Name: Manomi Koroth
Student ID: 501066400

CIND820 Big Data Analytics Project

Ryerson University

Fall 2021

A study on predicting prices of different Cryptocurrency using Machine Learning Algorithms.

1.Abstract

Nowadays, everyone has a desire to invest and earn an extra income outside of the regular day-to-day work schedule and for a while, investing in government bonds and stocks were the only way to achieve that. One of the most basic lessons in finance is the risk and reward are directly proportional, i.e., the greater the risk, the greater the reward. More recently, the spotlight has shone on cryptocurrency --which is a relatively more lucrative area of FinTech. Cryptocurrency is a new-age digital, decentralized and encrypted medium of exchange that is not governed by any form of institution which makes it more desirable as an asset. While cryptocurrency cannot replace fiat money, it alleviates the barriers of the exchange rates among national currencies. Technological advancement is uniformly influencing all parts of the economy; thereby also influencing the global markets to add cryptocurrency as another industry for people to invest in--besides company stocks. In this project, I aim to explore the forecasting short-term closing prices of each of the six different cryptocurrency companies and compare the predicted price to the actual price using appropriate machine learning algorithms.

The dataset I chose for this project is obtained from Kaggle Inc. website, which originally contains twenty-three csv files for each of the twenty-three different crypto currencies. However, I chose to save only csv files related to the six crypto currencies I aim to analyse

The goal of this project is to apply time-series with machine learning aspects- more specifically Long Short-Term Memory (LSTM) model on each of the six datasets and find out which one of them is profitable in the short-term trade.

Name: Manomi Koroth
Student ID: 501066400

2.Introduction

The digital influence has trickled into the financial trading markets and has been growing recognition among investors at a rapid pace since the early 2010s. According to a news report dated in early April of this year, the cryptocurrency market size reached approximately $2 trillion USD [1]. There are approximately 9775 currencies active today, Bitcoin (BTC) being the first and most popular and currently dominates close to 50% of the crypto market [2]. Since it is relatively new to the trading market, there has been limited literature research conducted on the matter, and by this, we can also assume that there is limited amount of data available to create accurate predictions on digital currency price. However, BTC has been active the longest, and the Kaggle dataset contains trading information from 2013, making it relatively easier for researchers to make price predictions compared to other cryptocurrencies [3].

3. Literature Review

As aforementioned, the novelty of cryptocurrency price prediction has contributed to the lack of many research papers on the subject, thereby also encouraging finance and data enthusiasts to experiment on different deep machine learning models that could apply to market data. When traditional time series methods such as AutoRegressive Integrated Moving Average (ARIMA) are applied to crypto currencies, it fails to grasp the limitations due to the non-linear and non-stationary patterns in the cryptocurrency data[3][4].

Despite the newness of the subject, researchers agree that frequent and high volatility of the prices causes an underlying chaos using traditional machine learning methods, thereby forcing Deep Learning Methods to come into play. Most review papers I came across suggested some DL methods are listed in Table 1. However, a few from the list go beyond the scope of this project due time and skill constraints.

Table 1

Methods mentioned in different papers
Long Short-Term Memory (LSTM)
Recurrent Neural Networks(RNN)

Regression-based
Support vector machines (SVM)
Tree-based
Artificial neural network (ANN)
Stacked artificial neural network (SANN)

3.1 Characteristics of the related works

The goal of the majority of the papers was to determine which of the listed DL methods is the most accurate to forecast the price of a cryptocurrency by using classification methods i.e., k-folds, Random Forest Trees to compare each and find out which classifies as the most accurate.

The next similar aspect of recent studies shows the analysis done mostly on BTC, it being the longest active currency to conduct analysis on. As mentioned in the introduction, it is the most popular.

The cited works focus on how each of the methods differ performance levels on the different time intervals of the prices. The papers differ between daily, weekly and monthly intervals of opening and closing prices. Besides forecasting prices based on historical data, the related works dwelled into researchers also releasing heavily on sentiment analysis. According to a few sources, researchers aimed to find the frequency of “Bitcoin” in tweets and if it related to the predictability of the process of the BTC, probably using Natural Language Processes (NLP) methods which will not be factored into this project.

4. Data

The dataset used was obtained from Kaggle Inc and is publicly available under Cryptocurrency Price History [5]. There are 23 csv files that contain historical prices for 23 different cryptocurrencies but the ones this project focuses on are those of Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), Dogecoin (DOGE), Cardano (ADA), and Binance Coin (BNB). Kaggle makes csv files for each of these cryptocurrencies individually, however, I merged all the individual files into one csv file saved under

crypto_market.csv. It was simple to merge all these files into one because each individual file has the same columns: *Serial Number, Name, Symbol, Date, High, Low, Open, Close, Volume, Market Capitalization.*

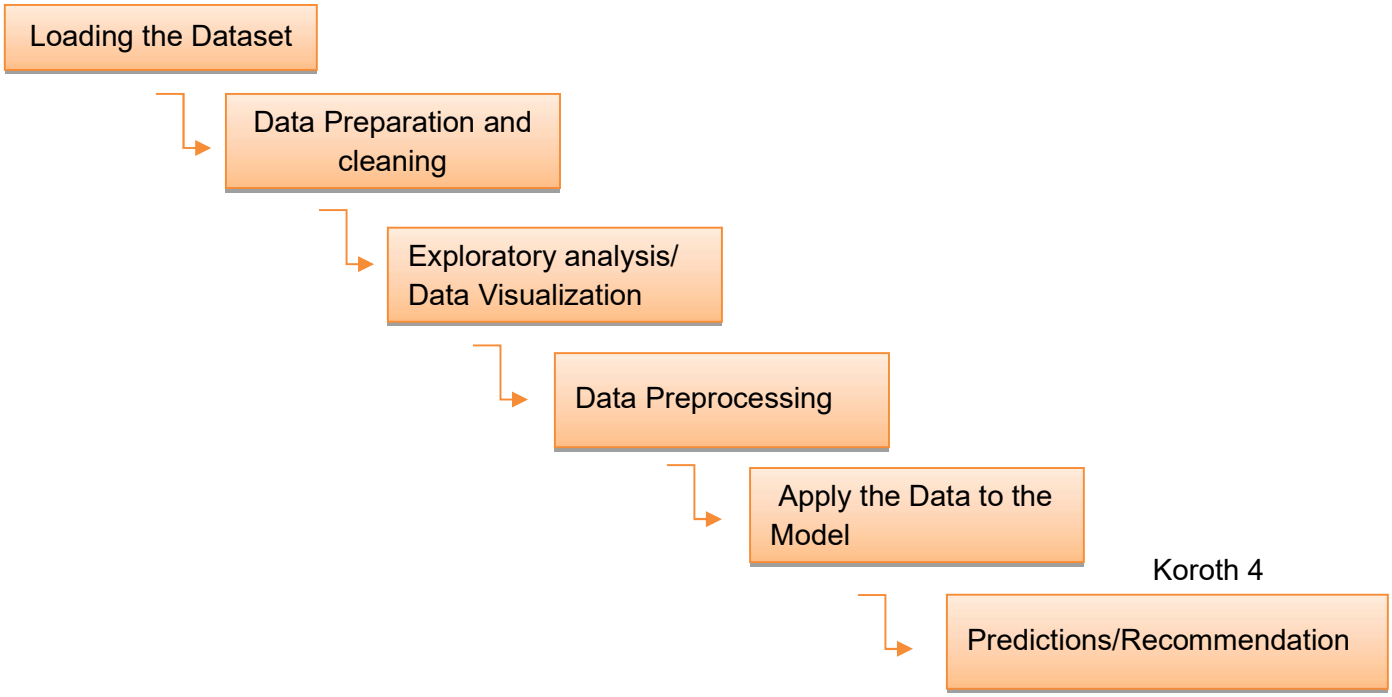
The initial dataset makes us focus on the non-uniform dates of each of the currencies. As mentioned, countless times before, BTC was active the longest, therefore pre-dates all the other currencies. As shown in Table 2, the starting trade dates of each currency differ and therefore would affect the predicted price of each.

Therefore during the analysis phases of the project, I will chose compare all currencies from 2017 to keep the prediction less biased and from a more uniform past data.

Table 2

Symbol	**Date**
BTC	2013-04-29
ETH	2015-08-08
DOGE	2013-12-16
ADA	2017-10-02
BNB	2017-07-26
LTC	2013-04-29

5. Methodology



Name: Manomi Koroth
Student ID: 501066400

6. Conclusion

I will be using Python as my programming language of choice because it will give me access to packages such as Sklearn, Numpy, Pandas and Keras. These packages not only contain tools for basic statistical tools for Exploratory Data analysis (EDA) but also deep neural network tools to perform LSTM, RNN or SVM models after the data preprocessing phase.

The aim of this project is predict prices of each of currencies by forecasting each 60 days into the future and then compare the profitability of each and lastly making a recommendation based on which was the most profitable for investor.

I noticed that most of the related works conducted so far have been published recently and therefore were virtually repetitive of the same information.

Name: Manomi Koroth
Student ID: 501066400

7. References

1: Kharpal, Arjun. “Cryptocurrency Market Value Tops $2 Trillion for the First Time as Ethereum Hits Record High.” *CNBC*, CNBC, 6 Apr. 2021, https://www.cnbc.com/2021/04/06/cryptocurrency-market-cap-tops-2-trillion-for-the-first-time.html.

2: “All Cryptocurrencies.” *CoinGecko*, https://www.coingecko.com/en/coins/all.

3: Mudassir, Mohammed, et al. “Time-Series Forecasting of Bitcoin Prices Using High-Dimensional Features: A Machine Learning Approach.” *Neural Computing and Applications*, Springer London, 4 July 2020, https://link.springer.com/article/10.1007/s00521-020-05129-6.

4: Pintelas E., Livieris I.E., Stavroyiannis S., Kotsilieris T., Pintelas P. (2020*) Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach*. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_9

5: Srk. “Cryptocurrency Historical Prices.” *Kaggle*, 7 July 2021, https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory.

6: Adcock, Robert, and Nikola Gradojevic. “Non-Fundamental, Non-Parametric Bitcoin Forecasting.” *Physica A: Statistical Mechanics and Its Applications*, North-Holland, 12 June 2019, https://www.sciencedirect.com/science/article/pii/S0378437119309859.

7: Yogeshwaran, S. & Kaur, Maninder & Maheshwari, Piyush. (2019). Project Based Learning: Predicting Bitcoin Prices using Deep Learning. 1449-1454. 10.1109/EDUCON.2019.8725091.

8: Simonetti, Juan. “Short Term Bitcoin Price Prediction with Deep Learning.” *Medium*, Geek Culture, 25 Aug. 2021, https://medium.com/geekculture/short-term-bitcoin-price-prediction-with-deep-learning-ab4386e84b5.