Final Report: Customer Voice Intelligence System

# 1. Executive Summary

This project focuses on analyzing large-scale customer reviews using Natural Language Processing (NLP) techniques. The goal is to understand customer sentiment, identify key topics, classify sentiment accurately, and provide actionable insights to enhance product strategy and customer engagement. We used Amazon Fine Food Reviews dataset as our source and implemented preprocessing, exploratory analysis, classification models, topic modeling, and entity recognition.

# 2. Text Analysis & Sentiment Insights

The reviews were preprocessed by removing noise, stopwords, and applying lemmatization. TF-IDF was used to vectorize text for modeling. WordClouds and N-gram analysis revealed key terms in positive and negative reviews.  
  
Top Positive Words: taste, love, delicious, perfect  
Top Negative Words: bad, worst, bland, waste  
  
Sentiment Distribution:  
- Positive: 70%  
- Negative: 30%

# 3. Classifier Performance & Findings

A Logistic Regression classifier was trained on TF-IDF features to classify review sentiment (positive/negative). The model achieved the following performance:  
  
- Accuracy: ~89%  
- Precision (Positive): 0.91  
- Recall (Positive): 0.87  
- F1-Score (Positive): 0.89  
  
Topic modeling with LDA revealed themes such as product taste, packaging issues, and delivery delays. NER with spaCy highlighted entities like brand names and locations.

# 4. Key Recommendations

Based on the analysis, the following strategic recommendations are proposed:  
- Improve packaging quality and delivery logistics based on negative topic trends.  
- Boost marketing around well-received products (e.g., those with high 'delicious' mentions).  
- Implement feedback loops to address complaints early using trend detection.  
- Launch loyalty programs for frequent reviewers with positive feedback.  
- Use NER results to monitor competitor mentions and track product names in reviews.