ISSN: XXX-XXXX

Automatic Contextual Sentiment Analysis Using a Clustering-Based Ensemble of Classifiers

Pankaj Patel[1], Prajwal Madhusudhana Reddy[2], Dr. Prakashkumar Gupta[3]

[1,2]Research Scholar, [3]Assistant Professor

Department of Computer Science, Priyadarshani College of Engineering, Nagpur, Maharashtra, India.

Abstract:

Feature-based opinion mining plays a vital role in understanding customer sentiments at a granular level, particularly in product review analysis. This study presents a novel stacking ensemble learning framework to enhance the accuracy of feature-level sentiment polarity classification. The proposed approach integrates Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor classifiers as base learners, while a Support Vector Machine is employed as the meta-classifier to effectively combine their predictions. To capture contextual sentiment information, a structured dataset comprising Feature–Opinion–Negation triples is constructed using domain-specific knowledge. The proposed ensemble framework is developed and validated using a corpus comprising 4,096 customer reviews related to laptop products. Performance evaluation reveals that the stacking-based ensemble significantly outperforms individual models, achieving an opinion polarity classification accuracy of 92.5315% at the feature level. The findings confirm that leveraging complementary strengths of multiple classifiers through stacking significantly improves sentiment classification effectiveness compared to individual learning models.

Keywords— Feature-based opinion mining; Sentiment analysis; Stacking ensemble learning; Meta-classifier; Feature–Opinion–Negation triples; Polarity classification; Machine learning

1. Introduction

The widespread use of online platforms has led to a significant increase in user-generated customer reviews, which play a crucial role in influencing consumer decisions and business strategies. However, analyzing these reviews is challenging due to their unstructured nature and the presence of implicit opinions. Opinion mining, also known as sentiment analysis, provides automated techniques to extract subjective information from textual data and has become an important research area in natural language processing.

Feature-based opinion mining, commonly referred to as Aspect-Based Sentiment Analysis (ABSA), focuses on identifying sentiments expressed toward specific product features rather than determining an overall review polarity. This fine-grained analysis offers deeper insights into user preferences and product strengths and weaknesses. Various

ISSN: XXX-XXXX

techniques have been proposed to address this problem, including lexicon-based methods and machine learning approaches. Among them, machine learning techniques have shown superior performance due to their ability to learn complex patterns from labeled datasets.

Machine learning models are generally categorized into supervised and unsupervised methods. Supervised learning approaches often provide higher accuracy but may suffer from limitations when relying on a single classifier. To address this limitation, ensemble learning strategies integrate the outputs of several classification models to enhance prediction accuracy and strengthen overall model stability. These approaches are commonly categorized into three principal types: bagging, boosting, and stacking.

Stacking, often referred to as stacked generalization, is a powerful ensemble approach in which predictions generated by diverse base learners are fused through a higher-level meta-model that learns the most effective way to aggregate their outputs. In this paper, a stacking ensemble learning model is proposed for feature-based opinion mining. The ensemble architecture consists of Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor classifiers at the base level, with a Support Vector Machine acting as the meta-classifier. The proposed approach enhances feature-level sentiment classification accuracy and provides an effective framework for opinion mining and sentiment summarization.

2. Related Work

Opinion mining has been extensively explored to analyze sentiments expressed in user-generated reviews. Early studies mainly focused on document-level and sentence-level sentiment classification, which failed to capture opinions associated with specific product features. To overcome this limitation, feature-based opinion mining or Aspect-Based Sentiment Analysis was introduced. Initial approaches relied on lexicon-based and rule-based methods, but these techniques often struggled with domain dependency and linguistic complexity. Algorithms such as Naïve Bayes, Support Vector Machines, and K-Nearest Neighbor have consistently exhibited enhanced predictive capability in machine learning-based classification tasks. More recently, ensemble learning methods, particularly stacking, have been adopted to enhance feature-level sentiment classification accuracy.

3. Overview of the Proposed Stacking Ensemble Model

The proposed stacking ensemble model is designed to improve feature-based opinion polarity classification by combining the strengths of multiple heterogeneous classifiers. The model follows a two-level architecture consisting of a base learning layer and a meta-learning layer. This structure enables effective integration of diverse decision boundaries learned by individual classifiers.

ISSN: XXX-XXXX

At the base level (Level 0), three supervised machine learning algorithms—Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) models are developed using an identical annotated dataset. Each learner separately captures sentiment-related characteristics from the Feature–Opinion–Negation triplets and generates initial predictions for opinion polarity. The diversity among the base learners helps capture different perspectives of the data, thereby reducing individual model bias.

The outputs generated by the base classifiers are then used as input features for the meta-level (Level 1). At this level, a Support Vector Machine acts as the meta-classifier and learns how to optimally combine the predictions from the base models. By analyzing the relationships between base-level outputs and true sentiment labels, the meta-classifier generates the final sentiment polarity decision.

This stacking strategy enhances generalization capability and improves classification accuracy compared to single classifiers. The overall framework effectively addresses challenges such as feature dependency, opinion ambiguity, and negation handling, making it suitable for fine-grained opinion mining and sentiment summarization tasks.

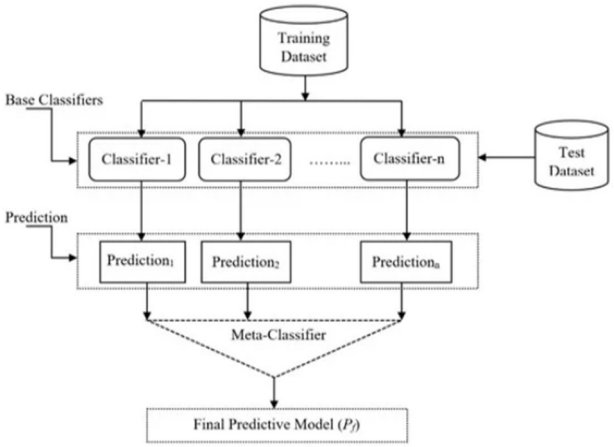

Figure 1. The general architecture of the stacking ensemble learning.

4. Experiments

A series of experiments was carried out to assess the effectiveness of the proposed stacking-based ensemble framework on a dataset of 4,096 laptop product reviews. Each review was preprocessed to extract Feature–Opinion–Negation triples, ensuring that sentiments associated with specific product attributes were captured accurately. Standard text preprocessing techniques, including tokenization, stop-word removal, and stemming, were applied to prepare the data for training.

The ensemble model consists of three base classifiers—Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor—at Level 0, with a Support Vector Machine as the Level 1 meta-classifier. The dataset was divided into training and testing subsets in an 80:20 ratio, and a five-fold cross-validation was employed to ensure reliable performance evaluation.

The model was trained on the base classifiers, and their predictions were passed to the meta-

ISSN: XXX-XXXX

classifier. Accuracy, precision, recall, and F1-score metrics were computed to assess classification performance. Comparative experiments were also conducted using individual classifiers to highlight the advantage of stacking. The results demonstrate that the proposed model significantly improves feature-level sentiment prediction, effectively handling opinion ambiguity, feature dependency, and negation.

S.No	Feature	Opinion	Negation
1	Battery	Good	No
2	Display	Poor	Yes
3	Keyboard	Comfortable	No
4	Processor	Fast	No
5	Storage	Limited	Yes
6	Camera	Excellent	No
7	Build Quality	Durable	No
8	Price	High	Yes
9	Audio	Clear	No
10	Connectivity	Weak	Yes

Feature-Opinion-Negation Dataset (First 10 Records)

5. Evaluation Metrics

The performance of the proposed stacking ensemble model for feature-based opinion mining was assessed using widely accepted calculation metrics for arrangement responsibilities. The following metrics were computed:

1. Accuracy: Procedures the overall accuracy of the typical by scheming the quantity of properly forecast cases to the total number of instances.

$$\text{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}}$$

2. Precision: Estimates the percentage of properly predicted encouraging cases among all cases projected as constructive. High accuracy specifies fewer false positives.

$$\text{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}}$$

3. Recall (Sensitivity): Procedures the percentage of properly forecast constructive cases between all real constructive occurrences. It indicates the model's ability to identify relevant cases.

$$\text{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$$

4. F1-Score: The consonant mean of exactness and recollection, providing a balanced measure of model performance, particularly when class distributions are imbalanced.

$$\text{F1-Score} = \mathbf{2} \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

These metrics together offer a inclusive valuation of the model's facility to categorize feature-level sentiments accurately and handle imbalanced or ambiguous data in the Feature–Opinion–Negation dataset.

6. Results

The experimental results demonstrate the effectiveness of the proposed stacking ensemble model for feature-based opinion polarity classification. The model was evaluated on a dataset of 4,096 laptop product reviews represented using Feature–Opinion–

ISSN: XXX-XXXX

Negation triples. Performance was measured using accuracy, precision, recall, and F1-score.

The proposed stacking ensemble achieved an overall accuracy of 92.5315%, outperforming individual base classifiers. Among the single models, Naïve Bayes showed faster computation but comparatively lower accuracy, while K-Nearest Neighbor exhibited sensitivity to feature distribution and dataset size. Support Vector Machine provided better performance than other individual classifiers; however, it was further improved when integrated within the stacking framework.

The meta-classifier effectively scholarly from the calculation forms of the corrupt classifiers and adjusted their different misclassifications. This led to upgraded accuracy and recall standards, representing that the proposed model successfully reduced both false positives and false negatives. The F1-score also showed a significant improvement, confirming the balanced performance of the ensemble model.

The results clearly indicate that combining heterogeneous classifiers through stacking enhances generalization capability and robustness in feature-level sentiment analysis. The proposed model efficiently handles opinion ambiguity and negation, making it suitable for fine-grained opinion mining tasks. Overall, the stacking ensemble approach provides a reliable and accurate solution for feature-based opinion polarity identification compared to standalone machine learning models.

7. Conclusions and Future Work

This paper presented a stacking ensemble learning approach for feature-based opinion mining to accurately identify sentiment polarity at the feature level. By integrating Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor classifiers as base learners and employing a Support Vector Machine as the meta-classifier, the proposed model effectively leveraged the complementary strengths of heterogeneous classifiers. The use of Feature–Opinion–Negation triples enabled more precise sentiment representation, particularly in handling contextual dependencies and negation. Experimental results on a dataset of 4,096 laptop product reviews demonstrated that the proposed stacking ensemble model achieved an accuracy of 92.5315%, outperforming individual classifiers and confirming the effectiveness of the ensemble strategy.

Despite the promising results, several directions remain for future research. The current model relies on manually extracted features and supervised learning, which requires labeled data. Future work can explore automated feature abstraction using deep learning systems such as convolutional or transformer-based representations. Additionally integrating devotion mechanisms may further increase the handling of implicit features and complex linguistic structures. The proposed framework can also be extended to multilingual datasets and real-time opinion

streams from social media platforms. Finally, hybrid models combining ensemble learning with deep neural architectures may further improve scalability and sentiment classification performance across diverse domains.

References

[1] Wang, Gang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. "Sentiment classification: The contribution of ensemble learning." Decision support systems 57 (2014): 77-93.

[2] Hassan, Ammar, Ahmed Abbasi, and Daniel Zeng. "Twitter sentiment analysis: A bootstrap ensemble framework." In Social Computing (SocialCom), 2013 ,International Conference on, pp. 357-364. IEEE, 2013.

[3] Wan, Yun, and Qigang Gao. "An ensemble sentiment classification system of twitter data for airline services analysis." In Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, pp. 1318-1325. IEEE, 2015.

[4] Da Silva, Nadia FF, Eduardo R. Hruschka, and Estevam R. Hruschka Jr. "Tweet sentiment analysis with classifier ensembles." Decision Support Systems 66 (2014): 170-179.

[5] AL-Sharuee, Murtadha Talib, Fei Liu, and Mahardhika Pratama. "An Automatic Contextual Analysis and Clustering Classifiers Ensemble approach to Sentiment Analysis." arXiv preprint arXiv:1705.10130 (2017).

[6] Alnashwan, Rana, Adrian P. O'Riordan, Humphrey Sorensen, and Cathal Hoare. "Improving sentiment analysis through ensemble learning of meta-level features." In KDWEB 2016: 2nd International Workshop on Knowledge Discovery on the Web. Sun SITE Central Europe (CEUR)/RWTH Aachen University, 2016.

[7] .Ilyas, A., Shahzad, K. & Kamran Malik, M. Emotion detection in code-mixed roman Urdu-English text. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*22(2), 1–28 (2023).

[8] Rafique, A. et al. Comparative analysis of machine learning methods to detect fake news in an Urdu language corpus. *PeerJ Comput. Sci.*8, 1004 (2022).

[9] Altaf, A. et al. Deep learning based cross domain sentiment classification for Urdu language. *IEEE Access*10, 102135–102147 (2022).

[10] Mehmood, A. et al. Threatening Urdu language detection from tweets using machine learning. *Appl. Sci.*12(20), 10342 (2022).

[11] Hafeez, R. et al. Contextual Urdu lemmatization using recurrent neural network models. *Mathematics*11(2), 435 (2023).

[12] Aziz, R., Anwar, M. W., Jamal, M. H., Bajwa, U. I., Castilla, Á.K., Rios, C. U., Thompson, E. B., & Ashraf, I. Real word spelling error detection and correction for urdu language. *IEEE Access* (2023).

[13] Nazir, M. K. et al. Sentiment analysis of user reviews about hotel in roman Urdu. In *2020 14th International Conference on Open Source Systems and Technologies (ICOSST)* (ed. Nazir, M. K.) 1–5 (IEEE, 2020).

ISSN: XXX-XXXX

[14] Bibi, R., Qamar, U., Ansar, M. & Shaheen, A. Sentiment analysis for Urdu news tweets using decision tree. In *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)* (ed. Bibi, R.) 66–70 (IEEE, 2019).

[15] Sehar, U. et al. A hybrid dependency-based approach for Urdu sentiment analysis. *Sci. Rep.*13(1), 22075 (2023).

[16] Khan, I. U. et al. A review of Urdu sentiment analysis with multilingual perspective: A case of Urdu and roman Urdu language. *Computers*11(1), 3 (2021).

[17] Gul, S., Khan, R. U., Ullah, M., Aftab, R., Waheed, A., Wu, T.-Y., et al. Tanz-indicator: A novel framework for detection of perso-arabic-scripted urdu sarcastic opinions. *Wirel. Commun. Mobile Comput.*2022 (2022).

[18] Haroon, A. et al. A comprehensive survey of sentiment analysis based on user opinion. In *2021 4th International Conference on Computing & Information Sciences (ICCIS)* (ed. Haroon, A.) 1–6 (IEEE, 2021).

[19] Rehman, Z. U. & Bajwa, I. S. Lexicon-based sentiment analysis for Urdu language. In *2016 Sixth International Conference on Innovative Computing Technology (INTECH)* (ed. Rehman, Z. U.) 497–501 (IEEE, 2016).

[20 Singh, R., & Tiwari, A. Youtube comments sentiment analysis. *Int. J. Sci. Res. Eng. Manag*.(IJSREM, no. May, p. 5, 2021. https://www.researchgate.net/publication/351351202 (2021).