CRYPTOCURRENCY MARKET PRICE PREDICTION

Dr J. Jospin Jeya –Associate Professor
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Harisha P
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Kiruthiga UL
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Abstract— Due to the rapid development of information and communication technologies, many activities in our daily life have been merged online and they become more flexible and more effective. A huge growth in number of online users has activated virtual word concepts and created a new business phenomenon which is cryptocurrency to facilitate the financial activities such as buying, selling and trading. Cryptocurrency represent valuable and intangible objects which are used electronically in different applications and networks such as online social networks, online social games, virtual worlds and peer to peer networks. The use of virtual currency has become widespread in many different systems in recent years.

Cryptocurrency, an encrypted , peer-to-peer network for facilitating digital barter, is a technology developed eight years ago. Bitcoin, the first and most popular cryptocurrency, is paving the way as a disruptive technology to long standing and unchanged financial payment systems that have been in place for many decades. Cryptocurrencies may revolutionize digital trade markets by creating a free flowing trading system without fees.

INTRODUCTION

One of the fields that benefit from these technologies and online connections is the financial and business sector. A growing number of online users has activated virtual world concepts and created a new business phenomena. Thus, new types of trading, transactions and currencies have been arising. One of the remarkable financial forms that have been emerged in the past few years is Cryptocurrency. Cryptocurrencies represent valuable and intangible objects which can be used electronically or virtually in different applications and networks such as online social

networks, online social games, virtual worlds and peer to peer networks.

A cryptocurrency is a digital asset designed to work as a medium of exchange wherein individual coinownership records are stored in a ledger existing in a form of a computerized database using strongcryptography to secure transaction records, to control the creation of additional coins, and to verifythe transfer of coin ownership. Cryptocurrency does not exist in physical form (like paper money) andis typically not issued by a central authority. Cryptocurrencies typically use decentralized control asopposed to a central bank digital currency (CBDC)..

SYSTEM ARCHITECTURE
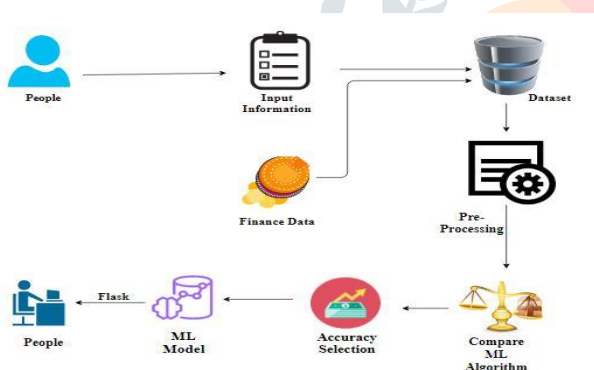
IMPLEMENTATION

- **Data Collection**
  - Data cleaning
  - Data transformation
  - Data selection
- **Pre-processing**
- **Training Dataset**
- **Testing Dataset**
- **Algorithm**

- **Evaluation**
- **Result**

Collecting Dataset

- Data Collection is one of the most important tasks in building a machine learning model.We collect the specific dataset based on requirements from internet.

- The dataset contains some unwanted data also. So first we need to pre-process the data and obtain perfect data set for algorithm.

Pre-processing

- However, some of the data may be noisy, i.e. may contain inaccurate values, incomplete values or incorrect values.Hence, it is must to process the data before analysing it and coming to the results.

Pre-processing techniques

- Data cleaning includes Fill in missing values, smooth noisy data, identify or remove outliers, and resolve in consistencies.

- Data transformation may include smoothing,aggregation, generalization, transformation which improves the quality of the data.

Training Dataset

Training dataset is a subset of the dataset used to build and fit predictive models. A training set is generated by building a training dataset script, which generates the training set features from the input options and the raw stock price data

Testing Dataset

Testing Dataset is a subset of the dataset to assess the likely future performance of a model. It is a good standard for evaluating the model. Testing set is used against the predicted dataset and testing the model that is trained. It is used for evaluation purposes.

■ALGORITHMS

[1] Logistic Regression

It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). 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.

Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

[2]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).

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). Hence, the name is Linear Regression.Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price,** etc.

[3] 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.

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.

[4] Random Forest Regression

Is a supervised learning algorithm that uses **ensemble learning** method for regression.

Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. Similar to my previous posts, I will be using data on House Sales in King County, USA.

From the sklearn package containing ensemble learning, we import the class **Random Forest Regressor**, create an instance of it, and assign it to a variable. The *.fit()* function allows us to train the model, adjusting weights according to the data values in order to achieve better accuracy. After training, our model is ready to make predictions, which is called by the *.predict()* method.

[5]Support Vector Machine or SVM

Support Vector Machines (SVM) are popularly and widely used for classification and regression problems in machine learning. SVMs solve binary classification problems by formulating them as convex optimization problems. The optimization problem entails finding the maximum margin separating the hyperplane, while correctly classifying as many training points as possible. SVMs represent this optimal hyperplane with support vectors. The sparse solution and good generalization of the SVM lend themselves to adaptation to regression problems. SVM generalization to SVR is accomplished by introducing an ε-insensitive region around the function, called the ε-tube.
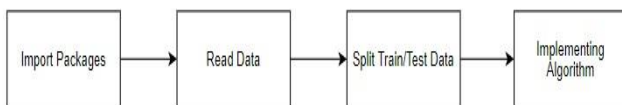
[6] Lasso Regression

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

The word "LASSO" stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. It is a statistical formula for the regularisation of data models and feature selection.

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean.

The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

BLOCK DIAGRAM

OUTPUTS:
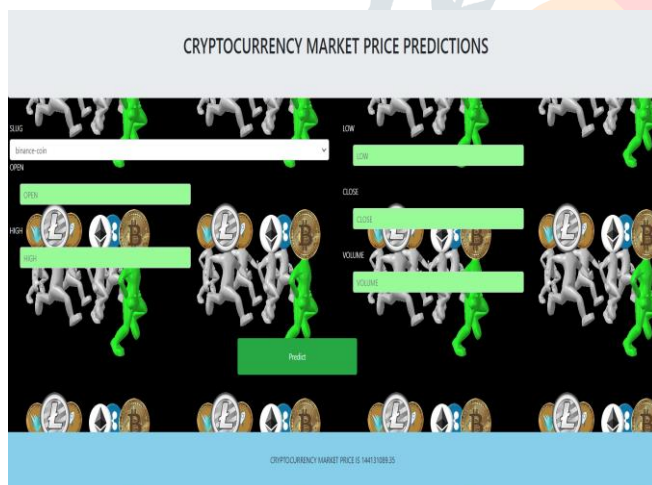
LASSO REGRESSION:

```
MEAN ABSOLUTE ERROR VALUE IS : 1365308447.269407

MEAN SQUARED ERROR VALUE IS : 8.717263191604509e+18

MEDIAN ABSOLUTE ERROR VALUE IS : 638446439.0957284

ACCURACY RESULT OF LASSO REGRESSION IS : 60.86985515990069

R2_SCORE VALUE IS : 0.6076703856956052
```

DECISION TREE:

```
MEAN ABSOLUTE ERROR VALUE IS : 157224058.3425926

MEAN SQUARED ERROR VALUE IS : 3.103164707620382e+17

MEDIAN ABSOLUTE ERROR VALUE IS : 23427197.5

ACCURACY RESULT OF DECISION TREE REGRESSOR IS : 98.86535214297673

R2_SCORE VALUE IS : 0.9886532194491098
```

RANDOM FOREST:

```
****************************
MEAN ABSOLUTE ERROR VALUE IS : 148472493.50012347

MEAN SQUARED ERROR VALUE IS : 1.6812837675140307e+17

MEDIAN ABSOLUTE ERROR VALUE IS : 29613494.079999566

ACCURACY RESULT OF RANDOM FOREST REGRESSOR IS : 99.20535219645542

R2_SCORE VALUE IS : 0.9920387540695182
```

RESULT AND COMPARISON

A performance analysis of our offered algorithm is demonstrated with other learning algorithms were the features are crafted by using speed up robust features.We have performed with Linear regression , Logistic regression , Decision Tree , Random forest , Lasso regression and Support Vector Machine algorithms and the accuracy we are getting is 97%.

CONCLUSION AND FUTURE WORK

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.

- ➤ Cryptocurrency Market Price prediction to connect with AI model.

- ➤ To automate this process by show the prediction result in web application or desktop application.

REFERENCES

1. C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. Conf. Weblogs Social Media (ICWSM)*, 2014, pp. 216–225. [Online]. Available: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

2. S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, 1998.

3. S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM J. Res. Develop.*, vol. 4, no. 1, pp. 66–82, Jan. 1960.