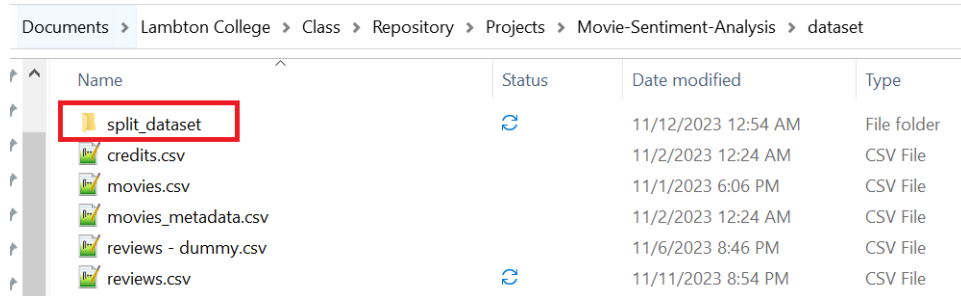


Sentiment generating process for review dataset

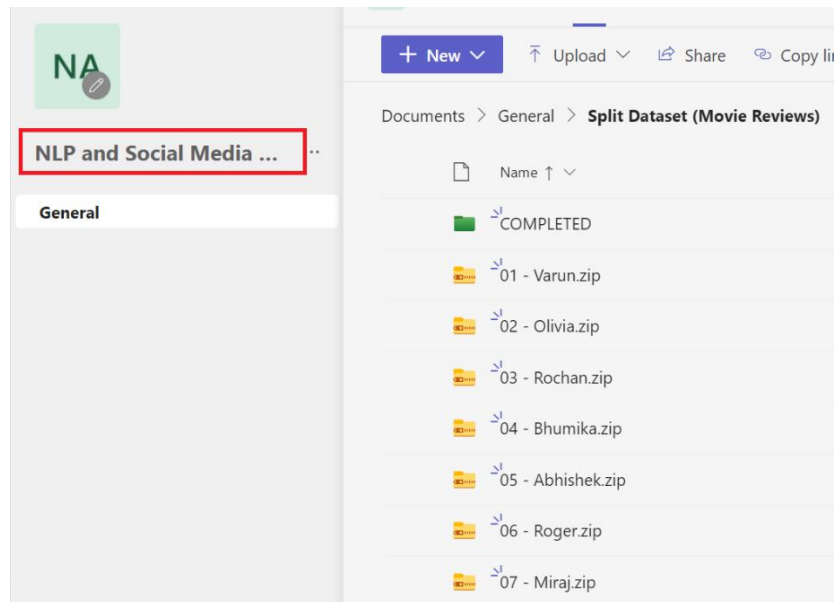
1. Create **split_dataset** folder inside your dataset folder.



Documents > Lambton College > Class > Repository > Projects > Movie-Sentiment-Analysis > dataset

Name	Status	Date modified	Type
split_dataset		11/12/2023 12:54 AM	File folder
credits.csv		11/2/2023 12:24 AM	CSV File
movies.csv		11/1/2023 6:06 PM	CSV File
movies_metadata.csv		11/2/2023 12:24 AM	CSV File
reviews - dummy.csv		11/6/2023 8:46 PM	CSV File
reviews.csv		11/11/2023 8:54 PM	CSV File

2. Check MS teams' group the assigned set of files (zipped file) for each person inside [Split Dataset \(Movie Reviews\)](#). Download the zipped file.



NA

NLP and Social Media ...

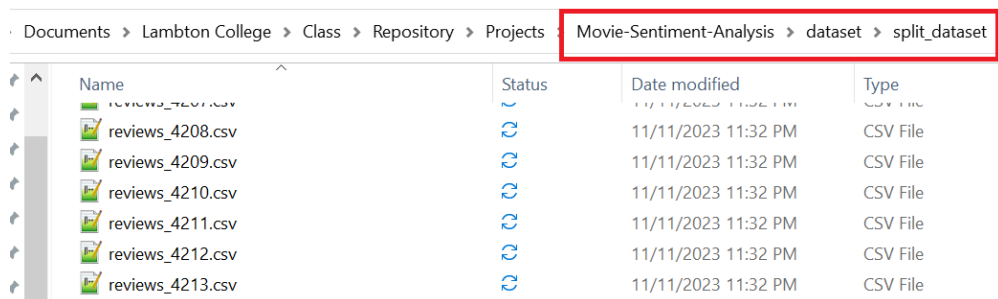
General

+ New Upload Share Copy link

Documents > General > Split Dataset (Movie Reviews)

Name
COMPLETED
01 - Varun.zip
02 - Olivia.zip
03 - Rochan.zip
04 - Bhumika.zip
05 - Abhishek.zip
06 - Roger.zip
07 - Miraj.zip

3. Unzip all files inside the 'split_dataset' directory



Documents > Lambton College > Class > Repository > Projects > Movie-Sentiment-Analysis > dataset > split_dataset

Name	Status	Date modified	Type
reviews_4207.csv		11/11/2023 11:32 PM	CSV File
reviews_4208.csv		11/11/2023 11:32 PM	CSV File
reviews_4209.csv		11/11/2023 11:32 PM	CSV File
reviews_4210.csv		11/11/2023 11:32 PM	CSV File
reviews_4211.csv		11/11/2023 11:32 PM	CSV File
reviews_4212.csv		11/11/2023 11:32 PM	CSV File
reviews_4213.csv		11/11/2023 11:32 PM	CSV File

4. Get the following files from dev branch (or you can simply pull all codes to your branch). Make sure that the version of text_preprocessing.py is the updated one since it's used by sentiment generator.

dev 12 branches 0 tags

Go to file Add file <> Code

This branch is 43 commits ahead of prod. Contribute

ovdlambton Merge pull request #70 from abccastro/task/acastro	f323e23 12 minutes ago	146 commits
dictionary	#62 rolled back changes on stopwords, i.e. separated it from spacy lemma	yesterday
docs	moved files	31 minutes ago
Generate Sentiments using Roberta.i...	#67 added try catch when processing files	13 minutes ago
Movie Sentiment Analysis.ipynb	#43 #44 #61 added code changes for NER including data extraction and...	6 hours ago
README.md	Merge branch 'main' into task/rmais	4 days ago
text_preprocessing.py	Merge branch 'task/vsharma' into task/acastro	34 minutes ago
utils.py	#62 rolled back changes on stopwords, i.e. separated it from spacy lemma	yesterday

5. Run the **Generate Sentiments using Roberta.ipynb**. Make sure all necessary Python libraries are installed.

Notes:

- a. This code will create two folders inside the **split_dataset**, namely **done** and **sentiment**
 - **done**: this folder will contain all old review files that were successfully processed
 - **sentiment**: this folder will contain all new review files with sentiment labelling
- b. The code is re-runnable, and you don't need to reprocess all the files inside the **done** folder

6. Once done with all the files, zip them and place the zipped file inside **COMPLETED** folder in MS teams.

