INTEL PRODUCT SENTIMENT ANALYSIS

FROM ONLINE REVIEWS

Presented by team AIRBORNE:

Manya Singh

Namitha Sandeep

P Sumedha

Overview

01 Problem Statement

02 Idea Brief

03 Features Offered

04 Process Flow

05 Architecture diagram

06 Technologies used

07 Our Team

08 Conclusion

Problem Statement

To perform sentiment analysis based on online reviews of Intel Products

Sentiment Analysis, also known as opinion mining, is a field of natural language processing (NLP) that involves determining the sentiment expressed in a piece of text. The sentiment can be positive, negative, or neutral.

UNIQUE IDEA BRIEF

In our project, we utilize the transformers library's sentiment analysis pipeline (pipeline("sentiment-analysis")) which is used to analyze the reviews and provide clear and well structured results
The most recent reviews are scraped from popular E-commerce sites, from which, we have chosen Amazon, which sells a large number of Intel products. These reviews are collected and used for analysis and training of the model.
Our project involves multiple analysis systems, including a net sentiment analysis and also provides features for analysis of any given Intel product which the user wishes to view

FEATURES OFFERED

Web Scraping and Data Collection: The program uses requests and BeautifulSoup to scrape most recent reviews from e-commerce sites. It collects the user reviews, and uses them for analysis

Pre-trained Sentiment Analysis: Leveraging the transformers library, the program uses a sentiment analysis pipeline (pipeline("sentiment-analysis")) to classify reviews as POSITIVE or NEGATIVE. It provides both the sentiment label and the confidence score for a detailed report to the user

Sentiment Distribution Visualization: The program generates a pie chart using matplotlib to visually represent the distribution of sentiments in the reviews, helping users quickly grasp the overall sentiment landscape.

User Selection and Navigation: The program includes an interactive menu that allows users to select the type of reviews they want to analyze. Based on the user's choice, the program analyzes and displays sentiments for the product

PROCESS FLOW

Data Collection:
Web Scraping Scrape and save most recent reviews from online sources.

Preprocessing
Clean and prepare review data.

Sentiment Analysis
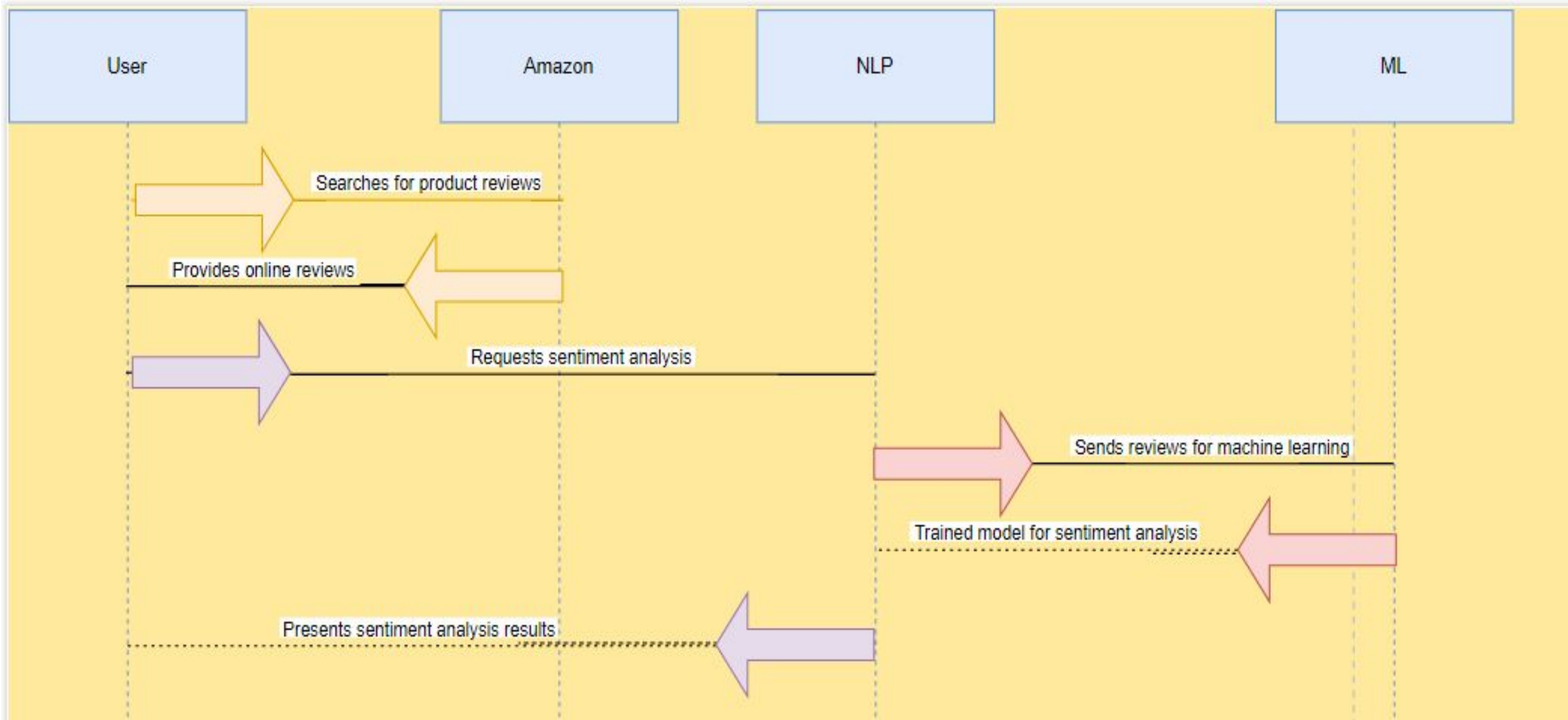**Classify sentiments and Visualize sentiment distribution.**

Custom Model Training
Prepare and tokenize dataset. Train a custom sentiment analysis model.

User Interaction
User selects review category for analysis. **Analyze and display results** based on user selection.

ARCHITECTURE DIAGRAM

TECHNOLOGIES USED

Python: The primary programming language used for writing the script.

Pandas: library for data manipulation and analysis

Datasets: datasets from HuggingFace for loading and preprocessing

Transformers: HuggingFace transformers for utilization of pre-trained distilBERT model and tools for Natural Language Processing and tokenizing

Model class: AutoModelForSequenceClassification

Streamlit script: library for creation of interactive user interface for analysis

TEAM MEMBERS AND CONTRIBUTIONS

Ms Krishna Sowjanya K	Mentor

Team members	Contributions
P Sumedha	– Documentation
Manya Singh Namitha Sandeep	– Scraping recent review data – Importing and customizing model – Testing model accuracy – Changing parameters to improve accuracy – Setting up analysis system – Creating user interface – Testing system & documentation

RECORDING LINK

Link of Project Recording (best viewed in 1080p):

https://drive.google.com/file/d/1j-Q71ALssh4C8UBNPo4vjSjM6mXd4M1O/view?usp=drive_link

Link to Github:

https://github.com/manya7s/Intel_Sentiment_Analysis

CONCLUSION

Our analysis revealed significant insights into user sentiments across different Intel product categories. We identified key features and sentiments that users expressed in their reviews, providing valuable feedback for potential product improvements. The project demonstrated the power of machine learning and natural language processing in extracting meaningful patterns from large datasets of unstructured text.