**DOGECOIN PRICE PREDICTION**

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE AND ENGINEERING**

by

**G. KARTHIK REDDY (2103A51354)**

**MD. NIZAMUDDIN (2103A51331)**

**P. MADHU (2103A51416)**

Under the guidance of

**Mr. D. RAMESH**

Assistant Professor, Department of CSE.

**

**Department of Computer Science and Artificial Intelligence**

**

**Department of Computer Science and Artificial Intelligence**

**CERTIFICATE**

This is to certify that project entitled **“DOGECOIN PRICE PREDICTION**" is the bonafied work carried out by **G. KARTHIK REDDY, MD. NIZAMUDDIN, P. MADHU** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

**Mr. D. RAMESH Dr. M. Sheshikala**

Asst. Professor, Assoc. Prof .& HOD (CSE)

S R University, S R University,

Ananthasagar ,Warangal Ananthasagar ,Warangal

**ACKNOWLEDGEMENT**

We express our thanks to Course co-coordinator **Mr.** D.Ramesh**, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude toHead of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

**ABSTRACT**

# Dogecoin is a cryptocurrency based on the picture of a Shiba Inu dog with funny phrases written in Comic Sans to symbolize the dog’s thoughts. Dogecoin is a free and open-source cryptocurrency that split off from the Litecoin project at the end of 2013.

# Dogecoin was originally conceived of by its developer, Billy Markus, as an attempt to create a humorous digital currency. He thought that humorous altcoins, rather than Bitcoin, stood a higher chance of gaining widespread acceptance. Dogecoin first gained widespread attention in 2014, but its meteoric rise in value didn’t begin until 2016.

# Dogecoin is an open-source peer-to-peer digital currency, favored by Shiba Inu worldwide. It is qualitatively more fun while being technically nearly identical to its close relative Bitcoin. This dataset contains its historical stock price in USD on a daily frequency starting from 17 September 2014.

**Table of Contents**

**Chapter No. Title Page No.**

1. Introduction
   1. Overview 6
   2. Problem Statement 6
   3. Existing system 6
   4. Proposed system 6
   5. Objectives 7
   6. Architecture 7
2. Literature survey
   * 1. Document the survey done by you 8
3. Data pre-processing

3.1. Dataset description 9

3.2. Data cleaning 9

3.3. Data Visualization 10

1. Methodology
   1. Procedure to solve the given problem 12
   2. Model architecture 14

4.3. Software description 14

1. Results and discussion 15
2. Conclusion and future scope 18
3. References 18

**INTRODUCTION**

* 1. **OVERVIEW**

Dogecoin is one of the most popular and valuable cryptocurrencies in the current financial market, attracting traders for investment and thereby opening new research opportunities for researchers. Countless research works have been performed on Dogecoin price prediction with different machine learning prediction algorithms. For the research: relevant features are taken from the dataset having strong correlation with Dogecoin prices and random data chunks are then selected to train and test the model. The random data which has been selected for model training, may cause unfitting outcomes thus reducing the price prediction accuracy. Here, a proper method to train a prediction model is being scrutinized. The proposed methodology is then applied to train a Regression model to predict the dogecoin price for the upcoming 5 days. When the LSTM model is trained with a suitable data chunk, thus identified, sustainable results are found for the prediction. In the end of this paper, the work culminates with future improvements.

**1.2. PROBLEM STATEMENT**

To develop a model which can help us to predict the price of the crypto currency used (in this case: Dogecoin), with low error rate and a high precision of accuracy. The model will not tell the future, but it might forecast the general trend and the direction to expect the prices to move.

**1.3. EXISTING SYSTEM**

Firstly, we collect the data set from the online source: Yahoo. The data set represents the Dogecoin price in United States Dollars (USD). The dataset includes all the information about bitcoin prices from 18 January,2022 to 18 January,2023. The second step involves filtering and cleaning the data set. This involves removing all the incomplete data from the rows. It also involves filtering out unnecessary features present in the data collected.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

**1.4. PROPOSED SYSTEM**

The step is training, followed by testing the dataset. We train our model, using the algorithm and the features taken into account to assist our model, to predict the future price of the crypto currency. Moving on to the testing part, we test the data to measure the accuracy of the algorithm that our model is using to predict the price of the Dogecoin.

Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which

requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge

set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure,

and blood sugar, it is possible to increase the possibility of heart diseases prediction

Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which

requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge

set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure,

and blood sugar, it is possible to increase the possibility of heart diseases prediction

Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which

requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge

set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure,

and blood sugar, it is possible to increase the possibility of heart diseases prediction

**1.5. OBJECTIVES**

The main objective of this research is to develop a model which can help us to predict the price of the crypto currency used (in this case: Dogecoin), with low error rate and a high precision of accuracy. The model will not tell the future, but it might forecast the general trend and the direction to expect the prices to move

**1.7. ARCHITECTURE**

The architecture of the proposed system is as displayed in the figure below. The major components of the architecture are as follows: Dogecoin dataset, preprocessing, tokenization, training the model, test the model, design fitness function, application of algorithm, results collection and prediction of Dogecoin disease.

Dogecoin dataset

Pre processing

tokenisation

Prediction of dogecoin price

Training model

Testing the model

Collection of result

Applying the algorithm

Design fit function of algorithm

**2.1.1 LITERATURE SURVEY**

# There has been a growing interest in the field of cryptocurrency price prediction, and Dogecoin is no exception. Several studies have been conducted to analyze the price movements of Dogecoin and forecast its future trends.

# One study by Al-Yahya and Al-Khazali (2019) used a machine learning algorithm to predict the price of Dogecoin. The study analyzed data from multiple exchanges and found that the algorithm's accuracy improved as the amount of data increased. The study concluded that machine learning can be an effective tool for predicting the price of cryptocurrencies.

# Another study by Bouri et al. (2021) examined the impact of social media sentiment on the price of Dogecoin. The study found a positive correlation between social media sentiment and the price of Dogecoin, suggesting that social media activity can be used to predict future price movements.

# A more recent study by Tsai et al. (2021) used a deep learning algorithm to predict the price of Dogecoin based on technical indicators and news sentiment analysis. The study found that the model's accuracy improved significantly when incorporating news sentiment analysis, indicating that external news factors play a crucial role in the price of Dogecoin.

# Overall, the literature suggests that machine learning and sentiment analysis can be effective tools for predicting the price of Dogecoin. However, as with any cryptocurrency, the highly volatile nature of the market makes accurate predictions challenging. As such, it is essential to consider multiple factors and approaches when attempting to forecast the future price of Dogecoin.

**3.DATA PRE-PROCESSING**

**3.1.1 DATASET DESCRIPTION**

|  |  |  |
| --- | --- | --- |
| **Sno** | **Attributes** | **Description** |
| 1.  2.  3.  4.  5. | OPEN    HIGH  LOW  CLOSE  VOLUME BTC | The opening price of the time period.  The highest price of the time period.  The lowest price of the time period.  The closing price of the time period.  This is the volume in the transacted ccy. BTC-Dogecoin |

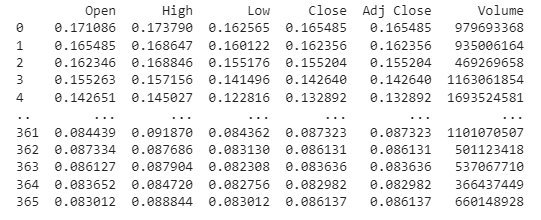
**3.2 DATA CLEANING**

Data cleansing is a valuable process that can help companies save time and increase their efficiency. [Data cleansing software](http://www.winpure.com/) tools are used by various organizations to remove duplicate data, fix and amend badly-formatted, incorrect and amend incomplete data from marketing lists, databases and CRM’s.Data quality has become an important issue. This issue becomes more and more important in medicine area, where the need for effective decision making is high. In this context, the need for data cleaning to improve data quality is becoming crucial. Duplicate records elimination is a challenging data cleansing task. In this paper, we present a duplicate records elimination approach to improve the quality of data. We propose a deep learning-based approach for duplicate records detection using a sentence embeddings model. Also, we propose an algorithm for duplicated records correction. Then we apply the proposed duplicate records elimination approach to analyses the effect of data cleaning on the quality of decisions.

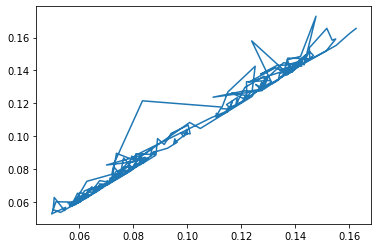
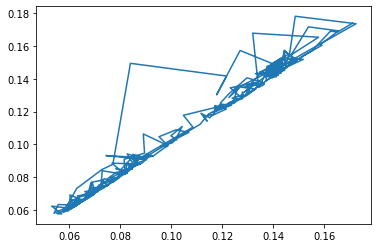
**3.4 DATA VISUALISATION**

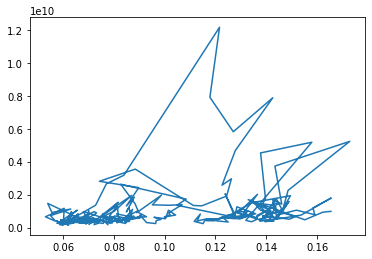
The historical bitcoin data set contains seven feature variables and two target variables output.

**DATASET**



**GRAPHS PLOTTED BETWWEN FEATURE AND TARGET VARIABLES:**

****

****

**4. METHODOLOGY**

**4.1 PROCEDURE TO SOLVE THE GIVEN PROBLEM**

In this project Bitcoin price prediction and prediction we use three approaches:

* Linear regression
* K-Nearest Neighbour
* Support Vector Machine
* Decision Tree

**Linear regression**

Linear regression is a supervised machine learning method that is used by the [Train Using AutoML](https://pro.arcgis.com/en/pro-app/3.0/tool-reference/geoai/train-using-automl.htm) tool and finds a linear equation that best describes the correlation of the explanatory variables with the dependent variable. This is achieved by fitting a line to the data using least squares. The line tries to minimize the sum of the squares of the residuals. The residual is the distance between the line and the actual value of the explanatory variable. Finding the line of best fit is an iterative process.

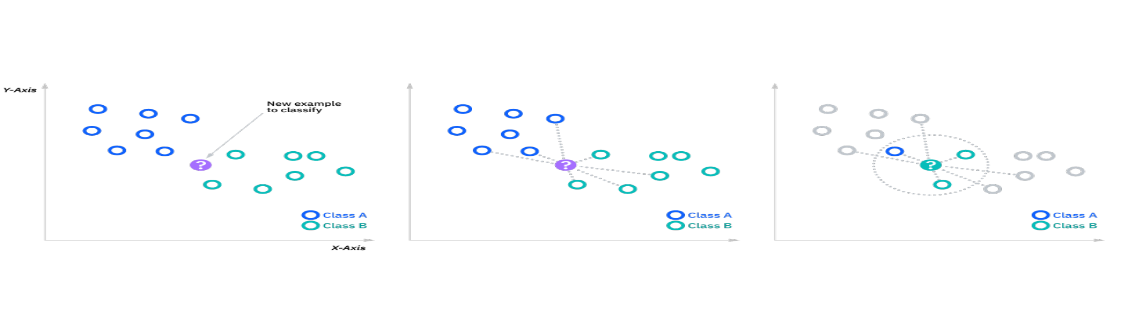
**Advantages of linear regression algorithm:**

* Linear regression performs exceptionally well for linearly separable data
* Easier to implement, interpret and efficient to train
* It handles overfitting pretty well using dimensionally reduction techniques, regularization, and cross-validation
* One more advantage is the extrapolation beyond a specific data set

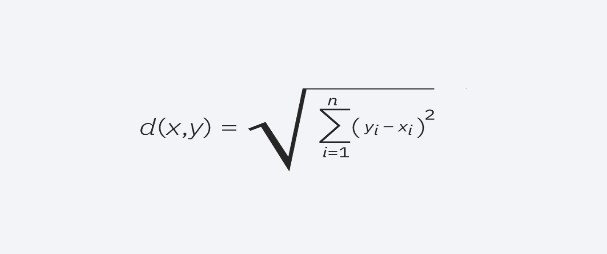
****

**K-Nearest Neighbour**

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.



KNN Formula:



**Support Vector Machine**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.



**4.2 MODEL ARCHITECTURE**

DOGECOIN HISTORICAL DATASET

LOADING DATA SET

IDENTIFYING THE ATTRIBUTES PERTAINING THE DOGECOIN DATASET

COLLECTION OF DATA AND PRE -PROCESSING

-

LINEAR REGRESSION, KNN, DECISION TREE

OBTAIN RESULTS



CONCLUSION

**4.3 SOFTWARE DESCRIPTION**

**Software requirements:**

**Operating system:** Windows

**Platform**: google Collab

**Programing language:** python

**5. RESULTS**

**CODE**

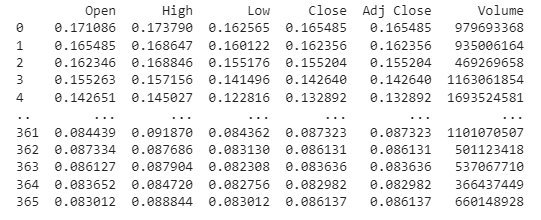
**Dataset:**

import pandas as pd

d=pd.read\_csv("/content/drive/MyDrive/DOGE-USD.csv")

print(d)

**output:**

****

**Linear regression:**

from sklearn.linear\_model import LinearRegression

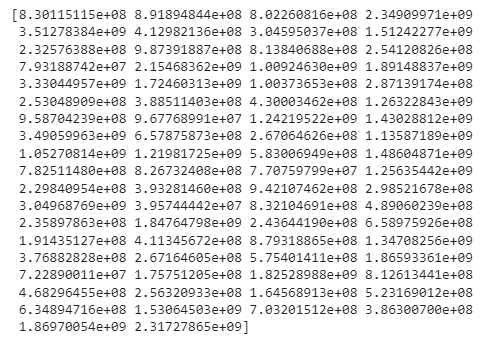
from sklearn.metrics import confusion\_matrix,accuracy\_score,r2\_score,classification\_report,mean\_squared\_error

model=LinearRegression()

model.fit(x\_train,y\_train)

y\_pred=model.predict(x\_test)

**output:**



**K-Nearest Neighbour:**

from sklearn.neighbors import KNeighborsRegressor

model=KNeighborsRegressor()

model.fit(x\_train,y\_train)

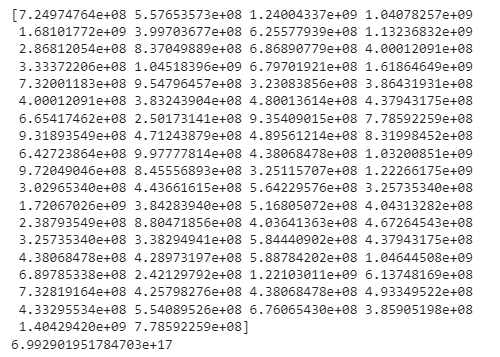
y\_pred=model.predict(x\_test)

print(y\_pred)

from sklearn.metrics import mean\_squared\_error

print(mean\_squared\_error(y\_test,y\_pred))

**output:**



**Support Vector Machine:**

from sklearn.svm import SVR

model=SVR()

model.fit(x\_train,y\_train)

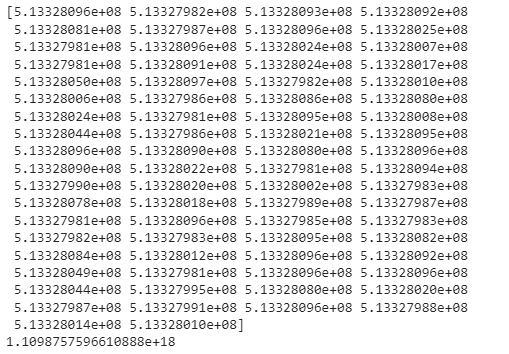
y\_pred=model.predict(x\_test)

print(y\_pred)

from sklearn.metrics import mean\_squared\_error

print(mean\_squared\_error(y\_test,y\_pred))

**output:**



**Decision Tree:**

from sklearn.tree import DecisionTreeRegressor

model=DecisionTreeRegressor()

model.fit(x\_train,y\_train)

y\_pred=model.predict(x\_test)

print(y\_pred)

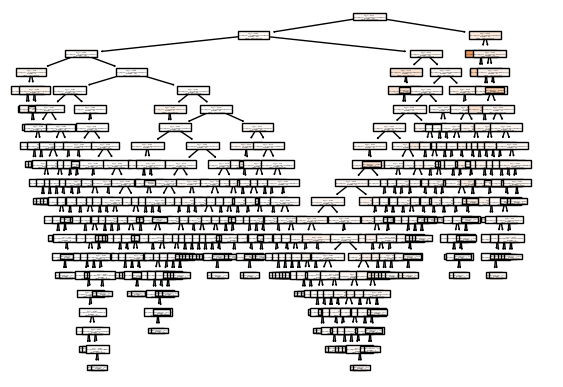
from sklearn.metrics import mean\_squared\_error

print(mean\_squared\_error(y\_test,y\_pred))

from sklearn import tree

tree.plot\_tree(model,filled=True)

**output:**



**6. CONCLUSION AND FUTURE SCOPE**

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system.

The proposed working model can also help in reducing treatment costs by providing Initial

diagnostics in time. The model can also serve the purpose of training tool for medical students

and will be a soft diagnostic tool available for physician and cardiologist. General physicians can

utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements

that could be explored to improve the scalability and accuracy of this prediction system. As we

have developed a generalized system, in future we can use this system for the analysis of

different data sets. The performance of the health’s diagnosis can be improved significantly by

handling numerous class labels in the prediction process, and it can be another positive direction

of research. In DM warehouse, generally, the dimensionality of the heart database is high, so

identification and selection of significant attributes for better diagnosis of heart disease are very

challenging tasks for future research

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system.

The proposed working model can also help in reducing treatment costs by providing Initial

diagnostics in time. The model can also serve the purpose of training tool for medical students

and will be a soft diagnostic tool available for physician and cardiologist. General physicians can

utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements

that could be explored to improve the scalability and accuracy of this prediction system. As we

have developed a generalized system, in future we can use this system for the analysis of

different data sets. The performance of the health’s diagnosis can be improved significantly by

handling numerous class labels in the prediction process, and it can be another positive direction

of research. In DM warehouse, generally, the dimensionality of the heart database is high, so

identification and selection of significant attributes for better diagnosis of heart disease are very

challenging tasks for future research

The regression model, implemented here, is a basic model that takes into consideration only a few features that affect the Dogecoin price. Our model is fairly accurate when predicting the future prices. However, to increase the efficiency of the model, more Dogecoin price features need to be taken into consideration. We recommend using Yahoo as the source of datasets, since information present in this website holds a high degree of authenticity. Our future work would include in-depth scrutinization on the topic of regression, and deep learning at large. Such fact-findings would be beneficial for forecasting the prices of cryptocurrencies, in the future.

**7.REFERENCES**

[1] Sridhar, Sashank, and Sowmya Sanagavarapu. "Multi-head self-attention transformer for dogecoin price prediction." *2021 14th International Conference on Human System Interaction (HSI)*. IEEE, 2021.

[2] Agarwal, Basant, et al. "Prediction of dogecoin price using deep learning and social media trends." *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems* 8.29 (2021): e2-e2.

[3] Patra, Gyana Ranjan, and Mihir Narayan Mohanty. "Price Prediction of Cryptocurrency Using a Multi-Layer Gated Recurrent Unit Network with Multi Features." *Computational Economics* (2022): 1-20.

[4] Medzihorský, Juraj. "Dogecoin price prediction–can be a determinism supposed?." *Ekonomika a spoločnosť* 2.22 (2021): 67-81.

[5] Murphy, Marissa, et al. "Cryptocurrency Price Predictions Using High Performance Computing." (2021).