**Predicting and Rebalancing Portfolios**

**Modules:**

**Scraper**

* scraping\_cleaning\_sec.ipynb notebook is used to scrape all sentiment data. It also takes care of cleaning and parsing HTML into raw text files and checkpoints its progress so extraction can resume later.

**Sentiment Analyzer**

* Sentiment\_Analysis\_SEC.ipynb contains all the code extract sentiment scores from SEC filings.
* Sentiment\_Stocks\_Visualization.ipynb is used to generate stock trends vs. the calculated sentiments.

**Stock Predictor**

* lstmstocks-sentiments.ipynb is used to train models with and without sentiments. This notebook also performs all the feature engineering such as windowing, cascading, merging and interpolation.\
* generate-predictions-csv.ipynb generates predicted data frames using saved models from earlier training.

**Portfolio Generator and Optimizer**

* generating\_portfolio.ipynb constructs stable portfolios using metrics such as correlation, covariance, Sharpe ratio and volatility. Also creates related visualizations that help select stable portfolios.
* Main\_CNN.ipynb optimizes the constructed portfolios by rebalancing them using a CNN model.

**Miscellaneous Files**

* These consists of various files which are used for data pre-processing and various visualizations.

**CASSANDRA**

* Download JAVA SE Runtime Environment using the .exe file from <https://www.oracle.com/java/technologies/javase-jre8-downloads.html>
* Download Apache Cassandra from <https://cassandra.apache.org/download/>
* Save the downloaded Cassandra files in program files.
* Install python 2.7 as Cassandra doesn’t support python 3.
* Open the command window in the directory of Cassandra and run Cassandra.bat in the bin folder
* Now run the cqlsh file in the command window while the bat file running, now we can type the commands and use Cassandra.
* To export the data in Cassandra directly into python files, install the Cassandra- driver and import it in your files.
* The code used for exporting the desired queries can be seen in the Cassandra\_exporter.py file.

**Libraries and requirements:**

* Jupyter
* NLTK VADER
* Keras
* Scikit-learn
* fastai datepart
* Pandas
* Numpy
* BeautifulSoup
* Stocker
* Quandl
* Plotly
* Tensorboard
* CASSANDRA
* CASSANDRA-DRIVER