

ABSTRACT

Sentiment classification has progressed from simple polarity detection to multilingual and cross-domain applications, yet domain shifts and linguistic variability remain major challenges to robust generalization. This paper presents a meta-fusion ensemble framework that integrates four transformer models—BERT, RoBERTa, DistilBERT, and XLM-RoBERTa—each fine-tuned on benchmark datasets including Sentiment140, IMDB, ABSA, and a multilingual corpus. Unlike static ensemble approaches such as majority voting or averaging, the proposed method employs a trainable Multi-Layer Perceptron (MLP) to dynamically fuse model logits, effectively capturing inter-model dependencies. Experiments conducted on a balanced evaluation set of 2,700 samples across diverse domains and languages demonstrate the effectiveness of this framework. The meta-fusion ensemble achieved an accuracy of 86.91% and a macro-F1 score of 85.67%, outperforming both individual transformer baselines and static fusion methods. These results confirm the advantage of learnable ensemble strategies for improving sentiment prediction under domain and language variability.