Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1691043

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 8 - Issue 4: 1134-1139 / July 2025

FORECASTING MUTUAL FUND PRICES ON TEFAS USING LIGHTGBM

Oktay ÖLMEZ[1]*

[1]*Afiniti, 34752 İstanbul, Türkiye*

Abstract: This paper investigates the application of the Light Gradient Boosting Machine (LightGBM) algorithm for predicting the prices of mutual funds traded on the Türkiye Electronic Fund Trading Platform (TEFAS). Given the increasing importance of mutual funds as an investment vehicle in Türkiye, this study explores the effectiveness of a state-of-the-art machine learning approach. Utilizing historical data from TEFAS, the LightGBM model is employed to capture complex patterns and non-linear relationships within the financial time series data. The research outlines the methodology for data preparation, feature implementation, data splitting, model configuration, training, and evaluation, including case studies to demonstrate the practical application and results.

Keywords: Machine learning, LightGBM, Tefas

***Corresponding author:** Afiniti, 34752 İstanbul, Türkiye
E mail: olmezoktay@gmail.com (O. ÖLMEZ)
Oktay ÖLMEZ https://orcid.org/0000-0002-9130-0038

Received: May 04, 2025
Accepted: June 11, 2025
Published: July 15, 2025

Cite as: Ölmez O, 2025. Forecasting mutual fund prices on TEFAS using LightGBM. BSJ Eng Sci, 8(4): 1134-1139.

1. Introduction

The landscape of financial markets is continuously evolving, presenting investors with a diverse array of instruments for wealth management and growth. Among these, mutual funds stand out as a popular investment vehicle, offering diversification and professional management across various asset classes. In Türkiye, the Türkiye Electronic Fund Trading Platform (TEFAS) plays a pivotal role in this ecosystem, serving as a centralized and transparent platform for the trading of a wide range of mutual funds. Established to provide investors with easy access to compare and trade funds from different management companies through a single account, TEFAS has significantly enhanced the accessibility and efficiency of the Turkish mutual fund market.

For both individual and institutional investors, the ability to anticipate future price movements of mutual funds is essential to make informed investment decisions, optimize portfolio performance, and manage risk effectively. However, mutual fund price prediction is a complex task influenced by a myriad of factors, including underlying asset performance, market sentiment, economic indicators, and fund-specific characteristics. The inherent volatility and dynamic nature of financial time series data make accurate forecasting a significant challenge.

Traditionally, statistical models and fundamental analysis have been employed for financial forecasting. While valuable, these methods often face limitations in capturing the intricate non-linear relationships and complex patterns present in modern financial data. The advent of machine learning techniques has opened new avenues for tackling these challenges, offering powerful tools capable of learning complex dependencies and making data-driven predictions.

This paper aims to explore the application of Light Gradient Boosting Machine (LightGBM), a highly efficient and effective gradient boosting framework, for the prediction of mutual fund prices available on the TEFAS platform. LightGBM has gained considerable attention in various domains, including financial time series forecasting, due to its speed, scalability, and ability to handle large datasets while delivering high predictive accuracy. Its tree-based learning algorithm and optimized techniques for gradient boosting make it particularly well-suited for capturing the complex and often non-linear patterns inherent in financial market data.

The study will leverage historical price data of selected mutual funds traded on TEFAS, along with potentially relevant features such as historical returns, trading volume, and market indices, to train and evaluate the LightGBM model. By focusing on the TEFAS platform, this research provides a specific and relevant context within the Turkish financial market. The application of LightGBM, a state-of-the-art machine learning algorithm, represents a contemporary approach to address the challenging problem of mutual fund price prediction.

In the rest of this paper, we will detail the data collection and preprocessing steps undertaken for this study and will describe the methodology for model development, training, and evaluation using LightGBM, including case studies on specific mutual funds. Finally, we will conclude the paper with a summary of the findings, potential implications, and directions for future research.

1.1. LightGBM for Financial Time Series Prediction

Financial time series prediction, particularly forecasting stock market price movements, remains a challenging yet crucial task for investors and financial institutions. The inherent volatility, non-linearity, and complex dependencies within financial data make accurate prediction difficult using traditional statistical methods. In recent years, machine learning techniques have shown significant promise in addressing these complexities by capturing intricate patterns and relationships within financial time series data.

Among the various machine learning algorithms, Gradient Boosting Machines (GBMs) have gained considerable attention due to their powerful predictive capabilities. Light Gradient Boosting Machine (LightGBM), an innovative framework developed by Microsoft, has emerged as a particularly efficient and effective GBM variant. LightGBM's key advantages lie in its speed and reduced memory consumption, achieved through techniques like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), making it well-suited for handling large-scale financial datasets.

Recent research has increasingly explored the application of LightGBM in financial forecasting with promising results. Studies have demonstrated its effectiveness in predicting stock prices and market trends. For instance, Hartanto, Kholik, and Pristyanto (2023) found that a LightGBM model was competitive with and could even outperform other boosting algorithms like XGBoost, AdaBoost, and CatBoost for stock price time series forecasting. This highlights LightGBM's capability in handling the temporal dynamics of stock data.Furthermore, researchers have explored hybrid models that leverage the strengths of LightGBM in combination with other techniques to enhance predictive performance. Guo, Li, and Xu (2021) proposed an LSTM-LightGBM hybrid model for predicting stock price rise and fall, where LSTM captured temporal dependencies and LightGBM further processed these outputs, resulting in improved prediction stability and performance. Similarly, Tian et. Al. (2022) also investigated a hybrid LSTM and LightGBM model for stock price prediction, indicating the benefits of combining these approaches.

Beyond traditional stock markets, LightGBM has also been successfully applied to forecasting trends in the cryptocurrency market, another highly volatile financial time series. Sun, Liu, and Sima (2020) developed a cryptocurrency price trend forecasting model based on LightGBM, demonstrating its robustness in this dynamic environment by incorporating both cryptocurrency data and economic indicators.

Moreover, LightGBM has shown its resilience and effectiveness even during periods of significant market stress. Guennioui, Chiadmi, and Amghar (2024) utilized a global forecasting approach with an optimized LightGBM model for predicting stock prices during the COVID-19 pandemic, showcasing its ability to handle heightened volatility and successfully forecast multiple stocks simultaneously.

These studies collectively underscore the growing importance of LightGBM as a powerful and efficient tool for financial time series prediction. Its ability to handle complex data, coupled with its computational efficiency, makes it a valuable algorithm for developing accurate and robust forecasting models in diverse financial markets. The continued exploration of LightGBM, both as a standalone model and in hybrid architectures, holds significant potential for advancing the field of financial forecasting.

2. Materials and Methods

This section outlines the methodology employed for predicting mutual fund prices on the TEFAS platform using the LightGBM algorithm. The process involves several key steps, including data preparation, feature engineering, data splitting, model configuration, training, and evaluation.

The initial step involves gathering historical data relevant to the mutual funds traded on TEFAS. Last 5 years of data is collected until April 23 of 2025. A primary data structure, a Pandas DataFrame, is generated to organize this information. This DataFrame includes, at a minimum, the following columns:

- date: The date of the observation.
- fund_price: The historical price on the given date.
- usd: The exchange rate of the US Dollar against the Turkish Lira on the given date.
- gold: The price of gold (e.g., per gram or ounce in Turkish Lira) on the given date.

Ensuring data quality, including handling missing values and ensuring proper data types, is a crucial prerequisite before proceeding to subsequent steps.

To simulate a real-world forecasting scenario, a specific prediction horizon is defined. This horizon represents the number of future periods (e.g., days) for which the model will make predictions. Additionally, a test size is specified, determining the proportion of the most recent data that will be held out for evaluating the trained model's performance on unseen data. The remaining data will be used for training.

To capture temporal dependencies and provide the LightGBM model with relevant information, a set of features is engineered from the raw data. These features are designed to reflect past price movements and trends. The generated features include:

- Lagged Price Values: The price of the mutual fund from previous periods (e.g., price from 1 day ago, 2 days ago, etc.). These features help the model understand the relationship between past and future prices.
- Price Change Percentage: The percentage change in the mutual fund's price over different

short-term intervals (e.g., 1-day, 2-day, 3-day percentage change). These capture short-term volatility and momentum.

- Simple Moving Averages (SMA): The average price of the mutual fund over defined windows (e.g., 5-day and 13-day simple moving averages). Moving averages help smooth out price fluctuations and identify trends.

These features are calculated based on the fund_price series in the DataFrame.

Recognizing that external factors can influence mutual fund prices, the exchange rate of the US Dollar (usd) and the price of gold (gold) are included directly into the feature set. These serve as exogenous variables that may capture broader market sentiment, economic conditions, or shifts in investor preference for alternative assets.

The prepared dataset, including the engineered features and external variables, is split into training and testing sets. A time-series split approach is adopted to ensure that the training data precedes the testing data chronologically. This prevents data leakage and provides a more realistic evaluation of the model's prospective forecasting ability. The size of the test set is determined by the previously defined test size parameter.

The Light Gradient Boosting Machine (LightGBM) algorithm, specifically the LGBMRegressor implementation, is chosen for the price prediction task. LightGBM is an ensemble learning method that builds a series of decision trees sequentially, with each new tree correcting the errors of the previous ones. Its efficiency and performance on tabular data and time series forecasting tasks make it a suitable choice.

The model is configured using the following parameters:

- objective: `regression_l1' specifies the objective function to be minimized during training. regression_l1 corresponds to Mean Absolute Error (MAE) regression, which is less sensitive to outliers compared to Mean Squared Error (MSE).
- metric: `mae' defines the evaluation metric to be monitored during training. Mean Absolute Error (MAE) measures the average magnitude of the errors between predicted and actual values.
- n_estimators: 1000 sets the number of boosting rounds or the number of trees to build. A higher number can improve performance but also increases the risk of overfitting and computation time.
- learning_rate: 0.05 determines the step size shrinkage after each boosting iteration. A smaller learning rate makes the training process more stable but requires more estimators.
- feature_fraction: 0.9 specifies the fraction of features to be randomly selected for training each tree. This technique, known as feature bagging, helps to prevent overfitting and increases the diversity of the trees.
- bagging_fraction: 0.8 determines the fraction of data samples to be randomly sampled for training each tree. This is also a form of bagging (data bagging) and helps to reduce variance and prevent overfitting.
- bagging_freq: 5 specifies the frequency for performing bagging. A value of 5 means that bagging will be performed every 5 boosting iterations.
- verbose: -1 controls the verbosity of the training output. A value of -1 silences the output.
- n_jobs: -1 specifies the number of parallel threads to use for training. Setting it to -1 utilizes all available CPU cores.
- seed: 42 sets the random seed for reproducibility of the results in a single run.

3.1. Model Training

The configured LGBMRegressor model is trained on the training dataset. During the training process, the algorithm iteratively builds decision trees, with each tree learning from the residuals of the previous trees, aiming to minimize the specified objective function (MAE).

The performance of the trained LightGBM model is evaluated on the unseen test dataset. The primary evaluation metric used is the Root Mean Squared Error (RMSE). RMSE is a widely used metric for regression tasks that measures the square root of the average of the squared differences between the predicted and actual values. It provides a measure of the typical magnitude of the prediction errors and is sensitive to large errors. The formula for RMSE is given by equation 1.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2} \quad (1)$$

where n is the number of observations in the test set, y_k is the actual price, and $\widehat{y_k}$ is the predicted price. A lower RMSE indicates better model performance.

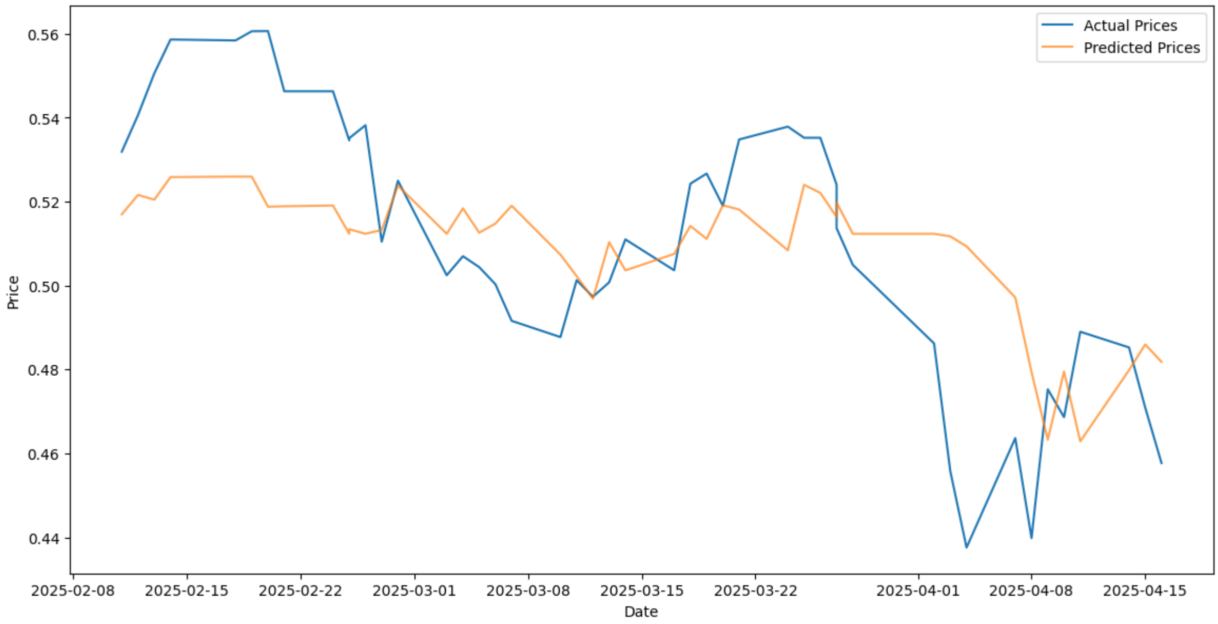

Figure 1. AFT Price Prediction using LightGBM.

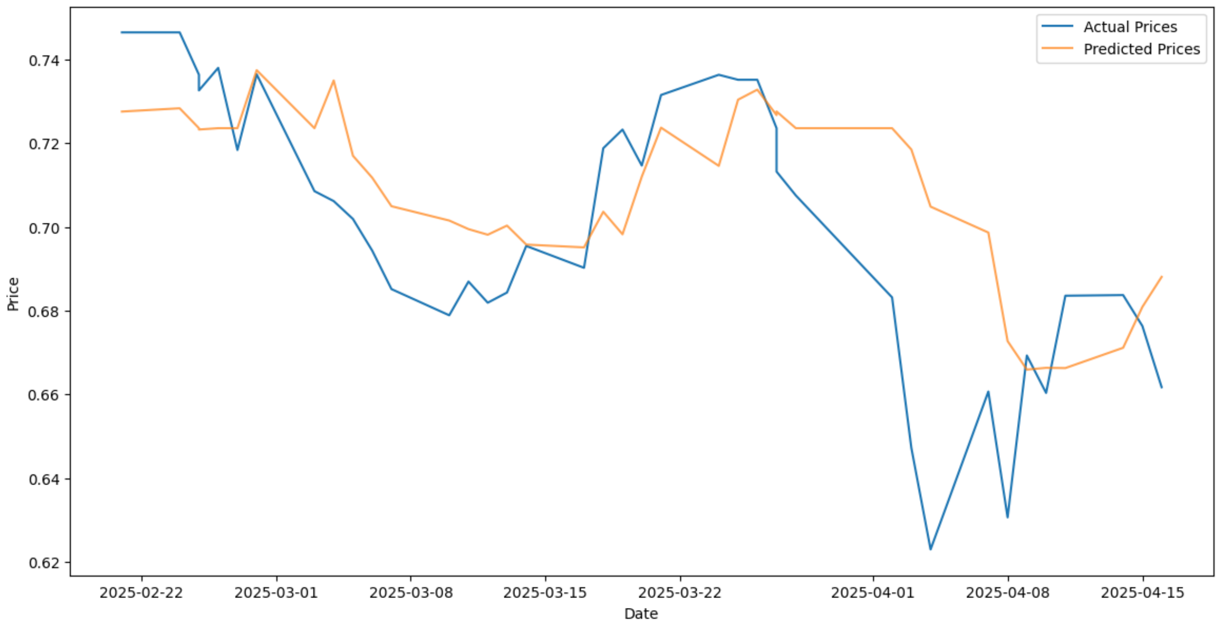

Figure 2. AFA Price Prediction using LightGBM.

3.1. Case Studies

To demonstrate the practical application of the proposed methodology, case studies were conducted using historical data for specific mutual funds available on TEFAS. The data for each fund was processed, and the LightGBM model was trained and evaluated as described in the preceding sections.

3.1.1. AFT mutual fund

For the AFT mutual fund, a single execution of the model with a fixed random seed yielded a Root Mean Squared Error (RMSE) of 0.02418357 on the test set (Figure 1.). In this setting, the predicted values for April 24 and April 25, 2025, were 0.48596486 and 0.48180393, respectively, while the corresponding actual values were 0.464522 and 0.494151. To evaluate the robustness and variability of the model's predictions, the training and prediction processes were repeated 1,000 times with different random seeds. The mean predicted values for April 24 and April 25 across these runs were 0.48851891 and 0.48317624, respectively. These results are summarized in the Table 1.

Table 1. The mean predicted values for April 24 and April 25

Date	April 24, 2025	April 25, 2025
AV	0.464522	0.494151
PV (Single Run)	0.48596486	0.48180393
MPV (1000 Runs)	0.48851891	0.48317624

AV= actual value, PV= predicted value, MPV= mean predicted value.

3.1.2. AFA mutual fund

For the AFA mutual fund, a single execution of the LightGBM model with a fixed random seed yielded a Root Mean Squared Error (RMSE) of 0.02486311 on the test set (Figure 2.). The predicted values for April 24 and April 25, 2025, were 0.68089192 and 0.68806604, respectively, while the corresponding actual values were 0.663933 and 0.694149. Following the same procedure as applied to the AFT mutual fund, the training and prediction processes were repeated 1,000 times with different random seeds to evaluate the stability and variability of the model's forecasts. The mean predicted values for April 24 and April 25 across these runs were 0.68074949 and 0.68883918, respectively. These results are summarized in the Table 2.

Table 2. The mean predicted values for April 24 and April 25

Date	April 24, 2025	April 25, 2025
AV	0.663933	0.694149
PV (Single Run)	0.68089192	0.68806604
MPV (1000 Runs)	0.68074949	0.68883918

AV= actual value, PV= predicted value, MPV= mean predicted value.

5. Conclusion and Future Work

The proposed methodology involved the preparation of historical data for TEFAS mutual funds, including the target fund's price along with relevant external factors like USD exchange rates and gold prices. A crucial step was the engineering of a comprehensive feature set, incorporating lagged prices, price change percentages, and moving averages to capture the temporal dependencies and trends in the time series data. The data was then split using a time-series approach to ensure realistic model evaluation. The LightGBM algorithm, configured with parameters optimized for regression and efficiency, was employed to train the predictive model. Model performance was primarily evaluated using the Root Mean Squared Error (RMSE).

The case studies conducted on the AFT and AFA mutual funds demonstrated the practical implementation of the methodology. The RMSE values obtained for both funds indicated a reasonable level of prediction error, suggesting that LightGBM can effectively learn from the historical data to forecast future price movements. Furthermore, the analysis of predictions from a single model run versus the average predictions over 1000 runs with different random seeds highlighted the importance of assessing model stability and reducing the impact of randomness in financial forecasting. The averaged predictions are considered more robust and reliable indicators of the model's true forecasting capability.

While the results are promising, this study represents a foundational step in applying LightGBM to TEFAS mutual fund price prediction. Several avenues exist for future research to further enhance the accuracy and sophistication of the forecasting model.

For instance, incorporating a wider range of potentially influential features could improve model performance. This includes macroeconomic indicators specific to Türkiye (e.g., inflation rates, interest rates, GDP growth), global market indices, commodity prices beyond gold. Note that a set of parameters was used in this study, more extensive hyperparameter tuning using techniques like grid search, random search, or Bayesian optimization could potentially lead to further improvements in model accuracy.

Author Contributions

The percentages of the author's contributions are presented below. The author reviewed and approved the final version of the manuscript.

	O.Ö.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Guennioui O, Chiadmi D, Amghar M. 2024. Global stock price forecasting during a period of market stress using LightGBM. Int J Comput Digit Syst, 15.1: 19-27.

Guo Y, Li Y, Xu Y. 2021. Study on the application of LSTM-LightGBM Model in stock rise and fall prediction. MATEC Web of Conferences, 336: 05011.

Hartanto A, Kholik YN, Pristyanto Y. 2023. Stock price time series data forecasting using the light gradient boosting machine (LightGBM) model. JOIV: Int J Inform Visualization, 7.4: 2270-2279.

Sun X, Liu M, Sima Z. 2020. A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Res Lett, 32: 101084.

Tian L, Feng L, Yang L, Guo Y. 2022. Stock price prediction based on LSTM and LightGBM hybrid model. J. Supercomput, 78.9: 11768-11793.