Predicting Bitcoin with Technical and Sentiment Analysis

Yoong Pei Xian 100334770

Abstract

There is no doubt that the crypto market is growing by leaps and bounds. The most popular cryptocurrency in the world, Bitcoin, has grown by more than 70% over the last year. The world's largest investment funds are actively investing in various crypto assets and looking for innovative approaches to maximize return on investment. Disruptive technologies such as machine learning transform the way of investing and put investors in a better position of their financial stand by fulfilling people's needs at a lower cost and minimum time with algorithm trading.

Initially, historical trading data such as closing prices and volumes are utilized predominantly for future price prediction due to the ease of implementation. Nonetheless, it is observed that the shift of public sentiment as a result of the changing political and economic policies also plays an important role in determining the price movement, thus raising the interest of many researchers in predicting the price using sentiment analysis in the recent years. Existing studies on the crypto price prediction focus mainly on standalone models using solely either technical or sentiment indicators while the papers that develop integrated models constructed using the integration of both technical and sentiment indicators are relatively limited.

Building on existing work on crypto price prediction, this research aims to i) compare the performance of standalone and integrated models in crypto price prediction, ii) examine the feasibility of the models in the real-life setting by implementing crypto trading on an automated trading bot, iii) draw a comparison on the performance of classification and regression models, iv) verify whether machine learning models can predict crypto price movement.

Python is to be used as the programming language to develop the machine learning models using its inbuilt libraries and supported modules. The steps involved in the methodologies include data collection, data pre-processing, model development, model evaluation, and trading on the automated bot developed using Python to examine the model performance in real-life trading settings. The data required for the research includes the technical indicators and sentiment indicators of different timeframes and intervals to be extracted from the open-source websites and processed using Python. Classification and regression models to be developed using Decision Tree, Random Forest, and Support Vector Machines and then evaluated using different metrics.

The data required for the research to be obtained from open-source platforms thus the research demonstrates no ethical concern. Several risks involved are assessed and mitigated.

1 General Description of the Proposal

The cryptocurrency market has achieved exponential growth in the recent years, with the market capitalization from around \$1 trillion to \$1.5 trillion over the past year (Leonard 2020), becoming one of the biggest unregulated markets in the world, where the number of global crypto users grew to over 200 million by June 2021 (Crypto 2021). Major cryptocurrencies have achieved massive growth in 2021. Bitcoin has surpassed \$1 trillion in market value in February 2021 (Locke 2021) and big companies including Tesla, Square, and MicroStrategy started to use their balance sheets to buy bitcoin. Ethereum's price is also up around 466%, about 6 times more than Bitcoin (Bradley 2021). Consequently, investors have started to include cryptocurrencies in their portfolios, perceiving them to have a prospective outlook on becoming the dominating financial instruments in the coming years.

Trading cryptocurrencies has always been considered a high-risk financial activity due to its high volatility and its state of still being as a nascent asset class in the price discovery phase (Sigalos 2022). Besides, changes in sentiment can send the prices to skyrocket or fall on a whim as unlike gold, cryptocurrencies are not pegged to tangible values. Therefore, investors have been looking for ways to conduct price predictions to form a judgment on whether to buy to desist from a specific cryptocurrency. Technology has changed the stock market such that today most of the transactions are performed by algorithms instead of being executed by humans. Algorithms constructed to perform price prediction are based on the indicators in either of these two aspects: (i) technical indicators: daily closing price, daily volume traded, moving averages, (ii) sentimental indicators: sentimental scores computed by the posts extracted on social media platforms.

Previously, technical indicators that are merely based on historical price values are used predominantly in the majority of the studies concerning crypto price prediction due to the ease of implementation. However, any external factors, for instance, changes in regulatory policies such as interest rate hikes, and shifts in market sentiments measured by the fear and greed index to examine the level of optimism about certain crypto adoption, are entirely not taken into consideration. The importance of these external factors is becoming increasingly powerful as the crypto markets grow bigger, validated by André Kostolany, one of the most successful investors of the 20th century, quoting that "facts only account for 10% of the reactions on the stock market; everything else is psychology." (Palmer 2021).

Given the recent massive volatility in the cryptocurrency market induced by President Joe Biden's executive order on digital assets, US inflation numbers coming up, and the ongoing war between Russia and Ukraine (Cryptopotato 2021), this further proves that cryptocurrencies are not only sensitive to historical price movement but also vulnerable to the expectations about future price movements and the shifts in public sentiments (Giaglis et al. 2015), providing more incentives for researchers to dive deeper into studying the impact of sentiments on the price movement on top of the technical analysis. Despite the huge amount of highly relevant studies to predict crypto price movement, there is only a limited number of papers focusing on the integration of technical and sentimental analysis to build models. Besides, most of these papers only measure the effectiveness of the model using metrics but do not build an actually automated trading bot based on the model developed to examine how the algorithms work in the real-life trading setting, prompting doubt on the models' abilities to generate promising returns on actual investment.

2 Research Aims and Objectives

This research aims to predict Bitcoin using different classification and regression models. For this purpose, a quantitative methodological approach to be utilized by collecting technical and sentiment data on the open-source platforms online. The process of data handling is further explained in the methodology section. The crypto selected to be Bitcoin as it is the most established digital asset in the crypto market with the highest market capitalization.

The scope of this research encompasses using two classes of features, considering a combination of technical and sentimental indicators as the input of classification and regression algorithms. Automated trading bot to be implemented using the models developed to examine the returns of investments that can potentially be generated in real-life trading. The performance of standalone models (composed of technical indicators or sentiment indicators) and integrated models (composed of technical and sentiment indicators) to be evaluated and compared. From the evaluation, it is possible to identify whether machine learning models can predict crypto price movement, or if the price is following a random walk process that does not show any patterns based on the result of validation metrics. The performance comparison of classification and regression models can also be examined. The concise aims and objectives of this research are listed below:

After examining the topics that are to be investigated in the study, the following research questions are proposed:

- To compare the performance of standalone and integrated models in crypto price prediction
- To examine the feasibility of the models on the real-life setting by implementing crypto trading on an automated trading bot
- To draw comparison on the performance of classification and regression models
- To verify whether machine learning models can predict crypto price movement

3 Literature Review

The literature review conducted by Jaquart et al. (2020) analyzing the existing research papers on cryptocurrencies revealed that classification and regression models are adopted equally in building the price prediction models, but there was a stronger inclination towards the regression models from 2018 onwards. Complex deep-learning algorithms such as recurrent neural networks (RNN) and long-short term memory (LSTM) neural networks are claimed to yield a more promising result (Jaquart et al. 2020, Alahmari 2020*a*), opposed to Chen, Li & Sun (2020)'s finding demonstrating that rather simple methods (e.g, logistic regressions) can outperform the complex algorithms when the high-dimensional feature sets are selected to compensate for the simplicity of the algorithms.

While there is a huge quantity of studies about cryptocurrencies, the most important research on the price predictability is reviewed based on the two distinct categories of indicators: 1) technical indicators often encompass historical market data (e.g. previous closing price, daily/weekly returns in percentage, the volume traded), 2) sentiment-based indicators yield information linked to the investors' emotions and opinions on internet (e.g. posts on Twitter, Reddit, Google Search). Most of the existing studies focus on using technical indicators due to their easy implementation (Huang et al. 2018). For instance, there have been extensive studies on the impact of moving average (MA) strategies in financial markets (Brock et al. 1992), and the comparison of MA strategies with the traditional buy-and-hold strategy (Shynkevich 2012) but the limitations of these studies are that most of them focused on the equity market such as the U.S. stocks instead of the cryptocurrencies. The fundamental differences between these two markets prompt doubt on whether the trading models that are proven to be effective can be applied the cryptocurrencies.

3.1 Technical Indicators

Huang et al. (2018) is among the first to examine the price predictability of cryptocurrencies. Instead of merely using the historical prices for prediction, they computed other high-dimensional trading indicators such as Moving Average (MA), Relative Strength Index (RSI), and Parabolic SAR (stop and return), claiming their strong predictive power. They utilized 124 complex technical indicators to predict the range of the next day's return using a simple decision tree algorithm and they yielded a satisfactory result showing the model's capabilities to even predict the narrow ranges of bitcoin daily returns. On contrary, Pintelas et al. (2020) questioned the use of traditional time series methods adopted in Huang et al. (2018)'s studies, squaring with the findings that these models cannot capture non-linear patterns of complicated problems. Instead, they adopted more complex deep learning algorithms including LSTM using only the historical prices as the indicators aiming to capture the non-linear patterns. Nevertheless, their findings concluded that even deep learning algorithms are inefficient and unreliable with the rationale that crypto prices follow almost a random walk process, supported by Stavroyiannis et al. (2019). There could be many reasons leading to the under-performing models to accurately predict the price movement, however, it is worth noting the use of price as the only indicator as it might result in weaker models as Resta et al. (2020) suggested that the integration of other crucial indicators such as trading volume can improve the model performance and yield more profits, confirmed by the studies of Balcilar et al. (2017). This makes their claims of the model's incapability to capture the trends questionable as they could have fed the algorithms with other technical indicators before jumping to the conclusion. Their claims of the random walk process are also objected to by Chen, Zhang & Lou (2020), demonstrating the hybrid of deep learning model sequence-to-sequence LSTM, convolutional neural network (CNN), and traditional time series model ARIMA can generate promising results by capturing both linear and non-linear patterns effectively.

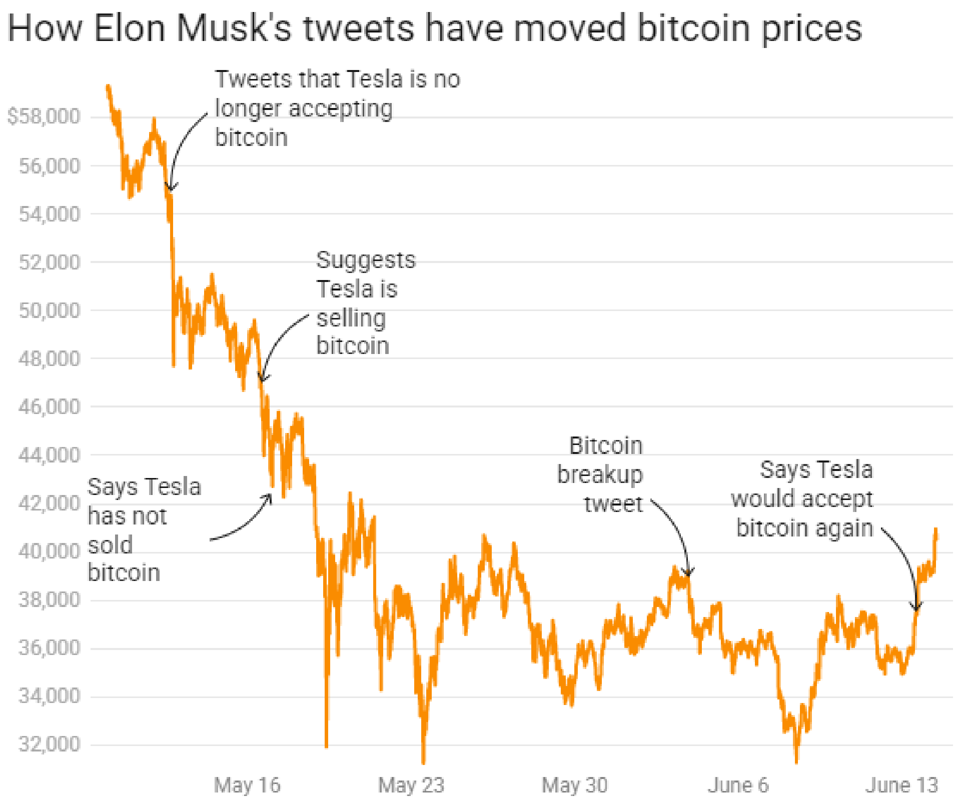
Despite this, Phaladisailoed & Numnonda (2018) criticized that the adoption of only technical indicators is not sufficient to provide an accurate prediction as the crypto price movement can be affected easily by macro-environmental factors such as new legislation and public sentiment. This can be easily justified by the rise of meme cryptocurrencies, which refers to the cryptocurrencies that skyrocket in price in a short period (often hours or days) because of a sudden surge in interest online or on social media. One of the most famous meme cryptocurrencies, Dogecoin, with a market cap of $18 billion, making it the 8th largest cryptocurrency on the planet, soared more than 800% in several hours after the encouraging tweets from Elon Musk (Greenopolis 2021).

3.2 Sentimental Indicators

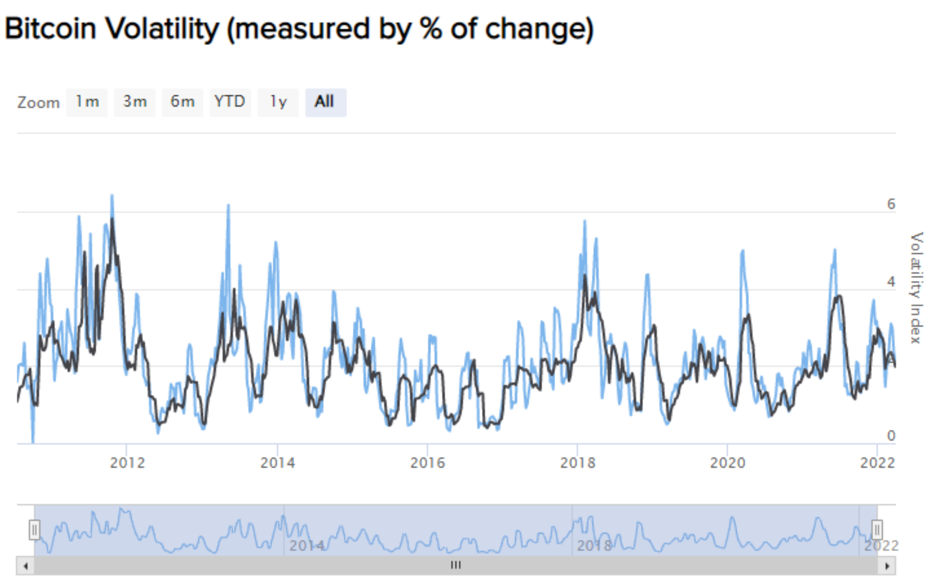
Years before the rising of Twitter and Reddit, researchers have already conducted experiments into the relationship between emotions and stock price movement. In 2004, Antweiler & Frank (2001) studied the impact of over 1.5 million messages posted on Raging Bull and Yahoo! Finance and they found that these messages are highly correlated to the trading volume and volatility. Kaminski & Gloor (2014), Mai et al. (2015)'s findings presented those posts with negative content have a higher impact on the price movement where the price would always end lower when there are many negative posts, supported by Giaglis et al. (2015). Nonetheless, Giaglis et al. (2015)'s research presents the limitations of only conducting the experiments in the short-run using 3-months historical prices of Bitcoin from 2014-to 2015 and since the research was conducted almost 7 years ago, it is unsure whether the models can still yield promising result in the today context, given the fast-changing behavior of the Bitcoin price.

Mai et al. (2015)'s also discovered that despite most of the studies using tweets as the dataset, the impact of forum posts exceeds the tweets as the price movement is mostly driven by the 'silent majority, supported by Stenqvist & Lönnö (2017) of using hottest topics on a Bitcoin-related forum as the basis of their research. Nonetheless, Kaminski & Gloor (2014) found that high trading volumes usually comes before the rise in positive tweets, meaning that the sentiment scores only mirror what is happening in the market, instead of predicting the market.

They challenge the findings of to citeUllah:2022 claiming that social media posts posted by influential characters such as celebrities, often spark immediate temporary exponential Bitcoin price fluctuations. One of the most prominent examples is the 'Musk Effect' of how Elon's tweets moved Bitcoin price drastically every single time, often seen as an unethical way of manipulating the market price, consistent with Choi (2020)'s findings that reliable people can move the market. It is worth highlighting that the underlying issue of these studies is that none of the factors the potentially linked events, for example, political and legislation changes, current phase in the crypto price cycle (Accumulation, markup, distribution, and markdown) (Special 2021), that happen on the same period into account when conducting the research, prompting doubts on the extent of influence as no metrics can measure it. Observing the following graph, it is noticeable that there was a steep fall or rise in price even before Elon made any announcements. For instance, many analysts perceived that Bitcoin was overbought, and a correction would happen in April. It was roughly the same time when Elon claimed that Tesla was no longer accepting Bitcoin, which many believed to be the reason for the sharp dip. However, it is highly likely that Elon's tweet just acted as a catalyst for the much-anticipated correction in the crypto market.



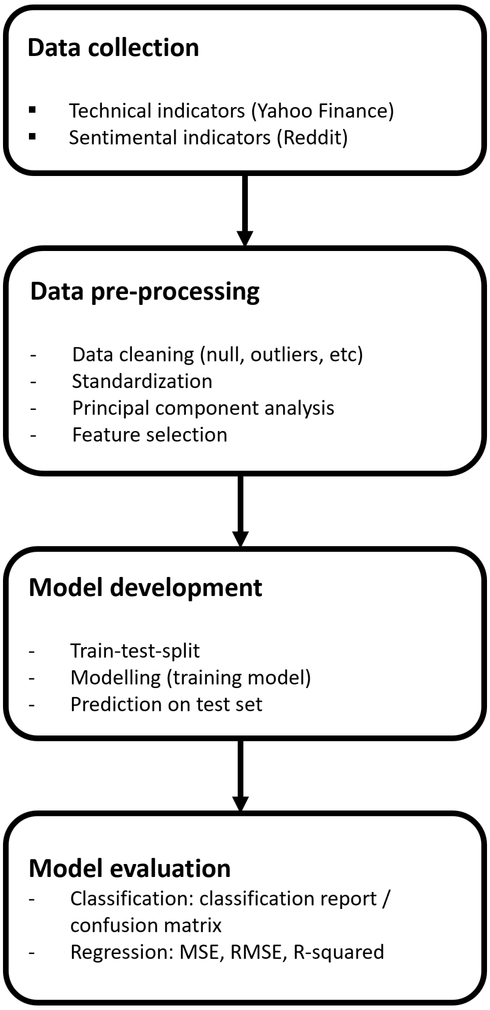
Many reasons can result in the variation of findings in different studies, for instance, the sample size, technical indicators selected, methodologies, algorithms, or even the prediction intervals, supported by Resta et al. (2020)'s research validating that the model performance can vary drastically for intraday and daily trading. They also pointed out the behavior of the Bitcoin price series during the considered period can affect the model performance to a great extent, especially when the price behavior is highly turbulent, and the lack of price consistency can make the research result differ much from the models developed in every paper adopted dataset from different time range. Studies that were conducted in the different years are highly likely to yield different results due to the constantly fluctuating volatility index, as shown in the image below, also other underlying factors that could potentially affect the price movement.



4 Methodology

This research aims to examine the predictability of the crypto price movement using different machine learning algorithms. It is designed to be quantitative research, in which public data is to be collected on various online platforms for analysis. Existing literature relating to crypto price prediction from different research journals, articles, and websites is also to be reviewed to enhance the understanding of different algorithms.

- Data collection from quandl.com
- Data pre-processing (e.g. generate new features, handle missing values etc.)
- Data splitting into training and testing set
- Build machine learning algorithms and tune the hyperparameters
- Train and test the models, and predict the values
- Validate the model performance using different metrics (classification report, confusion matrix, MAE etc.)
- Integrate the models to the automated trading bot to product returns based on a $100,000 of initial investment



4.1 Tools involved

Python to be the main programming language used in this paper. The following Python modules and packages to be involved in this study:

Packages	Purpose	Modules
pandas	To perform data manipulation and analysis	import pandas
numpy	To process array	import numpy
yfinance	To extract financial data	import yfinance
talib	To compute additional technical indicators	import talib
praw	To extract Reddit posts	import praw
ntlk	To perform sentimental analysis	from nltk.sentiment.vader import SentimentIntensityAnalyzer from nltk.tokenize import word_tokenize, RegexpTokenizer from nltk.corpus import stopwords from nltk.stem import SnowballStemmer nltk.download('vader_lexicon') nltk.download('punkt') nltk.download('stopwords') stop_words = stopwords.words("english") stemmer = SnowballStemmer("english")
matplotlib seaborn	To produce visualization	import matplotlib as mpl import matplotlib.pyplot as plt import seaborn as sns
scikit-learn	To pre-process data	from sklearn.preprocessing import StandardScaler, MinMaxScaler from sklearn.decomposition import PCA from sklearn.pipeline import Pipeline
scikit-learn	To build and train classification and regression models	from sklearn.model_selection import train_test_split from sklearn.model_selection import RepeatedStratifiedKFold from sklearn.model_selection import cross_val_score from sklearn.model_selection import GridSearchCV from sklearn.tree import DecisionTreeClassifier from sklearn.ensemble import RandomForestClassifier from sklearn.svm import SVC from sklearn.tree import DecisionTreeRegressor from sklearn.ensemble import RandomForestRegressor from sklearn.svm import SVR
scikit-learn	To evaluate models	from sklearn.metrics import classification_report, confusion_matrix from sklearn.metrics import mean_squared_error from sklearn.metrics import r2_score

4.2 Data collection

The following data to be collected for this study:

Target variables		Daily closing price of Bitcoin
Independent variables	Technical indicators	Open Close Volume Daily Return in Percentage Moving Average (MA) Relative Strength Index (RSI) Average Directional Index (ADX) Moving Average Convergence Divergence (MACD)
	Sentiment indicators	Daily total number of posts on Reddit Daily mean sentiment score

4.2.1 Timeframe

Bitcoin data of different timeframe and interval to be extracted.

- Set 1: 2015-01-01 to 2022-06-01
- Set 2: 2021-01-01 to 2022-06-01

4.2.2 Technical dataset

The data to be extracted from Yahoo Finance with the following features and stored as df.csv Datetime, Open, Close, Volume. Other technical indicators include Daily Return in Percentage, Moving Direction, Moving Average (MA), Relative Strength Index (RSI), Average Directional Index (ADX), Moving Average Convergence Divergence (MACD), Bollinger Bands derived from the daily closing price of Bitcoin are also stored inside the df.csv.

4.2.3 Sentimental dataset

A dataset consists of the sentiment score of different prediction intervals computed using Reddit posts are produced. Reddit posts consisting of keywords ”Bitcoin”, ”BTC” and “Bitcoins” along with their respective hashtags [”Bitcoin”, ”BTC” and “Bitcoins”] to be gathered by parsing the Reddit API on Python. These data are to be pre-processed using Natural Language Processing (NLP) pipelines consisting of the steps below by leveraging nltk and spacy library.

VADER, (Valence Aware Dictionary for Sentiment Reasoning) is to be adopted to perform text sentiment analysis to assign the data with polarity – positive or negative and the intensity of the emotion by using sentiment score between 1 and -1. Scores close to 1 indicate strong positive sentiment, and scores close to -1 indicate strong negative sentiment (Beri 2020).

The posts collected from Twitter will consist of irregular time intervals and thus, the mean sentiment score of each day is to be computed to obtain an evenly spaced time series that matches the economic-based indicators. Using VADER to compute sentiment score presents a limitation of being unable to take the weight of each tweet into account, meaning that the number of users reached is neglected.

4.3 Data pre-processing

The following steps to be involved in data pre-processing:

- **Data cleaning**: Studying the dataset to handle missing values, examine outlers and produce simple graphs to visualize the stock information.
- **Standardization**: Standardizing using StandardScaler to ensure all variables fall into a same range
- **Principal Component Analysis**: Reducing the dimensionality of dataset.
- **Feature selection**: Computing additional technical indicators and selecting the top correlated features using SelectKBest.

4.4 Model development and algorithms

The following steps to be involved in model building:

- **Train-test-split**: Splitting data into 70% train and 30% test set using Scikit Library
- **KFold cross-validation**: Measuring how well the predictions made by the model match the observed data
- **Modelling**: Training base models and hyper-parameters tuned models of different classification and regression algorithms.
- **Running prediction** on the models using the 30% test set

Classification models including Decision Tree Classifier (DTC), Random Forest Classifier (RFC), and Support Vector Machine (SVM) are developed to identify the upwards and downwards crypto price movement signaling buy or sell indicators. Regression models including Decision Tree Regressor (DTR), Random Forest Regressor (RFR), and Support Vector Regressor (SVR) are built on the purpose of predicting the next day's closing price of the crypto. A detailed description of each algorithm to be explored in the following section.

4.4.1 Decision Tree (DTC DTR)

DT is one of the most used algorithms to predict the value of the discrete-valued target in both classification and regression problems as it is fast and efficient compared to other classification algorithms. Like the tree-based algorithms, it does not require any feature transformation for non-linear data as multiple weighted combinations are not considered simultaneously. Multiple studies have proven the effectiveness of DT in predicting crypto price (Huang et al. 2019, Inamdar et al. 2019). The majority of the studies adopted Decision Tree Classifier (DTC) to predict the upward and downward trends of the crypto price movement but rarely used Decision Tree Regressor (DTC).

DTR follows the similar rules of the binary Recursive Partitioning (RP) algorithm, which uses continuous values of the predictors to split the data into different partitions repeatedly. Rules such as weight fraction for leaf node are identified to train the models (Alahmari 2020*b*).

Research suggests using multiple trees to ensure the reliability of the DT model due to its high sensitivity to the input data. This provides an incentive to adopt the Random Forest comprised of a varying number of trees.

4.4.2 Random Forest (RFC RFR)

RF reduces the risk of overfitting and accuracy is much higher than a single decision tree as a more generalized model is usually created (Cutler et al. 2011). Unlike DT which gives high importance to a particular set of features, RF does not depend highly on any specific set of features, allowing it to generalize over the data in a better way and produce a more accurate result. Caruana & Niculescu-Mizil (2006) claim that Random Forest Classifier (RFC) can yield higher accuracy than algorithms such as artificial neural networks and vector machines, supported by Inamdar et al. (2019) using Random Forest Regressor (RFR) to predict crypto price movement.

Like DTR, Random Forest Regressor (RFR) uses a Decision Tree for regression. As this algorithm can still perform relatively well when a significant percentage of data is missing, coupled with its relatively low computational cost, it is highly adopted in experiments where a large dataset is involved.

The only limitation of RF regardless of RFC or RFR is the higher training time and complicated interpretability.

4.4.3 Support Vector Machine (SVM SVR)

SVM finds a hyperplane in N-dimensional space (N — the number of features) that distinctly classifies the data points, in this case, either upwards or downwards. To handle non-linear models, SVM adopts different kernels such as Gaussian and the polynomial kernels to map the inseparable input data to a higher dimensional hyperspace where Gaussian kernels are claimed to yield desirable performance under general smoothness assumptions (Sebastiao & Godinho 2021). SVM often yields better performance compared to other algorithms (Sebastiao & Godinho 2021, Altan & Karasu 2019). In many cases, SVM is integrated with Particle Swarm Optimization (PSO) for feature selection where it often outperforms the individual SVM. For instance, Hitam et al. (2019) concludes that SVM-PSO can enhance the performance accuracy of the Bitcoin price prediction model significantly by up to 11% compared to the SVM. Strikingly, it can even outperform the complex Artificial Neural Networks (ANN) in a study conducted to predict the reservoir annual inflow (Hitam et al. 2019).

For regression, it is called Support Vector Regression (SVR), first identified by Vladimir Vapnik and his research team in 1992 (Guyon et al. 1992). This algorithm has received an increasingly wider recognition due to its strong nonlinear learning capability (Fan et al. 2013, Rustam & Kintandani 2019), which is highly suitable to the nonlinear dynamics of the Bitcoin series with a high degree of chaoticity and fractality. Nonetheless, Felizardo et al. (2019) criticizes the performance of SVR for long-term prediction to be highly inaccurate citing that the random component gets more important and affects accuracy. Fadil et al. (2021) recommends the implementation of hyperparameter tuning using Grid Search Optimization as it is proven to produce better accuracy.

The study might also include other algorithms such as bagging, boosting, and stacking to enhance the performance of the above-mentioned algorithms.

4.5 Prediction Interval

The temporal length window has been considered when predicting price movement as the news and public sentiment are generally not reflected on the price movement immediately, evidenced by Bollen et al. (2010), Valencia et al. (2019). Nonetheless, the temporal length window presented by researchers varies due to the different methodologies used. Bollen et al. (2010) also discovered that despite tweets being correlated to the price movement, it generally takes around 3 to 4 days for the change of public sentiment to be reflected on the price, supported by Valencia et al. (2019). On contrary, Pant et al. (2018)'s experiment presented a shorter time lag of only 1 day as they adopted a different Pearson coefficient method to compute the correlation. Due to the time lag, assumptions can be made that the accuracy of the model is higher when performing longer-term trading instead of intraday trading. This provides a rationale for a majority (79%) of the studies conducted that predicted price movement on a 24h interval but not any duration lesser than that.

In this research, a prediction interval of 24h, 48h, 72h, and 96h to be utilized to examine the most optimal interval for crypto price prediction.

4.6 Automated Trading Bot

Automated trading bot to be developed using Python, looping through the predicted signal column (1 indicating upward trend and 0 for downward trend) to conduct day trade and decide whether to sell or buy. An initial amount of investment to be set at $100,000 and the remaining amount after 30 days of trading using indicators generated by different algorithms are compared.

4.7 Model Evaluation

Classification: Classification Report and Confusion Matrix which allows the evaluation of a classification model through 4 key indicators: Accuracy, Precision, Recall, and F1-score to be used to measure the classification accuracy.

Regression: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared, which represents the coefficient of how well the values fit compared to the original values to be used to measure the regression accuracy.

5 Limitations

The study presents several limitations due to the time and resource constraints:

- Only posts from Reddit are extracted but there are other platforms where investors can share their views such as Twitter, Facebook, or trading platforms such as Etoro. This limits how far the findings can be taken and applied to real-life cases.
- Limitation on computation resources, especially the specification of hardware might limit the algorithms that can be examined due to the excessively long training time.

6 Ethical Analysis

There are to be no other individuals to be interviewed and no personal data collection to be performed.

The data required for the research includes the technical indicators (closing prices and volume traded of cryptocurrencies) is to be obtained on the open-source websites, and sentimental indicators (social media posts on Twitter and Reddit) are to extracted using Reddit and Twitter API through proper application to the platforms without any infringement of data privacy.

The discussion with the supervisor confirms that this research does not involve any ethical, social, legal or professional issues.

7 Risk Analysis

The risk identified is classified into the following categories:

Hardware and software risk: Failure might occur when executing the programs due to the device's malfunction and the system error that requires immediate vendor support to avoid delay in the research. Hardware failure might result in failing to upgrade systems and continuing to rely on end-of-life (EOL) hardware and software that can create major cybersecurity vulnerabilities, such as computer hacking that can cause information lost.

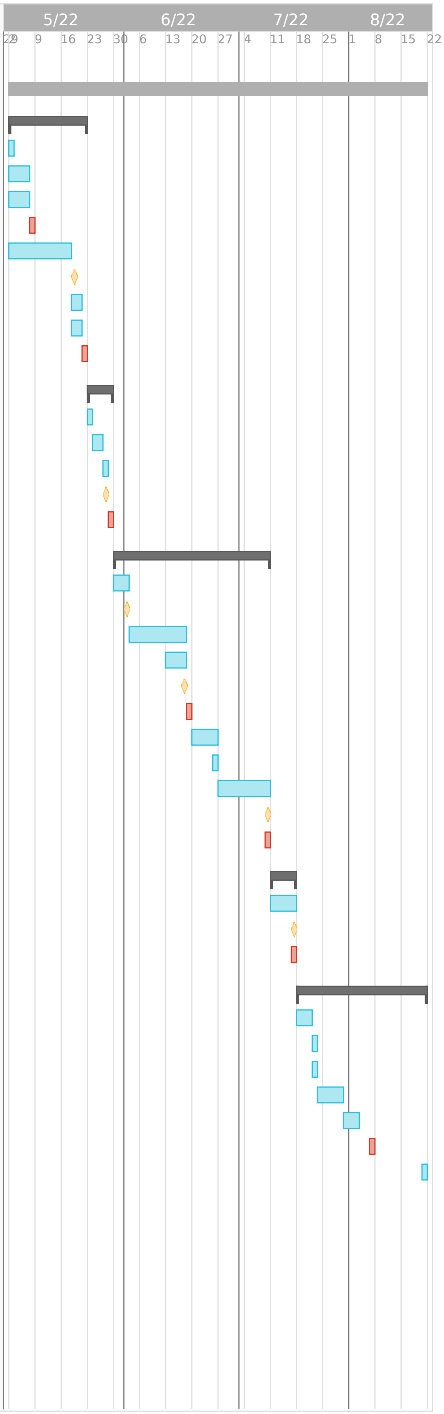
The risk can be minimized by preparing another desktop/laptop to serve as the backup during the failures and uploading the progressing research work to the clouds to be retrieved whenever these unforeseen events happen.

Health and safety risks: Health issues might arise due to the use of computing hardware for a long period. These include, but are not limited to postural problems, visual problems, and fatigue and stress. These could be overcome by suitable task design to incorporate breaks to ensure the posture and visual demands change.

Environmental risk: Uncontrollable natural disasters including earthquakes, tsunamis, flood might happen during the research period that can result in power shortage to support the running of computing hardware.

8 Project Plan

Dissertation	**0h**	**0%**
Phase 1 - Getting started	**0h**	**0%**
Set up research topic	0	0%
Review dissertation proposal	0	0%
Confirm research questions	0	0%
Supervisor meeting	0	0%
Research and review literature	0	0%
MIL: Complete the literature review section	0	0%
Finetune proposal if needed	0	0%
Confirm algorithms	0	0%
Supervisor meeting	0	0%
Phrase 2 - Data collection and analysis	**0h**	**0%**
Collect data required	0	0%
Conduct data cleaning if required	0	0%
Perform feature selection	0	0%
MIL: Finalize features for algorithms	0	0%
Supervisor meeting	0	0%
Phrase 3 - Research implementation	**0h**	**0%**
Write and methodology section	0	0%
MIL: Complete methodolody section	0	0%
Pilot study, build and review algorithms	0	0%
Analyze model performance	0	0%
MIL: Complete pilot study	0	0%
Supervisor meeting	0	0%
Finetune algorithms	0	0%
Finalize algorithms	0	0%
Demo trade using finalized algorithms	0	0%
MIL: Obtain research results	0	0%
Supervisor meeting to discuss findings	0	0%
Phrase 4: Results interpretation	**0h**	**0%**
Analyze research results	0	0%
MIL: Complete discussion section	0	0%
Supervisor meeting	0	0%
Phrase 5: Final report	**0h**	**0%**
Write limitation and conclusion	0	0%
Write abstract	0	0%
Write table of contents	0	0%
Edit final report	0	0%
Tentative task: write a publication	0	0%
Supervisor meeting	0	0%
MIL: Submission of final report	0	0%



References

Alahmari, S. (2020*a*), 'Predicting the price of cryptocurrency using support vector regression methods', *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES* **15**.

Alahmari, S. (2020*b*), 'Using nonlinear machine learning algorithms to predict the price of cryptocurrencies'.

Altan, A. & Karasu, S. (2019), 'The effect of kernel values in support vector machine to forecasting performance of financial time series and cognitive decision making', **4**, 17–21.

Antweiler, W. & Frank, M. Z. (2001), 'Is all that talk just noise? the information content of internet stock message boards," unpublished manuscript'.

Balcilar, M., Bouri, E., Gupta, R. & Roubaud, D. (2017), 'Can volume predict bitcoin returns and volatility? a quantiles-based approach', *ERN: Asset Pricing Models (Topic)* .

Beri, A. (2020), 'Sentimental analysis using vader', https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664/. [Accessed on 8 February 2022].

Bollen, J., Mao, H. & Zeng, X.-J. (2010), 'Twitter mood predicts the stock market', *Journal of Computational Science* **2**.

Bradley, S. (2021), 'What are the cryptocurrency trends to expect in 2022? experts share their opinions', https://www.chroniclelive.co.uk/business/what-cryptocurrency-trends-expect-2022-22593114. [Accessed on 8 February 2022].

Brock, W., Lakonishok, J. & LeBaron, B. (1992), 'Simple technical trading rules and the stochastic properties of stock returns', *The Journal of Finance* **47**(5), 1731–1764.

Caruana, R. & Niculescu-Mizil, A. (2006), 'An empirical comparison of supervised learning algorithms', *Proceedings of the 23rd international conference on Machine learning* .

Chen, Q., Zhang, W. & Lou, Y. (2020), 'Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network', *IEEE Access* **8**, 117365–117376.

Chen, Z., Li, C. & Sun, W. (2020), 'Bitcoin price prediction using machine learning: An approach to sample dimension engineering', *Journal of Computational and Applied Mathematics* **365**, 112395.

Choi, H. (2020), 'Investor attention and bitcoin liquidity: Evidence from bitcoin tweets', *Finance Research Letters* **39**, 101555.

Crypto (2021), 'Global cryptocurrency adoption doubled since january', https://blog.crypto.com/global-crypto-users-over-200-million/. [Accessed on 8 February 2022].

Cryptopotato (2021), 'Bitcoin volatility, biden's executive order, and russia-ukraine woes: This week's crypto recap', https://news.coinxhigh.com/2022/03/11/bitcoin-volatility-bidens-executive-order-and-russia-ukraine-woes-this-weeks-crypto-recap/. [Accessed on 8 February 2022].

Cutler, A., Cutler, D. & Stevens, J. (2011), *Random Forests*, Vol. 45, pp. 157–176.

Fadil, I., Helmiawan, M. & Sofiyan, Y. (2021), Optimization parameters support vector regression using grid search method, pp. 1–5.

Fan, G.-f., Qing, S., Wang, H., Hong, W.-C. & Li, H.-J. (2013), 'Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting', *Energies* **6**, 1887–1901.

Felizardo, L., Oliveira, R., Del-Moral-Hernandez, E. & Cozman, F. (2019), Comparative study of bitcoin price prediction using wavenets, recurrent neural networks and other machine learning methods, pp. 1–6.

Giaglis, G., Georgoula, I., Pournarakis, D., Bilanakos, C. & Sotiropoulos, D. (2015), Using time-series and sentiment analysis to detect the determinants of bitcoin prices.

Greenopolis (2021), 'A simple guide to the meme cryptocurrency: Dogecoin', https://greenopolis.com/a-simple-guide-to-the-meme-cryptocurrency-dogecoin/. [Accessed on 8 February 2022].

Guyon, I., Boser, B. E. & Vapnik, V. N. (1992), Automatic capacity tuning of very large vc-dimension classifiers, *in* 'NIPS'.

Hitam, N. A., Ismail, A. & Saeed, F. (2019), 'An optimized support vector machine (svm) based on particle swarm optimization (pso) for cryptocurrency forecasting', *Procedia Computer Science* **163**, 427–433.

Huang, J.-Z., Huang, W. & Ni, J. (2018), 'Predicting bitcoin returns using high-dimensional technical indicators', *The Journal of Finance and Data Science* **5**.

Huang, J.-Z., Huang, W. & Ni, J. (2019), 'Predicting bitcoin returns using high-dimensional technical indicators', *The Journal of Finance and Data Science* .

Inamdar, A., Bhagtani, A., Bhatt, S. & Shetty, P. (2019), Predicting cryptocurrency value using sentiment analysis, pp. 932–934.

Jaquart, P., Dann, D. & Martin, C. (2020), Machine learning for bitcoin pricing — a structured literature review, pp. 174–188.

Kaminski, J. & Gloor, P. (2014), 'Nowcasting the bitcoin market with twitter signals'.

Leonard, J. (2020), 'Best meme coins to buy – top 5 meme coins 2022', https://www.economywatch.com/cryptocurrency/best-meme-coins-to-buy. [Accessed on 8 February 2022].

Locke, T. (2021), 'From bitcoin hitting $1 trillion in market value to elon musk's dogecoin tweets: 12 key crypto moments from 2021', https://www.cnbc.com/2021/12/27/12-key-moments-that-fueled-cryptos-record-growth-in-2021.html. [Accessed on 8 February 2022].

Mai, F., Bai, Q., Shan, J., Wang, X. S. & Chiang, R. H. L. (2015), The impacts of social media on bitcoin performance, *in* 'ICIS'.

Palmer, C. (2021), 'How social listening and machine learning are used to predict bitcoin price volatility', https://bitcoinmagazine.com/markets/social-sentiment-and-the-bitcoin-price. [Accessed on 8 February 2022].

Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K. & Lama, B. K. (2018), Recurrent neural network based bitcoin price prediction by twitter sentiment analysis, *in* '2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)', pp. 128–132.

Phaladisailoed, T. & Numnonda, T. (2018), Machine learning models comparison for bitcoin price prediction, pp. 506–511.

Pintelas, E., Livieris, I., Stavroyiannis, S., Kotsilieris, T. & Pintelas, P. (2020), *Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach*.

Resta, M., Pagnottoni, P. & De Giuli, M. (2020), 'Technical analysis on the bitcoin market: Trading opportunities or investors' pitfall?', *Risks* **8**, 44.

Rustam, Z. & Kintandani, P. (2019), ‘Application of support vector regression in indonesian stock price prediction with feature selection using particle swarm optimisation’, *Modelling and Simulation in Engineering* **2019**, 1–5.

Sebastiao, H. & Godinho, P. (2021), ‘Forecasting and trading cryptocurrencies with machine learning under changing market conditions’, *Journal of Financial Innovation* **7**.

Shynkevich, A. (2012), ‘Performance of technical analysis in growth and small cap segments of the us equity market’, *Journal of Banking Finance* **36**, 193–208.

Sigalos, M. (2022), ‘Bitcoin’s wild price moves stem from its design — you’ll need strong nerves to trade it’, https://www.cnbc.com/2021/05/19/why-is-bitcoin-so-volatile.html. [Accessed on 8 February 2022].

Special, E. S. (2021), ‘Extent of elon musk’s influence on cryptocurrency; where is it headed?’, https://economictimes.indiatimes.com/markets/cryptocurrency/extent-of-elon-musks-influence-on-cryptocurrency-where-is-it-headed/articleshow/83037268.cms. [Accessed on 8 February 2022].

Stavroyiannis, S., Babalos, V., Bekiros, S., Lahmiri, S. & Uddin, G. S. (2019), ‘The high frequency multifractal properties of bitcoin’, *Physica A: Statistical Mechanics and its Applications* **520**, 62–71.
URL: *https://www.sciencedirect.com/science/article/pii/S0378437118315504*

Stenqvist, E. & Lönnö, J. (2017), Predicting bitcoin price fluctuation with twitter sentiment analysis.

Valencia, F., Gómez-Espinosa, A. & Valdes, B. (2019), ‘Price movement prediction of cryptocurrencies using sentiment analysis and machine learning’, *Entropy* **21**, 1–12.