

Project Progress Report

Progress made thus far

1. Set up Environment. Installed Python 3, made requirements.txt file and installed compatible libraries
2. Data Pre-Processing in progress- removed stop words, removed latin words. Tokenized words and removed punctuations. Performed POS tagging and extracted adjectives and nouns

Remaining Tasks

1. To execute LDA and extract topics
2. To perform sentiment analysis on each topic at a dataset level

Challenges faced

1. Library incompatibilities with Python version
2. UTF-8 encoding err - added diff type of encoding to read_csv in pandas
3. Spacy library incompatibility
4. No module named en-core-web-sm
5. Extracting meaningful words
6. Mapping lemmatisation functions to every row in data frames