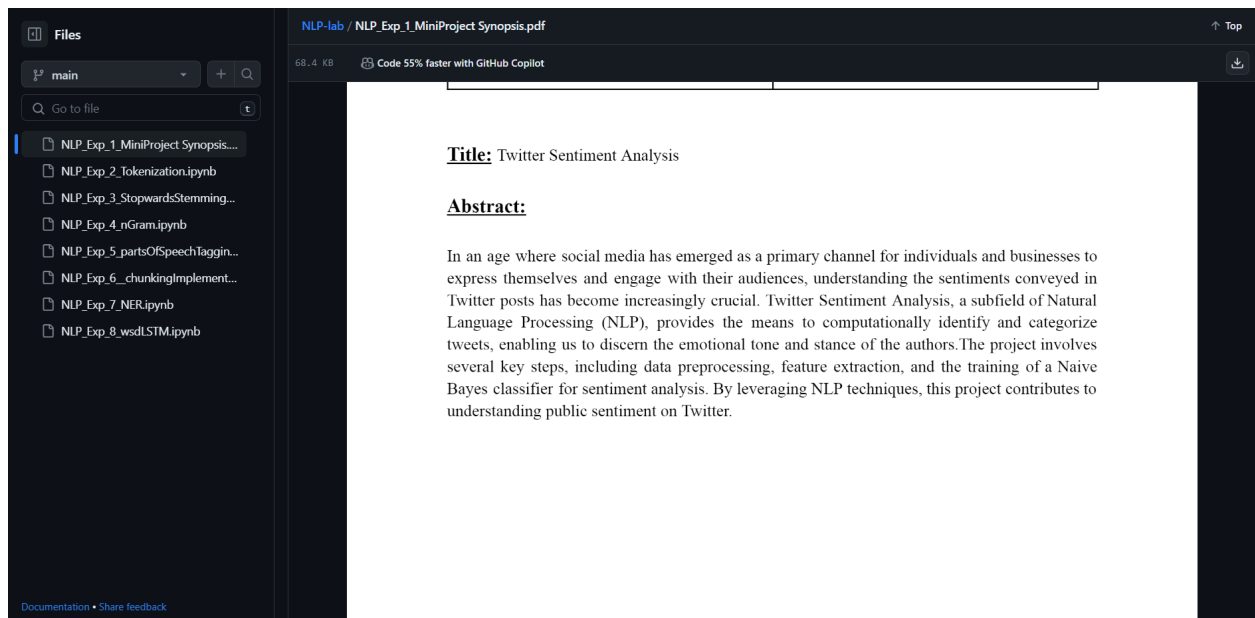
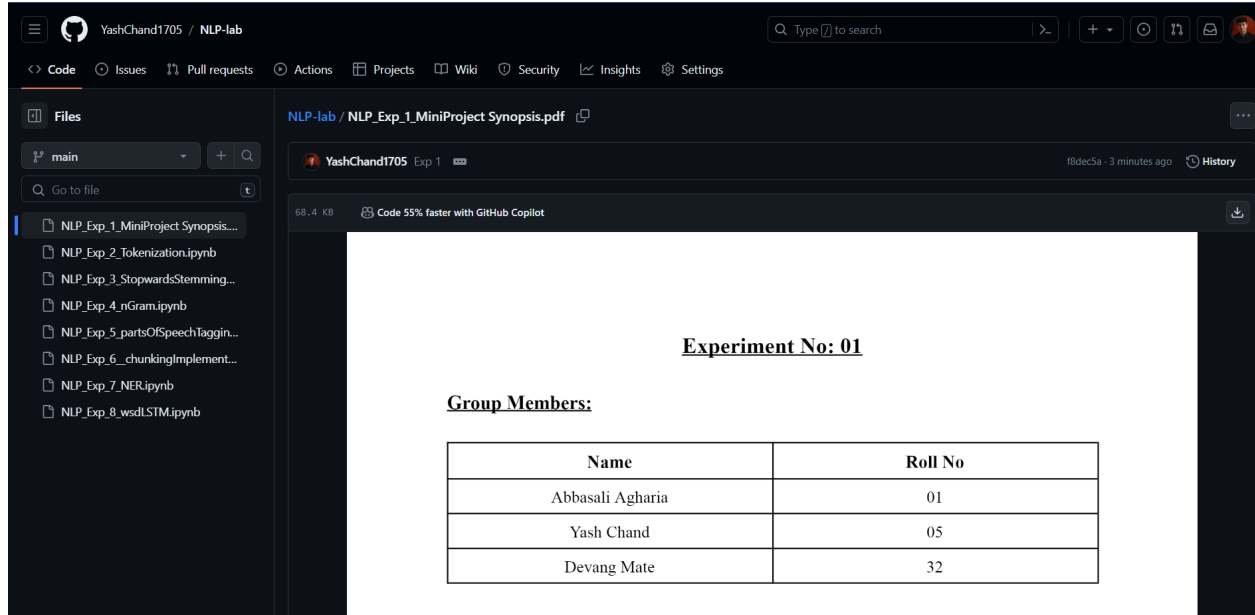


Experiment No: 01

URL:

https://github.com/YashChand1705/NLP-lab/blob/main/NLP_Exp_1_MiniProject%20Synopsis.pdf

Snapshot:



Files

main

Go to file

NLP_Exp_1_MiniProject Synopsis...

NLP_Exp_2_Tokenization.ipynb

NLP_Exp_3_StopwordsStemming...

NLP_Exp_4_nGram.ipynb

NLP_Exp_5_partsOfSpeechTaggin...

NLP_Exp_6_chunkingImplement...

NLP_Exp_7_NER.ipynb

NLP_Exp_8_wsdLSTM.ipynb

Documentation • Share feedback

NLP-lab / NLP_Exp_1_MiniProject Synopsis.pdf

68.4 KB Code 55% faster with GitHub Copilot

↑ Top

Methodology:

1. Data Collection and Preprocessing:

The project begins by importing the necessary libraries, including NumPy and Pandas, and reading a CSV file containing tweet data.

The data is filtered to retain only the 'text' and 'sentiment' columns.

The dataset is split into training and test sets, with neutral sentiments excluded from the training set.

2. Word Cloud Visualization:

Two word clouds are generated to visualize positive and negative words. These visualizations provide insights into the most frequently occurring words in tweets associated with each sentiment.

3. Text Tokenization and Filtering:

Tweets in the training set are tokenized, and words are filtered to remove URLs, mentions, hashtags, and the "RT" (retweet) tag.

Stop words from the NLTK library are removed to clean the text further.

4. Feature Extraction:

Word features are extracted by analyzing the frequency distribution of words in the cleaned tweets.

These features are used to create a feature set for training the sentiment analysis model.