Sentiment Detection on Customer Feedback Using Naive Bayes

Presented by: Shikhar Tripathi

Introduction

- • Analyze customer reviews using NLP
- • Classify text as Positive or Negative
- • Naive Bayes Classifier used for modeling

Tools & Technologies

- • Python
- • Scikit-learn
- • NLTK
- • Pandas, NumPy
- • Matplotlib/Seaborn

Dataset Summary

- • Source: Amazon Fine Food Reviews (Kaggle)
- • Total records: ~500,000 (used ~20,000 sample)
- • Labels: 4–5 = Positive, 1–2 = Negative
- • Neutral (score 3) ignored for binary classification

Preprocessing Steps

- • Lowercasing text
- • Remove punctuation & digits
- • Remove stopwords
- • Tokenization
- • TF-IDF Vectorization

Model Workflow

- 1. Text Cleaning
- 2. TF-IDF Extraction
- 3. Train/Test Split
- 4. Train Multinomial Naive Bayes
- 5. Evaluate model
- 6. Predict new reviews

Evaluation Metrics

- • Accuracy Score
- • Precision, Recall, F1-score
- • Confusion Matrix

- Example Accuracy: ~86%

Sample Predictions

- Review: 'I love the battery life!' → Positive
- Review: 'Broke after one day.' → Negative
- Review: 'Top-notch quality!' → Positive
- Review: 'Not worth the price.' → Negative

What’s Unique

- • Real-world dataset
- • Strong preprocessing pipeline
- • Accurate & explainable model
- • Extendable to Neutral class or app integration

Conclusion

- • NLP + Naive Bayes used for text classification
- • Achieved high accuracy with simple model
- • Scalable to real-world use cases
- • Thank you!