Bitcoin Price Prediction Using Machine Learning Models

Dr.Priyanka Singh
Dept of Computer science & engineering
SRM UNIVERSITY AP
Amaravati, India
priyanka$_s@srmap.edu.in$

Srinadh Doppalapudi
Computer science & engineering
SRM UNIVERSITY AP
Amaravati, India

Nandini Gogineni
Computer science & engineering
SRM UNIVERSITY AP
Amaravati, India

Meghana Bollam
Computer science & engineering
SRM UNIVERSITY AP
Amaravati, India

Shanmukh Mittapally
Computer science & engineering
SRM UNIVERSITY AP
Amaravati, India

***Abstract*—The volatile nature of cryptocurrency markets has led to a growing interest in developing accurate and effective prediction models for digital asset prices. However, prices of Bitcoin have highly fluctuated which makes them very difficult to predict. This study endeavors to enhance the precision of Bitcoin price forecasting by integrating diverse parameters influencing its valuation.. The task is achieved with varying degrees of success through the implementation of a Long Short Term Memory (LSTM) network where we have built an LSTM Model to predict the Close Price of Bitcoin. The LSTM model for time series forecasting is implemented as a comparison to the deep learning models. Our findings underscore the exceptional performance of the Mean Gamma Deviance Regression Loss (MGD), yielding a minimal loss measure of 0.002 and an impressively high R-Square (R2) coefficient of 99.5%. our study seamlessly integrates data collection, feature selection, and data preparation to enhance the precision of Bitcoin price forecasting.**

***Index Terms*—Bitcoin; cryptocurrency; Machine learning; Deep Learning; Long Short Term Memory**

I. Introduction

This research aims to discover the most efficient and highest accuracy model to predict Bitcoin prices from various machine learning algorithms. Bitcoin allows for peer-to-peer transactions without the need for a bank or governing authority [2]. Due to its acceptance in over 40 countries like Germany, Canada, and Croatia Bitcoin has gained popularity. Has also paved the way for other alternative cryptocurrencies. Moreover, Bitcoin is used as a medium for trading cryptocurrencies and conducting transactions for different products and services [11] [6]. Since its creation in 2009, Bitcoin has managed to remain secure against hacking attempts thanks to its blockchain technology. This technology encrypts each coin with a digital signature making it easier to track and promote trust. In this system, every owner signs off on a hash, from the transaction and adds the public key of the next recipient before transferring it. This secure and transparent process ensures that each Bitcoin transaction is genuine and maintains its integrity [13] [1]. Bitcoin was sold for $1,000 in January 2017. It reached $16,000 USD by the end of December 2017. In July 2021, Bitcoin was already worth $32,896 more USD by then. The market for cryptocurrencies is known to be highly volatile. Out of the numerous cryptocurrency markets, Bitcoin stands out as the most attractive to investors' eyes [10]. The reason behind this is that it has unique characteristics including that it offers some level of secrecy yet it is fully transparent in terms of its underlying system. Open, close, high, low – there are several factors influencing bitcoin's value and they will be discussed in the context of the present project. The dataset comprises daily data from 17th September 2014 until 15th August 2023. It begins by testing the data using some regression techniques before implementing several deep learning models which are preferred due to their effectiveness in scenarios where large amounts of data sets are available.

Identify applicable funding agency here. If none, delete this.

Discussion: The Data visualization shows the correlation between all the features and only the five selected features have a sharp correlation. Data is then fitted into the model using the predefined commands accessible to Python. These data models were trained and tested out with a limited number of data sets and provided the result. With the growing technology and the rise in the data sets we can still work on the model with various other alternative cryptocurrencies. The model shows a better prediction rate for LSTM but with a very slight difference compared to the linear regression model.

II. Background Work

This research uses some of the main libraries like Scikit-learn and Keras for analyzing data in order to create machine learning models. In this research, the Tensorflow library is also used to generate data flow graphs.

1) **Scikit-learn**: Scikit-learn is an open-source library for analyzing data mining. Python is used to analyze and

create models from various machine learning algorithms, such as classification, regression, and clustering. Scikit-learn can also be used for preparing data in several ways: normalization, standardization, and cleaning outlier data or missing data [6].

2) **Tensorflow:** TensorFlow is an open-source deep learning framework developed by Google.TensorFlow's versatility allows developers and researchers to work on a wide range of tasks, from image and speech recognition to natural language processing and more. The framework's defining feature is its flexible computation graph system, enabling users to define complex models and optimize them efficiently for various hardware platforms [12].

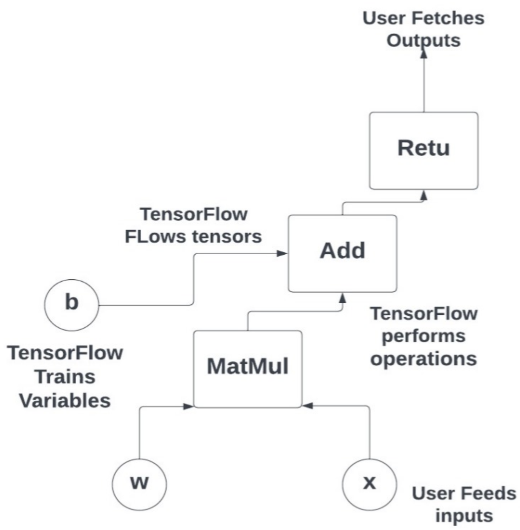

Fig 1: Example of data flow processing in Tensorflow.

3) **Keras:** Keras is an open-source library used for high-level NN. It's designed to provide a user-friendly and intuitive interface for building and training neural networks. Keras was initially developed as an independent project, but it has been integrated as the official high-level API into TensorFlow, making it a widely used tool for building neural network models [12].

4) **Matplotlib:** Matplotlib is a widely used data visualization library for Python. It provides a comprehensive set of tools for creating a wide range of static, animated, and interactive visualizations in Python scripts, notebooks, and applications. Matplotlib's flexibility allows users to create various types of plots, including line plots, scatter plots, bar plots, histograms, heat maps, and more [5].

III. Methodology

A. Data Collection

Data collection is the step, in any research project involving gathering information, for decision-making. Various methods are employed based on the data requirements with accuracy and reliability being paramount. In this scenario, the dataset comprises transactions spanning from September 9th, 2014 to August 15th, 2023. Initially, regression techniques were utilized to analyze the data however these methods did not yield results. Thus a deep-learning model was implemented—a form of machine learning of discerning patterns from data—which ultimately yielded more accurate outcomes than regression techniques.

```
maindf.head()
```

	Date	Open	High	Low	Close	Adj Close	Volume
0	2014-09-18	456.859985	456.859985	413.104004	424.440002	424.440002	34483200
1	2014-09-19	424.102997	427.834991	384.532013	394.795990	394.795990	37919700
2	2014-09-20	394.673004	423.295990	389.882996	408.903992	408.903992	36863600
3	2014-09-21	408.084991	412.425995	393.181000	398.821014	398.821014	26580100
4	2014-09-22	399.100006	406.915985	397.130005	402.152008	402.152008	24127600

Fig 2: Display of the Data collection.

B. Feature Selection:

The selection of the most pertinent features from a dataset is referred to as feature selection. The reason for this is that all characteristics do not hold the same value. Different characteristics may be relevant for various forecasts. We got better results using a smaller set of very important features by using data visualization to highlight those. Visualising Data- The Basics. It improves readability where data is written and creates a pattern or trend. One technique we employed was the use of different data visualizations like bar charts, line graphs, and scatter plots.In the process, this enables us to tell something about data relationships through images.Selection of a feature is crucial.It enhances the accuracy of the model, making it possible to work with fewer data. Feature selection and data visualization. It is helpful to convert data into something logical and to find out tendencies or trendiness. Nowadays, data visualization is an essential feature in our world. In this paper, I use the scikit-learn libraries to build a model only on its Closed, Open, High and Low when the Close price is to predict.

Features	Definition
Close	latest trade
Open	opening trade
High	highest trade during day
Low	lowest trade during day
Weighted price	mean Bitcoin price
Volume_(BTC)	total trade volume of day in BTC
Volume_(Currency)	total trade volume of day in USD
Timestamp	data recorded time

With the help of data visualization libraries, we can see the correlation between features and pinpoint the ones which we require. A sample image is shown below to show the correlation graph between the features in the given data set.

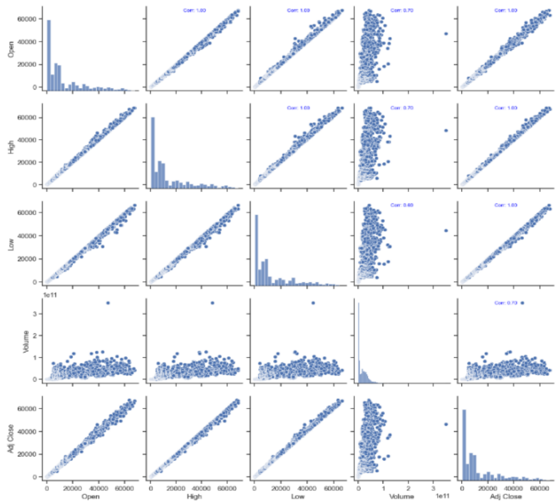

Fig 3: Correlation graph between the features.

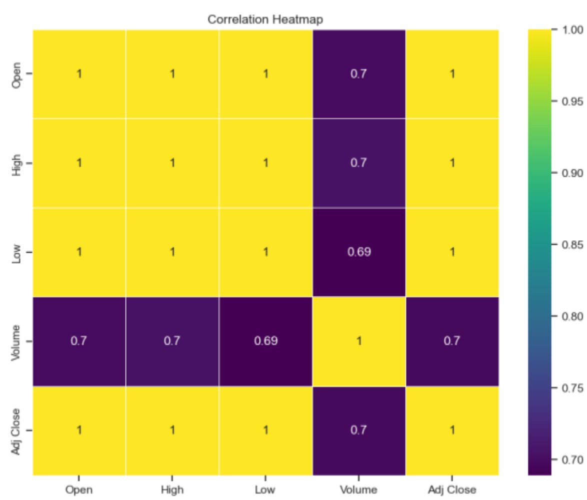

Fig 4: Heat map correlation coefficient values for each parameter of features.

C. Data Preparation:

When working with data that is measured in different scales, it is important to ensure that each variable contributes fairly to model fitting. Variables with varying scales can introduce bias into the learned model function, which can impact accuracy and outcomes [7]. To mitigate this, standardizing or normalizing the data is crucial for achieving better results. In Machine Learning and Deep Learning, where stability and speed are necessary for backpropagation, accurate data scaling becomes vital. By balancing and improving the efficiency of the learning process, accurate data scaling leads to enhanced model performance.

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

D. Algorithm Implemented:

Long Short-Term Memory (LSTM) LSTM is an important concept of deep learning, especially when it comes to RNN in solving the vanishing gradient problem. The errors may become too short or too long at training causing inadequate learning [3]. Unlike RNN, LSTM solves this problem by allowing error signals to freely flow backward through many unfolded layers thereby not allowed to evaporate or explode. In particular, LSTM has proven its worth for time-series data since events spread over many discrete time steps. This method operates on signals that have extended periods of delay between significant events and involves handling signals bearing low- and high-frequency components. LSTM has been successful when it comes to predicting time series data including stock predictions by researchers. The accuracy of the algorithm is much better than others hence is preferred for use in this field. The main formulation of the result in LSTM is based on Root Mean Square Error, R square score for regression, the equation.

This is the formula RSME = $\sqrt{1/N} \sum_{f=1}^{N} (x_i - \widehat{x_i})^2$

Where N is the total number of observations,x_i is the actual value; whereas,$\hat{x_i}$ is the predicted value. The main benefit of using RMSE is that it penalizes large errors. It also scales the scores in the same units as the forecast values [8].

E. Variance Regression:

The variance regression score indicates how closely the variance of the predicted values approximates the variance of the actual values. The larger variance regression score suggests that the model is more accurately forecasting the variance of the realized values in the field of predictive modeling. The R^2 score is one measure that indicates how closely the regression model "fits" the variability observed in the dependent variable. The value of R^2 is particularly relevant in predictive research as it is important in gauging the credible nature of any model prediction. A high R^2 score indicates that the model well represents underlying trends and patterns of the data, thus, it can be considered as an effective predictor. The resulting standard allows researchers not to measure only the model's performance but to take informed considerations concerning the choice of models, adjustments, and application of models in practice.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Where:
• n is the number of data points.
• y_i is the actual value of the dependent variable for the ith data point.
• $\hat{y_i}$ is the predicted value of the dependent variable for the ith data point.
• $\bar{y}$ is the mean of the actual values of the dependent variable.

F. Experimental Result:

After data analysis, we found out that exactly two features were suitable for this project's testing. Figure 5 shows the extracted features that were obtained after removing the irrelevant data.

```
In [68]: closedf = maindf[['Date','Close']]
         print("Shape of close dataframe:", closedf.shape)

         Shape of close dataframe: (3254, 2)
```

```
In [71]: closedf
Out[71]:
```

	Date	Close
1766	2019-07-20	10767.139648
1767	2019-07-21	10599.105469
1768	2019-07-22	10343.106445
1769	2019-07-23	9900.767578
1770	2019-07-24	9811.925781
...	...	...
3249	2023-08-11	29397.714844
3250	2023-08-12	29415.964844
3251	2023-08-13	29282.914063
3252	2023-08-14	29408.443359
3253	2023-08-15	29395.177734

1488 rows × 2 columns

Fig 5. Features selected are Close, Date.

G. Analysis of the year 2023:

The year 2023 was a volatile year for the Bitcoin market. The charge of Bitcoin fluctuated wildly, attaining a high of $23,919 in January and a low of $23,919 in July. The usual trend of the rate of Bitcoin in 2023 changed downward, however, there were several factors that could have contributed to the modifications in the charge. We have taken Bitcoin's monthly high and occasional prices as well as their open and close charge and combining this we've plotted an inventory analysis chart for the entire year.

Fig 6. Month comparison between Bitcoin open and close price.

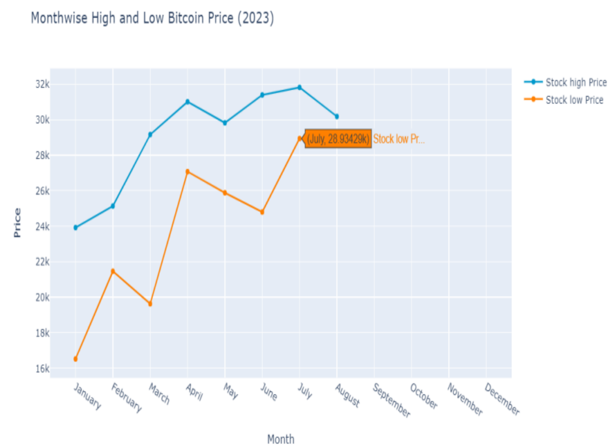

Fig 7. Month Wise High and Low Bitcoin Price (2023).

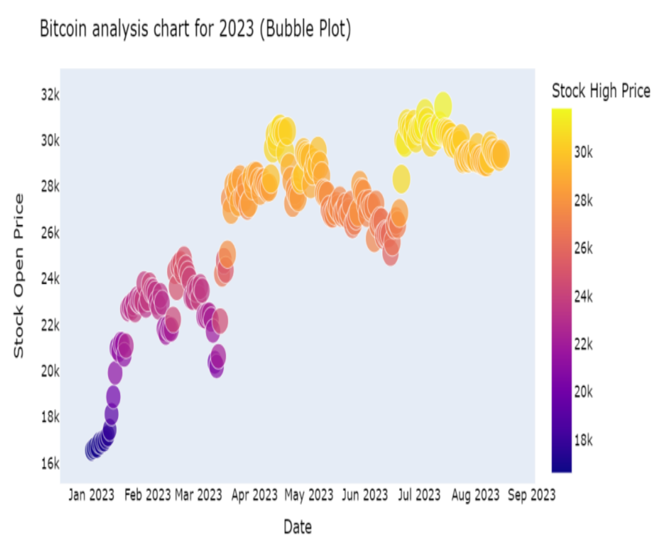

Fig 8. Bitcoin analysis chart for 2023.

H. Overall Analysis from 2014-2023:

The duration from 2014 to 2023 witnessed the superb evolution of Bitcoin, the arena's first cryptocurrency.

Starting as an insignificant curiosity in 2014, Bitcoin's adventure over the subsequent decade has been marked by intense volatility, regulatory challenges, meteoric rate surges, and a growing recognition in mainstream finance.

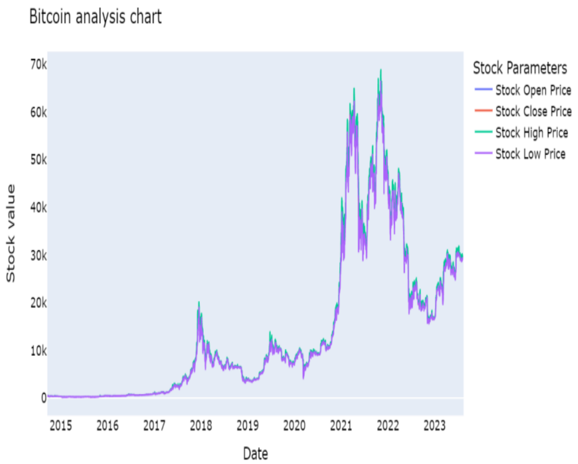

Fig 9. Overall Bitcoin analysis chart from 2014-2023.

Now the last day close the price data of Bitcoin before we train and test and predict the results.

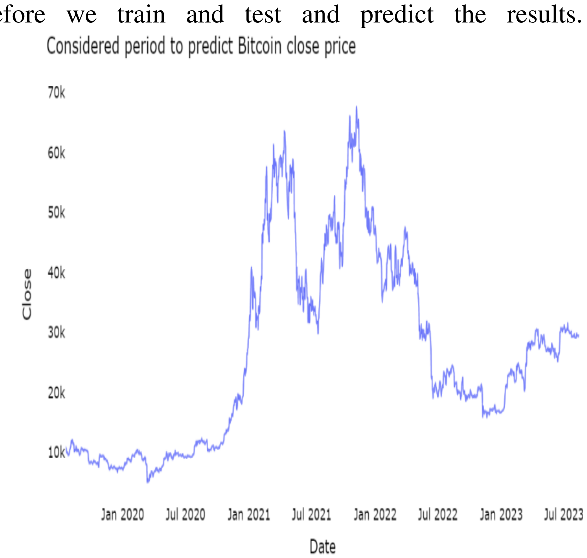

Fig 10. Considered period to predict Bitcoin close price.

For Step Split with training data we split 60 data period for training and 40 year for testing. Split at 1462 data in. If we split the data into trains and validation, we cannot use random splitting because that would lose the time component. We can see the effects of different models, one is a machine-learning model of Variance regression and the other is the R Square Score for regression. Variance regression usually works based on the mean square equation which tells us how accurate a linear graph of a continuous-time range data set is. We find that the accuracy of the training data is about 99.55 and the accuracy of the test data tends to be about 98.54 Accordingly in Figure 10a It is shown that for the R class, the accuracy of the training data is about 99.55 and the accuracy of the testing data is about 97.98.

```
In [85]: print("Train data explained variance regression score:",
               explained_variance_score(original_ytrain, train_predict))
         print("Test data explained variance regression score:",
               explained_variance_score(original_ytest, test_predict))

         Train data explained variance regression score: 0.9955710764955685
         Test data explained variance regression score: 0.9854746500234514
```

```
In [86]: print("Train data R2 score:", r2_score(original_ytrain, train_predict))
         print("Test data R2 score:", r2_score(original_ytest, test_predict))

         Train data R2 score: 0.9955224934607351
         Test data R2 score: 0.9798095070067485
```

Fig 10a: Accuracy is obtained from the training and testing data set using linear regression mode.

I. Evaluating Model Performance:

A Comparison of Original Bitcoin Close Prices and Predicted Close Prices. The comparison between original and predicted Bitcoin close prices serves as a vital diagnostic tool for evaluating the effectiveness of the prediction model. It aids in assessing whether the model's predictions align closely with the actual price movements or if there are significant disparities that might require further investigation or model refinement. Ultimately, this comparison sheds light on the model's predictive power, its ability to handle the inherent volatility of cryptocurrency markets, and its potential to provide valuable insights for traders, investors, and stakeholders in the Bitcoin ecosystem.

```
Train predicted data:  (1488, 1)
Test predicted data:  (1488, 1)
```

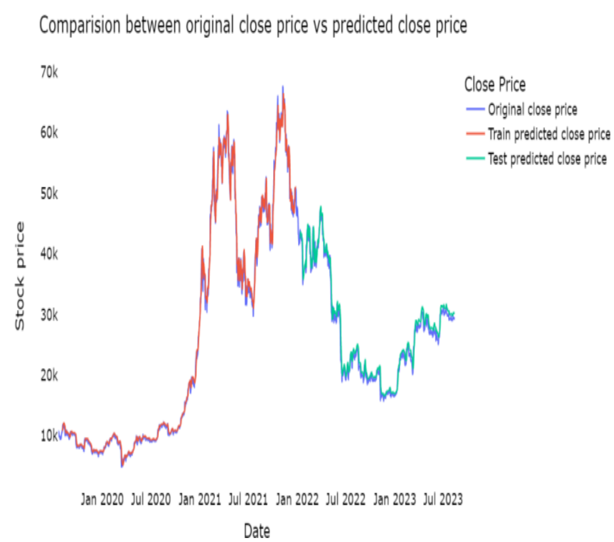

Fig 10b. Comparison between original close price vs predicted close price.

The last 15 days of the dataset and the next 30 days of predicted prices will then be plotted.The output of the code will be a plot of the closing prices, with the last 15 days of the dataset and the next 30 days of predicted prices highlighted.

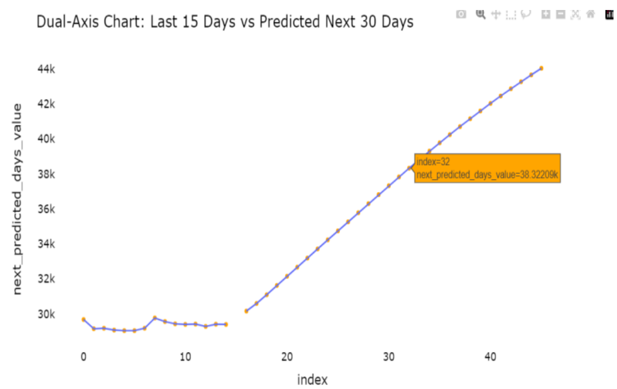

Fig 10c: Last 15 Days vs Predicted Next 30 Days.

Finally, we're producing a graph of the entire Closing Stock Price with subsequent 30 days period of prediction. The research paper presents a comprehensive visualization that merges historical tendencies with future projections. The entire Closing Stock Price trajectory is elegantly depicted, with a centered lens on the next 30 days of prediction. This method enables us to not handiest realise past market behavior but also anticipate capability tendencies, contributing to deeper information on the cryptocurrency market dynamics. The visualization serves as an effective tool for buyers and analysts alike, supplying insights that bridge the historical facts and predictive fashions, guiding choice-making tactics, and fostering knowledgeable strategies.

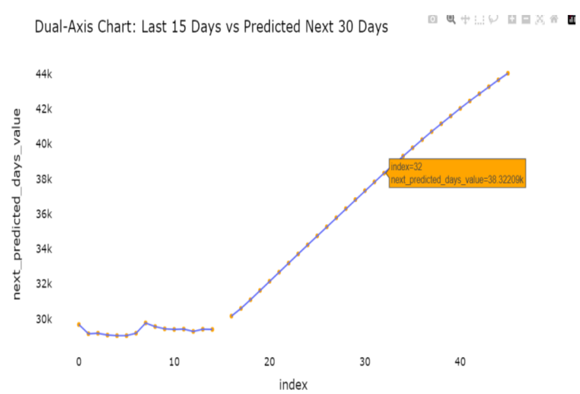

Fig 11. Plotting the whole closing Bitcoin price with prediction.

J. Proposed Work

Investing in cryptocurrencies, like Bitcoin is an option for growing your wealth in today's technology-driven world. Unlike currencies, cryptocurrencies are not tied to any government or country making them accessible to anyone worldwide without the concern of cross-border taxes [4]. The success of Bitcoin with its growth and value has led to the emergence of other cryptocurrencies with slight variations [3]. What makes crypto investing appealing is the convenience of buying and selling assets through user platforms like WazirX and Finance. Opening an account is as easy as creating an email or social media profile where you only need to provide details and proof of identity. In contrast to the stock market crypto trading operates on a peer-to-peer network eliminating the need for intermediaries like brokers or agents. It offers the flexibility to invest in fractions of assets. Allows people with schedules during daytime hours to participate equally since the crypto market operates 24/7.Investing in cryptocurrencies allows for participation by individuals with different schedules since the crypto market operates 24/7. Crypto exchanges are notably faster than ones due to blockchain technology ensuring instant transactions that provide transparency and efficiency. Overall investing in cryptocurrencies presents an opportunity for investors who seek financial growth. It offers decentralization, global accessibility, user-friendly platforms, fractional ownership options, and speedy blockchain transactions that redefine engagement [3] [9]

.

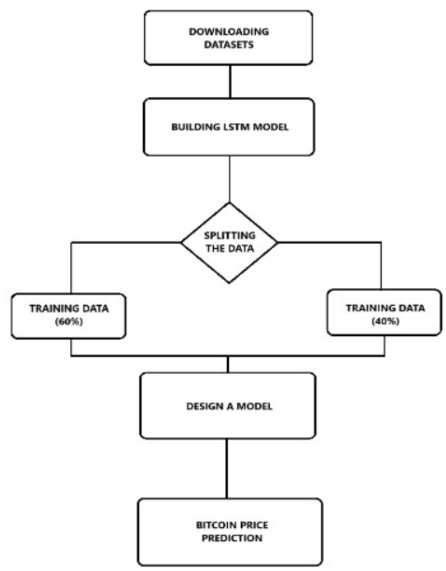

Fig 12: A Block Diagram for Price Prediction.

Limitations:

While cryptocurrency trading has become popular it faces hurdles with scalability being a concern. The growth of coins. The adoption of blockchain technology pales in comparison to the transaction volumes handled by industry giants like VISA. Crypto transactions are relatively slower. Still have a ways to go before matching the speed of established players such as VISA and MasterCard without infrastructure scaling. The cryptocurrency market is highly volatile and unpredictable with no guaranteed accuracy in predictions. It is influenced by sentiment and unexpected actions like a Bitcoin holder suddenly selling their entire asset can result in significant market fluctuations. With technology, predicting human emotions remains challenging. Technical analysis of charts typically revolves around three aspects; trend and momentum which indicate direction and strength; support and resistance levels that highlight stopping points; and overall patterns that shed light on market psychology. Cryptocurrencies lack the data available for traditional markets like stocks, currencies, and commodities making it difficult to identify reliable resistance levels and key support levels. This limited historical data complicates prediction. Trading practices, within the cryptocurrency market. The lack of data makes it more difficult to predict and engage in trading activities, within the crypto market. Although crypto trading has become increasingly popular it encounters obstacles in terms of scalability.

IV. Future Work:

The LSTM model, implemented here, is a basic model that takes into consideration only a few features that affect the Bitcoin price. Our model is fairly accurate when predicting future prices. However, to increase the efficiency of the model, more Bitcoin price features need to be taken into consideration. We recommend using Yahoo Finance is the source of datasets since the information present on this website holds a high degree of authenticity. Our future work would include in-depth scrutinisation on the topic of LSTM, and deep learning at large. Such fact-findings would be beneficial for forecasting the prices of cryptocurrencies with the help of LSTMs, in the future. The Data visualization shows the correlation between all the features and only the two selected features have a sharp correlation. Data is then fitted into the model using the predefined commands accessible to Python. These data models were trained and tested out with a limited number of data sets and provided the result. With the growing technology and the rise in the data sets, we can still work on the model with various other alternative cryptocurrencies.

V. Conclusion:

This study is used to compare the features: open, close, high, and low only, hence the result may differ if we tend to take various other features into consideration.Because the crypto market is volatile and influenced by social media and other external factors, data sets cannot be the only reason for forecasting. As technology advances, new data can be collected, analyzed, and practiced, resulting in better results for this experiment. This research underscores the importance of continuous model refinement and evaluation, as well as a holistic consideration of both technical and fundamental factors when making investment decisions based on price predictions. The integration of LSTM models in Bitcoin price prediction contributes to the ongoing discourse surrounding the application of machine learning techniques in financial markets and sets the stage for further innovation and research in this evolving field.

• To work on a better User Interface so that people can access these data easily and effortlessly.

• Implementing the IOT model for smart automatic analysis.

• Implementing more algorithms to find out the best method for predicting the cryptocurrency.

References

[1] Junwei Chen. Analysis of bitcoin price prediction using machine learning. *Journal of Risk and Financial Management*, 16(1):51, 2023.

[2] Ferdiansyah Ferdiansyah, Siti Hajar Othman, Raja Zahilah Raja Md Radzi, Deris Stiawan, Yoppy Sazaki, and Usman Ependi. A lstm-method for bitcoin price prediction: A case study yahoo finance stock market. In *2019 international conference on electrical engineering and computer science (ICECOS)*, pages 206–210. IEEE, 2019.

[3] Alvin Ho, Ramesh Vatambeti, and Sathish Kumar Ravichandran. Bitcoin price prediction using machine learning and artificial neural network model. *Indian Journal of Science and Technology*, 14(27):2300–2308, 2021.

[4] Patrick Jaquart, David Dann, and Christof Weinhardt. Short-term bitcoin market prediction via machine learning. *The journal of finance and data science*, 7:45–66, 2021.

[5] H Kavitha, Uttam Kumar Sinha, and Surbhi S Jain. Performance evaluation of machine learning algorithms for bitcoin price prediction. In *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, pages 110–114. IEEE, 2020.

[6] Ruchi Mittal, Shefali Arora, and MPS Bhatia. Automated cryptocurrencies prices prediction using machine learning. *Division of Computer Engineering, Netaji Subhas Institute of Technology, India*, 8:2229–6956, 2018.

[7] Darwin Nesakumar, Srinath Bhavyasri, Tanguturi Keerthana, Yelisetty Meghana, KN Pavithra, and P Mugilan. Bitcoin price prediction using deep learning and neural networks. In *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*, pages 1055–1058. IEEE, 2022.

[8] Thearasak Phaladisailoed and Thanisa Numnonda. Machine learning models comparison for bitcoin price prediction. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pages 506–511. IEEE, 2018.

[9] K Ramani, M Jahnavi, P Jagadeesh Reddy, P VenkataChakravarthi, P Meghanath, and SK Imran. Prediction of bitcoin price through lstm, arima, xgboost, prophet and sentiment analysis on dynamic streaming data. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 1514–1518. IEEE, 2023.

[10] Prachi Vivek Rane and Sudhir N Dhage. Systematic erudition of bitcoin price prediction using machine learning techniques. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 594–598. IEEE, 2019.

[11] Ashutosh Shankhdhar, Akhilesh Kumar Singh, Suryansh Naugraiya, and Prathmesh Kumar Saini. Bitcoin price alert and prediction system using various models. In *IOP Conference Series: Materials Science and Engineering*, volume 1131, page 012009. IOP Publishing, 2021.

[12] Sakshi Tandon, Shreya Tripathi, Pragya Saraswat, and Chetna Dabas. Bitcoin price forecasting using lstm and 10-fold cross validation. In *2019 International Conference on Signal Processing and Communication (ICSC)*, pages 323–328, 2019.

[13] Yike Xu. *Bitcoin Price Forecast Using LSTM and GRU Recurrent networks, and Hidden Markov Model*. University of California, Los Angeles, 2020.