

CAPSTONE PROJECT

SENTIMENT ANALYSIS AND OPINION MINING ON CUSTOMER REVIEWS USING NLP

(Kavya N – 23MIA1125, VIT Chennai)

EXECUTIVE SUMMARY

In today's digital environment, customers actively share their opinions through online reviews and social media platforms. These opinions strongly influence purchasing decisions and brand reputation. This project focuses on analysing customer reviews using Natural Language Processing (NLP) techniques to understand customer sentiment, identify recurring topics, and extract actionable business insights.

The project uses machine learning-based sentiment classification, exploratory text analysis, topic modelling, and trend analysis to convert unstructured textual data into meaningful insights. The results help organizations improve product quality, enhance customer engagement, and manage negative feedback effectively.

TEXT ANALYSIS & SENTIMENT INSIGHTS

Customer review data is unstructured and noisy, making preprocessing an essential step. The text data was cleaned by removing irrelevant elements and standardizing the content to improve model performance.

The following preprocessing steps were applied:

- Removal of HTML tags, punctuation, special characters, and numbers
- Conversion of text to lowercase
- Stop word removal
- Tokenization and lemmatization

Exploratory text analysis was conducted to understand customer opinions more clearly. Word cloud visualizations highlighted frequently occurring terms in positive, negative, and neutral reviews.

Key observations from text analysis:

- Positive reviews frequently mentioned quality, value, performance, and ease of use
- Negative reviews commonly focused on delivery delays, defects, poor service, and reliability issues
- Neutral reviews often contained factual or descriptive statements

N-gram analysis (bigrams and trigrams) revealed commonly repeated phrases, providing deeper insight into customer expectations and complaints. Additionally, analysis across star ratings showed a strong relationship between textual sentiment and numerical ratings, validating the effectiveness of sentiment analysis.

CLASSIFIER PERFORMANCE & FINDINGS

To enable sentiment classification, the cleaned text was transformed into numerical form using the TF-IDF vectorization technique, which captures the importance of words across the dataset.

Several supervised machine learning models were trained and evaluated:

- Logistic Regression
- Multinomial Naive Bayes
- Support Vector Machine (SVM)

The models were evaluated using:

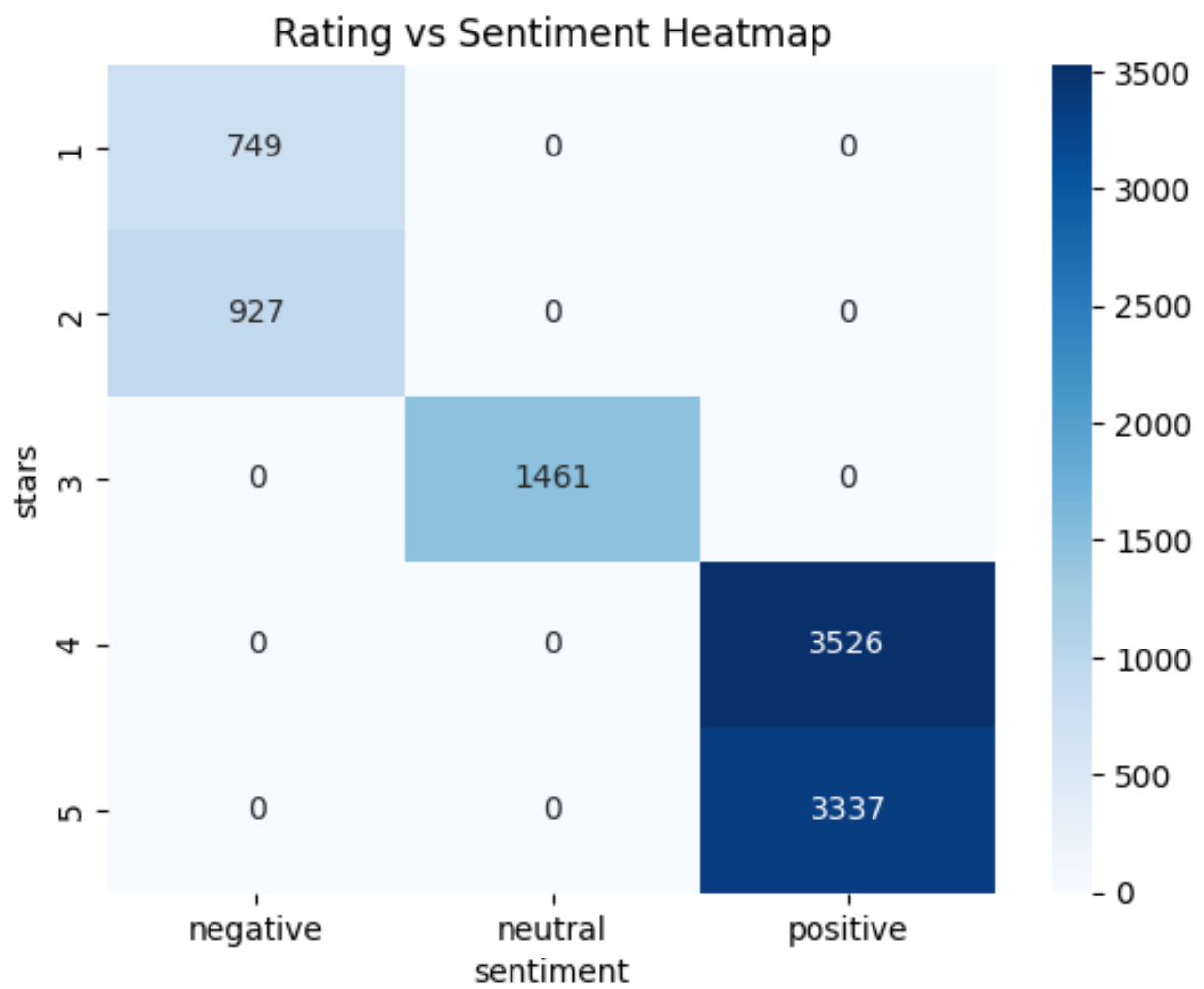
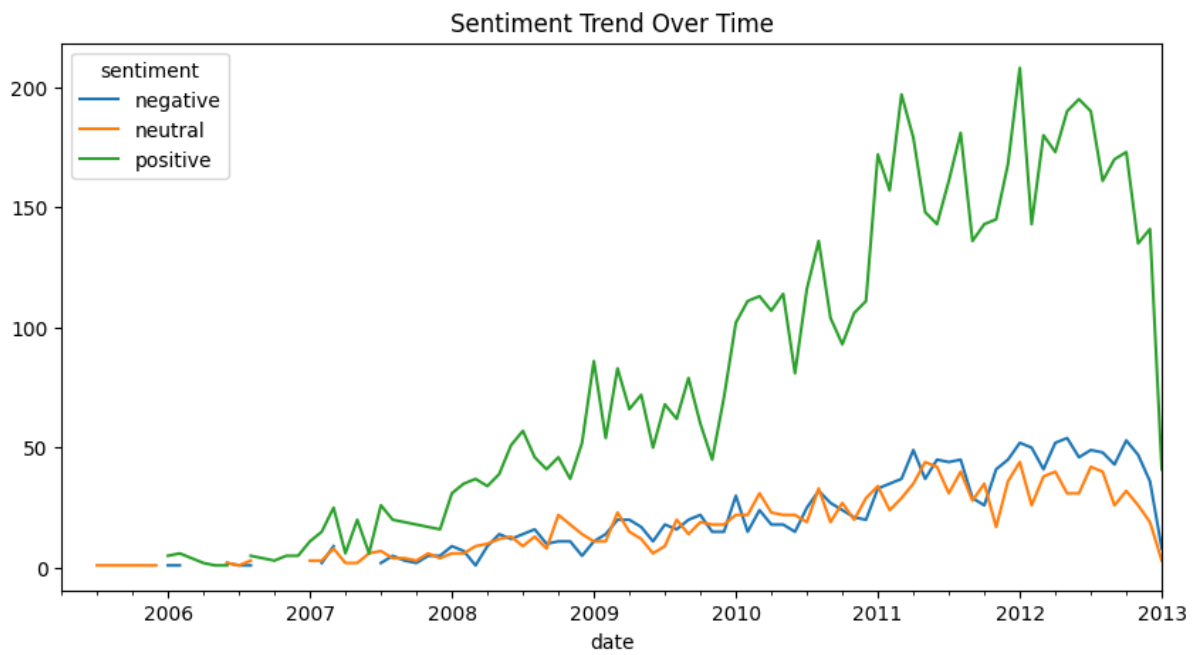
- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix
- AUC-ROC curve

Key findings from model evaluation:

- Logistic Regression achieved high accuracy and balanced performance across all sentiment classes
- Naive Bayes performed efficiently on large datasets but showed slightly lower accuracy due to its probabilistic assumptions
- SVM provided strong classification results, particularly for negative sentiment, but required more computational resources

Overall, the classification results confirmed that machine learning models can effectively identify customer sentiment from textual reviews.





KEY RECOMMENDATIONS

Based on sentiment analysis, topic modelling, and trend analysis, several actionable recommendations were identified.

Product and service improvements:

- Frequently criticized features should be prioritized for improvement
- Product defects and delivery issues require immediate operational attention

Customer engagement strategies:

- Highlight positively reviewed features in marketing campaigns
- Respond proactively to negative reviews to build customer trust

Crisis management:

- Monitor sentiment trends over time to detect sudden negative spikes
- Take quick corrective actions during periods of increased dissatisfaction

Continuous sentiment monitoring using NLP can help organizations remain responsive to customer needs and maintain a positive brand image.

CONCLUSION

This project demonstrates the effectiveness of Natural Language Processing and machine learning techniques in analysing large-scale customer review data. By combining sentiment classification, topic modelling, and trend analysis, the system provides valuable insights that support data-driven business decisions. The approach can be extended to real-time sentiment monitoring and advanced deep learning models in future work.