CRYPTOCURRENCY – TRADING AND INVESTMENT STRATEGY

HTTPS://GITHUB.COM/ALAKHAGR/CS419-PROJECT

Problem Statement - We plan to download tick-level trade data from the Bitmex exchange, which has the highest volume for cryptocurrency.We will run a script to process the trade data into candlestick price bins of different time intervals(1 min, 15 min, 30 min, 1H) along with various stock market indicators.

We divide the project into 2 parts:

A) We plan to identify candlestick patterns(signals) which indicate bullish or bearish sentiment of market and build trading strategies using these signals. However, these signals don’t work with very good probability of success. Our aim in this project is to increase our probability of success by filtering the earlier identified signals using xgboost algorithm. Our XGBOOST algorithm will take in various features like stock market indicators and other market indicators for each signal and predict whether we should act on that signal. It is a classification problem.

Problem Statement - (B) We are planning to build an RNN on the Bitmex data to predict the closing prices of securities for a given time interval. For this we plan on using LSTM library. The basic idea behind using RNN is that a perceptron doesn’t keep the history of price data. However, for correctly predicting the prices, we need to analyze the historical data and update the weight matrix accordingly. We also plan on plotting the graphs between the predicted and the actual prices and train in an efficient way such that the error is below a certain threshold.

Dataset Employed - Bitmex

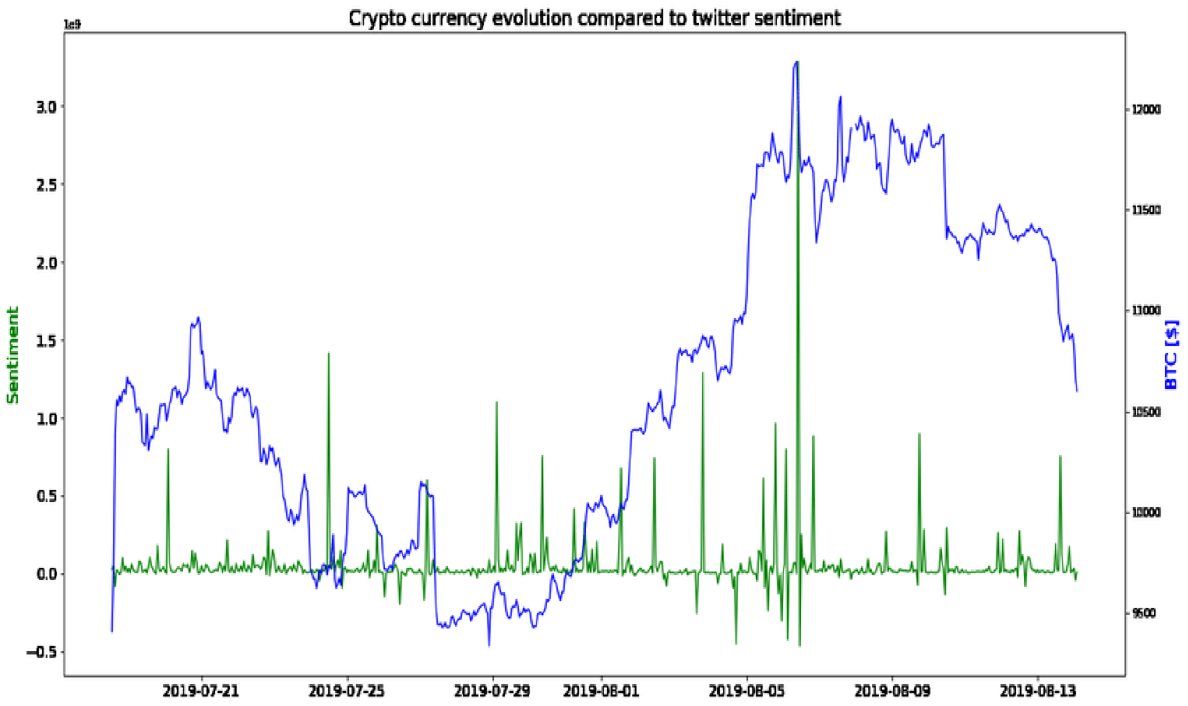
The Bitmex trading platform was used to collect the historical data of Bitcoin. It contains closing, opening, high and low prices as well as volume traded over different time frames.

Data Extraction

```
apiKey = 'LAqUlngMIQkIUjXMUreyu3qn'
apiSecret = 'chNOOS4KvNXR_Xq4k4c9qsfoKWvnDecLATCRlcBwyKDYnWgO'

#
# Simple GET
#
verb = 'GET'
path = '/api/v1/instrument'
expires = 1518064236 # 2018-02-08T04:30:36Z
data = ''

# HEX(HMAC_SHA256(apiSecret, 'GET/api/v1/instrument1518064236'))
# Result is:
# 'c7682d435d0cfe87c16098df34ef2eb5a549d4c5a3c2b1f0f77b8af73423bf00'
signature = HEX(HMAC_SHA256(apiSecret, verb + path + str(expires) + data))

#
# GET with complex querystring (value is URL-encoded)
#
verb = 'GET'
# Note url-encoding on querystring - this is '/api/v1/instrument?filter={"symbol": "XBTM15"}'
# Be sure to HMAC *exactly* what is sent on the wire
path = '/api/v1/instrument?filter=%7B%22symbol%22%3A+%22XBTM15%22%7D'
expires = 1518064237 # 2018-02-08T04:30:37Z
data = ''

# HEX(HMAC_SHA256(apiSecret, 'GET/api/v1/instrument?filter=%7B%22symbol%22%3A+%22XBTM15%22%7D1518064237'))
# Result is:
# 'e2f422547eecb5b3cb29ade2127e21b858b235b386bfa45e1c1756eb3383919f'
signature = HEX(HMAC_SHA256(apiSecret, verb + path + str(expires) + data))

#
```

The Bitmex python library along with the API key was used to extract data from the bitmex trading platform as a csv file

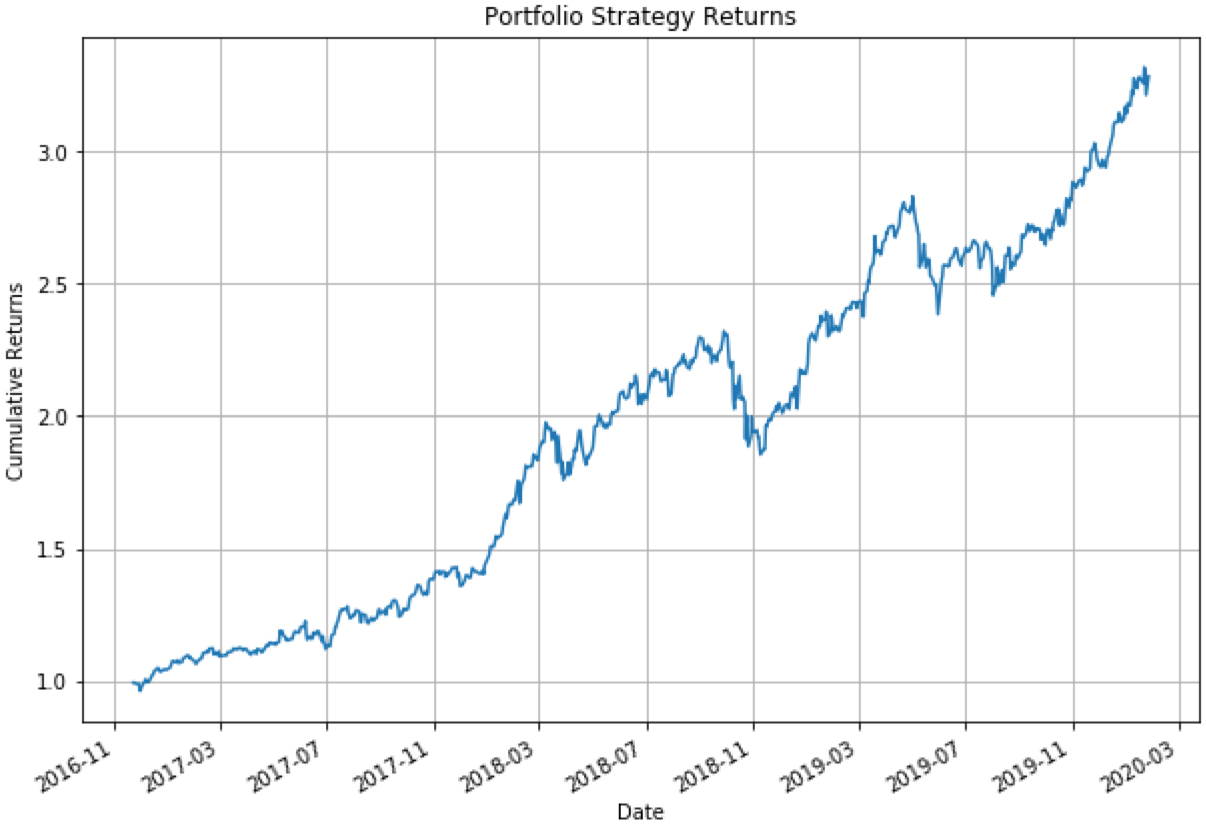
Sentiment Analysis and Strategy Making



- The attitude or overall view of investors, or their feeling for a single securities or the whole financial market, is known as market sentiment. This stock market mood study has an influence on the character of the market trend. Investors can use sentiment research to understand the price movement in the stock market. In volatile economic situations, the risk reward ratio assists investors in predicting the pulse of market activity.

- Social media channels, on-chain data, and other crypto indicators may all be used to get a sense of how a currency or project is doing. We are more likely to make better trading or investing judgments if we have a better grasp of the market's present sentiments.

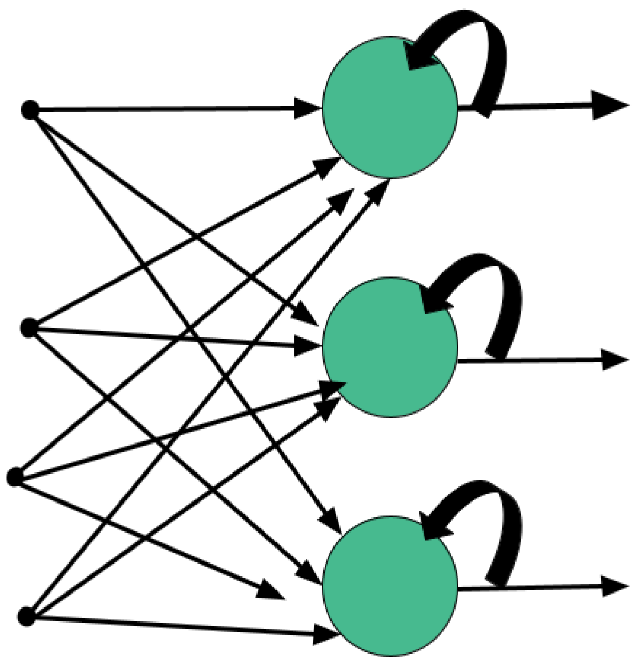
XgBoost



- XGBoost is a distributed gradient boosting toolkit that has been tuned for efficiency, flexibility, and portability. It uses the Gradient Boosting framework to create machine learning algorithms. XGBoost is a parallel tree boosting (also known as GBDT, GBM) algorithm that solves a variety of data science issues quickly and accurately. The same algorithm may tackle problems with billions of instances in a distributed environment (Hadoop, SGE, MPI).

- Gradient boosting is an iterative approach for converting weak learners to strong learners. The term XGBoost comes from the technical purpose of pushing the computing resources for boosted tree algorithms to their maximum. XGBoost has shown to be a very strong machine learning approach since its inception in 2014, and it is often the go-to algorithm in many Machine Learning contests.

RNN (Recurrent neural network)

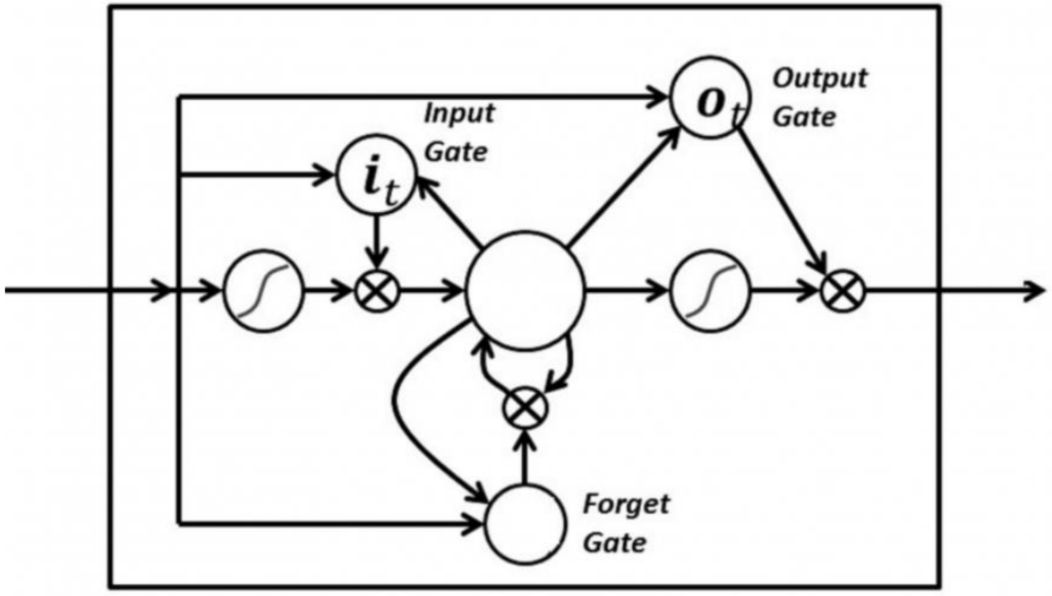


Recurrent Neural Network

- Apple's Siri and Google's voice search both employ recurrent neural networks (RNNs), which are the state-of-the-art method for sequential data. It is the first algorithm with an internal memory that remembers its input, making it ideal for machine learning issues involving sequential data.

- For sequential data such as time series, voice, text, financial data, audio, video, weather, and more, the algorithm performs admirably. In comparison to other algorithms, RNNs are capable of forming a considerably deeper grasp of a sequence and its context. The information in an RNN travels through a cycle. When making a decision, it takes into account the current input as well as what it has learnt from prior inputs.

LSTM (Long short-term memory)



- A recurrent neural network is a type of long short term memory. The output of the previous step is used as input in the current step in RNN. Hochreiter & Schmidhuber created the LSTM. It addressed the issue of RNN long-term dependency, in which the RNN is unable to predict words stored in long-term memory but can make more accurate predictions based on current data. RNN does not provide an efficient performance as the gap length rises.

- There are three gates in an LSTM: input, forget, and output. These gates decide whether fresh input should be allowed (input gate), whether it should be deleted because it isn't significant (forget gate), or whether it should have an influence on the output at the current timestep (impact gate) (output gate).

Sources

- https://www.analyticsvidhya.com/blog/2021/05/bitcoin-price-prediction-using-recurrent-neural-networks-and-lstm/
- https://towardsdatascience.com/forecasting-stock-prices-using-xgboost-a-detailed-walk-through-7817c1ff536a
- https://analyticsindiamag.com/hands-on-guide-to-lstm-recurrent-neural-network-for-stock-market-prediction/
- https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/

Contributions

- Chaitanya Langde (200040038) :- RNN and LSTM based prediction model and Data extraction script.
- Alakh Agarwal (200040018) :- XGBoost based strategy model and Data extraction script.
- Somya Sharma (200040140) :- RNN and LSTM based prediction model.
- Devang Tiwari (200040041) :- XGBoost based strategy model.
- Devanshu Saraf (200040042) :- RNN and LSTM based prediction model.