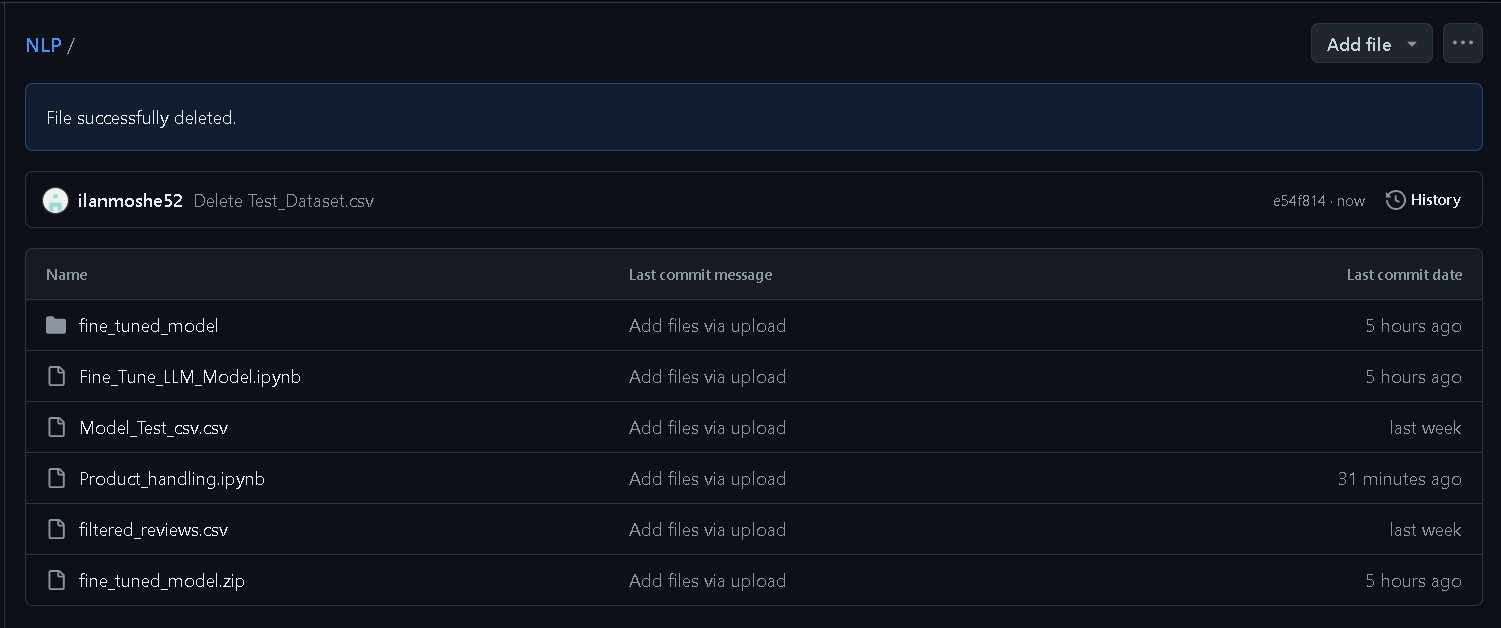
NLP Final Project Sentiment Analysis

Ilan Moshe (024176398)

1. File Description in the GIT:



Data Files:

* Model\_Test\_csv – The dataset used to test the model performance after the fine tune
* Filtered\_reviews.csv – Products customer reviews csv file. The link to the original Reviews file is:

<https://www.kaggle.com/code/mananhingorani/sentiment-analysis-from-customer-reviews/input?select=Reviews.csv>

I have done some processing and trimming to make it a more compact file.

* IMDB\_Dataset\_update.csv – (in the google drive, because of the file size it can’t be stored on the GIT), that is used for fine tuning the basic distilbert model. The original file is from:

<https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews> ,

but I have done some processing and trimming to make it a more compact file.

The Model Files:

* Fine\_tune\_model directory (and Fine\_tune\_model.zip) – is the model based on distilbert model after LLM fine training using a new dataset of sentiment analysis.

Original mode before fine tuning process is here:

[distilbert/distilbert-base-uncased-finetuned-sst-2-english · Hugging Face](https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english)

Code Files:

* Fine\_Tune\_LLM\_Model.ipynb – The file is used to fine tune the basic model with new dataset (IMDB\_Dataset\_update.csv) and evaluate the updated model performance. More detailed information is available in the documentation in the file itself.
* Product\_handling.ipynb – This file performs sentiment analysis of customers reviews per product and defines what are the products that are problematic from the customers point of view. More detailed information is available in the documentation in the file itself.

(There are many graphs in these code files, that are not seen with the GIT viewer, therefore these code files must be viewed in the Google Colab)

General Files

NLP Final Project - Ilan Moshe 024176398.doc – General project explanations