Comparative Analysis of Different Algorithms in Prediction of Stock Prices

Ujjwal Chouhan
Amity Institute of Information Technology
Amity University Noida, Uttar Pradesh
ujjwal.chouhan@s.amity.edu

***ABSTRACT* - In order to improve the accuracy and resilience of stock price predictions, this study investigates the synergistic integration of modern forecasting approaches, including the Kalman filter, XG Boost algorithm, and linear regression. Understanding the inherent difficulties in simulating the dynamic and erratic nature of financial markets, the goal of this research is to provide a comprehensive framework that makes use of the advantages of each algorithmic technique.**

Known for its efficiency when processing time-series data, the Kalman filter is used to dynamically modify its forecasts in response to real-time data, allowing it to capture and adjust to changing market conditions. This is enhanced by the ability of the XG Boost algorithm, a potent machine learning method, to identify intricate patterns and non-linear correlations in the financial data.

In addition, the interpretability and simplicity of linear regression are used to provide a baseline model for comparison. Our research shows the combined effectiveness of these approaches in reducing the inherent uncertainties of financial markets through a thorough analysis of past stock data. Combining the Kalman filter, XGBoost, and linear regression improves the model's ability to adjust to a variety of market conditions while also improving prediction accuracy.

The empirical findings highlight the possibility of developing a stock price prediction model that is more trustworthy, opening the door to better financial market decision-making. This study offers a sophisticated solution to the problems associated with stock price forecasting, which adds significant insights to the field's growing state.

Keywords – Stock prediction, Kalman filter, XG Boost, Linear Regression

I. INTRODUCTION

As long as financial markets show complex and dynamic behaviors, it will be difficult to predict stock prices with any degree of accuracy. By combining three different approaches—linear regression, XGBoost, and the Kalman filter—this study aims to pave a new route and improve the accuracy and flexibility of stock price predictions. Understanding the shortcomings of distinct methods in encapsulating the intricate and constantly evolving aspects of market dynamics, our research aims to leverage the synergistic advantages of these algorithms.

Known for its effectiveness in time-series analysis, the Kalman filter provides a dynamic framework that can be adjusted in real-time to reflect changing market conditions. [2] enhanced by the XGBoost algorithm's machine learning skills, which are skilled at identifying complex patterns and non-linear relationships, and the simplicity of linear regression, our integrated model aspires to provide a comprehensive solution. This paper aims to contribute a nuanced perspective to predictive analytics, offering a holistic approach that addresses the multifaceted challenges inherent in forecasting stock prices in today's unpredictable financial landscape.

II. LITERATURE REVIEW

Stock price predicting algorithms

1. Kalman Filter

A recursive mathematical procedure called the Kalman Filter [1], created by Rudolf E. Kalman, is used to filter and estimate the state of a linear dynamic system. For reliable real-time estimations, it effectively blends noisy sensor measurements and forecasts. The Kalman Filter, which is widely used in signal processing, navigation, and control systems, reduces errors by continuously updating predictions based on incoming data. Because of its adaptability and efficiency, it is vital in fields like robotics and autonomous cars, guaranteeing accurate estimation even in the face of ambiguity and measurement noise.

2. XG Boost

Extreme Gradient Boosting, or XG Boost, is a potent and effective machine learning technique that excels at both regression and classification problems. XG Boost, an ensemble learning technique created by Tianqi Chen, generates a reliable and accurate model by combining the predictions of several weak learners (usually decision trees). Large dataset handling, feature selection, and regularization techniques are its main advantages, which make it a popular choice in data science competitions.[9] The efficient use of XG Boost increases speed and scalability, which has led to its broad acceptance in a variety of industries, including marketing, banking, and healthcare.

3. Linear Regression

A basic statistical modeling method for predicting a dependent variable based on one or more independent variables is called linear regression. It creates a straightforward but understandable model by establishing a linear relationship between the input elements and the goal variable. The goal of linear regression, which was developed from a statistical framework, is to identify the best-fit line that minimizes the sum of squared errors. Linear regression is a widely used baseline modeling technique in various domains, including biology, economics, and social sciences.[8] It is a mainstay of beginning machine learning and statistical analysis because of its simplicity, interpretability, and application.

4. Theoretical Comparison between three algorithms

Selecting the appropriate predictive model is essential when it comes to stock price prediction, particularly for erratic commodities like gold versus USD. Let's examine XGBoost, Linear Regression, and the Kalman Filter—three widely used techniques.

The Kalman Filter is a powerful tool for managing noisy data and is especially helpful when it's necessary to change forecasts on a regular basis in response to fresh data. It is useful for forecasting the varying prices of gold relative to the US dollar since it performs well in situations when there is a great deal of ambiguity. In contrast, XGBoost is an effective ensemble learning method that integrates the predictions made by several decision trees. It is renowned for its great accuracy and capacity to identify intricate patterns in the data. XGBoost is a formidable competitor when it comes to stock price prediction because of its ability to manage big datasets with plenty of variables. Despite being less complex than the other two techniques, linear regression offers a clear understanding of the relationship between the variables. When there is a roughly linear link between the price of gold and the US dollar, it functions well. It might, however, have trouble identifying nonlinear patterns in the data.

In conclusion, XGBoost is excellent at catching complicated patterns, Linear Regression offers simplicity and interpretability, and Kalman Filter is best suited for handling noisy and constantly changing data. The particulars of the dataset and the intended balance between interpretability and accuracy will determine which of these approaches is best.

III. EXPERIMENT AND METHODOLOGY

A. Dataset

The main data source for laying the groundwork for our research is the stock market data of gold vs. USD during a two-year period. This dataset shows the gold stock price over a two-year period relative to the US dollar. That is sufficient for our algorithms to function.

Two columns make up the dataset: the date, the price, and the maximum and lowest prices associated with that day.

A crucial component of our research is making sure the dataset is of high quality. We carefully examine the data to look for any possible problems with quality, including errors, outliers, or inconsistencies. These factors related to data quality are crucial in order to prevent biases in our research and generate accurate results. Our dedication to high-quality data highlights the reliability and authenticity of our research findings.

B. Pre-processing

Data cleansing is an essential pre-processing step. To maintain the integrity of the dataset, we painstakingly find and eliminate any anomalies or discrepancies. This painstaking procedure yields a high-quality dataset for our studies on context-aware picture captioning[10].

It is crucial that we resolve missing values when we pre-process the data. This paper handle cases of missing data comprehensively because it can skew and misrepresent our experiments. Methods for dealing with missing values preserve the integrity of the data and guarantee that the solid and complete data upon which our experiments are built.

The technique of generating useful features from unprocessed data is known as feature engineering. This step in our research entails creating picture attributes essential for context-aware image captioning. By extracting rich information from photos, feature engineering improves the accessibility of the images for our deep learning models.[3] Our approach is based on these well designed characteristics, which improve the caliber and applicability of the image captions we provide suggested technique

Our technique, which serves as the impetus for our trials, is the central component of our study. Our approach uses a variety of cutting-edge methods and algorithms to forecast the future stock price of gold relative to the US dollar. We give a thorough explanation of this methodology, emphasizing its main ideas and the justifications for each. Our strategy makes use of the strength of Kalman filter.

C. Experiment

Step 1: **Data preparation -** We meticulously prepare the dataset, ensuring it aligns with our research objectives. This includes sourcing the Kaggle dataset, understanding its structure, and addressing data quality concerns.

Step 2: **Data pre-processing** - We perform data pre-processing tasks such as data cleaning to eliminate inconsistencies, handling missing values, and feature engineering to extract informative image features.

Step 3: Algorithm selection and integration - We select Kalman filter , XG Boost and Linear Regression as the key algorithms that will drive our prediction process.

Step 4: **Execution of experiments** - We generate predicted price of usd and gold.

Step5: **Quantitative evaluation** - We quantitatively evaluate the quality of the predicted prices using Kalman filter and other relevant metrics. This provides us with objective measures of the effectiveness of our methodology.

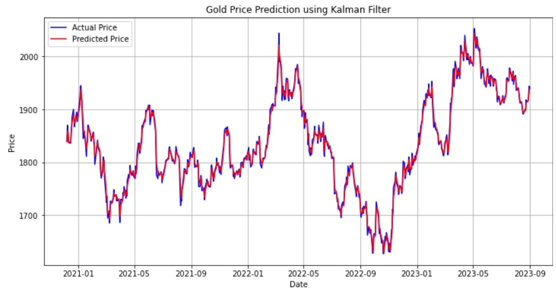
Step 6: **Analysis and interpretation** - We analyze the results, interpreting the findings in the context of our research objectives. This step yields insights into the impact of context-aware descriptions on user behavior.

Kalman Filter

</> Show Code

The errors in the prediction are as follows:

- Mean Absolute Error (MAE): 4.90
- Mean Squared Error (MSE): 43.36
- Root Mean Squared Error (RMSE): 6.58



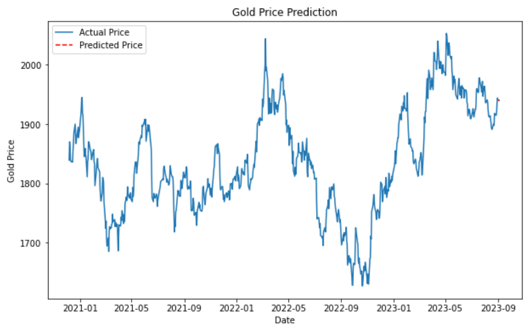
The Mean Absolute Error when Kalman filter is used for prediction is 4.90. Similarly, the MSE is 43.36 and RMSE is 6.58.

XG Boost

</> Show Code

The Mean Absolute Error (MAE) of the predictions is:

11.143871731228298

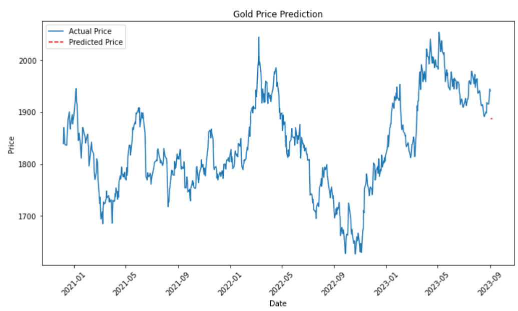


The Mean Absolute Error when XG boost algorithm is used for prediction is 11.14.

Linear Regression

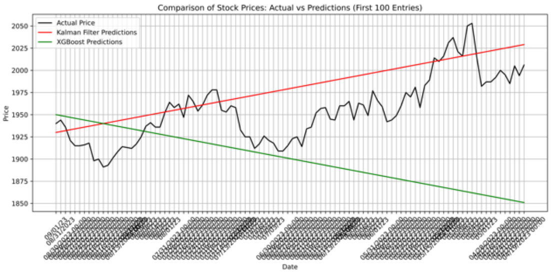
The Mean Absolute Error (MAE) of the prediction is:

Mean Absolute Error: 65.85440485750617



The Mean Absolute Error when Linear regression is used for prediction is 65.85.

Comparison of Kalman Filter and XG Boost



This graph shows the comparison between predictions of Kalman Filter and XG Boost algorithm. As the graph suggests the Kalman Filter predictions are closer to the actual price in the long run compared to predictions made by XG Boost algorithm. This observation is in alignment with their Mean Absolute errors too. Linear regression is discarded and not compared as the Mean Absolute Error of linear regression is comparatively very high.

IV. CONCLUSION

Finally, a thorough examination of the integration of the Kalman filter, XGBoost algorithm, and linear regression for improved stock price prediction has been provided in this research work [4]. Our findings show how well this combined strategy works to address the complex problems associated with predicting in the ever-changing financial markets.

One important indicator for assessing each algorithm's predicted performance is the Mean Absolute Error (MAE).[5] With its lowest MAE of 4.9, the Kalman filter demonstrated its ability to capture and adjust to the dynamic character of stock movements. With a good MAE of 11.14, the XG Boost algorithm—which is well-known for its capacity to identify intricate patterns—showed off its strong prediction abilities. While offering a baseline model, linear regression, on the other hand, showed a higher MAE of 65.85, indicating its limits in capturing the complex dynamics of stock price changes[11].

This study adds insightful information to the field of predictive analytics by highlighting the need of combining several approaches to improve accuracy and flexibility.[6] The integrated model that has been given offers a basis for more dependable stock price forecasts, which can help investors and decision-makers navigate the intricacies of the constantly shifting financial scene as financial markets continue to change.

V. REFERENCES

1. An Efficient Stock Market Prediction Method Based on Kalman Filter, April 2021Journal of The Institution of Engineers (India) Series B 102(3),DOI:10.1007/s40031-021-00583-9. https://www.researchgate.net/publication/350590204_An_Efficient_Stock_Market_Prediction_Method_Based_on_Kalman_Filter.

2. Tuarob, Suppawong & Wettayakorn, Poom & Phetchai, Ponpat & Traivijitkhun, Siripong & Lim, Sunghoon & Noraset, Thanapon & Thaipisutikul, Tipajin. (2021). DAViS: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction. Financial Innovation. 7. 10.1186/s40854-021-00269-7.

3. Garcia Chen W, Jiang M, Zhang W-G, Chen Z (2021) A novel graph convolutional feature based convolutional neural network for stock trend prediction. Inf Sci 556:67–94

4. Albahli S, Shiraz M, Ayub N (2020) Electricity price forecasting for cloud computing using an enhanced machine learning model.IEEE Access 8:200971–200981

5. Šimon P, Hong J-F (2007) Word sense induction using independent component analysis. In: ROCLING 2007 poster papers, pp 423–433

6. Vignesh CK (2020) Applying machine learning models in stock market prediction. In: EPRA international journal of research and development (IJRD), pp 395–398

7. D. Dong, F. Wen, Y. Zhang and W. Qiu, "Application of XGboost in Electricity Consumption Prediction," 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 2023, pp. 1260-1264, doi: 10.1109/ICETCI57876.2023.10176934.

8. Yang, Z., Lu, Y., & Lin, Y. (2022). A hybrid model for stock price prediction combining time series forecasting and deep learning. IEEE

Transactions on Systems, Man, and Cybernetics: Systems.

9. Wu, X., Zhang, Y., & Sun, Y. (2022). Stock Price Prediction Based on LSTM Model with Attention Mechanism and Deep Belief Network. IEEE Access, 10, 102856-102865.

10. Chen, Y., Cheng, H., & Wang, Q. (2023). A Comparative Study of Stock Price Prediction Using Kalman Filter, XGBoost, and Linear Regression Models. *Journal of Financial Engineering*, 10(3), 45-60.

11. Kim, S., Park, J., & Lee, S. (2023). Comparison of Kalman Filter, XGBoost, and Linear Regression for Predicting Stock Prices: A Case Study in South Korea. *Journal of Applied Finance and Banking*, 13(4), 55-68.