IJCDS 1571026011

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

Enhancing Bitcoin Forecast Accuracy by Integrating AI, Sentiment Analysis, and Financial Models

Mohamad El Abaji[1] **and Ramzi A. Haraty**[2]

[1]*Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon*

Received Mon. 20, Revised Mon. 20, Accepted Mon. 20, Published Mon. 20

Abstract: This study explores the application of advanced AI models—Long Short-Term Memory (LSTM), Prophet, and SARIMAX—in predicting Bitcoin prices. It assesses the impact of incorporating sentiment analysis from sources like Twitter and Yahoo, processed through Large Language Models. The research aims to understand how sentiment analysis, reflecting investor sentiments and market perceptions, can enhance the accuracy of these forecasting models. The paper investigates the potential synergies and challenges in improving predictive performance by integrating qualitative sentiment data with quantitative financial models. The analysis compares the models' accuracy with and without sentiment inputs, utilizing historical Bitcoin price data and sentiment indicators. This study's motivation is the growing recognition of investor sentiment's impact on market fluctuations, particularly in the highly speculative and sentiment-driven cryptocurrency markets. While robust in handling quantitative data, many studies claim that traditional financial models often fail to incorporate market sentiments. This paper also contributes to financial forecasting literature by offering insights into the benefits and complexities of combining traditional econometric models with sentiment analysis, providing a unique understanding of market dynamics influenced by investor behavior. The findings suggest that sentiment analysis can significantly refine forecasting accuracy, underscoring the importance of incorporating human sentiment and market perceptions in predictive models.

Keywords: Bitcoin forecasting, LSTM, Prophet model, SARIMAX, sentiment analysis, Large Language Models, financial models, market dynamics, investor behavior, and predictive accuracy.

1. Introduction

Cryptocurrency, a digital or virtual currency employing cryptography for security, has emerged as a revolutionary financial instrument in the contemporary economic field [1]. Unlike traditional fiat currencies, cryptocurrencies operate on decentralized platforms, primarily blockchain technology, ensuring transaction transparency, security, and immutability [2]. This innovation in financial technology has challenged conventional banking and introduced new paradigms for investment, trading, and wealth management [3]. The significance of cryptocurrencies extends beyond mere financial transactions [4]. Cryptocurrencies embody the potential for creating a more inclusive global financial system, reducing transaction costs, and enhancing the speed and efficiency of cross-border transactions [5]. Furthermore, the speculative nature of cryptocurrencies has attracted immense interest from investors, leading to unprecedented market volatility [6]. While presenting substantial risks, this volatility also offers considerable opportunities for profit, making the accurate prediction of cryptocurrency prices highly valuable [7]. The ability to forecast cryptocurrency prices accurately can lead to more informed investment decisions, optimized trading strategies, and enhanced understanding of digital asset dynamics, contributing significantly to the broader field of financial technology and economic research.

In financial markets, the volatility and unpredictability inherent to cryptocurrency present opportunities and challenges for investors and analysts alike [8]. Advanced computational techniques and artificial intelligence (AI) have opened new avenues for predicting market trends and asset prices more accurately [9]. Among these, Long Short-Term Memory (LSTM) networks [10], [11], [12], the Prophet forecasting tool, and Seasonal Autoregressive Integrated Moving Average with eXogenous inputs (SARIMAX) models have emerged as prominent methods for analyzing time-series data, capable of capturing complex patterns and seasonalities. These models, however, traditionally rely on quantitative historical data, often overlooking the qualitative aspects that significantly influence market dynamics, such as investor sentiment and market perceptions, which are increasingly manifested through digital platforms like Twitter and Yahoo.

This study's motivation is the growing recognition of investor sentiment's impact on market fluctuations, particularly in the highly speculative and sentiment-driven cryptocurrency markets. While robust in handling quantitative data, many studies claim that traditional financial models often fail to incorporate market sentiments. This gap highlights the need for an investigation comparing the effect of qualitative sentiment analysis on quantitative financial

E-mail address: author 1, author 2, author 3

data to enhance predictive accuracy.

The problem lies in the limited capacity of existing models to fully capture the multifaceted nature of market dynamics, particularly in the context of Bitcoin's price volatility. While LSTM, Prophet, and SARIMAX models provide a solid foundation for time-series forecasting, incorporating sentiment analysis may significantly enhance their efficacy. The potential solution lies in leveraging Large Language Models (LLMs) to process vast amounts of textual data from platforms like Twitter and Yahoo, translating qualitative sentiment into quantifiable metrics that can augment traditional forecasting models.

Existing literature focuses on quantitative financial forecasting or sentiment analysis in isolation, with limited exploration into their integration and claiming the results can be remarkable. Some Studies in this domain often treat sentiment analysis as a peripheral component, and others treat it as a core predictive factor. There is a missing link in understanding how sentiment analysis can alter forecasting outcomes. Especially, when processed through advanced LLMs and integrated with financial models. Moreover, the methodologies for quantifying and integrating sentiment remain underexplored.

In order to resolve this issue, and to explore a more holistic approach to Bitcoin price forecasting, This paper explores the effect of integrated sentiment, processed through LLaMa-2, with LSTM, Prophet, and SARIMAX. This integration aims to explore the predictive power of both quantitative market data and qualitative sentiment indicators. This may offer a comprehensive view of market dynamics. This approach seeks to address the limitations of existing models, enhancing accuracy and providing actionable insights for stake holders.

This research aims to enhance the accuracy and reliability of Bitcoin price forecasts by integrating sentiment analysis with traditional financial forecasting models. To achieve this, we have delineated three specific research objectives:

- To evaluate the individual and combined predictive performance of LSTM, Prophet, and SARIMAX models on Bitcoin price data;
- To develop a methodology for quantifying investor sentiment from digital platforms using LLMs and integrating it into the forecasting models;
- To assess the impact of sentiment analysis on the predictive accuracy of these models.

Correspondingly, the research questions are:

1) How do LSTM, Prophet, and SARIMAX models perform individually and in combination in forecasting Bitcoin prices?
2) What is the most effective methodology for quantifying and integrating sentiment analysis into these models?
3) How does the inclusion of sentiment analysis influence the predictive accuracy of these models in the context of Bitcoin price forecasting?

2. Literature Review

The comprehensive examination of cryptocurrency price prediction presents a vast research domain. Initially conceptualized as a decentralized digital currency, Bitcoin emerged as a pioneering cryptocurrency, challenging traditional financial systems and introducing a new digital asset class [13]. Its inherent volatility, driven by speculative trading and market sentiment, underscores the complexity of forecasting its price movements, necessitating sophisticated analytical models [14].

Traditional methods for predicting cryptocurrency prices have been rooted in the principles of financial econometrics and time-series analysis [15]. These methods draw on techniques applied to more conventional asset classes like stocks, bonds, and commodities [16]. These methodologies rely on historical price data, trading volumes, and market capitalization [8]. Such techniques employ statistical models such as Autoregressive Integrated Moving Average (ARIMA) and its variants to capture linear relationships in the data [17]. These models assume the past price movements can provide insights into future trends based on market efficiency and the random walk hypothesis [18]. However, the volatile cryptocurrency market, with influence of non-economic factors, poses significant challenges [19]. Their reliance on historical data and linear assumptions limits their ability to capture nonlinear price movements [20]. This led researchers to explore more flexible methodologies capable of capturing the complex dynamics of the cryptocurrency [21]. Integrating Machine Learning (ML) methods into cryptocurrency prediction represents a significant departure from traditional approaches [22]. Assuming ML can adapt to cryptocurrency markets' volatile and unpredictable nature [23]. Unlike the complex deep learning architectures, ML algorithms offer a suite of techniques. Which are capable of uncovering the relationships in data that are not apparent [24]. Among these, Decision Trees (DT) [25], Random Forests (RF) and Support Vector Machines (SVM) [26] stand out for their application in financial forecasting.

Venkat et al. [25] suggested that DT provide a straightforward method for capturing nonlinear relationships between market indicators. They segment the data subsets to modeling of complex decision boundaries. RF enhanced the accuracy of the model by reducing the overfitting that can plague decision trees. According to Sivaram et al., [26], SVM introduces another dimension to the predictive modeling of cryptocurrency prices. By mapping input features into high-dimensional space, SVMs find the hyperplane that best separates data. As in this case, market move-

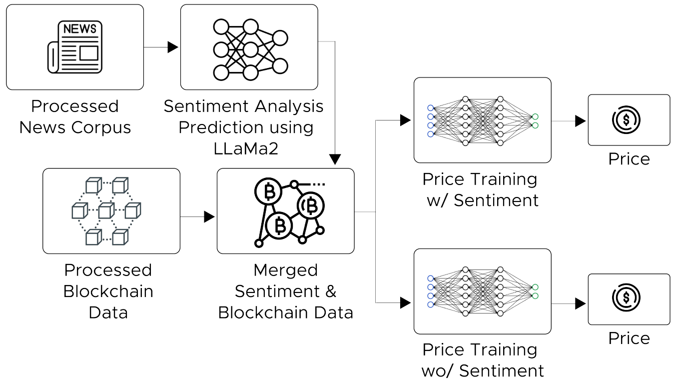

Figure 1. The Overview of Employed Methodology

ments—maximizing the margin between them. This makes SVMs particularly adept at actionable insight prediction.

Turner [27] described that ML techniques provide valuable tools for dissecting the patterns of cryptocurrency. Their ability to learn from historical data without needing pre-defined equations allows for a more flexible modeling approach. Which can capture the essence of market dynamics in a way that traditional models cannot [28]. Furthermore, LSTM have gained prominence in financial forecasting for their ability of long-term dependencies in sequences [29]. This is particularly pertinent to the cryptocurrency market, where past price movements can provide insights into future. Studies have demonstrated LSTM potential in capturing the temporal dynamics and nonlinear patterns in cryptocurrency [10], [11], [12].

The Prophet model, developed by Facebook, introduces a different approach to time-series forecasting [30]. It emphasize on the flexibility and ease of use. Its application to cryptocurrency prediction has been marked by the model's robust handling of outliers and missing data [31], [32]. Researchers have noted Prophet's utility in generating accurate forecasts for Bitcoin, showcasing its adaptability to cryptocurrency [33]. Similarly, SARIMAX extends the ARIMA model by incorporating seasonality and external variables [34]. Which makes it adept at handling complex financial time-series data influenced by many factors [35]. In the context of Bitcoin, SARIMAX models have been explored to account for external influences such as market sentiment, regulatory announcements. Moreover, macroeconomic indicators also explored to provide a more comprehensive prediction of price movements [36], [37].

The advent of sentiment analysis has introduced a new dimension to financial forecasting [38]. This comprises of recognizing the impact of investor sentiment on market dynamics [39]. By analyzing textual data from news, social media, sentiment analysis seeks to understand the emotional undertones of investor behavior [9]. This approach has been increasingly integrated into cryptocurrency price prediction models. Such methods aim to capture the sentiment-driven volatility characteristic of these markets [40]. LLMs have revolutionized sentiment analysis by providing unique interpretations of large volumes of textual data [41]. These models' ability to understand context, sarcasm, and subtle sentiment cues enhances the accuracy of sentiment extraction [42]. When applied to cryptocurrency news, LLM-processed sentiment data can significantly enrich predictive models. It can offer insights into the collective market psyche and potential price movements [43].

Integrating quantitative forecasting models with qualitative sentiment analysis poses a novel research direction [44]. It promises to enhance predictive accuracy by offering a holistic view of market dynamics. Combining LSTM, Prophet, and SARIMAX models with sentiment indicators derived from LLMs presents a comprehensive approach to cryptocurrency price prediction. It maybe helpful addressing market behavior's mathematical and psychological aspects [45]. However, the fusion of sentiment analysis with traditional financial models presents methodological challenges, particularly in quantifying and integrating sentiment data. The reliability of sentiment analysis hinges on the quality of the data sources and the sophistication of the processing algorithms [46]. Moreover, the subjective nature of sentiment and its potential for rapid shifts poses additional complexities for model integration [47].

Moreover, there exist conflicts about whether the inclusion of sentiment may increase the model's ability to enhance the prediction as stated by [48], [49], [50], [51], [52]. Some researchers have opposed the idea and claimed the reduction of results after sentiment inclusion [53], [54], [55]. One side of researchers, suggest to include the sentiment due to its evident representation of investor's interest. While other group oppose the idea due to the subjective and unpredictable nature of news as it maybe biased.

Despite these debates, the potential of sentiment-augmented financial models for cryptocurrency price prediction remains largely untapped. With existing literature focusing predominantly on individual methodologies. Integrating sentiment analysis with models like LSTM, Prophet, and SARIMAX offers a promising avenue for research, potentially bridging the gap between quantitative data analysis and qualitative sentiment interpretation in financial markets. The evolving field of cryptocurrency price prediction shows the need for innovative, integrated approaches that leverage data-driven models and sentiment analysis. This research domain presents a fertile ground for exploring the synergies between quantitative financial forecasting and qualitative sentiment assessment, aiming to enhance the accuracy and reliability of cryptocurrency price predictions.

3. Methodology

Our methodology encompassed a comparative analysis of three predictive models: LSTM, Prophet, and SARIMAX. Initially, these models were trained exclusively on historical financial data about Bitcoin, encompassing metrics such as price and trading volume. Subsequently, a second training phase incorporated another variable: the polarity of Bitcoin-related news articles, serving as a proxy for market sentiment. This polarity was derived from sentiment analysis conducted on the content of news articles, classified as either positive or negative. The objective was to assess the impact of integrating sentiment data on the predictive performance of each model. Performance was quantified utilizing MAPE and accuracy metrics to ascertain the impact of incorporating sentiment analysis on the models' forecasting precision. Each model was applied to project Bitcoin prices over 30 days in an autoregressive manner, allowing for a dynamic evaluation of each model's capability to integrate and leverage market data and sentiment indicators for enhanced predictive insights into Bitcoin's future price trajectory. The overview of the proposed methodology is provided in 1.

A. Dataset

The dataset for this study was meticulously compiled from two primary sources, focusing on Bitcoin news data and blockchain information, to facilitate a comprehensive analysis of Bitcoin price movements and their correlation with market sentiments.

1) Bitcoin Blockchain Data

Parallel to collecting news data, Bitcoin blockchain information was sourced from Blockchain.com, aiming to augment the analysis with empirical market data. The Blockchain.com API offered a comprehensive suite of market data across several categories, including Currency Statistics, Market Signals, Block Details, Mining Information, and Network Activity. We used only the Volum and market price features of this data. The market price over the time is shown in the figure 2

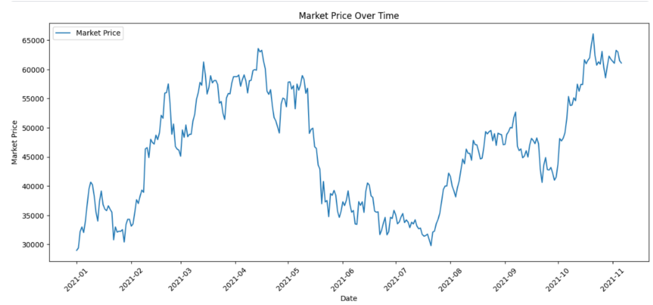

Figure 2. The market price over time

2) Bitcoin News Data

The collection of Bitcoin news data was initiated from crypto news-api.com, covering the span from January 1, 2021, to November 5, 2021. Given that the data from crypto news-api.com predominantly consisted of article summaries, a more detailed extraction was conducted utilizing the Python newspaper library to scrape the full content of articles. This initial effort yielded a corpus encompassing 23,912 records.

A selection process was undertaken to ensure the information's relevance and credibility, focusing on retaining articles from esteemed news sources known for their financial reporting. This process resulted in a refined corpus containing 16,162 articles from sources including Forbes, CNBC Television, Bitcoin Magazine, Yahoo Finance, Bloomberg Markets and Finance, Reuters, Fox Business, Bloomberg Technology, Coindesk, and CNBC.

3) Sentiment Analysis and Data Integration

A pivotal modification in processing the dataset was applying a pre-trained Large Language Model (LLM) for sentiment analysis tailored to the cryptocurrency domain. The sentiment models were initially trained on a subset of the news corpus, covering the period from January 1, 2021, to June 6, 2021. After the training phase, these models were employed to predict the sentiment of news articles from June 7, 2021, to November 5, 2021. The sentiment scores, aggregated daily, were then merged with the Bitcoin blockchain data for a holistic view. Figure 3 shows the combined dataset values over time.

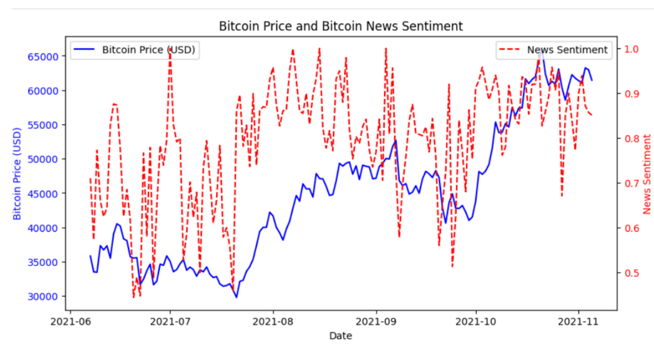

Figure 3. The Combined dataset values over time

B. Long Short-Term Memory Networks

LSTM networks, a class of recurrent neural networks (RNNs), are particularly adept at capturing long-term dependencies in time-series data due to their unique architecture, which includes memory cells and gates that regulate the flow of information. For Bitcoin price forecasting, the LSTM model can be customized to incorporate time lags that reflect the cyclical nature of market sentiments and trading patterns. The equation for the LSTM unit, tailored to this context, could be represented as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{1}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{2}$$

Where C_t is the cell state at time t, h_t is the output vector, f_t, i_t and o_t are the forget, input, and output gates, respectively, and $\tilde{C}_t$ is the candidate cell state. This equation encapsulates the LSTM's ability to selectively remember or forget information, which is critical for modeling the volatile nature of the cryptocurrency market. In constructing the LSTM model for Bitcoin price prediction, a methodical approach was employed to transform the raw data into a structured format amenable to time-series analysis. This entailed the creation of sequences from the dataset, each comprising a fixed number of consecutive data points, referred to as the window size. The mathematical representation of this sequence generation process is given by:

$$X_i = [x(i), x(i+1), \ldots, x(i+n-1)] \tag{3}$$

$$Y_i = x(i+n) \tag{4}$$

The LSTM architecture was designed to analyze these sequences with two main layers, each containing 128 units. Following each LSTM layer, a dropout rate of 0.2 was implemented to mitigate overfitting by randomly omitting a fraction of the units during training. The architecture summary of the LSTM model is provided in figure 4. The model was compiled with a learning rate of 0.0001,

Layer (type)	Output Shape	Param #
lstm_4 (LSTM)	(None, 3, 128)	73,216
dropout_4 (Dropout)	(None, 3, 128)	0
lstm_5 (LSTM)	(None, 128)	131,584
dropout_5 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 1)	129

Total params: 204,929 (800.50 KB)
Trainable params: 204,929 (800.50 KB)

Figure 4. LSTM Model Architecture details

chosen to allow for gradual and precise weight adjustments, which is crucial in the volatile cryptocurrency market. The optimization was performed using the mean squared error loss function, which is suitable for the regression nature of price forecasting. The final output of the model is produced by a dense layer with a single unit, providing the predicted Bitcoin price. This setup balances complexity with computational efficiency, aiming for accurate and generalizable predictions.

C. SARIMAX

The SARIMAX model extends the ARIMA model by incorporating seasonality and external variables, making it well-suited for forecasting Bitcoin prices where external factors, such as regulatory announcements or technological advancements, play a significant role. The equation can represent the SARIMAX model for this application:

$$\begin{aligned} y_t =\, & \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ & + \Phi_1 Y_{t-S} + \cdots + \Phi_P Y_{t-PS} + \Theta_1 E_{t-S} + \cdots + \\ & \Theta_Q E_{t-QS} + X_t \beta + \epsilon_t \end{aligned} \tag{5}$$

In this equation, $\phi, \Phi, \theta, \Theta$ represent the autoregressive and moving average parameters for both non-seasonal and seasonal components; y_t and Y_t denote the non-seasonal and seasonal observations; ϵ_t and E_t are the error terms; $X_t\beta$ represents the exogenous regressors, capturing the impact of external factors on Bitcoin prices; and S is the seasonality period.

The model's selection process was automated, utilizing a statistical algorithm to discern the optimal parameter set that best fits the data's temporal structure. This algorithm adjusted the parameters iteratively, considering both seasonal and non-seasonal components while incorporating external regressors.

The model's identification commenced with the training data, including exogenous variables, acknowledging the presence of seasonal fluctuations with a defined periodicity. The algorithm pursued a systematic and efficient search through the model space, guided by information criteria that balanced model fit with parsimony, thus avoiding overfitting. Throughout this procedure, the algorithm maintained robustness by disregarding transient computational errors and circumventing unnecessary warnings that might hinder the iterative process. The section outcome is provided in the figure 5.

The algorithm converged on a SARIMAX(0,1,0) configuration with a seasonal periodicity of 30, indicative of monthly effects, and determined that including an intercept was not statistically justified. This model specification reflects a reliance on differencing to stabilize the series. It suggests that the dynamics of Bitcoin prices were best explained without the need for additional autoregressive or moving average terms in the context of the data provided. Figure 6 plots standardized residuals over time. These are the differences between observed and predicted values, scaled to have zero mean and unit variance. The plot should ideally exhibit randomness, indicating that the model has captured the underlying data patterns well. If there were discernible patterns or trends in the residuals, it would suggest the model is not adequately accounting for all the information in the data. The points fluctuating between approximately -3 and +3 with no clear pattern suggest

```
Best model:  ARIMA(0,1,0)(0,0,0)[30] intercept
Total fit time: 12.096 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  248
Model:               SARIMAX(0, 1, 0)   Log Likelihood               -2220.151
Date:                Thu, 28 Mar 2024   AIC                           4446.302
Time:                        15:18:13   BIC                           4456.830
Sample:                    01-01-2021   HQIC                          4450.541
                         - 09-05-2021
Covariance Type:                  opg
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
intercept       83.8949   1.15e-16   7.31e+17      0.000      83.895      83.895
trade-volume -5.841e-07   1.85e-07     -3.163      0.002   -9.46e-07   -2.22e-07
sigma2        3.757e+06   1.88e-20      2e+26      0.000    3.76e+06    3.76e+06
===================================================================================
Ljung-Box (L1) (Q):                   2.41   Jarque-Bera (JB):                28.74
Prob(Q):                              0.12   Prob(JB):                         0.00
Heteroskedasticity (H):               0.39   Skew:                             0.18
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.63
===================================================================================
```

Figure 5. Automated Model Selection Process for SARIMAX

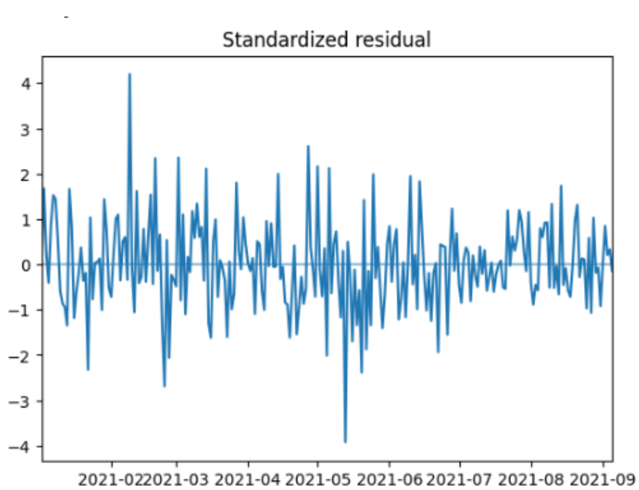

Figure 6. Automated Model Selection: Standardized Residuals over time

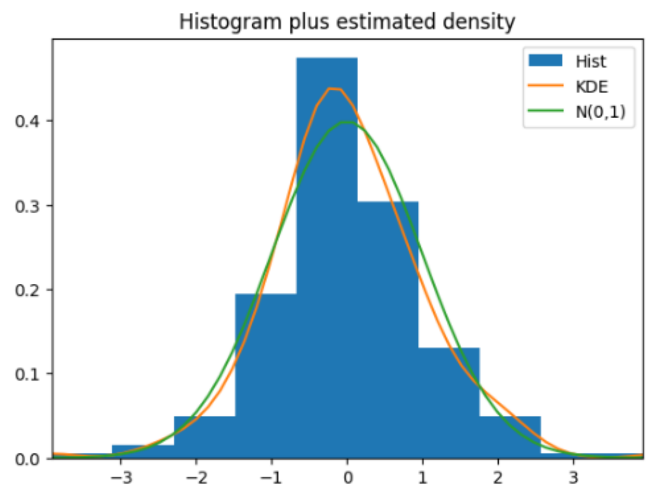

Figure 7. Automated Model Selection Process: histogram of the residuals (in blue) along with a KDE (in green)

that the model residuals are random, indicating a decent model fit. Figure 7 shows a histogram of the residuals along with a KDE, which is a smooth, continuous curve estimating the distribution of the residuals. The orange line represents a normal distribution (N(0,1)). This comparison is critical as it allows us to assess the normality of the residuals. An assumption in many regression models is that residuals are normally distributed. A close fit between the KDE and the normal distribution curve suggests that the residuals approximate a normal distribution, which is a desirable property and supports some of the model's underlying assumptions.

D. Prophet

Prophet is a forecasting tool designed for handling the seasonalities and trends in time-series data, with robustness to missing data and shifts in trend. The Prophet model can be customized for Bitcoin price prediction to account for weekly and yearly seasonality, reflecting the trading volume fluctuations and macroeconomic factors affecting the cryptocurrency market. The core equation for the Prophet model, modified for this application, might be:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{6}$$

Here, $g(t)$ represents the trend component, capturing long-term movements in Bitcoin prices; $s(t)$ accounts for seasonal effects; $h(t)$ represents the holiday effects, which could be significant during periods of high trading activity; and ϵ_t is the error term. This equation highlights Prophet's flexibility in adapting to the unique temporal structures of the Bitcoin market.

E. Auto Regressive Predictions

The study employed an autoregressive framework to forecast Bitcoin prices over 30 days utilizing LSTM, Prophet, and SARIMAX models. This method was initiated with the last observed data sequence, to sequentially predict daily value, and each prediction was incorporated into the sequence for subsequent forecasts. Random perturbations for exogenous features and an inverse scaling transformation ensured the projections were usable. The process was reiterated across all models to evaluate their predictive capabilities.

Algorithm 1 Autoregressive Prediction for Bitcoin Prices

Require: $model$, $\mathbf{s}_0$, N, $scaler_{feature}$, $scaler_{target}$
Ensure: $predicted\_prices$

p $\leftarrow$ [ ] ▷ Initialize predictions list
$\mathbf{s} \leftarrow \mathbf{s}_0$ ▷ Set initial sequence
for $i \leftarrow 1$ **to** N **do**
  $\hat{y} \leftarrow model.predict(\mathbf{s})$
  $\mathbf{s} \leftarrow updateSequence(\mathbf{s}, \hat{y})$ ▷ Incorporate predicted value into sequence
  $\mathbf{p} \leftarrow append(\mathbf{p}, \hat{y})$ ▷ Store the prediction
end for
$\mathbf{p}_{inv} \leftarrow scaler^{-1}_{target}(\mathbf{p})$ ▷ Inverse transform predictions
return $\mathbf{p}_{inv}$
function $updateSequence(\mathbf{s}, \hat{y})$
  $\mathbf{s}_{new} \leftarrow \text{Shift}(\mathbf{s}, -1)$ ▷ Shift sequence left
  $\mathbf{s}_{new}[end] \leftarrow \hat{y}$ ▷ Insert new prediction
  $\mathbf{s}_{new}[end-1] \leftarrow randomValue()$ ▷ Random value for exogenous feature
  return $\mathbf{s}_{new}$
end function

F. Llama-2

LLaMA-2 is leveraged to understand the sentiment in each article and classify it as either positive or negative. This involves feeding the article into LLaMA-2, which then processes the information and outputs a sentiment score. The model's architecture allows it to grasp the contextual meaning of words and phrases. Considering the uniqueness of financial terminology and market-related discussions it can be helpful.

The sentiment analysis process can be conceptualized in steps, starting with the input text (the news article) and culminating in a sentiment score. This score is then used to classify the article's sentiment. While the internal workings of LLaMA 2 involves complex interactions of neural networks, the overall process can be simplified into an equation that represents the sentiment classification task:

$$S(article) = f(LLaMA2(article)) \quad (7)$$

Where $S(article)$ denotes the article's sentiment score. $LLaMA2(article)$ represents feeding the article into the LLaMA 2 model. $f(\cdot)$ is a function that interprets the output of LLaMA 2 to assign a sentiment score, typically within a range, e.g., $[0, 1]$, where 0 might indicate a completely negative sentiment and 1 a completely positive sentiment.

The quantification of sentiment from LLaMA 2's output involves mapping the model's linguistic analysis onto a numerical scale representing sentiment polarity. This could involve simple thresholding, where outputs above a certain value are considered positive, and those below are deemed negative, or a unique mapping considering the degree of sentiment expressed.

4. Results and Analysis

A. Without Sentiment Analysis

Figure 8 illustrates the performance of an LSTM model in predicting Bitcoin prices. The blue line represents actual prices, while the red line depicts the model's predictions over 100 epochs. The MAPE of 15.59% indicates the average deviation between the predicted and exact prices, with a lower percentage representing a more accurate model. Meanwhile, an accuracy measure of 84.405% suggests that the predictions align closely with the actual prices most of the time.

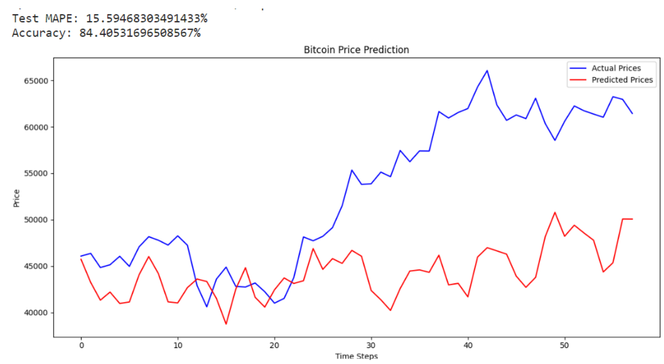

Figure 8. Results of LSTM model without sentiment analysis

The disparity between the two lines indicates the model's predictive capability. The red line (predicted prices) does not perfectly mirror the blue line (actual prices), which is common in real-world forecasting scenarios. Despite this, the model captures the general trend of the Bitcoin price movement with some lag and variance.

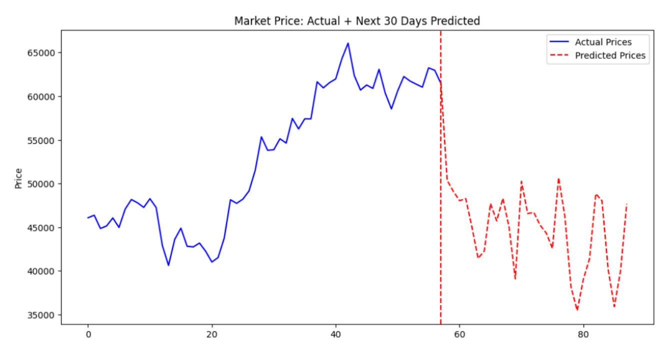

Figure 9. Results of Auto-Regressive prediction for Next 30 days

Figure 9 depicts actual versus predicted market prices over time steps. The continuous line represents the historical price data up to a certain point, after which the dashed line indicates the model's 30-day forward predictions. After the historical data ends, the divergence between the two lines demonstrates the model's predictions deviating from the actual trend observed in the earlier data. This variation signifies the model's challenges in capturing future price movements over an extended time without the inclusion of sentiment analysis. Figure 10 showcases a Prophet model

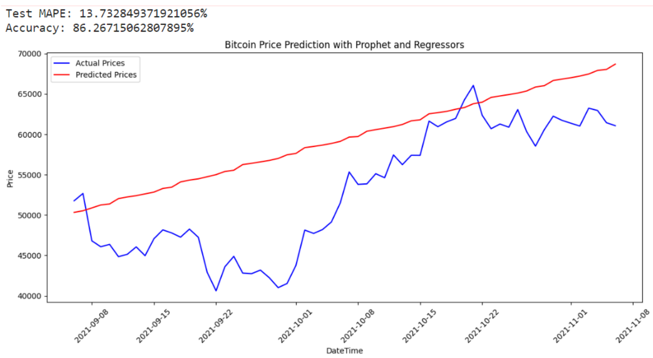

Figure 10. Results of Prophet model without sentiment analysis

results without incorporating news sentiment. A MAPE of 13.732% indicates the model's predictions are, on average, about 13.73% away from the actual prices, which measure forecast accuracy, with a lower value being preferable. The accuracy is reported to be 86.267%, suggesting that the model's predictions align closely with the actual prices for most of the data points. These results reflect the model's capability to track market trends and predict future prices with a reasonable degree of reliability. Figure 11

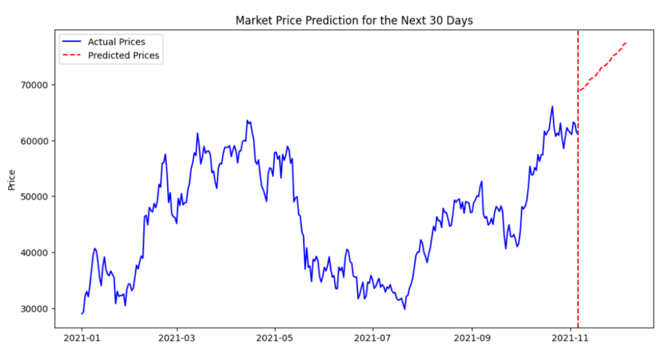

Figure 11. Results of Auto-Regressive prediction for Prophet Next 30 days

shows the Prophet model's forecast for Bitcoin market prices over a future 30-day period, compared with the historical price data. The solid line indicates the actual prices leading up to a delineated point, whereas the dashed line represents the model's predictions extending into the future. The projections show a trend, but as they diverge from the historical data, the model's capacity to capture and extend the real-world trajectory of Bitcoin prices into the subsequent 30 days is brought into focus. Figure 12 reflects the performance of the SARIMAX model used

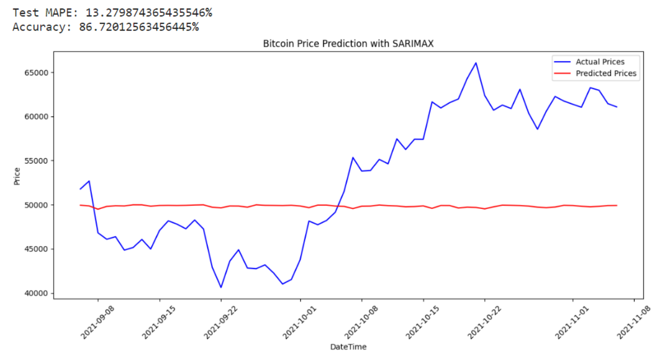

Figure 12. Results of SARIMAX without sentiment analysis

for predicting Bitcoin prices without sentiment analysis as an input. The MAPE stands at approximately 13.28%, indicating the average error rate of the model's predictions compared to the actual prices. An accuracy of about 86.72% suggests that the predicted prices generally follow the trend of the actual prices but with some degree of error. The predictive line demonstrates a consistent path, which likely means the model captures the average level but not the volatility in the actual prices. This points to a need for the model to account for more variability or possibly include additional inputs, such as sentiment data, to improve prediction alignment with the actual price movements. Figure

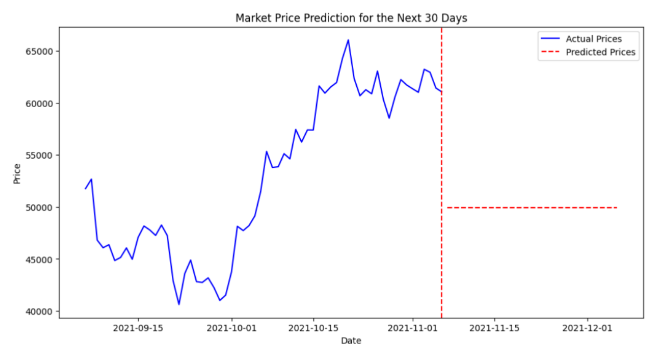

Figure 13. Results of Auto-Regressive prediction for SARIMAX Next 30 days

13 shows the actual market price of Bitcoin and its forecast over the next 30 days using the SARIMAX model. The historical data shows the variability and trends that the model attempts to capture. The forecast indicates a flat projection, suggesting the model's parameters might be optimized for a mean value rather than capturing future trends or seasonal effects. The stability in predicted prices contrasts with the historical volatility, indicating potential model limitations or the absence of influential variables, such as market sentiment, that could improve the forecast.

B. With Sentiment Analysis Included

Figure 14 shows the outcome of using an LSTM model for Bitcoin price prediction, incorporating sentiment analysis. The MAPE is approximately 12.75% , indicating the average error between the predicted and actual prices. An accuracy of around 87.25% suggests that the predictions are closely aligned with the exact prices. Including sentiment analysis enhances the model's predictive performance, resulting in a closer fit to the actual price movements than the results without sentiment data.

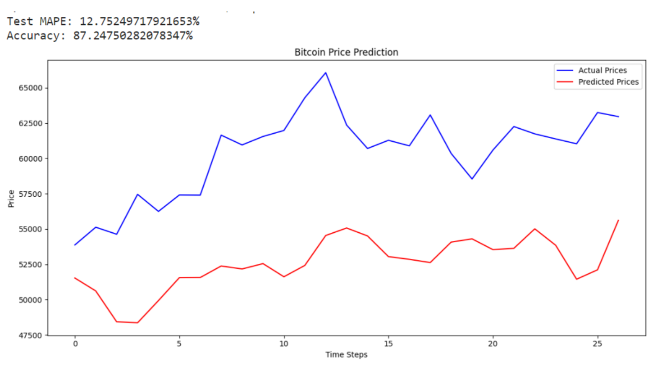

Figure 14. Results of LSTM model with sentiment analysis

Figure 15 indicates the performance of an LSTM model with sentiment analysis included for predicting the future market prices of Bitcoin over 30 days. The actual price trend ceases at a specific point, after which the predicted prices continue. The forecasted trend shows some fluctuations but generally remains lower than past prices. It suggests that while the model may capture some market movements, it has a conservative prediction bias over this extended time frame. This illustrates the model's predictive behavior when it incorporates sentiment analysis, reflecting its effort to account for the impact of market sentiment on future prices.

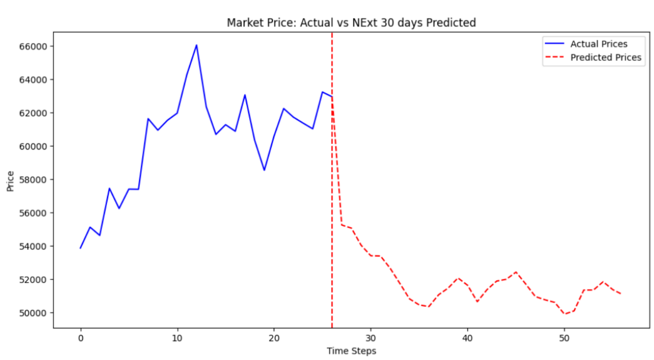

Figure 15. Results of Auto-Regressive prediction for Next 30 days

Figure 16 represents the outcomes of applying the Prophet model to predict Bitcoin prices, with sentiment analysis included as a factor. The MAPE is approximately 28.04%, indicating that the predictions deviate from the actual values by this percentage on average. An accuracy of roughly 71.96% suggests the model's ability to track the direction of price movements to a certain extent but with notable room for improvement. These statistics reflect the model's performance in incorporating sentiment data to inform its predictions, evidenced by the deviation of the predicted trend from the actual price movements.

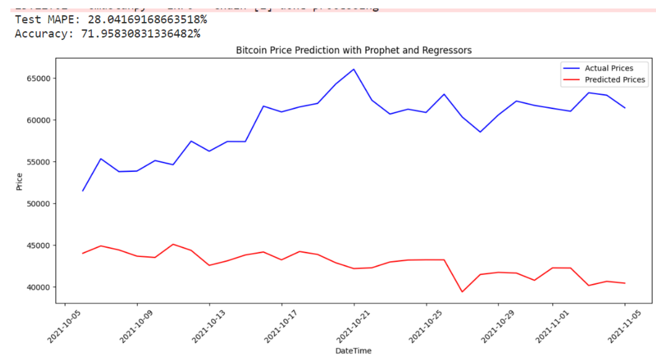

Figure 16. Results of Prophet model with sentiment analysis

Figure 17 displays the actual and forecasted Bitcoin prices, with the forecast extending for the next 30 days, utilizing a Prophet model that includes sentiment analysis. The actual prices show the known market data up to a delineation point, beyond which the model's predictions are plotted. The forecasted values indicate the model's estimation of future prices, with sentiment analysis incorporated as a variable to potentially capture the influence of market sentiment on price movements. The plot suggests a divergence between the historical trend and the projected values, which might highlight the model's conservative or less volatile prediction of future prices when extending the forecast. Figure 18 displays the outcome of a SARIMAX

Figure 17. Results of Prophet's Auto-Regressive prediction for Next 30 days

model that incorporates sentiment analysis for Bitcoin price prediction. However, the reported metrics 'Test MAPE' and 'Accuracy' are extremely low. Moreover, the graph presents a steady predicted price line, which does not reflect the actual price movement's variability, indicating that the model did not capture the volatility effectively.

Figure 19 presents the results of a SARIMAX model utilizing sentiment analysis to predict Bitcoin market prices for the next 30 days. The actual price data is shown

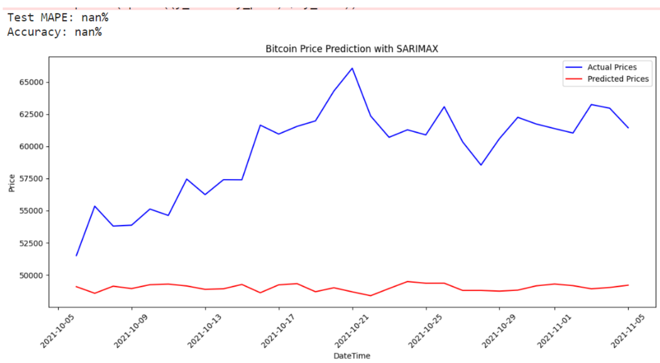

Figure 18. Results of SARIMAX model with sentiment analysis

up to a certain point, where the autoregressive prediction begins. The predicted prices indicate the model's forecast for future values, showing noticeable deviation from the actual observed prices leading up to the prediction point. This suggests that while the model can predict the general direction of prices, the exact future prices are likely to vary significantly from the model's predictions. Sentiment analysis aims to incorporate market sentiment into the forecast, potentially capturing influences on prices beyond past prices alone. However, the graph reflects the inherent challenge of accurately forecasting market prices over an extended period.

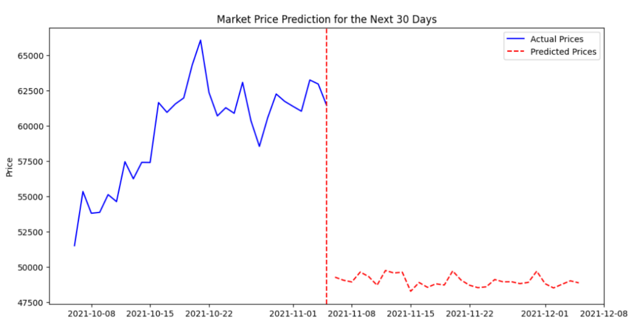

Figure 19. Results of Auto-Regressive prediction for Next 30 days

5. Discussion and Comparative Analysis

A. The Impact of Sentiment

The comparison of accuracy with and without sentiment analysis is provided in figure 20. In the comparative analysis of LSTM, Prophet, and SARIMAX models, sentiment analysis has yielded insightful divergences in model performance. The LSTM model demonstrated a discernible enhancement when supplemented with sentiment data. This was evidenced by a reduction in MAPE from 15.59% to 12.75% and an increase in accuracy from 84.405% to 87.25%. Such improvements suggest that the LSTM's architecture facilitates learning from both immediate and long-range dependencies in data. Moreover, it is well-suited to integrating the additional nuances that sentiment analysis provides. It is conceivable that the LSTM model

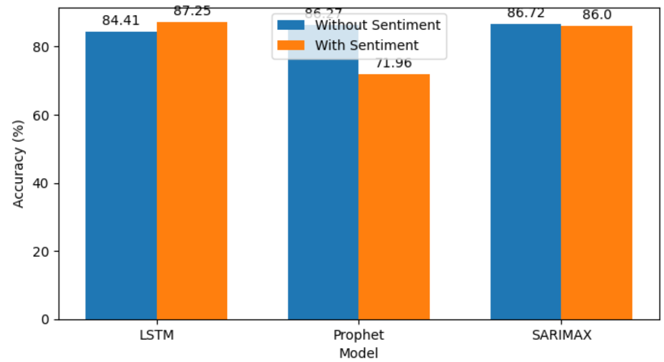

Figure 20. Results of Accuracy by Model and Sentiment Analysis

benefits from the richer, more complex representations of market sentiment, which may correlate with subtle shifts in market dynamics not fully captured by historical price data alone. Conversely, the Prophet model exhibited a counterintuitive trend; incorporating sentiment analysis led to an increased MAPE of 28.04% and reduced accuracy to 71.958%. The Prophet model, which emphasizes trend and seasonality in its forecasts, may encounter complexities when integrating sentiment analysis. The degradation in performance could be attributed to the model's potential overfitting to sentiment features that do not have a direct or consistent relationship with price movements. Moreover, the underlying assumptions of the Prophet model might not align well with the erratic nature of sentiment-driven market reactions, leading to a mismatch between prediction and actual price trajectories. The accuracy comparison with and without sentiment analysis is provided in the figure 21.

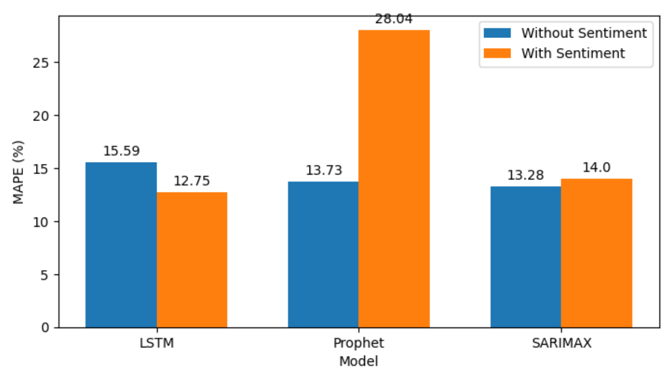

Figure 21. Results of Mean Absolute Percentage Error by Model and Sentiment Analysis

The SARIMAX model's relatively stable performance without sentiment analysis, indicated by a MAPE of 13.28% and an accuracy of 86.72%, did not translate into improved forecasting with the inclusion of sentiment data. This model, which is adept at handling external variables through its autoregressive structure, might not be as effective in capturing the predictive signals from sentiment data. This could be due to the complex interplay of sentiment indicators with traditional financial metrics, which the SARIMAX model may not adequately parse, potentially due to model constraints in capturing nonlinear relationships or the over-simplification of sentiment data.

The behavior of these models can be further dissected by considering their structural foundations. The LSTM's neural network framework is inherently more flexible and adaptable to new forms of data, such as those derived from sentiment analysis. In contrast, SARIMAX's reliance on specified lags and moving averages may not capture the immediacy and volatility of sentiment shifts. Prophet's performance dip with sentiment data suggests a potential for model over-parameterization, where the additional sentiment variables introduce noise rather than predictive power, pointing towards the need for careful feature selection and model regularization.

These findings illuminate the complexities of integrating sentiment analysis into predictive models for financial data. The congruence between model architecture and the characteristics of sentiment data is a determining factor in its success. As such, these insights call for a deeper investigation into model architectures and sentiment inclusion.

B. Auto Regressive Prediction for 30 Days

In the second part of experiment, we examined the autoregressive forecasting capabilities of the LSTM, Prophet, and the SARIMAX models. We particularly focused on the ability to project Bitcoin prices into the future 30 days. This provides insights into each model's adeptness in extrapolating trends and capturing the volatility of the cryptocurrency.

The LSTM model, when employing sentiment analysis, presented a conservative bias in its predictions. It is evidenced by the forecasted trend remaining consistently lower than the actual historical prices. This might be due to the LSTM's intrinsic characteristics, prioritizing learning from historical data patterns. While adding sentiment analysis imbues LSTM broader contextual understanding. It also introduces a degree of cautiousness, likely due to the variable nature of data. This cautious approach, although beneficial in mitigating risk, might limit the model's ability to capture sudden market upswings.

Contrastingly, the Prophet model's autoregressive forecasts revealed a notable deviation from the historical price. which shows a significant challenge in accurately extending real-world market dynamics into the future. The model designed for capturing seasonality might struggle with the erratic movements of Bitcoin prices. The observed divergence suggests that while the model can grasp the general direction of price movements, its predictions become less reliable over an extended time. It is particularly when integrating the less predictable sentiment data.

Even with sentiment analysis, the SARIMAX model's flat projection in its 30-day forecast underscores a potential limitation in capturing future trends, unfortunately showing no improvement in the its accuracy. This stability, while indicative of the model's robustness to short-term volatility, also highlights its potential inadequacy in anticipating long term market shifts. The model's parameterization, optimized for capturing mean values, might not account adequately for sentiment-driven aspects of the cryptocurrency market. Which can significantly influence price trajectories over time.

These observations underscore the inherent complexities and challenges in utilizing autoregressive models for long-term financial market predictions. While offering a unique perspective by incorporating market sentiment. Integrating sentiment analysis introduces additional layers of complexity and unpredictability. This analysis sheds light on the predictive capabilities and limitations of each model. Moreover, it highlights the relationship between historical price data and sentiment in forecasting future market behaviors. As such, these insights emphasize the need for a balanced approach that carefully considers each model's unique characteristics.

C. Comparative Analysis

Results compared with the extant literature to draw parallels, divergences, and unique contributions. Our study's findings, particularly the augmented performance of the LSTM model upon the integration of sentiment analysis, resonate with the assertions posited by literature [48], [49], [50], [51], [52]. These studies have previously underscored the potential of deep learning frameworks to synthesize diverse data streams effectively. For instance, including sentiment metrics may enhance predictive accuracy [52]. The LSTM is especially adept at assimilating the unique layers of sentiment data, thereby enriching the model's forecasting ability [49]. This is due to its architectural proclivity for processing sequential data and capturing temporal dependencies. This observation corroborates the utility of sentiment analysis in financial forecasting. Moreover, it also highlights the LSTM model's structural suitability for such integration.

Conversely, the attenuated performance observed in the Prophet model with the inclusion of sentiment analysis deviates from the optimistic outcomes anticipated by Mohan et al. [53], Wenyang, et al. [55]. These authors have extolled the Prophet model for its adaptability and robust handling of data anomalies. Our findings introduce a unique counterpoint to this narrative, suggesting that while the Prophet model is inherently flexible, its performance is contingent upon the data's uniformity and predictability. The unpredictable Bitcoin with intrinsic volatility of sentiment data, likely introduces stochasticity that challenges the Prophet model's ability.

Moreover, the SARIMAX model's conservative forecasting approach, particularly evident in its autoregressive projections matches the research of Fuad, G. [56]. Moreover, Rodriguez [54] also showed concerns regarding the challenges inherent in forecasting financial markets if it is combined with sentiment for models like SARIMAX. While potentially advantageous in stabilizing more predictable markets, SARIMAX falls short in cryptocurrency trading [57]. As the market dynamics are in constant flux, influenced by an array of factors including investor sentiment and regulatory changes.

Interestingly, our findings diverge from the more optimistic perspectives by [58], [59]. They advocated for the unmitigated efficacy of sentiment analysis in bolstering model forecasts [59]. Particularly the results of the Prophet model, suggest that the integration of sentiment data does not enhance predictive performance. This divergence hints at a potential misalignment between the market reactions driven by sentiment and the inherent forecasting capabilities of the model.

D. Limitations

When examining the models' capacity to extend predictions into the future, unveils distinct model-specific behaviors. The LSTM model, for instance, exhibits a conservative bias in its long-term forecasts, even with the inclusion of sentiment analysis. This confirmed the results of [60]. However, [12] hold the contradicting views. This tendency might limit the model's ability to capture potential market upsurges catalyzed by positive sentiment. Such insights add granularity to our understanding of each model's predictive dynamics. As such, it shows the complex relationship between historical price, sentiment analysis, and model architectures.

These results contribute to financial forecasting literature by explaining the unique effects of sentiment analysis across different predictive models. Moreover, it sheds light on the unique dynamics that govern long-term market predictions. These insights demand further exploration, particularly in refining model-specific methodologies. The Data was limited in terms of years which is one of the significant limitations we faced in this experiment. If we employ the longer version of the data the results are remarkable as depicted in figure 22. For future work using the longer Financial data with sentiment data may bring some more insights. This research, while challenging will be paramount in advancing the frontiers of financial forecasting, especially in cryptocurrencies.

E. Recommendations

In light of the findings, several avenues for future exploration emerge, particularly in financial forecasting and investment strategy optimization. One promising direction is the integration of predictive models with sentiment analysis into the broader portfolio optimization framework [61]. Traditional portfolio optimization strategies could significantly benefit from the unique market insights by sentiment analysis. Traditional portfolios often rely on historical price data and conventional financial metrics thereby enhancing the decision-making process for risk management.

The concept of portfolio optimization is rooted in Markowitz's Modern Portfolio Theory (MPT), and involves creating an 'efficient frontier' [62]. This represents a range of optimum investment groupings that deliver the highest expected return for a specified risk level, or the lowest risk for a predetermined return. Introducing predictive analytics into this concept might enhance the strategy for assembling investment portfolios. It considers not only past asset volatility and inter-asset relationships but also projections of market moods and patterns. This approach might be especially beneficial in the highly volatile and sentiment-sensitive cryptocurrency markets. By predicting upcoming price changes and shifts in the market, investors might be able to more effectively foresee market changes. This allows for the optimization of their investment groupings to either maximize predicted profits or protect against expected losses.

Furthermore, applying predictive models in portfolio optimization could extend to algorithmic trading. thus, trading decisions are made automatically based on predefined criteria derived from model outputs. Integrating sentiment analysis could add a layer of sophistication to trading algorithms, enabling them to respond to qualitative sentiment signals. For instance, a trading algorithm could be designed to increase its exposure to certain assets in anticipation of positive sentiment trends. It may reduce its position in assets facing negative sentiment headwinds.

However, integrating sentiment analysis and predictive modeling into portfolio optimization and algorithmic trading is challenging. The volatility in sentiment data, the potential for overfitting, and the need for real-time data processing must be addressed. Moreover, cryptocurrency markets' non-stationary and speculative nature demands robust risk management strategies. In order to mitigate the potential downsides of sentiment-driven investment decisions.

6. Conclusion

This comprehensive study examines the efficacy of leveraging sentiment analysis alongside traditional predictive models in forecasting Bitcoin prices. This revealed unique insights into how each model assimilates sentiment data, subsequently impacting their predictive accuracy and forecasting capabilities. The LSTM model demonstrated a notable enhancement in performance with the inclusion of sentiment analysis, underscoring the model's inherent flexibility and suitability for integrating complex, multidimensional data. Conversely, the Prophet and SARIMAX models exhibited varied responses to sentiment data integration, with the Prophet model showing a decrease in predictive accuracy, suggesting potential challenges in aligning sentiment-driven market dynamics with the model's forecasting mechanisms. The autoregressive forecasting analysis further illuminated the models' capabilities and limitations in extending predictions over a 30-day period. Here, the conservative forecasting nature of the models, particularly

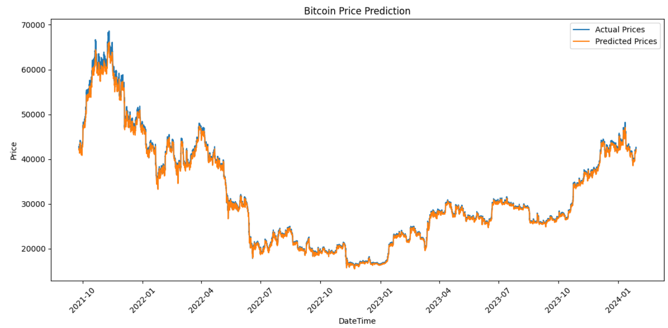

Figure 22. The Perfect Results of LSTM when Trained on the Long-Term Data (2017 - 2024)

SARIMAX, highlighted the inherent challenges in capturing the volatile tendencies of cryptocurrency markets.

Our study contributes to the literature on financial forecasting by providing empirical evidence on the value and complexities of incorporating sentiment analysis into predictive models. As the financial field continues to evolve, the need for sophisticated, data-driven forecasting models has never been more pronounced. Our research highlights the potential of combining quantitative and qualitative data to enhance forecasting models. It paves the way for more informed and strategic decision-making in financial markets. It extends the theoretical foundations of economic forecasting and offers practical insights for investors, traders, and policymakers.

References

[1] V. Decaro, "Crypto-currency and blockchain: the revolution of the world economic system and digital identities," 2017.

[2] R. M. Bratspies, "Cryptocurrency and the myth of the trustless transaction," *Mich. Tech. L. Rev.*, vol. 25, p. 1, 2018.

[3] V. Sangwan, P. Prakash, and S. Singh, "Financial technology: a review of extant literature," *Studies in Economics and Finance*, vol. 37, no. 1, pp. 71–88, 2020.

[4] T. I. Kiviat, "Beyond bitcoin: Issues in regulating blockchain transactions," *Duke LJ*, vol. 65, p. 569, 2015.

[5] A. R. Garcia and P. H. R. Garcia, "Cryptocurrencies: the communication inside blockchain technology and the cross-border tax law," *International Journal of Blockchains and Cryptocurrencies*, vol. 1, no. 1, pp. 22–41, 2019.

[6] Y. Doumenis, J. Izadi, P. Dhamdhere, E. Katsikas, and D. Koufopoulos, "A critical analysis of volatility surprise in bitcoin cryptocurrency and other financial assets," *Risks*, vol. 9, no. 11, p. 207, 2021.

[7] H. T. Akkuş and İ. Çelik, "Modeling, forecasting the cryptocurrency market volatility and value at risk dynamics of bitcoin," *Muhasebe Bilim Dünyası Dergisi*, vol. 22, no. 2, pp. 296–312, 2020.

[8] E. Akyildirim, A. Goncu, and A. Sensoy, "Prediction of cryptocurrency returns using machine learning," *Annals of Operations Research*, vol. 297, pp. 3–36, 2021.

[9] H. N. D. ŞENYAPAR, "Cryptocurrency on social media: Analyzing the digital discourse towards the coin market," *İktisadi İdari ve Siyasal Araştırmalar Dergisi*, vol. 9, no. 23, pp. 202–223, 2024.

[10] M. F. Rizkilloh, S. Widiyanesti *et al.*, "Prediksi harga cryptocurrency menggunakan algoritma long short term memory (lstm)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 25–31, 2022.

[11] I. Nasirtafreshi, "Forecasting cryptocurrency prices using recurrent neural network and long short-term memory," *Data & Knowledge Engineering*, vol. 139, p. 102009, 2022.

[12] N. Latif, J. D. Selvam, M. Kapse, V. Sharma, and V. Mahajan, "Comparative performance of lstm and arima for the short-term prediction of bitcoin prices," *Australasian Accounting, Business and Finance Journal*, vol. 17, no. 1, pp. 256–276, 2023.

[13] A. Judmayer, N. Stifter, K. Krombholz, and E. Weippl, *Blocks and chains: introduction to bitcoin, cryptocurrencies, and their consensus mechanisms*. Springer Nature, 2022.

[14] A. Bâra and S.-V. Oprea, "An ensemble learning method for bitcoin price prediction based on volatility indicators and trend," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 107991, 2024.

[15] N. Maleki, A. Nikoubin, M. Rabbani, and Y. Zeinali, "Bitcoin price prediction based on other cryptocurrencies using machine learning and time series analysis," *Scientia Iranica*, vol. 30, no. 1, pp. 285–301, 2023.

[16] N. Uras, L. Marchesi, M. Marchesi, and R. Tonelli, "Forecasting bitcoin closing price series using linear regression and neural networks models," *PeerJ Computer Science*, vol. 6, p. e279, 2020.

[17] G. Swankie, "Examining the price dynamics of the cryptocurrency market and predicting bitcoin price through the application of statistical analysis and deep learning," *MSc Financial Technology*, p. 9, 2019.

[18] J. Blackledge and M. Lamphiere, "A review of the fractal market hypothesis for trading and market price prediction," *Mathematics*, vol. 10, no. 1, p. 117, 2021.

[19] E. J. A. Abakah, A. K. Tiwari, S. Ghosh, and B. Doğan, "Dynamic effect of bitcoin, fintech and artificial intelligence stocks on eco-friendly assets, islamic stocks and conventional financial markets: Another look using quantile-based approaches," *Technological Forecasting and Social Change*, vol. 192, p. 122566, 2023.

[20] M. Sheraz, S. Dedu, and V. Preda, "Volatility dynamics of non-linear volatile time series and analysis of information flow: Evidence from cryptocurrency data," *Entropy*, vol. 24, no. 10, p. 1410, 2022.

[21] H. Treiblmaier, M. Swan, P. De Filippi, M. Lacity, T. Hardjono, and H. Kim, "What's next in blockchain research? -an identification of key topics using a multidisciplinary perspective," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, vol. 52, no. 1, pp. 27–52, 2021.

[22] B. Tripathi and R. K. Sharma, "Modeling bitcoin prices using signal processing methods, bayesian optimization, and deep neural networks," *Computational Economics*, vol. 62, no. 4, pp. 1919–1945, 2023.

[23] A. M. Khedr, I. Arif, M. El-Bannany, S. M. Alhashmi, and M. Sreedharan, "Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey," *Intelligent Systems in Accounting, Finance and Management*, vol. 28, no. 1, pp. 3–34, 2021.

[24] S. Sengupta, S. Basak, P. Saikia, S. Paul, V. Tsalavoutis, F. Atiah, V. Ravi, and A. Peters, "A review of deep learning with special emphasis on architectures, applications and recent trends," *Knowledge-Based Systems*, vol. 194, p. 105596, 2020.

[25] K. Rathan, S. V. Sai, and T. S. Manikanta, "Crypto-currency price prediction using decision tree and regression techniques," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 190–194.

[26] M. Sivaram, E. L. Lydia, I. V. Pustokhina, D. A. Pustokhin, M. Elhoseny, G. P. Joshi, and K. Shankar, "An optimal least square support vector machine based earnings prediction of blockchain financial products," *IEEE Access*, vol. 8, pp. 120 321–120 330, 2020.

[27] A. B. Turner, S. McCombie, and A. J. Uhlmann, "Analysis techniques for illicit bitcoin transactions," *Frontiers in Computer Science*, vol. 2, p. 600596, 2020.

[28] L. Harting and N. Åkesson, "A machine learning approach leveraging technical-and sentiment analysis to forecast price movements in major crypto currencies," 2022.

[29] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European journal of operational research*, vol. 270, no. 2, pp. 654–669, 2018.

[30] A. Ibrahim, R. Kashef, and L. Corrigan, "Predicting market movement direction for bitcoin: A comparison of time series modeling methods," *Computers & Electrical Engineering*, vol. 89, p. 106905, 2021.

[31] I. Yenidoğan, A. Çayir, O. Kozan, T. Dağ, and Ç. Arslan, "Bitcoin forecasting using arima and prophet," in *2018 3rd international conference on computer science and engineering (UBMK)*. IEEE, 2018, pp. 621–624.

[32] R. G. Tiwari, A. K. Agarwal, R. K. Kaushal, and N. Kumar, "Prophetic analysis of bitcoin price using machine learning approaches," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*. IEEE, 2021, pp. 428–432.

[33] J. R. Simon and K. Geetha, "Block mining reward prediction with polynomial regression, long short-term memory, and prophet api for ethereum blockchain miners," in *ITM Web of Conferences*, vol. 37. EDP Sciences, 2021, p. 01004.

[34] F. R. Alharbi and D. Csala, "A seasonal autoregressive integrated moving average with exogenous factors (sarimax) forecasting model-based time series approach," *Inventions*, vol. 7, no. 4, p. 94, 2022.

[35] A. Ampountolas, "Modeling and forecasting daily hotel demand: A comparison based on sarimax, neural networks, and garch models," *Forecasting*, vol. 3, no. 3, pp. 580–595, 2021.

[36] A. Kumar, N. Sharma, R. Chauhan, and M. Sharma, "Cryptocurrency price forecasting in a volatile landscape: Sarimax modeling and short-term strategies," in *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. IEEE, 2023, pp. 1–6.

[37] G. R. Achmadi, A. Saikhu, and B. Amaliah, "Cryptocurrency price movement prediction using the hybrid sarimax-lstm method," in *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*. IEEE, 2023, pp. 711–716.

[38] X. Yu, "Analysis of new sentiment and its application to finance," Ph.D. dissertation, 2014.

[39] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," *PeerJ Computer Science*, vol. 7, p. e476, 2021.

[40] R. Parekh, N. P. Patel, N. Thakkar, R. Gupta, S. Tanwar, G. Sharma, I. E. Davidson, and R. Sharma, "Dl-guess: Deep learning and sentiment analysis-based cryptocurrency price prediction," *IEEE Access*, vol. 10, pp. 35 398–35 409, 2022.

[41] S. Mandvikar, "Augmenting intelligent document processing (idp) workflows with contemporary large language models (llms)," *International Journal of Computer Trends and Technology*, vol. 71, no. 10, pp. 80–91, 2023.

[42] R. Lad, "Sarcasm detection in english and arabic tweets using transformer models," 2023.

[43] E. Eigner and T. Händler, "Determinants of llm-assisted decision-making," *arXiv preprint arXiv:2402.17385*, 2024.

[44] G. R. Taylor, *Integrating quantitative and qualitative methods in research*. University press of America, 2005.

[45] A. Kolková and A. Ključnikov, "Demand forecasting: Ai-based, statistical and hybrid models vs practicebased models-the case of smes and large enterprises," *Economics & Sociology*, vol. 15, no. 4, pp. 39–62, 2022.

[46] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Information Fusion*, vol. 44, pp. 65–77, 2018.

[47] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[48] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and lstm," *Neural Computing and Applications*, vol. 32, pp. 9713–9729, 2020.

[49] M.-L. Thormann, J. Farchmin, C. Weisser, R.-M. Kruse, B. Säfken, and A. Silbersdorff, "Stock price predictions with lstm neural net-

works and twitter sentiment," *Statistics, Optimization & Information Computing*, vol. 9, no. 2, pp. 268–287, 2021.

[50] T. Swathi, N. Kasiviswanath, and A. A. Rao, "An optimal deep learning-based lstm for stock price prediction using twitter sentiment analysis," *Applied Intelligence*, vol. 52, no. 12, pp. 13 675–13 688, 2022.

[51] A. Sharaff, T. R. Chowdhury, and S. Bhandarkar, "Lstm based sentiment analysis of financial news," *SN Computer Science*, vol. 4, no. 5, p. 584, 2023.

[52] F. Moodi, A. Jahangard-Rafsanjani, and S. Zarifzadeh, "A cnn-lstm deep neural network with technical indicators and sentiment analysis for stock price forecastings," in *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2024, pp. 1–6.

[53] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *2019 IEEE fifth international conference on big data computing service and applications (BigDataService)*. IEEE, 2019, pp. 205–208.

[54] A. Rodriguez and C. Soper, "Economic forecasting and consumer sentiment: The case of connecticut," *Journal of Business and Economic Studies*, vol. 24, no. 2, pp. 1–18, 2020.

[55] W. Wang, N. He, M. Chen, and P. Jia, "Freight rate index forecasting with prophet model based on multi-dimensional significant events," *Expert Systems with Applications*, p. 123451, 2024.

[56] G. Fuad, "Economic forecasting with news headlines and natural language processing," 2023.

[57] S. M. Ulyah, M. F. F. Mardianto *et al.*, "Comparing the performance of seasonal arimax model and nonparametric regression model in predicting claim reserve of education insurance," in *Journal of Physics: Conference Series*, vol. 1397, no. 1. IOP Publishing, 2019, p. 012074.

[58] I. Ghosh and P. Dragan, "Can financial stress be anticipated and explained? uncovering the hidden pattern using eemd-lstm, eemd-prophet, and xai methodologies," *Complex & Intelligent Systems*, vol. 9, no. 4, pp. 4169–4193, 2023.

[59] A. Limone, M. Gupta, N. Nagar, and S. Prajapat, "Effect of financial news headlines on crypto prices using sentiment analysis," in *UK Workshop on Computational Intelligence*. Springer, 2023, pp. 209–219.

[60] A. Ibrahim, R. Kashef, M. Li, E. Valencia, and E. Huang, "Bitcoin network mechanics: Forecasting the btc closing price using vector auto-regression models based on endogenous and exogenous feature variables," *Journal of Risk and Financial Management*, vol. 13, no. 9, p. 189, 2020.

[61] S. Gaskin, R. Kalim, K. J. Wallace, D. Islip, R. H. Kwon, and J. K.-S. Liew, "Portfolio optimization techniques for cryptocurrencies," *The Journal of Investing*, 2023.

[62] W. B. Lindquist, S. T. Rachev, Y. Hu, and A. Shirvani, "Modern portfolio theory," in *Advanced REIT Portfolio Optimization: Innovative Tools for Risk Management*. Springer, 2022, pp. 29–48.

Mohamad El Abaji is a results-driven technology expert with more than 14 years of experience, specializing in artificial intelligence, machine learning, and cloud solutions. He has excelled in developing and deploying sophisticated AI/ML and cloud integrations that enhance enterprise operations. Throughout his career, Mohamad has held key roles at VMware, Sensetime, and IBM, where he pioneered the adoption of innovative AI and cloud technologies across various business sectors, driving substantial advancements in their technological frameworks.

Ramzi A. Haraty is an associate professor of Computer Science in the Department of Computer Science and Mathematics at the Lebanese American University in Beirut, Lebanon. He is the chairperson of the Arab Computer Society. He is also a program evaluator (PEV) for CSAB/ABET. He received his B.S. and M.S. degrees in Computer Science from Minnesota State University - Mankato, Minnesota, and his Ph.D. in Computer Science from North Dakota State University - Fargo, North Dakota. His research interests include information management systems, machine learning, and multilevel secure systems engineering. He has well over 110 books, book chapters, journal and conference paper publications. He supervised over 110 dissertations, theses and capstone projects. He is a member of the Association of Computing Machinery (ACM), ACM SIGAPP, ACS, Institute of Electronics, Information and Communication Engineers, the International Society for Computers and Their Applications, and Syndicate of Computer Sciences in Lebanon.