Clear and unambiguous instructions on how to reproduce the predictions from start to finish including data pre-processing, feature extraction, model training and prediction generation. Point out the corresponding file names • • Data Files: o Which data files are being used? Provide all the relevant links. Clearly write out the downloaded file names.

| File Name | External file to use | Steps | how the model was developed and trained\* |
| --- | --- | --- | --- |
| RandomForSpacy.ipynb | rated\_GOOGL.csv | Run the file as is using google colab. SpaCy needs to be installed so colab had the easiest installation. Only need to upload the external file. | Random Forest was trained using the manual sentiment from the csv file. It was trained after the random forest was fine tuned in a different file using grid search. But the different is spaCy is the tokeinzer here. The split is 80/20.  Colab has its own CPU/GPU so no change was done from the default setting. |
| ML\_Models.ipynb | rated\_GOOGL.csv | 1. Imports Libraries 2. Loads Data 3. Cleans Data 4. Splits into test/train sets while tokenizing 5. Define Functions 6. Run through ML methods | Uses nltk word tokenizer to tokenize data. Uses predefined function to fit then evaluate the models (fitting with grid search). The tested models on this notebook are: Logistic Regression, SVM, KNN, Decision Tree, Random Forest, XG Boost, MLP.  -Trained on a 64-bit windows 10 system with 32GB of RAM using the CPU,AMD Ryzen 7 2700X. |
| TextBlob\_and\_Vadder.ipynb | rated\_GOOGL.csv | 1. Import libraries 2. Loads data 3. Cleans data 4. Creates Some Exploratory visualizations/wordclouds. 5. Runs through the lexicon methods 6. Evaluates accuracy | Uses Textblob and Vader to predict sentiment and creates some exploratory visualizations. |
| StockQuotes.ipynb | none | 1.import libraries  2. Set desired date range and interval of stock quotes  3. Enter in the Stock you wish to get financial data on | Scrapes historical stock data using the yahoo finance library. |
| Tweet\_Scraping.ipynb | none | 1. Import libraries 2. Set date range 3. Enter desired hashtag and file name | Scrapes twitter for the queried hashtag and date. |
| bertxstocks.ipynb | Training.1600000.processed.noemoticon.csv  AALprices.csv  BTC-USDprices.csv  GMEprices.csv  SPYprices.csv  WMTprices.csv | 1. Install data and dependencies 2. Rerun cell if np\_utils fails to load 3. Create csv of BERT sentiment scores as well as yahoo finance stock prices 4. Plot sentiment alongside actual delta with acc method | OS: OpenSUSE Tumbleweed (Kernel 5.12)  RAM: 16 GB  Disk Space: 100 GB  CPU: Intel i7-7700 |

\* how the model was developed and trained( including OS, memory (RAM), disk space, CPU/GPU used, and any required environment configurations required to execute the code):