CS-GY 6923 Machine Learning
Project Proposal

Runze Li
rl50xx@nyu.edu

March 8, 2024

1 What problem do you aim to tackle with? Why is it an important problem to you?

This project aims to solve the problem of stock market prediction. Specifically, we make predictions about Apple's future stock price. This is an important question because stock market predictions can help investors make more informed investment decisions, reduce investment risks, and increase investment returns. By using advanced machine learning models, we can try to capture the patterns behind stock prices, providing valuable insights to investors.

2 What dataset do you plan to use? Are you going to collect data yourself? If not, is the dataset already preprocessed?

I'm not going to collect the data myself. I will use Apple stock market data on Kaggle. This dataset includes daily opening price, highest price, lowest price, etc. It has already contained the key information needed to make predictions. However, this dataset isn't currently preprocessed, so I will perform necessary data cleaning and other preprocessing process.

Dataset: Apple Stock Price

3 Which ML models do you plan to use?

I plan to use the following three machine learning models for stock price prediction:

1. Support vector machine (SVM)

2. Long short-term memory network (LSTM)

3. Convolutional neural network (CNN)

4 Which accuracy/error measures do you plan to use?

To evaluate the performance of the model, I plan to use the following matrics:

1. Mean squared error (MSE)

2. Root mean squared error ($RMSE$)

3. Mean absolute error (MAE)

4. Coefficient of determination (R^2)

These metrics will help me evaluate the accuracy and reliability of the model's predictions.

5 Are there previously documented results on the same problem? If yes, how do you plan to improve previous results?

Existing researches have shown that machine learning models can be used for stock market prediction. So my goal is to improve the accuracy of Apple stock price predictions by comparing different models and using advanced machine learning techniques. I will try to improve the results by optimizing model configuration, enhancing data preprocessing, and other strategies.