Problem Statement

You are given monthly stock price data for company X. The problem is simple:

Use any machine learning technique to create a model that predicts whether the next month return for X is positive (+1) or negative (-1).

Data Format

The data format for both training and consumption by the final function is a csv with the following columns.

The Training set should be used to build your machine learning models. For the training set, we provide the target value (next month return sign) for each month. Your model will be based on the "features" open, high, low, close, return, volatility, volume. You can also use feature engineering to create new features.

The Test set will be used to see how well your model performs on the unseen data. For the test set, we do not provide the target value. For each row in the test set, use your model to predict the target value (return sign for the next month).

Feature columns:

- Open : *open price of the month*
- High: *highest price during the month*
- Low: *lowest price during the month*
- Close: *Closing price for the month*
- Return: *Monthly return*
- Volatility: *Monthly annualized vol*
- Volume: *Traded volume*
- Prev Volume: *Prev traded volume*

Target column:

- Next month return sign: *Is the return for the next month positive (+1) or negative(-1)*

Submission Format

You shall submit the following files:

- A word document describing your approach to the problem in a single page at the maximum as Q2_Explanation.docx
- A python file containing a function which consumes file paths for the data, prediction & model and saves your predictions to prediction file. A sample function is provided below as Q2.py.
- A pickled file containing all information you need from your trained model, we will run this trained model pickle file to generate predictions for out-of-sample data as Q2.pickle.
- The updated csv which was provided with this problem with your predictions in the format mentioned below as Q2.csv.It should contain 30 rows in one column without header containing your prediction.

0
1
0
1
1

Put the files in a zip and upload the zip file. Do note that you need not put the files in a folder when zipping.

A sample python file with the format has been provided to you as Q2.py, you need to fill your logic in the function generate_model that saves a pickle file and uses pickle file to generate results.

Evaluation Criteria

We shall measure submissions on the criteria:

$\Sigma\ \{all\ \text{dates}\}\ (PredictedScore - ActualScore)^2$

Primarily we shall use the prediction file provided for test to form a cutoff.

We shall use extra out-of-sample data not provided to you for final scoring.