Because of the size of our models and datasets, we didn’t include all the content in the zip file.

To run the different parts of this project, you should download the content of the project’s google drive:

https://drive.google.com/drive/folders/1brblSRUy-tGW83oCjk6CW75gtb4agHpK?usp=sharing

The different parts have been split so you can download the necessary part when running some parts of the project (eg. you need to download “pre-training\_and\_fine-tuning\_bertweet/best\_bertweet.pt” to use it to predict sentiment in “bertweet\_demo.ipynb”)

Organization of the entire folder

This folder contains all the resources required to reproduce the results mentioned in the report:

* the folder “bertweet-retrained” contains the re-trained model version of Bertweet on crypto-tweets.
* the folder “tweet\_datasets” contains all the tweet datasets required to reproduce our results
* the folder “final\_presentation” contains the presentation required as part of the submission
* the folder “fine-tuning\_roberta” contains the notebook to fine-tune roberta
* the folder “pre-training\_and\_fine-tuning\_bertweet” contains the notebook to pre-train Bertweet, the data used in pre-training and the notebook to fine-tune it
* the folder “task1” contains the necessary resources to predict the **sentiment** of manually labelled tweets for both Bertweet and Roberta
* the folder “task2” contains the notebooks to predict the sentiment over periods of time to compute the **causality with financial data**
* the “price” folder contains any financial data used in task 2 to compute the causality with sentiment
* the file “NLP Group 5 Project (2022).ipynb” is **the main notebook** to run when trying to reproduce the results
* the file “roberta\_demo.ipynb” is a **demo** notebook to predict the sentiment of some manually inputted tweets using the Roberta fine-tuned model
* the file “bertweet\_demo.ipynb” is a **demo** notebook to predict the sentiment of some manually inputted tweets using the Bertweet fine-tuned model

Instruction

Please follow below instructions to ensure smooth running of all the notebooks

* Unzip the zipped file and upload the folder at your **My Drive** directory in Google Drive

Demo

1. Please ensure you have turned on GPU runtime to try out the sentiment models
2. Demo Files can be found here
   1. BerTweet – bertweet\_demo.ipynb
   2. RoBerta – Roberta\_demo.ipynb
3. The demo functions can be found the end of each notebook and the function takes a list argument so an example input will be [“I love bitcoin HODL”, “Why is Elon Musk dissing BTC”] and this will output probabilities and predictions for each tweet.