Fontys University of Applied Sciences - ICT & Software engineering

Crypto Market

By Konstantin Mihaylov

08/03/2021

Table of Contents

Preface

This is a proposal document for the Genuine Challenge of Semester 4 ICT and Artificial Intelligence at Fontys University of Applied Sciences. Throughout this semester I will be working on an exploratory project as a part of Fontys Education. This will be a personal effort in which I will experience all the different phases of the AI Project Methodology resulting in a final delivery. Using a chosen dataset, I should develop a business intelligence solution that could help to forecast statistics for the crypto world.

The document starts with a brief description of the domain of the project in the face of planning. Afterwards, I will specify the project goal as well as the societal impact, which it would encompass. Then I will specify what is the received data, describe the working process and methodology, including our planning for each phase of the project, and I evaluate the potential risks involved in the process. Finally, I finish the document with a summarizing conclusion.

Glossary

ARIMA - **AutoRegressive Integrated Moving Average**

Blockchain - a system in which a record of transactions made in bitcoin or another cryptocurrency are maintained across several computers that are linked in a peer-to-peer network.

1. Abstract

This paper describes the construction of the short-term forecasting model of cryptocurrencies' prices using a machine learning approach. Using data time series and the autoregressive model ARIMA, I will make a short-term forecast (from 5 to 30 days) for the one of the most capitalized cryptocurrency: Bitcoin. I found that the proposed approach was more accurate in forecasting cryptocurrencies' time series both in the periods of slow rising (falling) and in the periods of transition dynamics (change of trend).

2. Introduction

The rapid development of digital currencies during the last decade is one of the most controversial and ambiguous innovations in the modern global economy. Significant fluctuations in the exchange rate of cryptocurrencies and their high volatility, as well as the lack of legal regulation of their transactions in most countries resulted in significant risks associated with investment into crypto assets. Machine learning models can likely give us useful insight to learn about the future of Cryptocurrency. It might tell us the general trend and direction of the prices.

A number of recent cryptocurrency market studies show that, unlike other financial assets, cryptocurrency prices are influenced by a number of specific factors that shape their demand, such as the number of Google trends searches, the number of posts in social networks and other mass media. These studies substantiated the feasibility of using non-typical factors as predictors.

3. Context Understanding

3.1. Impact on society

Clearly before directly speaking about the societal impact of this virtual currency, we have to define what is bitcoin and what determines the bitcoin's price?

Bitcoin is a cryptocurrency developed in 2009 by the unknown creator with the pseudonym, Satoshi Nakamoto. No one knows who Nakamoto is and has been able to figure out the true identity of this person or organization, but somehow the idea of the ambiguous nature of cryptocurrency and blockchain remains a mystery. Initially the blockchain technology was created to store bitcoin, but eventually, blockchain and bitcoin separated when people found a use for the blockchain outside the cryptocurrency. Using this technology the people recognise how blockchain may perform many different kinds of operations, such as transactions, smart contracts and even as a storage for applications. A blockchain is a decentralized ledger of all transactions across a peer-to-peer network. A decentralized ledger, in its simplest form, is a database that consists of and is updated by nodes or participants in a large network. Because blockchain is stored on a distributed ledger, it can safely and effectively record the transactions. Blockchain technology was given that name

because it consists of a large number of blocks of data linked together on a chain. Unlike investing in traditional currencies, bitcoin is not issued by a central bank or institution, but it is decentralized all across the network.

A singular block within a blockchain is a collection of data. The collection of data is stored in a block on a blockchain by connecting to other blocks in chronological order. With a peer-to-peer network, there is no need for a centralized authority to oversee and mediate transactions. Instead, it allows buyers, sellers, and other participants to execute trades directly to each other without the need for intermediaries using trading platforms.

Once the data is recorded, the information in each block cannot be updated or changed without altering the following blocks, which requires a consensus within the network. This consensus is how blockchain technology can prevent unauthorized online activity and is therefore not as vulnerable to hacking.

Conversary the traditional investments, bitcoin which is run on the decentralized blockchain is not like the traditional stock backed by an institution. Therefore, the monetary policy, inflation rates and economic growth measurements that typically influence the value of currency do not apply to bitcoin. However, bitcoin prices are influenced by the following factors:

- The supply of bitcoin and the market's demand for it
- The cost of producing a bitcoin through the mining process
- The rewards issued to bitcoin miners for verifying transactions
- The number of competing cryptocurrencies
- The exchanges it trades on
- Regulations governing its sale
- Its internal governance

Throughout the past decade, there has been an increase in the popularity of this cryptocurrency, because it seems like a technology of the future, but it has also revolutionized the way we store money, transfer, pay and do business. Additionally, it has improved the e-commerce system by making a better place for the users to make transactions more privately and efficiently.

A possible reason why the market for cryptocurrency gained popularity so drastically could be due to many individuals taking advantage of cryptocurrency as a lucrative investment opportunity. While many are still skeptical of cryptocurrency and lack an interest in investing time and money due to the uncertainty and mysterious nature of the market, some have invested by buying popular cryptocurrencies like Bitcoin, Litecoin, and Ethereum. Others have attempted to profit off of cryptocurrencies by mining bitcoins themselves, this has worked for some, but for others, this was unsuccessful, because, at a certain point, it costs more money to run the equipment that mines, calculates,, the cryptocurrency than the money gained by the mined

digital currency. Even still, countless individuals have made a small fortune from mining bitcoins.

Cryptocurrency has also gained popularity among many businesses and personal finances. Governments, businesses, and other large corporations are working to learn more about the cryptocurrency market and how they can utilize blockchain technology to their advantage. Many companies and businesses have begun researching blockchain technology to assess the practicality of incorporating this technology into their business operations. Since our society is becoming more and more digitally driven, financial service providers are becoming more interested in learning about cryptocurrency and potentially using it as a way to provide secure services in a more efficient and cost-effective manner.

By using the forecasts of the Crypto Market and analytics, those who use it can get an understanding where the price rates are pointing in a short term and why so many investors started looking at what kind of challenges bitcoin brings to the economy.

3.2. Sustainability

The Artificial Intelligence application would run like any software-based system. In other words, it would run on electricity which presents no more or less environmental impact than traditional software applications.

3.3. Privacy

In this section I will state how transparent the Crypto Market forecast could be. Everything will be open-source, which means every citizen will have public access to the code. Furthermore the dataset used for the predictions is based on the yahoo finance international system which is an open resource.

3.4. Stakeholders

Generally speaking the stakeholders are those who have a use of a wine recommendation system including all of the consumers, students and those who trade crypto stocks.

3.5. Target Group

Up till this year there are at least 35 million people with Bitcoin wallets. The main target group is 25–34 year olds, second highest group, 35–45. Mostly Male, but female investors (currently at about 5.3%) are doubling every two months or so. Low-tech to high-tech individuals. Global.

3.6. Ethics

In this section I will try to answer the most important question which is how risky investing in cryptocurrencies is. Crypto is risky and it is not like conventional investing in the stock market, but the main reason is that cryptocurrencies are unregulated by governments. The ethical part for the Crypto Market project is based on the fact that all the information about the price rate is publicly available for everyone. However, it must be stated that the prediction made from the model is based on forecast algorithms in a perfect non-realistic case. Everyone who's using the project should be aware that it is an approximate prediction which shouldn't depend upon.

3.7. Inclusivity

The way our perspectives and experiences are built shape directly the way products are designed. That's why it is very important when a software is built to be fair for everyone. It first starts within the organization. The technology should not be designed with people, but rather for people. We should empathize with user communities to understand their needs in order to challenge the traditional design thinking and human-centered product development.

4. Challenge Overview

4.1. Description

Cryptocurrencies are the most valuable investment of this decade. Bitcoin has been increasingly regarded as an investment asset. Because of its highly volatile nature, there is a need for good predictions on which to base investment decisions. While bitcoin is still the dominant option concerning market capitalization, alternative coins including Ethereum (ETH), Tether (USDT), Binance Coin (BNB), Cardano (ADA), and Polkadot (DOT) are among its closest competitors as of March 2021. What if we can predict the price rate for any cryptocurrency?

4.2. Project goal

I aim to produce reliable forecasts of time series related to the cryptocurrency Bitcoin. I will begin by introducing concepts of autocorrelation, stationarity, seasonality and proceed to apply one of the most commonly used methods for time-series forecasting, known as AutoRegressive Integrated Moving Average. By the end of the project I should have an overview of the price rate for the upcoming month.

4.3. Delivery

During the execution of the project the following sub tasks are going to be created.

1. Proposal
   In the following challenge I am going to do all four phases. The project proposal will contain an Exploratory Data Analysis and the potential impact assessment.

2. Provisioning
   In order to solve this problem an analysis in the form of a reproducible Jupyter notebook is going to be delivered.

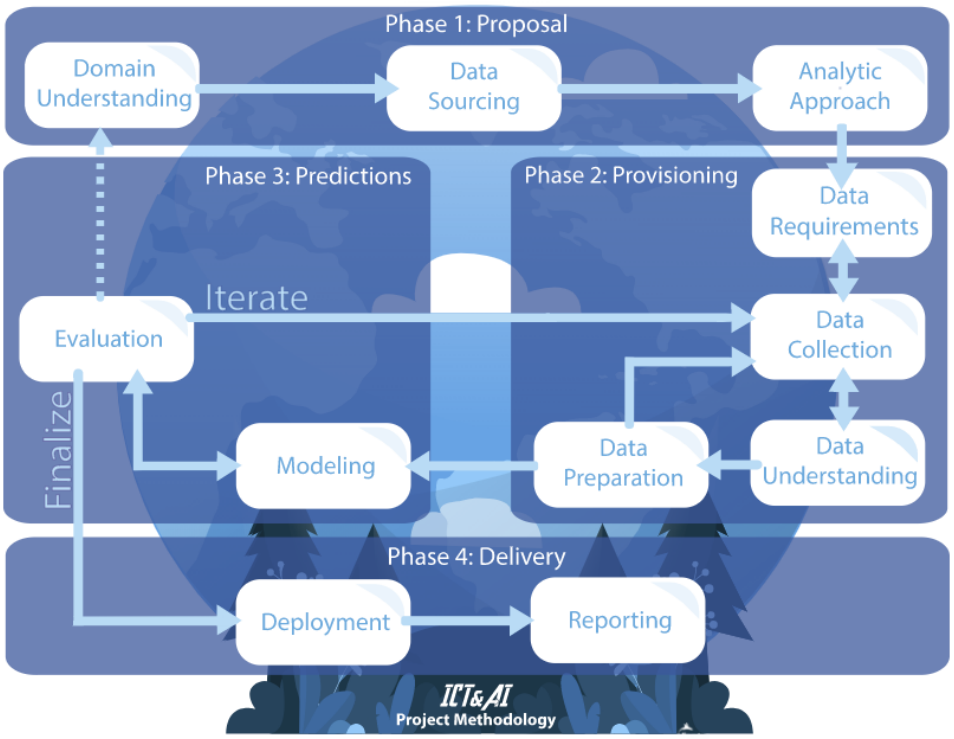
3. Prediction
   In this phase I will deliver a Result Document and Modeling Jupyter notebook attached.

4. Delivery
   In the following phase it will be delivered a Deployment document with Demonstration Materials attached.

5. Approach

5.1. Development methodology

During the semester was introduced to us the AI Project Methodology, which is loosely based on the IBM Foundation Methodology for Data Science. It consists of the following steps: Domain Understanding, Data Sourcing, Analytic Approach, which are part of a bigger domain named phase 1, Proposal.



After that there is the Provisioning phase 2, which gathers the data requirements and collection, data understanding, data preparation until the data is ready for the next phase. In the Prediction phase 3 a model is created, evaluated and if needed it should iterate through all the phases again, until the result is satisfying and evaluated. The final Delivery phase 4 is used for deployment of the model and reporting it.

In case the prediction results are evaluated negatively there is a possibility to work iteratively by going back from phase 3 to phase 2. This may be when it is understood that the models predict poorly based on the quality of the data that is used as input. Such iteration is likely to happen at least a few times before satisfactory results are reached, and the project is finalised. Also, every phase (and iteration of a phase) that your project passes through produces a deliverable.

5.2. Phases

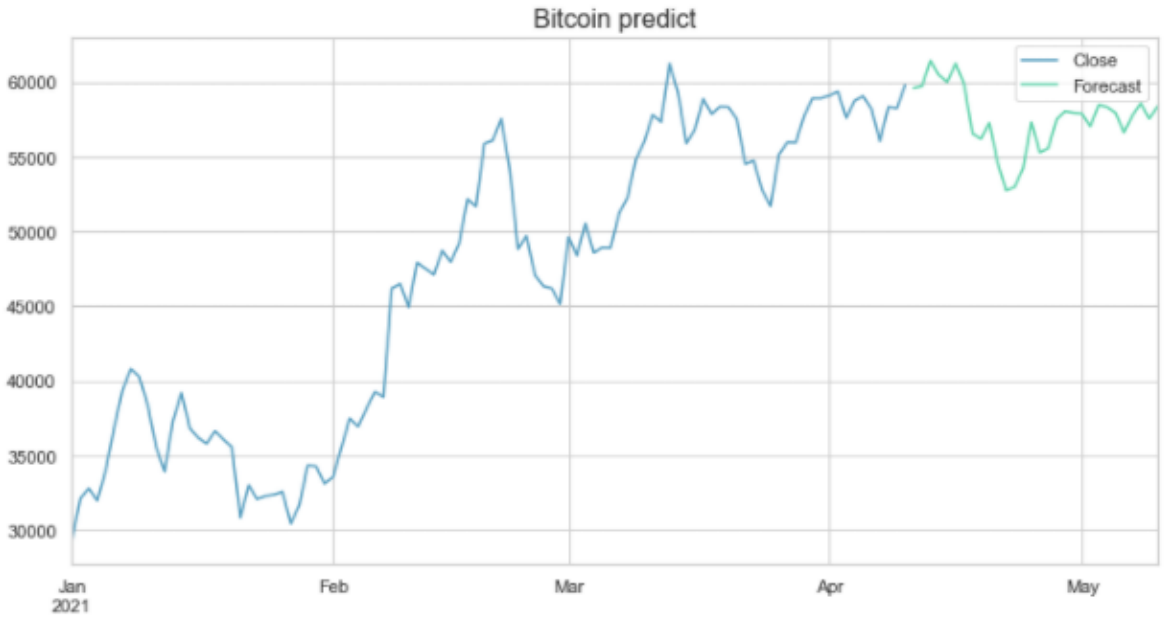
*	Phase(s)	Deliverable(s)	Deadline(s)
1	**PROJECT PROPOSAL**	📄 Project Proposal document with 📒 Exploratory Data Analysis notebook attached.	30th April
2	**PREPARED DATASET**	🖨️ Data Storage Solution and 📒 Preparation notebook.	1th May
3	**PREDICTION RESULTS**	📄 Results document with 📒 Modelling notebook attached.	2th May
4	**FINAL DELIVERY**	📄 Deployment Recommendation document with 📊 Demonstration Materials attached.	5th May

Each phase of the AI Project Methodology produces a deliverable. In the first project proposal phase the exploratory data analysis notebook, which is done with Jupyter notebook or an open document format based on json will demonstrate my attempt at finding opportunities for predictive analysis from a dataset. In this section there are no predictions yet. In the second phase the data storage solution will make the dataset easily accessible to the project in a suitable format for example it can de-serialize the data into a pickle file by using a Python module. The preparation notebook takes your raw data, cleans, prepares and stores it in such a way that the result is a data storage solution suitable for the project. The results document gives an overview of the quality of the prediction results from the modelling activities. The deployment recommendation document describes how I imagine the model could be applied in a real world example or examples, which do not have to be implemented, but I should provide a reasonable application that brings value. Value is not always

money. The deployment recommendations should be supported by a full impact analysis.

5.3. Analysis usability

Time series analysis is a statistical method of analyzing data from repeated observations on a single unit or individual at regular intervals over a large number of observations. Time-series analysis can be viewed as the exemplar of longitudinal designs. The most widely used approach is based on the class of models known as Autoregressive Integrated Moving Average (ARIMA) models.



Time series analysis had been more generally developed in areas such as engineering and economics before it came into widespread use within social science research. In time series analysis, dependence is assessed by calculating the values of the autocorrelations among the data points in the series. In contrast to a correlation coefficient, which is generally used to estimate the relationship between two different variables measured at the same time on multiple subjects, an autocorrelation estimates the relationships within one variable that is measured at regular intervals over time on only one subject in our case the price rate.

6. Data definitions

The dataset which I am going to use is based on Yahoo! Finance network, which provides financial news, data and commentary including stock quotes, press releases, financial reports. I am going to use Yahoo Finance provider to extract the latest historical data for cryptocurrencies such as Bitcoin (BTC-USD), Ethereum (ETH-USD) and Dogecoin (DOGE-USD). More info at:
https://finance.yahoo.com/cryptocurrencies/

The main data information about cryptocurrencies which you can find is here:

Field	Definition
Symbol	Type of stock.
Date	The date when the stock is registered.
Open	The price of the stock at the beginning of the trading day.
Hight	Highest price of the stock on that trading day.
Low	The lowest price of the stock on that trading day.
Close	The price of the stock at closing time.
Adj Close	The adjusted closing price amends a stock's closing price to reflect that stock's value after accounting for any corporate actions/ raw price.
Volume	The number of coins that have exchanged hands during a defined period of time, how much value of a coin has been bought and sold over the course of a day.

7. Deployment

In a machine learning project storing the models in order to log and manage its complexity sometimes is essential. Data version control or DVC connects the big data sets with the code by storing the content to Google Cloud Storage, Azure Blob Storage etc. DVC is an experiment management tool that takes advantage of the existing engineering toolset as Git and CI/CD. It can help to organize my data. As another solution I am going to containerize my reproducible model with Docker. It can open many options like training the model inside of a container and serving it in almost every environment as an endpoint solution.

8. Conclusion

Crypto Market is just a small project, Proof of Concept, which can be put into so many applications. Every trading platform has to use predictive models in a way that can recommend the best rates for the users needs. Nowadays these time series forecasts provide the way that you can predict future values or trends such as price flotations. Of course, there are countless factors that can influence cryptocurrencies and their investors — but looking at all of them may be much harder than it seems to be. The unregulated and anonymous nature of digital asset markets combined with the vulnerability of cryptocurrencies and other crypto assets to sentiment, emotion, and publicity make prices volatile. A lack of regulatory frameworks means there is a high degree of uncertainty like price volatility and manipulation. Investors and entrepreneurs are also concerned about the possibility of future restrictions, which may have a significant impact on the value of cryptocurrencies or end up, ultimately banning them altogether. Crypto regulations are complex and disorganized. It is risky to invest anywhere, but I believe that if we are looking further into the future rates we are most likely to be convinced of the stability of the trends. Considering all of the massive price projections and the forecasts suggest that cryptocurrencies have an interesting trading future.

References

[1] Predicting prices of bitcoin with machine learning, Data Science,
https://towardsdatascience.com/predicting-prices-of-bitcoin-with-machine-learning-3e83bb4dd35f

[2] What determines the value of bitcoin? , Investopedia Inc.,
https://www.investopedia.com/tech/what-determines-value-1-bitcoin/

[3] Time Series Analysis, Data Science,
https://www.researchgate.net/publication/229633091_Time_Series_Analysis

Image Credits

1. © Pearson Education, Inc