CRYPTOCURRENCY MARKET PRICE PREDICTIONUSING DATA SCIENCE PROCESS

1. HARITHA P

Assistant Professor, Dept of IT

Panimalar Engineering College

Chennai, India.

haritha81997@gmail.com

2. ANUSHRI C

Student, Dept of IT

Panimalar Engineering College

Chennai, India.

anushrichandran310@gmail.com

3. NANDHINI H

Student, Dept of IT

Panimalar Engineering College

Chennai, India.

nandhiniharibabus@gmail.com

4. PREETHI S

Student, Dept of IT

Panimalar Engineering College

Chennai, India.

preethiselvaraj12102000@gmail.com

Abstract - Cryptocurrency is a digital currency wherein the coin ownership records are stored in a ledger existing in a form of a computerized database using strong cryptography to secure transaction records and control the creation of additional coins and verify the transfer of coin ownership. Nowadays cryptocurrencies are used on large scale and there is a sudden rise and or decrease in their share and it is difficult to predict the price of the cryptocurrency. In this project, a machine learning model is built to predict the price of the cryptocurrency. The application of data science process is applied for getting a better model for predicting the outcome. Variable identification and data understanding is the main process of building a successful model. Different machine learning algorithms are applied to the pre-processed data and the accuracy is compared to see which algorithm performed better other performance metrics like precision, recall, and score are also taken into consideration for evaluating the model. The Machine learning model is used to predict the cryptocurrency outcome.

Keywords – Cryptocurrency, Random Forest Algorithm, Decision Tree, LASSO Algorithm, Regression Tree Algorithm, Data Science, Machine Learning.

I. INTRODUCTION

A. Objective of the project

The goal is to develop a machine learning model for Crypto Currency Prediction, to potentially replace the updatable supervised machine learning classification models by predicting leads to the shape of best accuracy by comparing supervised-algorithms.

B. Project Goals

- Exploration data analysis of variable identification
- Univariate data analysis
- Exploration data analysis of bi-variate and multi-variate
- Method of Outlier detection with feature engineering
- Comparing algorithms to predict the result

C. Scope Of The Project

Cryptocurrency could be a network-based medium of exchange within the style of digital assets and currencies that uses cryptographic functions to conduct financial transactions. Cryptocurrencies use block chain technology to realize decentralization, transparency, and immutability. The

most scope of the project is to search out the accuracy and find results from the flask framework deployment.

II. EXISTING SYSTEM

In other related works, the users presented a computational approach for identifying and characterizing crypto currency pump and dump operations that are disbursed on social media. they'd used financial and Twitter data per a selected coin, the strategy used was ready to detect, with reasonable accuracy, whether there's an unfolding attack on it coin on Telegram, and whether or not the resulting pump operation will reach terms of meeting the anticipated price targets. They also analyzed the activities of users involved in pump operations, and observe a prevalence of Twitter bots in cryptocurrency-related tweets in close proximity to the attack. Telegram was a well-liked choice for scammers to prepare and coordinate pump and dump operations. To analyze such activities, allow us to define the subsequent two notions one was a Pump Attempt which is the act of targeting a coin on Telegram by posting a pump message mentioning a "pump attempt". The second was Successful Pump Attempt which is that the pump attempt is successful if the particular price approaches the target price within a time window after the primary pump message has been posted.

Disadvantages

- They didn't use any specific algorithms for predicting the crypto-currency price
- The data they found is only accustomed deliberate to see the sole impact of social media.

III. PROPOSED SYSTEM

Cryptocurrency behaves differently and it's a little difficult to predict the chance. The proposed model ito s builds a model where the model is ready to predict the worth. The steps involved within the proposed model may be a process employed in data science from variable identification to putting together a model. the method starts from variable identification like dependent and experimental variables where we discover the target column.

Then the pre-processing techniques are applied coping with the missing values the pre-processed data is then accustomed to building a model by dividing the Infoset into a 7:3 ratio where 70% of the data is employed for training purposes that's model learns the pattern and therefore the remaining testing data is used to test the performance of knowledge. The regression model can also be accustomed predict the value of the cryptocurrency.

Advantages

- The machine learning algorithms are compared and also the performance metric also are calculated for better prediction.
- Machine learning model predictions allow businesses

to create highly accurate guesses.

IV. REVIEW OF LITERATURE SURVEY

- **Title :** A Research On Bitcoin Price Prediction Using Machine Learning Algorithms

  Author : Lekkala Sreekanth Reddy, Dr.P.Sriramya

  Year : 2020

In this paper, the researchers proposed to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. By gathering information from different reference papers and applying it in real time,I found the advantages and disadvantages of bitcoin price prediction. Each and every paper has its own set of methodologies for bitcoin price prediction. Many papers have an accurate price but some other don't, but the time complexity is higher in those predictions, so to reduce the time complexity here in this paper we use an algorithm linked to an artificial intelligence named LASSO(least absolute shrinkage selection operator. The other papers used different algorithms like SVM(support vector machine),coinmarkupcap, Quandl, GLM, CNN(Convolutional Neural Networks)and RNN(Recurrent neural networks), etc.. which do not have great time management, but in LASSO finding of the results from a larger database is quick and fast..so for this purpose, we draw a comparison between other algorithms and the LASSO algorithm, this survey paper helps the upcoming researchers to make an impact in their papers. The process that happens in the paper is the first moment of the research, we aim to understand and find daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. Our data set consists of various features relating to the Bitcoin price and payment network over the course of every year, recorded daily. By preprocessing the dataset, we apply the same data mining techniques to reduce the noise of data. Then the second moment of our research, using the available information, we will predict the sign of the daily price change with the highest possible accuracy.

- **Title:** Bitcoin Price Prediction Using Machine Learning

  Author: Neha Mangla

  Year: 2019

In this paper, the researchers tried to estimate the Bitcoin price precisely taking into consideration various parameters that affect the Bitcoin value.In our work, we pointed to understand and identify daily changes in theBitcoin market while obtaining to most sight intomost appropriatefeatures surrounding Bitcoin price. We will predict the daily price e change with highest possible accuracy. The market capitalization of publicly traded cryptocurrencies is currently above $230 billion.Bitcoin, the most valuable cryptocurrency, serves primarily as a digital store

of value, and its price predictability has been well-studied. These characteristics are outlined in the following subsection; the underlying details of Bitcoin,as they are described in depth in thecited papers.

- **Title:** Modeling and prediction of cryptocurrencyprices using machine Learning techniques

  Author: Alireza Ashayer

  Year: 2019

The contents of this chapter have been submitted to IEEE International Conference on Blockchain. The motivation for this review is to understand the trend of Blockchain research with respect to the machine learning field by studying and reviewing published articles. This understanding can help other researchers and practitioners with insight into the current state and future direction of research in this field. Given this motivation, the researcherswill review and verify the distribution of research papers bytheir year of publication and classify the research papers by the machine learning techniques used.

- **Title:** Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning

  Author: Franco Valencia, Alfonso Gómez- Espinosa and Benjamín Valdés-Aguirre

  Year:2019

Cryptocurrencies are becoming increasinglyrelevant in the financial wd and can be considered as an emerging market. The low barrier of entry and high data availability of the cryptocurrency market makes it an excellent subject of study, from which it is possible to derive insights into the behavior of markets through the application of sentiment analysis and machine learning techniques for the challenging task of stock market prediction. In this paper, the researchers propose the usage of common machine learning tools and available social media data for predicting the price movement of the Bitcoin, Ethereum, Ripple and Litecoin cryptocurrency market movements. The researchers compare the utilization of neural networks (NN), support act, or machines (SVM) and random forest (RF) while using elements from Twitter and market data as input features. The results show that it ispossible to predict cryptocurrency markets using machine learning and sentiment analysis.

- **Title:** Automated cryptocurrencies prices prediction using machine learning

  Author: Ruchi Mittal, Shefali Arora, and M.P.S Bhatia

  Year: 2018

Cryptocurrency is one of the trending areas of research among researchers. Many researchers may analyze the cryptocurrency features in several ways such as market price prediction, the impact of cryptocurrency in real life,

and so on. In this paper, the researchers focus on the market price prediction of the number of cryptocurrencies based on their historical trends. Their dataset consists of over nine features relating to the cryptocurrency price recorded daily over the period of 6 months. They applied some machine-learning algorithms to predict the daily price change of cryptocurrencies.

LIST OF MODULES

- Data Pre-processing
- Data Analysis of Visualization
- Comparing Algorithm with prediction within the typeof best accuracy result
- Deployment Using Flask

Modules description

Module – 1:

Data Pre-Processing

Validation techniques in machine learning are accustomed get the error rate of the Machine Learning (ML) model, which might be considered as near actuality error rate of the dataset. If the information volume is large enough to be representative of the population, you will not need the validation techniques. However, in real-world scenarios, to figure with samples of knowledge which will not be a real representative of the population of a given dataset. to seek out the missing value, duplicate value, and outline of information type whether it's float variable or integer. The sample of information is employed to produce an unbiased evaluation of a model fit on the training dataset while tuningmodel hyperparameters.

During the method of knowledge identification, it helps to grasp your data and its properties; this information will facilitate your choose which algorithm to use to create your model. variety of various data cleaning tasks use Python's Pandas library and specifically, it focuses on probably the most important data cleaning task, missing values and it's ready to more quickly clean data. It wants to spend less time cleaning data, and longer exploring and modeling. a number of these sources are just simple random mistakes. Other times, there may be a deeper reason why data is missing. It's important to know these differing kinds of missing data from a statistics point of view. the kind of missing data will influence the way to cope with filling within the missing values and detect missing values, and do some basic imputation and detailed statistical approaches for addressing missing data. Before, joint into code, it's important to grasp the sources of missing data. Here are some typical reasons why data is missing:

- User forgot to fill during a field.
- Data is lost while transferring manually from a legacy database.
- There was a software error.
- Users chose to not fill out a field tied to their beliefs about how the results would be used or interpreted.

Data Validation/ Cleaning/Preparing Process

Importing the library packages with loading the given dataset. to research the variable identification by data shape, and data type and evaluate the missing values, and duplicate values. A validation dataset may be a sample of knowledge held back from training your model that's wont to give an estimate of model skill while tuning models and procedures that you just can use to create the most effective use of validation and test datasets when evaluating your models. Data cleaning/preparing by renaming the given dataset and dropping the column etc. to analyze the uni-variate, bi-variate and multi-variate processes. The steps and techniques for data cleaning and correction will vary from every dataset to dataset. The primary goal of data cleaning and correction is to detect and remove errors and anomalies to increase the value of information in analytics and decision-making.

Module – 2:

Exploration data analysis of visualization

Data visualization is a crucial skill in applied statistics and machine learning. Statistics does indeed target quantitative descriptions and estimations of knowledge. this will be helpful when exploring and aiming to know a dataset and might help with identifying patterns, corrupt data, outliers, and far more. Sometimes data doesn't add up until it can take a look at in a very visual form, like with charts and plots. having the ability to quickly visualize knowledge samples et al. is a crucial skill both in applied statistics and in applied machine learning. it'll discover the numerous kinds of plots that you just will have to know when visualizing data in Python and the way to use them to higher understand your own data.

- How do chart statistic data with line plots and categorical quantities with bar charts?
- How do summarize data distributions with histograms and box plots?

To achieve better results from the applied model in the Machine Learning method the info should be in a proper manner. Some specified Machine Learning model needs information in an exceedingly specified format, for instance, the Random Forest algorithm doesn't support null values. Therefore, to execute a random forest machine learning algorithm, null values must be managed from the initial information set.

V. ENVIRONMENTAL REQUIREMENTS

Software Requirements:
Operating System : Windows

Tool   : Anaconda with Jupyter Notebook

Hardware requirements:

Processor   : Pentium IV/III
Hard disk   : minimum 80
GBRAM   : minimum 2 GB

VI.   SOFTWARE DESCRIPTION

Anaconda may be a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale processing, predictive analytics, etc.), that aims to simplify package management and deployment. The Anaconda distribution is employed by over 12 million users and includes quite 1400 popular data- science packages suitable for Windows, Linux, and MacOS. So, Anaconda distribution comes with over 1,400 packages furthermore because the Conda package and virtual environment manager called Anaconda Navigator and it eliminates the requirement to find out to put in each library independently. The open source packages may be individually installed from the Anaconda repository with the conda install command or using the pip install command that is installed with Anaconda. Pip packages provide many of the features of conda packages and in most cases they'll work together. Custom packages will be made using the conda build command and may be shared with others by uploading them to Anaconda Cloud, PyPI or other repositories. However, you'll be able to create new environments that include any version of Python packaged with conda.

Features Of Anaconda Navigator

Anaconda Navigator may be a desktop graphical interface (GUI) included in Anaconda distribution that permits you to launch applications and simply manage conda packages, environments, and channels without using command-line commands. Navigator can rummage around for packages on Anaconda.org or in an exceedingly local Anaconda Repository.Anaconda could be a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale processing, predictive analytics,etc.), that aims to simplify package management and deployment. Navigator is a simple, point-and-click thanks to work with packages and environments with no need to type conda commands in an exceedingly terminal window. you'll use it to search out the packages you would like, install them in an environment, run the packages, and update them – all inside Navigator.The following applications are available by default in Navigator:

- JupyterLab
- Jupyter Notebook
- Spyder
- PyCharm
- VSCode
- Glueviz
- Orange 3 App
- RStudio
- Anaconda Prompt (Windows only)
- Anaconda PowerShell (Windows only)

VII. CONCLUSION

The systematic process started from data cleaning and processing, missing value, exploratory analysis and eventually model building and evaluation. the simplest accuracy on public test data is taken into consideration and better accuracy score for the given test data is implemented within the machine learning model. This application can help to search out the Cryptocurrency market value. This project will improve that future idea of crypto currencies and it can even improve the market value of the cryptocurrency. during this paper we proposed a machine learning algorithm to search out the value of the given Cryptocurrency and even calculated the accuracy of assorted different Machine learning model. This application will help the users get more profit and also help the users to improve their trading ability.

VIII.   FUTURE ENHANCEMENTS

- Cryptocurrency market value prediction to attachwith AI model.
- To automate this process by show the predictionlead to web application or desktop application.
- To optimize the work to implement in AI environment.
- Show the statistical trend in cryptocurrency for 1week.
- Automate this process and show the predictionend in android application or IOS application.
- Even improve the result using computer science environment.
- Add more coins and predict its value.

REFERENCES

[1] Zhang Y, Chan S, Chu J, Sulieman H (2020) On the market efficiency and liquidity of high-frequency cryptocurrencies in a bull and bear market.

[2] Lekkala Sreekanth Reddy, Dr.P. Sriramya (2020), A Research On Bitcoin Price Prediction Using Machine Learning Algorithms.

[3] Neha Mangla (2019), Bitcoin Price Prediction Using Machine Learning.

[4] Alireza Ashayer (2019), Modeling and prediction of cryptocurrency prices using machineLearning techniques.

[5] Franco Valencia, Alfonso Gómez-Espinosa and Benjamín Valdés-Aguirre (2019), Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and MachineLearning.

[6] Mallqui DC, Fernandes RA (2019) Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques.

[7] Ji Q, Bouri E, Lau CKM, Roubaud D (2019), Dynamic connectedness and integration in cryptocurrency markets.

[8] Ruchi Mittal, Shefali Arora and M.P.S Bhatia (2018), Automated cryptocurrencies prices prediction using machine learning.

[9] Vo.A, Yost-Bremm C (2018) A high-frequency algorithmic trading strategy for cryptocurrency.

[10] Thomas Gunter Fisher, Christopher Krauss and Alexander Deinert(2017), Statistical Arbitrage in Cryptocurrency Markets.