Forecasting and Trading Crypto with Deep Neural Networks

ACTUPS5842 Advanced Data Science in Finance and Insurance

Course Project

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Submitted By

Emily Xiao

Yuyang Zhao

May 1, 2023

Supervised By

Yubo Wang

Table of Contents

Introduction **3**

Cryptocurrencies **3**

Data Exploration **4**
- Pricing/Volume Data 4
- Economic/Technical Indicators 4
- Alternative Data 4

Natural Language Process **5**

Predictive Modeling **6**
- Logistic Model (Baseline Model) 6
- Multilayer Perceptron Neural Network 6
- Long Short-Term Memory Neural Network 7
- Modeling Performance 8

Backtest Strategy **9**
- General Strategy Ideology 9
- Max 10 Trading Strategy 9
- Momentum Trading Strategy 10
- Strategy Performance 10

Future Work **12**

Introduction

In the age of big data, there are numerous different data sources to consider that may influence changes in cryptocurrency prices in addition to the trade data provided by exchanges. Both social media and financial news may both have an impact on the cryptocurrency market. For example, the price of Bitcoin raised when Tesla announced that it would accept Bitcoin payments. In addition, newly released regulations and direction of public opinion would greatly impact its prices, when the Federal Reserve claimed to increase the interest rates for consecutive 75 basis points during this year, the cryptocurrency market was seen as an unstable and risky market that investors are withdrawing their money for more stable investments in such turmoiled era, so the price of the overall cryptocurrencies is dropping drastically this year. Deep learning models becomes the best match due to its ability to absorb the multi-modal data, vast capacity, flexibility, and the abundance of support from a fast-growing research community.

The objective of this project is to create a deep neural network model that uses trading data, news, social media feeds, and indicators to forecast cryptocurrency returns. And build a backtester on historical data to tune and increase profitability on a long-short equity strategy. This paper is going to cover basic information about current cryptocurrency trends, methodology of the research, data collection process, language processing models, deep learning prediction results, and trading strategy along with its profitability.

Cryptocurrencies

There are mainly four reasons that made Bitcoin volatile: supply and demand, investor sentiments, government regulation, and public attention (media propagate).

Take Bitcoin for example, Bitcoin's market price is primarily influenced by supply and demand. There is a limited number of Bitcoins can be mined (21 million tokens), and number of coins in the circulation is limited. As there are more people interested in Bitcoin mining and Bitcoin investing, the demand is rising as supply drop. Thus, as the supply is reaching to its limit, the prices are very likely to climb. However, if this is always the case, the market value of Bitcoins should continuously increase instead of fluctuating. This is because it's strenuous to predict the price of Bitcoin when all tokens are mined, and when the limit is reached. It's possible that the price goes high due to the short supply and people are urged to trade Bitcoin to preserve its value. It's also possible that Bitcoin loss its value as there's no profit from mining.

Moreover, as the financial player uses financial instruments or other operation techniques to compete for ownership of Bitcoin, the price fluctuates as response to their actions. The investors behavior and sentiments are also very influential to the market price of Bitcoins. As Bitcoins are more popular and less Bitcoins are to be mined, more actions are likely to take by investors and thus more volatile the Bitcoin price.

Publications, predictions, views from public and private sectors are another cause of Bitcoin's volatile. Since Bitcoin is growing its popularity among the public, more people are growing interest in Bitcoin and are hoping to reward from an investment. People often pay close attention to the unverified information present by the news such as rumors about regulations on Bitcoin, which will impact Bitcoin's price in the short run. Regarding government regulations, the Internal Revenue Service (IRS) considers Bitcoin a capital asset. Meaning if one mined a Bitcoin or used it as an investment instrument, it has to be reported as an income and taxes must be paid (tax paid will be based on the market value of the coin on the day one collects it). Although this doesn't directly cause any price changes of Bitcoin, it's possible to have an influence if the regulation causes investor concerns.

This leads to the reason of web scraping and text mining. We would like to bring in alternative data such as news feeds and social media posts as predictors, to enhance the prediction accuracy of cryptocurrency prices. However, the effectiveness is still to be determined.

Data Exploration

Pricing/Volume Data

As the most popular and widely used data aggregator, YahooFinance has crypto open, high, low, close, and volume data from as early as October 30, 2014. This means that more technical indicators can be calculated to support model learning.

Economic/Technical Indicators

Technical indicators are mathematical calculations from historical price, volume, and OHLC (open, high, low, close) data. They serve the need to forecast financial market direction as they integrate data from more dimensions. Some of the indicators added to the data are Simple Moving Average, Exponentially-Weighted Moving Average, Bollinger Bands, Return Strength Index, and Force Index.

Simple Moving Average (SMA) is the average stock price over a specified period. SMA represents the long-term trend of the stock and ignores the intra-day and daily price movements. This reduces the noise that deviates the trader from real trend, and allows traders to see the trajectory over a large time horizon.

Bollinger Bands is a momentum indicator that dictates whether the stock is being oversold or overbought. It is a signal of overbought when the price touches the upper bound multiple times in a given period, and a signal of oversold when the price touches the lower bound multiple times in a given period. A signal of overbought marks the line to sale as the stock is trading above is intrinsic value, and vice versa.

The relative strength index (RSI) also signals bullish/bearish, overbought/oversold price trends. The Force Index measures the amount of power used to move the price of an asset. It is calculated as the price change in a time period multiply by the volume over that time period. A larger Force Index means strong price moves, and serves as a great predictor of the future prices.

Economic Indicators are in broader scale that could indicate potential future investment possibilities.

Inflation rates are measured through the Consumer Price Index (CPI). It is the price of a weighted average of a basket of consumer goods and services that are considered necessities. And the change in CPI is a measure of inflation rate. Given the cases that inflation reaches four decades high, investors are switching to more stable and less risky investments to hedge the uncertainty in the financial market right now, which contributes to investors retreat from the cryptocurrency market and leading Bitcoin prices slumped in 2022. We interpolate the monthly inflation rate data to fit into our data framework.

U.S. treasury yields signals expectations on future economic activity. The macroeconomic index could serve to represent the proxy of the mortgage rate, and it is also seen as a sign of investor sentiment about the economy. A rising yield indicates the falling demand for the treasury bonds, which also indicates that investors are willing to enlarge portfolios in higher-risk, higher-reward investments. Treasury yields with different maturities (5, 10, and 30 years) are added to the dataset.

There are also other technical and economic indicators we won’t explain in detail.

Alternative Data

Twitter is a popular microblogging social media platforms. The posts and comments on the platforms could provide the real time discussion or opinions of the users who are interested in cryptocurrency or are real life investors. The spread of the subjective opinions will more likely cause the changes of prices of the discussed cryptocurrency compared to stocks and bonds in traditional finance. As a result, an important part of features contains numerical representation of sentiments from tweets since cryptocurrency was available for trading. For the exact types of cryptocurrencies in our pool, Bitcoin and Ethereum are chosen given the two cryptocurrencies dominate the crypto market and also the most discussed cryptocurrencies. Therefore, simply search for "Bitcoin" and "Ethereum" when web scraping is enough. There are minimal posts and news related to other cryptos, however. The uneven size of data would harm the prediction accuracy for the cryptos with smaller market capitalization. A solution is to use "crypto" as keyword when extracting news and feeds and let it represent the general trend in crypto market. Then all other cryptos can use this data as its alternative data, to make up for their lessen discussions on the internet.

To enhance the quality of scrapped tweets, we employed a methodology of a 3-layered filtering so that the cleanest and most valuable tweet discussions could be kept while deceptive advertisement as well as meaningless words were excluded.

```
spam_words = ['chance', 'Follow', 'follow', 'free', 'link', 'LINK', 'bio', \
              'BIO', 'Update', 'last price', 'Current {} Price'.format(topic),\
              'ONLY', 'BLOOMBERG', "report", "Bollinger", 'chart', 'unlock',\
              'Poll', 'live'] #filter spam tweets by banning keywords

spam_emojis = ' -👋 -🟩 -🟢 -🟥 -🔍 -📊 -⬅️ -➡️ -🤖 -🔴 -✅ -📈 -💹 -🤖 -🔻 -👆 -📉 -⬇️ -👇 -❓ '
```

Natural Language Process

Besides basic pricing, volume data along with technical and economic indicators, we draw great focus on text data analysis through the public sentiment on Twitter, as well as media propagation. In order to transform unreadable text data into the readable format by the model, we adopt a simple transformer model to extract the sentiment within the text data.

Our transformer model was trained on a dataset of about 1.6 million tweets, covering a range of topics and sentiments. The training dataset is from Kaggle, called sentiment140. When given a piece of text, the model returns one of three sentiment statuses: positive, neutral, or negative, along with their corresponding predicted probabilities. For example, if we input the text "Covid cases are increasing fast!", the model output a negative sentiment with a higher probability on negative. This information can be useful in a range of applications, from sentiment analysis of customer feedback to tracking public opinion on various topics.

Since training a full language processing model takes too much time, we used TextVectorization to change Tweets to vectors. It is a build-in Karas layer that does preprocessing steps such as tokenizing, encoding, and padding to the input text. We have it as the first layer of our model. It is run separated from all the other layers. Then we have the build_transformer method. It is built with TensorFlow, which will be specifically talked about later.

After training the model, we input bitcoin tweets to our model and output subjectivity and polarity, which will be used as a part of our variables for model prediction. This function used Tensorflow to build a transformer model in Keras.

The function begins by creating an Input layer with a shape of (sequence_length,). Then, followed by an Embedding layer with vocab_size and embedding_dim, which is used to transform the input tokens into continuous vectors and fed into the transformer blocks. Next, the model enters a loop where it adds transformer_blocks instances of the custom TransformerBlock layer. We will talk about how we customize this layer in next slide. Throughout the loop, each block applies multi-headed attention and feed-forward neural networks to the input data, allowing the model to learn complex relationships between the input tokens. After processing the input through the transformer blocks, a GlobalAveragePooling1D layer is applied to reduce the sequence dimension, which extracts a single vector representing the entire input sequence. Finally, a Dense layer with num_classes output units and a softmax activation function is added to produce the final output probabilities for each class. Once all the layers have been defined, the function returns a Keras model that is built using the defined input and output layers. The TransformerBlock class inherits from tf.keras.layers.Layer to create a custom layer.

We wrote 4 layers in constructor method. The first one is the most important layer of performing our transformer model. It computes multi-headed attention over the input sequence. This allows the network to focus on different parts of the input at the same time, improving the ability to process complex data. The Add layer computes the element-wise sum of two input tensors. It is used to combine the results of the multi-headed attention layer with the original input tensor. To ensure that the network performs well, the Normalization layer is applied to the sum output. This layer applies layer normalization to the input tensor, which helps to improve training stability and accuracy.

Lastly, the Sequential container is used to apply two Dense layers sequentially to produce the final output tensor. These layers are used to compute the final output of the TransformerBlock, which can then be passed on to the next layer in the network. In the call method of TransformerBlock, we define how to apply these layers and operations to the input tensor.

Predictive Modeling

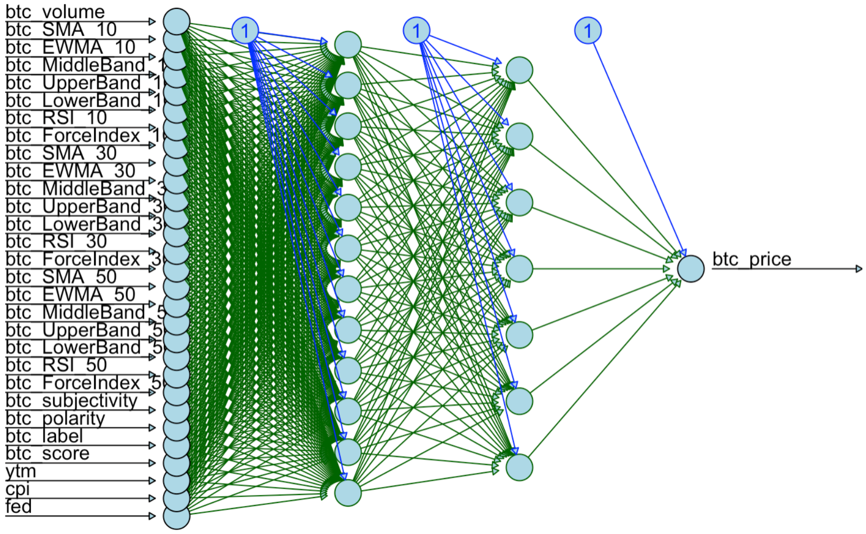
Logistic Model (Baseline Model)

We decided to make this into a classification problem where the prediction will be positive or negative price changes. This is because we are not doing any market making, so we don't care about the price at the next ticker. As long as we can get an idea of the price trend, we should be able to proceed trades accordingly. Due to this classification nature, we chose logistic regression to benchmark other model performances.

The resulting probabilities can be categorize using certain thresholds (typically 0.5 threshold), to get ultimate return labels. The model results in an accuracy of 52%, slightly higher than random guess. It's not a sophisticated model and might left out many important details, so we expanded to neural network models, which are more mimicking human decision making process and thus are more suitable for complicated dataset like this.

Multilayer Perceptron Neural Network

Artificial neural network based on MLP are feed forward nets with one or more layer of nodes between its input and output layers, and due to the nonlinearity activation function with each nodes the MLP is capable of forming arbitrarily complex decision function in the pattern space (Devi Rani Guha, 2010). Given multilayer perceptron has many layers with many neurons stacked together, this data structure could pick out features at different scales or resolutions and combine them into higher-order features, for example, from lines to collections of lines to shapes. Also, MLP model could fit into the scenario of outputting binary predictions.



The above figure the general idea of MLP neural network, not the true model this project created. To obtain the best set of parameters for prediction, a grid search of parameters is performed.

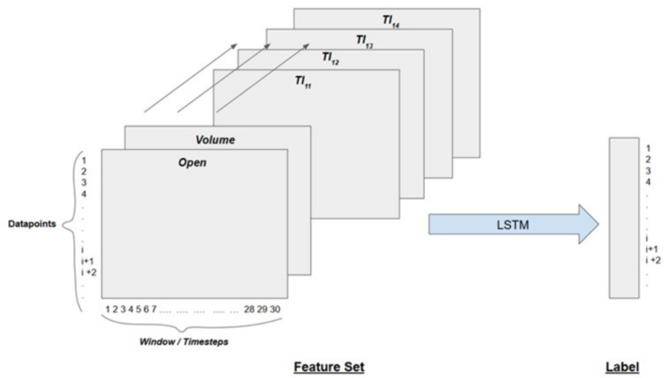
In this MLP, there are 3 hidden layers, each size of 50, 100, and 50. Weights are initialized randomly and assigned to each input variable. The weighted inputs are summed and passed through the ReLu activation function. The ReLu activation function returns the max between 0 and input. The major benefit is that ReLu doesn't activate all the neurons at the same time. If the weighted linear transformation given initial weights is less than zero, the neuron is not activations. This largely increases computation and learning efficiency compared to the other activation functions. Then the training algorithm uses stochastic gradient descent to proceed with forward and backward propagation. A given row of data will go through the network with initial assigned weights and produce an output. The output of the network is then compared to the true value. Then the error is back-propagated through the network while updating the weights based on the amount they contributed to the error. This process is repeated for all training data, and one round of updating the network is an epoch. In addition, the alpha serves as a penalty term and avoids overfitting.

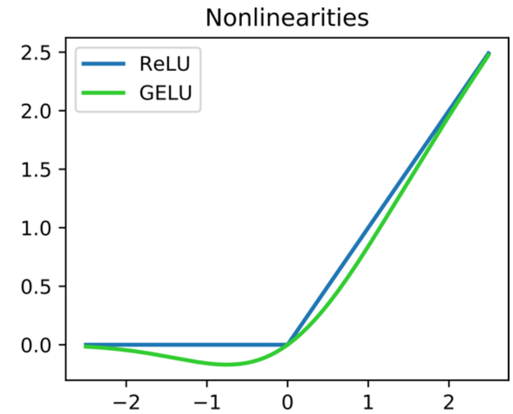
Once the model is fully trained by completing the predefined number of epochs, predictions can be made by feeding inputs to the network and generate outputs. The major disadvantage is that MLP is fully connected, thus includes too many parameters. This amount of connection turns into a dense web, and resulting in redundancy and inefficiency

Long Short-Term Memory Neural Network

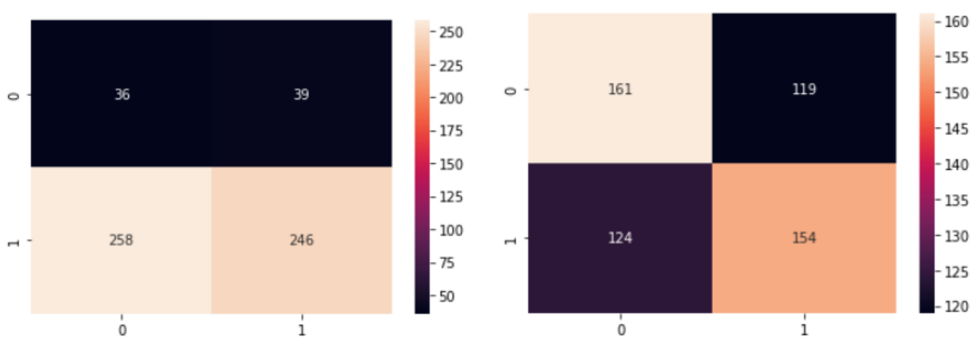
As a special case of recurrent neural networks (RNN), LSTM resolved the vanishing gradient problem of RNN, which is the shortcoming of traditional RNN that it fails to learn in the presence of time lags greater than 5-10 discrete time steps between relevant input events and target signals. LSTM model can learn to bridge minimal time lags in excess of 1000 discrete time steps by enforcing constant error flow through "constant error carrousels", which makes it more well-suited to classify, process and predict time series given time lags of unknown duration. The selective function of LSTM model makes it more powerful than conventional recurrent neural networks and it is applied to solve different areas of problems, from language modeling, machine translation to image captioning and generation.

The LSTM model will learn a function that maps a sequence of past observations as input to an output observation. As shown in the figure below, the graphical representation of the 3-dimensional dataframe, each variable is being expanded to 2-dimension by adding a lookback period. There are 30 features in total and we chose a lookback period of 21-days, this yields a 30x21 matrix for each sample. The next day return will be estimated based on this two-dimensional matrix.





For the activation function, GeLu is being adopted as a smoother version of ReLu. As discussed above, ReLu = max(0, input), can hurt the results when a significant number of neurons are set to 0. Too many "dead" neurons make the weights not being calibrated enough. GeLu resolves this problem by weighting inputs by their percentiles. It is differentiable in all ranges and makes sure that all neurons contribute to weight calibration during forward and backward propagation. The drawback of ReLu can be seen in the results of MLP network. There is a highly uneven split between 1 and 0 predicted results where more buy signals were predicted, as shown in the figure below (left). This is majorly caused by the incomplete calibration of network weights and unused neurons, thus largely damaged PnL when we proceed with trading. After using GeLu activation function, there is an even split between buy and sell predictions (as shown in the figure below, right), and made the trades more profitable.



Modeling Performance

Before embarking on strategy formulation and backtesting, we choose the classification results to be the standard to decide which model could produce better trend prediction in the given test period, the more predicted class status corresponds to the trend status in real life, the better the model functions well to predict.

For baseline model, taking basic price and volume information, economic and technical indicators and also sentiment output from the social feeds, we could achieve a total 52% accuracy for predicting the correct up and down price trend, but the recall of predicting the upward trend only just hit 50%, which also means that the model has the same amount of downward trend predictions that are wrongly classified as the amount of correctly classified upward trend predictions.

For Multilayer perceptron model, by taking into all the features and conduct the scale pipeline to the dataset, we achieve a total 49% accuracy giving the best tuned model predicting the correct price trend of Bitcoin. By looking into the detailed recall and precision of the two predicted directions, the recall of the predicted downward status only gets 0.12, which means that there are over 80% of wrongly classified downward predictions over the whole predicted downward predictions and this could cause a great loss when the cryptocurrency's price is going up while we classify it as going down.

For Long Short-Term Memory model, we achieve a total accuracy of 56%. Leveraging the Gelu activation function, the final hypertuned model actually achieved a comparably balanced and better prediction accuracy in both trends and outperform two other models. The precision of upward and downward trend predictions both are 0.56, and the recall of the downward price trend prediction reaches 0.57, which means that there are less parts of upward trend predictions compared with the correctly predicted downward trend, and could be an edge of formulating the later strategy, given there are more times in the 2022 that Bitcoin prices are slumping given the continuous interest rate raises and investors' uncertainty about this volatile asset.

<u>Backtest Strategy</u>

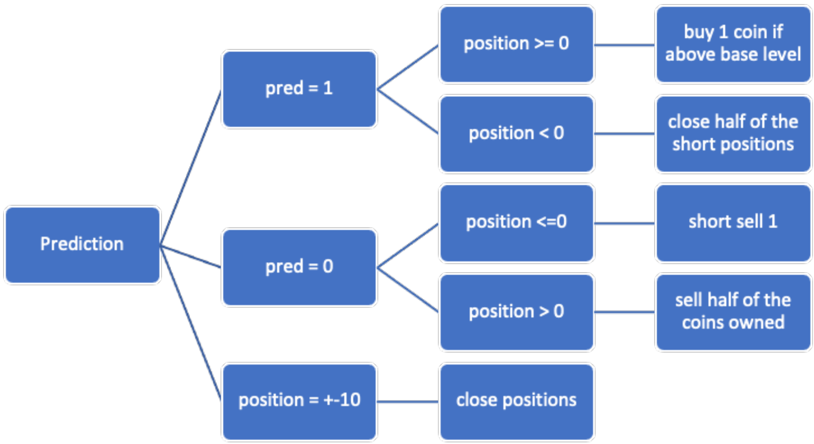
General Strategy Ideology

In order to formulate a trading strategy based on our cryptocurrency portfolio, we leverage the predicted upward and downward status of a given cryptocurrency to decide the direction of buying or selling. To make it more sophisticated under the real market situation, we divide investor's accounts into the following: cash account, coin account and also the margin account, besides basically buying and selling, we will maintain our margin account to be satisfied with the given rules. The cash account is used to keep track of the amount of cash remaining. It is dangerous that the trader runs out of cash to maintain short positions and in our implementation, we set up a cash base level so that no trading activity will be taken place with cash is below that amount. The coin account is used to keep track of the value of the cryptocurrencies owned. The margin account is for when short positions are taken and since the money in the account will fluctuate as the cryptocurrency value does, it is used to constantly check whether money will be needed to add when margin call happens. In addition, whenever we trade the exact coins in a daily frequency, we need to pay the commission fees to the exchanges; whenever we need to short the cryptocurrency under the case that we do not have any, we need to offer the collateral and pay the interests in the given borrowing period.

In terms of numerical values in this project, we set the commission fee to be 0.2% of the dollar amount traded. In terms of collateral, it is common that for the trader who wants to open a short position, 50% of the amount traded will be required. For example, if one intends to short one bitcoin, the collateral will be the dollar amount equal to half a bitcoin. Meanwhile, initial margin percentage, maintenance margin percentage to be respectively 50% and 30% of the amount that is short. Since the counterparty charges interests for traders with short positions, the interest rate per day is set to be 0.02%. In other words, the interest due for the trader will accumulate at the value of 0.02% of the dollar amount of the short position. The cash base level will be 5000 dollars, meaning that once the available cash falls below that level, no trading action will be taken as a means to prevent from forced clearance due to liquidity issues.

Max 10 Trading Strategy

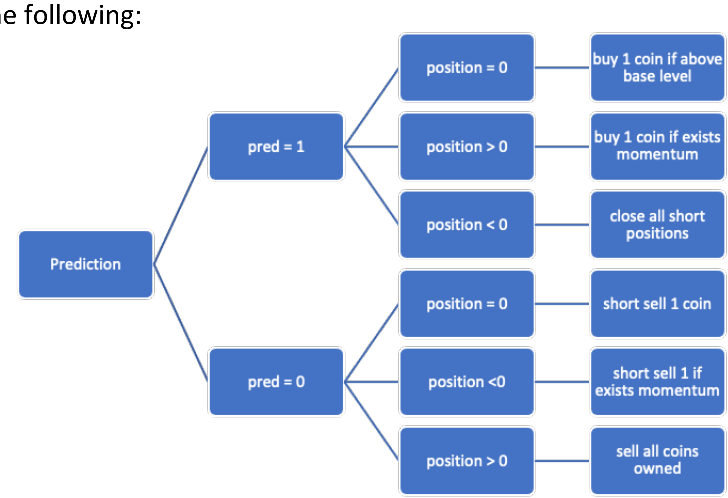
The trading logic is the following:



As its name implies, if the position reaches 10 or -10 – the case of either owning 10 coins or short selling 10 coins – we think that the position is risky enough that it is time to close the position and calculate the profit and loss.

Momentum Trading Strategy

The trading logic is the following:



Momentum stands for the idea that if the underlying value has increased over the last couple days, it is likely that the trend will continue. In our case, if the value of the coin has solely increased for the past 4 days, added with the predicted output of 1, we are confident enough to be more aggressive and buy one more coin if cash is above the cash base level. Otherwise, since the model predicts the coin value to fall, it would be a good time to sell the coins and hopefully walk away with the gain. According to our test data, it is rarely the case that the coin has been solely increasing or decreasing for 4 days in a row. Hence, A position size with absolute value greater than 1 is often not seen. Therefore, the strategy value will have much smaller fluctuation.

Strategy Performance

Given the trained model by our final Long Short-Term Memory model, we leverage the predicted upward and downward status of the cryptocurrency from March 2021 to October 2022 to build trading strategies to compare the returns from our two trading strategies and the simple buy and hold strategy.

Initial cash $300,000	Max 10	Momentum	Buy and Hold
Total Return	112.60%	19.95%	-64.72%
PnL	337795.88	59848.21	-194158.93
Sharpe Ratio	1.91	1.06	-0.63
Maximum Drawdown	-11.51%	-5.58%	-72.55%
Calmar Ratio	5.54	2.26	-0.68
Sortino Ratio	2.60	1.63	-0.89

For simple buy and hold strategy, we set the initial cash account as 300,000 dollars, and it ends up losing 64% of its original value due to the crash of cryptocurrency market in the year of 2022. Sharpe ratio, Sortino ratio, Calmar ratio all end up being negative and the maximum drawdown reaches horrifying 72%, meaning about maximum 216,000 dollars could be lost at a cost for the worst trading action.

For max 10 trading strategy, by holding 300,000 dollars in the beginning, we accumulate about 638,000 equivalent dollars amount at the end of the backtesting period, achieving 113% excessive return given the initial cash value. Also, it obtains a Sharpe ratio of 1.9 and a Sortino ratio of 2.6, which indicates the negative return trading days' volatility is smaller than the positive return trading days' volatility, and the numeric value of Sharpe ratio near 2 could be generally considered as very good investment opportunity; Besides, max 10 trading strategy achieves a maximum drawdown of 11%, which means the greatest loss during the time could only be around 30,000 dollars; the last metric Calmar ratio of 5.5, which is generally considered as a very good investment given ranges of 3.0 to 5.0.

For momentum trading strategy, by holding 300,000 dollars in the beginning, we accumulate nearly 360,000 equivalent dollars amount at the end of the backtesting period, achieving about 20% excessive return given the initial cash value. Also, it obtains a Sharpe ratio of 1.0 and a Sortino ratio of 1.6, which also indicates the negative return trading days' volatility is smaller than the positive return trading days' volatility, and the numeric value of Sharpe ratio above 1 could be considered as a good investment opportunity; Besides, momentum trading strategy achieves a maximum drawdown of 5%, which means the greatest loss during the time could only be around 15,000 dollars and the greatest loss could be minimized through this trading strategy; the last metric Calmar ratio of 2.2, which means it could achieve very steady profit by controlling the risk to its minimum level.

Comparing with the benchmark buy and hold strategy, both of our formulated trading strategies outperform the buy and hold strategy by obtaining nearly 6 times and 3.5 times of the total value compared with benchmark in the end of the backtesting period. Also overall, the more aggressive max 10 trading strategy outperform momentum strategy by acquiring the higher Sharpe ratio, Sortino ratio, and also Calmar ratio, even though max 10 strategy's maximum drawdown is bigger than momentum strategy, but it could better seize the opportunity of the short selling in year 2022 and profit quite a lot from it.

Future Work

Although this project manages to solve a variety of challenges faced when predicting cryptocurrency prices and creating trading strategies, there are several problems that requires more research. Regarding data preparation, the quality of alternative data can be improved by transformer. Crypto-related posts and feeds are mainly spams and advertisements, thus have no analyzing significance, nor contributing to cryptocurrency price predictions.

Regarding model creation and parameters tuning, due to the high volatile feature of cryptos, the models are sensitive to parameter changes. As mentioned in the model creation section, the neural network model initiates a matrix of weights for the input variables and calibrate weights during backpropagation to achieve best fitted model. However, there is no guarantee that the derived network is global optimal because of the randomness component in weight initialization. As tested during model creation, seeds selection has a large impact on the resulting accuracy where a bad initial set of weights result in a less predictive model, and thus harder to be profitable from. Thus, more improvements are to be made on the current model to mitigate such uncertainly and dependency on seeds. In addition, a more stable model is to be discovered.