October 9, 2022

MAIS 202 – Project deliverable 1

Project Idea:

This project idea was inspired by the books titled "Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with python, 2nd Edition" and "Advances in Financial Machine Learning." The project: **Machine learning-driven quantitative trading bot**," will follow much of the methodologies presented in the previously mentioned textbooks, alongside diverse resources on the topic of quantitative trading strategies. Given the complexity and intricacy of this project, the scope, features, and particular methodologies may vary and change throughout the development of the model in order to fit the timelines of the bootcamp. The model's objective is to predict returns on US stocks and ETFs.

Datasets:

The three datasets will be the Intraday Minute Bar for the S&P 500, Stock market Data for all NASDAQ, NYSE, and S&P500 listed companies, and the US funds dataset. These datasets give all relative financial information related to US ETFs and publicly traded companies.

Methodology:

a) Data Preprocessing: Given that the datasets range in different periods (years/days), the data will first need to be arranged so as to match the periods available in all three datasets.

b) Machine Learning Model: In order to measure the portfolio performance, we want to be able to estimate the returns and the risk of the investment. This will be done through "ratios that compute a measure of return per unit of risk." Decision trees will be used for prediction of "asset returns and price moves," alongside Bayesian ML to quantify uncertainty.

c) Evaluation metric: Given that decision trees will be used for a regression problem, the Mean squared error will be used as a metric. Additional metrics may be added as the machine-learning model gets finalized.

Application:

For the web app, given that the financial data is not up to date, the user will have the option to input a specific point in the past to which the model will analyze the data leading to that day and output the best investment as well as the possible return and risk as a ratio.