**ReadMe File**

**Stock Recommendation System**

List of Executables

1. swarm.py
2. fractal.py
3. ClusteringBasedOnPopularity-Kmeans+Hierarchical+MeanShift.ipynb
4. ClusteringBasedOnVolatility-Kmeans+hierarchical+MeanShift.ipynb
5. kvalue+PAM.R
6. cleanAndMergeFiles.ipynb
7. featureSelection.r

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**fractal.py**

Language : python

Run From : Terminal

Files Required : allstocks.csv

To Run:

1. Open Terminal
2. Change to project directory using cd <directory\_name>/Code/Swarm+FractalDimensions/
3. Run the following command : python fractal.py

Estimated time to run : few seconds

Output:

The terminal will output a scaled version of slope values for each stock.

The output includes the log-log plot of a random stock on a popup.

Attributes:

Can be changed in the attributes dictionary on line 194.

* scales : list of scale values to be used (1,n/5), update logscales accordingly if this is updated
* logscales : list of log base 2 of scales list, update scales accordingly if this is updated
* n : total number of days being considered (0,900)
* stock : to plot log-log plot for i’th stock

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**swarm.py**

Language : python

Run From : Terminal

Files Required : allstocks.csv

To Run:

1. Open Terminal
2. Change to project directory using cd <directory\_name>/Code/Swarm+FractalDimensions/
3. Run the following command : python swarm.py

Estimated time to run : 1 hour

Output:

Swarming will occur on terminal screen. Program will output a 2D array every few seconds which corresponds to a time period in the swarming. Due to rapid printing the effect of swarming is noticed.

Attributes:

All attribute values can be changed in the attributes dictionary in line 508.

* attw - weight for attraction vector
* aligw - weight for alignment vector
* oriw - weight for momentum
* threshold - threshold value to find similar birds
* tpsize - time period
* tpmode - mean or median value for the interval of size tpsize
* mindist - minimum separation distance, it is recommended that this value is set to 1
* maxdist - maximum distance upto which a bird can see.

Simulation duration can be chnaged on line 659 by calling simulate for any value between 1 and 1000/(tpsize+1).

Upto 4 similarity types can be chosen and each one shows a significant difference. Canbe changed by changing ‘simtype’ on line 522 to any value in {1,2,3,4}.

‘simtype’ : 1: function defined on line 264

‘simtype’ : 2: function defined on line 289

‘simtype’ : 3: function defined on line 294

‘simtype’ : 4: function defined on line 316

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ClusteringBasedOnPopularity-Kmeans+Hierarchical+MeanShift.ipynb**

Language : python

Run From : Jupyter Notebook

Files Required : Popularity\_voltraded.xlsx

To Run:

1. Open the file ClusteringBasedOnPopularity-Kmeans+Hierarchical+MeanShift.ipynb in Based\_On\_Popularity directory in jupyter Notebook
2. Run all the commands

Output:

Clusters graph based on Popularity using K-Means, Hierarchical and MeanShift Clustering

**ClusteringBasedOnVolatility-Kmeans+hierarchical+MeanShift.ipynb**

Language : python

Run From : Jupyter Notebook

Files Required : volatility\_data.xlsx,filenew2.xlsx

To Run:

1. Open the file ClusteringBasedOnVolatility-Kmeans+hierarchical+MeanShift.ipynb in Based\_On\_Volatility directory in jupyter Notebook
2. Run all the commands

Output:

Clusters graph based on Volatility using K-Means, Hierarchical and MeanShift Clustering

**kvalue+PAM.R**

Language : R

Run From : RStudio

Files Required : new\_data.csv

To Run:

1. Open the file kvalue+PAM.R in ChooseAppropriateKvalue+PAMClustering directory in RStudio
2. Run the file

Output:

Clusters graph based on Popularity and Volatility using PAM clustering.

Obtaining optimal k value.

Note: please add the full directory path if the R code gives a file not found error.

**cleanAndMergeFiles.ipynb**

Language : python

Run From : Jupyter Notebook

Files Required : All stocks csv file

To Run:

1. Open the file cleanAndMergeFiles.ipynb in Stock CSV's-20181205T032255Z-001 directory in Jupyter Notebook.
2. Run the file.

Output:

Merged csv file

**featureSelection.R**

Language : R

Run From : RStudio

Files Required : featureSelection.csv

To Run:

1. Open the file featureSelection.R in featureSelection directory in RStudio
2. Run the file

Note: please add the full directory path if the R code gives a file not found error.