Trading using Machine Learning

Description

The objective of this mini project is to implement a trading algorithm making use of machine learning. In this project we develop a random forest model to predict if the return will be positive or negative over next 5 days based on the price history and twitter sentiment.

Features

Many features were tried for the modeling and the features listed below were found to be useful and used to build final model

- Last 20 day daily return
- Last 20 day twitter sentiment
- Difference between previous day price and last 20 day moving average

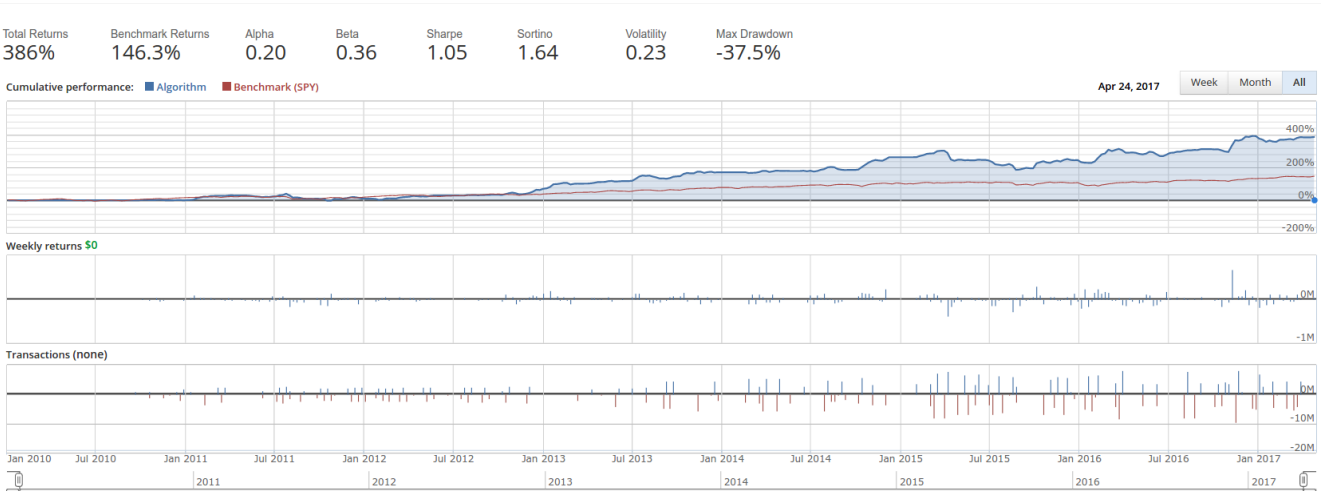
Number of features: 41

Trading Strategy

Classification models are created to predict return using the price and sentiment features (listed in previous section) for the stocks being considered. One random forest model is built for each of the stock and the stock For each stock 150 historical observations are considered for modeling. The next 5 day return is the outcome that is predicted. If the 5 day return is more than 0.5% then the outcome is considered as Positive, if it is less than -0.5% then outcome is Negative and the outcome is neutral otherwise. Initially the model is trained on 80% of the data and scored on 20% of the data. The stock for which the model score is highest is selected for trading. The model is then retrained on 100% data and based on the model prediction a long or short position is established. The predictions are generated every 5 days (i.e the number of days for which the model classifies the return) and trades are placed as per model suggestion.

Back Testing

The trading algorithm is implemented in quantopian and a back test is conducted from Jan 4,2010 to Apr 25, 2017. The snap shot of result is provided below. The algorithm performed fine and produced returns higher than the overall market



Summary

The algorithm was able to reasonably identify if the next 5 day return would be positive and negative. The return produced by the algorithm was better than over all market.