USING SOCIAL MEDIA INDICATORS FOR CRYPTOCURRENCY PRICE PREDICTION

By

Usama Bin Tariq

00000363524

A thesis submitted to the faculty of Computer Science Department,
Military College of Signals, National University of Sciences and Technology,
Islamabad, Pakistan, in partial fulfillment of the requirements for the degree of MS in
Computer Science

August 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that the final copy of MS/MPhil thesis entitled "Using Social Media Indicators for Cryptocurrency Price Prediction" written by Usama Bin Tariq, (Registration 00000363524), of Military College of Signals (MCS) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: Naima

Name of Advisor: **Assoc. Prof. Dr. Naima Iltaf**

Date:

Signature (HoD): Brig Head of Dept of CSE Mil College of Sigs (NUST)

Date: 22/8/23

Signature (Dean/Principal):

Date: 23/8/23 Brig Dean, MCS (NUST) (Asif Masood, Phd)

Declaration

I, Usama Bin Tariq declare that this thesis titled "Using Social Media Indicators For Cryptocurrency Price Prediction" and the work presented in it are my own and have been generated by me as a result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a Master of Science degree at NUST
- Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
- Where I have consulted the published work of others, this is always clearly attributed
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
- I have acknowledged all main sources of help
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Usama Bin Tariq
NUST00000363524

DEDICATION

The thesis is devoted to

MY FAMILY, TEACHERS & FRIENDS

for their love, all out encouragement and support

ABSTRACT

A cryptocurrency is a digital currency based on a decentralized blockchain network. Transactions in cryptocurrencies are performed without a central authority or single administration. The transactions are secured by the strong hashing algorithm (SHA-256). There are more than 22,000 cryptocurrencies in the market, with over a $1 Trillion market cap. BTC (Bitcoin) is a famous cryptocurrency, designed to be a decentralized and secure form of digital cash. Cryptocurrencies can be used to purchase goods and services, transfer funds, and even as investments. The use of cryptocurrency has increased in the last few years, cryptocurrencies are commonly used for investment purposes.

This research focuses on developing a price prediction model for Dash coin and Bitcoin Cash. The prices of cryptocurrencies are highly unstable, which makes it challenging to forecast future prices. Researchers used Twitter sentiments, news, and previous market data with the help of NLP, Machine Learning (ML), and Deep Learning (DL) techniques to predict the future prices of different currencies. Our research uses state-of-the-art DL techniques to build a prediction model. The inclusion of technical indicators such as the Relative Strength Index and Moving Average along with the Fear & Greed Index and historic data helps to capture market sentiment and improve the overall accuracy of the model. We trained Gradient Recurrent Unit (GRU) model for Bitcoin Cash, and Dash cryptocurrencies, and the results show that our approach has better results than others.

Our work focuses on predicting future close prices of cryptocurrencies. Our main objective is to incorporate additional key features such as technical indicators and the Fear & Greed Index data can lead to improved accuracy for existing market models available. This research will help investors to better understand the direction in which the market is moving.

ACKNOWLEDGEMENTS

I am thankful to ALLAH who has blessed me with the strength & the passion to pursue the subject thesis and I am obliged to Him for His benevolence and mercy. Without his consent, I could not have indulged myself in this demanding work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Bitcoin	BTC
Etherium	ETH
BCH	Bitcoin Cash
Litecoin	LTC
Decentralized finance	DeFi
Machine Learning	ML
Extreme Gradient Boosting	XGBoost
Deep Learning	DL
Recurrent Neural Network	RNN
Long Short-Term Memory	LSTM
Gated Recurrent Unit	GRU
Bi-Directional Long Short-Term Memory	Bi-LSTM
Bi-Directional Gated Recurrent Unit	Bi-GRU
Fear and Greed Index	FGI
Moving Average	MA
Relative Strength Index:	RSI
Moving Average Convergence Divergence Line	MACD
Mean Squared Error	MSE
Root Mean Squared Error	RMSE
Mean Absolute Error	MAE

Chapter 1

INTRODUCTION

Cryptocurrencies are digital currencies that are completely computerized based, there is no physical existence of these currencies, and it eliminates the regular means of physical currency. Cryptocurrencies are based on decentralized blockchain networks, and transactions in cryptocurrencies are secured by SHA-256 and MD5 hashing algorithms. Transactions in a cryptocurrency network are transparent to all users for verification and no one controls the transactions or flow of money. The first cryptocurrency, Bitcoin (BTC) was published in 2008 by an unknown creator named Satoshi Nakamoto. The starting price of BTC was $0 but it jumped to $0.09 in the first year. The price of BTC was raised constantly. At the end of 2021, the price of a BTC was $65K [1]. The total market volume of cryptocurrencies is around 872 Billion dollars as of Nov 2022 [2]. Bitcoin is the main contributor to the cryptocurrency pool with 36.67% of the market capitalization as of Nov 2022 [3]. Cryptocurrencies are becoming a new way of financial transactions. Etherium (ETH) is the second most popular cryptocurrency. ETH development was started back in 2013 and in 2015 the ETH went live [4]. Bitcoin was mainly developed for making transactions, however, the Etherium architecture was not just to make financial transactions, etherium was designed for the development of decentralized applications which is also known as dApp. DeFi is one example of dApp where you can buy, sell and trade digital assets such as pictures.
The future of cryptocurrencies is bright, as the growing acceptance of digital cryptocurrencies is driving more adoption of cryptocurrencies. Cryptocurrencies are becoming more accepted as a legitimate form of payment because of their growing popularity and usage. However there are many challenges in cryptocurrencies, there are many countries where cryptocurrency is not recognized officially because of money laundering and security issues.
Cryptocurrency prices are highly volatile, making it challenging to invest in them. Investors make use of various techniques to monitor price movements and trends, using technical indicators to forecast future closing prices. Many factors can influence cryptocur-

rency price fluctuations, including but not limited to market demand and supply, regulatory changes, news, media coverage, and changes in technology or technical developments [5]. In the current era, social media has the potential impact on cryptocurrency prices, with social media influencers encouraging followers to either pump or dump the cryptocurrency market, resulting in uptrends or downtrends in the respective cryptocurrency markets [6].

Ref	Year	Description	Cryptocurrencies	Method	Dataset
[7]	2019	BART model was found to be more accurate than the simple ARIMA model.	BTC, ETH, XRP	Arima, Decision Tree	YFinance
[8]	2022	Twitter sentiment and historic prices used to predict future prices.	DASH, BCH	GRU, LSTM	Twitter sentiment, historic data
[13]	2020	Stochastic probability technique is used to predict prices.	BTC, ETH, LTC	MLP, LSTM	Twitter API, coinmarketcap.com
[14]	2022	Using past data, the price for the next 1-day period was predicted.	ETH, DOGE, EOS, BTC	LSTM	YFinance
[15]	2019	News data sentiment score historical prices are used to predict cryptocurrency prices.	ETH	LSTM	NewsNow, poloniex.com
[16]	2021	BTC data was used as an up/down indicator predict LTC and ZCH prices.	LTC, ZCH	LSTM, GRU	Investing.com
[17]	2021	LTC & XMR historic data is used to predict future prices.	LTC, XMR	RNN, LSTM, GRU	-
[27]	2022	The future prices of BTC, ETH, BNB, ADA, and USDT have been predicted using historical data	BTC, ADA, ETH, BNB, USDT	Bi-LSTM, LSTM	YFinance

Table 1.1: Relative Comparison

1.1 Problem Statement and Objectives

Cryptocurrency is a digital currency in a decentralized network, it is becoming more popular as means of investment and buying and selling of goods. The price of the cryptocurrency is highly volatile which makes it difficult to predict the future close prices. Because of the volatile nature of cryptocurrencies, it's hard for investors to predict the direction of the market and as a result, the investors end up losing their money. Previous studies used different techniques to forecast the future closing prices of cryptocurrencies including ML and DL algorithms. Most of the researchers focused on pre-historic data, technical data of cryptocurrency, social media sentiments, and news sentiments, those research showed impressive results. There are a lot of factors involved in fluctuations of cryptocurrency price including but not limited to social media, historic data, trends, surveys, and news. However, the impact of social media influencers and technical indicators has not been fully explored. Therefore, the development of a system that can predict future cryptocurrency prices using historical data and other key factors is critical. The purpose of this research is to create a system that can forecast the future close prices of Bitcoin Cash and Dash using historical data, social media indicators, and technical indicators. This research will contribute to building a more accurate model for the prediction of cryptocurrency prices, and as a result, this research will help investors to have an idea of cryptocurrency trends before time. The aim is to present a deep learning-based system designed for cryptocurrency price prediction The research Objectives are as follows:

- To present Deep Learning based system for predicting the cryptocurrency price.

1.2 Contributions

The contributions of our work are as under,

- This research focuses on deep learning solutions to predict the future close prices of BCH and Dash.

- To enhance the model's capability to correctly predict the future close prices of BCH and Dash.

Only an outline of the proposed framework at a conceptual level is provided here, more details are provided in later chapters.

Technical Indicators: Technical indicators are mathematical calculations and statistical measurements used to analyze the trends and price movement in Stock Market or Cryptocurrency Market. Technical indicators are one of the important factors which investors used while investing in the cryptocurrency market. A study [18] was carried out which considered technical indicators while predicting the price of the cryptocurrency. It was found that technical indicators have a positive impact on predicting future close prices. The neural network model with technical indicators had significantly better results than the rest of the neural network without technical indicators.

Gated Recurrent Unit: LSTM, a variant of the Recurrent Neural Network (RNN), is a type of neural network. It is designed to develop models which require long-term and short-term memory in sequential data [19]. LSTM incorporates input, forget, and output gates. Additionally, the Gated Recurrent Unit (GRU) also falls into the category of neural networks and is a variation of the traditional RNN. It is mainly used for long-term memory in sequential data. GRU utilizes only update and reset gates. In comparison, LSTM is more complex than GRU.

Applications: Cryptocurrency has emerged as a new way of investing, selling, and purchasing goods. The emerging adoption of the cryptocurrency market has resulted in the creation of systems that can assist investors in understanding market trends. Cryptocurrency prices are highly volatile because of different factors including but not limited to social media activity, pre-historic price, government regulations, and technical parameters, because of the volatility it is hard to predict the trend of the cryptocurrency price whether it going upward or downward. A lot of investors were at a loss investing in the market because they didn't have an idea about the direction of the market. This research will help investors to have an idea of the close prices of the following month based on prehistoric prices, social media sentiments, and technical indicators. As a result, investors can decide better which cryptocurrency to invest in.

1.3 Thesis Outline

The thesis is broken down into five parts:

- Chapter 1: Chapter 1 consists of an introduction and research objectives. It also contains the contributions that we have made in this report.

- Chapter 2: In Chapter 2, a background study and literature review are provided along with a brief description of existing techniques and quantitative measures used in this thesis report are given.

- Chapter 3: This chapter focuses on the data collection, data cleaning, and preprocessing, descriptive statistical analyses of the dataset, relevant feature selection, and then calculates the technical indicators parameters which later can be utilized as features when implementing the machine learning classification model in the next chapter.

- Chapter 4: We have done our experiment and performance evaluation in this chapter. Studying and testing different deep learning models for predicting future close prices is proposed along with experimental results and a comparison with existing techniques is provided.

- Chapter 5: This chapter concludes the report and proposed future work.

Chapter 2

PRELIMINARIES

2.1 Cryptocurrency

Cryptocurrencies are digital currencies that exist only in a computerized form. They do not have a physical presence and replace traditional forms of currency. Cryptocurrencies operate on decentralized blockchain networks, and transactions are secured utilizing hashing algorithms like SHA-256 and MD5. The blockchain is a decentralized ledger that exists on multiple nodes throughout the network. Each block in this network is unique and verified by all nodes, ensuring its immutability. As a result, transactions are permanent and transparent to all nodes. Cryptocurrencies utilize cryptography techniques for transactions. Each user has their own wallet protected by a public and private key. A transaction includes the amount, the recipient's public key, and a digital signature generated using the sender's private key. The transaction is then sent to the cryptocurrency network, where nodes, acting as miners, validate it by performing complex mathematical operations. After multiple nodes verify the transaction, it is subsequently added to the blockchain network. This addition is accomplished through mechanisms such as proof-of-work and proof-of-stake. In proof-of-work, miners solve complex mathematical problems, while in proof-of-stake, a certain number of validators validate the transaction. Miners are rewarded with the respective cryptocurrency for solving these complex mathematical operations [20].

Cryptocurrencies work on a decentralized network, ensuring that no single central authority controls the flow of the cryptocurrency. The distributed network makes it difficult to hack or tamper with because all nodes have the same copy of the network. If a single node is tampered with, other nodes will reject the changes, preserving the integrity of the main blockchain. Cryptocurrency owners store their assets in digital wallets, which can be hardware devices, software applications, or exchanges like Binance [21]. Investors use cryptocurrency exchanges such as Binance [21], Coinbase [22], and Kucoin [23] to perform investments in specific cryptocurrencies. Over time, cryptocurrencies and blockchain technology have continued to evolve. Researchers are working on stability and scalability solutions to handle increased transaction volumes, as well as addressing privacy and energy consumption concerns.

2.2 Bitcoin & Altcoins

The first cryptocurrency, Bitcoin (BTC) was published in 2008 by an unknown creator named Satoshi Nakamoto. The starting price of BTC was $0 but it jumped to $0.09 in the first year. The price of BTC was raised constantly. At the end of 2021, the price of a BTC was $65K [1]. The total market volume of cryptocurrencies is around 872 Billion dollars as of Nov 2022 [2]. Bitcoin is the main contributor to the cryptocurrency pool with 46.31% and Etherium follows with 19.69% of the market capitalization as of the most recent data of May 2023 [24] as presented in Figure 2.1. Cryptocurrencies are becoming a new way of financial transactions. Etherium (ETH) is the second most popular cryptocurrency. ETH development was started back in 2013 and in 2015 the ETH went live [4]. Bitcoin was mainly developed for making transactions, however, the Etherium architecture was not just to make financial transactions, etherium was designed for the development of decentralized applications which is also known as dApp. DeFi is one example of dApp where you can buy, sell and trade digital assets such as pictures.

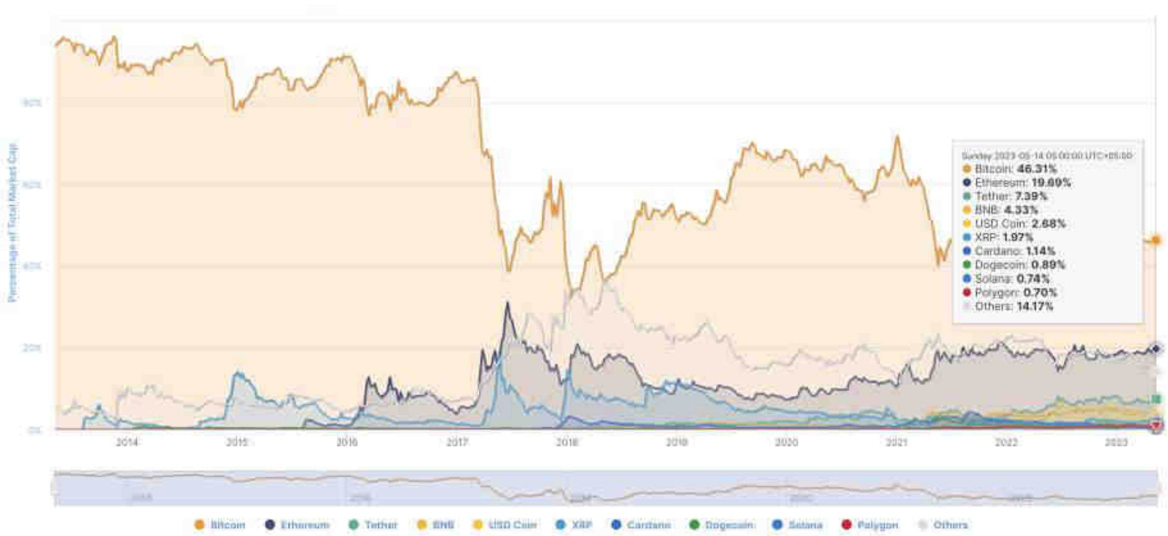

Figure 2.1: Cryptocurrency Market Cap (2013–2023).

Any cryptocurrency other than Bitcoin is known as an altcoin. As of May 2023, there are over 23,000 cryptocurrencies, and this number continues to gradually increase over time.

Altcoins represent approximately 54.69% of the total market capitalization of cryptocurrencies. Investors often prefer to invest more in altcoins compared to Bitcoin, as altcoins tend to offer higher returns but also come with higher risks. On the other hand, Bitcoin is considered a safer investment [25]. In terms of market capitalization ranking, Bitcoin is followed by Ethereum, Tether, BNB, USD Coin, XRP, Cardano, and so on [3].

2.3 Social Media Impact on Cryptocurrency

Cryptocurrencies prices are linked to the cryptocurrencies production rate factor but are not limited to it, it is also linked with the momentum and trends in the market as well as the investor's attention to the market [9]. A study founds that hackers attacked different crypto exchanges and make the cryptocurrency more or less volatile by taking the cryptocurrency out of the market which results in a loss of confidence in the cryptocurrency market [10]. To evaluate the impact of social media on cryptocurrency prices, a study was carried out solely on Elon Musk an individual billionaire tweet. It was found out that the price of bitcoin jumped +3.58% when Elon Musk add "#bitcoin" in their bio, the price of Dogecoin increased abnormally to 12.5% in 2 Min and 26.5% in an hour when Elon tweets about Dogecoin [11]. Elon Musk revealed in May 2021 that he was working together with Dogecoin developers to enhance the system's efficiency. Shortly after his tweet, the price of Dogecoin increased [12]. Figure 2.2 shows the trading view of the DOGE-USDT pair, it can be seen that the price of Dogecoin jumped from 0.39USDT to 0.73USD in May 2021.

Figure 2.2: Elon Musk Impact on Doge Coin

Social media has a potential impact on cryptocurrency prices, people on social media talk about cryptocurrencies on different forums such as Twitter, Facebook, Reddit, and more, from where other people get an idea of market movement. A study [26] used Bitcoin Talk Forums to find out the relation between social media and price fluctuations of cryptocurrency, the researcher used tokenized comments and LSTM was then used to forecast the future closing prices of Bitcoin, and it was found out that those comments have a potential effect on the price fluctuations.

There are a good number of cryptocurrency exchanges available in the market including Binance, Bitrue, Coinbase, and more. The cryptocurrency exchange market has a potential role in moving the price of cryptocurrency upward or downward. Binance is top of all other cryptocurrency exchange markets, whenever a new coin is listed on finance, there is a 90% chance that the coin price will jump high even for a few minutes, it is known as Binance Effect. Binance is not directly responsible for the fluctuations in the price, because the binance is one of the top markets, and it has a huge user base and trading volume which results in these fluctuations.

Figure 2.3: Binance Effect

As seen in figure 2.3 the price of 1Inch Cryptocurrency was 0.2000USDT at the time of opening trading at the binance platform, but as soon as the trading opened at binance the price jumps to 3.0885USDT and it closed at the 2.954USDT in the first day. This is just one example, in any currency listed on binance the price will always go up.

Whenever a new cryptocurrency project starts, the owner of the project starts the marketing on social media, the marketing campaign gets a hike when the top social media influencers take part in the campaign which results in gathering a lot of audience for the respective cryptocurrency, which results in the price hike of a particular currency.

2.4 Technical Indicators & Cryptocurrencies

Technical indicators are mathematical calculations and statistical measurements used to analyze trends and price movements in the stock market and cryptocurrency market. These indicators play an important role for investors when making investment decisions. There are many technical indicators available, including Moving Average, Relative Strength Index, Bollinger Bands, Moving Average Convergence Divergence (MACD) line, and many more. These indicators provide insights into price movements and trends in the respective markets, helping investors gain a better understanding of market direction. A study [18] was conducted that used technical indicators along with historical price data, using a deep learning algorithm. The results show significant improvements compared to models that only use historical price data as input. The results show that using technical indicators can improve the predictive accuracy of the model

Moving Average:

Moving average is the most commonly used technical indicator in the cryptocurrency market because of its simplicity and indications. Moving averages smooth the price data over a period of time by calculating the average value. There are two most commonly used types of Moving Average.

Simple Moving Average (SMA): SMA calculates the average price over a specified number of periods. SMA simply sums up the closing price of N number of days and then divides the result by N. Each day, one old value is dropped and the latest one is added to the calculation. SMA results are slow to recent price changes but it is useful in identifying long-term trends. Figure 2.4 shows an example of SMA with blue and orange lines for 9 days and 18 days as input respectively.

Exponential Moving Average(EMA): EMA is more reactive than SMA, EME assigns more weight to the close prices and makes it more responsive to the current price move-

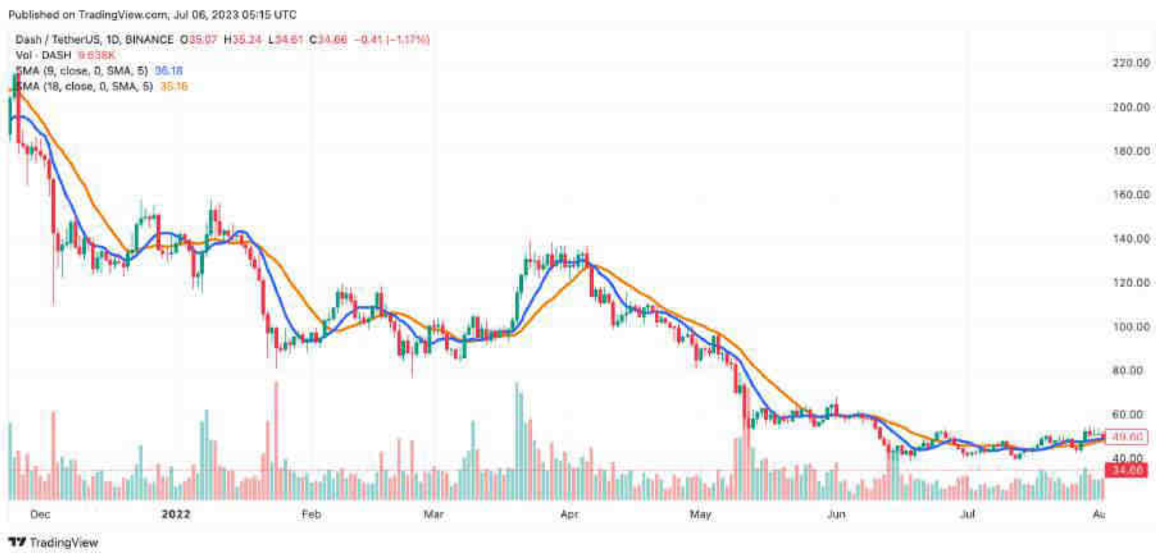

Figure 2.4: DASH-USD SMA

ment. EMA calculation is done in a way that provides more importance to the new data compared to the old data. Traders use EMA for short-term trading to generate signals.

When the prices cross the moving average line, it signals a bullish trend as a buy signal, on the other hand when the price moves below the moving average line it indicates a bearish trend. Figure 2.5 shows an example of EMA applied to the DASH-USDT trading pair. The graph displays two lines: one in blue and another in orange. These lines represent the EMA calculations for 18 and 36 days, respectively.

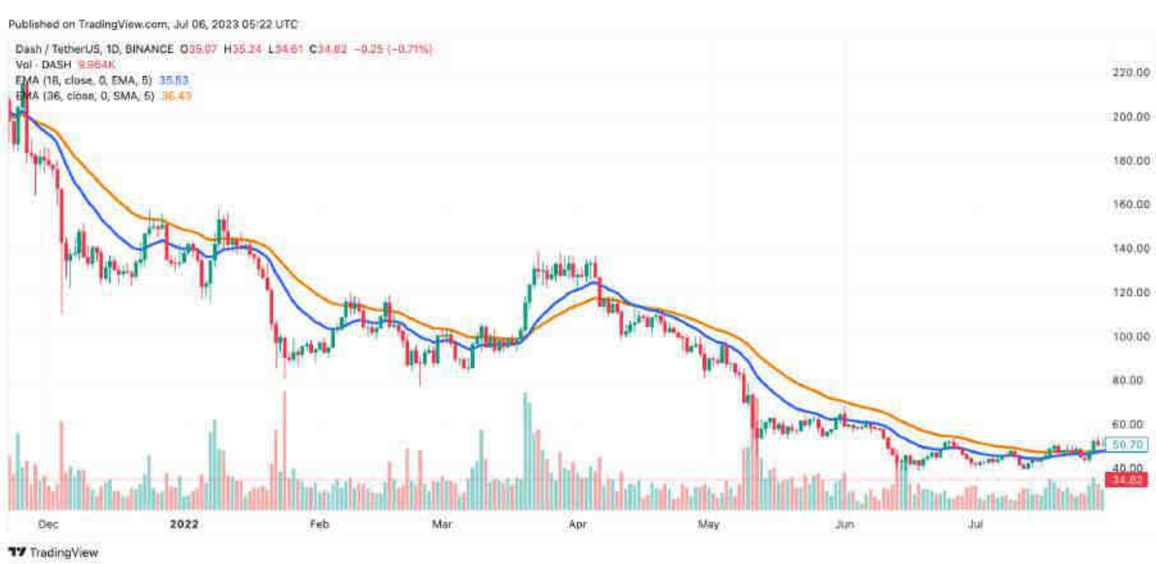

Figure 2.5: DASH-USD EMA

Relative Strength Index(RSI):

RSI is another widely-used indicator used by the traders in cryptocurrency market. The indicator measures both the rate and change of price and is used in the identification of overbought or oversold conditions.

RSI value is calculated based on average gains and losses recorded during a predefined period. The formula compares the average gain during the up periods and the average loss during the down periods and as a result, it generates a value between 0 - 100. A value above 70 indicates an overbought market, while a value below 30 indicates an oversold market. Figure 2.6 illustrate an example of RSI for the BCH-USD trading pair. The bottom purple line shows the RSI value for 14 days, it is also seen that there are two spots in the trading view where the market is overbought.

Figure 2.6: DASH-USD EMA

Moving Average Convergence Divergence Line (MACD):

The MACD is a technical indicator that identifies the buy and sell signals. The MACD indicator consists of the MACD line and the signal line. The signal line is used to generate the trading signal. When the MACD crosses above the signal line, it indicates a bullish market, and when it crosses below the signal line, it indicates a bearish market. The calculation of the MACD is given in upcoming chapters.

2.5 Technical Cryptocurrency Data

Each cryptocurrency has some technical data associated with it, technical data can also affect the price of the cryptocurrency. A few technical data are listed below

- **Volume:** Volume of the cryptocurrency represents the count of units for a particular cryptocurrency traded within a time period such as 1 Hour, 12 Hours, 24 Hours, 1 Week, or 1 Month. Higher market volume indicates that more people are interested in cryptocurrency. Higher trading volume can contribute to higher price volatility.

- **Hash Rate:** Hash rate is the speed of solving complex mathematical operations for a particular cryptocurrency on the transaction, the higher the rate of hash means that more miners are there it indicates that the network is secure and more investors are interested towards it.

- **Mining Difficulty:** Mining difficulty is how much it is challenging to find a valid hash for the transaction. It affects the rate of production of new coins.

- **Network Congestion:** Network congestion effect the transaction fees, when there are more pending transactions in the queue then investors pay more to make a transaction in priority which results in a higher fee. Network congestion also affects the price of the cryptocurrency.

- **Transaction Fee:** The transaction fee also affects the cryptocurrency price, the higher the transaction fee the lower the will be trading. If the transaction fee is higher, the user will try to use other cryptocurrencies instead.

- **Regulations:** The government regulations also impact the price of the cryptocurrency, positive news such as acceptance of cryptocurrency at the government level can increase the price.

- **Technology Upgrades:** Technology upgrades such as improvements in the way blockchain works, may also may result in a rise in the cryptocurrency price. Positive advancements can increase the confidence and adoption of particular cryptocurrencies.

2.6 Cryptocurrency Price Prediction Techniques

Cryptocurrency emerged as a new way of financial transactions, it became popular in the last decade and also took the researcher's intention to work on this. Investors used different techniques such as market analysis to forecast the close price of cryptocurrencies. Research also has used different techniques such as ML and DL to predict the future closing price of the cryptocurrency. This work will be very useful for new investors who want to have an idea of the price movement of the Dash Coin and Bitcoin Cash cryptocurrencies.

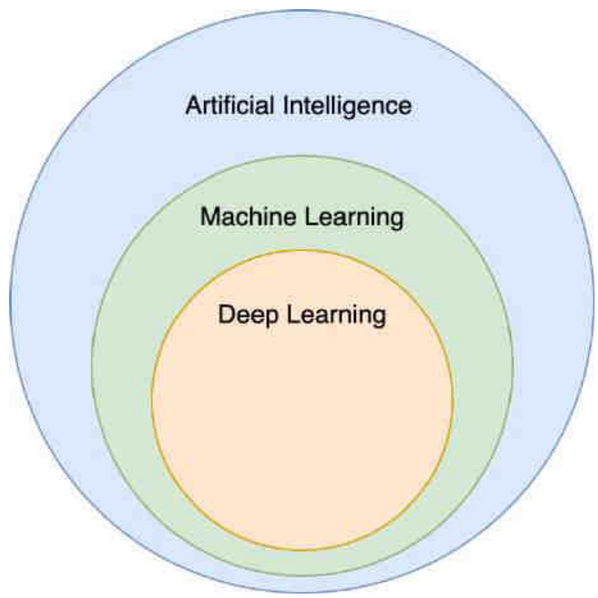

Figure 2.7: AI, ML, and DL Comparison

As shown in Figure 2.7 Artificial intelligence has multiple subtypes including ML. Deep Learning is considered a subcategory of Machine Learning. Furthermore, ML and DL are further divided into different techniques. Each technique is used for different data types and output needs.

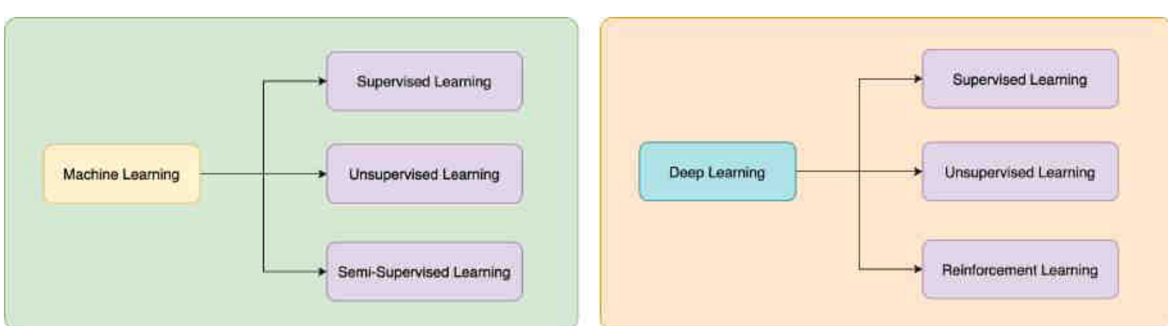

Figure 2.8: Machine Learning & Deep Learning Types

As shown in Figure 2.8, machine learning is divided into 3 broad subtypes which are supervised, unsupervised, and semi-supervised learning. Supervised learning is utilized with labeled data, an unsupervised learning technique is performed with unlabeled data and semi-supervised learning techniques are performed when some data is labeled and some is unlabeled. Whereas deep learning can't be categorized as machine learning but deep learning also solves the same problem in a different and efficient way. When we have a specific prediction problem to solve on a dataset, the very first step is to choose whether we need to go for Machine Learning or Deep Learning. After choosing the correct field the next step would be to choose the algorithm/technique which we will use to predict. Choosing an algorithm is a crucial step because our results solely depend on the algorithm. No one can tell which algorithm is better without trying it first. Each algorithm has its own pros and cons, not every dataset can be trained in every algorithm.

The researchers used different state-of-the-art ML and DL techniques to predict the future close pricing of different cryptocurrencies.

2.6.1 Machine Learning Techniques:

Machine learning has many classifications, and clustering regression algorithms, Figure 2.9 shows different categorized machine learning algorithms.

Random Forest Algorithm:

Random Forest is a well-known machine learning algorithm used for both classification and regression tasks. It relies on the Decision Tree Algorithm as its fundamental building block. A decision tree is a hierarchical structure like a tree, where each node represents a feature extracted from the dataset. Random Forests use a technique called bootstrap aggregating, or "bagging," to create multiple subsets of the original data. This process creates different training sets for every decision tree in the collection. During the construction of a decision tree, instead of considering all features, a random subset of features is selected for each split point. The random forest creates a set of predefined decision trees. Each tree is grown independently using training data and a feature set. When the algorithm is completed running it results in a forest, the predictions are derived by summing up the forecasts from all the individual trees. For regression tasks, the predictions of all trees are averaged to get the final prediction.

In study [33] researchers have taken different features including network data, market data,

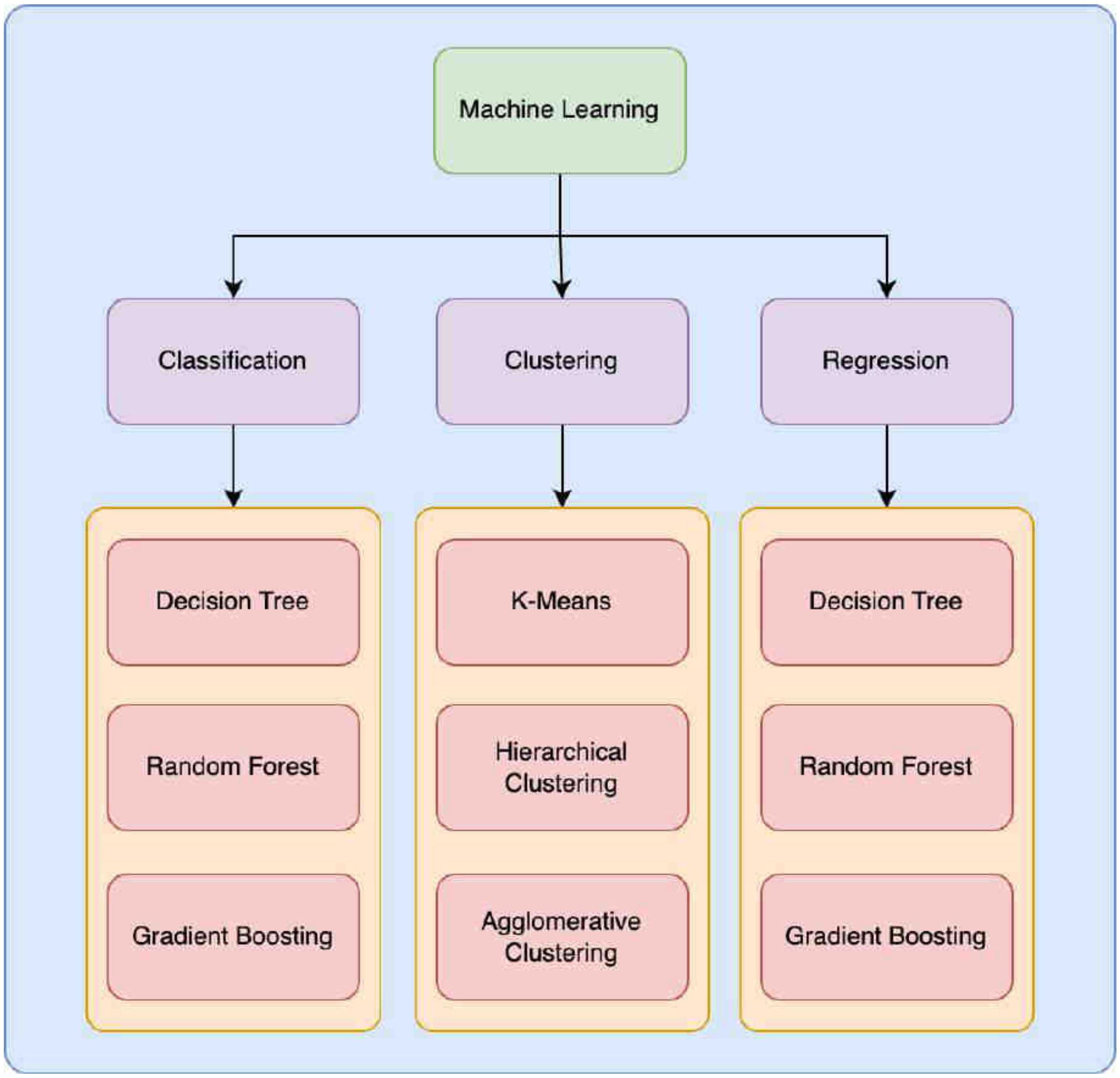

Figure 2.9: Popular Deep Learning Algorithms

and gold prices to forecast the close prices for Bitcoin, the researcher has included the usage of the Random Forest model in the price forecasting model.

Pros:

- Random forest can handle a large amount of data and it is less pruned to overfitting problems.
- It can handle numerical and categorical data and the output can be class or regression.
- It is robust and it can handle missing data.

Cons:

- Random forest can be computationally expensive when the number of trees is large.
- Data that has high dimensions and is sparse may not fit well in Random Forest.

Gradient Boosted Trees (GBT):

GBT is an ML algorithm that uses Decision Tree with gradient boosting. It is built on a decision tree algorithm, GBT follows an ensemble learning approach, and it is a way to build a stronger model by incorporating multiple weak models. GBT is started using a shallow decision tree, and the new tree is constructed which optimizes the loss function. The predictions from the newly created decision tree are added to the ensemble, with a learning rate that controls the contribution of each tree. The creation of a new tree is repeated until it achieves the specified number of trees. In GBT each new tree fixes the mistake which is done by the previous tree and in this way, the gradient decent approach is taken to provide the final tree.

In study [34] researcher had taken the open, low, high, close, volume, and market capital of Bitcoin (BTC), Bitcoin Cash (BCH), Etherium (ETH), Doge, Litecoin (LTC), IOTA, NEM, and NEO to predict the closing price using Gradient Boosted Tree, Neural Net, Ensemble, and KNN. As a result, GBT for Bitcoin has a Root Mean Squared Error of 3.849.

Pros:

- GBT has high accuracy for multiple ML tasks.
- GBT can handle categorical and numerical data.
- GBT can is used for unsupervised learning and regression tasks.

Cons:

- GBT is computationally expensive, it takes more computational power to solve and build up the model.
- GBT is prone to overfitting when the number of trees is large.

Extreme Gradient Boosting (XGBoost):

XGBoost is a well-known ML algorithm for various tasks. Classification and regression tasks can be solved by XGBoost. XGBoost is based on the ensemble technique, which incorporates multiple weak models to build up a stronger predictive model. The main idea

behind gradient boosting is to sequentially introduce new models to the ensemble, with each new model trained to correct errors made by the preceding models. This iterative process helps improve the overall accuracy and predictive power of the model.
In study [35] researchers have used the XGBoost, CNN, ARIMA, MLP, LSTM, and Baseline algorithms to forecast the future close prices of Ether, LTC, and XMR. XGBoost outperformed the other models.

Pros

- XGBoost is high in performance, it uses parallel processing and optimized algorithms.
- This algorithm is high in predictive accuracy as compared to other algorithms.
- XGBoost can handle numerical and categorical data.

Cons

- XGBoost can be resource expensive when dealing with large data.
- XGBoost hyperparameters can be set which can result in a time-consuming task to find the best one.

2.6.2 Deep Learning Techniques:

Figure 2.10 shows different deep learning algorithms that are common and most popular. For our use case of prediction of cryptocurrency close prices, RNN is most suitable, RNN has further different variations including LSTM and GRU.

Recurrent Neural Network (RNN)

RNN is a feed-forward neural network having a memory in it. Compared to conventional neural networks, RNN inputs and outputs are interdependent, it is designed to process sequential data and allow temporal data to persist in the network's hidden layers. The core concept of RNN is to get information from the previous output and include it in the next input.
RNN structure is like the conventional neural network, it has input, output, and hidden layers. Additionally, RNN has a time step layer that processes each sequential data at one

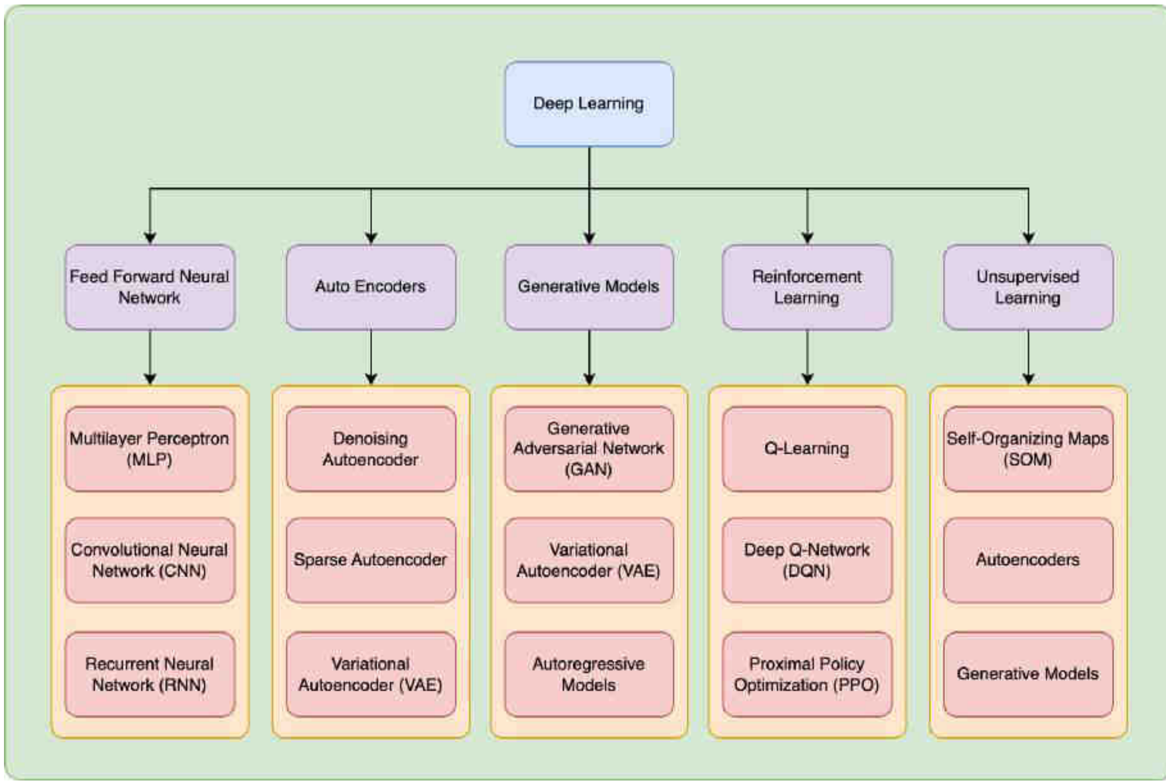

Figure 2.10: Popular Deep Learning Algorithms

timestamp, at each time stamp the input of the layer is combined with the output of the previous layer.

Some researchers focused on using RNN for the prediction of cryptocurrencies prices, in study [17] have incorporated LSTM, and GRU to forecast the future prices of Bitcoin, Litecoin, Ripple, Monero, Thether, and IOTA for 1 day window period. As the cryptocurrency data is sequential data so the data would best fit in the RNN model, as the RNN has a vanishing gradient problem so we need to change the algorithm to have better results.

Gradient Descent

In deep neural network training, the Gradient Descent algorithm is applied to optimize the model by iteratively adjusting the weights and biases to minimize the loss function. The single goal of the gradient descent aims to discover a set of weights and basis that results in a minimum difference between predicted and actual values. Equation 2.1 shows the calculation of the gradient descent. Gradient descent is all about finding the local minima, as shown in Figure 2.11.

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \nabla J(\theta^{(t)}) \tag{2.1}$$

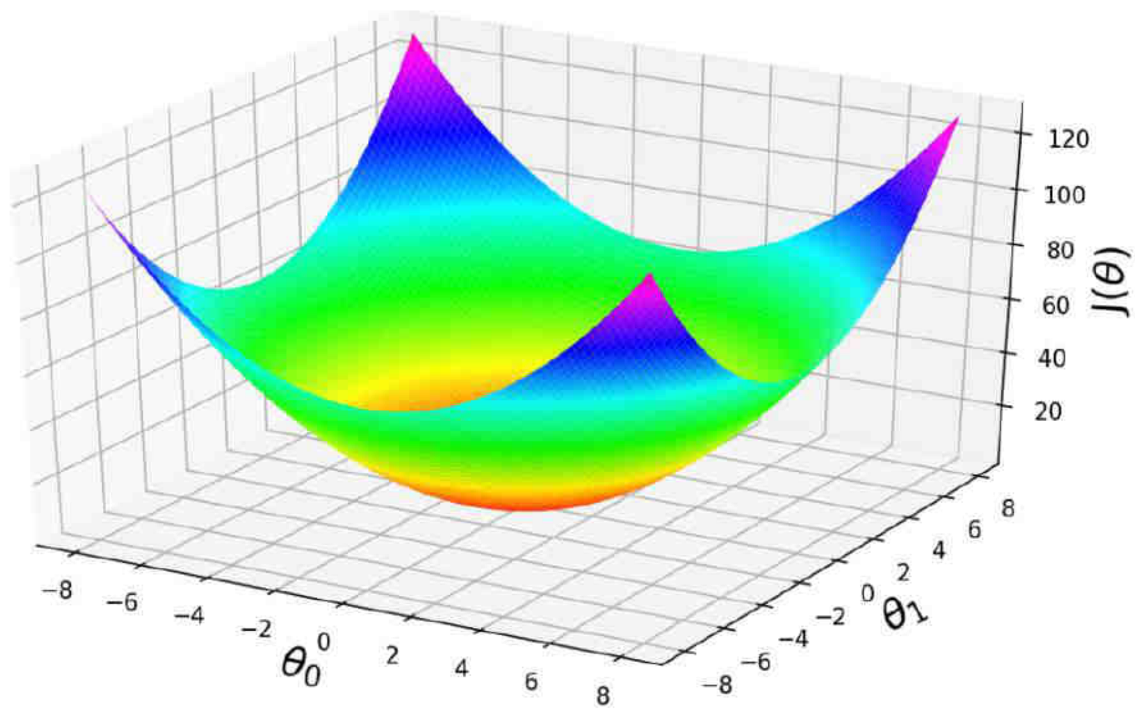

Figure 2.11: Gradient Descent

Vanishing Gradient

A vanishing gradient is a problem that occurs during the training of the deep neural network with many layers, when using the gradient descent for weight and basis the gradient may become very small or very big as they backpropagate to earlier layers. When the gradient becomes smaller, the changes in the weights and basis also become smaller and result in slow divergence and no update at all because of that the earlier layer receives a weak gradient and the weights and basis are not adjusted to learn from the data.
RNNs with a large number of recurrent connections have a vanishing gradient problem when the gradients are backpropagated through time, and during this process, the gradients can either shrink which results in the vanishing of the gradient, or become large which resulted in explode of the gradient. To overcome the vanishing gradient problem new architecture LSTM and GRU were developed.

Long Short Term Memory (LSTM)

LSTM is also a type of Artificial Neural Network (ANN) designed to solve the vanishing gradient problem and to build up long-term memory. LSTM has memory cells as the main building blocks of its architecture. The memory cell is responsible for storing and propagating information across different time steps. LSTM uses a gate mechanism to retain and forget information. 2.12 is the visualization of the inner architecture of the LSTM.

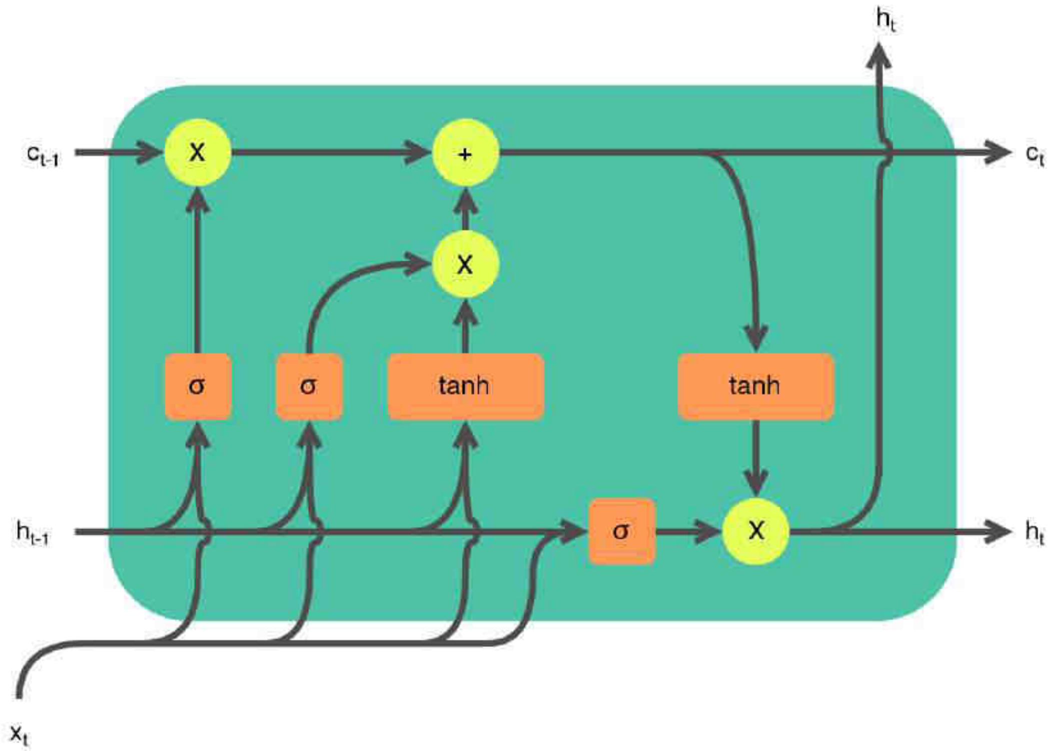

Figure 2.12: LSTM Inner Architecture

- **Input Gate:** $a_t = \sigma(U_{xi}x_t + U_{hi}h_{t-1} + b_i)$
- **Forget Gate:** $b_t = \sigma(U_{xf}x_t + U_{hf}h_{t-1} + b_f)$
- **Output Gate:** $c_t = \sigma(U_{xo}x_t + U_{ho}h_{t-1} + b_o)$
- **Candidate Cell State:** $\tilde{D}t = \tanh(Uxcx_t + U_{hc}h_{t-1} + b_c)$
- **Cell State Update:** $E_t = b_t \odot E_{t-1} + a_t \odot \tilde{D}_t$
- **Hidden State:** $h_t = c_t \odot \tanh(E_t)$

Many studies focused on using LSTM for the prediction of the close price of cryptocurrencies. In study [8], researchers have used LSTM and GRU for price prediction of DASH and BCH cryptocurrencies. They used Twitter sentiments, Bitcoin data as parent coin, and historical data of DASH and BCH to predict future close prices. Still, the researchers didn't consider other social media factors such as Google trends, surveys, Reddit forums, etc, and they also didn't consider using technical indicators as a helper to predict prices.

In study [14] used LSTM for the prediction of Ethereum, Bitcoin, EOS, and Doge cryptocurrency close prices for the next 1 day, the researcher just considered the historic close prices in order to predict the future close prices, they haven't considered using any technical indicators and social media indicators.

Gated Recurrent Unit (GRU):

Figure 2.13 shows the inner structure of the GRU, which is a type of RNN used primarily for processing sequence data, such as time series and audio signals. GRU offers several advantages over traditional RNNs, including the ability to address common vanishing gradient problems in RNNs [32]. GRU consists of an update gate and a reset gate, which are calculated as shown in equations 2.2 and 2.3, respectively.

$$U_z = sigmoid(W_z * [h_{t-1}, x_t]) \tag{2.2}$$

$$U_r = sigmoid(W_r * [h_{t-1}, x_t]) \tag{2.3}$$

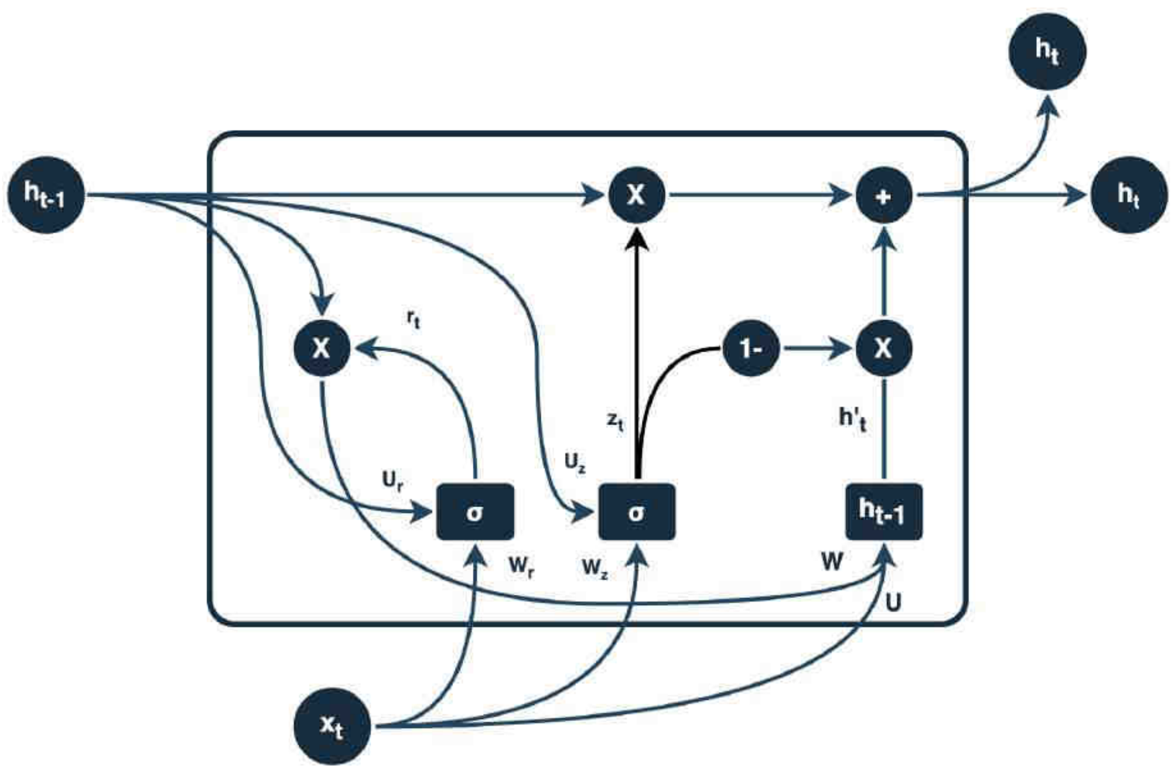

Figure 2.13: GRU Unit

Researchers have used GRU along with LSTM to predict the future close prices of different cryptocurrencies. In study [16] authors gathered historic data for Litecoin and ZCH to predict their future close prices from historic data, they also introduced a new technique to predict the future prices, they used LTC and ZCH historic data along with historic BTC data as an up/down indicator to predict the close prices of LTC and ZCH, they considered LTC and ZCH as a dependent on the BTC prices. In study [17], authors used a hybrid approach to predict the future prices of LTC and XMR for 1, 3, and 7 Days windows using historic data. They have used GRU and LSTM together in the prediction of future close prices of LTC and XMR. But neither they considered any social media indicators nor they used any technical indicators.

Bi-Directional GRU (Bi-GRU):

Bi-GRU is also a variant of RNN architecture that combines two GRU to process sequential data forward and backward. It is an extension of standard GRU to capture the context of data from the past and future. In Bi-GRU two GRU operate independently, one works in the backward direction while the second one works in the forward direction. The forward GRU works from the start of the data to the end, while the backward GRU does the reverse. Bi-GRU has a potential advantage if we compare it with the simple GRU. Bi-GRU is a powerful architecture that offers a contextual representation by considering both the past and the future.

Researchers have also used Bi-GRU in the prediction of cryptocurrency prices, the price movement of any currency depends on its historic movement along with current affairs. In study [27] researchers have used LSTM, Bi-LSTM, and GRU for the price forecast of Bitcoin, ADA, Etherium, BNB, and USDT for a 1-day window, they only used historic data to predict the future prices of respective cryptocurrencies, there a lot of factors which should be considered while predicting future close prices including social media indicators.

Activation Functions:

An activation function is a mathematical formula that removes the linearity from the output of the neuron in a neural network. It determines whether the neuron should be activated or not based on the outcome of the neuron. Selecting the correct activation function is a crucial step in solving deep learning problems, and the wrong choices can lead to unexpected results. There are a number of activation functions, but a few of them are described below.

Sigmoid Function: The sigmoid function is defined as in Equation 2.4, it converts the output value to 0 and 1, which makes it suitable for classification problems. Figure 2.14 illustrate the output of the sigmoid function.

$$(f(x) = \frac{1}{1 + e^{-x}}) \tag{2.4}$$

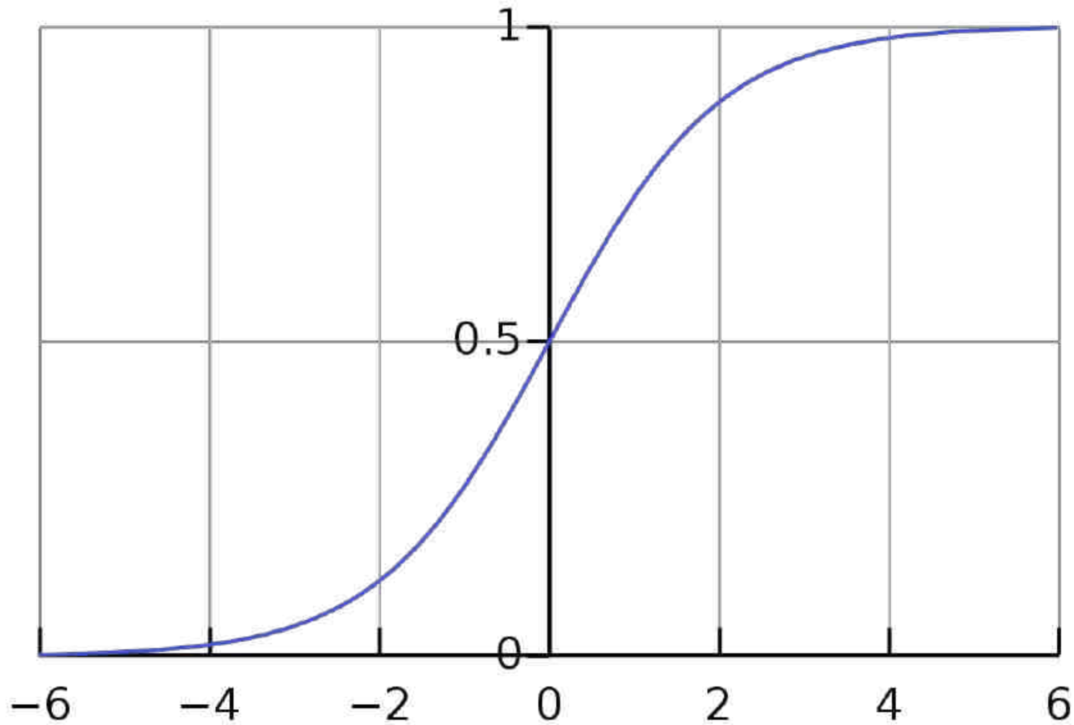

Figure 2.14: Sigmoid

Hyperbolic Tangent: Hyperbolic tangent is also known as the Tanh function, this function is defined in Equation 2.5. It converts the output of the neuron to the (-1, 1) range. Figure 2.15 illustrate the output of the sigmoid function.

$$(f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}) \tag{2.5}$$

Rectified Linear Unit (ReLU): The ReLU function is represented in Equation 2.6, ReLU returns the input value if it is possible otherwise it returns 0. In DL, ReLU stands out as the most widely used activation function as it has the ability to tackle vanishing gradient problems. Figure 2.16 describes the outcome of the ReLU activation function.

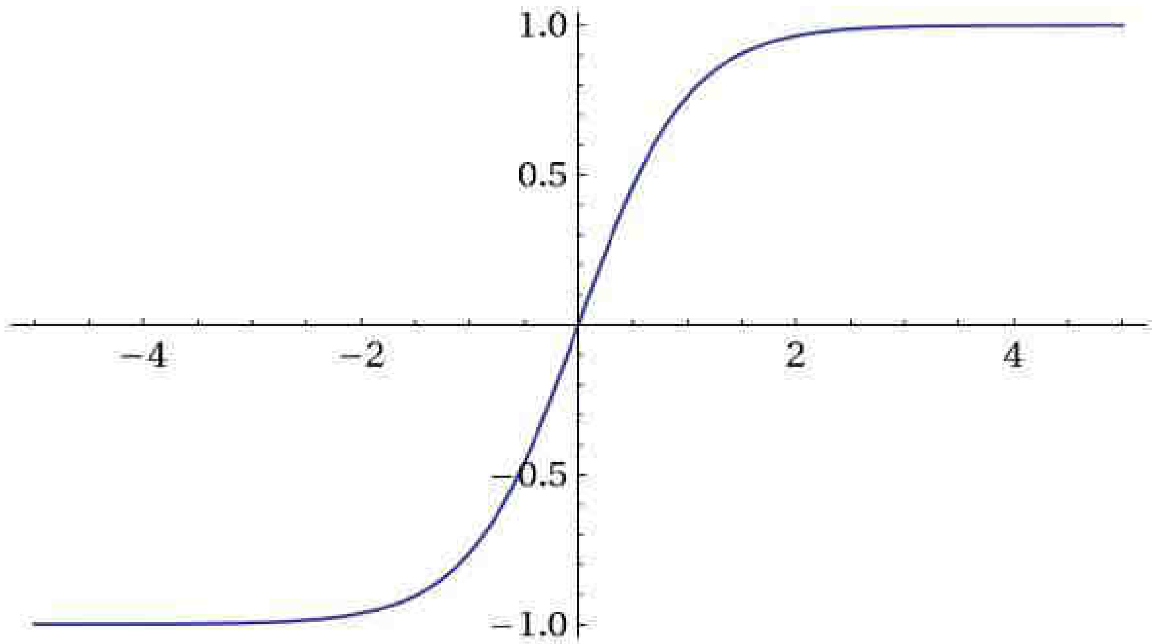

Figure 2.15: TanH

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases} \tag{2.6}$$

Linear Function: Linear function is the most simple function, it outputs the same value as the input and doesn't introduce any linearity in the data, it is defined in Equation 2.7, and it can be visualized in Figure 2.17 .

$$f(x) = x \tag{2.7}$$

The sigmoid function is commonly used in binary classification, it is mostly used in probabilistic scenarios. However, sigmoid functions have a vanishing gradient problem. On the other hand, TanH is mostly used in symmetric data, the ReLU is most widely used because of its simplicity and elimination of the vanishing gradient problem. The output of the neuron remains unchanged by the linear activation function so it is mostly used in solving regression problems.

Performance Metrics:

Performance evaluation is an important part of any research when applying machine or deep learning techniques. It allows us to infer how accurate our model is, choosing the cor-

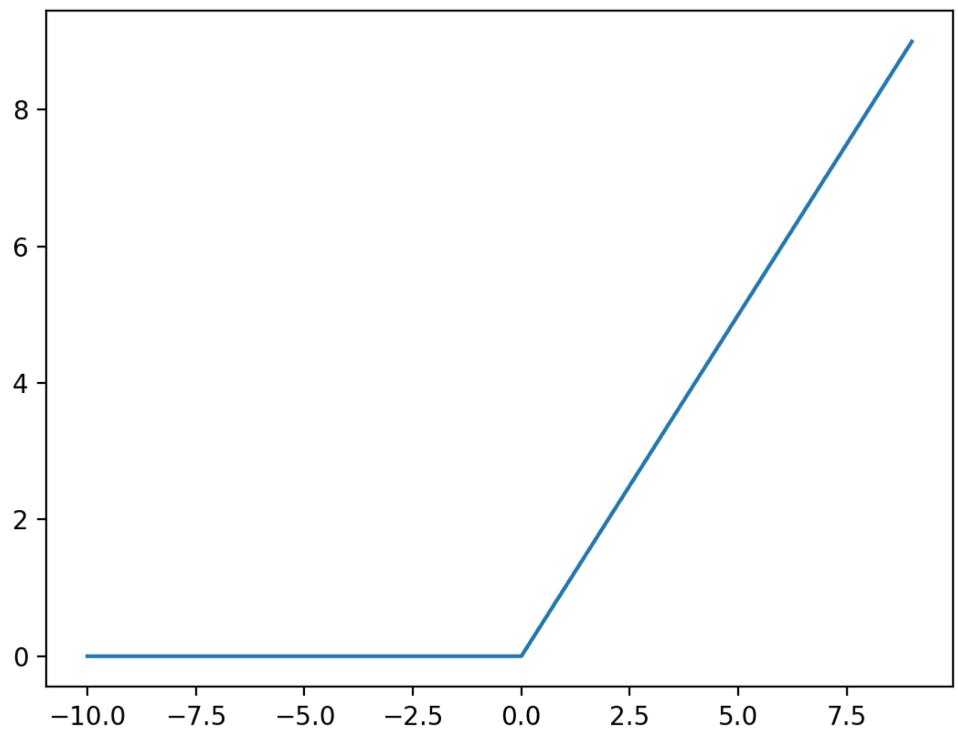

Figure 2.16: ReLU

rect metrics is the important part, and differs from study to study. There are a lot of evaluation metrics, but we are only going to use 3 metrics named MSE, RMSE, and MAE.

Mean Squared Error (MSE):

The MSE is calculated using the equation 2.8, which involves computing the average squared difference between the values generated by predictions and actual data. The MSE indicates how closely the predicted value matches the actual value.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.8}$$

Root Mean Squared Error (RMSE):

RMSE is computed using the equation 2.9, which involves applying the square root operation on the MSE. This metric provides insight into the magnitude of the difference between the values generated by predictions and actual data.

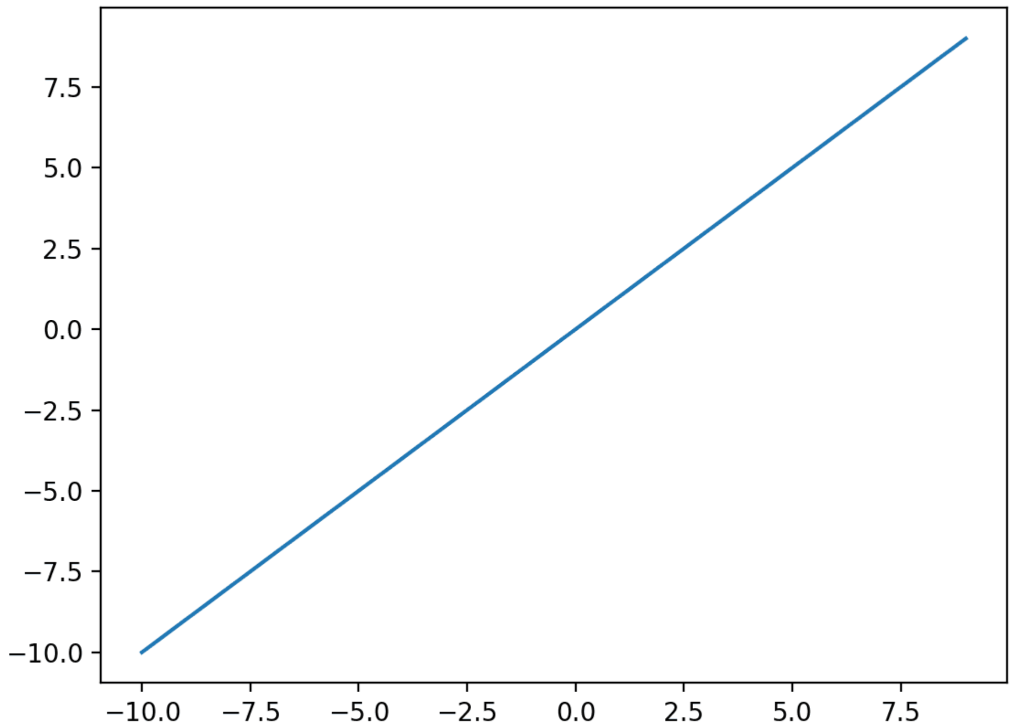

Figure 2.17: Linear Function

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.9}$$

Mean Absolute Error: MAE is calculated as specified in equation 2.10, it is calculated by taking the average absolute difference between the values generated by predictions and actual data. It is the simplest measure that tells us the average magnitude of the error made by the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.10}$$

Summary:

In this chapter, Cryptocurrency and the prediction of Cryptocurrency and their techniques have been discussed in detail along with the background study of existing state-of-the-art DL approaches, the inner structure of deep learning techniques including LSTM, GRU, and BiGRU has also been discussed along with the activation functions and their usage. Characteristics of the cryptocurrency technical data have been explored and discussed the impact of technical data on the price of the cryptocurrency as well as technical indicators for cryptocurrency trading have also been explored in detail, it is also been discussed how to use full technical indicators can while predicting the future close prices of cryptocurrencies.

Chapter 3

Data Collection & Processing

This chapter focuses on the data collection, data cleaning, and preprocessing, descriptive statistical analyses of the dataset, relevant feature selection, and then using different formulas to generate technical indicators data from historical data which will be used as features while implementing the deep learning model in the next chapter.

3.1 Dataset Required

Selecting the data is a crucial step in solving deep learning problems. The end results depend entirely on the chosen dataset. We need to gather two types of data: firstly, prehistoric cryptocurrency price data for Bitcoin Cash and DASH, and secondly, the Fear & Greed Index data. Once we have collected the prehistoric cryptocurrency data, we can process the prehistoric data to generate technical index data, this data is essential for our approach to developing a prediction model.

3.2 Data Collection

A crucial part of the deep learning problem is the data on which the proposed algorithm/technique can be applied, and as a result, the performance will be measured based on the results. Dataset selection is also a crucial step while solving deep learning problems, the results will be totally based on the data we choose. There are a lot of cryptocurrency historic datasets available in the market according to our area of interest and the nature of the problem. A research question is closely related to datasets to experiments and gets effective and accurate results. In this research we have used Yahoo Finance [36] to collect our required data using a Python library named yfinance [37]. Almost every historical cryptocurrency data is available on Yahoo Finance, but we are only interested in Bitcoin Cash and Dash historical data. The data is available in directory form in Python with UTF-8 encoding. The second part of the data which is the Fear & Greed (FGI) index is not available on Yahoo Finance, the historical FGI data is taken from a website [38], and the data is available in JSON form and it can be loaded in Python with the API exposed by the website. The data is available in UTF-8 encoding.

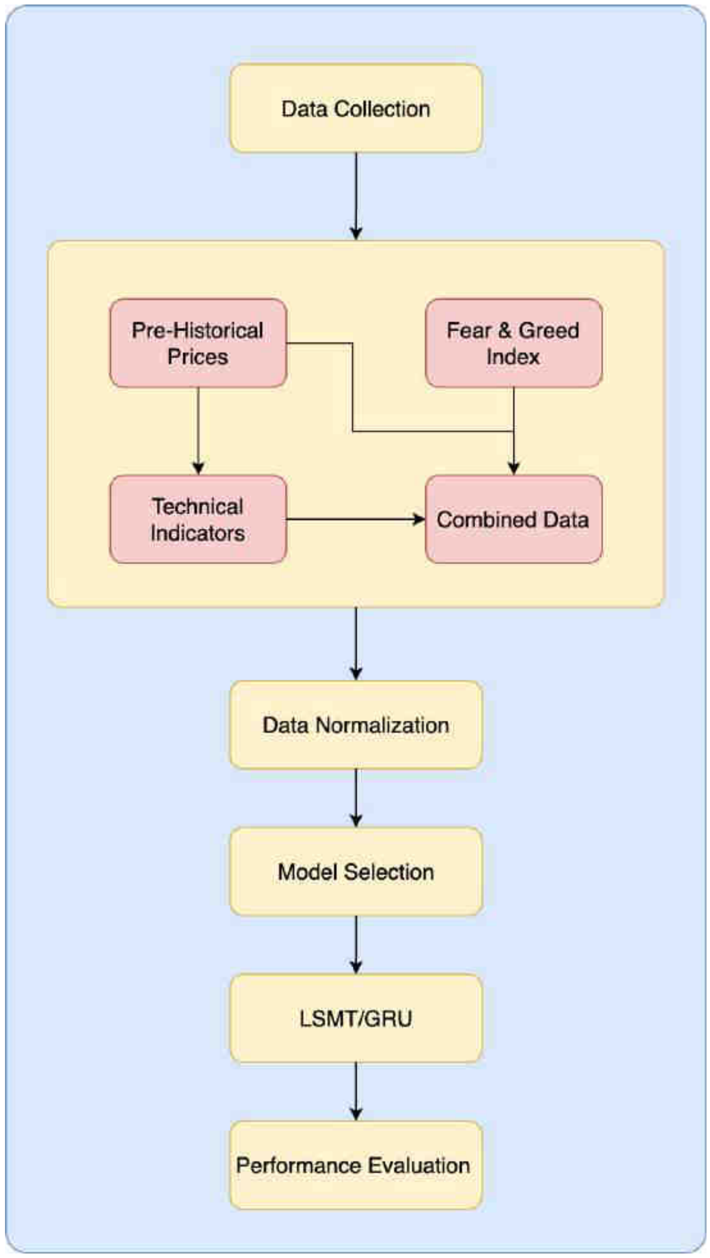

Figure 3.1: Proposed System Model

3.3 Data Cleaning & Preprocessing

The price data were collected from Yahoo Finance [37]. Yahoo Finance provides real-time data as well as historical data for daily, weekly, 15-day, and monthly periods. We collected daily Bitcoin Cash (BCH) and DASH prices from May 6th, 2018, and November 6th, 2017 to January 29th, 2023, respectively. The collected data includes the following information:

- Open: The opening price for the present day
- Close: The closing price for the present day.

- Low: The lowest price for the present day.
- High: The highest price for the present day.
- Volume: The number of units traded for the present day.

Dataset Feature	Dataset Description
DASH Prices	Historical DASH coin price data including open, high, low, close & volumes.
DASH Technical Indicators	DASH Technical Indicators including Moving Average Convergence Divergence (MACD) derived from historical prices of DASH.
BCH Prices	Historical BCH coin price data including open, high, low, close & volumes.
BCH Technical Indicators	BCH Technical Indicators including Moving Average Convergence Divergence (MACD) derived from historical prices of BCH.
Fear & Greed Index (FGI)	The FGI is calculated by combining multiple factors, including volatility, market momentum, social media sentiment, surveys, dominance, and trends, to determine the current level of fear or greed in the market.

Table 3.1: Dataset Description

3.3.1 Technical Indicators

After historical data collection, the next step is to generate Moving Average Convergence Divergence as technical indicators. Which is calculated as.

$$k = \frac{2}{N+1} \tag{3.1}$$

$$EMA_c = price_c * k + EMA_p * (1 - k) \tag{3.2}$$

In this formula, k represents the smoothing factor, which is calculated as 2 divided by the number of periods (N) plus 1. The EMA_c is the current value of the Exponential Moving Average (EMA), taking the weighted average of the current close price ($price_c$) and the previous EMA value (EMA_p), where the weight given to the current price is controlled by the smoothing factor k.

$$MACD = EMA_{12} - EMA_{26} \tag{3.3}$$

To calculate the 12 and 26-period EMAs

$$EMA_{12} = \frac{2}{13} \cdot (price_c - EMA_{11}) + EMA_{11} \tag{3.4}$$

$$EMA_{26} = \frac{2}{27} \cdot (price_c - EMA_{25}) + EMA_{25} \tag{3.5}$$

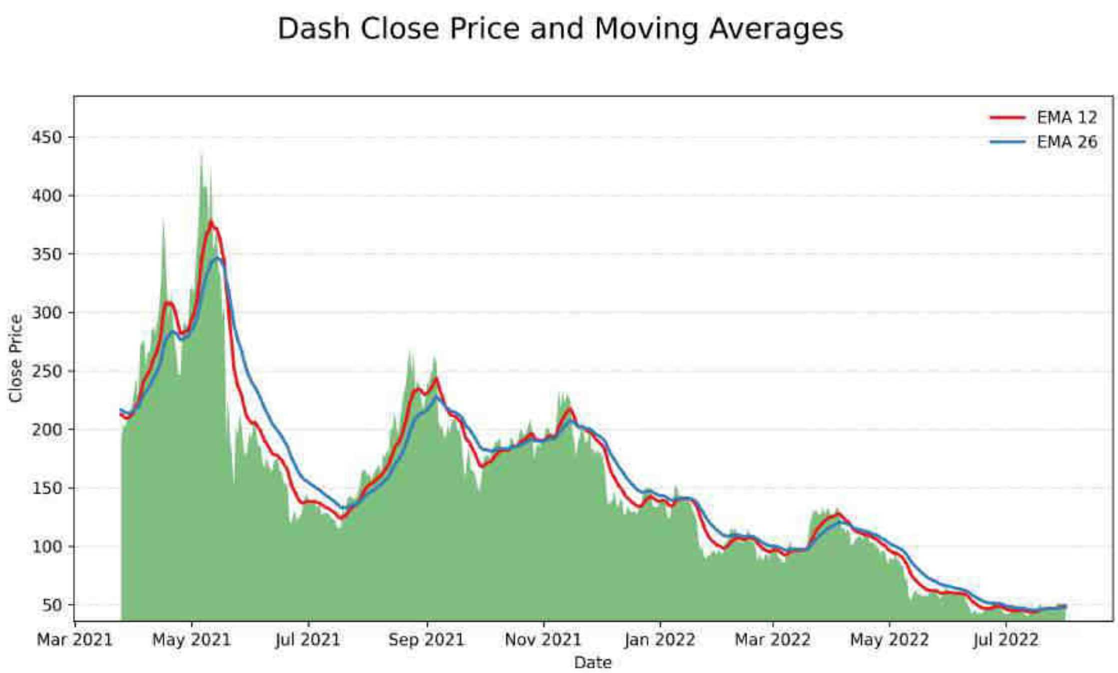

Figure 3.2: System Model

In these formulas, $price_c$ represents the closing price for the current period, and EMA_{11} and EMA_{25} are the Exponential Moving Averages (EMAs) from the previous periods (calculated using the same formulas recursively). The multipliers 2/13 and 2/27 are used to give more weight to the current period's price in the calculation of the EMAs. The result is a MACD. The Moving Average (MA) is a useful tool for indicating market trends and predicting future price movements. By analyzing the EMA of the 12 and 26 periods, we can gain insight into the strength of the trend. Figure 3.2 provides a visual representation of the EMA lines.

3.3.2 Fear and Greed Index

The FGI is a key feature for predicting cryptocurrency prices. The FGI is a score that is updated on a daily basis, calculated by taking multiple factors into consideration. We obtain our data from [38].

The FGI score is determined by the following factors:

- Volatility (25%): Calculated by comparing current and maximum volatility to the average volatility of the past 30 to 90 days.
- Market Momentum (25%): Measured by current volume and momentum in comparison with the average momentum of the past 30 to 90 days.
- Social Media Sentiment (15%): Calculated by analyzing Twitter and Reddit posts (Reddit data is still experimental) to determine overall sentiment.
- Surveys (15%): Polls are taken to gauge how people view the market (This measure is not always included).
- Dominance (10%): Market capital share of the entire cryptocurrency market.
- Trend (10%): Trends data is collected from Google related to cryptocurrency searches.

3.3.3 Dataset Generation

After gathering historical, technical, and FGI data, all the information is merged based on the corresponding date. The resulting data include open, close, low, high, volume, MACD, and FGI features. To predict prices over the next 30 days, we need to input data for the previous 30 consecutive days. Algorithm 1 describes how we generated 30 days of sequential data.

The algorithm 1 takes as input the Open, High, Low, Close, and Volume data belonging to either DASH or BCH cryptocurrency, as well as window size (W) and the Fear & Greed Index which is calculated by combining various factors. The MACD technical indicator is calculated using the equation 3.3. After calculating MACD, all of the data is merged into a single dataset based on corresponding dates. Finally, the main dataset is calculated using a window size (W) of 30 in our case. The algorithm generates new features from 0 to W, each time shifting the data back by the current index. This way, we obtain new features from Open, High, Low, Close, Volume, FGI, and MAC from 0 to the next 30 days in a single row, with $next_{close}$ serving as the class feature for the W_{ith} day.

Algorithm 1 Generate Dataset

Historic Data: $H \in [\{o, h, l, c, v, m, f, d\}]$
FGI Data: $F \in [\{v, d\}]$
Input Window: $W = 30$
Output: $O \in [\{o_0, \ldots, f_0\}, \ldots, \{o_{30}, \ldots, f_{30}\}]$

procedure GENERATEDATA(H, F, W)
  $T \leftarrow$ Calculate(MACD, H)
  $I \in [\{o, h, l, c, v, m, f\}]$
  $I \leftarrow Combine(H, T, F)_{on \rightarrow date}$
  $O \leftarrow$ Empty(Dataset)
  for $ind \in [0, W]$ **do**
    $O \leftarrow$ new_col(o_{ind}, shift(I_o, $-ind$))
    $O \leftarrow$ new_col(h_{ind}, shift(I_h, $-ind$))
    $O \leftarrow$ new_col(l_{ind}, shift(I_l, $-ind$))
    $O \leftarrow$ new_col(c_{ind}, shift(I_c, $-ind$))
    $O \leftarrow$ new_col(v_{ind}, shift(I_v, $-ind$))
    $O \leftarrow$ new_col(m_{ind}, shift(I_m, $-ind$))
    $O \leftarrow$ new_col(f_{ind}, shift(I_f, $-ind$))
    $O \leftarrow$ new_col(nc_{ind}, shift(I_c, $-ind - W$))
  return O

For the simple model, there were a total of 1786×150 and 1695×150 input data points for BCH and DASH respectively. For the Technical & FGI models, there were 1786×180 and 1695×180 data points for DASH and BCH respectively.

3.4 Descriptive Statistical Analyses

We have performed descriptive statistical analyses on the historical prices and Fear & Greed Index to get some meaningful insight from the data and to refine our research questions and follow a better direction. Exploratory data analyses have been done before the development of the actual prediction model to have a good understanding of the data before.

Figure 3.7 shows the correlation between BCH Closes prices and the Fear & Greed index. The line plot shows a comprehensive overview of the BCH Close prices movement while showing the Fear & Greed Index trends simultaneously. The line for BCH Close price displays its fluctuations over time, illustrating significant peaks and valleys that reflect market dynamics. The FNG line reveals the sentiment trends over time surrounding the price of BCH, helping to have an overall idea of the market sentiment and investor behavior. It is found that the price of the BCH price increased when the sentiment score increased.

Figure 3.8 show the correlation between the DASH price and the Fear & Greed index. It is

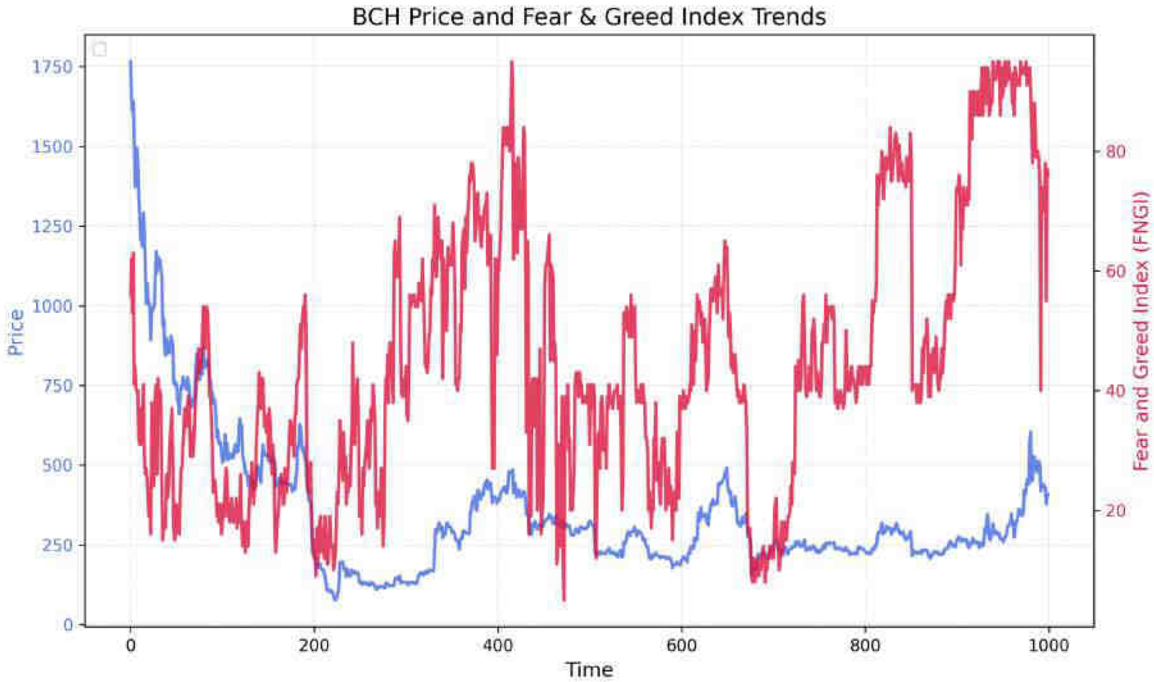

Figure 3.3: Correlation of BCH Price and Fear & Greed Index

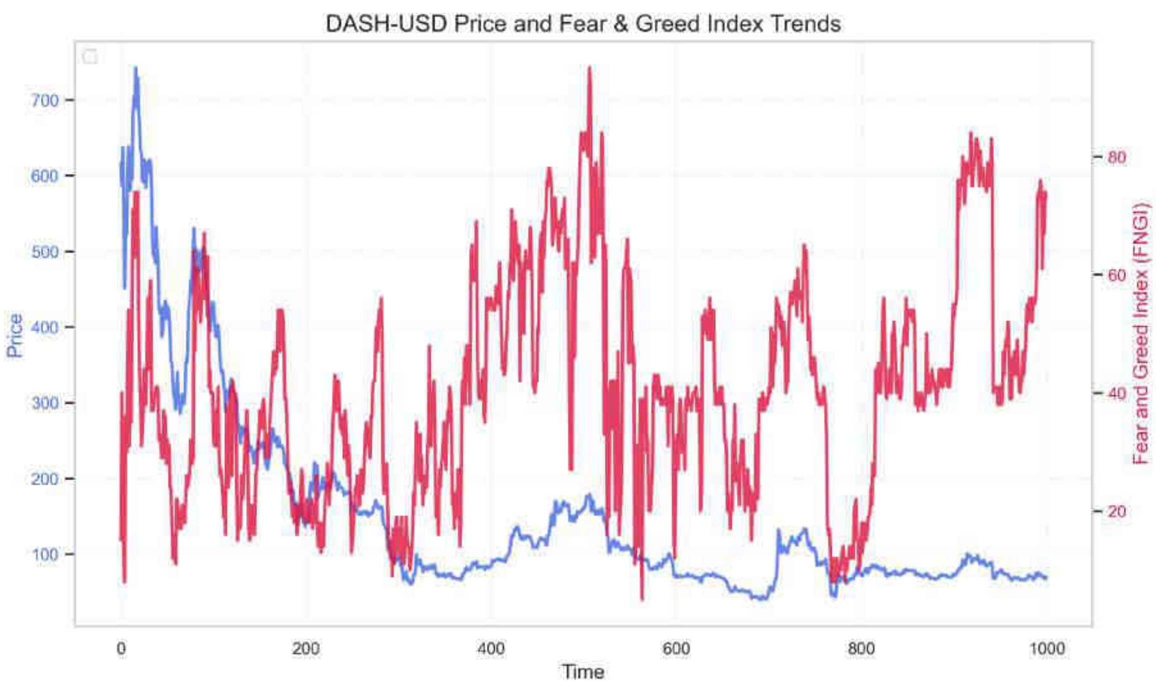

Figure 3.4: Correlation of Dash Price and Fear & Greed Index

seen that the price and sentiment have positive relationships.

Figure 3.5 is a scatter plot with a trend line to examine the relationship between BCH price and FNGI sentiment. The X-Axis of the plot shows the sentiment score of the FNGI which is ranging from 0 to 100 and while the Y-axis displays the BCH price. The scatter plot allows us to observe the distribution of the data and to find the potential pattern of the data.

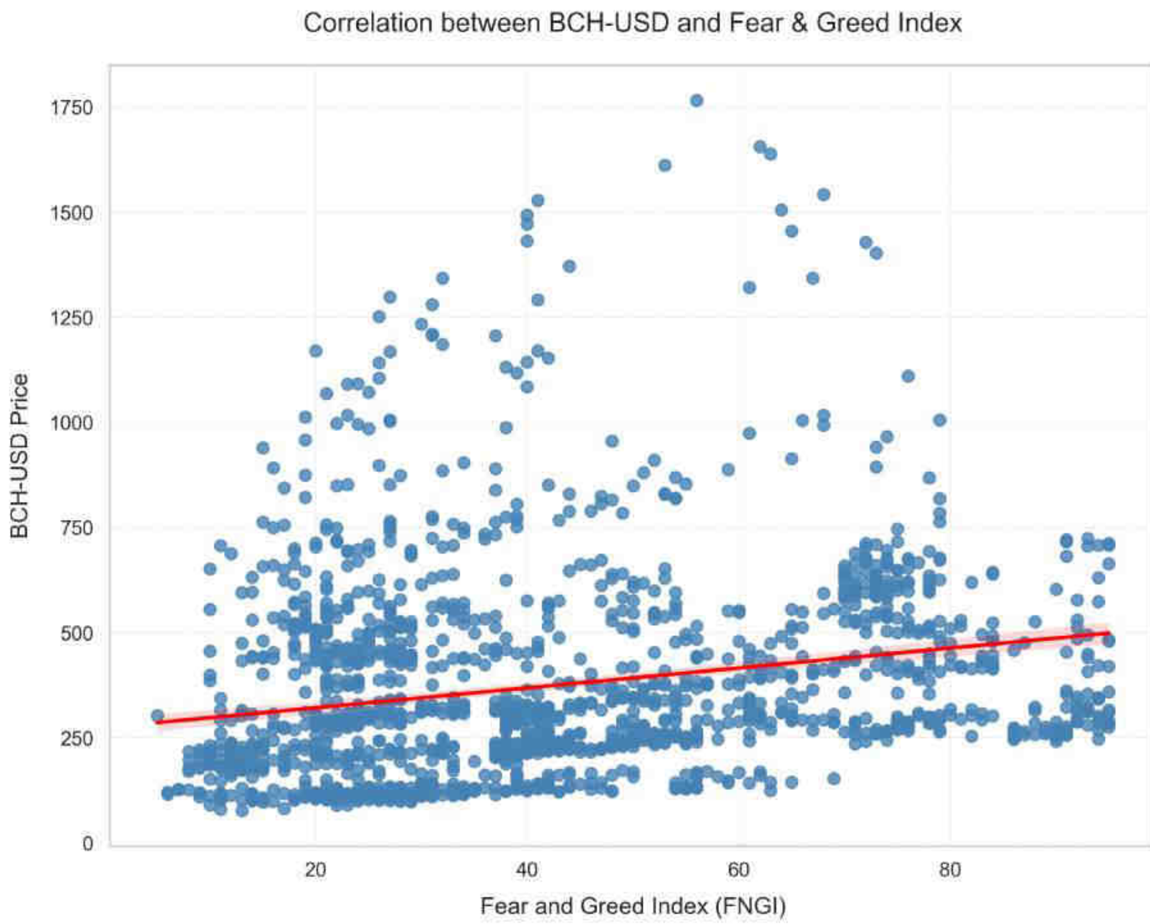

Figure 3.5: Scatter Correlation of BCH Price and Fear & Greed Index

The trend line helps to visually capture the direction and strength of the relationship between the FNGI and BCH price. This upward trend suggests that as the FNGI sentiment increases, there is a corresponding increase in the BCH price.

Figure 3.6 is a scatter plot with a trendline to examine the relationship between DASH Price and FNGI Sentiment. Same as BCH, the upward trend lines show that the FNGI Sentiment and DASH price has a positive relationship.

Figure 3.7 shows the correlation analyses of various factors including BCH Price, trading volume, Sentiment (FNGI), and MACD (Moving Average Convergence Divergence). The correlation coefficients were computed to quantify the strength and direction of these relationships. The correlation between Volume and MACD is 0.38 which indicates that they have a strong relationship. The sentiment and MACD correlation value is 0.54 which indicates a very strong relationship, it suggests that changes in sentiments tend to align with the changes in market trends as indicated by the MACD.

Figure 3.8 shows the correlation analyses of Dash price, volume, FNGI, and MACD. The

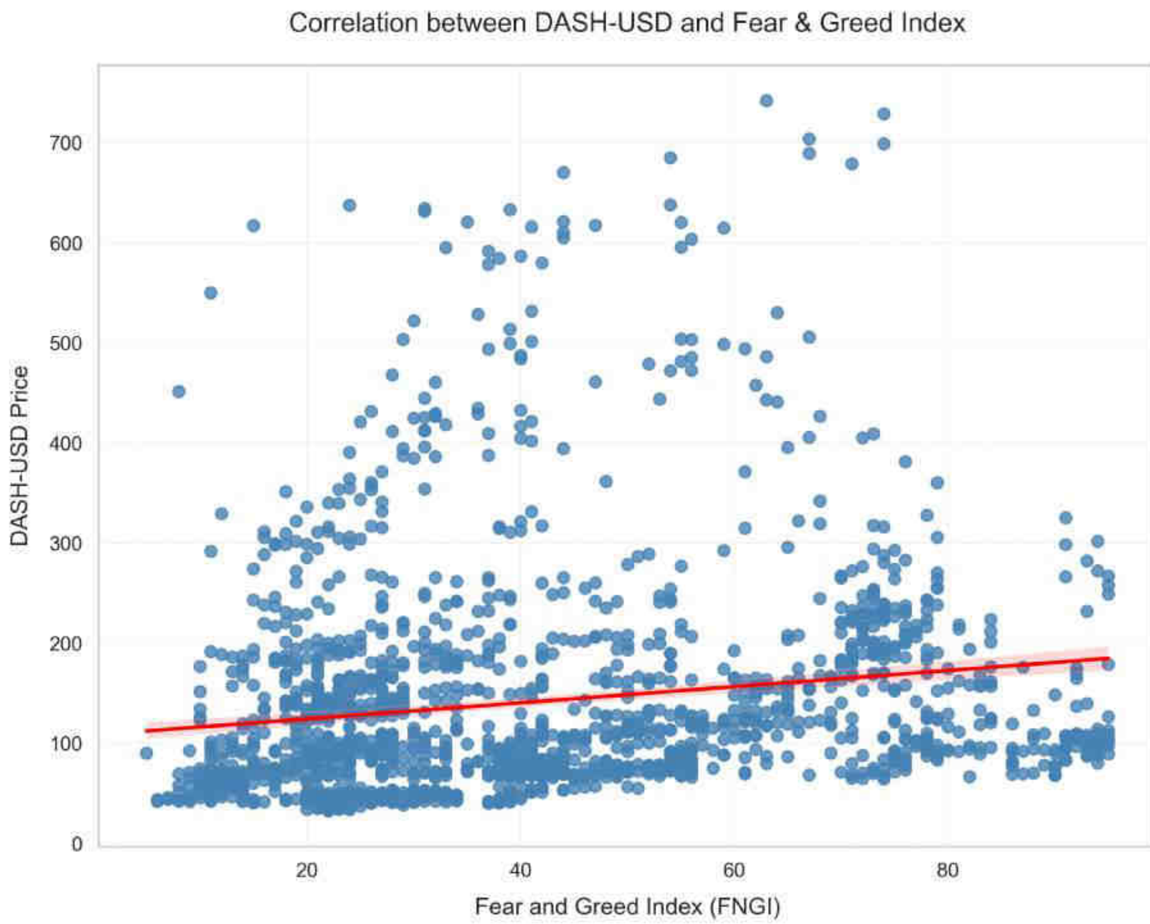

Figure 3.6: Scatter Correlation of Dash Price and Fear & Greed Index

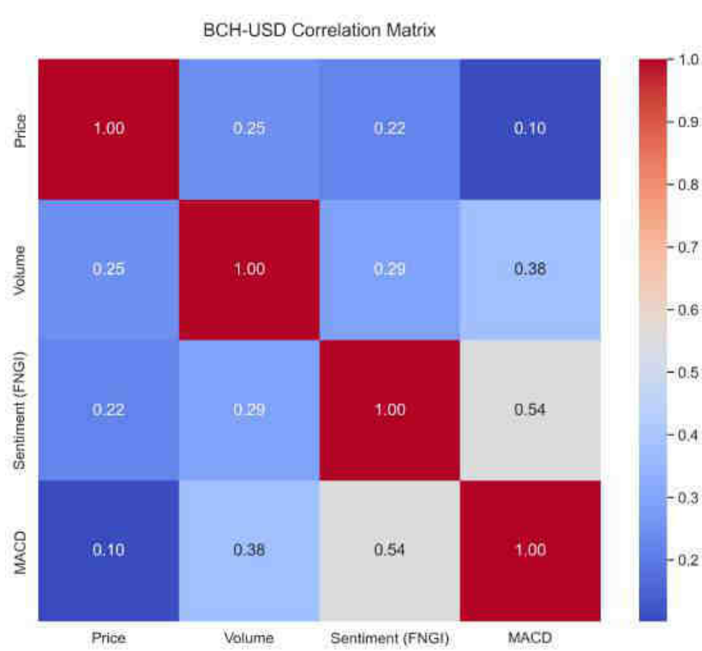

Figure 3.7: Correlation of BCH Prices, Sentiments, Volume, and Technical Indicators

sentiment and MACD show a strong relationship, the same as the BCH correlation it shows that changes in FNGI sentiments tend to align with the changes in market trends as indi-

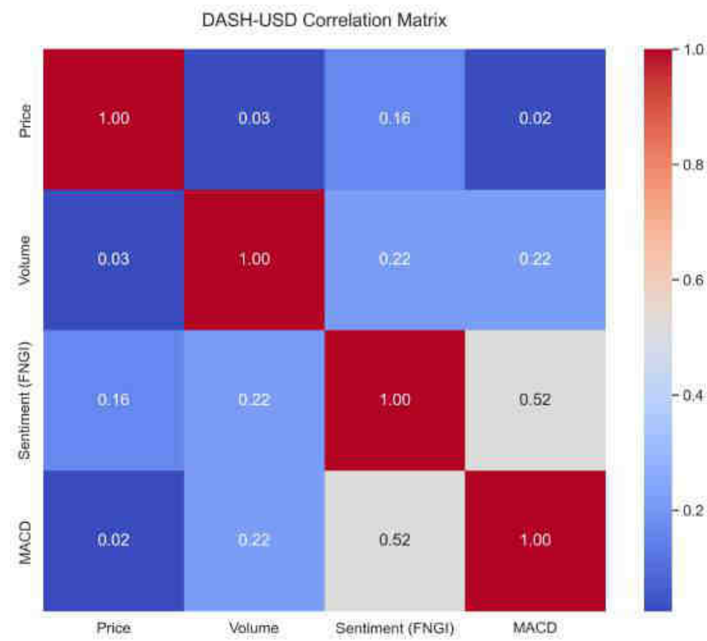

Figure 3.8: Correlation of Dash Prices, Sentiments, Volume, and Technical Indicators

cated by MACD.

3.5 Data Normalization

Feature selection plays a crucial and vital role in DL problems. After feature selection, the following step would be to perform data normalization, as the data values may be high in range as compared to other features which may cause inappropriate results. The prices of DASH and BCH are relatively high compared to their FGI and MACD values. Such a significant difference in values may lead to unexpected outcomes. To tackle this problem we have different scalers which are described below.

3.5.1 Min Max Scaler

The min scaler is the most commonly used scaler in machine and deep learning problems. The min-max scaler transforms the data into a specific range like 0 to 1.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \cdot [1 - 0] \quad \text{where } x_{new} \in [0, 1] \tag{3.6}$$

3.5.2 Standard Scaler

The standard scaler is also a popular and most commonly used scaler in machine and deep learning problems. The Standard Scaler makes a dataset by transforming each feature in the dataset, such that it has an average of 0 and a standard deviation of 1.

$$z = \frac{x - \mu}{\sigma} \quad \text{where} \quad z \in (-\infty, +\infty) \tag{3.7}$$

The Equation 3.7 is the calculation of the Standard Scaler. In this equation, z denotes the standardized value, x represents the original value, μ indicates the feature's mean, while σ represents the standard deviation of the feature.

To ensure that all features are within the same range, we utilized a min-max scaler with a range of 0 to 1 to scale the values of all features. Equation 3.6 illustrates how we can normalize the values.

After the data normalization, the data will be split into test and train data. Train data is where the model will be trained, the model can have amazing results if we have a good amount of data. Usually, there will be a split of 70/30 for train and testing. But in our case, we used an 80/20 Split, furthermore, the train data will have a split of 80/20 for the validation purpose which will be carried out during the model training.

Summary:

In this chapter, we have completed the collection of pre-historic cryptocurrency data. After obtaining the historical data, we derived the technical indicators from it. Following the computation of the technical indicators, we collected the Fear Greed Index data. Subsequently, the historical, technical, and Fear Greed Index data were combined together to make a final dataset. To have better understanding of the data, we selected specific features and conducted in-depth analyses. Additionally, we generated various graphs and visualizations to understand the relation between features. Afterward, the data was normalized to ensure that all features are within the same range, enabling our model to learn from the data more effectively.

Chapter 4

Price Prediction using Social Media Indicator

This chapter is dedicated to creating a forecasting model for cryptocurrencies using prehistoric, technical, and social media indicators using state-of-the-art DL approaches. We have trained the model for GRU using the normalized data using TensorFlow [31]. After training, the model is evaluated using test data for MSE, RMSE, and MAE Matrics. After evaluation, the model is also compared with the existing state-of-the-art deep learning approaches and it was found that our model outperforms.

4.1 Problem Statement

Cryptocurrency has become an emerging way of trading and making transactions for digital goods. Despite this, the price fluctuations of cryptocurrencies create difficulties for a trader to have an idea of price changes in the future. ML and DL techniques have been used by researchers to build different models for predicting cryptocurrency prices using historic data, new's sentiments, technical indicators, and Twitter sentiments, but they didn't consider using Social Media Indicators, Technical Indicators, and historic prices in a single model for price prediction, this lack of feature integration hampers the effectiveness of cryptocurrency price prediction.

4.2 Proposed Solution

Cryptocurrency prices are highly volatile, and one of the main reasons for huge price changes is what people talk on social media about particular cryptocurrencies. To better predict the price of cryptocurrency we are going to include social media indicators such as the Fear & Greed index, Technical indicators Moving Average Convergence Divergence Line, and Historic Price.

4.2.1 Selecting the Optimal Deep Learning Algorithm

This section includes the detail of the DL algorithm selection for our prediction problem. Researchers have used different methods to forecast the future close prices of cryptocurrencies. Our data is time series data, and there are a lot of DL algorithms that can handle this problem. We selected Gated Recurrent Unit (GRU) which is prominent for our use case.

4.2.2 Experimental Setup: GRU Model Training and Evaluation

In this section, we will discuss the model building, training, and evaluation process in detail. Our chosen model for price prediction has been trained using the GRU algorithm, and we have selected this approach after studying other state-of-the-art DL approaches. Researchers and data scientists have also used RNN, LSTM, and other ML algorithms to predict the prices of cryptocurrency.

In our specific implementation of the GRU model we have used TensorFlow [31] by Google. As for the programming language Python 3.11 is used, along with Jupyter Notebook. Jupyter Notebook is very handy, it allowed us to use Interactive Python. The model was trained on a Mac M1 system with 100 epochs with a batch size of 32 and using the Adam optimizer.

The deep learning model tuned and optimized on these parameters is shown in Table 4.1.

Parameter	Value
batch_size	32
Optimization Algorithm	adam
learning_rate	0.001
validation_split	0.2
epochs	100

Table 4.1: Parameters for GRU Model

4.2.3 Neural Network Architecture

The architecture of the proposed neural network model for predicting cryptocurrency prices is shown in Figure 4.1. The model begins with an input layer, followed by a Bidirectional GRU (BiGRU) layer with 128 units. In order to solve the issue of overfitting, a dropout layer is used with a dropout rate of 0.2. The dropout layer randomly selects inputs and sets them to zero based on the probability outlined in equation 4.1.

$$prob_{drop} = 1/(1 - rate) \tag{4.1}$$

After the dropout layer, the model continues with a bidirectional GRU layer comprising 128 units, which are followed by an additional dropout layer with a dropout rate of 0.2. The subsequent dense layer has 32 units and uses the ReLU activation function. Finally,

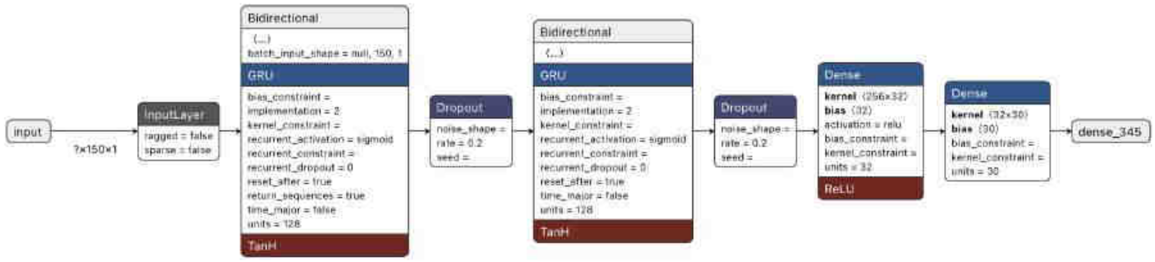

Figure 4.1: Model

the model adds a dense layer with S units, where S corresponds to the window size of the cryptocurrency price prediction, using the linear activation function shown in Equation 2.7. Figure 4.2 illustrate the steps of how the research was conducted.

4.3 Performance Evaluation

In this section, the performance of the trained model is evaluated in detail and can be visualized graphically as well.

4.3.1 Evaluation Metric

The performance of our model was measured using metrics such as MSE, RMSE, and MAE. They are the simplest and most commonly used metrics.

4.3.2 BCH Results and Analysis

There are three different models developed for predicting the price of BCH. The first model, known as the "simple" model, includes features such as $open_0$, $high_0$, low_0, $close_0$, and $volume_0$, up to $open_30$, $high_30$, low_30, $close_30$, and $volume_30$ within the window size. The second model, called "with_macd", is an enhanced version of the simple model that includes all of the basic features and adds the MACD indicator, denoted as all_simple_0 to all_simple_{30}, and $macd_0$ to $macd_{30}$ within the window size. The third model, "with_fngi", uses all of the basic features from the simple model but adds the Fear & Greed Index (FGI) feature, denoted as all_simple_0 to all_simple_{30}, and fgi_0 to fgi_{30} within the window size.

To evaluate the performance of the BCH model we have used different error measures, which provide valuable insights into its accuracy. Figure 4.3 presents a comparison of three measures MSE, RMSE, and MAE used for the calculation of losses.

The simple model for predicting the future prices of BCH Coin only used historical prices

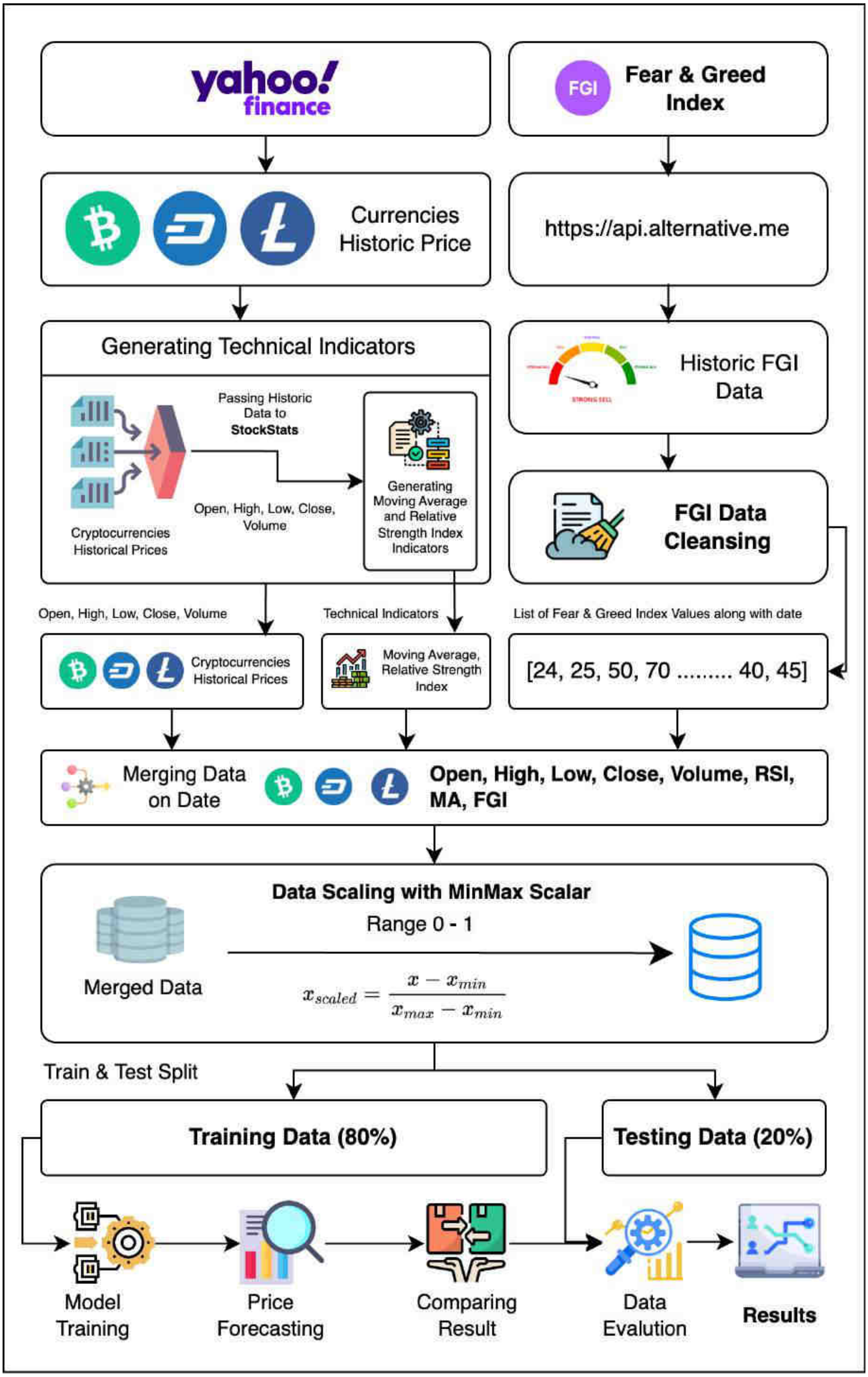

Figure 4.2: System Model

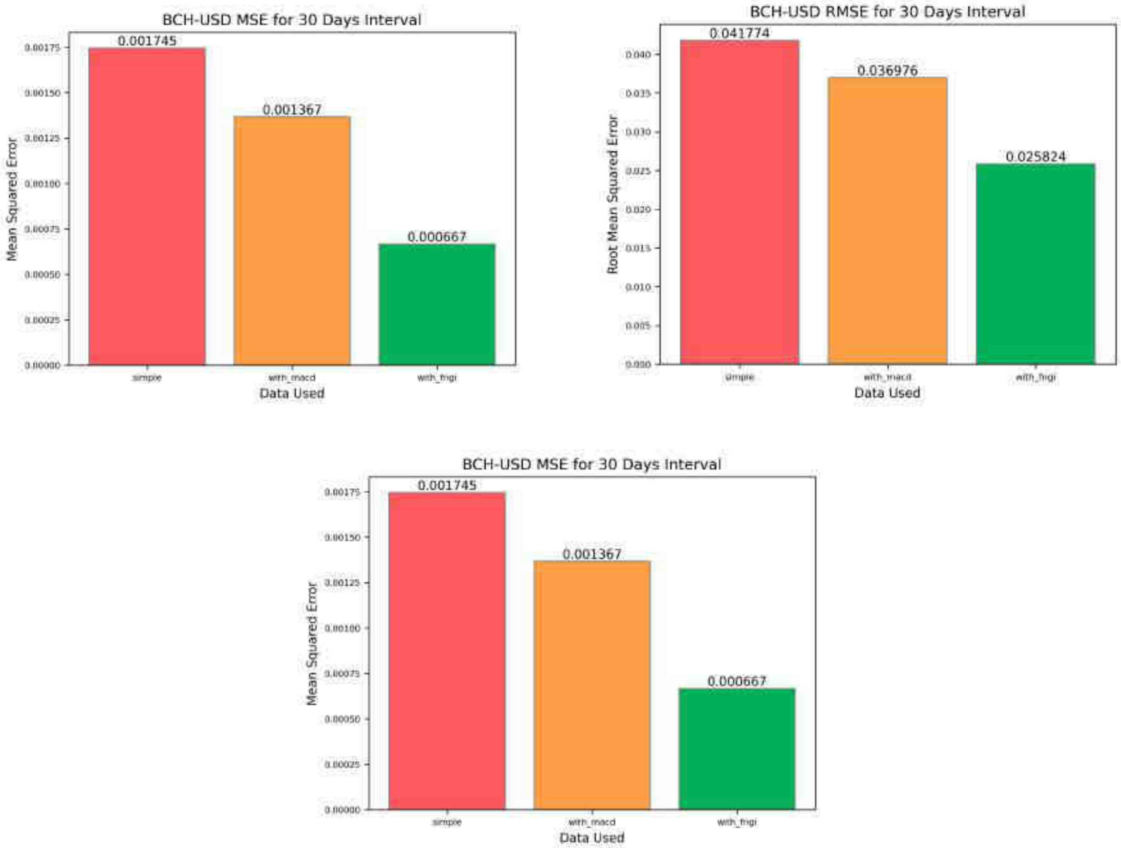

Figure 4.3: Comparative Analysis of Bitcoin Cash (BCH) Prediction Error

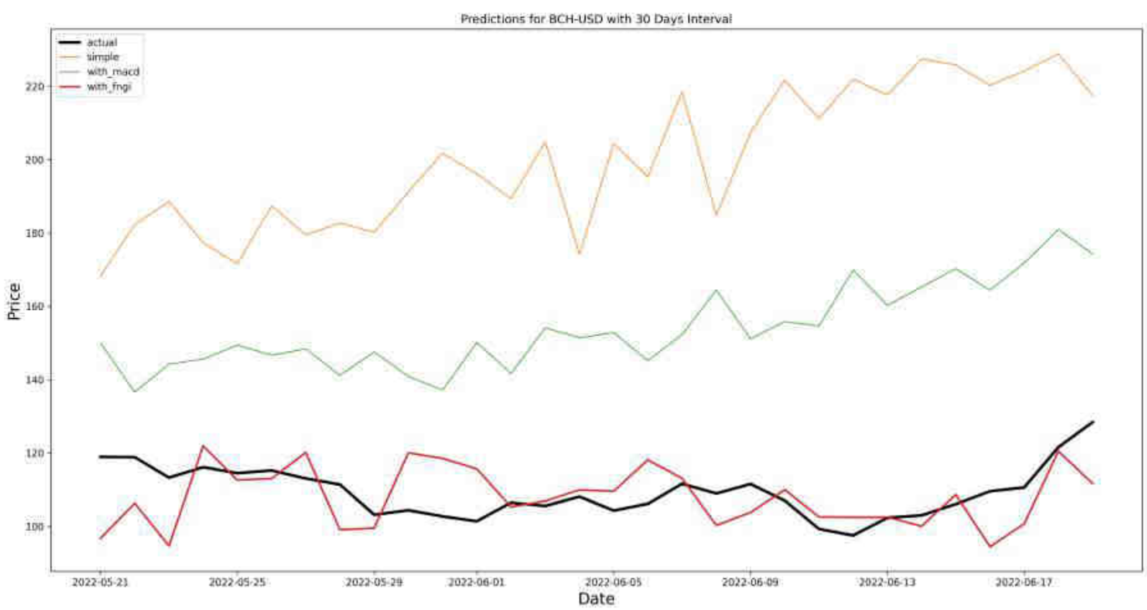

Figure 4.4: BCH Prediction

and volume as input, resulting in lost values of 0.0017, 0.0417, and 0.0340 for MSE, RMSE, and MAE, respectively. To improve performance, technical indicators MACD was included, which leads to lower lost values of 0.0013, 0.0369, and 0.0304 for MSE, RMSE, and MAE, respectively. However, after the addition of the Fear & Greed index to the initial model, its performance exceeded that of the other models, resulting in the lowest lost values of 0.0006, 0.0258, and 0.0201 for MSE, RMSE, and MAE, respectively, as shown in Figure 4.3 also described in Table 4.2. The models successfully predicted the future prices of BCH Coin for the next 30 days, as shown in Figure 4.4 which compares the prediction results from all models with the actual values.

Model	**Error**		
	MSE	RMSE	MAE
simple	0.00174510	0.04177444	0.03401276
with_macd	0.00136720	0.03697569	0.03049348
with_fngi	0.00066689	0.02582422	0.02017940

Table 4.2: BCH Prediction Error

4.3.3 Model 2: DASH Price Prediction

The Dash price prediction follows the same methodology as the BCH price prediction. There are three different models used for Dash price prediction. The first model is simple and considers only the variables $open_0$, $high_0$, low_0, $close_0$, and $volume_0$, up to $open_{30}$, $high_{30}$, low_{30}, $close_{30}$, and $volume_{30}$ within a given window size. The second model incorporates technical indicators, specifically the MACD, along with the variables from the simple model, such as all_simple_0 to all_simple_{30} and $macd_0$ to $macd_{30}$. The final model, with_fngi, includes all the features from the simple model as well as the Fear and Greed Index value, resulting, in the end, features all_simple_0 to all_simple_{30} and fgi_0 to fgi_{30} within the window size.

The performance of the three models was evaluated using the MSE, RMSE, and MAE as specified in Equations 2.8, 2.9, and 2.10, respectively.

Similar to the BCH Price Prediction model, the Dash model incorporating the Fear and Greed Index outperformed the other models. The simple model, which uses only historical prices and volume as inputs, resulted in lost values of 0.0040, 0.0638, and 0.0511 for MSE, RMSE, and MAE, respectively. The second model improved on this performance by

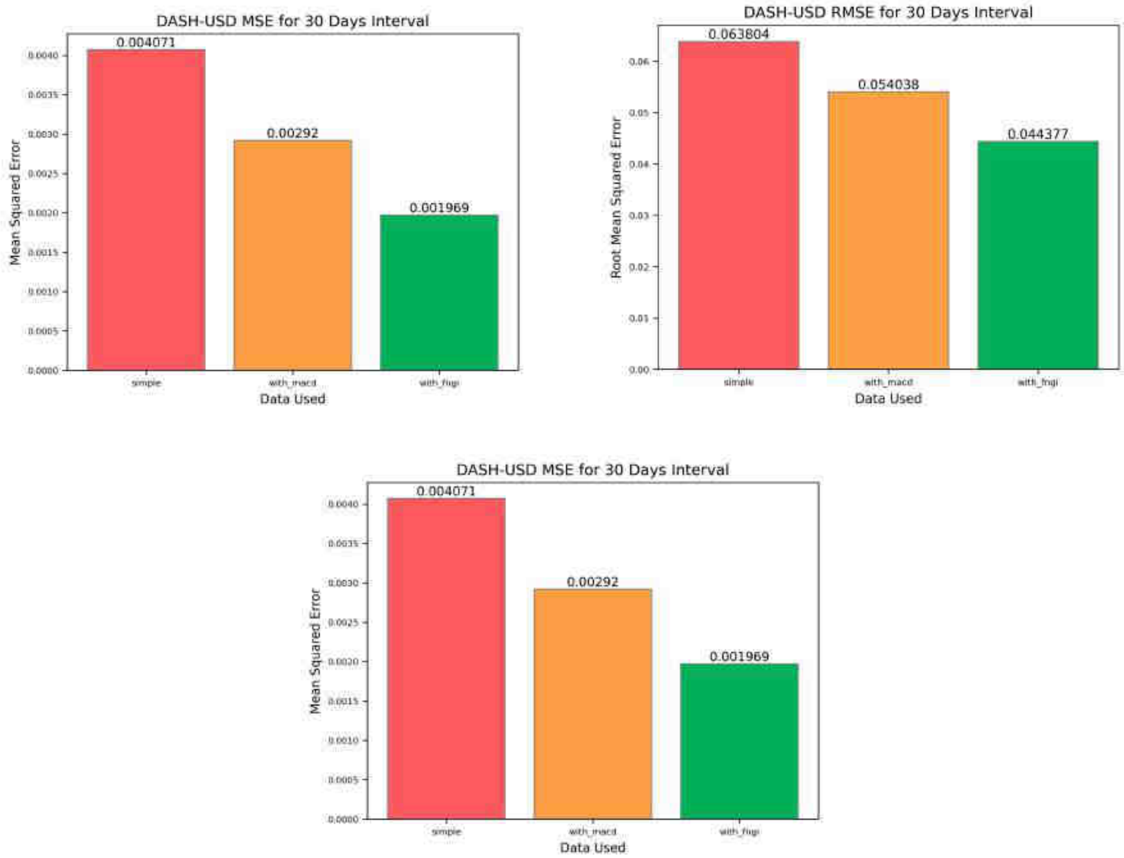

Figure 4.5: Comparative Analysis of Dash Coin Prediction Error

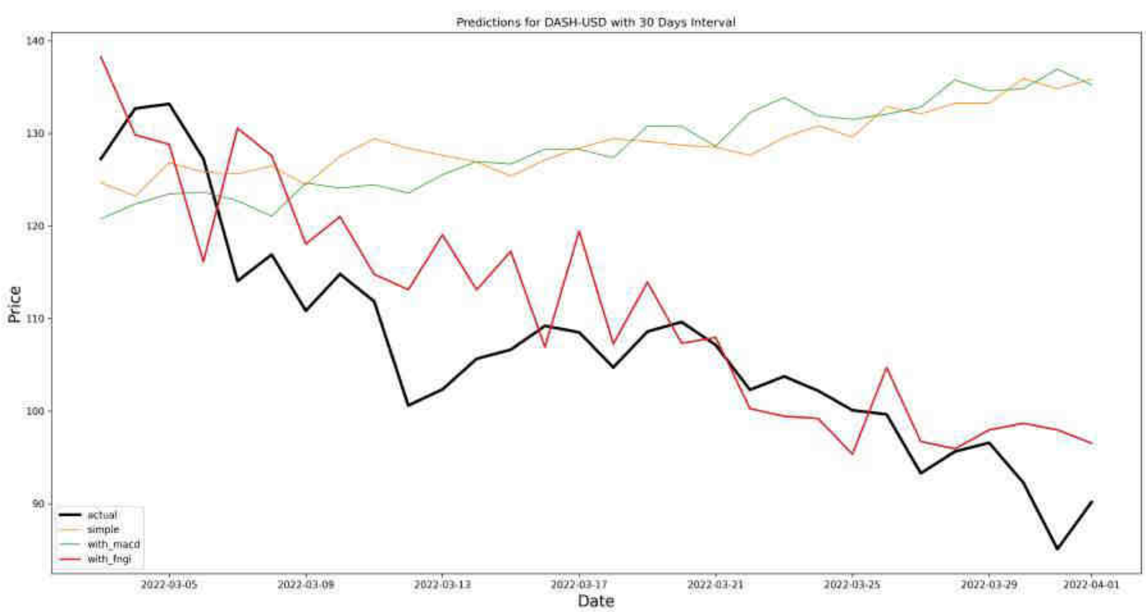

Figure 4.6: DASH Prediction

Model	Error		
	MSE	RMSE	MAE
simple	0.00407090	0.06380359	0.05115422
with_macd	0.00292007	0.05403766	0.04338968
with_fngi	0.00196933	0.04437718	0.03806518

Table 4.3: Dash Prediction Error

incorporating technical indicators and resulted in lower lost values of 0.0029, 0.0540, and 0.04338 for MSE, RMSE, and MAE, respectively. The last model, which incorporated the Fear and Greed Index with the simple model's features, resulted in the lowest lost values of 0.0019, 0.0443, and 0.0380 for MSE, RMSE, and MAE, respectively. A comparison of the lost values is shown in Figure 4.5 and summarized in Table 4.3. All of these models successfully predicted the future prices for the next 30 days of Dash coin from the previous 30 days data. Figure 4.6 shows a comparison of the actual values and predictions from these three models.

Summary:

In this chapter, we have completed the training and testing of the model. After studying many types of research, and trying out different algorithms we proposed this approach which ends up providing the best results. We have completed our experiments, and performance evaluation after the training of the model. The performance was evaluated based on MSE, RMSE, and MAE matrics. After evaluation of the performance we also tried to predict the future close prices of BCH and Dash coins and include the visualization for that. BCH ends up with 0.0017, 0.0013, and 0.00066 MSE for simple, with_macd and with_fngi respectively. Dash ends up with 0.004070, 0.0029. 0.0019 MSE for simple, with_macd and with_fngi respectively. Our approach which is with_fngi outperformed the simple and with_macd.

Chapter 5

CONCLUSION AND FUTURE DIRECTIONS

This consists of the research conclusion along with suggestions for future directions to enhance the performance of the existing model.

5.1 Conclusion

In this research paper, we studied the existing approaches for predicting crypto prices. Cryptocurrency prices are highly volatile and depend on multiple factors, including social media activity. We built a deep neural model that can predict the prices of BCH and DASH coins by incorporating both technical and social media indicators for future price prediction. To evaluate the model's performance, we predicted the prices of BCH and DASH coins and compared the results using a loss function with existing models. Our results showed that our approach outperformed existing methods for predicting crypto prices.

5.2 Future Directions

Include News Impact

In this research paper, we have focused on social media sentiments, which are primarily based on what people say on social media platforms. However, incorporating news sentiment may provide additional insights and positively impact price prediction accuracy.

Consider Economic Factors

The countries with the highest number of cryptocurrency traders are likely to be impacted by their economic conditions. Factors such as inflation, interest rates, and government policies regarding cryptocurrencies can have a significant impact on cryptocurrency prices.

Consider Other Cryptocurrencies

We have only considered applying this technique to DASH and BCH, there are a lot of other currencies in the market that have more consumer base. Considering other currencies will help investors in making decisions.

Considering other Markets

This technique we have used in predicting future close prices of cryptocurrencies can be used to predict other stock markets such as PSX and NYSX.

BIBLIOGRAPHY

[1] NextAdvisor. (2022, October 31). Bitcoin Price History: 2009 to 2022. NextAdvisor With TIME. https://time.com/nextadvisor/investing/cryptocurrency/bitcoin-price-history/

[2] CoinGecko. (n.d.). Cryptocurrency Prices, Charts, and Crypto Market Cap. https://www.coingecko.com/

[3] Crypto Market Cap Charts. (n.d.). CoinGecko. https://www.coingecko.com/en/global-charts

[4] Wikipedia Contributors, "Ethereum," Wikipedia, Apr. 10, 2019. https://en.wikipedia.org/wiki/Ethereum

[5] How cryptocurrency price moves in the market? (2022) Appinventiv. Available at: https://appinventiv.com/blog/understanding-cryptocurrency-price-shift-in-market/ (Accessed: 14 May 2023).

[6] Manning, J. (2022) The role of social media in crypto prices: Mawson, Mawson Infrastructure Group. Available at: https://mawsoninc.com/role-of-social-media-in-crypto-prices/ (Accessed: 14 May 2023).

[7] V. Derbentsev, N. Datsenko, O. Stepanenko, and V. Bezkorovainyi, "Forecasting cryptocurrency prices time series using machine learning approach," SHS Web of Conferences, vol. 65, p. 02001, 2019.

[8] Parekh, R. et al. (2022) "DL-guess: Deep learning and sentiment analysis-based cryptocurrency price prediction," IEEE Access, 10, pp. 35398–35409. Available at: https://doi.org/10.1109/access.2022.3163305.

[9] Liu, Y. and Tsyvinski, A. (2020) "Risks and returns of cryptocurrency," The Review of Financial Studies, 34(6), pp. 2689–2727. Available at: https://doi.org/10.1093/rfs/hhaa113.

[10] Corbet, S. et al. (2020) "The destabilising effects of cryptocurrency cybercriminality," Economics Letters, 191, p. 108741. Available at: https://doi.org/10.1016/j.econlet.2019.108741.

[11] Ante, L. (2023) "How elon musk's twitter activity moves cryptocurrency markets," Technological Forecasting and Social Change, 186, p. 122112. Available at: https://doi.org/10.1016/j.techfore.2022.122112.

[12] Malwa, S. (2023) Dogecoin-like spike in Milady nfts after Elon Musk's tweet, but will it last?, CoinDesk Latest Headlines RSS. Available at: https://www.coindesk.com/markets/2023/05/12/dogecoin-like-spike-in-milady-nfts-after-elon-musks-tweet-but-will-it-last/ (Accessed: 06 July 2023).

[13] Jay, P. et al. (2020) "Stochastic Neural Networks for cryptocurrency price prediction," IEEE Access, 8, pp. 82804–82818. Available at: https://doi.org/10.1109/access.2020.2990659.

[14] Fleischer, J.P. et al. (2022) “Time series analysis of cryptocurrency prices using long short-term memory,” Algorithms, 15(7), p. 230. Available at: https://doi.org/10.3390/a15070230.

[15] Vo, A.-D. (2019) “Sentiment analysis of news for effective cryptocurrency price prediction,” International Journal of Knowledge Engineering, 5(2), pp. 47–52. Available at: https://doi.org/10.18178/ijke.2019.5.2.116.

[16] Tanwar, S. et al. (2021) ‘Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations’, IEEE Access, 9, pp. 138633–138646. doi:10.1109/access.2021.3117848.

[17] Patel, M.M. et al. (2020) ‘A deep learning-based cryptocurrency price prediction scheme for Financial Institutions’, Journal of Information Security and Applications, 55, p. 102583. doi:10.1016/j.jisa.2020.102583.

[18] Alonso-Monsalve, S. et al. (2020) ‘Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators’, Expert Systems with Applications, 149, p. 113250. doi:10.1016/j.eswa.2020.113250.

[19] Understanding LSTM networks (no date) Understanding LSTM Networks – colah’s blog. Available at: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (Accessed: 17 May 2023).

[20] Kaspersky (2023) What is cryptocurrency and how does it work?, www.kaspersky.com. Available at: https://www.kaspersky.com/resource-center/definitions/what-is-cryptocurrency (Accessed: 18 May 2023).

[21] Cryptocurrency exchange for Bitcoin, Ethereum & Altcoins (no date) Binance. Available at: https://www.binance.com/en (Accessed: 18 May 2023).

[22] (No date) Coinbase - buy and sell Bitcoin, Ethereum, and more with trust. Available at: https://www.coinbase.com/ (Accessed: 18 May 2023).

[23] Crypto Exchange: Bitcoin Exchange: Bitcoin trading (no date) KuCoin. Available at: https://www.kucoin.com/ (Accessed: 18 May 2023).

[24] Global cryptocurrency market charts (no date) CoinMarketCap. Available at: https://coinmarketcap.com/charts/ (Accessed: 18 May 2023).

[25] Roberta Silverglate, C.B.J. (2023) Bitcoin vs. Altcoins: Which is the better investment?, California Business Journal. Available at: https://calbizjournal.com/bitcoin-vs-altcoins-which-is-the-better-investment/ (Accessed: 18 May 2023).

[26] M., P. et al. (2021) ‘Global cryptocurrency trend prediction using social media’, Information Processing & Management, 58(6), p. 102708. doi:10.1016/j.ipm.2021.102708.

[27] Hansun, S., Wicaksana, A. and Khaliq, A.Q. (2022) ‘Multivariate cryptocurrency prediction: Comparative analysis of three recurrent neural networks approaches’, Journal of Big Data, 9(1). doi:10.1186/s40537-022-00601-7.

[28] Hamayel, M.J. and Owda, A.Y. (2021) 'A novel cryptocurrency Price prediction model using GRU, LSTM, and bi-LSTM machine learning algorithms', AI, 2(4), pp. 477–496. doi:10.3390/ai2040030.

[29] Zhuang, C. (no date) Stockstats, StockStats. Available at: https://pypi.org/project/stockstats/ (Accessed: March 12, 2023).

[30] Kostadinov, S. (2019) Understanding GRU networks, Medium. Towards Data Science. Available at: https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be (Accessed: March 21, 2023).

[31] Google (no date) Tensorflow, TensorFlow. Available at: https://www.tensorflow.org/ (Accessed: April 4, 2023).

[32] Kostadinov, S. (2019) Understanding GRU networks, Medium. Available at: https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be (Accessed: 17 May 2023).

[33] Chen, Z., Li, C. and Sun, W. (2020) 'Bitcoin price prediction using machine learning: An approach to sample dimension engineering', Journal of Computational and Applied Mathematics, 365, p. 112395. doi:10.1016/j.cam.2019.112395.

[34] Chowdhury, R. et al. (2020) 'An approach to predict and forecast the price of constituents and index of cryptocurrency using Machine Learning', Physica A: Statistical Mechanics and its Applications, 551, p. 124569. doi:10.1016/j.physa.2020.124569.

[35] Shahbazi, Z. and Byun, Y.-C. (2022) 'Knowledge discovery on cryptocurrency exchange rate prediction using machine learning pipelines', Sensors, 22(5), p. 1740. doi:10.3390/s22051740.

[36] Yahoo Finance - Stock Market Live, quotes, Business & Finance News (no date a) Yahoo! Finance. Available at: https://finance.yahoo.com/ (Accessed: 28 May 2023).

[37] Aroussi, R. (2017) Market data from Yahoo! Finance's API, GitHub. Available at: https://github.com/ranaroussi/yfinance (Accessed: March 12, 2023).

[38] Alternative.me (no date) Crypto Fear & Greed index - bitcoin sentiment, Bitcoin Sentiment - Alternative.me. Available at: https://alternative.me/crypto/fear-and-greed-index/ (Accessed: March 12, 2023).

[39] Adam (no date) Adam - Cornell University Computational Optimization Open Textbook - Optimization Wiki. Available at: https://optimization.cbe.cornell.edu/index.php?title=Adam (Accessed: 08 June 2023).