BITCOIN PRICE PREDICTION USING MACHINE LEARNING

M.Jeevana Sujitha[1], V.Anjani Kranti[2], P.Saroja[3]

1 Assistant Professor, Department of CSE,SRKR College of Engineering, Bhimavaram, West Godavari, AP, India.
2 Assistant Professor, Department of CSE,SRKR College of Engineering, Bhimavaram, West Godavari, AP, India.
3 Assistant Professor, Department of CSE,SRKR College of Engineering, Bhimavaram, West Godavari, AP, India.

---***---

Abstract -The objective of this paper is to create a predictive model using past Bitcoin prices that can precisely predict its future prices. The model should be capable of identifying the intricate patterns and trends in Bitcoin's value fluctuations, which can be impacted by numerous factors such as market sentiment, regulatory changes, adoption rates, and other macroeconomic factors.The model should be able to learn from past data and make accurate predictions for future prices, which can help traders and investors make informed decisions and manage their risk exposure.

***Key Words*:** Bitcoin, Cryptocurrency, Priceprediction,LSTM(LongShort-Term Memory),Model Evaluation,Data Preprocessing Feature Selection,Model Performance Metrics

I.INTRODUCTION

Bitcoin is a digital currency that operates on a decentralized peer- to- peer network. It's not controlled by any central authority, and its value is determined by request forces of force and demand. The price of Bitcoin has shown tremendous volatility over the times, with frequent sharp rises and falls. In 2017, the popularity of crypto currencies increased significantly due to their rapidly growing request capitalization for several consecutive months. Presently, there are over 1,500 crypto currencies being traded, with a combined value of over $300 billion. In January 2018, the total request capitalization peaked at over $800 billion. A recent survey revealed that between 2.9 and 5.8 million investors, both private and institutional, are involved in various sales networks. With time, it has become easier to access these crypto currencies.

Machine literacy is a type of artificial intelligence that enables computers to learn from data and make prognostications or opinions without being explicitly programmed. In the environment of Bitcoin price validation, machine literacy models can be trained on literal price data and other applicable features, similar as trading volume, news sentiment, social media exertion, and mining difficulty, to identify patterns and trends in the data. There are colorful machine literacy ways that can be used for Bitcoin price validation, similar as direct retrogression, support vector machines, neural networks, and ensemble styles. The choice of fashion depends on the nature and complexity of the data, as well as the performance conditions and computational coffers available.

Once a machine literacy model is trained and validated, it can be used to prognosticate the unborn price of Bitcoin with a certain position of accuracy. Such prognostications can help investors and dealers make informed opinions about buying or dealing with Bitcoin, and can also give perceptivity into the underpinning factors that drive Bitcoin's price movements,. Overall, a Bitcoin price vaticination ML design can be a precious and grueling bid, as it requires moxie in machine literacy, finance, and cryptocurrency, as well as a deep understanding of the request dynamics and factors that influence Bitcoin's price.

II.EXISTING SYSTEM

Research into stock market prediction techniques has demonstrated their adaptability for forecasting cryptocurrency prices. Nonetheless, machine learning applications within the cryptocurrency sphere have primarily focused on Bitcoin. Various algorithms, including random forests, Bayesian neural networks, and long short-term memory neural networks, have been leveraged to analyze Bitcoin prices. These studies have exhibited varying degrees of success in predicting Bitcoin price fluctuations, with neural network-based algorithms outperforming others.

Over the span of a year, deep reinforcement learning has proven to be more efficacious than the conventional buy-and-hold strategy when it comes to forecasting the prices of a diverse array of twelve cryptocurrencies.

III. PROPOSED SYSTEM

- The objective of our study was to evaluate the effectiveness of the models in forecasting the daily prices of 1,681 crypto currencies. Out of the three models, two were developed using gradient boosting decision trees while the third one utilized long short-term memory (LSTM) recurrent neural networks. The data shows that all three types of models are more effective than a basic model that uses a simple moving average to predict currency prices based on past prices. Additionally, the model that utilizes long short-term memory recurrent neural networks consistently generates the highest profit on investment.

IV. SYSTEM ARCHITECTURE

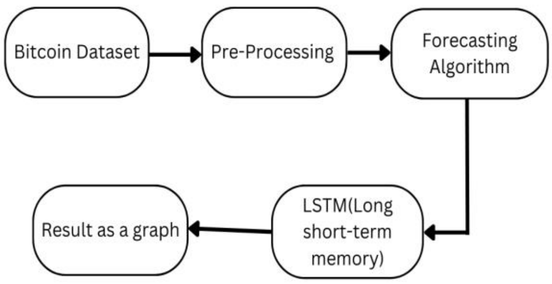

Fig.1. Architectural design for Bitcoin Price Prediction

The main aim of this project is to examine the degree of accuracy in predicting the direction of bitcoin prices through machine learning methods. The research problem deals with time series prediction. Although numerous studies have been conducted on various machine learning techniques in time series prediction, there is a lack of research on bitcoin specifically. Bitcoin is currently in a transitional phase, which makes it more unstable compared to other currencies like the USD. Despite this, it is the most successful currency in the past five years, making it an attractive subject for research. A study of existing literature has revealed that using a GPU instead of a CPU can lead to significant improvements in machine learning algorithm performance. To investigate this, the RNN and LSTM networks are trained and benchmarked using both the GPU and CPU. This analysis provides valuable insights.

4.2 MODULES:

1. **Dataset Collection**
2. **Data Processing**
3. **Modelling**

4.2.1 DATASET COLLECTION:

To begin our analysis, we utilized Bitcoin market data that was publicly available on Kaggle2.The dataset provided a historical record of Bitcoin's performance from December 1st, 2014 to January 8th, 2018, with information collected every minute. This timeframe spanned over 1.5 million minutes and included data on opening and closing values, highest and lowest values, volume traded, and weighted price for each minute. Our primary objective was to analyze polarity trends in the market, so we initially focused on developing a model by labeling the dataset as true if the price increased at the end of the minute timestamp and false if it remained the same or decreased.

Link for Bitcoin Dataset:
https://www.kaggle.com/datasets/adityasakare/bitcoin-dataset

4.2.2 DATA PRE-PROCESSING:

When gathering information through scraping, we obtain a 2D tensor that comprises m samples and n features. In order to transform this data into windows data, we employed a time-series transform, using a window size of w=50. This transformation yielded a 3D tensor with a shape of (m-w) samples by n features by w day window size. To elaborate further, the initial data point for m=0 had a 2D tensor with m features for the first 49 days. Once the data was normalized, we separated it into input and output data, with the last day being excluded and used as the output data.

Three common data pre-processing steps are:

Formatting: If the data you have selected is not suitable for your needs, it could be because it is stored in a relational database when you require a flat file, or it is in a proprietary file format, and you need it in a text file or a relational database. Before starting work, it is essential to check that the data format matches your requirements.

Cleaning: Cleaning data involves the elimination or correction of absent data. It is possible that some data cases may be deficient and may not possess the information required to tackle the issue. In such cases, it may be necessary to eliminate them. Furthermore, some attributes may contain confidential data that may require anonymization or complete removal from the data.

Sampling: It is possible that there is an excess of data available for your use, which can lead to longer processing times for algorithms and increased memory requirements. To speed up exploring and creating prototypes, it may be beneficial to select a smaller, representative sample of the data before working with the entire dataset.

4.2.3 MODELLING:

During the process of developing a model, Model Evaluation plays a crucial role in determining the best model that accurately represents the data and predicts its future performance. It is not recommended to evaluate the model's performance using the same data used for training as it may lead to overfitted and overly optimistic models. In this particular work, a 3-layer bidirectional RNN was utilized to forecast Bitcoin's closing price by taking into account a range of data from the preceding days. To determine whether to keep or discard data, the use of LSTM layer was implemented taking into account the input, output and memory from previous occurrences. Moreover, the Dense Layer was utilized to calculate input by applying weight matrix and bias (if necessary) and then activating it using a certain function. Sigmoid activation function was chosen in this case because it results in either 0 or 1. The optimization method employed was Adam, and the loss function was RMSE.

4.3 PERFORMANCE ANALYSIS:

The root mean square error (RMSE) is a common method for assessing prediction accuracy. It calculates the distance between predicted and actual values using Euclidean distance. To derive the RMSE, the difference between predicted and true values is computed for each data point, and the residual norm for each data point is determined. The average of the residuals is then calculated and the square root of that mean is taken. RMSE is frequently utilized in supervised learning settings where true measurements are required for each predicted data point. It is a valuable metric for evaluating predictive models.

Root mean square error can be expressed as $RMSE = \sqrt{\sum_{i=1}^{N} \quad \| y(i) - \hat{y}(i) \|^2 / N}$

where N is the number of data points, y(i) is the i-th measurement, and y ̂(i) is its corresponding prediction.

V. IMPLEMENTATION TECHNIQUES

LSTM (Long Short-Term Memory): Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that has gained significant success in time series forecasting tasks, particularly in predicting the prices of Bitcoin. LSTM model can analyze historical price data of Bitcoin to recognize patterns that can be used to forecast future prices by considering the long-term dependencies between past prices. The model works by processing input data in a sequence of time steps while maintaining a memory state that allows it to remember crucial information from previous time steps. This memory state is updated at each time step based on the input data and the output from the previous time step. The model employs gates to regulate the flow of information into and out of the memory state, allowing it to selectively remember.

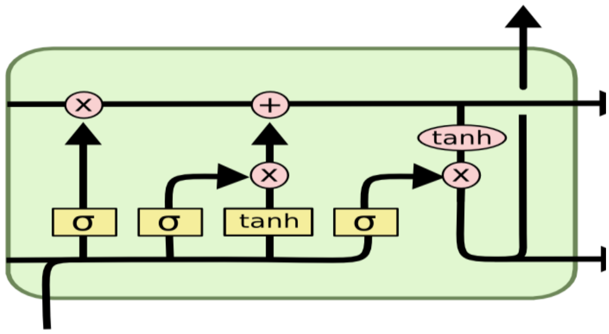

Fig.2. Internal working of LSTM

In order to predict Bitcoin prices using the LSTM model, the first step is to preprocess historical price data and input it into the model. A portion of the data is used to train the model, while another portion is used for testing and evaluating its performance.

Once the model has been successfully trained, it can be utilized to make predictions on new and previously unseen data. The LSTM model has demonstrated promising outcomes in Bitcoin price prediction and is widely used in this field. Nevertheless, it should be emphasized that predicting Bitcoin prices is a complicated task and no model can guarantee perfect accuracy.

VI. RESULT

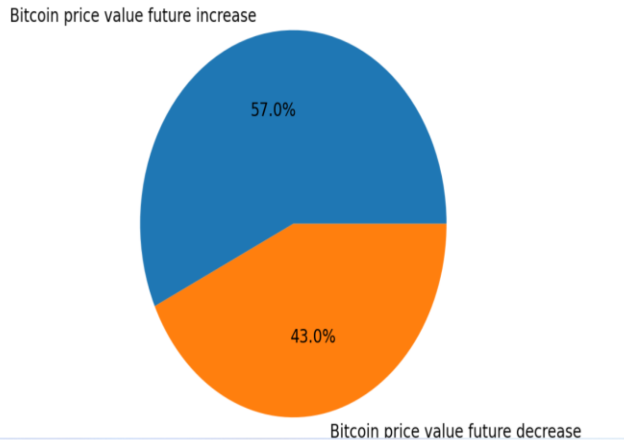

Fig.3 Bitcoin price future prediction

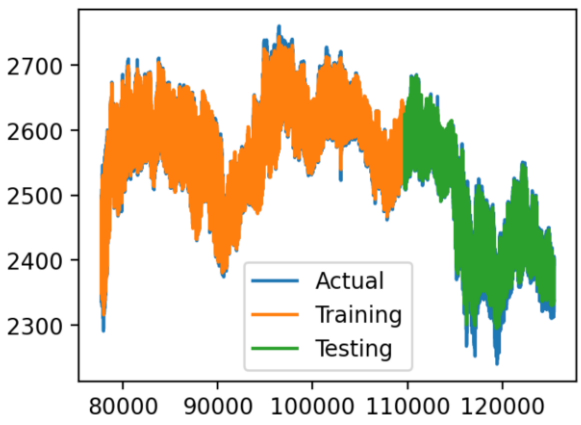

Fig.4. Graph for Actual - Training – Testing of Bitcoin

VII. CONCLUSION

Using LSTM (Long Short-Term Memory) for Bitcoin price prediction can be an effective approach due to its ability to capture long-term dependencies and temporal patterns in time-series data. The LSTM model can learn and adapt to complex relationships between past and present Bitcoin price movements, which makes it useful for predicting future price changes. However, it's important to note that Bitcoin's price is influenced by various factors, including global economic conditions, government regulations, and adoption rates. As a result, it can be difficult to accurately predict future prices using only historical price data. However, we managed to achieve the RMSE value as low as 0.2 for training and 0.45 for testing.

Therefore, while LSTM models can be helpful in forecasting Bitcoin prices, they should not be relied upon solely for investment decisions. It is important to consider multiple factors and conduct thorough research before making any financial decisions.

REFERENCES

[1] Huisu Jang,Jaewook Lee,"An Empirical Study on Modelling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information", in IEEE Early Access Articles, 2017.

[2] F. Andrade de Oliveira, L. Enrique ZÁrate and M. de Azevedo Reis, C. Neri Nobre "The use of artificial neural networks in the analysis and prediction of stock prices",2011.

[3] M. Daniele, A. BUTOI,"Data mining on Romanian stock market using neural networks for price prediction",2013.

[4] D. Shah, K. Zhang,"Bayesian regression and Bitcoin",2014.

[5] McNally, S,Roche, Caton, "Predicting the Price of Bitcoin Using Machine Learning",2016.

[6] Enas Al Kawasmi, Erin Arnautovic, Davor Svetinovic,"Bitcoin-Based Decentralized Carbon Emissions Trading Infrastructure Model",2015.

[7] Erik Casagrande, Selamawit Woldeamlak, Wei Lee Woon, , H. H. Zeineldin, and Davor Svetinovic,"NLP-KAOS for Systems Goal Elicitation Smart Metering System Case Study",2013.

[8] Zheshi chen,Chunhong Li,Wenjun Sun "Bitcoin price prediction using machine learning: An approach to sample dimension engineering", 2020.

[9]Ahmed M Khedr, Ifra Arif, Magdi El-Bannany, Saadat M alhashmi, Meenu Sreedharan, "Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey", 2021.

[10]Aniruddha Dutta, Saket Kumar, Meheli Basu, "A gated recurrent unit approach to bitcoin price prediction",2020.

[11]Apoorva Aggarwal, Isha Gupta, Novesh Grag, Anurag Geol, "Deep learning approach to determine the impact of socio economic factors on bitcoin price prediction",2019.

[12] Jure Leskovec, Kevin J. Lang, Michael W. Mahoney, "Empirical Comparison of Algorithms for Network Community Detection", 2010.