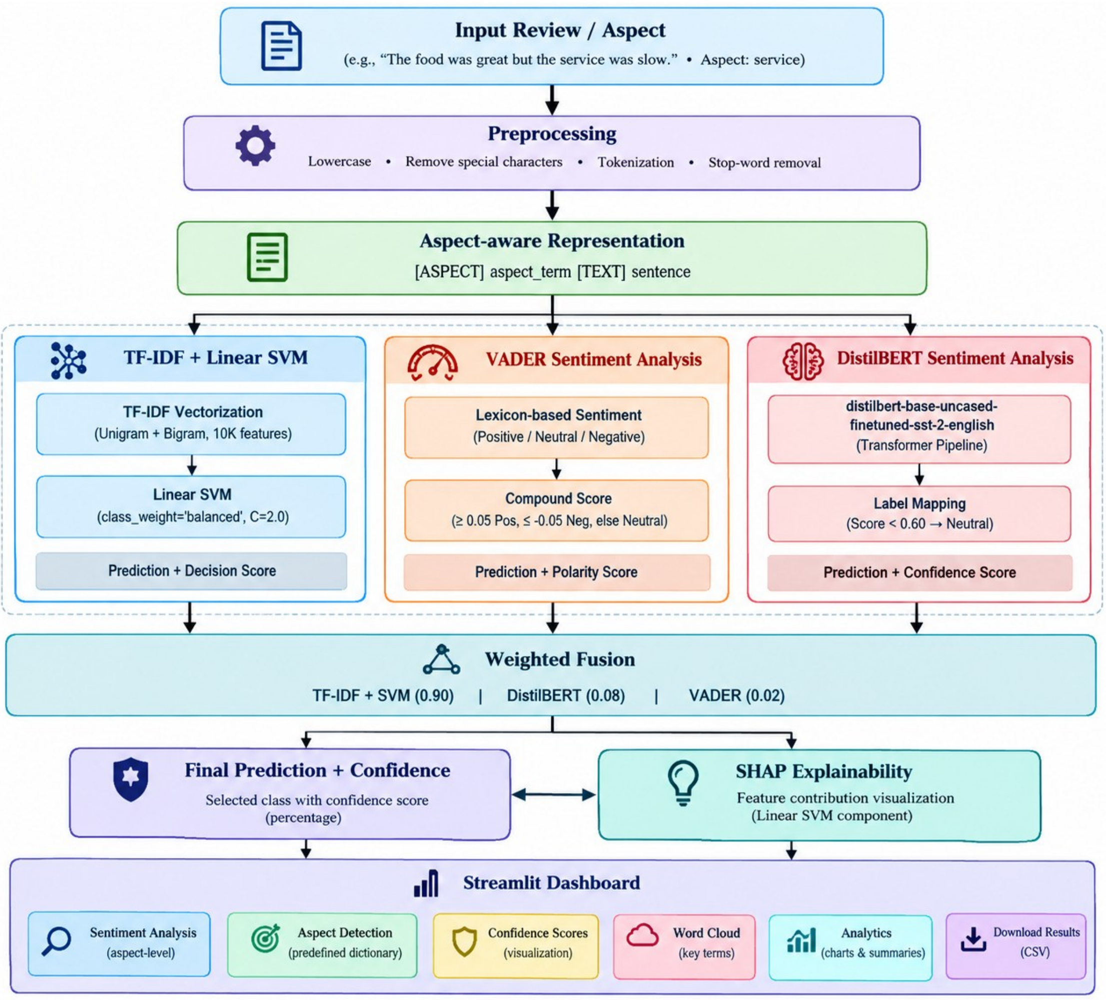

Figure 1. Overall archiecture of the proposed hybrid explainable sentiment analysis framework.

6.2 TF-IDF Representation

Term Frequency–Inverse Document Frequency (TF-IDF) converts text into numerical features by considering how informative a term is within a document and across the collection. The implementation includes both single words and two-word sequences. Sublinear term-frequency scaling is used so that repeated occurrences of the same term have a reduced incremental effect.

6.3 Linear Support Vector Machine

A LinearSVC classifier is trained on the TF-IDF feature matrix. The implementation uses class_weight='balanced' and C=2.0. The SVM acts as the principal supervised component of the hybrid framework. Its decision function is also used to derive a normalized confidence-like score for the SVM component.

6.4 VADER Sentiment Analysis

VADER contributes a rule- and lexicon-based sentiment signal through positive, neutral, negative, and compound scores. In the implementation, compound values of at least 0.05 are assigned positive, values of at most -0.05 are assigned negative, and values between these limits are assigned neutral. The polarity information is also included in the hybrid confidence calculation.

6.5 DistilBERT Sentiment Analysis

The implementation uses the distilbert-base-uncased-finetuned-sst-2-english sentiment pipeline as an additional sentiment signal. Because the pretrained model is binary-oriented, scores below 0.60 are mapped to a neutral outcome in the implementation; otherwise the POSITIVE or NEGATIVE label is retained. This provides a simple neutral-handling mechanism within the project framework.

ISSN: 2456-9348 (Online)
SJIF Impact Factor: 8.232

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

6.6 Weighted Fusion

The three model outputs are combined through weighted voting. TF-IDF with Linear SVM receives 0.90, DistilBERT receives 0.08, and VADER receives 0.02. For each prediction, the assigned weight is added to the class selected by that component, and the class with the greatest accumulated value becomes the final sentiment. The displayed confidence is that selected-class total multiplied by 100.

Table 1. Fusion weights used in the proposed framework.

Component	Fusion Weight
TF-IDF + Linear SVM	0.90
DistilBERT	0.08
VADER	0.02

6.7 Confidence Score

The confidence value is included to show the relative strength of the class selected by the fusion process. It is calculated from the accumulated fusion score and displayed in percentage form. Because it is not statistically calibrated, it should be regarded as a confidence-like application score rather than a probability.

6.8 SHAP Explainability

SHAP is incorporated as a post-hoc explanation technique for the linear SVM component. To control memory requirements, the implementation uses a small sample of ten training instances and limits the displayed feature space to the first 100 TF-IDF features. SHAP LinearExplainer is applied to the trained SVM, and the resulting explanation is saved as a visualization. SHAP does not alter the model prediction; it explains feature contributions.

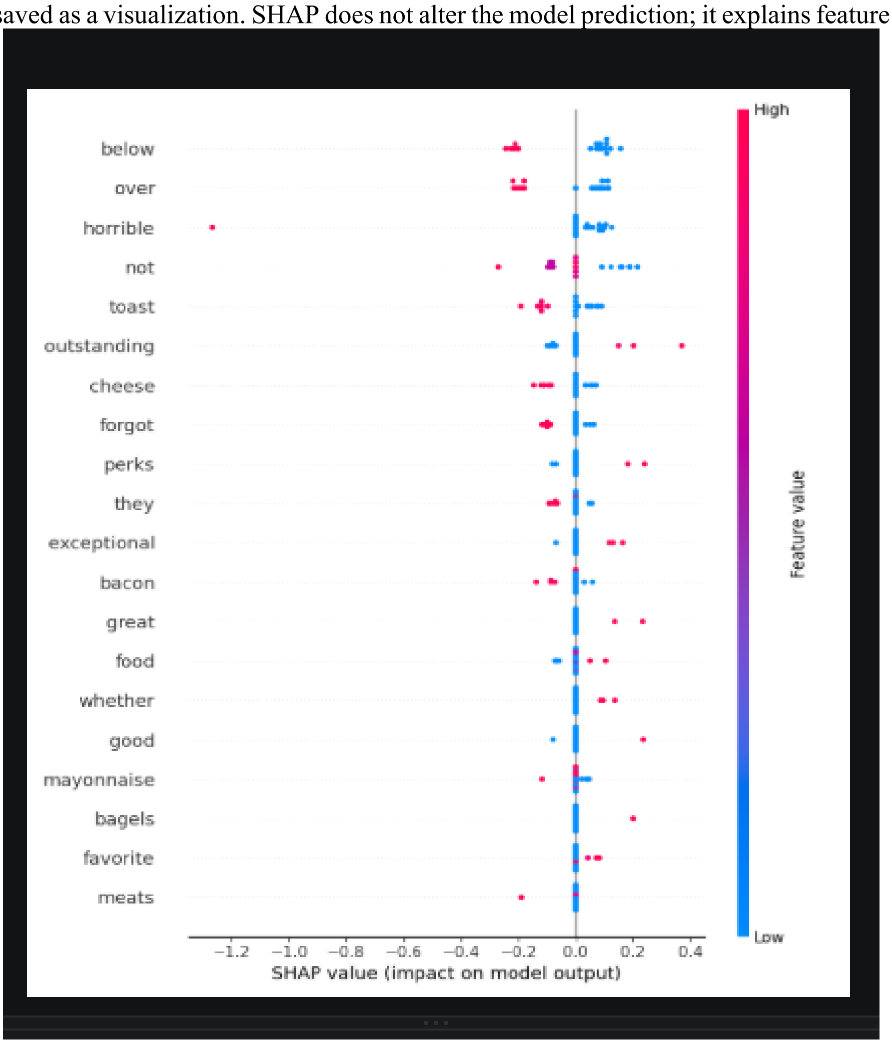

Figure 2. SHAP summary plot showing feature contributions to the Linear SVM sentiment prediction.

Illustrates the SHAP summary plot generated for the Linear SVM component.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

The horizontal axis represents the SHAP value, indicating the direction and magnitude of each feature's contribution to the model output. Features positioned toward the positive side contribute more strongly toward the corresponding model output, while features on the negative side contribute in the opposite direction. The color scale represents the relative feature value. This visualization provides an interpretable view of how individual TF-IDF features influence the model prediction.

6.9 Streamlit Dashboard

The Streamlit dashboard provides an interactive interface for restaurant-review analysis. Its views include Sentiment Analysis, Aspect Detection, Confidence Scores, Word Cloud, Analytics, and Download Results. A predefined aspect dictionary identifies common restaurant aspects including food, service, price, ambience, and cleanliness. For each detected aspect, the hybrid prediction function produces sentiment and confidence information.

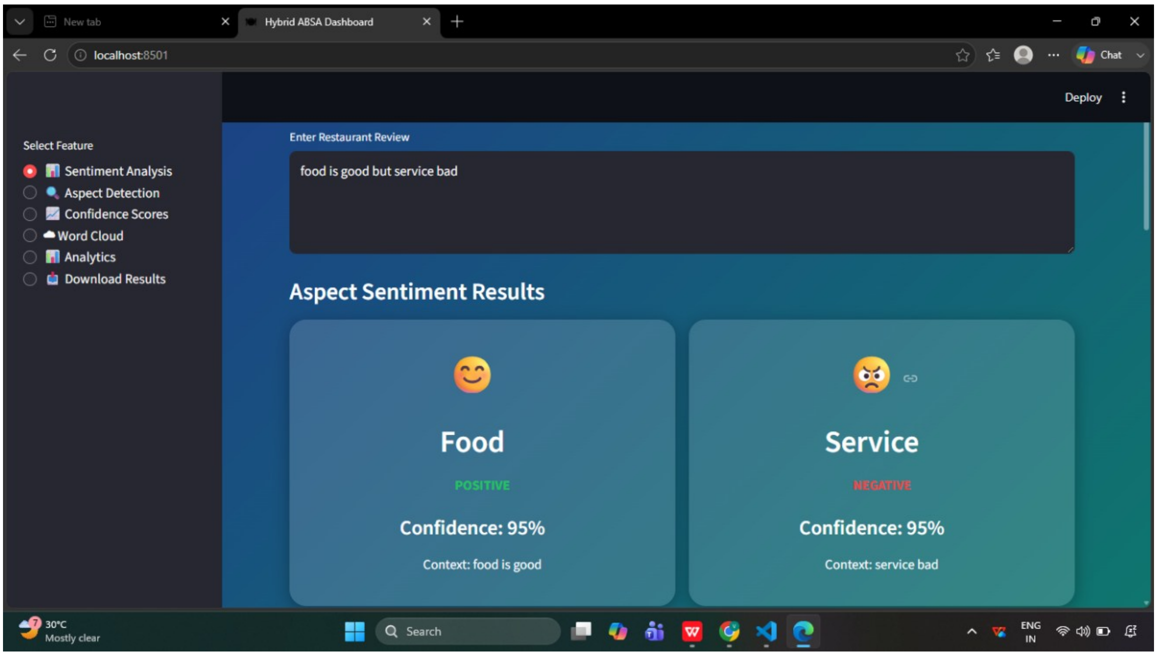

Figure 3.Streamlit Dashboard interface

The analytics view summarizes positive, negative, and neutral confidence values and provides graphical summaries. The application also supports a detected-aspect table and CSV download functionality, allowing the model output to be used in a practical analytical workflow.

7. IMPLEMENTATION

The system is implemented in Python using standard data-science and NLP libraries. Pandas is used for tabular data handling, scikit-learn provides TF-IDF and LinearSVC, VADER supplies lexicon-based sentiment analysis, the Transformers pipeline provides DistilBERT inference, SHAP provides explainability, and Streamlit provides the interactive interface.

The implementation is organized as an end-to-end pipeline: XML parsing, preprocessing, aspect-aware representation, model training or inference, prediction fusion, confidence calculation, explanation generation, and dashboard presentation. The design emphasizes practical integration of heterogeneous sentiment signals rather than dependence on a single algorithm.

8. RESULTS AND EVALUATION

The project evaluation records 63.69% for VADER, 93.28% for TF-IDF with SVM, and 93.28% for the hybrid configuration. These figures are retained as reported project evaluation results. Because the available metric calculation uses the training DataFrame, they are reported as training-set/project evaluation results and are not claimed as independently validated held-out test accuracy.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

Table 2. Reported training-set/project evaluation results.

Method	Reported Accuracy
VADER	63.69%
TF-IDF + Linear SVM	93.28%
Hybrid Fusion	93.28%

The reported figures show that TF-IDF with SVM supplies the main predictive contribution in the present implementation. The hybrid and SVM values are equal in the reported evaluation, which is consistent with the dominant 0.90 weight assigned to the SVM component. VADER and DistilBERT still provide additional signals and can affect individual decisions when their predicted classes agree with or differ from the SVM output.

The explainability component addresses a different requirement from predictive accuracy. SHAP provides feature-level evidence for the SVM decision, while the confidence score communicates the strength of the fused class. The dashboard then makes these outputs accessible to users through visualizations and downloadable results.

8.3 Unseen Phase-B Data Analysis

To complement the training-set/project evaluation, the supplied Restaurants_Test_Data_PhaseB.xml file was processed as an unseen inference dataset. The file contains 800 unique restaurant-review sentences with aspect information, but it does not provide gold sentiment polarity labels. Therefore, supervised evaluation metrics cannot be calculated for this dataset without additional annotation.

Aspect-aware processing generated 1,328 prediction records from the 800 sentences because some sentences contain more than one aspect. The trained TF-IDF and Linear SVM classifier was applied to these unseen records using the same aspect-aware representation used during training.

The resulting prediction distribution was 857 positive records (64.53%), 316 negative records (23.80%), and 155 neutral records (11.67%). These values are descriptive model-output statistics and must not be interpreted as accuracy, precision, recall, F1-score, or ground-truth sentiment proportions.

Predicted sentiment	Count	Percentage
Positive	857	64.53%
Negative	316	23.80%
Neutral	155	11.67%
Total	1,328	100.00%

Table 3. Unseen Phase-B prediction distribution.

The unseen-data analysis indicates that the trained classifier assigns the largest share of Phase-B aspect records to the positive class. This analysis demonstrates inference on data not used for the training-set metric calculation and provides an additional practical validation of the deployment pipeline, while preserving the distinction between prediction distribution and labeled-test performance.

9. DISCUSSION

The reported training-set/project evaluation shows that TF-IDF with Linear SVM provides the dominant predictive signal in the current implementation. The hybrid configuration has the same reported accuracy because the SVM component receives the largest fusion weight. VADER and DistilBERT therefore act mainly as complementary signals in the present design. The practical value of the framework also comes from the explainability and dashboard layers: SHAP provides feature-level evidence for the SVM component, while the confidence-like score and Streamlit interface make the results easier to inspect. However, the reported accuracy should not be interpreted as independent generalization performance because the current evaluation code calculates the metrics using the training DataFrame. The available Phase-B file is useful for unseen-data inference, but it contains no gold polarity labels and therefore cannot support accuracy, precision, recall, or F1 calculations. A future labeled test evaluation should report these metrics together with class-wise results and a confusion matrix.

10. ADVANTAGES

- Combines lexical, classical machine-learning, and transformer-based sentiment signals.
- Provides aspect-aware sentiment analysis rather than only overall review polarity.
- Uses weighted fusion to integrate heterogeneous predictions.
- Provides confidence-like information with the final prediction.
- Adds SHAP-based post-hoc explainability.
- Provides an interactive Streamlit interface for practical analysis.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

- Supports visual analytics and downloadable results.

11. LIMITATIONS

- The reported training-set/project evaluation is not an independent held-out evaluation; the Phase-B file contains no gold sentiment labels.
- The dashboard aspect detector uses a predefined keyword dictionary and may miss unseen expressions or synonyms.
- VADER is a general-purpose lexicon and may not capture all restaurant-domain language.
- DistilBERT is pretrained for SST-2 and is not specifically fine-tuned for three-class restaurant ABSA.
- The fusion weights are manually specified rather than learned from validation data.
- SHAP is computed on a small sample and reduced feature space to control memory usage.
- The dashboard's overall rating is rule-based and should be treated as an application-level summary.
-

12. FUTURE SCOPE

Future work can investigate learned fusion weights, domain-specific transformer fine-tuning, stronger aspect extraction using sequence labeling or transformer-based ABSA models, calibration of confidence scores, and evaluation on multiple public datasets. Further experiments can report macro-F1, class-wise precision and recall, robustness to negation and mixed sentiment, and computational cost. SHAP explanations can also be expanded using optimized sampling strategies when sufficient hardware is available.

13. CONCLUSION

This work developed a hybrid, explainable ABSA prototype that combines TF-IDF, Linear SVM, VADER, DistilBERT, weighted fusion, confidence scoring, SHAP, and a Streamlit dashboard. The design brings together supervised machine learning, lexicon-based sentiment analysis, and transformer inference, followed by an explanation layer and an interactive interface. The reported project evaluation results are 63.69% for VADER and 93.28% for both TF-IDF with SVM and the hybrid configuration. In addition, the unseen Phase-B analysis produced a prediction distribution of 857 positive, 316 negative, and 155 neutral records from 1,328 aspect-level prediction records. Because Phase-B contains no gold polarity labels, these counts describe model outputs rather than test accuracy. The implementation demonstrates how prediction, explainability, confidence information, and visualization can be integrated for aspect-level restaurant-review analysis.

EVALUATION NOTE

The Phase-B XML used in the unseen-data analysis was supplied for project inference and contains sentence/aspect information without gold sentiment polarity labels. Accordingly, the Phase-B results are presented as prediction distributions rather than accuracy-based test results. The reported 93.28% value remains explicitly identified as training-set/project evaluation.

REFERENCES

1) M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 27–35, 2014. doi: 10.3115/v1/S14-2004.
2) C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, 2014. doi: 10.1609/icwsm.v8i1.14550.
3) C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995. doi: 10.1007/BF00994018.
4) T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in *Machine Learning: ECML-98*, pp. 137–142, Springer, 1998. doi: 10.1007/BFB0026683.
5) J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019. doi: 10.18653/v1/N19-1423.
6) V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2019.
7) S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, pp. 4765–4774, 2017.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
https://ijetrm.com/

8) M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016. doi: 10.1145/2939672.2939778.
9) F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, J. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
10) W. McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2017.
11) S. Raschka, Y. Liu, and V. Mirjalili, *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*. Birmingham, UK: Packt Publishing, 2022.
12) Streamlit, "Streamlit Documentation." [Online]. Available: https://docs.streamlit.io/

AUTHOR INFORMATION

Author: VIDYADHARI S
M.Tech. Information Technology
Andhra University College of Engineering (A)
Andhra University
Guide: Prof. B. Prajna

