Capstone Project Report: Review Dish

# Overview

This capstone project, titled 'Review Dish', is a comprehensive Natural Language Processing (NLP) system designed to analyze large-scale customer reviews. The system classifies sentiment, identifies key discussion topics, tracks trends over time, and generates business insights using advanced ML and NLP techniques.

# Objectives

- Understand customer sentiment across reviews

- Identify frequently mentioned topics and keywords

- Classify reviews by sentiment and intent

- Recommend data-driven business improvements

- Provide visual reports through interactive dashboards

# Tech Stack

Language: Python

Notebook: Google Colab / Jupyter

NLP Libraries: NLTK, spaCy, TextBlob, BERTopic

Vectorization: TF-IDF, Word2Vec, BERT

ML Models: Logistic Regression, SVM, Naive Bayes

Topic Modeling: BERTopic

Trend Analysis: Pandas, Matplotlib, Seaborn

Dashboards: Plotly

NER: spaCy

# Project Pipeline

1. Data Preprocessing: Removed noise, lemmatized text, and vectorized using TF-IDF, Word2Vec, BERT

2. Exploratory Text Analysis: Word clouds, N-gram analysis, ratings by product & time

3. Sentiment Classification: Models - Logistic Regression, Naive Bayes, SVM

4. Topic Modeling: Used BERTopic to extract main themes and correlate with review scores

5. Trend Analysis: Time-based sentiment patterns and spike detection

6. Named Entity Recognition: Used spaCy to extract brands, products, and locations

7. Insight Generation: Highlighted most common praise/criticism and improvement areas

8. Reporting & Visualization: Built Streamlit dashboard with interactive visualizations

# Dataset

The dataset used was the Amazon Fine Food Reviews from Kaggle, comprising over 568,000 customer reviews. Each entry included review text, star ratings, and timestamps. The data enabled a deep dive into consumer sentiment trends and patterns.

<https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews>

# Results and Screenshot

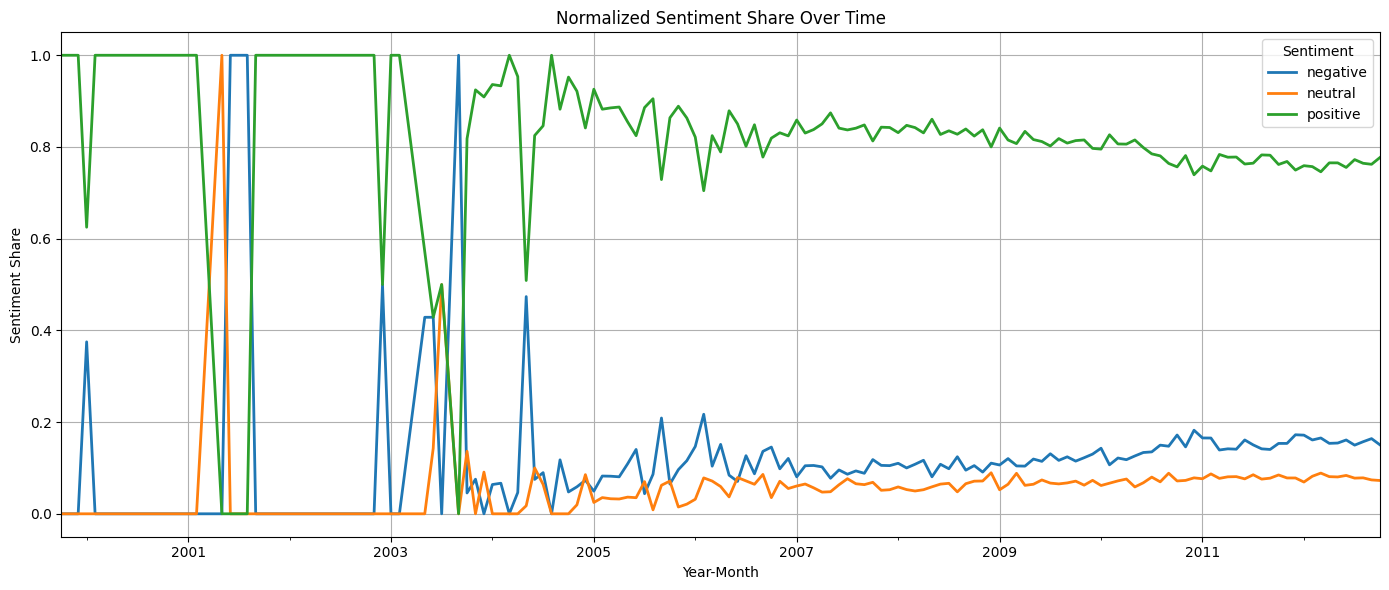
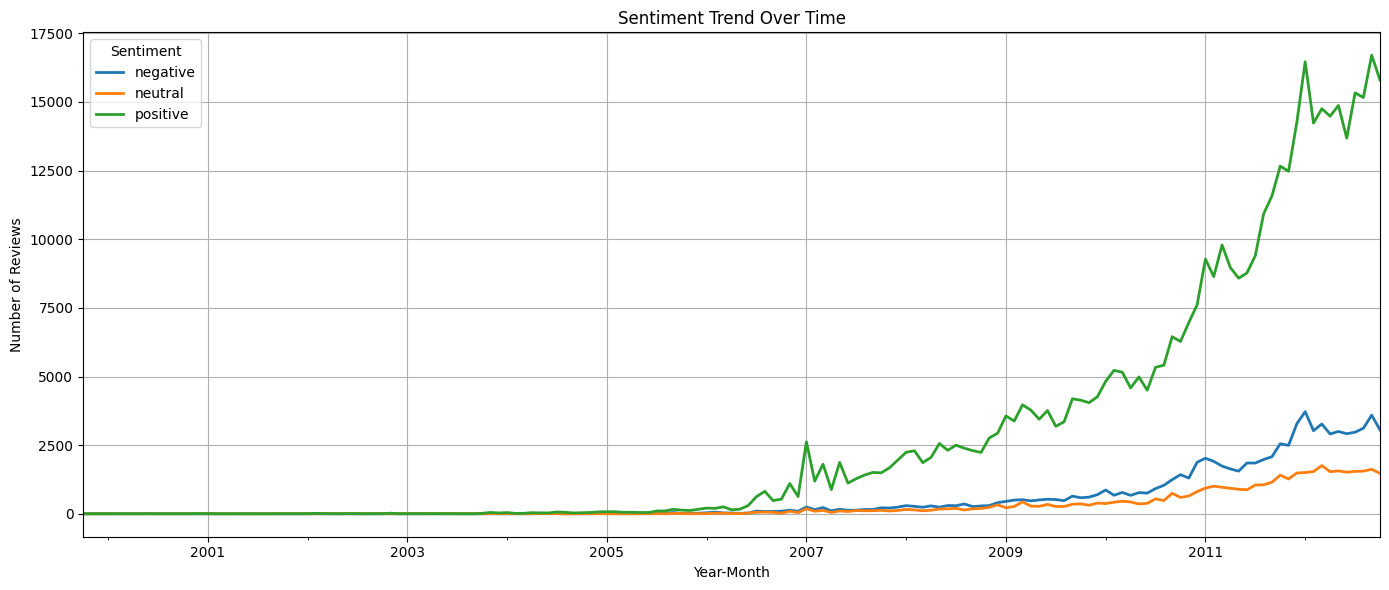
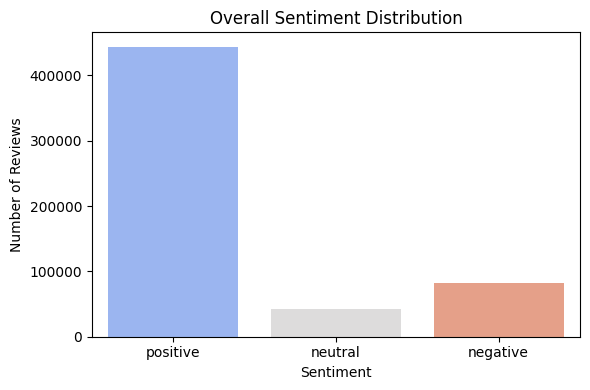
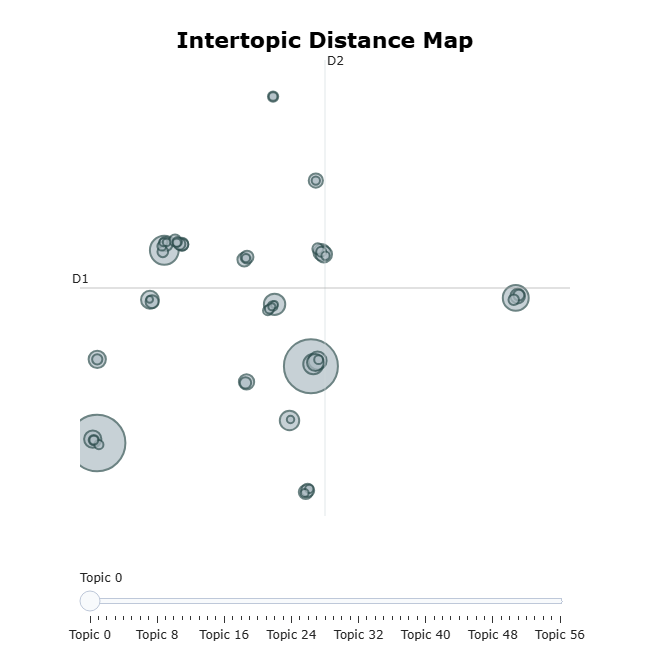
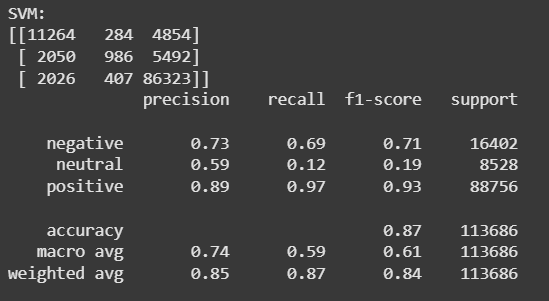
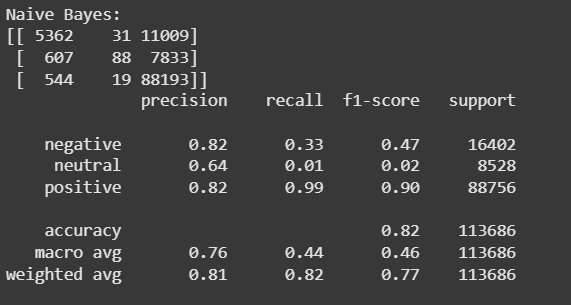
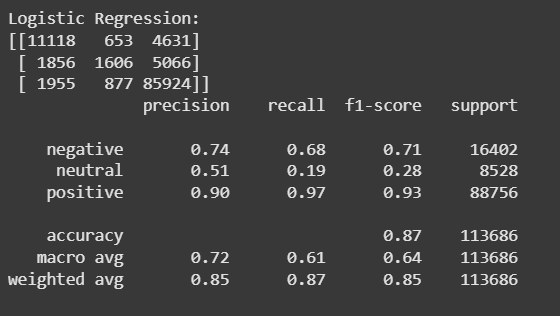
Word cloud



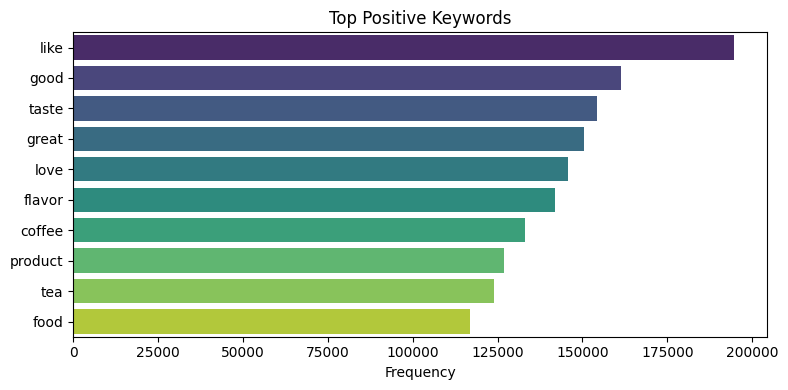
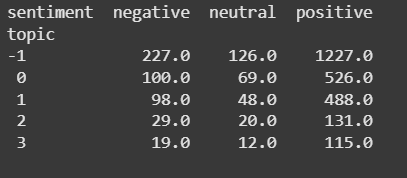
# p wc

# ne wc

Classifiers



Sentiment per topic



# Future Work

- Use multilingual datasets

- Integrate audio/video sentiment (AssemblyAI)

- Enable real-time dashboard updates

- Deploy a full-scale app with a feedback API