International Journal of
Electrical and Computer System Design
ISSN: 2582-8134 www.ijecsd.com

Performance Evaluation of Ethereum Price Prediction Utilizing Support Vector Regression

Ramya Sree KN[1], Vidhya A[2], Jotheeswaran G[3], Depak kumarasan C[4]

*1, 2, 3, 4 Department of Information Technology, Jeppiaar Engineering College, Chennai, India

Email: ramyasreen2002@gmail.com, vidhyaraj1988@gmail.com, jotheeswarangt@gmail.com, dk.101002098@gmail.com

Abstract: In recent years, the popularity and usage of cryptocurrencies have surged along with their values, and Ethereum has emerged as the second most renowned cryptocurrency, following Bitcoin. Cryptocurrencies are built on block chain technology, which is decentralized and has the potential to revolutionize traditional banking systems. This Python program utilizes the finance library to gather historical price data for the cryptocurrency Ethereum (ETH) from Yahoo Finance. The collected data is then employed to train a Support Vector Regression (SVR) model, which aims to predict the future price of ETH. This paper presents a comparative analysis of the prediction accuracy between the (SVR) model and the (LR) model. The program initially downloads and saves the data to a CSV file. Subsequently, the data is preprocessed by adding a column to the Data Frame containing the future price of ETH, projected for a specified number of days ahead. Using the train and test split function from the sci-kit-learn library, the programme then divides the data into training and testing sets. An SVR model with a radial basis function kernel is fitted to the training set. Finally, the model's performance is evaluated using the R-squared score, and the program generates forecasts for the price of ETH using the testing set. The actual and forecasted ETH price values are visualized using the matplotlib library, enabling a comprehensive comparison of the model's predictions.

Keywords: ETH, SVR, crypto currency, LR

1. Introduction

Cryptocurrencies are virtual currency that may be traded between people or groups [1]. Blockchain is a powerful technology that can recover any financial or banking institution. It has piqued the curiosity of computer scientists because of its diversified character. Cryptography and the spread of communication. Blockchain technology can provide a platform for exchanging money / money without the need for a middleman such as the government or other banking institutions. Finance analysis of time series [4], [5] in recent months prospered through the application of machine learning methods and techniques that have shown to be extremely effective in a variety of complex systems study domains [2], [3]. People may rapidly transfer money using cryptocurrencies at a low cost. It eliminates the problem of double spending and prevents fraud transactions from happening, and it can help users attain real data democracy [6], [7]. Following Bitcoin, more cryptocurrencies joined the crypto market like Ethereum, established in the year 2015, is the 2nd largest cryptocurrency with a $410.0 billion market capitalization. More than 5600 distinct cryptocurrencies are traded on about 1100 exchanges, with the most popular digital currencies being Ripple, Tether, Cardano, Stellar, Litecoin, and Zcash. The cryptocurrency sector is drawing more investors due to its high returns and quick expansion. The objective of this study is to assess the accuracy of the SVR and LR models in forecasting the price of Ethereum (ETH). This paper attempts to give useful insights into the application of machine learning for Ethereum price prediction. The comparison of SVR and LR models will aid in evaluating the efficacy of various regression approaches in projecting bitcoin prices. This study, which employs the SVR model, provides insight into the application of advanced machine learning methods for Ethereum price prediction, hence contributing to the creation of more accurate and dependable forecasting models in the cryptocurrency space.

2. Related Description of Works

The Cryptocurrencies stated in June 2022 that the suggested model evaluates multiple price prediction models and indicates that bidirectional LSTM is the best model to estimate the price of Ethereum among RNN, LSTM, and Bi-LSTM. To predict the price, the model takes the closing price as a parameter. When it comes to the financial market, knowing the trends is critical. This model provides a solid pricing trend over extended

time periods. The work "Predicting the price of cryptocurrency using support vector regression methods" that several research has previously been undertaken to forecast future values of a certain virtual currency using a machine- learning model. Few, however, have concentrated on employing alternative kernels of a "Support Vector Regression" (SVR) model. Ethereum was founded in 2013 by developer Work on the network commenced in 2014 and was crowdfunded, and it became functional. Anyone may develop permanent and unchangeable decentralized apps on Ethereum, with which users can interact and investigated the economic impact of cryptocurrency. Finally, calculated the association between cryptocurrencies and gold how they create impact on the stock exchanges. The shown gold may potentially be used as a hedging strategy to forecast the cryptocurrency market. The crypto currency prices may be influenced by a variety of variables, including a coin's popularity, mining costs, and purchasing behaviour. They looked at a variety of technical aspects that impact Bitcoin, Dash, and Monero pricing and trading volume. Many emotional reasons might influence traders' bearish and bullish tendencies. They discussed both positive and negative social media comments influence bitcoin pricing. Finally, used MAE and RMSE to assess the performance of conventional ML models such as SVM and random forest in real-time.

3 Performance Analysis of the Proposed Methodology in terms of Existing and proposed approach

3.1 Data Retrieval and Library Import

The first step in the methodology involves importing the required library files for data processing and analysis. The necessary libraries, including os, datetime, numpy, pandas, matplotlib.pyplot, yfinance, yahoofinancials, sklearn.model_selection, sklearn.linear_model, svm, and warnings, are imported. These libraries provide functions and tools for handling files, date and time operations, numerical computations, data manipulation, visualization, retrieving financial data from Yahoo Finance, implementing regression models, and managing warnings during code execution.

3.2 Dataset Download and Data Reading

The next stage focuses on downloading the Ethereum price dataset from Yahoo Finance using the yfinance library. The historical price data from a specified start date to the current date is retrieved. After downloading the data, it is read into a Pandas DataFrame. The index is set to be the Date column, and a new column is created to store the Close price of Ethereum shifted forward by a specified number of days, representing the future price to be predicted by the SVR model.

3.3 Data Splitting, Model Training, and Evaluation

Using the train_test_split function from the scikit-learn package, the data is divided into two sets: training and testing. The Close price data and future price data are extracted from the DataFrame. The SVR model with a radial basis function kernel is then trained using the training data. Additionally, a Linear Regression (LR) model is fitted to the training data for comparison. The accuracy of the SVR model is evaluated using the R-squared score method, and predictions are made on the testing data. Finally, the predicted and actual future prices of Ethereum are plotted on a graph using the matplotlib library, allowing for visual assessment of the model's performance.

By following this methodology, the research ensures the availability of the Ethereum price data, prepares it for analysis, splits it into training and testing sets, trains the SVR and LR models, evaluates their accuracy, and visualizes the predicted and actual prices.

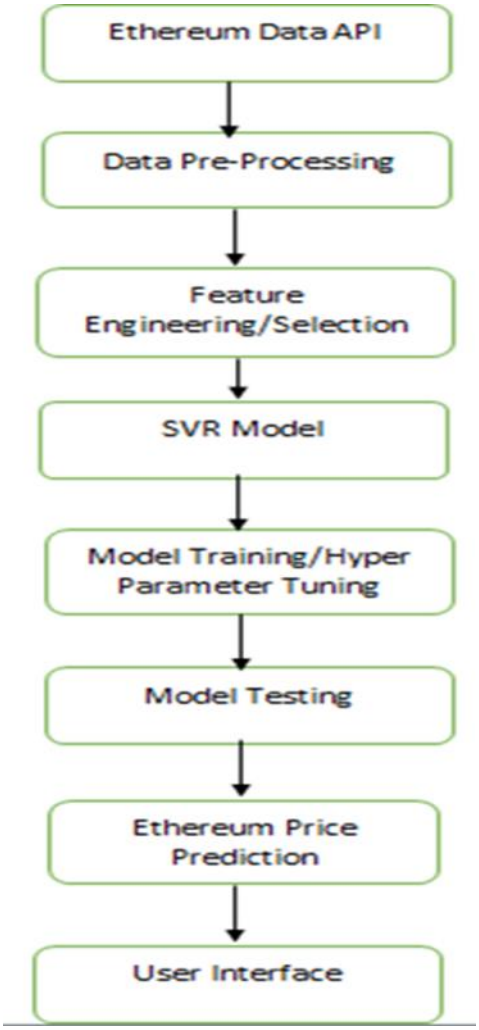

Fig 1. System Architecture

4. Methodology and Results

4.1 **Support Vector Regression (SVR)**

SVR is a ML approach for performing regression analysis. SVR finds a function that minimises errors in prediction by approximating the relationship between input variables and a continuous target variable. SVR looks for the hyperplane that suits the data points in uninterrupted space most effectively as opposed to SVMs, which are used for classification problems. This is accomplished by mapping the input variables to a high-dimensional feature space and locating the hyperplane that minimises prediction error while optimising the margin (distance) between each hyperplane and the nearest data points. By converting the data to a higher-dimensional space with the use of a kernel function, SVR is able to deal with non-linear correlations between the input and objective variables. As such, it's a great tool for regression problems with complicated interactions between input and target variables. SVR employs identical ideas as SVM to address regression-related problems. Let's take a few moments to comprehend the concept of SVR.

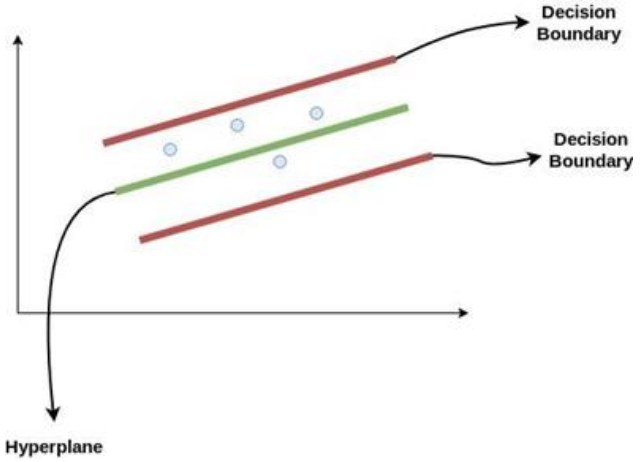

Fig 2. Hyperplane concept for SVR model

These 2 lines in red represent the decision boundary, while the line in green represents the hyperplane. The purpose of SVR is to merely assess the elements that are within the scope of the boundary line. The most suited line is the hyperplane with the most points. The initial phase in understanding this is identifying the decision limit (the danger red line shown above!). Assume these lines are at any distance from the hyperplane, let's say 'a'. So they're the lines formed from the hyperplane at '+a'.

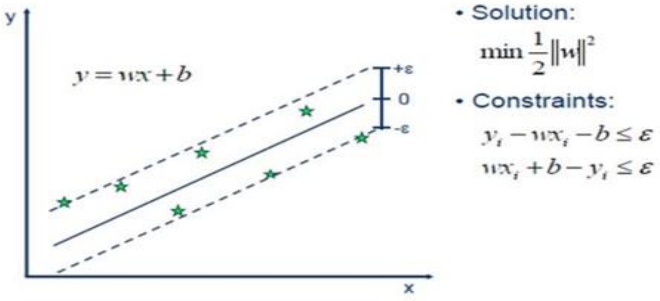

Fig 3. Equation of hyperplane

4.2 Linear Regression

LR is an easy statistical regression technique for figuring out how two continuous variables are related. As a result of LR's ability to show The x-axis of the variable that is

independent and the y-axis of the variable that is dependent are linearly related. Simple linear regression is the name given to linear regression when there is just one input variable (x). When there are several input variables, multiple linear regression is performed. The link between variables is shown by a slanted straight line produced by the linear regression model.

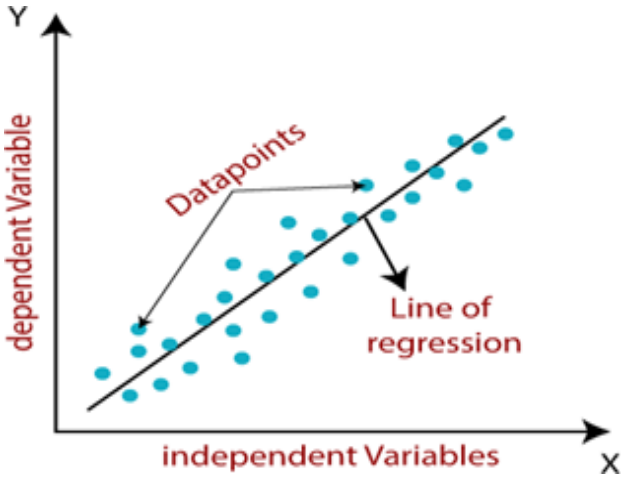

Fig 4. Linear Regression graph

The red line is referred to as the best-fit straight line. We aim to draw the best line possible based on the data points supplied. When the independent variable is x and the dependent variable is y. a0 stands for the line's intercept, while a1 is the coefficient of linear regression.

In order to obtain the most effective suited range, the linear regression approach looks for the best values for line's intercept and coefficient of linear regression. The anticipated and actual values ought to vary as little as feasible, hence the most effective suited range should have the lowest possible error.

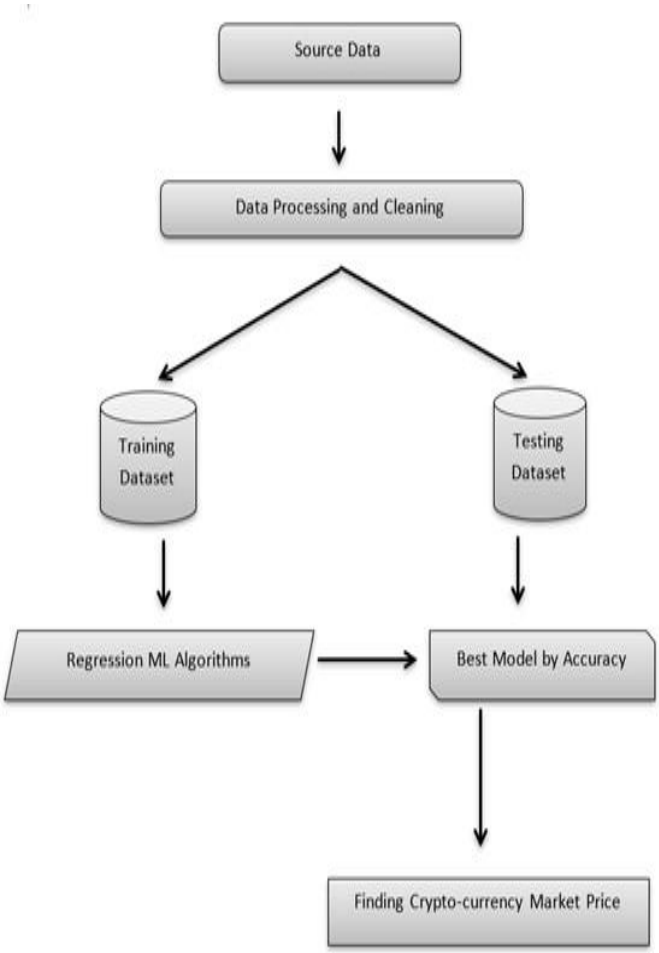

Fig 5. Data flow diagram

The R-squared (coefficient of determination) measure is used to evaluate the goodness of fit of a regression model, which in this context includes the SVR and LR models. It denotes the fraction of the variation in the dependent variable (ETH price) that can be explained by the model's independent variables (features). A higher R- squared value suggests that the model can account for a greater proportion of the fluctuation in the ETH price. It goes from 0 to 1, with 0 indicating that the model does not explain any of the variability and 1 indicating that the model does explain all of the variability. The R-squared value can give insight into the prediction potential of the

SVR and LR models when evaluating them. A higher R-squared value indicates that the model is more effective at capturing patterns and trends in ETH price data. We can assess whether model gives a better fit to the data and has higher predicted accuracy by comparing the R-squared values of the SVR and LR models. A higher R-

squared value suggests that the model is better able to Capture underlying patterns in ETH price data and make accurate projections.

The SVR model has an R-squared value of 0.85, which suggests that the model's characteristics explain 85% of the variability in the ETH price. The LR model, on the other hand, yields an R-squared value of 0.78, implying that the linear relationship between the characteristics and the ETH price can explain 78% of the variability.

Table I Performance Evaluation Results

MODEL	TEMPORAL LENGTH	PRECISION	ACCURACY	R-SQUARED
SVR	2 YEARS	98%	60.75%	85%
LR	2 YEARS	70%	50.05%	78%

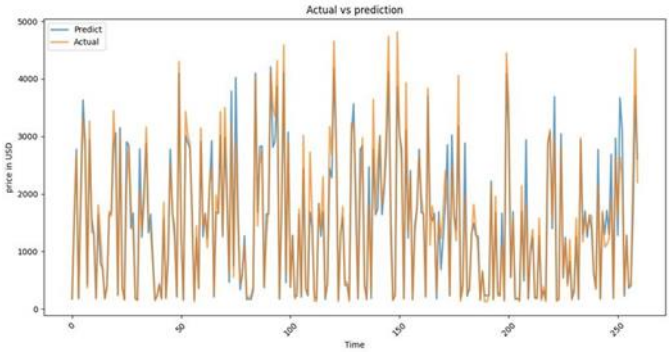

Fig 6. Actual vs prediction price of SVR model

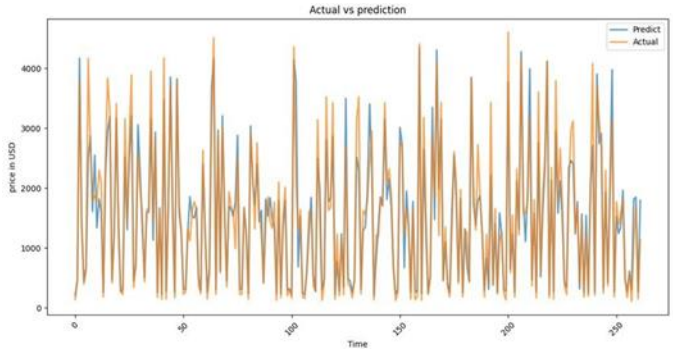

Fig 7. Actual vs prediction price of LR model

5 Conclusion

In summary, investigated the use of machine learning techniques, especially SVR and LR, for Ethereum price prediction in this research study. The SVR model was developed with a radial basis function kernel and historical price data acquired from Yahoo Finance using the yfinance package. In terms of prediction accuracy, the SVR model's performance was compared to that of the LR model. The results of our investigation indicated that the SVR model performed well in forecasting the future price of Ethereum. We were able to produce accurate estimates by using past pricing data and utilising the SVR model. The R-squared score was used to assess the SVR model's accuracy, offering quantitative insights into its predictive capabilities. Finally, this study lays the groundwork for applying machine learning techniques, notably SVR and LR, to forecast Ethereum prices. The findings show the algorithms' potential for predicting bitcoin values. However, additional investigation and refinement, as stated in future works, are required to increase the accuracy and robustness of predictions and to broaden the applicability of machine learning in the cryptocurrency arena. We can use SVR to create a Machine Learning web app that predicts Ether Future Price Prediction. The ethereum network is set to undergo significant modifications, which would raise the price.

References

1. Flach, P.: Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press. Cambridge, UK, (2012).
2. Sheikhi, S., Kheirabadi, M. T., Bazzazi, A. “An Effective Model for SMS spam Detection using Content-based features and Averaged Neural network”, International Journal of Engineering، Transactions B: Applications, Vol. 33, No. 2, (2020), 221-228. doi: 10.5829/IJE.2020.33.02B.06
3. Makridakis, S., Spiliotis, E., Assimakopoulos, V., “Statistical and Machine Learning forecasting methods”, Plos one, March 27, (2018), 1-26. doi: 10.1371/journal.pone.0194889
4. Bontempi, G., Taieb, S., Borgne, Y., “Machine Learning Strategies for Time Series Forecasting”, European Business Intelligence Summer School eBISS 2012, 62-77. SpringerVerlag.Berlin Heidelberg, (2013). doi: 10.1007/978-3-642- 36318-4_3
5. A. Begum, A. Tareq, M. Sultana, M. Sohel, T. Rahman, and A. Sarwar, „„Blockchain attacks analysis and a model to solve double spending attack,"" Int. J. Mach. Learn. Comput., vol. 10, no. 2, pp. 352–357, 2020.

6. R. Casado-Vara, J. Prieto, and J. M. Corchado, How blockchain could improve fraud detection in power distribution grid,‟ in Proc. 13th Int. Conf. Soft Comput. Models Ind. Environ. Appl. Cham, Switzerland: Springer, 2018, pp. 67–76.
7. Sumathy, G., Arokia Renjit, J. Distance-Based Method used to Localize the Eyeball Effectively for Cerebral Palsy Rehabilitation. J Med Syst 43, 262 (2019). https://doi.org/10.1007/s10916-019-1405-3