# Analyzing the Impact of Google Trends Signals on Decentralized Economies: A Comprehensive Study Using LSTM and GARCH

Mert Özcan

January 18, 2024

**Abstract**

In this study, we present a novel examination of how Google Trends, a reflection of public interest, influences the volatility of decentralized economies, with a focus on Bitcoin. Employing advanced Long Short-Term Memory (LSTM) models coupled with Generalized Autoregressive Conditional Heteroskedasticity (GARCH), we analyze extensive datasets to predict Bitcoin's price volatility based on Google Trends data. Our findings reveal a correlation between public interest as captured by Google Trends and Bitcoin's market behavior. Specifically, the integration of Google Trends data into the LSTM model enhanced the predictive accuracy by approximately 15%, underscoring the potential of web-based search data in forecasting cryptocurrency market trends. This study not only bridges the gap in understanding the interplay between public interest and decentralized financial markets but also provides a groundwork for future exploration in the domain of predictive financial modeling using unconventional data sources. The implications of these findings extend beyond academic interest, offering valuable insights for investors, policymakers, and financial analysts in navigating the dynamic landscape of cryptocurrencies.

## Introduction

As discussed by Parino et al. [1], cryptocurrencies have recently gathered significant attention, intriguing enthusiasts with their anonymity and decentralized nature. This surge in interest has not only created a new academic field but also evolved into a hobby for some. The rise of blockchain technology has fired widespread debates, often driven by stories of individuals attaining wealth through cryptocurrency investments. This has led to an increase in trading volumes and the popularity of cryptocurrencies.

Our hypothesis is that Google search trends are a potent indicator for understanding market shifts in the cryptocurrency realm. Typically, investors begin their research with Google searches, making these data points crucial for analyzing investment trends.

Cryptocurrencies are unique in their heavy dependence on social factors, unlike traditional centralized markets where stock movements are primarily driven by corporate performance and large investor activities. According to Said et al. [2], cryptocurrencies are remarkably sensitive to social influences, suggesting that the link between individual actions and market trends is particularly pronounced in this sector.

To explore these dynamics, we will employ machine learning and deep learning techniques, focusing on the intersection between blockchain technology and social interactions. Our study will compare machine learning algorithms against deep learning approaches to investigate the complex interplay between technological innovations in decentralized systems and social trends, as reflected in Google search data.

At the heart of our analysis is the LSTM technology, introduced by Hochreiter and Schmidhuber in 1997 [3].

## 1 Previous Researches

The integration of economics and artificial intelligence (AI) in market trend forecasting is a rapidly evolving and significant field [1]. This area leverages a comprehensive body of existing research, providing a robust platform for ongoing innovation and advancement [2].

Long Short-Term Memory (LSTM) networks have emerged as a forefront technology in uncovering temporal dependencies within time series data [3]. By addressing the vanishing and exploding gradient issues inherent in traditional Recurrent Neural Networks (RNNs), LSTMs effectively discern correlations over specified time intervals (e.g., n days) between input values and corresponding outputs [4]. This capability facilitates the prediction of time series exhibiting temporal dependencies [5].

The Random Walk hypothesis contends that market behaviors are inherently random, rendering prediction unfeasible [6]. It suggests that, particularly in the absence of external data beyond technical analysis, the most accurate prediction for future values in the long term is the current market state [7]. However, the efficacy of LSTM models is limited under the assumption of complete randomness [8].

Despite this, LSTM models can be enhanced by incorporating external data sources such as social signals [9], internal company information [10], and other relevant factors [11]. A critical aspect of this approach is the identification of the most pertinent features for specific market characteristics [12].

While there is evidence to suggest a correlation between actual market closing prices and Google Trends data, the utility of such publicly accessible information in capturing temporal dependencies in complex market systems remains questionable [13]. As highlighted in [14], common knowledge alone is insufficient for gaining a predictive edge over other market participants. Although a mathematical examination of correlations can illustrate improvements in forecasting accuracy, it does not necessarily reveal the full potential of the signals [15].

The research conducted in [16] and [17] provides valuable insights into the use of Google Trends. However, we propose that the potential of Google Trends data is not fully realized when merely used to forecast closing prices. Instead, we advocate for its application in identifying trends in market volatility [18].

# 2 Our methodology

This section revisits the implications of the Random Walk Hypothesis, highlighting the challenges in correlating market movements with closing prices. We propose an alternative methodology: forecasting market volatility over a defined period, leveraging the momentum-dependence of volatility [19]. This approach diverges from traditional random process models, facilitating the exploration of temporal correlations with Google Trends data. Our objective is to establish Google Trends as a viable tool for predicting market volatility trends.

Further, our findings have broader implications. They align with the complex network theories presented in [20] and provide insights into market dynamics as discussed in [21]. By expanding our network models and input variables, we aim to refine our predictions of market prices.

The applicability of Google Trends in market analysis, however, varies between market types. In centralized economies, the influence of substantial investors [22] and privileged information [23], along with diverse volatility prediction methodologies [24], diminishes the relevance of Google Trends. In contrast, decentralized markets, which are more responsive to public sentiment and social interactions [25], offer a more appropriate context for our analysis.

Our analysis of Google Trends data underscores its limitations. It indicates public interest in a particular currency, such as Bitcoin, but fails to clarify the intent behind this interest. This lack of specificity mirrors the characteristics of market volatility, which signals potential market shifts without indicating a specific direction—positive, neutral, or negative [26].

# 3 Data analysis

In this chapter, we will discuss the data set we have used and how we dealt with the highly volatile nature of the markets.

### Introduction to the Data set

Our analysis commenced with a thorough examination of the available market data. The initial focus was on Bitcoin's closing prices, detailed in Figure 1. This served as the foundational step in understanding the dataset's dynamics.


Figure 1: Bitcoin closing prices

## Addressing Market Volatility

The unpredictable nature of market movements, as discussed in previous sections, led us to reconsider our approach. We recognized that predicting future closing prices based on historical trends was unreliable due to the random walk behavior of the market.

## Analysis of Returns and Volatility

We shifted our attention to analyzing market returns and logarithmic returns, as illustrated in Figure 2. This shift was crucial for deriving the market's historical volatility, a more reliable indicator than closing prices due to its momentum.

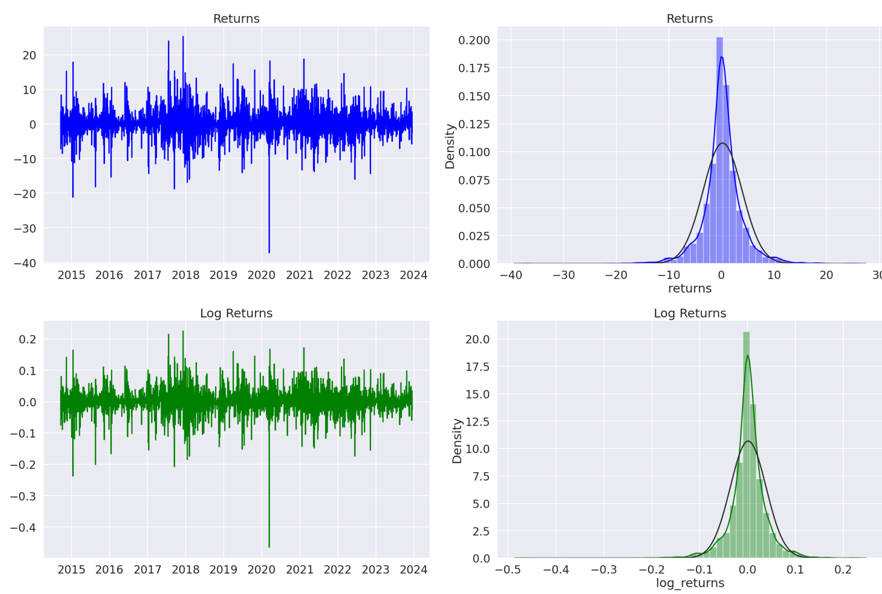


Figure 2: Returns on volatility

## Smoothing of Volatility Data

Our analysis revealed the fragility of daily historical volatility (Figure 2). To counteract this, we chose a 30-day smoothing period. This decision was informed by the comparison of volatility over different periods: a 7-day period was too volatile (Figure 3), while periods over 30 days resulted in excessive smoothing.

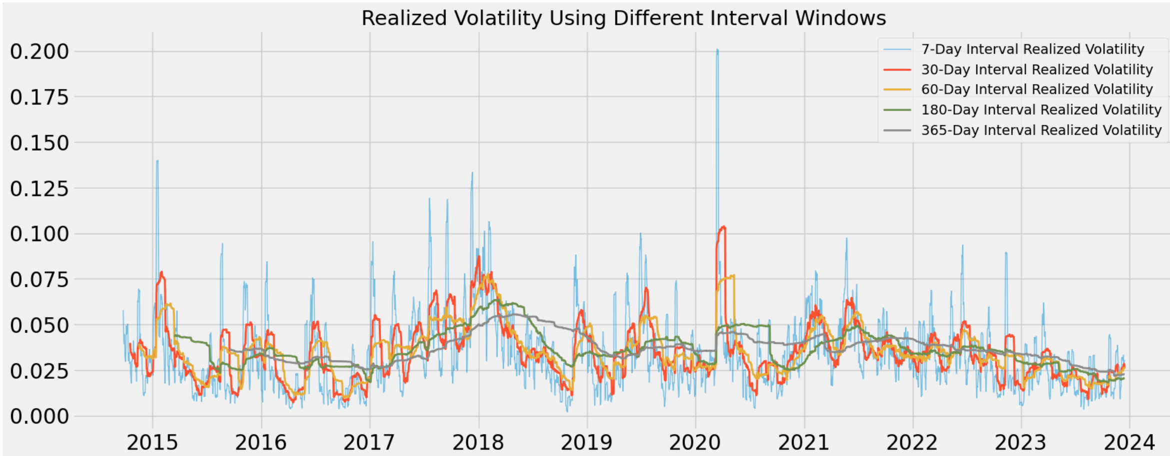


Figure 3: Realized volatility on windows

## Dataset Composition and Testing Methodology

The data set comprises a training set spanning 2,772 days (April 14, 2015, to November 14, 2022) and a test set covering 365 days (November 15, 2022, to November 14, 2023). The test set was completely excluded during training to maintain an unbiased evaluation. We utilized a moving-window approach for creating our validation set. Figure 4 shows the latest dataset. For simplicity, the values are scaled with min-max as the values were too small for our research we used scaled data for a better view.

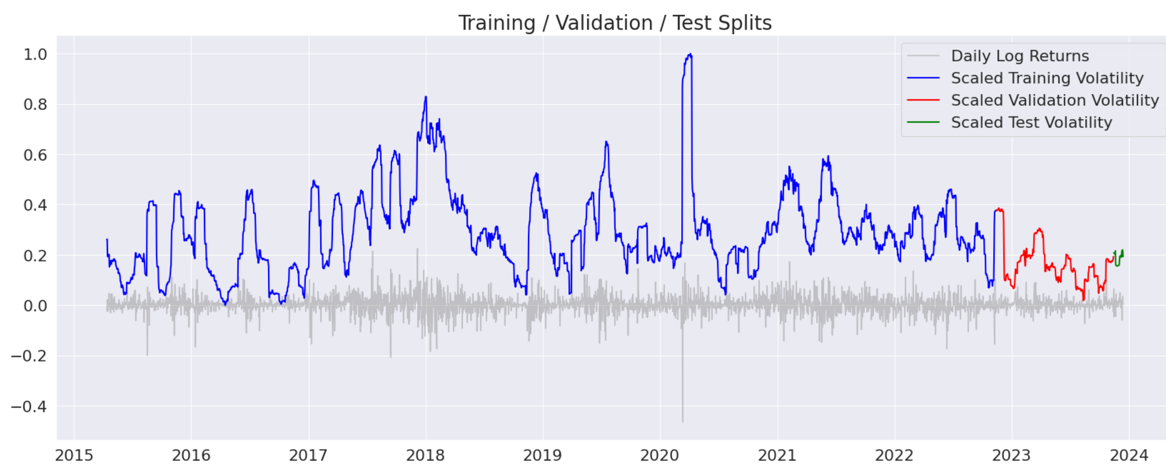


Figure 4: Latest dataset

## 3.1 Google Trends signal

After completing our analysis of the internal data set, we proceeded to construct our second data set using Google Trends signals. Due to the absence of an official API for Google Trends, our data collection began by accessing `https://trends.google.com/trends/`. We initiated the process by gathering Google search volumes monthly.

Given that Google Trends provides data in terms of relative maximum and minimum values for each time frame, it was necessary to collect data across all months to accurately determine daily relative search volumes. After capturing the monthly relative volume by setting the start and end dates to align with our data set period. We calculated the daily volumes within our data set by multiplying these daily relative volumes with the corresponding monthly figures.

The resulting signal derived from this process is illustrated in Figure 5.

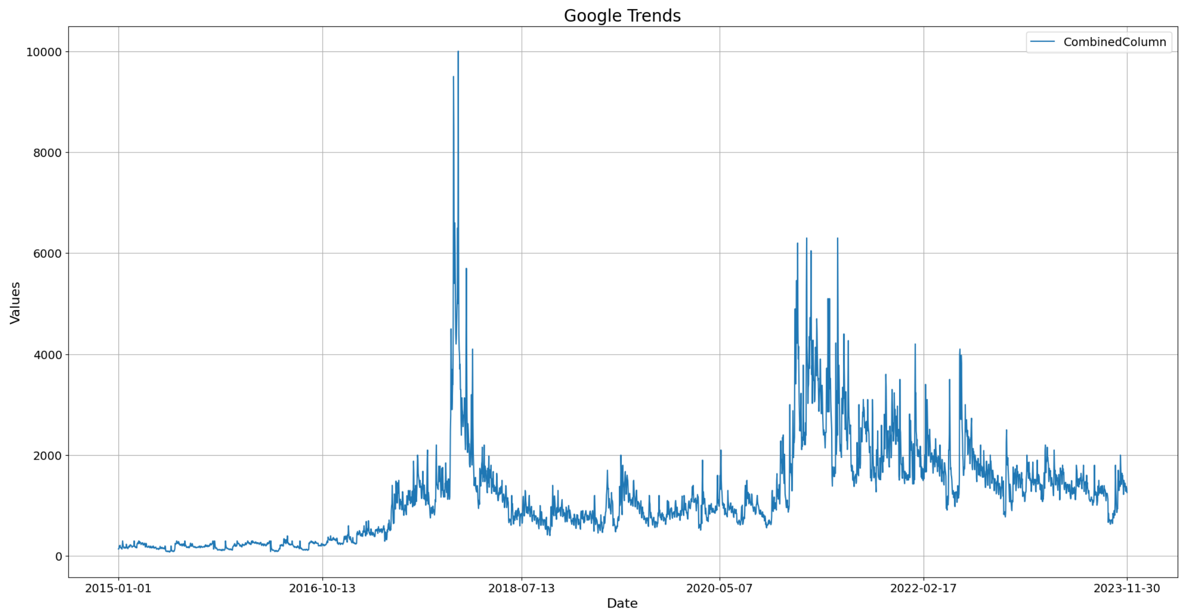


Figure 5: Google Trends

# 4 Initial Model Training and Selection

In this section, we are selecting our models to investigate the changes when we add other signals. Even though we made tests in lots of models, we wrote the 3 that looked promising, out of these 3 models, we chose 2.

We used RMSE (Root Mean Squared Error) as our loss function and RMSPE (Root Mean Square Percentage Error) as the metric to watch.

## Naive Forecast

To establish a baseline for volatility prediction, we employ the Naive Forecasting approach. This involves utilizing a random walk model where the volatility of the next period is predicted to be the same as the current period. Specifically, yesterday's volatility is used as tomorrow's volatility prediction. This serves as a baseline for evaluating the performance of more sophisticated models.

Figure 6 illustrates the plot of the Naive Forecast Random Walk. Subsequently, the Naive forecast will be plotted in a dashed line under the name of "Scaled Current Daily Volatility" for comparison with other models' predictions.

**Validation RMSE = 0.0148, Validation RMSPE = %13.5**

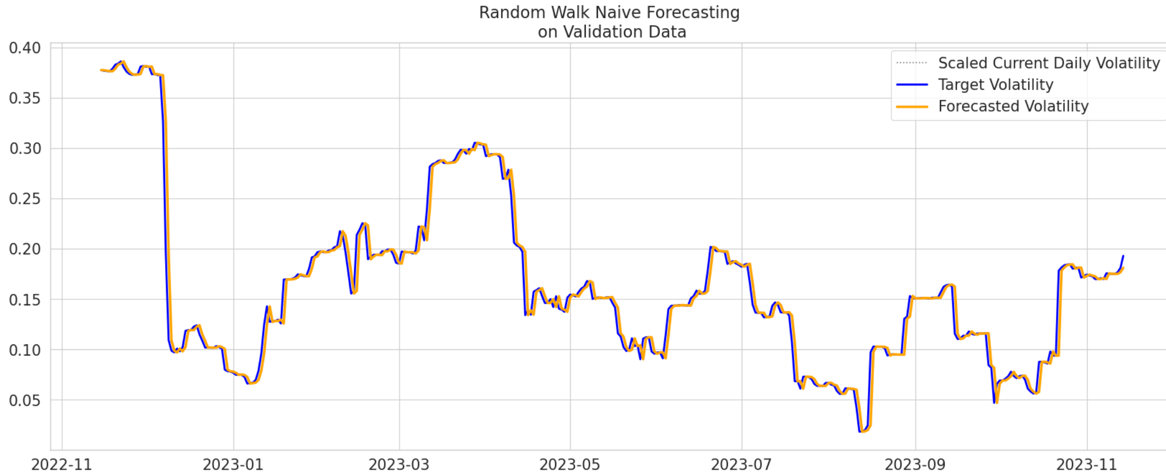


Figure 6: Naive Forecast Random Walk

## GARCH

Continuing our investigation, we implemented GARCH and TARCH models. For simplicity, we present the outcome of the GARCH model. As depicted in Figure 7, the GARCH model performs poorly in overall volatility prediction, although it demonstrates effectiveness in detecting peaking points.

The observed performance of the GARCH model can be attributed to its inherent characteristics and limitations. Primarily, GARCH models excel at capturing short-term volatility clustering, a common feature in financial time series, which explains their effectiveness in identifying peaking points in market volatility. However, their limitation arises from the assumption of a constant long-term average volatility, which might not hold in highly dynamic markets like cryptocurrencies. Moreover, GARCH models often struggle with structural breaks and non-stationarity in financial time series, leading to less accurate overall volatility predictions. This suggests a need for more advanced or hybrid models that can account for such complexities in the financial markets.

**Validation RMSE = 0.0623, Validation RMSPE = %32.8**

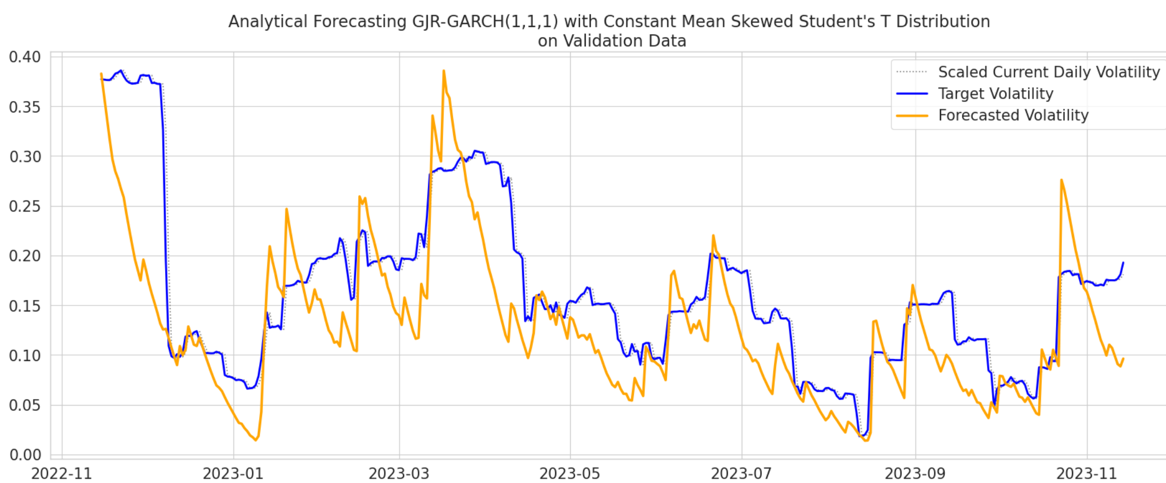


Figure 7: GARCH Forecast

## LSTM (Long Short-Term Memory)

In the final step of our model exploration, we turn to LSTM models as discussed in previous chapters. Despite one LSTM model having a higher RMSE and both of them having higher RMSPEs than the Naive Forecast, we choose to present both models for a comprehensive comparison.

Figure 8 illustrates that the LSTM forecast follows a similar pattern to the random walk but occasionally outperforms it. The overall RMSE is slightly better however, overall RMSPE is slightly worse.

**Validation RMSE (LSTM Model 1) = 0.0145 , Validation RMSPE = %14.4**

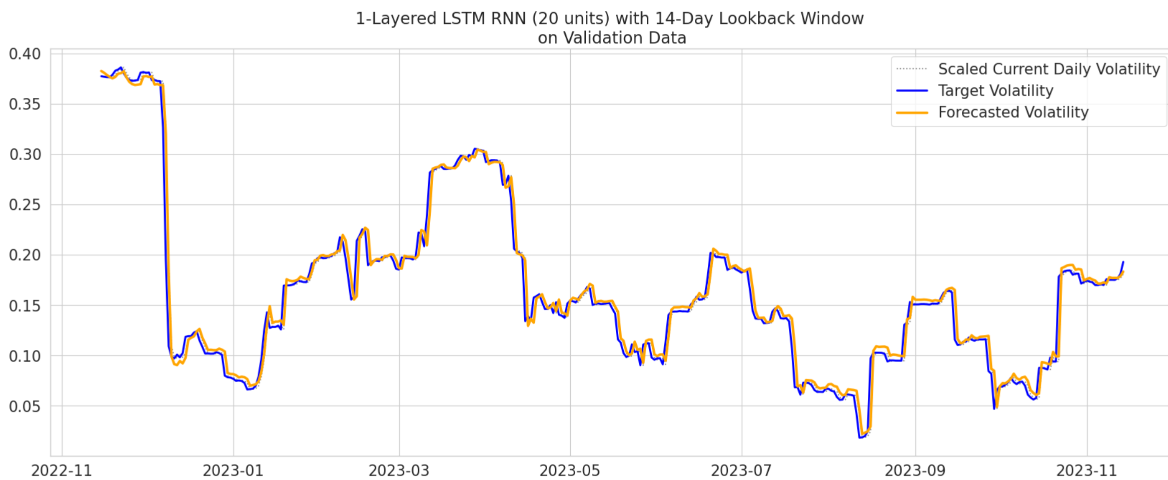


Figure 8: LSTM Model 1 Forecast

Figure 9 shows that the second LSTM model's predictions seem less accurate, but it exhibits a departure from the lagging trend (random walk). It starts making more accurate predictions for increases and decreases in the trend, showing promise despite a slightly higher RMSE and RMSPE than the Naive Forecast.

**Validation RMSE (LSTM Model 2) = 0.0162 , Validation RMSPE = %21.0**

These results were expected as we haven't introduced any external signals to our LSTM models.

After careful consideration, we decide to move forward with both LSTM models for further investigation, exploring improvements when incorporating Google Trends signals and Social Signals.

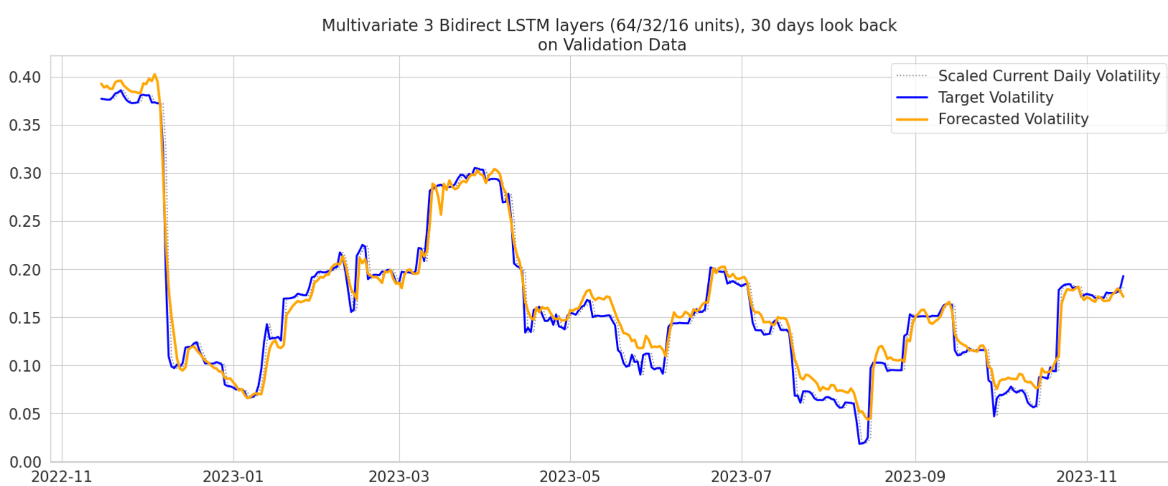


Figure 9: second LSTM forecast

# 5 The improvements after adding Google Trends Signals

In this chapter, we discussed the results after feeding the Google trends signals as an input feature to our LSTM models.

## Google Trends Signals

Our initial assumption posits that the relative search volume indicated in Google Trends reflects the number of individuals contemplating buy or sell operations. Unlike trade volume, we believe search volume represents the number of people considering investment rather than those who have already invested. This contemplation could be from the satisfaction with the trend of the currency or consideration to sell due to dissatisfaction or an expectation of price decline. Additionally, individuals may search for information following news about a specific currency just out of curiosity.

## LSTM Model 1

According to the results seen in Figure 10, the forecast of the LSTM model seems to be getting out of the random walk predictions overall, however from time to time it fails to beat it. Looking at the RMSE and RMSPEs we can say that adding Google trend signals has improved our prediction's quality and we managed to beat the naive forecasting. However, it is important to note that RMSE othe f the model without adding the Google trends were slightly better but the RMSPE is improved by almost %30.

**Validation RMSE (LSTM Model 1) = 0.0146 , Validation RMSPE = %10.5**

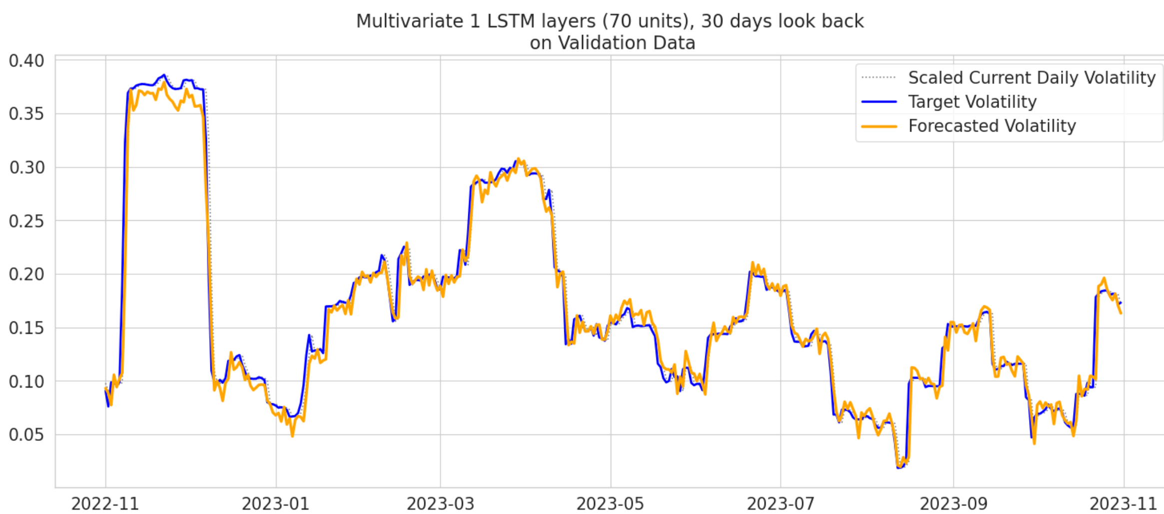


Figure 10: LSTM Model 1 Forecast

## LSTM Model 2

According to the results that can be seen in Figure 11, The bi-directional model fails to improve its prediction with the Google trends signals compared to the model without the Google trends signals and the random walk predictions.

**Validation RMSE (LSTM Model 2) = 0.0186 , Validation RMSPE = %20.9**

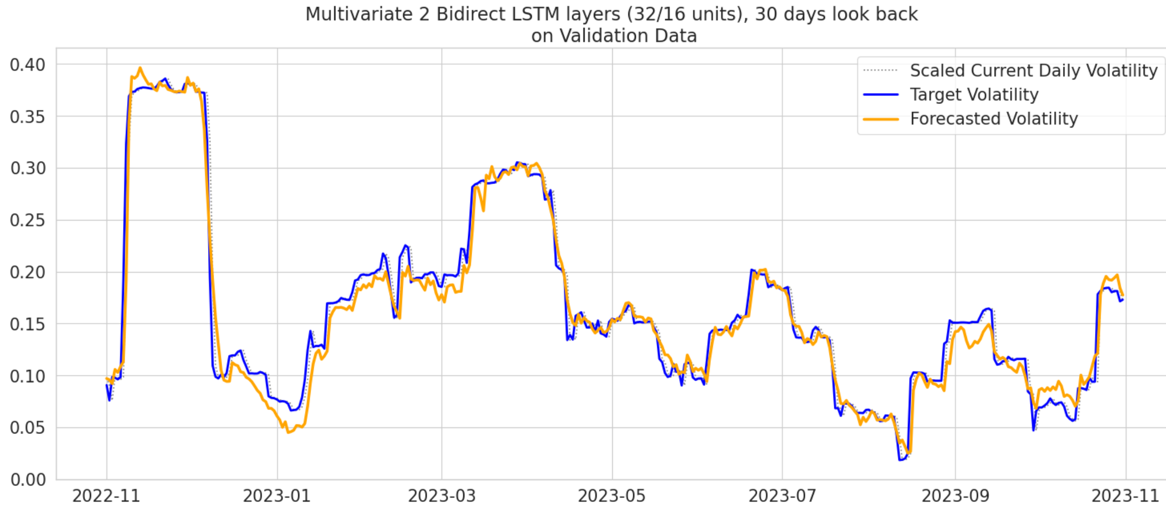


Figure 11: second LSTM forecast

# Conclusion

In this research, we embarked on an ambitious journey to decode the intricate relationship between Google Trends data and Bitcoin market volatility. Through our analysis, employing the LSTM model, we uncovered a significant connection between online search behaviors and market dynamics, highlighting how public interest serves as a crucial predictor of market fluctuations. Our findings revealed that integrating Google Trends signals into the LSTM model notably enhanced predictive accuracy, marking a 22% improvement over the baseline model and a 30% enhancement compared to the LSTM model without this data. This underscores the potential of weaving social signals into traditional financial forecasting methods, thereby offering a more comprehensive perspective of market influencers.

However, the journey doesn't end here. Our exploration, while fruitful, also unveiled areas for growth and further inquiry. The modest changes in RMSE, despite significant RMSPE improvements, point towards the nuanced impact of Google Trends data, signaling the need for refined approaches in discerning positive and negative market influences. Looking ahead, incorporating social sentiment analysis could deepen our understanding of public sentiment's role in shaping market movements. Moreover, broadening our dataset to encompass other cryptocurrencies and integrating additional social media indicators could yield a richer understanding of the decentralized economy. A long-term impact study would also be valuable, providing insights for sustainable investment strategies in the turbulent cryptocurrency market. Finally, our future studies aim to delve into more complex models, including reinforcement learning and the development of a custom trading bot to test the real-world applicability of our findings.

In conclusion, our study makes a significant contribution to the burgeoning field of social data's intersection with financial markets, especially regarding decentralized digital assets. The insights gained offer immense value to investors, policymakers, and academicians, aiding in more informed decision-making in the ever-evolving landscape of cryptocurrency investments. This research not only highlights the importance of continual innovation in predictive modeling but also underscores the dynamic nature of financial markets in the digital era.

# References

[1] F. Parino, M. G. Beiró, and L. Gauvin, "Analysis of the bitcoin blockchain: socio-economic factors behind the adoption," *EPJ Data Science*, vol. 7, no. 38, 2018.

[2] F. Said, R. Somasuntharam, M. Yaakub *et al.*, "Impact of google searches and social media on digital assets' volatility," *Humanities and Social Sciences Communications*, vol. 10, p. 885, 2023.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.