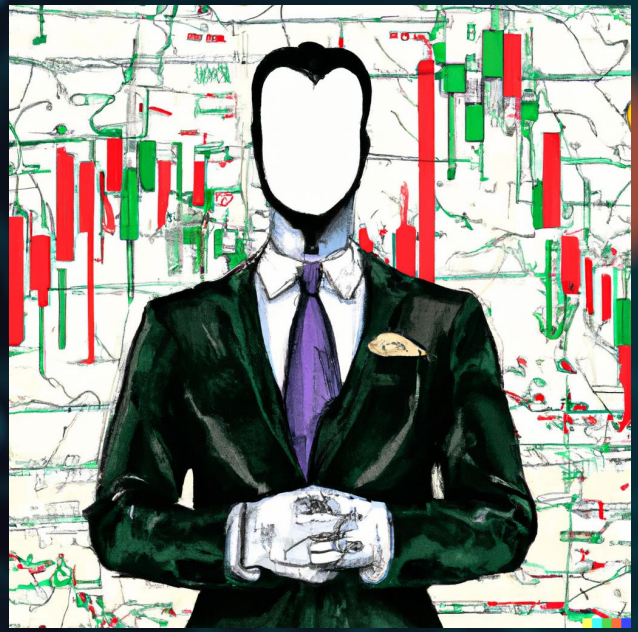
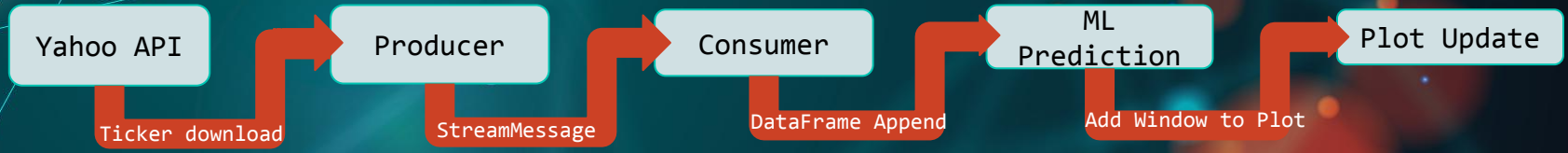


Yahoo Finance Streaming

Sebastian Fux ♦ Marcin Ciesiul ♦ Steven Albiez

Structured Streaming



Structured Streaming

Combination of streaming and batch processing. Using the Yahoo Finance python package stock prices are sent to the producer (batch process). The consumer works as a live stream and processes the data continuously (streaming). Minute-level data is aggregated by the consumer to windows with a length of 5 minutes which contains an open, high, low & closing stock price. As soon as the consumer has 5 new data points a new window is created and stored locally. The ML Prediction/Plot Update are again set up as a batch process. A file watcher is implemented and as soon as new data becomes available a a new prediction is conducted and the candlestick plot updated.

Storing the Output

The producer generates a lot of redundant data, since Yahoo Finance package delivers minute-by-minute data for at least one day. In order to keep the ‘clean’ and tractable the consumer appends the data, i.e. every 5th minute a new window is added to the structured stream on which the sql operations are done.

Data

Yahoo Finance API provides:

- stock quotes
- historical prices
- financial statements
- **for our analysis we rely on adjusted close and time stamp from yahoo finance**

We use data minute-by-minute data updates from the last 5 days and summarize this data in 5 minute windows.

```
%pip install yfinance

def get_stock_prices(ticker, counter):

    period = '5d'
    if counter != 0:
        period = '1d'

    df = yf.download(tickers=ticker, period=period, interval='1m')
    df['id'] = df.index.strftime('%Y-%m-%d %H:%M:%S')
    # Extract date from the index
    df['date'] = df.index.strftime('%Y-%m-%d')
    # Extract time from the index
    df['time'] = df.index.strftime('%H:%M:%S')

    df.to_csv('stock_data/' + ticker + '_df.csv', index=False, header=False)
    print(f"New DataFrame saved for ticker: {ticker}")
    return
```

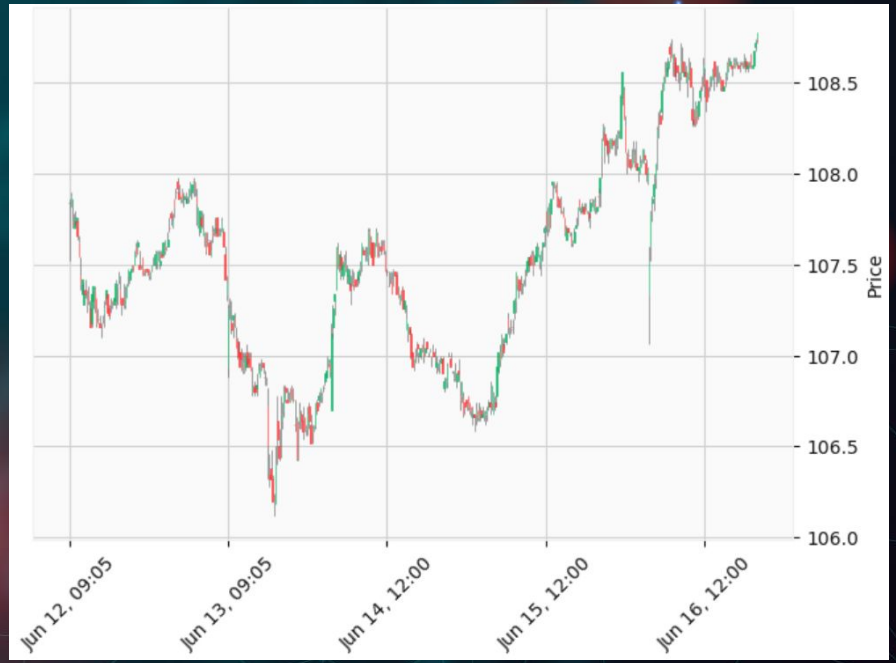
Results - Candlestick Plots

Near-time candlestick plots

Stock prices, which are updated every minute, are summarized in non-overlapping windows each having a length of 5 min. For each window candlestick plots are generated which show the open, high, low and close of that window.

All 5 minutes a new candle stick is added to the plot on the right side. If the stock price has increased during the trading window it is colored in green and if stock price decreased it is colored in red.

The candle sticks can be used to generate trading signals, i.e. whether a stock should be sold or bought.



Results - Trading Strategies I

Trading strategies

We leverage the Spark ML frame and generate trading signal using the Logistic Regression (LR) approach. The LR is an approach to classify variables, i.e. it can be used to classify a stock whether it is worth to buy or not.

In our example, we used to LR to classify whether the open from the subsequent window will be larger than the open from the current window. If the prediction showed that:

Opening Stock Price (t+1) > Opening Stock Price(t)

This was considered a buy signal and the stock was bought at the opening stock price for window t+1 and sold at the closing price for the same window.

If the prediction showed that:

Opening Stock Price (t+1) < Opening Stock Price(t)

it was treated as a short selling signal meaning that we sold the stock for the opening price and bought it back at the closing price at the end of the window.

In order to predict the opening stock price for next window, we relied on the open stock price of the previous windows and a moving average of the opening prices for either 5 and 15 windows. Based on this data, the ML approach predicted, whether the next opening stock price will be higher than the current one.

All data was used for training except the most recent observation window which was used as testing data.

Results - Trading Strategies II

Trading strategy for Nestle

The trading strategy was run for NESN.SW ran from 15:00-16:00 on 16.06.2023. One hour results in 12 trading windows each lasting for 5 min. The summary of trading is as follows:

Total number of prediction:	12
Correct predictions:	4
% of correct predictions:	33%
Return for each of trading windows:	-0.036%, -0.018%, -0.018%, -0.018%, 0.037%, -0.055%, 0.0734%, 0.000%, -0.109%, 0.000%, 0.018%, -0.055%
Cumulative return for 1 hour of trading:	-0.18%

```
Short sell signal generated!

Lucky you - the prediction was very good. The return for that window is: 0.03737%. Congrats!

The cumulative return over observed windows is: -0.06%!
```