Predicting Bitcoin Prices Through the Integration of Sentiment Analysis and ARIMAX Models

1st Muhammad Abrar Triyadi
School of Computing
Telkom University
Bandung, Indonesia
mabrartt@student.telkomuniversity.ac.id

2nd Narita Aquarini
E ´cole Doctorale Science Economics
Universit ´e de Poitiers Intervenant Finance
La Rochelle, France
aquarinin@excelia-group.com

3rd Indwiarti
School of Computing
Telkom University
Bandung, Indonesia
indwiarti@telkomuniversity.ac.id

4th Putu Harry Gunawan
School of Computing
Telkom University
Bandung, Indonesia
phgunawan@telkomuniversity.ac.id

Abstract—**With the massive development of investment, cryptocurrency has become one of the world's most popular and widely traded digital investment instruments today. However, this investment is also affected by high price volatility, making it challenging for investors to make the right decision. Social media platforms like X play a significant role in facilitating investor discussions, where sentiment from the community often influences investment choices. This research analyzes the role of sentiment analysis in influencing Bitcoin price forecasting predictions by integrating the sentiment scores of X tweet data regarding "Bitcoin" and "BTC" with Bitcoin price using the ARIMAX method. Three scenarios were created for exogenous variable input features, including current sentiment score t, sentiment score input from n time steps, and average sentiment score input from n time steps. The results obtained from the ARIMAX(9, 1, 9) model with scenario 2 (sequence of past sentiment scores), which uses sentiment score input from n time steps, is more optimal than other scenarios as an exogenous variable with RMSE evaluation accuracy of 47799.91 and MAE of 40936.45. These results highlight the practical importance of sequential sentiment data in improving prediction accuracy and providing investors with actionable insights into market behaviour. This study recommends adopting sequential sentiment analysis as a key feature in forecasting models to enhance decision-making in cryptocurrency investments.**

Index Terms—**bitcoin, sentiment analysis, ARIMAX**

I. Introduction

Financial investment has become one of the most popular ways to achieve long-term and short-term financial goals. Investment is increasingly growing with the presence of one of the digital financial assets, namely cryptocurrency. Cryptocurrency is designed as a medium of exchange built on the blockchain system. It offers a transparent system, guaranteed security, decentralized and not bound by intermediaries such as banks or other traditional financial institutions [1]. Bitcoin is included as the most prominent cryptocurrency in terms of market capitalization. From 2014 to the present, bitcoin has experienced a significant price increase of 11900%; this makes Bitcoin one of the most widely used digital investment instruments among the public [2].

Although crypto, especially Bitcoin, has enormous potential in developing financial investment instruments, it is necessary to understand that crypto prices have a high level of volatility [3]. This volatility challenges investors as they carry out their investment activities. Many factors influence crypto price volatility, including regulation, technology adoption, global market dynamics, and public sentiment. This shows that the cryptocurrency market is investor-driven. Thus, the movement of cryptocurrencies is influenced by socially constructed opinions and the future expectations of current and future crypto investors.

Public sentiment shows how people view digital assets through social media, news, or online discussions. One of the social media that is widely used to express opinions and become a forum for discussion about Bitcoin is Twitter or what is now known as X. Most people trust the opinions of others in the online communities they follow to determine their crypto investment activities [4]. In recent years, X has become one of the platforms that influence others and shape the perception of the crypto market. In making the right decisions regarding investment activities, it is essential to consider a combination of public sentiment analysis and crypto price prediction forecasting to serve as a reference to see the market's direction and future price fluctuations. In this research, the sentiment method used is the VADER (*Valance Aware Dictionary and Sentiment Reasoner*), a Lexicon-based method with unsupervised techniques to determine the sentiment value of X tweet data.

As for Bitcoin price forecasting, the ARIMAX or *AutoRegressive Integrated Moving Average with eXogenous variables* method will be used. This method uses ARIMA techniques with exogenous variables. In this study, the exogenous variable is the sentiment score, which is calculated using three different scenarios. This scenario will be explained in section III. This method is used because it is able to model the relationship between dependent variables, such as Bitcoin price, and external variables, such as sentiment analysis results. A study conducted by [5] showed that the ARIMAX approach successfully forecasted the bee population with external variables of weather data. Other research [6] shows that the ARIMAX model can improve accuracy compared to the usual ARIMA model, with an increase in RMSE from ARIMA is 0.0373 then ARIMAX obtained an RMSE of 0.0301. Research [7] comparing the ARIMAX method with Vector Autoregressive (VAR) for strategic commodity price forecasting shows that the ARIMAX model can predict the consumer price of rice with a MAPE of 0.15%. This is 15.27% better than the VAR model. The effectiveness of the ARIMAX method in forecasting makes it a powerful tool for forecasting Bitcoin prices.

Besides the existing findings in previous studies, gaps can be developed in the literature regarding the use of sentiment data from social media, especially tweets, to improve the accuracy of bitcoin price prediction. Other studies have been limited in exploring the integration of various sentiment input scenarios, such as current sentiment score and average historical data as exogenous variables in ARIMAX models. This research seeks to bridge this gap by examining the impact of various sentiment analysis scenarios on the accuracy of Bitcoin price predictions.

II. Literature Review

A. Cryptocurrency

Cryptocurrency or crypto is a digital financial instrument that runs on a cryptographic security system. The system of ownership information is stored in a digital database integrated with cryptographic technology that also regulates the creation of new coins and verification of ownership transfers. Unlike conventional currencies, cryptocurrencies have no physical form and are not bound by intermediaries such as banks or other traditional financial institutions [2].

Today, cryptocurrencies have grown to become one of the growing and popular digital financial assets for people in the world. The rapid rise and fall of crypto prices shows that crypto has a high level of volatility [8]. With the growing potential, many countries and investors are interested in this digital currency. One of them in 2021, El Savador became the first country to announce that Bitcoin is legal as a medium of exchange [9]. In short, cryptocurrencies work through a combination of blockchain technology, consensus mechanisms, and market dynamics, creating a decentralized and secure method for conducting transactions [10].

B. Sentiment Analysis

Sentiment is one part of the natural language processing (NLP) field [11]. This field of study analyzes a person's attitude, opinion, emotion, evaluation, and assessment of various things such as an event, a product, an organization, a topic, an issue, etc. [12]. The highly volatile crypto prices are partly influenced by opinions and sentiments on social media and news, many investors rely on information spread on social media. One social media platform widely used to discuss and share opinions is X or Twitter. As cryptocurrencies continue to rise in price, many people speculate and share their opinions about them in tweets on X [13]. As more and more people delve into crypto analysis and publish their technical analysis and thoughts on the market in tweets on X, this can make the majority of other people follow the results of the analysis and use it as their reference material in crypto investment activities [14].

C. VADER (Valence Aware Dictionary and Sentiment Reasoner)

There are two ways to perform sentiment analysis: rule-based or machine learning. The rule-based approach uses a collection of words in a lexicon and specific rules to classify a tweet. VADER is a lexicon and rule-based sentiment analysis method that does not require prior training. In general, VADER calculates a vector-shaped sentiment score representing positive, negative, and neutral labels with normalized polarity in the range of 0 to 1 [15]. This method, developed by C.J. Hutto & E.E. Gilbert at the Georgia Institute of Technology, is a lexicon-based method tailor-made for understanding social media text [14].

This model incorporates a rule-based approach to improve the accuracy of sentiment analysis [16]. This research analyses sentiment analysis on tweets about crypto using sentiment calculation and the VADER model algorithm. VADER generates text sentiment scores into positive, negative, and neutral polarities. VADER performs an integration of dictionary-based and rule-based approaches, which enables better performance results compared to pure dictionary-based methods [17].

D. ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables)

The ARIMAX model basically uses the ARIMA model with the addition of exogenous variables or external variables. Based on research by Aggraeni et al. [7], the ARIMAX model effectively utilizes exogenous variables to improve prediction accuracy, making it a suitable choice for forecasting Bitcoin prices with sentiment scores as exogenous inputs. The ARIMA model is a time series model that is often used to perform future forecasting based on historical data. This model was introduced by George E. P. Box and Gwilym M. Jenkins in 1976. This model consists of 3 main components: *auto-regression*, *integration*, and *moving average* [18]. Here is the general form of ARIMA (p, d, q) (1) [19]:

$$\phi_p(B)(1-B)^d(y_t-\mu)=\theta_q(B)\epsilon_t \tag{1}$$

In the first stage, the data must be transformed in order to achieve stationarity, which is an important step to build an ARIMA(p, d, q) model. In 1979 Dickey and Fuller [20] introduced a stationary test called "testing for unit roots". They introduced the Augmented Dickey-Fuller (ADF) as the default method for evaluating whether or not a unit root exists. Furthermore, white-noise tests are used to reduce data noise, using techniques such as the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) [19]. Then the following are the three main components of ARIMA (p, d, q) [18]:

1) **AR**(p) is part of the *Autoregressive* component of ARIMA. This parameter determines the number of lags used in the model and is used to account for the historical usage of the time series. For AR(2), the equation is (2):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + e_t \tag{2}$$

Where Φ_1 and Φ_2 are the model parameters, and e_t is the error.

2) **I**(d) it represents the *Integrated* component of ARIMA. The value of the degree of differencing is determined by this parameter, which basically calculates the difference between the current value and the previous value. This parameter is used when the time series is not stationary and needs to be transformed to become stationary.

3) **MA**(q) this refers to an error as a combination of the previous error terms (3):

$$y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{3}$$

Below is the form of the linear equation resulting from combining the three parameters (4):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \quad (4)$$

4) **X**(eXogenous variables) The formula of ARIMAX [6] is shown in equation (5):

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{r} \beta_k X_{t-k} + \epsilon_t \quad (5)$$

This research is limited to only focusing on the use of ARIMAX, not using hybrid methods.

E. Evaluation

At this stage, we will evaluate the ARIMAX model that has been built. The model evaluation will use several evaluation methods, including [21]:

1) *Average Absolute Error* (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - f(x_i)| \quad (6)$$

The average deviation of each data value in a data set. The formula for MAE, as shown in Equation (6), calculates the average absolute difference between the predicted and actual data values.

2) *Mean Squared Error* (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 \quad (7)$$

Equation (7) obtained by squaring the difference between the actual value and the predicted value, then summing the squared differences and calculating the average of the squared differences.

3) *Root Mean Squared Error* (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2} \quad (8)$$

This method calculates the average difference between the predicted value and the actual value. Equation (8) calculated by dividing the number of observations (n) by the square root of the squared deviation.

III. Methodology

A. Dataset

In this research, the taken dataset uses the *scrapping* method with the python programming language and the yfinance or yahoo finance library. The following is an explanation of each of the data sets used in this study.

1) Cryptocurrency Dataset: The bitcoin price datasets were retrieved using the yfinance library, an open source library that allows users to retrieve price data such as stocks, cryptos, indices, currencies, and other financial data for the purpose of financial analysis or application development [22]. Bitcoin data is taken from November 2014 to November 2024 with the amount of data successfully retrieved as much as 3561 data. Each column in the data is "Date", "Adj Close", "Close", "High", "Low", "Open", "Volume". The data are saved in CSV file format.

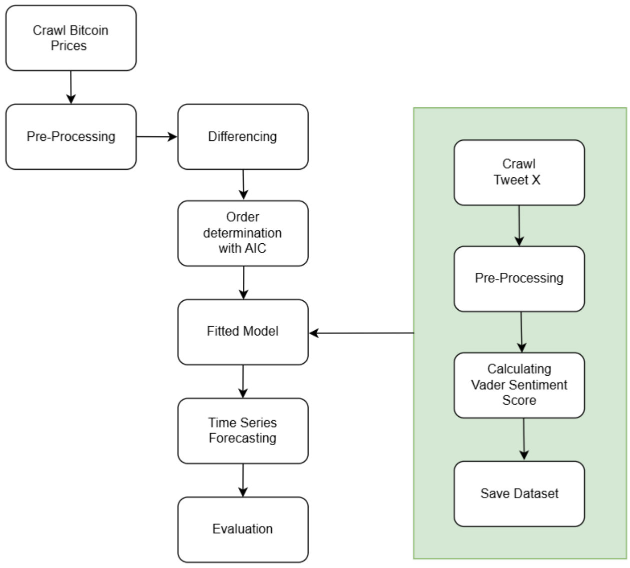

Fig. 1. Dataset flowchart.

2) Tweet Dataset: The tweet data is retrieved using the TweetHarvest library, an open source library commonly used to retrieve tweet data from X social media with specific keywords and the limit on the number of tweets that can be set [23]. Each data is retrieved with the keywords Bitcoin and BTC. The data is taken from the beginning of January 2020 until November 2024 with a total of 4191 tweets, the data is saved in CSV file format.

B. Data Preprocessing

Data preprocessing is an important part of ensuring that the data to be used meets the standards. This ensures maximum model results and accuracy.

1) Tweet Dataset PreProcessing: Before sentiment analysis, the tweet data is preprocessed by removing URLs, hashtags, mentions, removing symbols other than letters and numbers, and converting all text to lowercase using the NLTK libray.

2) Crypto Dataset Preprocessing: In the Bitcoin datasets, there is a "date" column that needs to be converted to date time, this is done to facilitate the process of building ARIMA models, which tend to require time series data. Then, to see the correlation of crypto price volatility with public sentiment, the crypto dataset is slightly manipulated by taking the closing value at the end of each month.

C. Merge dataset

After the cleaning process is carried out on each data, the next step is to merge the dataset that has been given a sentiment score with Bitcoin price data (see Fig. 1), merging this dataset based on the lag feature that has been determined with lag t and $t-n$, using the Bitcoin price from the previous 10 days as a predictor.

D. Scenarios for feature input

To see how sentiment analysis can affect the forecasting accuracy of Bitcoin price, three different scenarios have been designed regarding the integration of sentiment analysis for input features in the construction of ARIMAX models as exogenous variables. The following are the equations for each scenario:

1) **Using the current score (t) sentiment input** (1).

$$X_t \tag{9}$$

Where X_t is the input sentiment score at time t.

2) **Using input sentiment scores from n time steps** (10).

$$\begin{aligned} Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \\ \beta_n X_{t-n} + \theta_1 \epsilon_{t-1} + \phi_1 Y_{t-1} + \epsilon_t \end{aligned} \tag{10}$$

Where:

- X_{t-i} Sentiment score of the previous i time step.
- n Number of time steps considered.

3) **Using the average input sentiment score n time steps** (11).
The goal is to see how the average sentiment scores can affect the predicted closing price.

$$Y_t = \beta_0 + \beta_1 \left(\frac{1}{n} \sum_{j=1}^{n} X_{t-j} \right) + \theta_1 \epsilon_{t-1} + \phi_1 Y_{t-1} + \epsilon_t \tag{11}$$

Where:

- $\frac{1}{n}\sum_{j=1}^{n} X_{t-j}$ The average of the sentiment scores over the previous n time steps used as features for the model.

As can be seen in equation (11), which represents scenario 3 in determining the optimal features in integrating sentiment analysis as an exogenous variable in bitcoin price prediction.

E. VADER Sentiment Model

Compute sentiment scores using the VADER method or *Valence Aware Dictionary and Sentiment Reasoner*. This VADER produces a sentiment score in the form of a vector value representing positive, negative, and neutral with normalized polarity in the range 0 to 1 [15]. Furthermore, after obtaining the sentiment score, a combination is performed to see if public sentiment has an effect on the price volatility of each crypto. Sentiments are also validated and scored using vader.

F. ARIMAX (p, d, q) Model

Before building the model, it is first necessary to ensure that the data used are stationary with the ADF test [20] [19] If not, then differencing must be done first, differencing will produce a value for the parameter d, if differenced once, then $d = 1$. Then, the data is transformed to be normally distributed using the Box-Cox method; the goal is to improve normality and to equalize the variance in the data [24]. In addition, the parameters p and q are obtained by testing the AIC (*Akaike Information Criterion*) testing, which is a type of evaluation tool commonly used to assess how good the parameters are for ARIMA models. Introduced in 1973 by Hirotsugu Akaike [25], AIC evaluates the level of fit (*goodness of fit*) and also the simplicity or efficiency (*simplicity/parsimony*) in a model in terms of statistical numbers by comparing between models, and then models with lower AIC values are considered more optimal [26]. After obtaining the best order with AIC, input sentiment features with three different scenarios for the fitted model. Each scenario is used to build models and forecasts.

IV. Results and Discussion

A. Sentiment Analysis Model

In this section, the first process is to calculate the sentiment score using the VADER method, which analyzes text data with multiple positive, negative, and neutral labels, using composite score calculation rules. The sentiment score is calculated by using rules that calculate the intensity of the words in the text data, resulting in multiple values that reflect the emotional atmosphere contained therein. The result of the analysis is the composite score, which is the combined overall sentiment score of the text. In Fig. 2 is System Flowchart. Below is the composite score rule for determining sentiment labels:

1) If the composite score reaches a value of 0.05 or above, the text is categorized as having a positive sentiment, indicating that the content of the text tends to carry favourable or optimistic connotations.
2) If the composite score is -0.05 or less, the text is categorized as having a negative sentiment, indicating that the content of the text has a bad or pessimistic connotation.
3) Finally, if the composite score falls between these two values, the text is categorized as having a neutral sentiment, meaning that no strong sentiment can be detected.

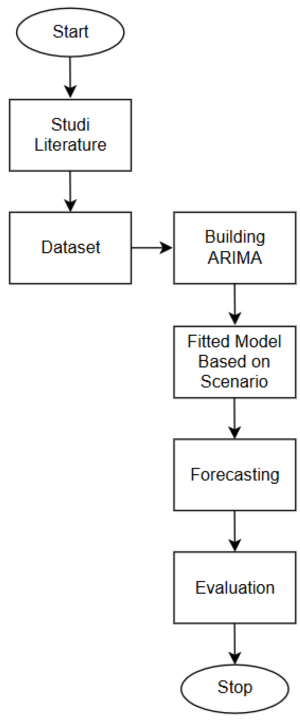

Fig. 2. System flowchart.

B. Crypto ARIMA Model

1) ***Augmented Dickey-Fuller (ADF) Test Result:*** The Augmented Dickey-Fuller (ADF) test was performed to test the stationarity of the closing prices, as shown in Table I. The ADF test results show that Bitcoin have a p-value greater than 0.05, which means that the data is not stationary and it needs to be differenced first.

2) ***Box-Cox Transformation:*** The Box-Cox transformation for BTC results in a lambda value of 0.477.

TABLE I
ADF TEST RESULTS FOR BITCOIN (BTC)

Paramater	BTC
ADF Statistic	-0.4501
p-value	0.9014
Optimal Number of Lags	13
Critical Values	
1%	-3.4341
5%	-2.8632
10%	-2.5676

*3) **ACF and PACF Plots**:* The autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to understand the correlation patterns in time series data and can also help identify appropriate statistical models for analyzing and predicting time-series data.

- The autocorrelation function (ACF)
  ACF measures the correlation between the value of a variable and the value of the variable at the previous time lag, helping to measure the correlation in general.
  ACF lags for BTC: 0, 8, 9, 24, 28, 31, 33, 36, 40, 44
- The partial autocorrelation function (PACF)
  The extent to which the previous value can have an immediate effect on the current value, after controlling for the effects of lags in between.
  PACF lags for BTC: 0, 8, 9, 24, 28, 33, 36, 40, 44, 46

From the information provided by the ACF and PACF plots, we can select possible values for the ARIMA(p,d,q) model. This identification involves observing the patterns in the ACF and PACF to determine the values of p (lag in the AR model) and q (lag in the MA model).

*4) **AIC Value Identification**:* Identifying the p,d, and q parameters using AIC, it was found that the best ARIMA model is (9, 1, 9), with the best AIC value being 10628.672078939779. This ARIMA model will be used to fit models in different scenarios.

*5) **Parameter Estimation**:* The ARIMA(9, 1, 9) model is used to estimate parameters with additional exogenous variables in different scenarios.

- The first scenario perform parameter estimation by adding scenario 1, which uses the current sentiment score t as an input feature.

TABLE II
PARAMETER ESTIMATION FOR ARIMAX SCENARIO 1

Evaluation Metrics	Value
AIC (Akaike Information Criterion)	10628.672
BIC (Bayesian Information Criterion)	10738.269
HQIC (Hannan-Quinn Information Criterion)	10669.161
Log-Likelihood	-5294.336

- Second scenario Parameter estimation is run using input features in the form of sentiment scores from n time steps.

TABLE III
PARAMETER ESTIMATION FOR ARIMAX SCENARIO 2

Evaluation Metrics	Value
AIC (Akaike Information Criterion)	30334.160
BIC (Bayesian Information Criterion)	30465.609
HQIC (Hannan-Quinn Information Criterion)	30382.728
Log-Likelihood	-15143.080

- Third scenario Parameter estimation is run using input features in the form of the average sentiment score of n time steps.

TABLE IV
PARAMETER ESTIMATION FOR ARIMAX SCENARIO 3

Evaluation Metrics	Value
AIC (Akaike Information Criterion)	30235.927
BIC (Bayesian Information Criterion)	30345.411
HQIC (Hannan-Quinn Information Criterion)	30276.385
Log-Likelihood	-15097.963

Based on the AIC value category, it can be seen that the first scenario shown by Table II is the best in sentiment score integration.

C. Fitted Model and Forecasting

Table V shows that scenario 2 produces the best performance among the other scenarios, with each evaluation metric showing the lowest value. This result shows that the scenario using the input sentiment scores from n time steps is more optimal than the other scenarios as exogenous variables. This performance suggests that multiple time-step scenarios allows the model to capture trends or sentiment patterns that may not be visible at a single time. This aligns with Formula 6, where n can be adjusted to better capture historical sentiment.

Another point obtained is the comparison of scenarios 1 and 3, where scenario 1 uses sentiment data at the current time or $X(t)$ and scenario 3 uses the average sentiment score n time steps. The evaluation results show that the approach in Scenario 3 can improve the accuracy of the model, with an RMSE value in Scenario 1 of 48651.58, which then decreases to 48345.92 in Scenario 3. This shows that the sentiment data that has been combined and calculated as the average value can improve the accuracy of using only the sentiment data at time t for future forecasting.

TABLE V
EVALUATION OF ARIMAX IN THREE SCENARIOS

Metrics	Scenario 1	Scenario 2	Scenario 3
MSE	2366975990.06	2284831765.03	2337327812.00
RMSE	48651.58	47799.91	48345.92
MAE	41964.65	40936.45	41560.99

Furthermore, Figure 3 is a visualization of the fitted results of the ARIMAX(9,1,9) model in scenario 2 to predict the Adjusted Close Price of Bitcoin. The model is a fairly good fit for the overall Bitcoin price trend. The orange line, which represents the model's prediction results, follows the pattern of the blue line, which is the actual price, with a fairly high level of accuracy. The model successfully captures historical price fluctuations and provides estimates that are close to the actual data.

However, while the prediction results appear to be in line with the price trend, there are some minor deviations, especially in periods where Bitcoin price volatility is very high. This suggests that while ARIMAX(9,1,9) is able to capture the main patterns in the data, the model still requires further tuning. These adjustments are important to improve prediction accuracy, especially in the face of the unpredictable nature of Bitcoin market volatility.

The visualization results of Bitcoin price prediction with the ARIMAX(9,1,9) model (see Fig. 4) show a flat pattern,

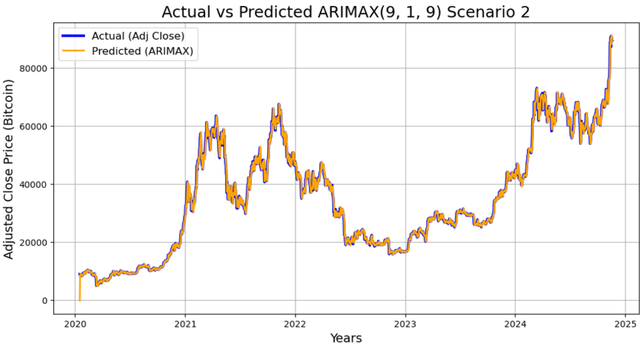

Fig. 3. Close price ARIMAX (9,1,9) based on scenario 2.

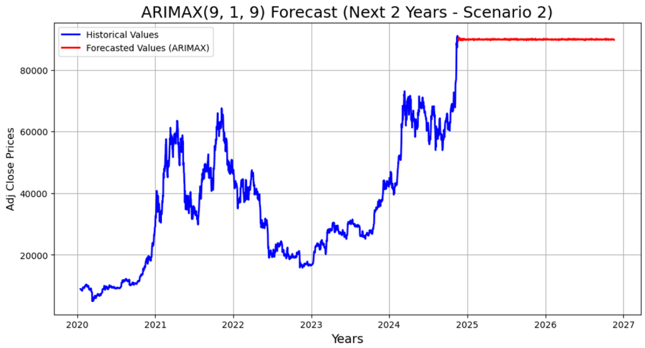

Fig. 4. Forecasting ARIMAX (9,1,9) based on scenario 2.

indicating that the model is less effective at capturing future price fluctuations. This may be due to the high volatility in the historical data, which is difficult for the model to replicate. Despite the lack of forecasting accuracy, the model evaluation shows that sentiment positively affects predictions. Scenario 2, which uses sentiment scores from the previous time step, has the lowest RMSE, demonstrating its effectiveness in capturing the relationship between sentiment and price. Scenario 3, which uses the average of the sentiment scores, also improves accuracy compared to scenario 1, which uses only the current scores.

V. Conclusion

This research analyzes the role of sentiment analysis in influencing Bitcoin price forecasting predictions by integrating the sentiment scores of X tweet data regarding "Bitcoin" and "BTC" with Bitcoin price using the ARIMAX method. The results show that making sentiment analysis an input feature of Bitcoin price forecasting significantly improves the prediction accuracy.

In the experiments conducted, various scenarios were tested to evaluate the effectiveness of using sentiment analysis. It was found that scenario two, which uses sentiment score input from n time steps, is more optimal than other scenarios as an exogenous variable. This scenario also showed better performance in the evaluation metrics than the other scenarios. This performance suggests that sequential sentiment data can be a powerful feature for forecasting Bitcoin price.

Although sentiment analysis acting as an exogenous variable can improve prediction accuracy, the ARIMAX model still shows high AIC values and other evaluation metrics. This suggests that although the integration of sentiment analysis makes a positive contribution, there is still a need for adjustments and the development of a more optimal model to improve prediction accuracy.

Further research suggests applying a hybrid approach by combining ARIMAX models or other time series models with machine learning algorithms. This approach is expected to produce a more robust and accurate model in forecasting Bitcoin prices by considering more complex aspects of the available data.

References

[1] B. Zohuri, H. T. Nguyen, and M. Moghadam, "What is the cryptocurrency? is it a threat to our national security, domestically and globally?" *International Journal of Theoretical & Computational Physics*, vol. 3, no. 1, pp. 1–14, Jan. 2022, submitted: 3 Jan 2022; Published: 5 Feb 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:248729619

[2] S. Corbet, B. Lucey, A. Urquhart, and L. Yarovaya, "Cryptocurrencies as a financial asset: A systematic analysis," *International Review of Financial Analysis*, vol. 62, pp. 182–199, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1057521918305271

[3] N. A. Huseynova and A. Yusifov, "Investing in cryptocurrency as an alternative way of financial investments," *AGORA International Journal of Economical Sciences*, vol. 17, no. 2, pp. 10–16, 2023. [Online]. Available: https://univagora.ro/jour/index.php/aijes/article/view/6436/2037

[4] R. Rahuf and N. Maknickienė, "Investigating an individual's opinion on social media about the cryptocurrency market," in *Proceedings of the 12th International Scientific Conference "Business and Management 2022"*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249212070

[5] T. Sahu, N. Chouksey, S. Kumar, S. K. Rana, S. Kanaujia, and R. Yadav, "Forecasting of honey bee population by arimax model using weather variables," *International Journal of Statistics and Applied Mathematics*, vol. 8, no. 5S, pp. 682–686, 2023.

[6] A. Pokhrel and R. Adhikari, "Leveraging exogenous insights: A comparative forecast of paddy production in nepal using arima and arimax models," *Economic Review of Nepal*, vol. 6, no. 1, p. 52–69, Dec. 2023. [Online]. Available: https://www.nepjol.info/index.php/ern/article/view/67970

[7] W. Anggraeni, K. B. Andri, Sumaryanto, and F. Mahananto, "The performance of arimax model and vector autoregressive (var) model in forecasting strategic commodity price in indonesia," *Procedia Computer Science*, vol. 124, pp. 189–196, 2017, 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050917329149

[8] S. Alghamdi, S. Alqethami, T. Alsubait, and H. Alhakami, "Cryptocurrency price prediction using forecasting and sentiment analysis," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2022.01310105

[9] A. Gaikwad and S. Mavale, "The impact of cryptocurrency adoption as a legal tender in el salvador," *International Journal of Engineering and Management Research*, vol. 11, 12 2021.

[10] W. K. Härdle, C. R. Harvey, and R. C. G. Reule, "Understanding cryptocurrencies," *SSRN Electronic Journal*, 2019. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3360304

[11] K. L. Tan, C. P. Lee, and K. M. Lim, "A survey of sentiment analysis: Approaches, datasets, and future research," *Applied Sciences*, vol. 13, no. 7, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/7/4550

[12] K. Du, F. Xing, R. Mao, and E. Cambria, "Financial sentiment analysis: Techniques and applications," *ACM Comput. Surv.*, vol. 56, no. 9, Apr. 2024. [Online]. Available: https://doi.org/10.1145/3649451

[13] S. Arslan, "Bitcoin price prediction using sentiment analysis and empirical mode decomposition," *Computational Economics*, vol. 1, no. 1, pp. 1–15, May 2024, open Access. [Online]. Available: https://link.springer.com/article/10.1007/s00712-024-00723-4

[14] A. Raheman, A. Kolonin, I. Fridkins, I. Ansari, and M. Vishwas, "Social media sentiment analysis for cryptocurrency market prediction," 2022. [Online]. Available: https://arxiv.org/abs/2204.10185

[15] H. G. B and S. N. B, "Cryptocurrency price prediction using twitter sentiment analysis," in *Natural Language Processing, Information Retrieval and AI*. Academy and Industry Research Collaboration Center (AIRCC), Feb. 2023, p. 13–22. [Online]. Available: http://dx.doi.org/10.5121/csit.2023.130302

[16] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[17] J. Miao, "Research on tweet sentiment analysis based on vader in the field of cryptocurrency," in *Proceedings of the 3rd International Conference on Digital Economy and Computer Application (DECA 2023)*. Atlantis Press, 2023, pp. 256–264. [Online]. Available: https://doi.org/10.2991/978-94-6463-304-7_28

[18] R. Mittal, R. Gehi, and M. Bhatia, "Forecasting the price of cryptocurrencies and validating using arima," in *Proceedings of the 4th International Conference on Computers & Management (ICCM)*. New Delhi, India: Netaji Subhas Institute of Technology, 2018.

[19] J. T. Tsoku, D. Metsileng, and T. Botlhoko, "A hybrid of box-jenkins arima model and neural networks for forecasting south african crude oil prices," *International Journal of Financial Studies*, vol. 12, no. 4, 2024. [Online]. Available: https://www.mdpi.com/2227-7072/12/4/118

[20] D. Dickey and W. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *JASA. Journal of the American Statistical Association*, vol. 74, 06 1979.

[21] J. Sega, Shivam, Y. Gautam, and K. D. Yadav, "Improving the prediction and fine-grained approach for predicting and forecasting cryptocurrency price using arima," in *2024 4th International Conference on Intelligent Technologies (CONIT)*, 2024, pp. 1–7.

[22] S. V. Chanda and A. A, "Web scraping in finance using python," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 5, pp. 2249–8958, June 2020. [Online]. Available: https://www.ijeat.org/

[23] M. Ghufron, M. Arsyada, M. Lukman, Y. Putra, and N. Rakhmawati, "Analisis sentimen pengguna twitter terhadap pemilu 2024 berbasis model xlm-t," *J-INTECH*, vol. 11, pp. 307–315, 12 2023.

[24] G. E. P. Box and D. R. Cox, "An analysis of transformations," in *Research Methods Meeting of the Society*. University of Wisconsin and Birkbeck College, University of London, 1964, read at a Research Methods Meeting of the Society, April 8th, 1964, Professor D. V. Lindley in the Chair.

[25] J. E. Cavanaugh and A. A. Neath, *Akaike's Information Criterion: Background, Derivation, Properties, and Refinements*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 26–29. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_111

[26] M. D. Angelo, I. Fadhiilrahman, and Y. Purnama, "Comparative analysis of arima and prophet algorithms in bitcoin price forecasting," *Procedia Computer Science*, vol. 227, pp. 490–499, 2023, 8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923017179