VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JNANA SANGAMA", BELAGAVI-590018, KARNATAKA

A Project Report on

"CRYPTOCURRENCY PRICE PREDICTION USING MACHINE LEARNING"

Submitted in partial fulfillment of the requirement for the award of degree of

Bachelor of Engineering

In

Computer Science and Engineering

JAGADISH V GAIKWAD (4NN18CS021)
NILESH SHRENIK HOSURE (4NN18CS031)
SHRAVANI M R (4NN18CS042)
NIKITA KULLOLI (4NN18CS053)

Under the guidance of

Mr. Madhusudhan H S

Assistant Professor

Department of CSE

ESTD-2008

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NIE INSTITUTE OF TECHNOLOGY

MYSURU-570018

2021-2022

NIE Institute of Technology

Mysuru

(Affiliated to Visvesvaraya Technological University, Belagavi)

ESTD-2008

Department of Computer Science and Engineering

Certificate

This is to certify that the project work entitled **"CRYPTOCURRENCY PRICE PREDICTION USING MACHINE LEARNING"** is carried out by **JAGADISH V GAIKWAD (4NN18CS021), NILESH SHRENIK HOSURE (4NN18CS031), SHRAVANI M R (4NN18CS042), NIKITA KULLOLI (4NN18CS053)** the bonafide students of **NIE INSTITUTE OF TECHNOLOGY** in partial fulfilment for the award of **Bachelor of Engineering in Computer Science & Engineering** of the **Visvesvaraya Technological University, Belagavi -590018, Karnataka** during the academic year **2021-2022**. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering degree.

Signature of the Guide	**Signature of the HOD**	**Signature of the Principal**
Mr. Madhusudhan H S Asst. Professor, Dept. of CSE, NIEIT, Mysuru.	Smt. Usha M S Associate Professor &Head, Dept. of CSE, NIEIT, Mysuru.	Dr. Archana N V Principal NIEIT, Mysuru.

External Viva

Name of the examiners	Signature with Date
1.__________________	1. _________________
2.___________________	2. _________________

DECLARATION

We, **JAGADISH V GAIKWAD, of** USN **4NN18CS021, NILESH SHRENIK HOSURE,** of USN **4NN18CS031, SHRAVANI M R,** of USN **4NN18CS042, NIKITA KULLOLI** of USN **4NN18CS053** hereby undertake that the project work entitled **“CRYPTOCURRENCY PRICE PREDICTION USING MACHINE LEARNING”** is carried out by us independently under the guidance of **Mr. Madhusudhan H S**, Asst. Professor, Department of Computer Science and Engineering, NIEIT, Mysuru-18, in partial fulfillment of the requirement for the award of Bachelor of Engineering in Computer Science and Engineering by the Visvesvaraya Technological University, Belagavi-590018. The project has been our original work and has not formed the basis for the award of any degree, associate ship, fellowship or any other similar titles.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose consistent guidance and encouragement crowned my efforts with success.

We consider ourselves proud to be a part of **NIE Institute of Technology Mysuru** family, the institution which stood by our way in all our endeavors.

We wish to express our gratitude to **Dr. Archana N V,** Principal, **NIE Institute of Technology Mysuru,** for providing a congenial working environment.

We express our sincere thanks to **Smt. Usha M S**, Associate Professor & Head, Department of Computer Science and Engineering, for her support and encouragement.

We would like to thank our guides**, Mr. Madhusudhan H S** and **Mrs. Lakshmi B S**, Assistant Professors, Department of Computer Science and Engineering, for their inspiration, guidance, constant supervision, direction and discussions in successful completion of this project work.

We are thankful to the Department of Computer Science and Engineering staff members and non-staff members for their co-operation extended towards this work.

Finally, our heartfelt gratitude to our **family members, relatives** and **friends** for their constant support, motivation and encouragement throughout this project. Above all I would like to thank God, Almighty, for having showered his blessings on us.

JAGADISH V GAIKWAD (4NN18CS021)
NILESH SHRENIK HOSURE (4NN18CS031)
SHRAVANI M R (4NN18CS042)
NIKITA KULLOLI (4NN18CS053)

ABSTRACT

After the boom and bust of cryptocurrencies' prices in recent years, it has been increasingly regarded as an investment asset. Because of its highly volatile nature, there is a need for good predictions on which to base investment decisions. Although existing studies have leveraged machine learning for more accurate Cryptocurrency price prediction, few have focused on the feasibility of applying different modeling techniques to samples with different data structures and dimensional features. To predict Cryptocurrency price at different frequencies using machine learning techniques, we first download the dataset from a trusted website which keeps all the data of various cryptocurrencies then we classify various Cryptocurrencies by the dataset that is according to the available price. We extract the basic trading features acquired from a cryptocurrency exchange are used for 1 month price prediction. Machine learning algorithms including ARIMA and SVR models for Cryptocurrency's daily price prediction with high-dimensional features achieve an accuracy of 93% and 94% respectively, outperforming more complicated machine learning algorithms. Compared with benchmark results for daily price prediction, we achieve a better performance, with the highest accuracy of the machine learning algorithm of 97%. Our Hybrid Machine learning model including Support Vector Regression and Autoregressive integrated moving average for One month's Cryptocurrency price prediction is superior to other Machine learning methods, with accuracy reaching 97%. Our investigation of Cryptocurrency price prediction can be considered a pilot study of the importance of the sample dimension in machine learning techniques.

LIST OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

A cryptocurrency, crypto-currency, or crypto is a digital currency designed to work as a medium of exchange through a computer network that is not reliant on any central authority, such as a government or bank, to uphold or maintain it. There are 18000 various cryptocurrencies', the most popular among them is Bitcoin. Bitcoin is a crypto currency used worldwide for digital payment or simply for investment purposes. Bitcoin is decentralized i.e., it is not owned by anyone. Transactions made by bitcoins are easy as they are not tied to any country. Investment can be done through various marketplaces known as "bitcoin exchanges." These allow people to sell/buy bitcoins using different currencies. The largest bitcoin exchange is Mt Gox. Bitcoins are stored in a digital wallet which is basically like a virtual bank account. The record of all the transactions, the timestamp data is stored in a place called Blockchain. Each block contains a pointer to a previous block of data. The data on blockchain is encrypted. During transactions the users name is not revealed, but only their wallet ID is made public. Many cryptocurrencies entered into the crypto market after Bitcoin, for example, Ethereum, launched in 2015, is the second-largest cryptocurrency with a $410 billion market capitalization. More than 5,600 different cryptocurrencies are traded in around 1,100 exchanges and Ripple, Tether, Cardano, Stellar, Litecoin, and Zcash are the most popular digital currencies. Back in June 2016, the total market capitalization of all cryptocurrencies was approximately 12.22 billion dollars and it fluctuated in 2017. It increased to $1.75 trillion in June 2021 [12], with an all-time high of $2 trillion. It will reach nearly $8 trillion by 2030. The daily volume of the crypto market is around $117 billion and more than 100 million people are using these currencies

Predicting the future is no easy task. Many have tried and many have failed. But many of us would want to know what will happen next and would go to great lengths to figure that out. Imagine the possibilities of knowing what will happen in the future! Imagine what you would have done back in 2012 when Bitcoin was less than $15 knowing that it would surpass $18,000! Many people may regret not buying Bitcoin back then but how were they supposed to know in the first place? This is the dilemma we now face in regards to cryptocurrency. We do not want to miss out on the next jump in price but we do not know when that will or will not happen. So how can we potentially solve this dilemma? Maybe machine learning can tell us the answer.

Machine learning models can likely give us the insight we need to learn about the future of cryptocurrency. It will not tell us the future but it might tell us the general trend and direction to expect the prices to move. Let's try and use these machine learning models to our advantage and predict the future of Bitcoin by coding them out in Python!

1.1 Related Work

After the boom of cryptocurrencies' prices in recent years, cryptocurrency has been increasingly regarded as an investment asset. Because of its highly volatile nature, there is a need for good predictions on which to base investment decisions. Although existing studies have leveraged machine learning for more accurate Bitcoin price prediction, few have focused on the feasibility of applying different modeling techniques to samples with different data structures and dimensional features. To predict Bitcoin price at different frequencies using machine learning techniques, many have tried predicting the prices of various cryptocurrencies' using various Machine learning algorithms and some also using mathematical and statistical methods. Most of these predictions resulted in failure or inconsistency in the models. Some of them have also considered various external parameters including twitter sentiment analysis or the effect of covid-19 on the rise and fall of various cryptocurrencies. We have studied specific models which we needed for our predictions, there has also been news about movement of prices of cryptocurrencies based on political mis-beliefs on cryptocurrencies, thus resulting in the fall of prices.

- **Deep Learning Models:** Appropriate design of deep learning models in terms of network parameters is imperative to their success. The three main options available when choosing how to select parameters for deep learning models are random search, grid search and heuristic search methods such as genetic algorithms. Manual grid search and Bayesian optimization were utilized in this study. Grid search, implemented for the Elman RNN, is the process of selecting two hyperparameters with a minimum and maximum for each. One then searches that feature space looking for the best performing parameters. This approach was taken for parameters which were unsuitable for

Bayesian optimization. This model was built using Keras in the Python programming language.

Bayesian optimization was chosen for selecting LTSM parameters where possible. This is a heuristic search method which works by assuming the function was sampled from a Gaussian process and maintains a posterior distribution for this function as the results of different hyperparameter selections are observed. One can then optimize the expected improvement over the best result to pick hyperparameters for the next experiment. The performance of both the RNN and LSTM network are evaluated on validation data with measures to prevent overfitting. Dropout is implemented in both layers, and we automatically stop model training if its validation loss hasn't improved in 5 epochs.

- **ARIMA (Autoregressive Integrated Moving Average):**

The basic principle of the ARIMA model is to estimate the trend and the seasonality of the series and to remove them from the series in order to obtain a stationary series. In this series, statistical forecasting techniques can be used. The final step would be to convert the forecast values to the original scale by applying constraints on trend and seasonality. Trend – mean varying over time. For example, we can see an average increase in the number of bitcoin price over time. Seasonality – time frame variations. For example, people may tend to buy cars in a given month due to pay increments or festivals. Static forecasts are performed and the RMSE is calculated to compare with other models. ARIMA model was implemented to compare its predictability with the LSTM and figure out which is the most suitable method for time series data which has huge fluctuations. ARIMA (Auto regression integrated moving average) is a class model that captures a suite of different standard temporal structures in time series data which include trend, seasonality, cycles, errors and nonstationary data. This allows it to exhibit dynamic temporal behavioral a time sequence. The data preparation phase is done similar to the LSTM model approach [2].

- **SVM (Support Vector Machine):**

Predictive models for classification and regression problem is a Support Vector Machine (SVM) works by creating a higher-dimensional model which assigns each new data provided to one category or another. Decision is obtained from the verge that which maximizes the functional and geometric margins between classes. SVM model is best for resolving classification problems [7].

- **LSTM (Long Short-Term Memory):**

  Long Short-Term Memory Network Model (LSTM) is unique type of Recurrent Neural Network (RNN) specifically used to cease long-term dependency problem. Information is stored for longer period; this marks as their default behavior. It selects the necessary information to be store and discards the irrelevant information. The sequential characteristics in the time series data is ignored by many of the machine learning algorithms. This problem is identified in rolling window LSTM. Application of any other machine learning classifiers on time series data usually overlook the sequential data. Rolling window LSTM builds model structure for time series prediction by not considering weights. LSTM takes longer time to train as compared to RNN model, the parameters and activation functions increased the computation leading to long term memory.

1.2 Aim

Cryptocurrencies are the most complex currency whose value change every second. Investing money in cryptocurrency is more risk and less profit. Our aim is to predict the cryptocurrency price accurately taking into consideration various parameters that affect the cryptocurrency based on the dataset consisting of various features relating to the price and payment of cryptocurrencies over the course which may affect the future prices, and thus resulting into losses.

1.3 Scope

Today cryptocurrency is a secure transaction system that has a valuable impact on capital. They are awarded under a restriction in which customers offer their computer authority to register and listing trades with the coins. The purchase and sale of these coins in different currencies is carried out in an alternative workplace where "purchase" or "sell" requests are placed in the ordered e-book. "Buy" or "bid" offers to talk about the purpose of purchasing certain cryptocurrency measures at a few costs while "provide" or "ask" offers to talk about the expectation of providing certain measures at a certain cost. The change is ordered through the coordination of pricing requests from the arrangement of e-books to a valid exchange between customers and suppliers.

1.4 Motivation

Considering various cryptocurrencies, the one which largely matters to everyone is Bitcoin. The large price fluctuation of Bitcoin was caused by many factors, which can be roughly divided into two categories. First, Bitcoin market is a newly developing market. There is no physical representation linked with this kind of virtual asset. Meanwhile, a large number of individual investors can be easily affected by market manipulation, consequently making unreasonable decisions. All these issues (fake news, manipulation or other reasons) lead to a large price fluctuation of Bitcoin. Second, Bitcoin market lacks government regulation. Regulators in traditional financial markets are basically missing in the field of cryptocurrencies. For instance, fake news frequently affects the decisions of individual investors. BSV (Bitcoin Satoshi's Vision (BSV) is a variant of Bitcoin since 2018) increased from $125 to $251 on 2019.05.29 with the news that BSV would be backed by the Binance currency exchange. However, the price fell back to $130 when the CEO of Binance clarified that this announcement was a fake news. In addition, Bitcoin is a global product that is affected by regulation around the world. For example, the sharp reduction in Bitcoin price by almost 50% in early 2018 was mostly caused by government regulations in South Korea and China, who forbid the initial coin offerings (ICOs).

1.5 Organization of Report

The sequel of pages and their hierarchical arrangement play a pivotal role in structuring the project report properly and interlinking the vital elements of the report in the best possible format. This project report consists of 8 chapters as mentioned below:

- Introduction
- Literature survey
- System Requirements
- System analysis
- System methodology
- System implementation
- System Testing
- Results, Conclusion & Future scope and References

Introduction: Provides the background information about the project and the basic idea of what this project is expected to do. Literature survey: gives the detailed study of all the existing systems and its disadvantages.

System requirements: Talks about the detailed description on system requirements including both hardware and software. System analysis provides a detailed description about the system analysis, why is it required, method of analysis of existing system, proposed system and its components.

System design: It is involved in giving a description on how the system is going to be designed, how exactly the system would be developed. System implementation: It is all about the implementation part of the project that describes the critical coding of the project.

Testing: gives information about testing the project in the real time scenarios and determines the efficiency of the system.

Snapshots: It consists of snapshots of software and hardware modules.

Conclusion & Future scope: It includes the extensions that could be made to this project.

References: consists of the papers, books, and websites we have referred to.

Chapter 2

LITERATURE SURVEY

A literature survey or review which combines both summary and synthesis of specific conceptual categories. Literature survey gives conclusion about how one can analyze and understand gaps exist in how a problem has been researched to date.

2.1 Survey Papers

A literature review surveys, scholarly articles or any other resources which are relevant to our area of interest in the research provides a brief description and critical evaluation of works which are related to our research problem. Literature survey provides an overview of sources that we have explored or referred during our research.

In this paper, they have used the LSTM version of Recurrent Neural Networks, pricing for Bitcoin. To develop a better understanding of its price influence and a common view of this good invention, they have First given a brief overview of Bitcoin again economics. After that, they define the database, including data from stock market indices, sentiment, blockchain and Coin market cap. Further in this investigation, they demonstrate the use of LSTM structures with the series of time mentioned above. In conclusion, they have drawn the Bitcoin pricing forecast results 30 and 60 days in advance.[1]

The objective of this paper was to determine the predictable price direction of Bitcoin in USD by machine learning techniques and sentiment analysis. Twitter and Reddit have attracted a great deal of attention from researchers to study public sentiment. They have applied sentiment analysis and supervised machine learning principles to the extracted tweets from Twitter and Reddit posts, and they have analyzed the correlation between bitcoin price movements and sentiments in tweets. We explored several algorithms of machine learning using supervised learning to develop a prediction model and provide informative analysis of future market prices. They have used LSTM, ARIMA and RNN to analyze the price movements.[2]

The goal of this paper is to ascertain with what accuracy the direction of Bitcoin price in USD can be predicted. The price data is sourced from the Bitcoin Price Index. The task is achieved with varying degrees of success through the implementation of a Bayesian optimized recurrent neural network (RNN) and a Long Short-Term Memory (LSTM) network. The LSTM achieves the highest classification accuracy of 52% and a RMSE of 8%. The popular ARIMA model for time series forecasting is implemented as a comparison to the deep learning models. As expected, the non-linear deep learning methods outperform the ARIMA forecast which performs poorly. Finally, both deep learning models are benchmarked on both a GPU and a CPU with the training time on the GPU outperforming the CPU implementation by 67.7%.[3]

In this proposed work, it was studied to forecast the Bitcoin price precisely considering different parameters that influence the Bitcoin price. This study first handles, it was identified that the price trend on day-by-day changes in the Bitcoin price while it gives knowledge about Bitcoin price trends. The dataset till current date was taken with open, high, low and close price details of Bitcoin value. Exploiting the dataset machine learning module is introduced for prediction of price values. The aim of this work was to derive the accuracy of Bitcoin prediction using different machine learning algorithm and compare their accuracy. Experiment results are compared for decision tree and regression model.[4]

In this paper, three kinds of features are presented for the price fluctuation prediction, including basic features, traditional technical trading indicators, and features generated by a Denoising autoencoder. They evaluated these features using an Attentive LSTM network and an Embedding Network (ALEN). In particular, an attentive LSTM network can capture the time dependency representation of Bitcoin price and an embedding network can capture the hidden representations from related cryptocurrencies. Experimental results demonstrate that ALEN achieves superior state-of-the-art performance among all baselines. Furthermore, they have investigated the impact of parameters on the Bitcoin price fluctuation prediction problem, which can be further used in a real trading environment by investors.[5]

. This paper conducts an in-depth study on evolution of Bitcoin and also a systematic review is done on various machine learning algorithms used for predicting the prices. Comparative analysis envisions to select optimal technique to forecast prices more precisely the solutions provided in many of the existing system gives 60- 70% accuracy. Although some techniques are not considered as the accuracy obtained by them is very less, the overall study is very promising and can help investors to invest accordingly. The accuracy of NARX Model is the best model in predicting the Bitcoin prices. The analysis conducted in this work can be further extended by more research on upcoming advance methods. Hence more thorough picture of forecasting prices can be obtained.[6]

The wide price range of digital currencies highlights the need for reliable preparation for predicting the currency's price. A new model is a situation in which this paper presents a new way of forecasting digital value for money by considering several variables, such as stock market capitalization, volume, distribution, and high-end delivery. To include training results, active LSTM networks, and overview of long-term organizations are considered. The proposed technique is used for the benchmark data sets. The results indicate the efficiency of the forecasting of digital currency.[7]

This paper compares the performance of four different machine learning models on predicting the Bitcoin return rate and price trend. Data are formulated to four input feature sets, including: (1) Historical Bitcoin exchange data; (2) Historical Bitcoin exchange data + COVID-19 data (recovery, confirmed, death); (3) Historical Bitcoin exchange data + Twitter data; (4) Historical Bitcoin exchange data + COVID-19 data (recovery, confirmed, death) + Twitter data. The four machine learning models implemented are: (1) Random Forest; (2) Decision tree; (3) AdaBoost; (4) Support vector machine. They found that: (1) Twitter data can improve the performance of models; (2) People consider information within 5 days when they make decisions on investments; (3) Support vector machine does not perform well in predicting Bitcoin return rate or price trend; (4) COVID-19 data does not help improve the prediction. However, we have very limited COVID-19 data, so future research with more COVID-19 data may help confirm if the last statement is correct or not.[8]

This paper proposes the novel method of the construction of prediction model using deep learning approach. The proposed approach was found to be more accurate than the Machine learning models used for prediction as the deep learning model consider the non-linear nature of price. The results verify the applicability of model and give a direction to investors on how deep learning techniques can be used in decision making.[9]

In this paper, they have proposed a deep-learning-based hybrid model (includes Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM)) to predict the price of Litecoin and Zcash with inter-dependency of the parent coin. The proposed model can be used in real-time scenarios and it is well trained and evaluated using standard data sets. Results illustrate that the proposed model forecasts the prices with high accuracy compared to existing models.[10]

In this paper, they have undertaken to develop systems that can effectively predict price movements in the cryptocurrency market, they display significant efficiency gaps, which this paper further explores. The authors then attempt to learn from past studies and construct a more holistic approach to a predictive price model for the cryptocurrency market. This focuses on assessing key factors that affect the volatility of the market – public perception, trading data, historic price data, and the interdependencies between Bitcoin and Altcoins - and how they can be best utilized from a technological aspect by applying sentiment analysis and machine learning techniques, to increase the efficiency of the process.[11]

In this paper, they have proposed to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value, here in this paper they use an algorithm linked to artificial intelligence named LASSO (least absolute shrinkage selection operator. In LASSO, finding of the results from a larger database is quick and fast. So for this purpose they drew a comparison between other algorithms and the LASSO algorithm, this survey paper helps the upcoming researchers to make an impact in the their papers. The process happens in the paper is first moment of the research, they aim to understand and find daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. Their 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, they apply some data mining techniques to reduce the noise of data. Then the second moment of the research, using the available information, they predicted the sign of the daily price change with highest possible accuracy.[12]

Chapter 3

SYSTEM REQUIREMENTS

System requirements are the configuration that a system must have in order for a hardware or software application to run smoothly and efficiently. Failure to satisfy these requirements may result in installation problems or performance problems. The former may prevent a tool or application from getting installed, whereas the latter may cause a product to malfunction or perform below expectation or may be to hang or crash. System requirements are also referred to as minimum system requirements. System requirements can be broadly classified as hardware requirements, software requirements, functional requirements and non-functional requirements. The hardware system requirements often specify the OS version, processor type, memory size, available disk space and extra peripherals, if any, needed. The software system requirement consistent of all necessary requirements required for project development. A functional requirement defines a function of a system or its component, where a function is described as a specification of behavior between outputs and inputs. Non-functional requirements impose constraints on the planning or implementation like performance engineering requirements, quality standards, or design constraints.

3.1 Hardware requirements

- PROCESSOR: Intel(R) Core (TM) i5-8250U CPU @ 1.60GHz, 1801MHz
- RAM: 8 GB
- Graphics: NVIDIA GEFORCE
- OS-type: 64-bit operating system
- Hard disk:30 GB

3.2 Software requirements

- Operating System: Windows 10
- Tools: Xampp, Jupyter
- Languages: MySQL, Python

3.3 Functional requirement

A functional requirement defines a function of a system or Its component, where function is described as a specification of behavior between outputs and inputs.

3.4 Non-Functional requirements

A non-functional requirement is a requirement that specifies criteria which will be used to judge the operation of a system, instead of specific behaviors.

- **Usability:** The system must be simple that people like to use it, but not so complex that people avoid using it. The user must be familiar with the user interfaces and should not have problems in migrating to a new system with a new environment. The menus, buttons and dialog boxes should be named in a manner that they provide clear understanding of the functionality. Several users are going to use the system simultaneously, so the usability of the system should not get affected with respect to individual users.
- **Reliability**: The system should be trustworthy and reliable in providing the functionalities. Once a user has made some changes, the changes must be made visible by the system. The changes made by the Programmer should be visible both to the Project leader and to the Test engineer as well.
- **Security:** Apart from bug tracking the system must provide necessary security

and must secure the whole process from crashing. As technology began to grow in fast rate the safety became the main concern of an organization. Millions of dollars are invested in providing security. Bug tracking delivers the utmost security available at the very best performance rate possible, ensuring that unauthorized users cannot access vital issue information without permission. Bug tracking system issues different authenticated users their secret passwords so that there are restricted functionalities for all the users.

- **Scalability:** The system should be scalable enough to add new functionalities at a later stage. There should be a standard channel, which may accommodate the new functionalities.
- **Performance:** Ensure software applications will perform well under their expected workload. The response time of the system is good.
- **Maintainability**: The system monitoring and maintenance should be simple and objective in its approach. There should not be too many jobs and roles running on different machines such that it gets difficult to watch whether the jobs and roles are running without errors.
- **Portability:** The system should be easily portable to another system. This is required when the web server, which is hosting the system gets stuck due to some problems, which requires the system to be taken to another system.
- **Flexibility:** The system should be flexible enough to allow modifications at any point of time.
- **Reusability**: The system should be divided into such modules that it could be used as a part of another system without requiring much of work.

Chapter 4

SYSTEM ANALYSIS

System development can generally be thought of having two major components: System analysis and System design. System design is that the process of planning a replacement system or one to exchange or complement of an existing system. But before this planning can often done, we must thoroughly understand the existing systems and determine how computers can be best used to make its operation simpler and more effective. System analysis is the process of gathering and interpreting facts, diagnosing problems, and using the knowledge to recommend improvements to the system. This is the job of the system analyst.

The principal objective of the system-analysis phase is to specify what the system needs to do to meet the wants of end users. System analysis is conducted for the aim of studying a system or its parts so as to spot its objectives. It is a drag solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose.

The process of system analysis consists of various steps:

• **Collecting Information about existing systems:** This step involves collection of all information about the existing systems by research, observations about working operations of the systems in various scenarios and reviews about the systems by the users.

• **Identifying the inputs, outputs and processes of the existing systems:** Every system has inputs and outputs, the system analyst needs to identify the data input to the existing system, and the data output given by the existing system. This is because any new system that is designed will need to affect similar inputs and outputs as the existing system. Any new system that is created will need to take in the same input data and will have to produce similar kind of outputs. The analyst also has got to understand how the existing system works

• **Identifying issues in the existing systems:** No system is perfect and it is job of the system analyst to try and identify where and what are the problems in the existing system. If these problems are often fixed, the system will work more efficiently and, Within the case of a business, be more profitable.

• **New system requirements specifications:** After the problems with existing system are understood, the system analyst can begin to plan how the new system will fix those problems. The analyst specifies an inventory of requirements for the new system. This list is usually called the Requirements Specification.

• **Type of hardware and software required:** The system analyst will now need to decide hardware and software which will be required for the new system. The Hardware requirements would include how many computers to use, what type of network to use, how many servers required, etc. The Software requirements would include decisions such as whether to use ready- made off-the-shelf software or use a custom-written software.

4.1 Existing System

Bitcoin system is unique from any other asset on the financial market and thereby creates new possibilities. This work can serve as an exploratory beginning for several techniques, descriptive analysis of Bitcoin prices. As the network is large, small as well as large transactions are carried out in this network resulting in volatility of prices which needs to be maintained. The results show survey of techniques that used and also the technique which suits best for the prediction. The solutions provided in many of the existing system gives 60- 70% accuracy. Although some techniques are not considered as the accuracy obtained by them is very less, the overall study is very promising and can help investors to invest accordingly.

The accuracy of NARX Model is the best model in predicting the Bitcoin prices. The analysis conducted in this work can be further extended by more research on upcoming advance methods. Hence more thorough picture of forecasting prices can be obtained.

It was found that various deep learning algorithms were successful in prediction ranging from GRU, LSTM and CNN then comparing them with machine learning algorithm SVM. It can be inferred that deep learning models take long time to train. Prediction of bitcoin price was a complicated task as it is based on large dataset. Machine learning model or deep learning model without real time deployment are of no use.

4.2 Proposed System

The proposed model has transformed the data in a suitable form which is passed to the machine learning algorithms. Features of the data are extracted and used for prediction. Our hybrid model combines both the models (SVR and ARIMA), we use .merge method to merge both the data frames. Firstly, we collect data from websites (coinmarketcap.com and finance.yahoo.com) then we process the data and split the data and run both the models separately and test them then we apply the .merge to combine both the data frames, it only considers the test values of both the algorithms that are same, other values are deleted in this method. Thus, providing us with the better results of both models.

Chapter 5

ARCHITECTURE

5.1 System Architecture

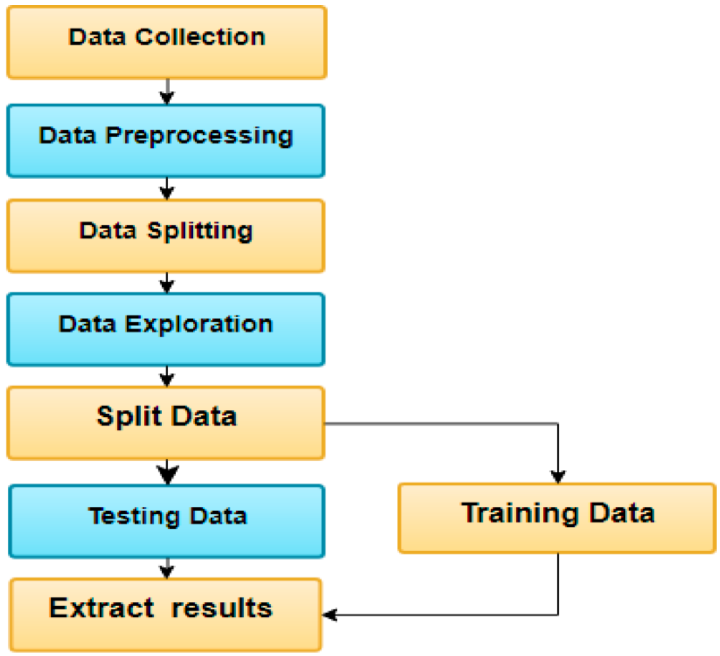

Figure 5.1 System Architecture

Various software, mathematical formulas, and algorithm-based methods have been proposed for monitoring, predicting and detecting the trends and movement of various cryptocurrencies, this would have succeeded people to earn more profits and reduce their losses. This system architecture includes various steps starting from Data Collection, preprocessing splitting the data into training and testing then applying various Machine learning or deep learning algorithms. Firstly, the data is collected from trusted sources who keep the flow of data, download the historical data and then convert the date format in the csv file. The date, if it is in timestamp format then convert the date, delete all the values with NaN (empty entries) and then split the data into training and testing in the ratio 7:3.

Then apply the machine learning algorithms to get suitable accuracy, train the model to give out better accuracy. In our case, we have applied two machine learning algorithms to predict the output.

5.2 Methodology

The proposed approach for building the machine learning model to predict cryptocurrency price prediction involved various steps.

1. **Data Collection:** The dataset requirement entirely depends on the project requirement. The dataset can be collected from various sources such as file, database or even a sensor. The dataset of the various cryptocurrencies' prices used in this work to build machine learning model was collected from finance.yahoo.com and some of the rows and columns of the dataset are shown in fig 5.2. The complete dataset has 7 columns for each cryptocurrency. The CSV file of dataset has prices based upon various factors such as open, high price, Adj. close price, closing market price, market volume.

Date	Open	High	Low	Close	Adj Close	Volume
17-09-2014	465.864	468.174	452.422	457.334	457.334	21056800
18-09-2014	456.86	456.86	413.104	424.44	424.44	34483200
19-09-2014	424.103	427.835	384.532	394.796	394.796	37919700
20-09-2014	394.673	423.296	389.883	408.904	408.904	36863600
21-09-2014	408.085	412.426	393.181	398.821	398.821	26580100
22-09-2014	399.1	406.916	397.13	402.152	402.152	24127600
23-09-2014	402.092	441.557	396.197	435.791	435.791	45099500
24-09-2014	435.751	436.112	421.132	423.205	423.205	30627700
25-09-2014	423.156	423.52	409.468	411.574	411.574	26814400
26-09-2014	411.429	414.938	400.009	404.425	404.425	21460800
27-09-2014	403.556	406.623	397.372	399.52	399.52	15029300
28-09-2014	399.471	401.017	374.332	377.181	377.181	23613300
29-09-2014	376.928	385.211	372.24	375.467	375.467	32497700
30-09-2014	376.088	390.977	373.443	386.944	386.944	34707300

Figure 5.2 Dataset

2. **Data Preprocessing:** This is the most important step in machine learning which helps in building the model more accurately to perform the analysis. In this step, data collected from the site is converted to the clean data set. Then, the data is split into testing set and the training set. For e.g., Data can be divided into 70% of training data while 30% of testing data.
3. **Data Scaling Phase:** In this step the data is scaled according to model requirements. It reshapes data to make the model more suitable.

4. **Model Building Phase:** The preprocessed data is used to build the best performing model.
5. **Model Learning Phase:** After the training data is defined, it is configuring with the defined model to start the learning phase. After the fully configured machine learning algorithms are defined, the data is passed to the model for training. This is achieved by calling fit () method.
6. **Evaluation**: This is the integral part as it helps in finding the best model that represents data and how well the prediction is achieved. Input values are passed to the model and the output is the predicted values. The output is compared with the testing data to calculate accuracy and the RMSE values.

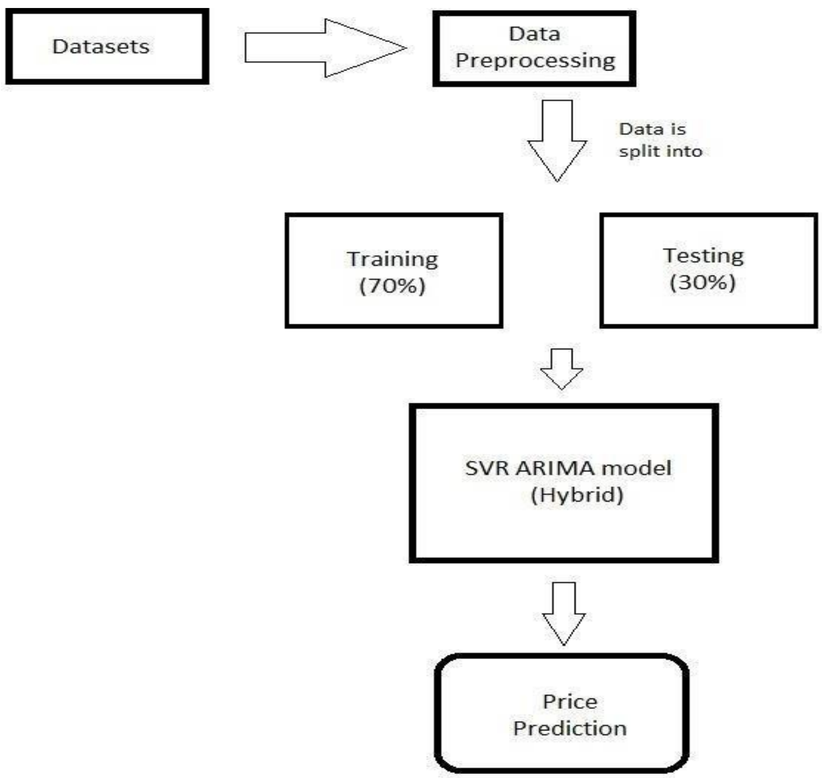

Figure 5.3 Flow chart of our Model

5.2.1 Post-Processing

The post-processing mechanism involves the process of applying machine learning algorithms on the processed dataset. We have used 2 machine learning algorithms to predict the future prices of cryptocurrencies. They are ARIMA and SVR, the mathematical equations of both are described below.

ARIMA Algorithm:

The Autoregressive Integrated Moving Average (ARIMA) is a linear model, which combines an AR process, a MA process and an integrated component, which involves differencing the time series to convert it into a stationary process.

Step 1: The auto-regressive (AR) model expresses the time series xt at time t as a linear regression of the previous p observations, that is:

$$x_t = \alpha + \sum_{i=1}^{p} \phi_i \, x_{t-i} + \varepsilon_t$$

and where εt is the white noise residual term and ϕi are real parameters.

Step 2: The Moving averages (MA) use dependency between residual errors to forecast values in the next period. The model helps to adjust to unpredictable events. The q th order moving average model denoted by MA (q) is defined as follows:

$$x_t = \alpha - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where α and θi are real parameters. The ARMA model combines the power of AR and MA components together. This way, an ARMA (p, q) model incorporates the p th order AR and q th order MA model, respectively.

Step 3: We denote by Φ and θ the AR and MA coefficients vectors. The α and εt captures the intercept and the error term at time t. The complete ARIMA (p, q) model can be seen in detail in equation, that is

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

SVR Algorithm:

The SVR is an extended technique of the "Support Vector Machine" (SVM). Researchers have found that SVM offers excellent performance of time series predictions. The SVM finds an optimal hyper-plane to separate two classes of patterns. SVR is formulated as an optimization problem that tries to minimize the prediction error. Refer as the formulation of the optimization function of SVR:

SVR formula

$$\min \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

assume that ξi , ξi* ∗ ≥ 0 and w is the undetermined parameter vector.

One advantage of using the SVR is the ability to apply different kernel methods which they form different formula. Refer below equations, these define the applied kernels method as follow:

$$\text{Linear} = \mathrm{k}\left(x_i, y_j\right) = x_i . y_j$$

where xi, yi are datasets.

$$\text{Polynomial} = \mathrm{k}\left(x_i, y_j\right) = ((x_i . y_j) + p)^q$$

Where p and q are the kernel parameters and satisfy the condition p ≥ 0, q ∈ N.

$$\text{Gauss Kernl(RBF)} = \exp(- \|x_i - x_j\|^2 / Q$$

where Q ≥ 0.

Chapter 6

IMPLEMENTATION

6.1 Data collection

Dataset collection is a process of collecting required data from all the relevant sources as needed for the research problem, test the hypothesis and evaluate the outcomes.

Yahoo Finance: Yahoo! Finance is a media property that is part of the Yahoo! network. It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. It also offers some online tools for personal finance management. In addition to posting partner content from other web sites, it posts original stories by its team of staff journalists. The website has provided us with the historical dataset of various cryptocurrencies.

CoinMarketCap: CoinMarketCap is a cryptocurrency industry utility that aggregates and reports recently-traded prices for hundreds of cryptocurrencies traded on hundreds of platforms around the world. For each currency it reports the total value of the outstanding currency (the "market capitalization"), its total trading volume and its rank by trading volume over the past month and the past 24 hours.

6.2 Algorithms Used

ARIMA (Autoregressive Integrated Moving Average):

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

An ARIMA model can be understood by outlining each of its components as follows:

- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

- Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

ARIMA Parameters: Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

- p: The number of lag observations in the model; also known as the lag order.
- d: The number of times that the raw observations are differenced; also known as the degree of differencing.
- q: the size of the moving average window; also known as the order of the moving average.

In a linear regression model, for example, the number and type of terms are included. A 0 value, which can be used as a parameter, would mean that particular component should not be used in the model. This way, the ARIMA model can be constructed to perform the function of an ARMA model, or even simple AR, I, or MA models.

SVR (Support Vector Regression):

Support Vector Regression (SVR) works on similar principles as Support Vector Machine (SVM) classification. One can say that SVR is the adapted form of SVM when the dependent variable is numerical rather than categorical. A major benefit of using SVR is that it is a non-parametric technique. Unlike SLR, whose results depend on Gauss-Markov assumptions, the output model from SVR does not depend on distributions of the underlying dependent and independent variables. Instead, the SVR technique depends on kernel functions. Another advantage of SVR is that it permits for construction of a non-linear model without changing the explanatory variables, helping in better interpretation of the resultant model. The basic idea behind SVR is not to care about the prediction as long as the error (ϵ_i) is less than certain value. This is known as the *principle of maximal margin*. This idea of maximal margin allows viewing SVR as a convex optimization problem. The regression can also be penalized using a cost parameter, which becomes handy to avoid over-fit. SVR is a useful technique provides the user with high flexibility in terms of distribution of underlying variables, relationship between independent and dependent variables and the control on the penalty term.

Hybrid Model: Our hybrid model combines both the models (SVR and ARIMA), we use .merge method to merge both the data frames. Firstly, we run both the models separately and test them then we apply the .merge to combine both the data frames, it only considers the test values of both the algorithms that are same, other values are deleted in this method. Thus, providing us with the better results of both models.

Chapter 7

SYSTEM TESTING

System testing is any activity aimed at evaluating an attribute or capability of a program or system and determining that it meets its required results. Although crucial to software quality and widely deployed by programmers and testers, software testing still remains an art, due to limited understanding of the principles of software. The difficulty in software testing stems from the complexity of software: we cannot completely test a program with moderate complexity.

Testing is more than just debugging. Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Test techniques include, but are not limited to the process of executing a program or application with the intent of finding software bugs (errors or other defects). The purpose of testing can be quality assurance, verification and validation, or reliability estimation. Software testing can be stated as the process of validating and verifying that a computer program/application/product:

- Works as expected
- Can be implemented with the same characteristics
- Satisfies the needs of stakeholders.
- Meets the requirements that guided its design and development

Testing can be used as a generic metric as well. Software testing is usually a trade-off between budget, time and quality. Software testing, depending on the testing method employed, can be implemented at any time in the software development process. Traditionally most of the test effort occurs after the requirements have been defined and the coding process has been completed, but in the agile approaches most of the test effort is on- going. As such, the methodology of the test is governed by the chosen software development methodology.

7.1 Testing Methodologies

Software Testing can be of different types, depending on the purpose and entities being tested. The various kinds are:

- **Unit Testing:** It involves testing individual software components or modules. It's typically done by the programmer and not by testers, as it requires detailed knowledge of internal program design and code. It may require developing test driver modules or test harnesses.
- **Integration Testing:** It involves the testing of integrated modules to verify combined functionality after integration. Modules tested are typically code modules, individual applications, client and server applications on a network, etc. This type of testing is especially relevant to client/server and distributed systems.
- **System Testing:** The entire system is tested as per the requirements. It is a type of Functional testing that is based on overall requirements specifications and covers all combined parts of a system.
- **End-to-end Testing:** Similar to system testing, it involves testing of a complete application environment in a situation that mimics real-world use, such as interacting with a database, using network communications, or interacting with other hardware, applications, or systems if appropriate.
- **Acceptance Testing:** Normally this type of testing is done to verify if system meets the customer specified requirements. In engineering and its various sub disciplines, acceptance testing is a test conducted to determine if the requirements of a specification are met. It may involve chemical tests, physical tests or performance tests.

7.1.1 System Testing

- **System Testing:** The entire system is tested as per the requirements. It is a type of Functional testing that is based on overall requirements specifications and covers all combined parts of a system. The system tests performed on the project are:

Module	Input	Expected output	Actual output	Result
Data preprocessing	Preprocessed csv file	Successfully cleansed data	Taking dataset to project	pass
Prediction using ARIMA	Preprocessed csv file with reduced feature	Accuracy according to ARIMA	Accuracy according to ARIMA	pass
Prediction using SVR	Preprocessed csv file with reduced feature	Accuracy according to SVR	Accuracy according to SVR	pass
Classification using Hybrid Model	Preprocessed csv file with reduced feature	Accuracy and Prediction according to Hybrid Model	Accuracy and Prediction according to Hybrid Model	pass

Table 7.1: System Testing

CONCLUSION

All in all, predicting a price-related variable is difficult given the multitude of forces impacting the market. Add to that, the fact that prices are by a large extent dependent on future prospects rather than historic data. However, using machine learning algorithms has provided us with a better understanding of cryptocurrencies, and ARIMA architecture. Other features can be considered It is essential to predict the future prices of cryptocurrencies as they will help in making more profits and reducing losses. Thus, our proposed model which gives accurate results, can predict the prices of coins, the highest and lowest prices during that period.

REFERENCES

[1] Bitcoin price prediction using machine learning.

https://www.researchgate.net/publication/352063200_BITCOIN_PRICE_PREDICTION_USING_MACHINE_LEARNING

[2] Real – Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis.

https://www.researchgate.net/publication/342464470_Real-Time_Prediction_of_BITCOIN_Price_using_Machine_Learning_Techniques_and_Public_Sentiment_Analysis

[3] Predicting the Price of Bitcoin using Machin Learning.

https://ieeexplore.ieee.org/document/8374483

[4] Crypto-Currency price prediction using Decision Tree and Regression techniques.

https://ieeexplore.ieee.org/document/8862585

[5] Enhancing bitcoin price fluctuation prediction using attentive LSTM and embedding network.

https://scholars.ln.edu.hk/en/publications/enhancing-bitcoin-price-fluctuation-prediction-using-attentive-ls

[6] Systematic Erudition of Bitcoin Price Prediction using Machine Learning Techniques.

https://ieeexplore.ieee.org/document/8728424

[7] Cryptocurrency Price Prediction Using Neural Networks and Deep Leaning.

https://ieeexplore.ieee.org/document/9441872

[8] Bitcoin price prediction in the time of COVID-19.

https://ieeexplore.ieee.org/document/9382525

[9] Bitcoin Price Prediction Using Deep Learning and Real Time Deployment.

https://ieeexplore.ieee.org/document/9362735

[10] A Predictive Model for the Global Cryptocurrency Market: A Holistic Approach to Predicting Cryptocurrency Prices.

https://ieeexplore.ieee.org/document/8699292

[11] Deep Learning-Based Cryptocurrency Price Prediction Scheme With Inter-Dependent Relations.

https://ieeexplore.ieee.org/document/9558869

[12]A Research On Bitcoin Price Prediction Using Machine Learning Algorithms.

https://www.ijstr.org/final-print/apr2020/A-Research-On-Bitcoin-Price-Prediction-Using-Machine-Learning-Algorithms.pdf