Machine Learning Project

Long Short / Weekly Buy Sell Signals for S&P 500

Team 1: -

Achal Shah - ads925

Prajwal Pitlehra - pp2696

Tania Suresh - tns9572

Data

The S&P 500, is a stock market index tracking the stock performance of 500 large companies listed on stock exchanges in the United States.

Source of Data - Yahoo Finance(yfinance) is a module in Python, which is used to collect online data, and with it, we can collect the financial data of Yahoo. It's an open-source tool that uses Yahoo's publicly available APIs, and is intended for research and educational purposes.

Training Data Period - 1928 to 2008 with Weekly Frequency

Testing Data Period - 2009 to 2021 with Weekly Frequency

Source - https://pypi.org/project/yfinance/

Feature Engineering

1. Fractional Differentiation - General Practice is to take returns/log prices for stationarity but that leads to memory loss on the other hand price has memory but is not stationary. Therefore, for trade-off used fractional differentiation instead of integer differentiation.
   Library - https://github.com/simaki/fracdiff

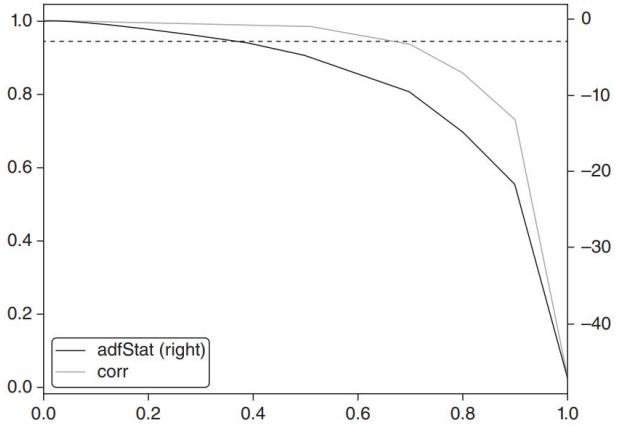


FIGURE 5.5 ADF statistic as a function of *d*, on E-mini S&P 500 futures log-prices

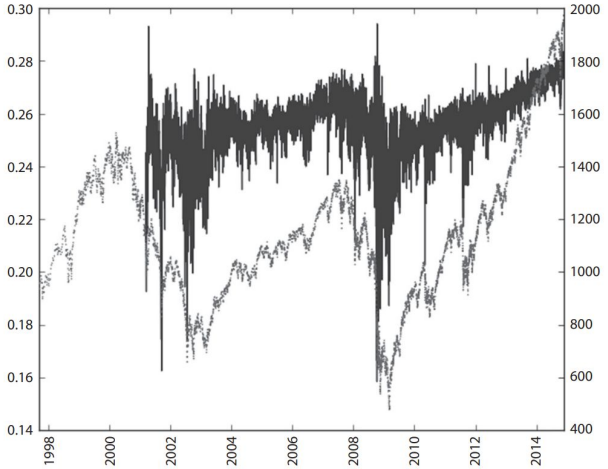


FIGURE 5.4 Fractional differentiation after controlling for weight loss with a fixed-width window

NYU TANDON

Feature Engineering

2. Technical Indicators: -

a. Average True Range - The true range indicator is taken as the greatest of the following: current high less the current low; the absolute value of the current high less the previous close; and the absolute value of the current low less the previous close. The ATR is then a moving average. (Volatility)
b. Relative Strength Index - The index compares a security's strength on days when prices go up to its strength on days when prices go down. (Momentum)
c. On-Balance Volume - It is a technical indicator of momentum, using volume changes to make price predictions. (Volume)
d. Weighted Close Price - Weighted Average Closing Price means the average of the closing bid and asked prices of the Common Stock quoted in the over-the-counter market summary or the last reported sale price. (Price)
e. Hilbert Transform - In mathematics and in signal processing, the Hilbert transform is a specific linear operator that takes a function, u(t) of a real variable and produces another function of a real variable H(u). (Trend vs Cycle)
f. Kalman Filter - For statistics and control theory, Kalman filtering, also known as linear quadratic estimation (LQE), is an algorithm that uses a series of measurements observed over time, including statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone. (minimizes the state of error variance)
Libraries - https://mrjbq7.github.io/ta-lib/
https://pykalman.github.io/

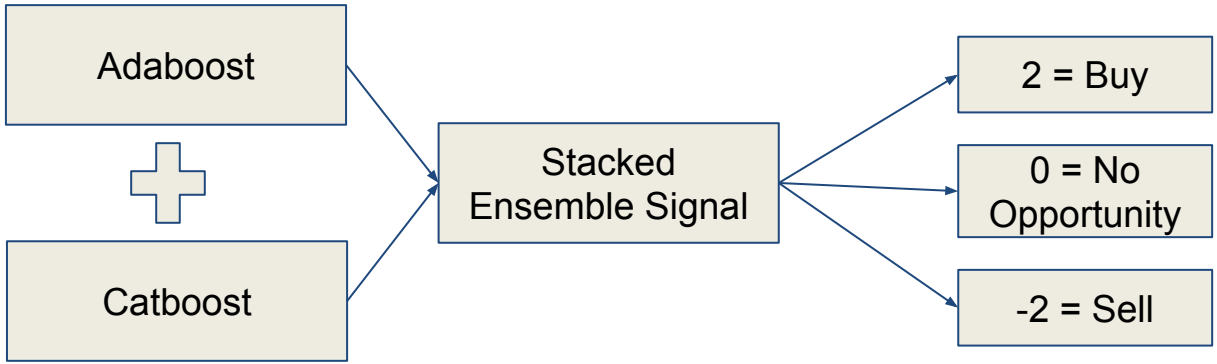
Source - https://www.investopedia.com/terms/t/technicalindicator.asp

Model Stacked Ensemble and Predictions

Independent variables - Open, High, Low, Adjacent Close, Volume, Average True Range, Relative Strength Index, On-Balance Volume, Weighted Close Price, Hilbert Transform, Kalman Filter
Dependent variable - Shift Weekly Price and for positive change transform into 1 and for negative change transform into -1 for classification.

Models Used - Adaboost Classifier and Catboost Classifier

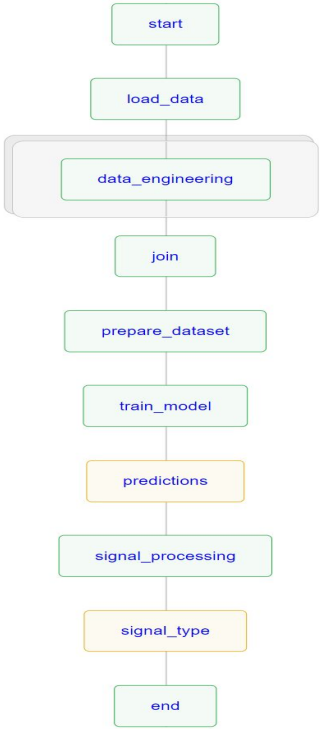
Signals -

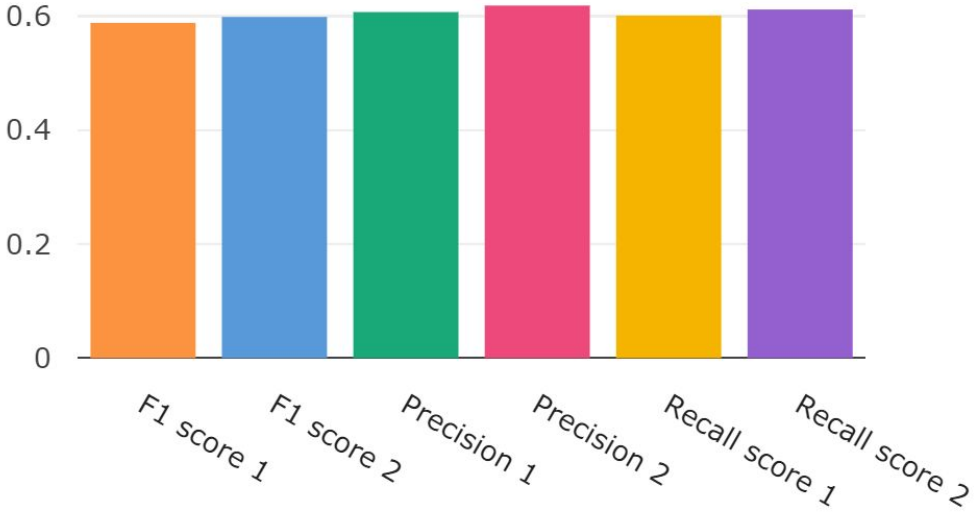


NYU TANDON

Source - https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/

Metaflow, Metrics and Webpage Link





score 1 - Adaboost Model
score 2 - Catboost Model

Google Cloud Platform Link - https://mlfin-370922.ue.r.appspot.com

NYU TANDON

Source - https://www.youtube.com/watch?v=xcODUk0o6tU/

Thank you

Source - https://pypi.org/project/yfinance/