Cryptocurrency exchanges and stock exchanges forecasting

Josep Serra Mas, Mar Bulló Porta, Llorenç Burgas Nadal , Joaquim Massana Raurich

September 7, 2024

Abstract

Accurate stock market and cryptocurrencies prediction is challenging due to the volatility driven by factors like news, and other external influences. Traditional statistical and econometric models struggle with non-stationary time series data, showing the need for more robust prediction methods. Machine learning algorithms offer a promising approach by utilizing information from financial news, search interests, and related data (e.g., gold and gas prices) to enhance prediction accuracy. This paper explores the impact of these external factors on stock market prediction, focusing on improving performance through feature selection and analyzing the influence of search interest on investor behavior and prices.

1 Introduction

The financial markets, encompassing both traditional stocks and emerging cryptocurrencies, have always been characterized by volatility and complexity. Predicting price movements within these markets is a crucial task for investors, as even small fluctuations can lead to significant gains or losses. Historically, various methods have been employed to forecast market trends, ranging from fundamental and technical analysis to more sophisticated statistical and econometric models. However, these traditional approaches often fall short when faced with the dynamic nature of financial data.

The advent of artificial intelligence (AI) and machine learning has opened new avenues for financial prediction, offering the potential to process and analyze vast amounts of data. Unlike traditional models, AI-based approaches can adapt to changing market conditions and identify complex patterns in data that are not immediately apparent. This makes AI particularly well-suited for predicting price movements in both stocks and cryptocurrencies, where external factors such as news events, social media sentiment, and global economic indicators play a significant role.

Cryptocurrencies, in particular, present unique challenges and opportunities for AI-based prediction. Unlike traditional assets, cryptocurrencies operate in a relatively unregulated and rapidly evolving market, with prices often driven by speculative behavior, technological developments, and geopolitical events. The decentralized nature of cryptocurrencies, combined with their 24/7 trading environment, further complicates prediction efforts, as traditional market closure times do not apply. This continuous trading cycle results in a highly responsive market that can react instantly to new information, making accurate and timely predictions even more critical.

In this paper, we propose a AI-driven approach to predict price changes in both stocks and cryptocurrencies. By using machine learning algorithms, we aim to integrate various data sources, including financial news, social media trends, search interests, and macroeconomic indicators. Our approach seeks to enhance prediction accuracy but also to provide a comprehensive understanding of the factors influencing market behavior.

2 Related work

The prediction of financial market movements, including both stock and cryptocurrency prices, has been a focal point of research for many decades. Early efforts to forecast these markets primarily relied on traditional statistical and econometric models, which, despite their usefulness, often struggled with the volatile and non-stationary nature of financial data.

2.1 Machine Learning Approaches

In recent years, the advent of machine learning has significantly advanced the field of financial market prediction. Researchers have explored a variety of machine learning techniques to predict stock price movements, with many studies focusing on the integration of social media data. For instance, opinion analysis of Twitter feeds has been utilized to gauge public sentiment and predict stock market trends. **Bollen et al. (2011)** were among the pioneers in this area, using sentiment analysis from News to predict stock market movements .

2.2 Neural Networks and Sentiment Analysis

Building on this foundation, other researchers have employed neural networks in conjunction with social media data to enhance prediction accuracy. **Nguyen et al. (2015)** used neural networks alongside daily Twitter feeds to forecast the daily up and down changes in the closing values of the Dow Jones Industrial Average, demonstrating the potential of combining these data sources . Furthermore, **Jiang et al. (2018)** found that Long Short-Term Memory (LSTM) networks, when combined with Twitter sentiment analysis, outperformed other machine learning models, such as Support Vector Machines (SVM), in predicting stock prices .

2.3 Sentiment Analysis in Diverse Applications

The application of sentiment analysis is not limited to financial markets. It has also been successfully employed in various other domains, such as predicting movie revenues and analyzing public sentiment towards U.S. presidential candidates during the 2012 election . For English sentiment analysis, the Valence Aware Dictionary and sentiment Reasoner (VADER) has been widely used to classify sentiment in tweets .

In this paper, we build on these advancements by developing a comprehensive AI-driven approach to predicting both stock and cryptocurrency price movements. Our model integrates various external data sources, including financial news, social media trends, and macroeconomic indicators, with sentiment analysis techniques to enhance prediction accuracy.

3 Methodology

This section outlines the steps involved in our proposed framework for predicting stock and cryptocurrency price movements. The methodology encompasses data collection, preprocessing, feature extraction, and the application of machine learning models to make accurate predictions.

3.1 Data collection

Data collection is a critical step in our framework, as the data directly influence the prediction accuracy. In this subsection, we describe the sources of our collected data and the structure of the datasets used in our analysis.

3.1.1 Yahoo Finance stock data

To obtain historical price data for stocks, we rely on Yahoo Finance. The stock price data is collected for selected stock markets over a specific time period, and the data is downloaded in `.csv` format. Each file contains seven key features:

- **Date:** The specific date of the recorded data.
- **Open:** The opening price of the stock on that day.
- **High:** The highest price reached by the stock during the trading day.
- **Low:** The lowest price reached by the stock during the trading day.
- **Close:** The closing price of the stock at the end of the trading day.

- **Volume:** The total number of shares traded during the day.
- **Adjusted Close:** The closing price adjusted for dividends, stock splits, and other corporate actions.

3.1.2 Price of raw materials

We collect data on the prices of valuable raw materials, such as gold and natural gas, as these commodities can significantly influence financial markets. The data is sourced from Investing.com, which provides both live feeds of current pricing and extensive historical data spanning several years.

3.1.3 Financial news data

In addition to historical price data, we collect financial news to capture external factors that may influence stock prices. For this purpose, we use the `NewsAPI`, a tool for accessing a wide range of news sources and articles in real-time. Financial news articles are gathered based on keywords related to the selected stocks and cryptocurrencies within the same time frame as the historical price data.

3.1.4 Google trends data

To further enhance our prediction model, we incorporate search interest data from Google Trends. Google Trends provides insights into the relative search popularity of specific keywords over time, reflecting public interest and sentiment toward certain topics. We collect data on search trends for keywords related to the selected stocks and cryptocurrencies.

3.2 Preprocessing

3.2.1 Yahoo Finance stock data

The historical stock data collected from Yahoo Finance is typically well-structured, but it may still contain missing or erroneous entries due to market holidays or data retrieval issues. We handle missing values by applying backward-fill.

3.2.2 Price of raw materials

Although the raw material price dataset is generally reliable, it is checked for any missing entries. When missing values are detected, they are addressed using backward-fill to maintain the continuity of the time series data.

3.2.3 Financial news data

For the financial news articles collected via `NewsAPI`, we remove any duplicate articles and irrelevant content that does not contribute to the prediction task.

3.3 Sentiment analysis

Sentiment analysis in our framework is conducted using the Stanford Sentiment Analysis package from Stanford NLP (Socher et al., 2013), which goes beyond traditional methods by analyzing the full structure of sentences rather than just individual words in isolation. This advanced tool leverages Recursive Neural Tensor Networks (RNTNs) to understand the relationships between words, capturing complex linguistic features like negation and sarcasm. The sentences are then parsed into syntactic trees, allowing the tool to assign sentiment scores that reflect the emotional tone of the content. These scores, ranging from strongly negative to strongly positive, are aggregated daily for each stock or cryptocurrency and incorporated as features in our prediction models.

4 Implementation

When we had the data ready, we proceeded to build the LSTM model. We began by splitting the data into training and test sets, ensuring that the model had enough data to learn while also retaining sufficient data for independent evaluation. The LSTM model we developed included several layers, along with Dropout layers to prevent overfitting, a common issue in complex models. After training the model, we evaluated its performance using tools such as the confusion matrix, classification reports, and accuracy metrics to ensure that the model was learning effectively.

5 Evaluation

When we tested our model for the first time, it resulted in an accuracy of 50%, so we then focused on improving it. We experimented with different neural network architectures, such as LSTM with and without Dropout, and other models like SVM, Random Forest, BPNN, GRU, and RNN. However, none of these significantly improved the accuracy of our initial model. We also explored the possibility of adding new variables, retaining those that showed good correlation with Bitcoin's closing price and eliminating others that did not contribute to improvements. As a final measure, we conducted several tests and expanded our dataset to cover three months, one year, and five years of data. We discovered that increasing the dataset to a longer period was the only way to improve accuracy. This decision meant we had to drop the sentiment analysis variable from news articles, as it limited our time window to one month of data. Ultimately, we decided to use five years of data, which allowed us to achieve an accuracy of 88%

Model	Accuracy 3 months	Accuracy 5 Years
LSTM	0.6	0.88
SVM	0.27	0.44
RF	0.5	0.66
BPNN	0.36	0.66
GRU	0.5	0.55
RNN	0.5	0.55

6 Conclusion and future work

In conclusion, we believe that one of the major obstacles to conducting this type of research is data collection, as long periods are required to gather information, and there are often significant time constraints (for example, some locations only allow data collection during the last month). Despite these challenges, we believe we have conducted good research, achieving an accuracy that ranges between 0.66 and 0.88. In the future, we would like to reduce the variability in accuracy and increase it, as while a result of 0.88 is very good, a result of 0.66 is less satisfactory.

To summarize, this research has demonstrated that AI-based models can offer significant improvements in financial market forecasting, although there is still room for enhancement. The combination of machine learning techniques with external data sources shows considerable potential, but it is essential to continue refining the models and addressing current limitations. We are confident that with further research and development, it will be possible to achieve even more accurate and reliable results in the future.

References

Sean McNally, Jason Roche, Simon Caton . (2018). Predicting the Price of Bitcoin Using Machine Learning. Google Scholar . https://ieeexplore.ieee.org/abstract/document/8374483

Sean Reaz Chowdhury , M. Arifur Rahman, M. Sohel Rahman , M.R.C. Mahdy . (2020). An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. Google Scholar . https://www.sciencedirect.com/science/article/pii/S0378437120302703

Zheshi Chen Chunhong Li Wenjun Sun,. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering, Journal of Computational and Applied Mathematics. Google Scholar . https://www.sciencedirect.com/science/article/pii/S037704271930398X

Akyildirim, E., Goncu, A. Sensoy, A. Prediction of cryptocurrency returns using machine learning. Ann Oper Res 297, 3–36 (2021). https://doi.org/10.1007/s10479-020-03575-y

A. Radityo, Q. Munajat and I. Budi, "Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, Indonesia, 2017, pp. 433-438, doi: 10.1109/ICACSIS.2017.8355070. keywords: Bitcoin;Artificial neural networks;Training;Investment;Genetic algorithms;Backpropagation;c neural network (ANN),

Huang, X. et al. (2021). LSTM Based Sentiment Analysis for Cryptocurrency Prediction. In: Jensen, C.S., et al. Database Systems for Advanced Applications. DASFAA 2021. Lecture Notes in Computer Science(), vol 12683. Springer, Cham. https://doi.org/10.1007/978-3-030-73200-$4_4$7

Phumudzo Lloyd Seabe Claude Rodrigue Bambe Moutsinga Edson Pindza. (2023). Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. Google Scholar . https://www.mdpi.com/2504-3110/7/2/203

M. Rafi, Q. A. K. Mirza, M. I. Sohail, M. Aliasghar, A. Aziz and S. Hameed, "Enhancing Cryptocurrency Price Forecasting Accuracy: A Feature Selection and Weighting Approach With Bi-Directional LSTM and Trend-Preserving Model Bias Correction," in IEEE Access, vol. 11, pp. 65700-65710, 2023, doi: 10.1109/ACCESS.2023.3287888. keywords: Predictive models;Bitcoin;Forecasting;Data models;Long short term memory;Machine learning;Blockchains;Blockchain;cryptocurrency;machine learning,

Hansun, S., Wicaksana, A. Khaliq, A.Q.M. Multivariate cryptocurrency prediction: comparative analysis of three recurrent neural networks approaches. J Big Data 9, 50 (2022). https://doi.org/10.1186/s40537-022-00601-7

Khan, W., Ghazanfar, M.A., Azam, M.A. et al. Stock market prediction using machine learning classifiers and social media, news. J Ambient Intell Human Comput 13, 3433–3456 (2022). https://doi.org/10.1007/s12652-020-01839-w

M. Nabipour P. Nayyeri H. JabaniA. Mosavi E. Salwana Shahab S. (2020). Deep Learning for Stock Market Prediction. Google Scholar . https://doi.org/10.3390/e22080840

X. Ji, J. Wang and Z. Yan, "A stock price prediction method based on deep learning technology," in International Journal of Crowd Science, vol. 5, no. 1, pp. 55-72, April 2021, doi: 10.1108/IJCS-05-2020-0012. keywords: Deep learning;Wavelet transforms;Social networking (online);Biological system modeling;Time series analysis;Predictive models;Feature extraction;Text mining;Deep learning;Financial social media;Stock price prediction,

Li, S., Liao, W., Chen, Y., Yan, R. (2023). PEN: Prediction-Explanation Network to Forecast Stock Price Movement with Better Explainability. Proceedings of the AAAI Conference on Artificial Intelligence, 37(4), 5187-5194. https://doi.org/10.1609/aaai.v37i4.25648

Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar,. (2020). Stock Closing Price Prediction using Machine Learning Techniques,. Google Scholar . https://doi.org/10.1016/j.procs.2020.03.326