**Predicting Cryptocurrency Prices Using**

**Machine Learning Algorithms**

Literature review

Table of Contents

[Abstract 2](#_Toc127712952)

[Introduction 4](#_Toc127712953)

[Literature Review 4](#_Toc127712954)

[Approach/Methodology 6](#_Toc127712955)

[Overview of the Data 7](#_Toc127712956)

[Conclusion 7](#_Toc127712957)

[Reference 8](#_Toc127712958)

[1 8](#_Toc127712959)

[2 8](#_Toc127712960)

# Abstract

**Title:**

**Predicting Cryptocurrency Prices Using Machine Learning Algorithms.**

With the advancement and modernization of technology, various industries are rapidly evolving to adapt to the latest global trends. One such change that has been taking place is the increasing prevalence of cryptocurrency as a medium of exchange and investment. This shift towards digital currencies represents a significant departure from traditional fiat currencies and stock exchange practices and It’s the time to understand what digital money really means for everyone’s future **[1], [2]**.

**Research Questions:**

What are the most effective predictive and time series analysis techniques for forecasting short-term closing prices of cryptocurrencies? Which features are influential predictors for classifying the short-term closing prices of selected cryptocurrencies?

What is the correlation between the predicted prices generated by machine learning algorithms and the actual prices of the chosen cryptocurrencies?

**Scope of the Research:**

The project aims to use machine learning algorithms to predict short-term closing prices of different cryptocurrency companies. The dataset for this project is obtained from Kaggle Inc. which contains historical data for the chosen cryptocurrencies. The objective is to compare the predicted prices with the actual prices and identify which cryptocurrency presents the most profitable opportunity for short-term trading.

To answer the research question, we will explore different machine learning algorithms and time-series analysis techniques, including LSTM and ARIMA. We will compare the efficiency and stability of these techniques to identify the most effective ones for our purpose.

**Data Source:**

The data set used in this project is obtained from Kaggle Inc. which contains historical data for the chosen cryptocurrencies. The data is related to the closing prices of each of the six different cryptocurrency companies.

**Limitations of the Research:**

The scope of this study is restricted to specific cryptocurrency companies and the historical data that is accessible for them, and thus, it may not accurately reflect the overall cryptocurrency market. The machine learning algorithms utilized in the analysis rely on historical data, and the future value of cryptocurrencies can be influenced by unpredictable factors, such as changes in regulations, market sentiment, and global events. The precision of the forecasts could be influenced by the quality and comprehensiveness of the data used.

**Background Information:**

Cryptography is used to secure cryptocurrency, which is a digital or virtual form of money. It is decentralised and not under the jurisdiction of a single entity, such as a government or bank, unlike conventional currencies. It is a distributed ledger used to record cryptocurrency transactions and is used to secure and authenticate user data. The worldwide financial system has been significantly impacted by this innovative technology, and its future growth potential is enormous. The first cryptocurrency, Bitcoin, was released in 2009, and since then, the market has expanded to encompass several other cryptocurrencies, with a market capitalization of over $1 trillion. Despite having only recently emerged, cryptocurrencies have already had a significant impact on the financial landscape and are predicted to continue doing so.

**Specific Area of Research:**

The specific area of research in this project is the utilization of machine learning algorithms in time-series analysis to predict the future prices of selected cryptocurrencies and identify profitable opportunities for short-term investment.

**Data Set:**

The project aims to use data to achieve the goal **[5]**.

<https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory>

**Gethib Link:**

<https://github.com/shahgem/CIND-820>

# Introduction

The field of cryptocurrency price prediction is an exciting and rapidly evolving area of research, with many promising avenues to explore. Given the newness of this topic, the number of research papers available is limited, making it an ideal area for finance and data enthusiasts to explore and experiment with various deep machine learning models.

Conventional time series techniques, such as ARIMA, have proved to be ineffective in capturing the non-linear and non-stationary patterns that are commonly present in cryptocurrency data. These limitations underscore the need for alternative approaches to analyze this unique and dynamic market.

One of the major challenges in predicting cryptocurrency prices is the high volatility and frequency of price fluctuations, which can lead to underlying chaos. As such, Deep Learning Methods have emerged as a critical tool in this field, as they can capture complex patterns in the data that other methods might miss.

While several DL methods have been suggested in review papers, including LSTM, RNN, SVM, ANN, and SANN, only a few are relevant to this project, given the constraints on time and expertise. Hence, this project will focus on a select subset of DL methods that are suitable for forecasting cryptocurrency prices.

In addition to analyzing the historical data of cryptocurrencies, researchers have also explored the role of sentiment analysis in price prediction. Some studies have analyzed the frequency of "Bitcoin" in tweets to determine if it correlates with the predictability of the BTC price, using Natural Language Processing (NLP) methods. However, this project will not consider sentiment analysis due to the limitations of the scope and the complexity of NLP techniques. One of the studies “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis” **[**3**]** concludes People who tweet about cryptocurrencies even when their prices drop have an interest in them beyond investment opportunity making the tweets biased towards positive.

In summary, this project aims to explore and evaluate the performance of several Deep Learning Methods to forecast the prices of cryptocurrencies. While previous researchers have conducted extensive investigations on this topic, I will approach it from a fresh perspective and provide additional insights. **[4]**. Through this project, we hope to contribute to the growing body of research in this exciting and rapidly evolving field.

# LiteratureReview

The emergence of cryptocurrencies has created an entirely new asset class that is characterized by high volatility and lack of regulation. As a result, the study of cryptocurrency markets has become increasingly important to both investors and financial analysts. The field of cryptocurrency price prediction has garnered significant attention, yet there is still a dearth of research papers on the subject. This shortage of information has motivated finance and data enthusiasts to explore the application of various deep machine learning models to market data.

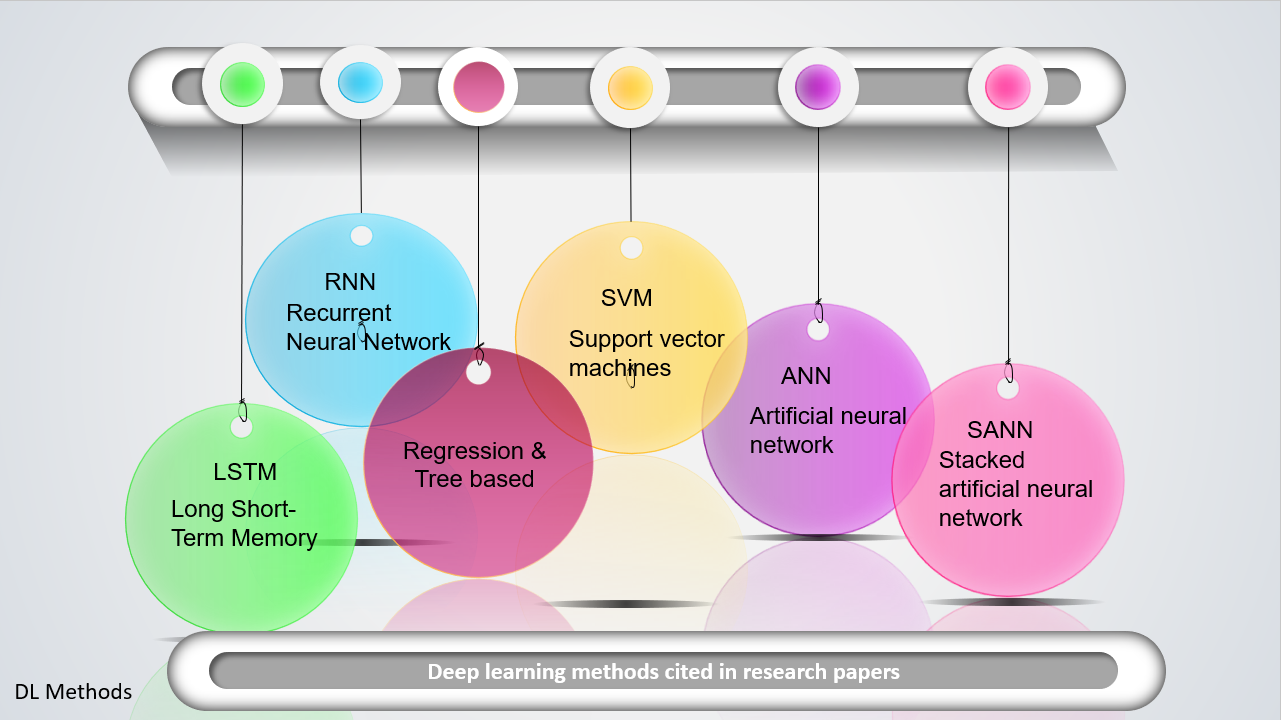
Conventional time series techniques such as AutoRegressive Integrated Moving Average (ARIMA) have limitations in capturing the non-linear and non-stationary patterns that exist within cryptocurrency data, highlighting the need for alternative approaches. Deep learning methods have shown great promise in addressing these issues, owing to their ability to handle complex, high-dimensional datasets. Researchers agree that the frequent and high volatility of cryptocurrency prices can lead to underlying chaos, necessitating the use of Deep Learning Methods.

Several research papers have investigated which deep learning method is most accurate in forecasting cryptocurrency prices. Classification methods such as Random Forest Trees and k-folds were employed to compare the accuracy of each method. The studies have primarily focused on analyzing Bitcoin (BTC), given its status as the longest active currency for conducting analysis and its dominance in the cryptocurrency market.

The reviewed papers primarily investigate the performance differences of deep learning methods across various time intervals of cryptocurrency prices. These time intervals include daily, weekly, and monthly opening and closing prices. In addition to forecasting prices based on historical data, some studies have also incorporated sentiment analysis.

Researchers have analyzed the frequency of the term "Bitcoin" in tweets and explored its relationship to BTC predictability. However, it is worth noting that the utilization of natural language processing (NLP) methods for sentiment analysis is not relevant to this project.

Overall, the study of cryptocurrency price prediction is a nascent field that requires further research. Deep learning methods have shown great promise in addressing the unique challenges posed by cryptocurrency markets, and several studies have explored the application of these methods to BTC. This project aims to build on this research by investigating the performance of various deep learning methods for forecasting the prices of other popular cryptocurrencies.



# Approach/Methodology

1. Data Collection/Preparation: Explain the data collection process, specifying the types of data collected, the sources, and the relevant timeframes **[5]**.
2. EDA/Visualization: This step involves exploring the data that have been collected and getting a better understanding of its properties and characteristics. This may involve creating visualizations of the data to help you identify patterns, trends, and anomalies **[6]**.
3. Data Processing: This step involves processing the data in a way that prepares it for use in the predictive model. This may involve feature engineering, which is the process of selecting and transforming the variables that will be used to make predictions **[7]**.
4. Apply the data to the model: This step involves building a predictive model that uses the data to make predictions.
5. Predict: This step involves using the trained model to make predictions on new, unseen data. The goal is to use the model to make accurate predictions that can be used to inform decision-making.
6. Conclusion and Future Work: Summarize the study's findings, discuss the implications for cryptocurrency price prediction, and suggest possible future research directions.



# Overview of the Data

The dataset utilized in this project was sourced from Kaggle Inc, and is publicly accessible under the title "Cryptocurrency Market Data Historical Cryptocurrency Price for All tokens!" **[5]**.

During this course of study, I will emphasis on Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), Dogecoin (DOGE), Cardano (ADA), and Binance Coin (BNB). Although Kaggle provides separate CSV files for each cryptocurrency, the author of this project has amalgamated all the individual files into a single CSV file named "crypto\_market.csv". This amalgamation was possible due to the identical columns of each file, i.e., Serial Number, Name, Symbol, Date, High, Low, Open, Close, Volume, Market Capitalization.

The initial dataset brings to the fore the non-uniformity of the trading dates of the various currencies. As previously noted, BTC has been in circulation for the longest period, pre-dating all the other cryptocurrencies in the dataset. Since the start dates for each currency, which may significantly impact their predicted prices. To mitigate this potential bias, it is advisable to limit the analysis to the period from 2017 onwards, as this provides a more consistent and recent historical data for all the selected currencies. This filtered dataset has been saved under the file name "sixcryto.csv".

In light of this, during the analysis phase of this project, it is prudent to compare all currencies from 2017 onwards, so as to minimize prediction bias and incorporate a more uniform and recent historical data.

# Conclusion

This project aims to predict the prices of various currencies using Python's vast range of packages such as Sklearn, Numpy, Pandas, and Keras. The goal is to forecast the prices of each currency 60 days into the future and compare their profitability to identify the most profitable investment option for potential investors.

The existing literature in this field tends to reiterate the same information repeatedly. However, this project will approach the problem from a new perspective and leverage recent trends and developments in the field of cryptocurrency and financial forecasting.

To achieve this goal, we will follow a comprehensive approach that involves collecting relevant data, preprocessing the data, performing feature engineering, and using various machine learning techniques to model the data. The performance of each model will be evaluated using appropriate evaluation metrics, and the best-performing model will be selected to forecast the prices of each currency.

In summary, this project will contribute to the existing literature in this field by providing a new perspective on the problem of predicting the prices of various currencies. Additionally, it will leverage the power of Python's advanced machine learning techniques to forecast the prices of each currency and recommend the most profitable investment option for potential investors.

# Reference

# 1

Hern, A. (2018). *Bitcoin and cryptocurrencies – what digital money really means for our future*. Retrieved from the Guardian: https://www.theguardian.com/technology/2018/jan/29/cryptocurrencies-bitcoin-blockchain-what-they-really-mean-for-our-future

# 2

*How blockchain could disrupt banking*. (2022, Oct 18). Retrieved from CNBC: https://www.cbinsights.com/research/blockchain-disrupting-banking/#:~:text=Payments%3A%20By%20establishing%20a%20decentralized,time%20transactions%20between%20financial%20institutions.

# 3

Jethin Abraham, D. H. (2018). *The SMU Data Science Review*. Retrieved from chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039&context=datasciencereview

# 4

JYOTI SARPAL, T. R. (n.d.). Retrieved from chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.ir.juit.ac.in:8080/jspui/bitstream/123456789/3602/1/Cryptocurrency%20Price%20Prediction%20Using%20Deep%20Learning.pdf

# 5

*Cryptocurrency Historical Prices*. (n.d.). Retrieved from Kaggle: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory

# 6

Santosh, S. (2021). *A Comprehensive Guide On Data Visualization In Python*. Retrieved from AnalyticsVidhya: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-data-visualization-in-python/

# 7

Rençberoğlu, E. (2019). *Fundamental Techniques of Feature Engineering for Machine Learning*. Retrieved from Towards Data Science: https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114