Pontifícia Universidade Católica do Rio de Janeiro

Caio Rodrigues

Forecasting Financial Asset Prices Using Prophet

Monografia de Final de Curso

Orientador: Luiz Scramento

Declaro que o presente trabalho é de minha autoria e que não recorri, para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor.

Rio de Janeiro, Novembro de 2024

Acknowledgements

First and foremost, I would like to express my deepest gratitude to God. Your divine guidance and blessings have been my strength and inspiration throughout this journey. Thank you for providing me with the wisdom, courage, and perseverance to complete this work.

I would also like to extend my heartfelt thanks to my advisor, Luiz Sacramento, for your invaluable guidance, support, and encouragement. Your expertise and insights have been instrumental in shaping this work, and I am deeply grateful for your mentorship.

To my parents, thank you for always believing in me and for providing the love and stability that allowed me to pursue my studies with passion and determination. Your sacrifices and guidance have shaped who I am today, and I am forever grateful for everything you have done.

To my brother, thank you for being a constant source of inspiration, motivation, and intellectual growth. Your insights and encouragement, whether in moments of joy or challenge, have been invaluable, and I am truly fortunate to have you by my side.

This work is as much a reflection of your love, dedication, and inspiration as it is of my own efforts.

Keywords

PROPHET; STOCKS; CRYPTOCURRENCIES; RETURN; FORECAST; DATA

CONTENTS

1 Introduction

The primary question driving this research is how effectively the Prophet algorithm can forecast the returns of financial asset prices, specifically focusing on stocks and cryptocurrencies. Accurate forecasting of financial asset prices is a critical aspect for investors, financial analysts, and policymakers. It enables them to make informed decisions, manage risks, and optimize investment strategies. Given the inherent volatility and complexity of financial markets, developing robust predictive models is essential to navigate these challenges.

The Prophet algorithm, developed by Facebook, is known for its ability to handle time series data with seasonal components and missing values. This makes it a suitable choice for financial data, which often exhibits such characteristics. The main accomplishments of this work include the collection and preparation of a comprehensive dataset, the implementation of the Prophet model, and the evaluation of its performance using established metrics.

The data used in this study comprises adjusted closing prices for stocks and cryptocurrencies from 2022 to April 2024. This data was stored in a PostgreSQL database and included various variables such as company name, date, volume, ticker, adjusted closing prices, and stock returns. For the cryptocurrency segment, additional variables such as the name of the cryptocurrencies, volume, price, and return were included. The dataset was designed to cover a time span that includes various economic cycles and shocks, thereby improving the robustness of the stock predictions. Stocks from major indices of various countries, including the S&P 500 for the United States, and prominent indices from Brazil, Mexico, China, India, England, France, German, and Spain, were selected to ensure high transaction volumes and liquidity. For cryptocurrencies, the study included Bitcoin, Ethereum, Solana, Binancecoin, and Dogecoin.

The Prophet algorithm was chosen for its unique features, such as the ability to model seasonal components and handle missing data. The algorithm decomposes time series data into trend, seasonal, and holiday components, along with an error term. This decomposition allows for a more nuanced understanding of the underlying patterns in the data. The performance of the Prophet model was evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics provide a clear measure of the average prediction error and the average magnitude of the errors, respectively.

The results of this study indicate that the Prophet model demonstrates promising performance in forecasting the returns of financial assets. For stock price returns, the mean observed return was 0.001030081, with an RMSE of 0.00016 and an MAE of 0.012141492. These results suggest that the model's predictions are relatively close to the observed returns, with a high degree of accuracy in capturing the overall trend and behavior of stock price returns. However, the larger MAE indicates that there are frequent smaller errors in the predictions, which may need further refinement of the model.

For cryptocurrency price returns, the mean observed return was 0.00402214, with an RMSE of 0.000948873 and an MAE of 0.030671942. These results suggest that while the Prophet model does a reasonable job of predicting cryptocurrency returns, it is not perfectly accurate in capturing the magnitude of the fluctuations. The higher RMSE and MAE values for cryptocurrencies reflect the greater market volatility inherent to this asset class.

The findings of this study contribute to the broader understanding of financial asset price forecasting by highlighting the efficacy of the Prophet algorithm. The comparative analysis with other predictive models provides insights into the relative performance of the Prophet model, informing future research and model selection. The practical implications of this study are significant for investors and financial analysts, offering guidance on model selection and application for forecasting financial asset prices.

This monograph is structured as follows: The data chapter describes the data sources, variables used, and provides descriptive statistics of the dataset. The methodology chapter details the Prophet strategy, including the modeling of trend, seasonal, and holiday components, and discusses potential problems encountered during the study. The results chapter presents the findings of the stock and cryptocurrency price return predictions, including a comparative analysis of the model's performance. Finally, the conclusion chapter discusses the conclusions drawn from the study, highlights limitations, and suggests directions for future research.

In summary, this study demonstrates that the Prophet model is a valuable tool for forecasting financial asset prices, particularly in capturing trends and minimizing large errors. However, the model's performance could be further improved with a larger dataset and additional refinements. Future work could involve expanding the dataset to improve the model's robustness and prediction accuracy, as well as conducting a benchmark comparison with other predictive models to provide valuable insights into the relative strengths and weaknesses of the Prophet algorithm in forecasting volatile assets like cryptocurrencies.

2 Data

For this study, we utilized the adjusted closing prices of financial assets, including stocks, from 2022 until April, 2024 and cryptocurrencies, from 2022 until April, 2024. The database was stored in PostgreSQL, and included variables such as company name, date, volume, ticker, adjusted closing prices, stock return, and other descriptive attributes. For the cryptocurrency segment, the data was also stored in PostgreSQL but included different variables such as the name of the cryptocurrencies, date, volume, price, and return. In this study, we aim to train the model using a time span that includes various economic cycles and shocks to improve stock predictions.

We selected stocks from the major indices of various countries to ensure high transaction volumes and liquidity. For stocks from the United States, we used the S&P 500 index. Similarly, we chose prominent indices for other countries, including Brazil, Mexico, China, India, England, France, and Spain. Regarding cryptocurrencies, we included Bitcoin, Ethereum, Solana, Binancecoin and Dogecoin.

2.0.1 Data sources

All the data and variables of stocks and cryptocurrencies, that include volume, ticker name, open price, closing price, high, low, close, adjusted close, volume of each financial asset in each day and the variable country that indicates the market where the financial assets are, all this variables was obtained from the Yahoo Finance API and integrated into PostgreSQL servers. With that i was able to create other column to calculate the adjusted price return that is the variable that we want to project and estimate. Important to remember what is the variable Adjusted Price Return

$$\text{Adjusted Price Return} = \frac{\text{Adjusted close price}_t - \text{Adjusted close price}_{t_0}}{\text{Adjusted close price}_{t_0}}$$

This integration enabled us to create a comprehensive database containing essential details such as ticker prices, high and low values, adjusted closing prices, and adjusted price returns, as well as descriptive statistics tables and graphics used in our analysis.

2.0.2 Data descriptive

Descriptive statistics are essential tools in any data analysis, providing simple summaries about the sample and the measures. These statistics help to understand the basic features of the data and form the foundation for further statistical analysis. Descriptive statistics summarize and organize characteristics of a data set, making it easier to interpret and communicate the information.

In this section, we will present the descriptive statistics for our data set, focusing on the variable price return. We will provide measures of central tendency such as the mean and median, which give us an idea of the typical value within our data. Additionally, we will discuss measures of dispersion, including the standard deviation, range, maximum, and minimum, which describe the spread and variability of the data points. These statistics will give us a comprehensive overview of our data, highlighting key patterns and outliers that may be present.

Statistic	Mean	Median	Standard Deviation	Maximum	Minimum
Return of Stock Price	0.000346	0.0	0.020579	1.005783	-0.666158

Table 1 – Descriptive statistics for the return of stock prices.

The mean return is 0.000346, indicating that on average, the stock returns are slightly positive but very close to zero, suggesting a stable overall performance with minor gains over time. The median return is 0.0, meaning that half of the returns are below zero and half are above, reflecting a balance between gains and losses.

The standard deviation of 0.020579 shows a moderate level of variability around the mean, indicating that while the average returns are close to zero, there is significant fluctuation in individual returns. The maximum return observed is 1.005783, highlighting that there was a period of extremely high returns, likely due to a significant positive event affecting the stock market or a particular stock. Conversely, the minimum return is -0.666158, pointing to a period of substantial negative returns, possibly caused by a major adverse event.

Statistic	Mean	Median	Standard Deviation	Maximum	Minimum
Return of crypto	0.001205	-0.000141	0.041432	0.449432	-0.422809

Table 2 – Descriptive statistics for the return of cryptocurrencies prices.

The mean return of 0.001205 suggests that, on average, investing in cryptocurrencies yields a small positive return. However, the negative median of -0.000141 indicates that more than half of the individual returns are slightly negative, which could imply a skew in the distribution of returns due to a few large positive outliers. The high standard deviation of 0.041432 reflects considerable volatility in cryptocurrency returns, meaning that while the average return is positive, individual returns can vary widely from this average. The maximum and minimum returns (0.449432 and -0.422809, respectively) demonstrate the extreme fluctuations in cryptocurrency prices, underscoring the potential for both significant gains and losses in this market. In this study, we employed several statistical measures to evaluate the prediction results of the Prophet and ARIMA models. The chosen metrics include RMSE (Root Mean Square Error), MAE (Mean Absolute Error).

RMSE (Root Mean Square Error): RMSE was selected because it provides a clear measure of the average prediction error, allowing us to assess how closely the model's forecasts match the actual values. It emphasizes larger errors due to its squaring nature, making it particularly useful for identifying significant deviations.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

MAE (Mean Absolute Error): The Mean Absolute Error (MAE) is a measure of prediction accuracy commonly used in regression and time series forecasting. It quantifies the average magnitude of the absolute errors between predicted values and observed values, without considering the direction of the errors. Essentially, MAE provides a straightforward interpretation of the model's average error.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

which the y variable is the stock price return, and "n" is the number of observations.

By using these metrics, we aimed to provide a comprehensive assessment of the predictive performance of both models, allowing for a clearer understanding of their strengths and weaknesses in forecasting.

2.0.3 Analysis

By analyzing these graphics, we can identify trends, compare performance across different cryptocurrencies, and gain insights into market fluctuations over time. This visual representation not only enhances our understanding of the data but also facilitates effective communication of key findings.

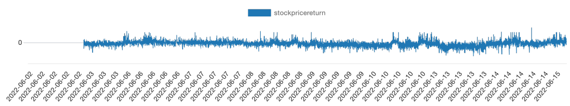

Figure 1 – Stocks price return

Observing the first graphic depicting the price returns of stocks, we notice that the mean is very close to zero, consistent with the descriptive statistics table. There are periods where returns were more negative and others where they were more positive. Comparing this graphic with that of cryptocurrencies, it makes sense that this other financial asset carries more risk and exhibits greater volatility, indicated by a higher standard deviation. It's interesting to observe all of this by comparing the graphics of stocks and cryptocurrencies, which aligns with what we observe in the descriptive statistics table.

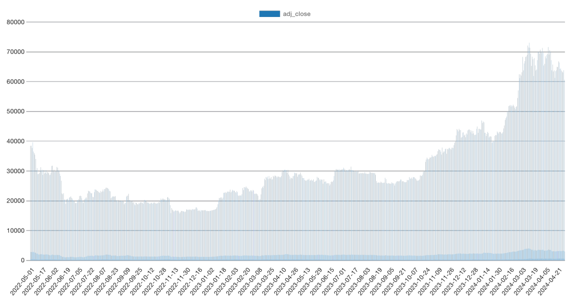

Figure 2 – Cryptocurrencies close price

Here we present a bar chart depicting the closing prices of cryptocurrencies from our database, which includes prominent cryptocurrencies such as Bitcoin, Solana, Ethereum, Dogecoin, and Binancecoin. The main objective of including these cryptocurrencies in our research is to explore their diverse attributes and characteristics. Cryptocurrencies like Ethereum and Bitcoin are known for their high transaction volumes, while others have lower transaction volumes and are considered more speculative. Our goal is to analyze how different algorithms forecast these cryptocurrencies, taking into account their varying levels of volatility, in order to refine each model and conduct comparative analyses. This trend component suggests that midst the short-term fluctuations, there is a discernible pattern or direction in the close prices of the cryptocurrencies over time. Such patterns can provide insights into long-term market movements and investor sentiment, potentially indicating underlying factors driving the cryptocurrency prices beyond immediate volatility. Analyzing these trends can be crucial for understanding market behavior and making informed investment decisions. It is important to note that recognizing these patterns is essential for training the Prophet algorithm to forecast future price movements accurately.

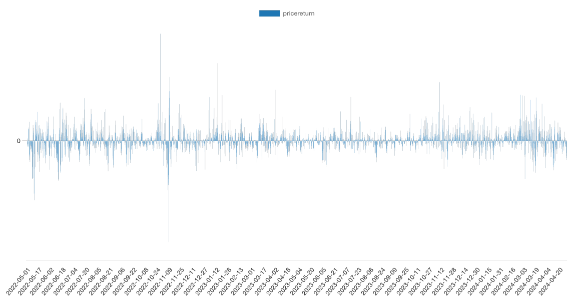

Figure 3 – Cryptocurrencies close price

Observing this last graphic, not only in the period from October 24, 2022, to November 25, 2022, but in the 2022 in general the price returns of cryptocurrencies exhibited notable fluctuations, marking the global high and low within this timeframe. This period stands out as a time of increased volatility, where the market experienced rapid changes in cryptocurrency prices.

3 Methodology

3.0.1 Prophet strategy

We focus on forecasting the returns of financial assets with Prophet and observes the efficiency of the algorithm using RSME and other statistics parameters by evaluating Prophet's performance, we aim to identify its strengths and limitations in forecasting the prices of various financial assets. The FB-Prophet algorithm is a time series forecasting that suits better when are missing data in the time series, and when the time series have a seasonal component. This algorithm can be represented by that equation:

$$Y(t) = g(t) + s(t) + h(t) + e_t \tag{3.1}$$

Where:

- $y(t)$ represents the observed value of the time series at time t, in our scenario, this refers to the return value of the financial assets.

- $g(t)$ represents the trend component, which captures the overall direction of the time series data.

- $s(t)$ represents the seasonal component, which captures periodic patterns or seasonal effects.

- $h(t)$ represents the effects of holidays or other special events.

- e_t represents the error term, which accounts for any noise or randomness in the data.

The trend component g(t) is often modeled using a piece wise linear or logistic function. The seasonal component s(t) is typically modeled using Fourier series to capture periodic fluctuations. The effects of holidays or other special events h(t) are represented using indicator variables.

Overall, the FB Prophet model aims to decompose the time series data into these components and then forecast future values by extrapolating the trend, seasonal, and holiday effects. The parameters of the model are estimated using a Bayesian approach, which allows for uncertainty estimation in the forecasts. The efficiency of forecasting with Prophet is influenced by two main factors: the algorithm's training and the availability of up-to-date data.

When i was fitting the model was choose training with data to 2022 until April 2024 for stock and cryptocurrencies prediction, putting the components of the model, we assume daily, weekly and year seasonality as true, and holiday effects as false.

3.0.2 Potential Problems

The main issue is that I could likely achieve more accurate predictions if the dataset covered a longer time horizon, which would have probably led to better results. However, since i used more than 560,000 observations, From a machine learning and time series perspective, a dataset of this magnitude also helps to reduce the risk of over-fitting, as the model generalizes better across a diverse set of observations. In financial modeling, particularly for stock predictions, this large dataset can capture rare events and fluctuations in market behavior. I tried using data from additional years, but the hardware I have available couldn't handle such a large dataset.

Another issue I encountered was the closure of financial markets on certain days in different countries. In my initial prediction, the adjusted closing prices predicted by the algorithm were inaccurate because on days when the market is closed, the price remains the same as the previous day's. I resolved this by filtering out holidays and weekends, ensuring that only valid trading days were included in the prediction.

4 Results

Here, I present the results of our stock price return predictions using the Prophet algorithm. The model was trained on data spanning from June 2, 2024, to December 31, 2023, to forecast returns for the period from January 1 to April 30, 2024. The table below shows statistics comparing the observed stock price returns with the predicted values for this forecast period in 2024.

I used Root Mean square error(RMSE) and Mean absolute error(MAE) primarily because the article (AKINSOMI et al., 2016) in the references suggests that RMSE is a recommended statistic for evaluating predictions of stock prices and stock returns. This is because the errors in these variables are typically normally distributed, and RMSE performs well in such cases. I employed the MAE to assess the magnitude of errors without considering their direction, as it does not account for outliers. In contrast, the RMSE incorporates a penalty for outliers and larger errors.

Table 3 – Stock Price Return

Metric	Observed Mean	RMSE (Predictions)	MAE (Predictions)
Stock Price Returns	0.001030081	0.00016	0.012141492

The results from the Prophet model's prediction of stock price returns reveal several important insights:

Mean Observed Return : The mean observed return of 0.001030081 indicates a positive average return over the period analyzed. This suggests that, on average, the stock prices have shown a slight upward trend. However, the magnitude of this return is relatively small, which could indicate low volatility or modest market movements during the given time frame.

RMSE : RMSE of 0.00016 is quite low, indicating that the model's predictions are relatively close to the observed returns. RMSE is particularly sensitive to large errors, so this value suggests that the model does not make large predictive errors. The low RMSE implies that the model captures the overall trend and behavior of stock price returns with a high degree of accuracy.

MAE : MAE of 0.012141492 is considerably larger than the RMSE, which may initially seem unusual, as typically the RMSE tends to be larger than the MAE. This discrepancy can be attributed to the scale of the errors and the fact that MAE is less sensitive to the magnitude of errors. The MAE value suggests that, on average, the model's predictions deviate from the actual returns by approximately 0.012 units. While this is a relatively large error compared to the RMSE, it may indicate that there are smaller, more frequent prediction deviations that do not heavily impact the overall RMSE due to the square root transformation.

The results suggest that the Prophet model performs relatively well in forecasting stock price returns, particularly in terms of minimizing large errors (as evidenced by the low RMSE). However, the larger MAE indicates that there are frequent smaller errors in the predictions that may need further refinement of the model or is possible that with a larger data base we could have better results.

Using the cryptocurrency dataset, I assessed the model's ability to predict the price returns of Cryptocurrencies over the same period as the stocks.

Table 4 – Cryptocurrencies Price Return

Metric	Observed Mean	RMSE (Predictions)	MAE (Predictions)
Cryptocurrencies Price Returns	0.00402214	0.000948873	0.030671942

The mean observed return of 0.00402214 suggests that, on average, the cryptocurrency prices have shown a positive return during the observed period. This return is notably higher than the observed stock price return from the previous analysis (0.001030081), which indicates that, on average, the cryptocurrency market has been more volatile or dynamic than the stock market during this period.

RMSE of 0.000948873 is relatively low but higher than the RMSE value from the stock price prediction (0.00016). This indicates that the Prophet model's predictions for cryptocurrency returns are fairly close to the observed values, but there is a higher level of variability in its predictions when compared to stock returns. RMSE being higher for cryptocurrencies could be a result of greater market volatility inherent to the cryptocurrency market, where price fluctuations tend to be more pronounced.

MAE of 0.030671942 is considerably larger than both the RMSE and the mean observed return. This suggests that the model's predictions are, on average, deviating significantly from the actual observed returns. While the RMSE indicates that the model avoids large errors to some extent, the MAE value reflects that the model consistently makes moderate-to-large errors in its predictions. The larger MAE could be indicative of the high volatility in the cryptocurrency market, where even small price movements can result in larger prediction errors, especially when compared to the smaller, more stable stock price returns.

The results suggest that while the Prophet model does a reasonable job of predicting cryptocurrency returns (as evidenced by the low RMSE), it is not perfectly accurate in capturing the magnitude of the fluctuations. The higher MAE indicates that the model's predictions tend to deviate more substantially from actual observed returns, likely due to the inherent volatility in the cryptocurrency market.

5 Conclusion

This study set out to evaluate the effectiveness of the Prophet algorithm in forecasting the returns of financial asset prices, specifically focusing on stocks and cryptocurrencies. The research involved collecting and preparing a comprehensive dataset of adjusted closing prices from 2022 to April 2024, implementing the Prophet model, and assessing its performance using metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The study also included a comparative analysis to highlight the strengths and limitations of the Prophet model.

The results demonstrated that the Prophet model is capable of capturing the overall trends in financial asset prices with a high degree of accuracy, as evidenced by the low RMSE values. For stock price returns, the model showed a mean observed return of 0.001030081, with an RMSE of 0.00016 and an MAE of 0.012141492. For cryptocurrency price returns, the mean observed return was 0.00402214, with an RMSE of 0.000948873 and an MAE of 0.030671942. These findings suggest that while the Prophet model performs well in minimizing large errors, it tends to have frequent smaller errors, particularly in the more volatile cryptocurrency market.

The implications of these results for the general question raised in the introduction are significant. The study confirms that the Prophet algorithm is a valuable tool for forecasting financial asset prices, particularly in markets with seasonal components and missing data. The ability of the Prophet model to handle such complexities makes it a robust choice for financial forecasting. However, the higher MAE values indicate that there is room for improvement, especially in capturing the magnitude of fluctuations in more volatile markets like cryptocurrencies. This suggests that while the Prophet model is effective in identifying trends, additional refinements are needed to enhance its predictive accuracy.

Several open questions and possible extensions arise from this study. One key question is how the Prophet model compares to other advanced predictive models, such as those based on machine learning and deep learning techniques. Future research could involve a benchmark comparison to provide a more comprehensive understanding of the relative strengths and weaknesses of different models. Additionally, expanding the dataset to include a longer time horizon and more diverse financial assets could improve the robustness and accuracy of the predictions. Another potential extension is to explore the integration of external factors, such as macroeconomic indicators and market sentiment, into the Prophet model to enhance its forecasting capabilities.

In conclusion, this study demonstrates that the Prophet model is a promising tool for forecasting financial asset prices, particularly in capturing trends and minimizing large errors. However, the model's performance could be further improved with additional data and refinements. Future research should focus on comparing the Prophet model with other predictive models, expanding the dataset, and integrating external factors to enhance the model's accuracy and reliability. These extensions would not only improve the predictive performance of the Prophet model but also contribute to a deeper understanding of financial market dynamics and the factors driving asset price movements

Bibliography

AKINSOMI, O. et al. Real estate returns predictability revisited: novel evidence from the us reits market. *Empirical economics*, Springer Berlin Heidelberg, Berlin/Heidelberg, v. 51, n. 3, p. 1165–1190, 2016. ISSN 0377-7332.

CHENG, J. et al. Forecasting bitcoin prices using artificial intelligence: Combination of ml, sarima, and facebook prophet models. *Technological forecasting social change*, v. 198, p. 122938, 2024. ISSN 0040-1625.

GHOSH, I.; CHAUDHURI, T. D. Integrating navier-stokes equation and neoteric iforest-borutashap-facebook's prophet framework for stock market prediction: An application in indian context. *Expert systems with applications*, Elsevier Ltd, v. 210, p. 118391, 2022. ISSN 0957-4174.

LOLEA, I.-C.; PETRARIU, I.-R.; GIURGIU, A. Arima vs. machine learning in terms of equity market forecasting. *Analele Universităţii din Oradea. Ştiinţe economice*, University of Oradea, v. 30, n. 2, p. 299–308, 2021. ISSN 1222-569X.

MALLIKARJUNA, M.; RAO, R. P. Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*, v. 5, n. 1, p. 40, 2019. Disponível em: ⟨https://doi.org/10.1186/s40854-019-0157-x⟩.

SONKAVDE, G. et al. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, v. 11, n. 3, 2023. ISSN 2227-7072. Disponível em: ⟨https://www.mdpi.com/2227-7072/11/3/94⟩.

(CHENG et al., 2024) (GHOSH; CHAUDHURI, 2022) (SONKAVDE et al., 2023) (LOLEA et al., 2021) (AKINSOMI et al., 2016) (MALLIKARJUNA; RAO, 2019)