ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 9, Issue 8, August 2022

Data Science Technique Applied to Crypto Currency Market Price Prediction

Mrs Savithri S
Department of MCA
Dhanalakshmi Srinivasan College of Engineering and Technology

Ms. Charitha P
Department of MCA
Dhanalakshmi Srinivasan College of Engineering and Technology

Abstract: Crypto currency 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 to control the creation of additional coins, and to verify the transfer of coin ownership. Nowadays crypto currency are used in large scale and there is a sudden rise or decrease in their share and it is difficult to predict the price of the crypto currency. In this project a machine learning model is built to predict the price of crypto currency. The application of data science process is applied for getting the better model for predicting the outcome. Variable identification and data understanding is the main process in building the successful model. Different machine learning algorithms are applied on the pre-processed data and the accuracy are compared to see which algorithm performed better other performance metrics like precision, recall, score are also taken in consideration for evaluating the model. The machine learning model is used to predict the crypto currency outcome.

I. INTRODUCTION

They presented a computational approach for identifying and characterizing crypto currency pump and dump operations that are carried out in social media. They had used financial and Twitter data pertaining to a particular coin, the method used was able to detect, with reasonable accuracy, whether there is an unfolding attack on that coin on Telegram, and whether or not the resulting pump operation will succeed in terms of meeting the anticipated price targets. They also analysed 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 popular choice for scammers to organize and coordinate pump and dump operations. The second was Successful Pump Attempt which is the pump attempt is successful if the actual price approaches the target price within a time window after the first pump message has been posted.

Disadvantages:

1. They did not use any specific algorithms for predicting the crypto-currency price
2. The data which they found are only used to consider to see the only the impact from social media

II. PROPOSED SYSTEM

Cryptocurrency behaves differently and it is little difficult to predict the possibility. The proposed model is build a model where the model is able to predict the price. The steps involved in the proposed model is a process used in data science from variable identification to building a model. The process start from variable identification like dependent and independent variable where we find the target column.

Then the pre-processing techniques are applied like dealing with the missing values the pre-processed data then used to build a model by dividing the dataset into 7:3 ratio where 70% of the data is used for training purpose that is model learns the pattern and the remaining testing data is used to test the performance of data.

Advantages:

1. The machine learning algorithms are compared and the performance metric are also calculated for better prediction.
2. Machine learning model predictions allow businesses to make highly accurate guesses.

III.INTRODCUTION MACHINE LEARNING

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

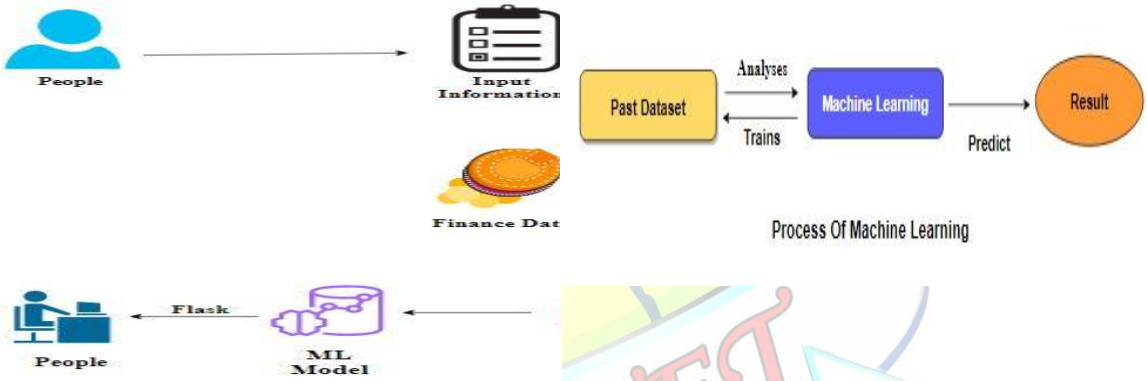
International Journal of Advanced Research Trends in Engineering and Technology

Vol. 9, Issue 8, August 2022

with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine

Process of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. Machine learning can be roughly separated in to three categories. There are supervised learning, unsupervised learning and reinforcement learning.

SYSTEM ARCHITECTURE



LIST OF MODULES

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

IV. MODULE DESCRIPTION

MODULE - 1

Data Pre-processing

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

Data Validation/ Cleaning/Preparing Process

Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine.

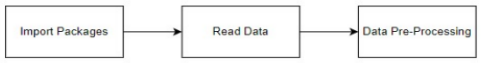
Preparing the Dataset :

This dataset contains 1079 records extracted, which were classified into 4 classes:

- Bitcoin
- Ethereum

Importing the library packages with loading given dataset. To analyzing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A validation dataset is a sample of data held back from training your model that is used to give an estimate of model skill while tuning model's and procedures that you can use to make the best use of validation and test datasets when evaluating your models.

MODULE DIAGRAM:



GIVEN INPUT EXPECTED OUTPUT

input : data

output : removing noisy data

MODULE – 2:

Exploration data analysis of visualization

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 9, Issue 8, August 2022

Sometimes data does not make sense until it can look at in a visual form, such as with charts and plots. Being able to quickly visualize of data samples and others is an important skill both in applied statistics and in applied machine learning.

MODULE DIAGRAM:

GIVEN INPUT EXPECTED OUTPUT
input : data
output : visualized data

- identify the patterns from given dataset.

• Data manipulation can be done easily with data frames

MODULE – 3:

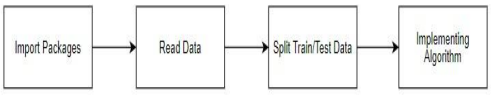
Logistic Regression

The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables.

Linear Regression:

Linear Regression is a machine learning algorithm based on supervised learning. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output).

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT
Input : Data
Output : Getting Accuracy

MODULE – 4:

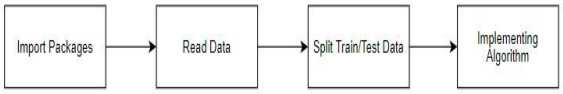
Decision Tree Regression:

Decision Tree - Regression. Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

Random Forest Regression:

Random Forest Regression is a supervised learning algorithm that uses **ensemble learning** method for regression.

MODULE DIAGRAM:



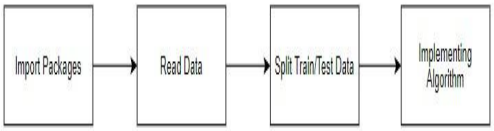
GIVEN INPUT EXPECTED OUTPUT
Input : Data
Output : Getting Accuracy

MODULE – 5:

Support Vector Regression:

Support Vector Machines (SVM) are popularly and widely used for classification and regression problems in machine learning.

MODULE DIAGRAM:



GIVEN INPUT EXPECTED OUTPUT
input : data
output : getting accuracy

V. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This application can help to find the Cryptocurrency Market Price.