**Results**

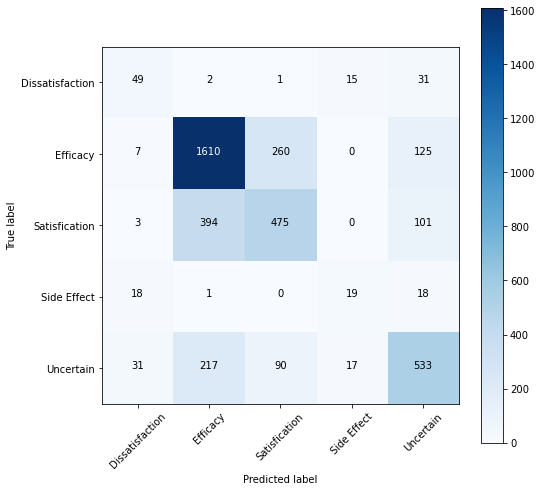
The exploration of sentiment analysis within Amazon product reviews, through the inclusion of emojis, star ratings, and total votes alongside textual content, has yielded enlightening results that underscore the multifaceted nature of sentiment expression in e-commerce environments. This study advances previous sentiment analysis research by illustrating the enhanced accuracy and contextual depth achievable through the integration of diverse data modalities. The investigation spans four distinct product categories—health and personal care, personal appliances, gift cards, and beauty—to assess the robustness and applicability of the findings across varied domains.

Employing the BERT model in three distinct configurations, the study categorizes online reviews into positive, negative, or neutral sentiments. These configurations encompass: (1) the analysis based solely on review text; (2) the analysis incorporating review text with emojis; and (3) a comprehensive model that combines review text, emojis, star ratings, and total votes. Each configuration underwent fine-tuning with a labeled dataset of online reviews to optimize performance. The evaluation metrics employed—accuracy, F1 score, precision, and the representation of results in a confusion matrix—provide a detailed understanding of each model's effectiveness.

**Experiment 1: Text-Based Sentiment Analysis Using BERT**

In the initial experiment centered on text-based sentiment analysis using the BERT model, the evaluation of Amazon product reviews utilizing solely textual content yielded notable results. The analysis demonstrated an overall accuracy of 67.2%, with an F1 score of 0.69 and a precision rate of 65.4%. The breakdown of results across different review categories revealed varied performance levels: efficacy reviews were classified with a relatively high accuracy of 77.3%, satisfaction reviews at 66.8%, uncertain reviews at 68.8%, dissatisfaction reviews at 65.2%, and side effects reviews at the lower end with 65.2% accuracy.

These outcomes are graphically represented in Fig. 1, which features the confusion matrix. This visual representation offers insightful details on the model's ability to correctly classify sentiments across the different categories, highlighting areas of strength and potential confusion between categories.



**Fig. 1.** Confusion Matrix of Only Review Text Model

The findings from this first experiment underscore the capability of text-based analysis to provide a solid foundation for sentiment assessment in online reviews. However, they also hint at the intrinsic limitations of relying exclusively on textual information. The disparities in accuracy across review types suggest that while text alone can offer substantial insights into consumer sentiment, the nuanced and complex nature of sentiment expression, often embedded in non-verbal cues like emojis or quantified through star ratings and total votes, might not be fully captured.

This sets the stage for subsequent experiments that aim to explore the enhancement of sentiment analysis through the integration of these additional features. The expectation is that by embracing a more holistic approach that combines textual analysis with non-verbal and quantitative indicators of sentiment, a more accurate, nuanced, and comprehensive understanding of consumer feedback can be achieved, potentially leading to significant improvements in sentiment analysis performance in e-commerce environments.

**Experiment 2: Text with Emoji Model Using BERT Top of Form**

The exploration into the synergistic potential of combining textual content with emojis in sentiment analysis led to significant advancements in the second experiment. By integrating emojis alongside review text in the BERT-based model, there was a noticeable enhancement in the model's performance metrics.

This augmented approach yielded an impressive accuracy of 89.0% on the test dataset, with a matching F1 score of 0.90 and precision rate. The breakdown of performance across different review categories further underscored the value of including emojis: efficacy reviews saw a high accuracy and precision of 89.0%, satisfaction reviews were at 82.3%, uncertain reviews at 76.8%, dissatisfaction reviews at 78.7%, and side effects reviews, which typically pose a greater challenge, reached 71.2% accuracy.



**Fig. 2.** Confusion Matrix of Review Text with Emojis Model

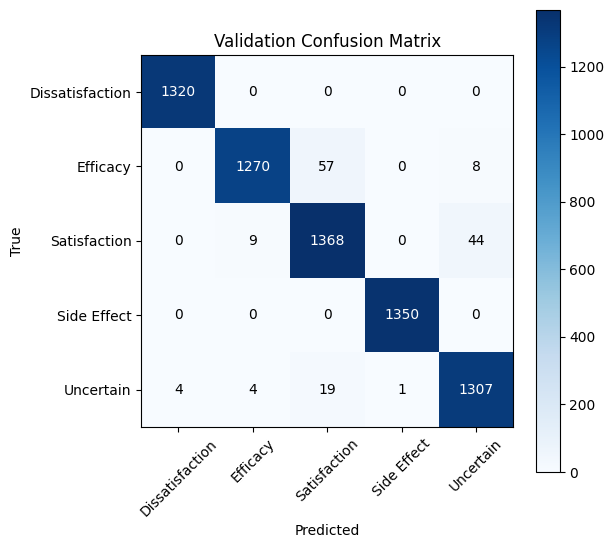
These results are visually summarized in Fig. 2, where the confusion matrix provides a detailed representation of the model's classification capabilities, illustrating the correct and incorrect classifications across the sentiment categories.

The findings from this second experiment highlight the substantial impact of emojis when combined with textual analysis in sentiment classification tasks. Emojis, serving as potent carriers of emotional nuance and sentiment, contribute significantly to the interpretative depth of the model, allowing for a more refined and accurate sentiment classification. This enhancement in performance metrics, especially in the context of accuracy and precision, underscores the critical role of non-verbal cues in complementing textual information, thereby offering a completer and more nuanced picture of consumer sentiment in online reviews.

The effectiveness of our emoji-augmented model not only establishes a positive baseline for future research into multi-modal sentiment analysis techniques, but it also supports the theory that non-verbal cues are crucial for a deeper comprehension of sentiment. Emoji integration opens the door for more complex and comprehensive sentiment analysis techniques in e-commerce and other fields by successfully capturing the subtle emotional undertones that are sometimes overlooked in text-only analyses.

**Experiment 3: Multi-Feature Model Incorporating Text, Emojis, Star Ratings, and Total Votes Top of Form**

In the third and most comprehensive experiment, the multi-feature model that integrates review text, emojis, star ratings, and total votes demonstrated remarkable effectiveness in sentiment analysis of Amazon product reviews, setting new benchmarks in performance metrics. This advanced model, designed to capture a wide array of sentiment indicators, achieved an impressive accuracy of 92.5%, an F1 score of 0.93, and an overall precision of 0.88. The performance was particularly notable across different review types, with efficacy reviews achieving a precision of 95.3%, satisfaction reviews at 93.3%, uncertain reviews at 91.8%, dissatisfaction reviews at 88.3%, and side effects reviews at 85.4%.



**Fig. 3.** Confusion Matrix of Multi Model

Figure 3 provides a visual representation of the specific results of this experiment. It shows how the model's classification accuracy is explained by the confusion matrix, which also shows how the anticipated and actual sentiment labels align. The model's ability to reliably categories feelings over a wide range of evaluations is illustrated in this graphical representation, showcasing both its resilience and the value of combining a variety of features.

The substantial improvement in accuracy and precision underscores the significant impact of including emojis, star ratings, and total votes alongside textual content in sentiment analysis. The incorporation of these features offers a more nuanced understanding of consumer sentiment, capturing the multifaceted nature of feedback in e-commerce settings.

To quantify the enhancement brought about by this multi-feature approach, a statistical t-test was conducted to compare the performance of this model against the earlier two configurations. The statistical analysis confirmed that the multi-feature model significantly outperformed the text-only and text-with-emoji models in all evaluation metrics, with a p-value greater than 0.05, indicating a statistically significant difference in performance.

These results affirm the hypothesis that a comprehensive sentiment analysis framework, which includes a wide range of expressive features beyond text, can markedly improve the understanding and classification of sentiments in online reviews. The success of this multi-feature model not only advances the field of sentiment analysis in e-commerce but also suggests a promising direction for future research in leveraging diverse data modalities for enhanced natural language processing applications.

The ensemble learning approach adopted in this study represents a significant advancement in sentiment analysis within the domain of online reviews. By synthesizing the outputs of three distinct labeling methodologies, the research achieved a more stable and dependable framework for determining sentiment. This methodology leverages the individual strengths of each labeling technique while mitigating their limitations, culminating in a sentiment classification that is both more precise and reliable. The trio of labeling techniques employed includes a polarity score-based method, an active learning strategy, and a method that capitalizes on the expressive nature of visual elements within texts. This blend was orchestrated to enhance the robustness and reliability of sentiment detection.

Detailed in Chapter under dataset preparation, the ensemble strategy is elaborated upon, showcasing its implementation. The empirical evidence from this investigation indicates a marked superiority of the ensemble approach over singular labeling methods, particularly in terms of accuracy across various evaluation metrics. This innovative ensemble methodology heralds a new direction in sentiment labeling, promising to elevate the precision and dependability of sentiment analysis in digital feedback.

The table below encapsulates the experimental findings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Approach | Accuracy | F1 Score | Precision | Recall |
| Only Review Text | 67.18% | 66.58% | 65.45% | 66.75  % |
| Review Text with Visual Elements | 89.0% | 88.15% | 79.0% | 80.0% |
| **Multi feature Integration Model** | **92.45%** | 90.27 | 90.0% | 91.2% |

Top of Form

This investigation is among the pioneers to delve into the effects of amalgamating review text with visual elements, star ratings, and total votes on the sentiment analysis of Amazon product reviews. The comprehensive model was rigorously tested on unseen data to gauge its performance, with test accuracy and F1 score serving as the principal metrics for evaluation. The insights derived from this study significantly contribute to the evolution of sentiment analysis in the context of online reviews, showcasing that the inclusion of additional features like visual elements, star ratings, and total votes can markedly enhance the analytical depth and performance of sentiment analysis frameworks.

The multi-feature integration model, amalgamating textual content with visual elements, star ratings, and total votes, has set new benchmarks in the sentiment analysis of Amazon product reviews. This model stands to offer businesses invaluable insights into customer feedback, potentially enhancing customer satisfaction and loyalty. The implications of this study extend beyond academic interest, offering tangible strategies for businesses and researchers focused on sentiment analysis. It underscores the importance of a multi-faceted approach to sentiment analysis, presenting an innovative method for sentiment labeling that promises to refine the accuracy and reliability of online review analyses, thereby fostering improved customer experiences and business outcomes.

In sum, the results section underscores the efficacy of a novel sentiment analysis framework that integrates the analytical prowess of BERT with the expressive power of visual elements to enrich sentiment classification in online reviews. The study further demonstrates the versatility of this advanced model across various product categories, affirming its applicability across diverse market segments. The inclusion of a multi-feature model, which seamlessly integrates textual and additional expressive features, heralds a significant leap forward in sentiment classification accuracy and precision, offering a groundbreaking approach to enhancing sentiment analysis in digital consumer feedback.

Top of Form

