Sentiment

Analysis

Data Collection

Data collection process involved collecting comments from the official page of Modern Family.

This was done using scraping code.

1300 comments were collected.

This was collected for training and testing purposes. The user should collect new data for their use case.

```
great@Greatness-HP MINGW64 ~/Documents/DATA44001 (main)
$ python simple_scraper.py

get_auth_twitter_pg()
        waiting for login, sleeping for 3 seconds
        waiting for login, sleeping for 3 seconds
        waiting for login, sleeping for 3 seconds
        waiting for login, sleeping for 3 seconds
        authenticated
        sleeping for 3 seconds

get_timeline_tweets(): ModernFam
        throttling/scrolling, then sleeping for 5 second

        extracted 1 tweets
        extracted 2 tweets
        throttling/scrolling, then sleeping for 5 second

        extracted 3 tweets
        extracted 4 tweets
        extracted 5 tweets
        extracted 6 tweets
        extracted 7 tweets
        extracted 8 tweets
        extracted 9 tweets
        extracted 10 tweets
        throttling/scrolling, then sleeping for 5 second

        extracted 11 tweets
```

Data Preprocessing

Clean data by removing URLs, mentions, special characters.

Tokenized the data.

```
$ python preprocessing.py data/twitter_comments.json data/unlabeled_comments.json
```

Sentiment Analysis

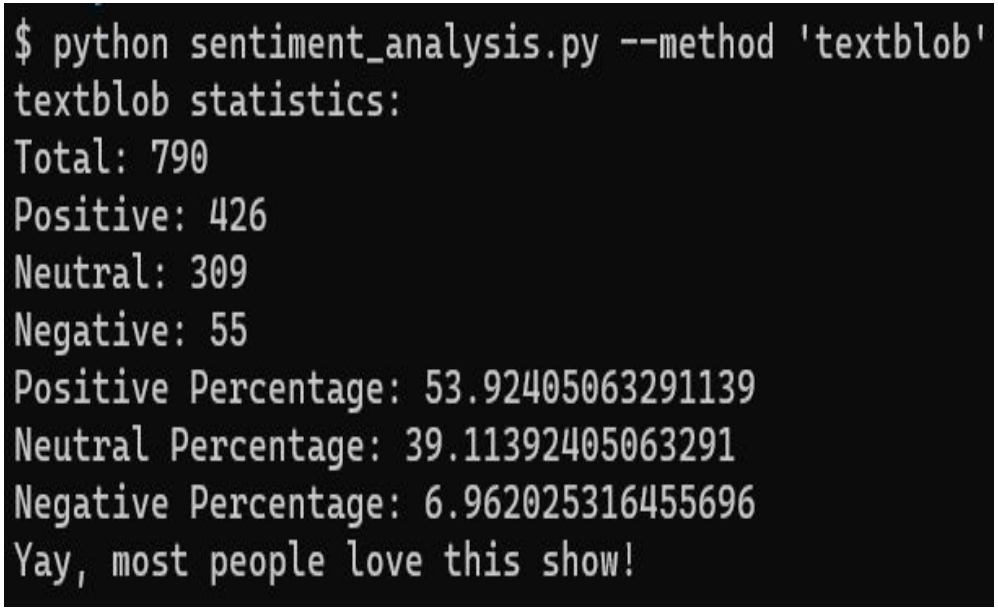
Classifies comments as positive, negative, or neutral

Uses two approaches: rule-based (TextBlob) and machine learning (Naive Bayes)

Sentiment Analysis cont'd

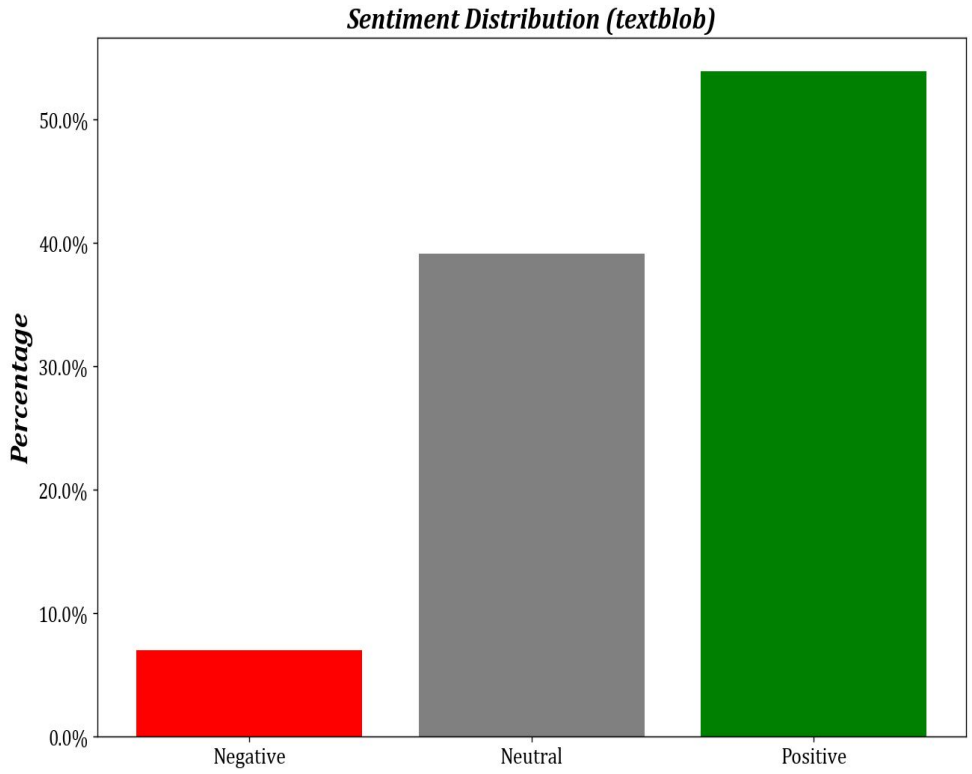
User has the option to choose which method to implement.

Output statistics (totals, percentages, etc).



Sentiments Visualizations

Bar chart



Sentiments Visualizations cont'd

Word cloud