Sentiment Analysis on Amazon Fine Foods

Alagesan, Aishwariya, alagesan.a@northeastern.edu

Gajula, Saiteja Reddy, gajula.sai@northeastern.edu

Hansda, Sheela, hansda.s@northeastern.edu

Project Overview

Objective:

Conduct sentiment analysis on Amazon Fine Food Reviews to classify customer sentiments as positive, neutral, or negative.

Goals:

- Uncover trends in customer sentiment
- Improve sentiment prediction accuracy
- Provide insights into customer preferences

Dataset

Source: Amazon Fine Food Reviews dataset

Size: Over 500,000 reviews

Key fields

- reviewText
- score (1-5)
- summary
- metadata (product ID, user ID, timestamps)

Data Preparation and Preprocessing

Data Acquisition

- Exploratory Data Analysis (EDA)

- Analysis of rating distribution

- Word clouds for positive and negative sentiments

Preprocessing Techniques

- Data cleaning (removing HTML tags, special characters, etc.)

- Text normalization (lowercase conversion, stopword removal)

- Feature engineering (sentiment labeling, word count extraction)

Data Splitting

- 80% training

- 10% validation

- 10% testing

Model Architecture

Base Models :

- Naïve Bayes

- Logistic Regression

Main models :

- Recurrent Neural Network (RNN)

- Long Short-Term Memory (LSTM)

Results

LSTM Model Performance:

- Accuracy *78.25%*

- Precision

  Positive: 82%, Neutral: 76%, Negative: 74%

- Recall

  Positive: 78%, Neutral: 75%, Negative: 72%

- F1-Score

  Positive: 80%, Neutral: 75%, Negative: 73%

RNN Model Performance

- Accuracy: *82.26%*

- Precision:

  Positive: 86%, Neutral: 80%, Negative: 80%

- Recall:

  Positive: 83%, Neutral: 83%, Negative: 80%

- F1-Score:

  Positive: 84%, Neutral: 81%, Negative: 80%

CONCLUSION

The RNN model showed better overall accuracy and consistency across sentiment classes, while the LSTM model demonstrated improved performance in handling longer-term dependencies.