Problem Statement:

Intuition Technologies, a quantitative trading firm, is interested in developing an algorithmic trading platform capable of identifying sizeable opportunities in the stock market and generating returns greater than those available from simply investing in the S&P 500.

Context:

Hedge Funds, high frequency trading firms, quant firms and big brokerage houses all participate in algorithmic trading strategies to maximize risk-adjusted returns. Intuition Technologies is dedicated to developing technologies that exceed the expectations of its investors. Intuition is interested in building and continuously improving their algorithmic trading models with machine learning to predict the directional movement of assets in the stock market. The data we will use is comprised of daily historical stock price data for multiple S&P 500 companies.

We will transform the historical stock price datasets to return series, which will ultimately act as a signal for us to process and extract information from. We will generate performance metrics, technical indicators, and back tests for our project to ensure positive returns in our algorithmic trading. Our goal as a data scientist is to create an execution strategy using the aforementioned tools to inform the program of when to buy or sell a stock. An important aspect of this project is to ensure the model or process can generalize to future market environments.

Success Criteria:

Develop a model to predict the directional movement of a stock price (up or down) and an execution strategy for making the trade (buy or sell) at the opportune time. The execution strategy should be capable of sustaining returns greater than the S&P 500 for at least a year cycle of the fund, and should account for any risk associated with trades, such as executions fees. Fundamentally, the model will represent a quantitative investing strategy limited to unlevered long-only stock portfolio.

Scope + Risks:

The model should be able to identify and alert the analyst of any anomalous behaviors inherent to the strategy, such as unusually short or long holding periods that result in subpar returns. The returns of the execution strategy could prove to be reliant on a period greater than one year to beat the returns of the S&P 500.

Decision Maker/Other Stakeholders:

Fund Manager

Senior Quantitative Researcher

Data Scientist – Matt Miller

Constraints:

- Limitations of the historical data may hinder the model from identifying the economic environment
- Size of the data may require using a limited number of features or technical indicators
- An unlevered long-only stock portfolio will be constructed, and therefore we may inadvertently expose ourselves to risk otherwise hedged with a short portfolio

Data Sources:

The data is a compilation of end of day stock prices for one hundred different companies. The data is stored in one hundred different .csv files. Each .csv file contains the date, open, close, high, low, and volume for each day from the beginning of 2010 until the end of 2019. In addition, we have the end of day data for the SPY from 2010 to 2019. Overall the data is sized at around 11.0 MB.