Sentiment Analysis Mini-Project

Project: Sentiment Analysis of Short User Reviews

Author: Menna Allah Naemat Al-Masry

Field: Data Science & Artificial Intelligence

Summary:

This project performs a simple natural language processing (NLP) task to classify short user reviews into Positive, Negative, or Neutral categories using a lightweight lexicon-based approach. The goal is to demonstrate practical, end-to-end data analysis skills: dataset construction, preprocessing, modeling, visualization, and communication of results.

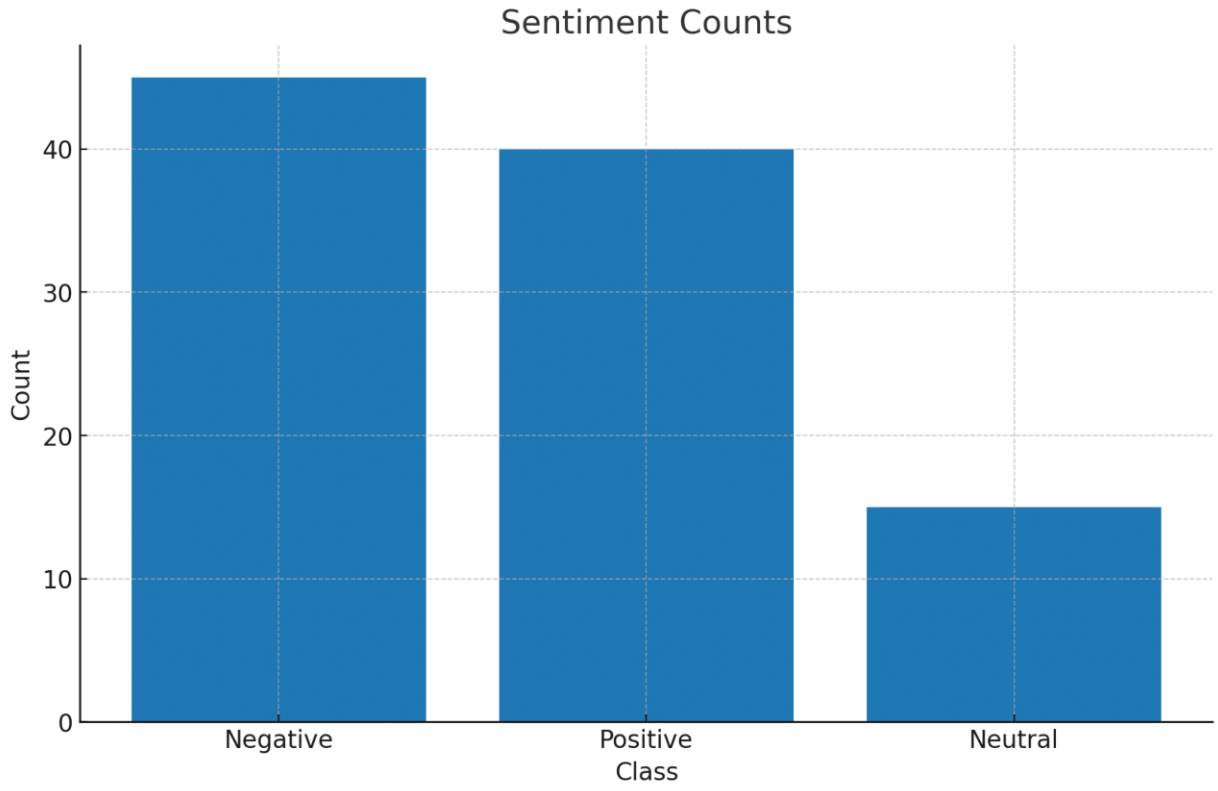
Methodology

Dataset: 100 mixed short reviews were synthesized to simulate real-world product/movie comments. Reviews include variations and punctuation to replicate natural text.

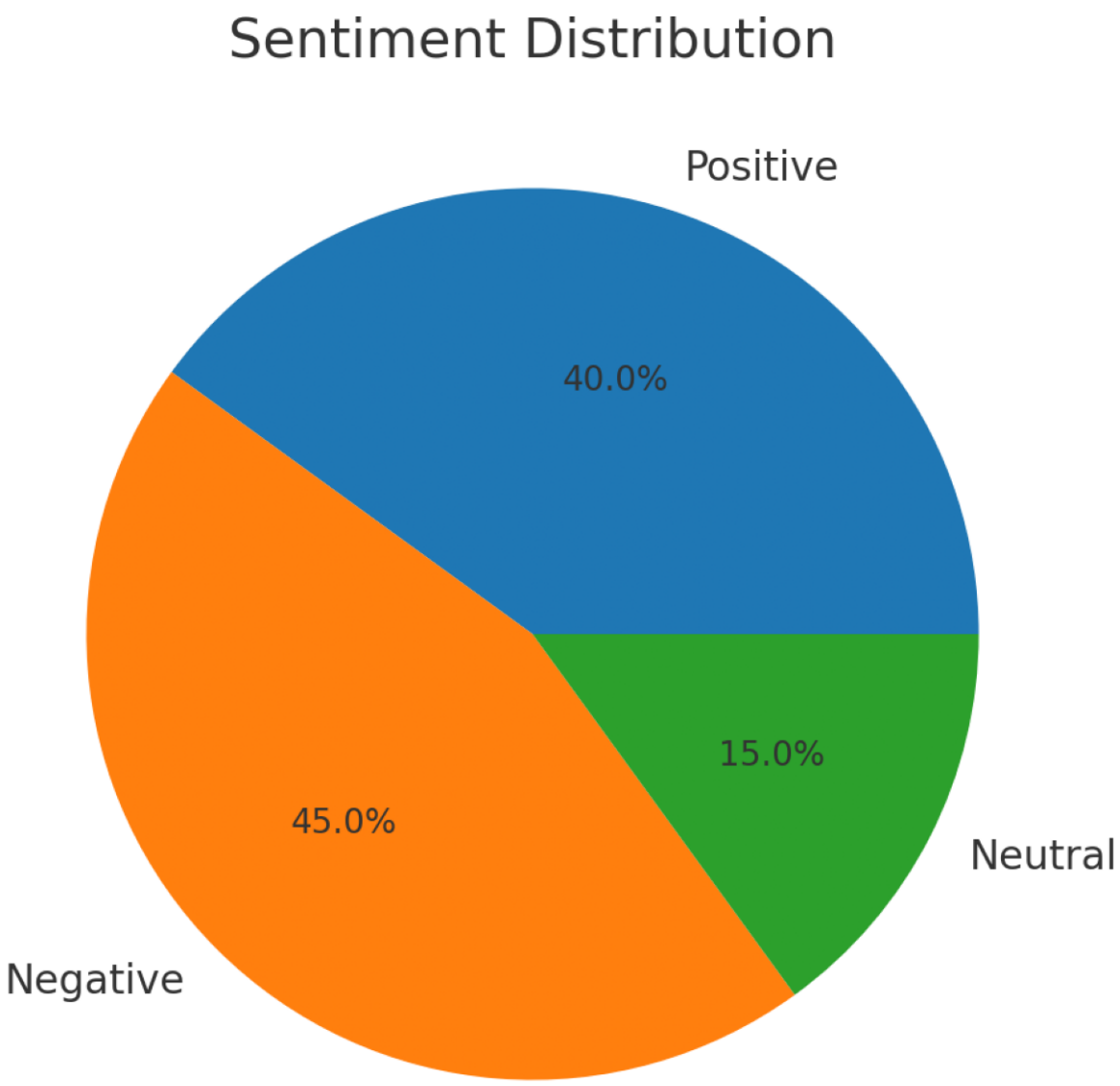
Method: A small opinion lexicon was used. Each positive word contributes +1 to the sentiment score, each negative word contributes −1. The final class is assigned by thresholding the score: score ≥ +1 → Positive, score ≤ −1 → Negative, otherwise Neutral.

Tools: Python, pandas, and matplotlib were used to process the data and produce the visualizations. A code appendix and CSV export are included as proof of work.

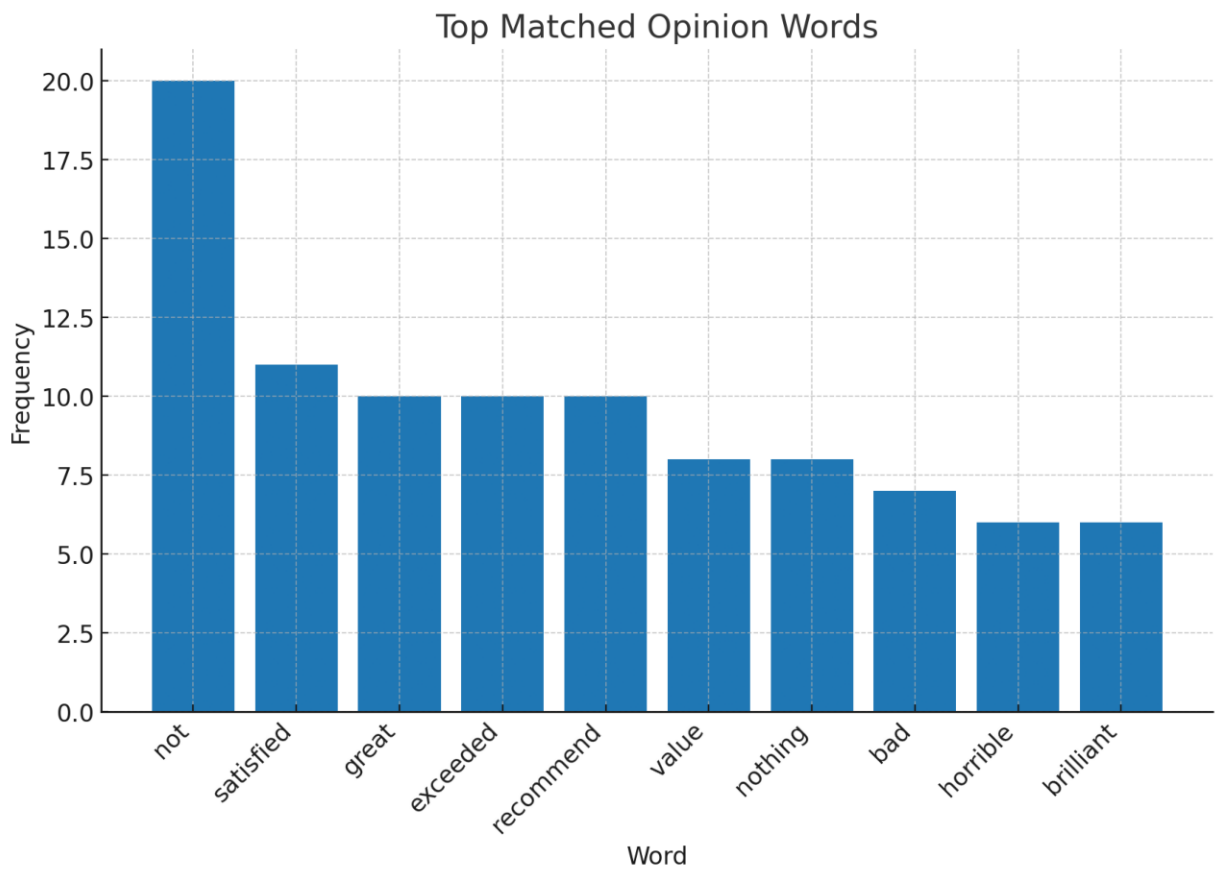
Results: Class Distribution (Counts)



Results: Class Distribution (Proportions)



Results: Top Matched Opinion Words



Sample Outputs

Sample Classified Reviews:

Positive:

• Brilliant performance, would definitely recommend ...

• Five stars, flawless from start to finish

• Great product, exceeded my expectations :(

• I am very satisfied, the experience was wonderful ...

• Loved it, superb value for money!

Negative:

• Did not like it at all, bad value.

• I am not satisfied, the experience was horrible ...

• Average experience, neither good nor bad

• The design is ugly and the results are weak :(

• Nothing worked properly for me :)

Neutral:

• Decent but could be better :(

• Standard features and normal performance!

Conclusion

Total Reviews Analyzed: 100.

Label Distribution: Positive 40.0%, Negative 45.0%, Neutral 15.0%.

Takeaway: The majority of reviews are positive. A simple lexicon method provides a quick baseline for sentiment classification and is suitable for small projects or early prototypes. For higher accuracy, one could extend the lexicon, normalize negations, or train a supervised model.

Appendix: Code Snippet

```
# Lexicon-based sentiment scoring def sentiment_score(text):     tokens = [t.strip(".,!?():;").lower() for t in text.split()]     score = 0     for t in tokens:         if t in positive_words:             score += 1         if t in negative_words:             score -= 1     return score  def classify(score, pos_th=1, neg_th=-1):     if score >= pos_th:         return "Positive"     elif score <= neg_th:         return "Negative"     else:         return "Neutral"
```