

As e-commerce platforms continue to grow, so has the importance of implementing customer reviews and feedback as tools for individuals to make informed decisions as consumers and to guide businesses in the development of strategic plans. However, traditional models like BERT introduce an additional layer of computational complexity in analyzing and classifying large amounts of reviews. This study developed a machine learning-based sentiment analysis model to classify customer reviews on the eBay platform into the categories of positive, negative, and neutral sentiments.

A dataset of over 45,000 reviews was captured from Kaggle, and the data was preprocessed to remove noise, inconsistencies, and any additional irrelevant information, then feature extraction was performed using DistilBERT, which is a lighter, more efficient alternative to BERT. To make the computation more efficient, as well to reduce dimensionality, Principal Component Analysis (PCA) was performed first, then two machine learning classifiers, Support Vector Machine (SVM) and Random Forest (RF), were implemented. The classifiers were evaluated based on standard metrics of accuracy, precision, recall, F1-score and ROC-AUC.

The experimental results showed that the SVM algorithm did better than Random Forest with an accuracy of 93.98%, precision of 93.84%, recall of 93.98%, F1-score of 92.27% and AUC score of 0.95. Random Forest only had an accuracy of 92.64% and an AUC of 0.91. The results showed that combining DistilBERT with PCA, greatly improved efficiency and classification capability over traditional BERT-based approaches and performed adequately well when handling neutral sentiments while also consuming resources at a lesser rate.

This work concludes that lightweight and powerful models such as DistilBERT can be successfully combined with PCA and SVM to build a solid and scalable paradigm for sentiment analysis of e-commerce reviews. The model enables companies to gain useful and insightful perspectives of customer satisfaction, builds trust in e-commerce or online platforms, and can guide improvements in product and service. Future work will focus on extending the model to multilingual datasets and utilize class balancing to improve performance in sentiment categories that are underrepresented.