

Documentation

Hrithik Singh, Anirudha Anekal

June - August 2023

Abstract

We are creating a GAN model and using social media data to find anomalies in the stock market. We are focusing on Indian stocks. We are also building a website to run and provide easy access to the model.

1 Technologies and modules:

1.1 GAN Modules

- haweslib: This is to generate the values from the stock data. We are taking buy and sell orders and performing hawkes process on them to generate probaility values. These values are then used to generate images.
- PIL: This is used to generate images from the values generated by haweslib.
- matplotlib: This is used to generate the various graphs. We have utilised it to judge the model's error functions.
- pandas: This is for data management. This is the module that provides us with functionality of a dataframe.
- numpy & scipy: This is for the various mathematical functions and processes required by us. A few other python modules also use them internally.
- Various tensorflow modules: We have used a wide variety of modules from tensorflow. Modules choosen as per requirement specifications.
- The GAN Module: This is the actual GAN module. It takes the images generated by PIL and trains the GAN model on them.

1.2 Front End Technologies

- HTML: This is used to create the basic front end of the website.
- Bootstrap + CSS: This is used to style the website.
- Highcharts: This is a JavaScript library that is used to create graphs. It has been used to make the graph.

1.3 Middle and Back End Technologies

- Vue: This is the mainframe work used in this project. It is used to create a dynamic front end of the website. It also acts as the middleware, connecting the front end to the back end.
- Flask: This is used as a server. It accepts requests from Vue and processes them. It has been used mainly to call nselib. It also calls postgresQL.
- Express: This is also running as a server. Its main role is to perform calls to nse india.
- PostgreSQL: This is used to create the database for the website. It mainly stores news data.
- nselib: This is used to retrieve data about NSE stock market. We are using it to get the list of companies and their stock symbols. We use the class captial_market to retrieve day prices of a stock.
- snsrape: This is used to make search requests to twitter.
- praw and pushshift: These are used to make search requests to reddit.

2 Discarded modules and technologies

- streamlit: This was used to create the first iteration of the front end of the website. It was discarded because it was not flexible enough.
- pyhawkes: This was an alternative to hawkeslib. It was not used as hawkeslib proved to be enough.
- Selenium: This was used to make special requests to the nse api. However this was removed as the method was a form of CSRF (Cross Site Request Forgery). We are instead using an open-source database of news articles.