

Stock Price Models

Members and Emails

| | |
|-------------------------|--------------------------------|
| Hou, Matthew Zheng-Ming | matthewhou732@gmail.com |
| Iglesias, Martin | 100384068@alumnos.uc3m.es |
| Li, Xiang | xli1026@terpmail.umd.edu |
| Nash, Robert Franklin | rnash@terpmail.umd.edu |
| Zhang, Ruilin | 472075013ruilinzhang@gmail.com |

Problem: Efficient Market Hypothesis - Market is perfectly priced, always random
Irrationality - The market is irrational, based on irrational human behaviors, ML likes reliable data

Hypothesis:

LSTM will be the most consistent and accurate in predicting future stock prices

Motivation:

Predicting stock prices has the potential to yield great profits. By analyzing historical stock prices, we can train a model to predict future stock prices which could help us to determine whether buying a particular stock would be good or not. Prediction can also forecast the market as a whole, helping people determine the best financial decisions today.

Learning method:

Linear regression, KNN, Multilayer Perceptron, Support Vector Machine, LSTM (Long short term model)

Language, environment, Methodology:

- Python (scikit-learn)
- Tensorflow - R
- SVM
- LSTM → commonly used to predict future values based on past data, also commonly used for chatbots and other NLP applications

An LSTM model is primarily used to predict the future, based on past data. The input data is a 3D array where the first dimension is the number of observations, second dimension represents the time series, and the third dimension represents the number of features. The output data would be a 2D array where each element would represent the resulting output of each time series.

Source of data:

Google Finance

<https://datahub.io/collections/stock-market-data>

<https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs>

Evaluation Design:

In order to evaluate our model, we need to split our data into one test set and one training set. The test set will be used to monitor the models performance and provide us with feedback on how to improve the model. We will compare all models by RMSE .