Project Documentation

Project Title:

Sentiment and Opinion Mining System using Dash and NLP.

- **Overview**:
  - This project is designed to analyze textual data, such as social media comments or product reviews, to extract both general sentiment (positive, negative, neutral) and detailed opinions on specific aspects or features using Natural Language Processing (NLP) techniques.
- **Goal**:
  - Classify text data based on sentiment polarity.
  - Extract aspects (specific entities) using NLP for detailed opinion analysis.
  - Visualize the results interactively through a Dash-based dashboard.

2. Project Features

- **Sentiment Analysis**:
  - Uses **VADER (Valence Aware Dictionary and sEntiment Reasoner)** to classify text sentiment into positive, negative, or neutral. VADER computes sentiment polarity based on the compound score derived from the text.
  - Sentiment scores (positive, negative, neutral) are applied to individual comments or sentences.
- **Opinion Mining**:
  - **Aspect Extraction**: Identifies specific aspects of the text (e.g., “camera,” “battery life”) using **spaCy**'s noun chunk extraction capabilities.
  - **Aspect Sentiment Classification**: For each extracted aspect, VADER is used to classify the sentiment (positive, negative, neutral) surrounding the aspect.
- **Interactive Dashboard**:
  - Built with **Dash**, the dashboard allows users to visualize sentiment distribution (pie chart) and aspect-based sentiment analysis (bar chart). Users can interactively explore sentiment and opinions related to specific aspects.

3. Project Structure

- **Directory Structure**:
  - app.py: Main script for the interactive dashboard.
  - sentiment_analysis.py: Script for handling sentiment analysis logic using VADER.
  - opinion_mining.py: Handles aspect extraction and opinion mining using spaCy and VADER.
  - utils.py: Helper functions for text processing (tokenization, stopword removal).

- comments.csv: Input CSV file containing textual data for sentiment and opinion analysis.
- requirements.txt: List of dependencies and libraries for the project.
- config.yaml: Configuration file containing project parameters (e.g., sentiment thresholds, paths).

4. Installation and Setup

- **System Requirements**:
  - **Python**: Version 3.x.
  - Required libraries include **pandas**, **nltk**, **spacy**, **dash**, and **plotly**.
- **Installation**:
  - Install the required packages using:
  - pip install -r requirements.txt
  - Download the necessary data for **NLTK** and **spaCy**:
  - python
  - nltk.download('vader_lexicon')
  - nltk.download('stopwords')
  - nltk.download('punkt')
  - python -m spacy download en_core_web_sm

5. Data Processing

- **Loading Data**:
  - Load comments or text data from a CSV file with a **body** column.
- **Text Preprocessing**:
  - Clean and preprocess text data by lowercasing, removing punctuation, and stopwords using NLTK functions.
  - Example of raw text: "The battery life is amazing!"
  - Example of processed text: "battery life amazing"

6. Sentiment Analysis

- **VADER Sentiment Analysis**:
  - **VADER** calculates sentiment scores: Positive, Negative, Neutral, and Compound. The compound score is used to classify the text into positive, negative, or neutral sentiment categories.
  - Example:
    - Input: "The camera quality is excellent."
    - Sentiment Scores: Positive (0.8), Negative (0.0), Neutral (0.2), Compound (0.85).
    - Classified Sentiment: Positive.

7. Opinion Mining

- **Aspect Extraction**:
  - Aspects (nouns or phrases) are extracted using **spaCy**'s noun chunking. These aspects represent key features or entities mentioned in the text.
- **Aspect Sentiment Classification**:
  - For each aspect, its surrounding context is analyzed using **VADER** to assign sentiment (positive, negative, neutral) for that specific aspect.
  - Example:
    - Comment: "The battery life is terrible, but the screen is great."
    - Extracted Aspects: ["battery life", "screen"]
    - Sentiment Classification:
      - "battery life" → Negative.
      - "screen" → Positive.

8. Interactive Dashboard

- **Sentiment Visualization**:
  - A pie chart displays the overall sentiment distribution (positive, negative, neutral) for all the comments.
- **Aspect Sentiment Visualization**:
  - Users can select specific aspects (e.g., "camera", "battery life") to view a bar chart of sentiment distribution for that aspect.
- **User Interactions**:
  - Interact with the dashboard by selecting aspects from a dropdown menu to dynamically update the aspect sentiment chart.

9. Results and Analysis

- **Sentiment Results**:
  - The sentiment analysis displays the overall sentiment breakdown for the dataset.
  - Example: "70% of the comments are classified as Positive, 20% as Neutral, and 10% as Negative."

output.csv:

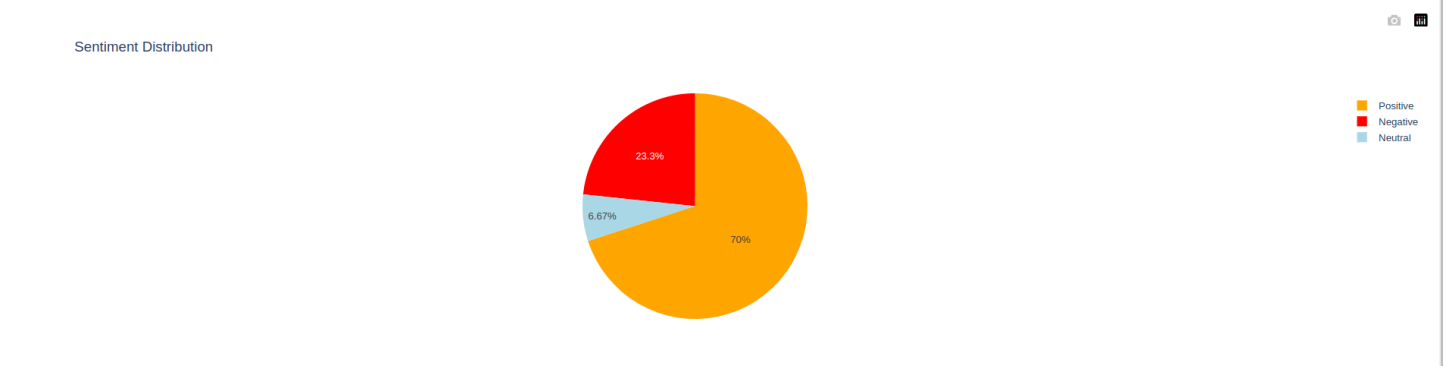
app.py | output.csv U | main.py

Process > output.csv > data

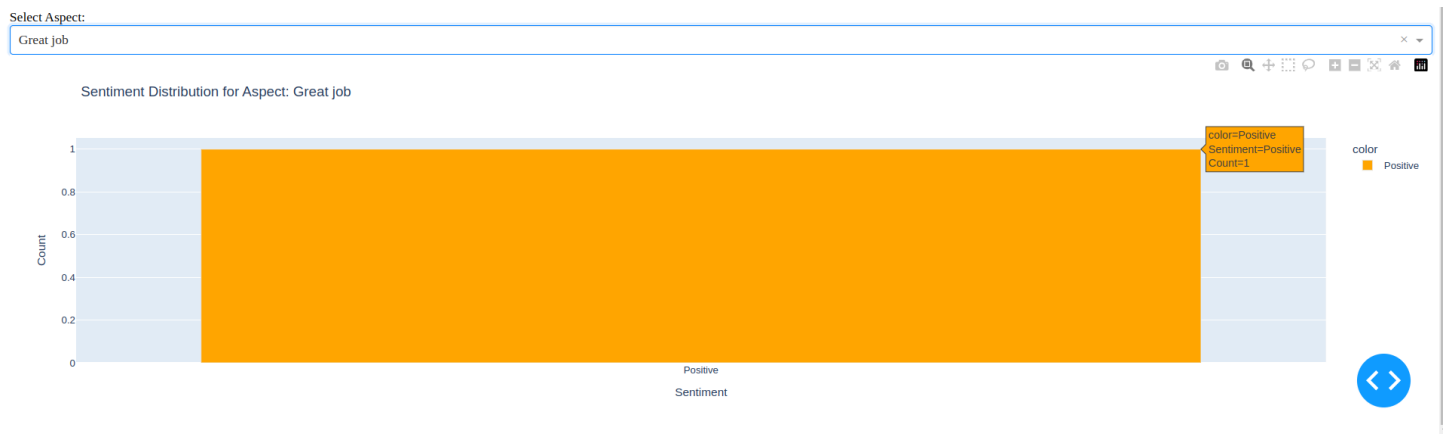
```
comment,aspects,aspect_sentiments
This video is amazing! I learned so much.,"['This video', 'I']","{'This video': 'Positive', 'I': 'Neutral'}"
I didn't find this helpful at all. The explanation was too vague.,"['I', 'this', 'The explanation']","{'I': 'Positive', 'this': 'Positive', 'The explanation': 'Negative'}"
Great tutorial! Very clear and easy to follow.,['Great tutorial'],{'Great tutorial': 'Positive'}
"Could have been better, but still decent.",[],{}
"Absolutely terrible, I wouldn't recommend it.","['I', 'it']","{'I': 'Negative', 'it': 'Neutral'}"
"Informative content, but the pacing was too slow.","['Informative content', 'the pacing']","{'Informative content': 'Neutral', 'the pacing': 'Neutral'}"
Fantastic! I will definitely share this with my friends.,"['I', 'this', 'my friends']","{'I': 'Positive', 'this': 'Positive', 'my friends': 'Positive'}"
"Not bad, but I expected more details.","['I', 'more details']","{'I': 'Neutral', 'more details': 'Neutral'}"
"I loved the visuals in this video, they were very engaging!","['I', 'the visuals', 'this video', 'they']","{'I': 'Positive', 'the visuals': 'Positive', 'this video': 'Pos
The speaker could have been more enthusiastic.,['The speaker'],{'The speaker': 'Positive'}
Superb explanation! Cleared up all my doubts.,"['Superb explanation', 'all my doubts']","{'Superb explanation': 'Positive', 'all my doubts': 'Negative'}"
"This was a waste of time, too much irrelevant information.","['This', 'a waste', 'time', 'too much irrelevant information']","{'This': 'Negative', 'a waste': 'Negative',
"Decent, but the background music was too distracting.",['the background music'],{'the background music': 'Negative'}
"I really enjoyed this video, the editing was top-notch!","['I', 'this video', 'the editing']","{'I': 'Positive', 'this video': 'Neutral', 'the editing': 'Neutral'}"
"Meh, I've seen better tutorials on the same topic.","['I', 'better tutorials', 'the same topic']","{'I': 'Positive', 'better tutorials': 'Positive', 'the same topic': 'Ne
Very insightful! I feel like I finally understand the concept.,"['I', 'I', 'the concept']","{'I': 'Positive', 'the concept': 'Neutral'}"
"I struggled to keep up with the pace, too fast for beginners.","['I', 'the pace', 'beginners']","{'I': 'Negative', 'the pace': 'Neutral', 'beginners': 'Neutral'}"
Outstanding work! Will recommend it to my team.,"['O                    ork', 'it', 'my team']","{'Outstanding work': 'Positive', 'it': 'Neutral', 'my team': 'Neutral'}"
"This video is all over the place, not structured we                    video', 'the place']","{'This video': 'Negative', 'the place': 'Negative'}"
"So boring, I couldn't finish watching it.","['I', 'it']","{'I': 'Neutral', 'it': 'Neutral'}"
"Good content, but the quality of the audio could be better.","['Good content', 'the quality', 'the audio']","{'Good content': 'Positive', 'the quality': 'Positive', 'the
Loved it! The examples were spot on and easy to understand.,"['it', 'The examples']","{'it': 'Positive', 'The examples': 'Positive'}"
"The speaker was monotone, hard to stay focused.",['The speaker'],{'The speaker': 'Positive'}
"This is exactly what I needed, thank you!","['This', 'exactly what', 'I', 'you']","{'This': 'Positive', 'exactly what': 'Positive', 'I': 'Positive', 'you': 'Neutral'}"
"The video feels rushed, not enough attention to details.","['The video', 'not enough attention', 'details']","{'The video': 'Neutral', 'not enough attention': 'Neutral',
"Helpful, but some points were oversimplified.",['some points'],{'some points': 'Neutral'}
Great job! I learned new techniques I hadn't seen before.,"['Great job', 'I', 'new techniques', 'I']","{'Great job': 'Positive', 'I': 'Neutral', 'new techniques': 'Neutral
This tutorial is way too basic for advanced users.,"['This tutorial', 'advanced users']","{'This tutorial': 'Positive', 'advanced users': 'Positive'}"
"The visuals were stunning, but the explanation was weak.","['The visuals', 'the explanation']","{'The visuals': 'Negative', 'the explanation': 'Negative'}"
"I appreciate the effort, but this could use more depth.","['I', 'the effort', 'this', 'more depth']","{'I': 'Positive', 'the effort': 'Neutral', 'this': 'Neutral', 'more
```

Col 2: aspects

Sentiment and Opinion Mining Dashboard



- **Aspect Sentiment Results**:
  - The dashboard also shows sentiment results specific to aspects (e.g., product features).
  - Example: "For the aspect ‘battery life’, the sentiment is mostly Negative (80%), indicating dissatisfaction."
-



10. Conclusion

- **Summary**:
  - The project combines **Sentiment Analysis** and **Opinion Mining** using NLP to provide a comprehensive understanding of both general sentiment and opinions on specific aspects.
- **Key Learnings**:
  - Challenges in extracting meaningful aspects from noisy text data and handling mixed sentiments were overcome by fine-tuning text preprocessing steps.
- **Future Work**:
  - Expand the system by integrating **machine learning models** for more accurate sentiment classification.
  - Improve aspect sentiment analysis by using advanced techniques such as **Aspect-Based Sentiment Analysis (ABSA)** models.

Existing Features

- Sentiment analysis using VADER.
- Opinion mining through aspect extraction using spaCy.
- Text preprocessing for clean and normalized data.

New Features Added

- Interactive dashboard for visualizing both general sentiment and aspect-specific sentiment distributions.
- Enhanced opinion mining by visualizing sentiment tied to specific aspects.
- Dynamic content handling from CSV files to generate real-time interactive insights.

Github link of my project

Moneshai2004/**NLP-for-sentiment-analysis-and...**

sentiment analysis and opinion mining is a Natural Language Processing (NLP) tims to determine the sentiment (positive, negative, or neutral)...

1 Contributor | 0 Issues | 0 Stars | 0 Forks