

CRYPTO CURRENCY MARKET PRICE PREDICTION USING DATA SCIENCE TECHNIQUE

ARIVUNIDHI. A

Department of Computer Science & Engineering, SRMIST, Vadapalani, Chennai, India.
Email: ai7710@srmist.edu.in

M. RITU RAJU

Department Of Computer Science & Engineering, SRMIST, Vadapalani, Chennai, India.
Email: mr7828@srmist.edu.in

R. B. BALAJI KAMAL

Department Of Computer Science & Engineering, SRMIST, Vadapalani, Chennai, India.
Email: bk3890@srmist.edu.in

MR. N. MUTHURASU

Assistant Professor, SRM University of Science and Technology, Vadapalani, Chennai.

ABSTRACT

“Cryptocurrency is a sort of digital currency that is distinguished by the fact that records of coin ownership are kept in a ledger that takes the form of a computerised database. Strong cryptography is utilised in order to protect transaction records, regulate the production of extra coins, and validate the sale or transfer of currency ownership. Cryptocurrency is a form of digital currency. Furthermore, due to its volatile value, the price of cryptocurrency is hard to forecast. A model based on machine learning is being developed for the purpose of predicting the price of the cryptocurrency as part of this project. The process of data science is applied in order to obtain a more accurate model for predicting the outcome of the experiment. The primary steps in developing a reliable model are determining the relevant variables and gaining an understanding of the data. Not only is accuracy considered, but also other performance indicators like precision, recall, and score, when assessing the model's worth. These metrics are applied to the data after it has been preprocessed and a comparison is made to determine which machine learning algorithm performed the best. In order to forecast the outcome of the cryptocurrency, a machine learning model is utilised.”

Keywords: Crypto, RainForest, CNN, Machine Learning, Deep Learning, etc.

1. INTRODUCTION

‘The field of data science is an interdisciplinary field that makes use of scientific methods, procedures, algorithms, and systems to extract value from data in a wide variety of contexts. This includes, but is not limited to, the retrieval of information and findings from both structured and unstructured data.’ Peter Naur first suggested using the term "data science" in 1974 as a different name for the field of computer science. This is where the term "data science" can be traced back to. The definition was still subject to change. This paper examines the cryptocurrency forecasting approaches now in use in an effort to identify the shortcomings of existing tactics including ANNs, Bayesian regression, and SVMs. When presented with fresh data, supervised machine learning models like ANN and SVM may categorise it. The

foundation of an ANN is a group of synthetic neurons. The key problem with the model ANN was dealing with over-fitting data, handling several parameters at once, whereas SVM is concentrated on data classification. As a result, predicting accuracy was poor. This paper examines the cryptocurrency forecasting approaches now in use in an effort to identify the shortcomings of existing tactics including ANNs, Bayesian regression, and SVMs. When presented with fresh data, supervised machine learning models like ANN and SVM may categorise it. The foundation of an ANN is a group of synthetic neurons. The key problem with the model ANN was dealing with over-fitting data, handling several parameters at once, whereas SVM is concentrated on data classification. As a result, predicting accuracy was poor.

2. METHODOLOGY

‘Syntax and Semantics’:

Python's goal is to be an approachable programming language. It has a straightforward design, and it employs English terms where other people may put punctuation. It is not like many other languages in that it does not utilise curly brackets to separate blocks of code, and although though semicolons are technically allowed following statements, their use is quite uncommon. Instead, it uses a different method. When compared to C and Pascal, it features a lower number of syntactic exceptions and special instances.

‘Indentation’:

“Main article: Python syntax and semantics & Indentation”

Instead of using keywords or curly brackets to delimit blocks, Python makes use of the indentation provided by whitespace. After specific statements, the amount of indentation should be increased; when the amount of indentation should be decreased, it indicates that the current block has reached its conclusion. Thus, the semantic structure of the code is precisely reflected in the program's visual structure. This feature, found also in other languages, is frequently known as the off-side rule [1]; however, indentation does not have any meaningful connotation in most languages. The indent size of four spaces is the standard recommendation.

“Statements and control flow”:

‘Python's statements include’:

Assigning a value to a variable is what the equals sign (=) represents in an assignment statement. The conditional execution statements (elif, else, and if) and the if statement. One such expression is the for statement, that loops around an iterable object and stores its contents in a local variable. The while statement keeps running a piece of code as long as the condition it checks remains true. The Try statement, which causes any exceptions to be caught by the except clauses linked to it and guarantees that any cleaning code in a finally block will be run no matter how the code block that contains it terminates. The raise statement is used to throw an exception or to re-throw an exception that has already been caught. In object-oriented programming, a block of code can be executed and associated with a class using the class statement. Creates a method or function definition using the def statement. The break statement

can be used to exit a while loop. This iteration is bypassed due to the continue statement. Since the Del command removes a variable, it can't be used without producing an error. It is possible to reassign a variable that was previously removed. A pass statement for the NOP. Syntactically, a code block must be empty. Assert checks for conditions when debugging. An output is produced by a generator function and returned by the yield statement. The co-routine form. Function return statement. The import statement imports module functions and variables into the current programme. The assignment operator (=) is used to affix a given name to a dynamically allotted object. Variables can be redirected to any object. Python variable names are generic reference holders with no fixed data type. A variable refers to a typed object at a specific time. Programming languages that are statically typed, as contrast to dynamically typed languages, restrict variables to only storing values of a single type. Guido van Rossum says Python won't allow tail call optimization or first-class continuations. [80] [81] Python's generators give improved co-routine-like capability.

2.1 Random Forest Algorithm:

‘Random forests, often called random decision forests, are a type of supervised learning technique that may be applied to a variety of tasks including classification, regression, and others. In order for them to function properly, they must first construct a large quantity of decision trees during the training phase. After that, the output class is determined by selecting the category that corresponds to the mode of the classes (when used for classification) or the mean prediction (when used for regression) of the individual trees. This technique is also known as random choice forests and random forests, to name just a couple of its many other aliases. Random choice forests are preferable to decision trees due to the fact that they do not suffer from the issue of overfitting to the training data.’

2.2 Working of Random Forest

“The Random Forest algorithm is a type of supervised machine learning that is heavily reliant on ensemble learning. This type of machine learning may be found in the field of artificial intelligence. Training in the format of ensemble learning refers to a type of learning in which several instances of the same algorithm or various kinds of algorithms are combined to generate a more reliable prediction model. This type of learning is sometimes referred to as "machine learning." The random forest algorithm creates what is known as a "forest of trees" as a result of mixing many algorithms of the same kind, or multiple decision trees; this is how the method got its name, "Random Forest." The random forest technique can be applied to solve problems involving regression as well as classification.”

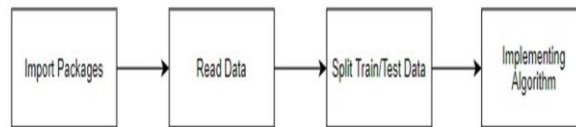
- From the given dataset, choose N random records
- ‘Construct a decision tree using these N records as the input.’
- ‘Repeat steps 1 and 2 until you reach the amount of trees that you wish to include in your algorithm.’

‘If there is a regression problem, every tree in the forest makes a guess at Y whenever a new record is added (output). The median of all the numbers predicted by the individual trees in the

forest may be used to come up with a conclusion figure. If there is a problem with categorization, every tree in the forest will estimate which class the newly collected data falls under based on the classifications of its neighbours. After tallying up the votes, the new record will be filed under the most popular category.

“MODULE DIAGRAM”

“GIVEN INPUT EXPECTED OUTPUT”



“Input: Data

Output: Getting accuracy”

3. SYSTEM ARCHITECTURE

Overview of the system

Every cryptocurrency's all-time open, high, low, and closing prices, as well as trading volume and market capitalization data. To make the code compatible with CRAN, I had to go through it line by line, and as a result, there have been major improvements made to the way that some of the field conversions have been carried out, as well as to the data that has been cleaned. The formatting of numbers or the unusual processing of scientific notations should no longer be a problem as a result of this change.

3.1 Objectives

‘The purpose of this project is to develop a supervised machine learning predictive model of crypto currency, with the hope that this model would eventually replace the existing supervised machine learning categorization methods in terms of accuracy.

3.2 Scope of the Project

The term "cryptocurrency" refers to a form of online currency that takes the form of digital assets and is transacted using cryptographic processes. This type of currency is based on the internet. Using blockchain technology, cryptocurrencies are able to achieve decentralisation, transparency, and immutability in their transactions. The task of determining the accuracy will dominate the majority of the work for this project.

“Feasibility study:”

“Data Wrangling”

“In this part of the report, the data will be loaded, the cleanliness of the dataset will be checked, and then the dataset will be trimmed and cleaned so that it can be analysed. Ensure that the actions are thoroughly documented, and that cleaning decisions are adequately justified.”

Data collection

“The data set that was collected for the purpose of making predictions based on given data is divided into two parts: the Training set and the Test set.”As a rule of thumb, the Training set and the Test set are split in a 7:3 ratio [9].”The Data Model is applied to the Training set, which was created using the “Random Forest, logistic, and Decision Tree methods, and the Support vector classifier (SVC)”. Then, depending on the accuracy of the test results, a prediction is made on the Test set.

Preprocessing

The data that was collected could have values that were not recorded, which could result in inconsistency. In order to obtain better results, the data must first be preprocessed, which will also improve the algorithm's overall efficiency. Both the outliers and the variables need to be converted, and the first step is to get rid of the outliers.

Building the Regression model

The estimation of the price of the cryptocurrency market, the following are some of the reasons why a prediction model based on the Random Forest Algorithm is effective: The findings of the regression problem are improved as a result.

Construction of a Predictive Model

Gathering data and having a significant amount of historical data are both requirements for machine learning. The data collection should include sufficient historical data as well as raw data. Raw data cannot be used for anything without first being processed in some way. It is utilised in the preprocessing stage of whatever type of algorithm is being modelled. Training and testing this model to ensure it works correctly and makes accurate predictions with a minimum of errors [2]. A tuned model involves making adjustments from time to time in order to achieve greater accuracy.

4. HARDWARE AND SOFTWARE

REQUIREMENTS

Hardware Requirements:

The hardware of a computer, sometimes known as the physical computer resources, is the most typical collection of requirements specified by any software application. For this project, the following hardware is needed:

“Processor	: Pentium IV/III
Hard disk	: minimum 80 GB
RAM	: minimum 2 GB”

Software requirements:

Software requirements are concerned with laying out what a piece of software needs in terms of hardware and other resources before it can be installed on a computer and work properly. This can be done by listing the software prerequisites. The majority of the time, these prerequisites are not included in the installation package for the application and must be installed independently prior to being able to install the product. For this project, the following software prerequisites are necessary:

Operating System	: Windows
Tool	: Anaconda with Jupyter Notebook.

5. RESULTS AND DISCUSSIONS

We are currently fitting our testing data into our trained model in the form of x train and y train as a step similar to training dataset in order to make predictions utilising our model.

The length of the data and the number of epochs had a significant influence on the testing outcomes after our training data provided a range of findings. For instance, the outcomes would be as follows if we changed the dataset for the model.

‘Since the data is divided into training data and data used for prediction, the close market price and the test data are included. ‘The Rain Forest model is trained using test data obtained over a long period of time in order to preserve its correctness. A results study shows that as the size of the dataset increases, both models perform more accurately. ‘The first step in the analytical process was the preparation and processing of the data, followed by the analysis of “missing value, exploratory analysis, model creation, and model evaluation. ‘On the open-ended test” the participant with the best prediction score will be identified. This programme can assist in determining the price of the cryptocurrency market. The model may produce more patterns, and new data can be used to fine-tune the layer weights. The model's ability to accurately anticipate the closing market rate of any given cryptocurrency is demonstrated by the fact that there is a very small difference between the projected data and the actual data. The vast amount of data is normalised to speed up training while keeping the model’s accuracy.

6. CONCLUSION

Researchers and academics have recently shown an interest in blockchain technology as well as machine learning. Only a few instances of how they are used in the actual world are self-driving vehicles and widely used, anonymous cryptocurrency-based payment systems. The aim of this study is to improve the knowledge of academics and practitioners on the use of well-known time-series prediction models on cryptocurrencies. The research results discussed in

this thesis have a number of significant ramifications. First, for predicting the future prices of cryptocurrencies, and to a greater extent, any financial asset, and sophisticated models are not always better than a simple persistence model. The ARIMA model performs second best overall based on mean squared error rate. “The LSTM has the highest accuracy in classification at 52% and the lowest RMSE at 8%.”[7]. LSTM and other neural network-based models fare better in predicting the direction of price movement. Overall, however, the Rain Forest technique combined with machine learning and data science yields the greatest accuracy for both short- and long-term prediction.

For further investigation, this work might be expanded in a variety of directions. Cryptocurrency intra-day values may be utilised to employ additional data and give researchers new ideas in this field. Additionally, future study can concentrate on the use of other machine learning techniques not included in this paper. On a public test set, the highest-scoring accuracy score will be selected.”The price of a cryptocurrency may be determined with the help of this application [5].

REFERENCE

1. Neha Mangla, Akshay Bhat , Ganesh Avabratha Narayana Bhat, Bitcoin Price Prediction Using Machine Learning (May 2019)
2. AUTOMATED CRYPTOCURRENCIES PRICES PREDICTION USING MACHINE LEARNING submitted by Ruchi Mittal, Shefali Arora and M.P.S Bhatia, DOI: 10.21917/ijsc.2018.0245
3. Lekkala Sreekanth Reddy, Dr.P. Sriramy, A Research On Bitcoin Price Prediction Using Machine Learning Algorithms (April 2020)
4. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning, Franco Valencia, Alfonso Gómez-Espinosa and Benjamín Valdés-Aguirre (14 June 2019)
5. Modeling and Prediction of Cryptocurrency Prices using Machine Learning Techniques submitted by Alireza Ashayer (April 2019)
6. Predicting the Price of Bitcoin Using Machine Learning| Sean McNally ; Jason Roche ; Simon Caton; Ireland, Dublin , IEEE 2018
7. Bitcoin Volatility Forecasting with a Glimpse into Buy and Sell Orders| Tian Guo ; Albert Bifet ; Nino Antulov- Fantulin ; IEEE 2018
8. F. Andrade de Oliveira, L. Enrique ZÃ¡rate and M. de Azevedo Reis; C. NeriNobre, —The use of artificial neural networks in the analysis and prediction of stock prices,| in IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2151-2155.
9. M. Daniela and A. BUTOI, —Data mining on Romanian stock market using neural networks for price prediction. informaticaEconomica, 17,2013.
10. Bitcoin Cost Prediction using Deep Neural Network Technique| Kalpanasonika ,Sayasri S , Vinothini , SugaPriya , IEEE 2018
11. D. Yermack, “Is Bitcoin a Real Currency?”, The Handbook of Digital Currency, 2013.

12. A. Badev and M. Chen, "Bitcoin: Technical Background and Data Analysis", Available at: <https://www.federalreserve.gov/econresdata/feds/2014/files/2014104pap.pdf>.
13. M. Iwamura, Y. Kitamura and T. Matsumoto, "Is Bitcoin the Only Cryptocurrency in the Town? Economics of Cryptocurrency and Friedrich A. Hayek", Discussion Paper Series, Institute of Economic Research, Hitotsubashi University, pp. 1-13, 2014.
14. C. Gunduz and B. Donald, "A Theoretical Foundation for Technical Analysis", Journal of Technical Analysis, Vol. 59, No. 5, pp. 1-22, 2003.
15. L. Andrew, H. Mamaysky and J. Wang, "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation", Journal of Finance, Vol. 55, No. 4, pp. 1705-1765, 2000.