Gram Gold/TL Time Series Analysis

Egehan Sözen
Bilişim Sistemeleri Mühendisliği
221307013
egehansozenn@gmail.com

Emirhan Söbüoğlu
Bilişim Sistemleri Mühendisliği
221307014
emirsbgl@gmail.com

Abstract—This paper presents a comprehensive study of time series analysis on gold price (Gram Altın/TL) using five different transformer models: Linear Regression, XGBoost, Temporal Fusion Transformer (TFT), Informer, and ARIMA. The study encompasses data preprocessing steps, model training, and performance evaluation metrics, including Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-Squared (R^2), and time efficiency (training and inference times). A dataset comprising 3488 records was scraped via Selenium and analyzed in Python using Google Colab. Experimental results highlight the strengths and limitations of each model for time series forecasting.

Keywords—Gold price forecasting, transformer models, time series analysis, machine learning, Python, web scraping

I. Introduction (*Heading 1*)

Gold prices have a significant impact on economies and financial markets, making their prediction an essential task for traders and analysts. This paper focuses on time series forecasting of Gram Altın/TL prices using a combination of traditional and advanced machine learning models. The study aims to evaluate the effectiveness of various transformers in forecasting tasks and to identify the most efficient model in terms of accuracy and computation time.

II. Methodology

A. Data Collection and Preprocessing

A Dataset comprising 3488 gold price records was scraped from Investing.com using Selenium. Nine technical indicators, such as Moving Averages and Relative Strength Index (RSI), were calculated and added to the dataset. Data normalization was performed to standardize the input for model training, and the processed data was saved in a CSV file for consistency across experiments.

B. Models Used

- **Linear Ragression:** A statistical model that establishes a linear relationship between features and the target variable.
- **XGBoost:** An ensemble learning model known for its robustness and speed
- **TFT**(Temporal Fusion Transformer)**:**A transformer model specifically designed for time series forecasting, capable of handling temporal dependencies.
- **Informer:** An efficient transformer model optimized for long-term dependency modeling in time series.
- **ARIMA**(Autoregressive Integrated Moving Average): A classical statistical model used for univariate time series analysis.

III. Result and Discussion

A. Evaluation Metrics

The performance of the models was assessed using the following metrics:

- **MSE**(Mean Squared Error)**:** Measures the average squared difference between predicted and actual values.
- **MAPE**(Mean Absolute Percentage Error)**:** Computes the average magnitude of prediction errors.
- **MAE**(Mean Absolute Error):Computes the average magnitude of prediction errors.
- **RMSE**(Root Mean Squared Error)**:** Highlights the magnitude of prediction errors in the same units as the target variable.
- **R-Squared**(R^2)**:** Indicates the proportion of variance explained by the model.
- **Time Efficiency:** Training and inference times were recorded to assess computational efficiency.

B. Results

The results for each model are summarized below:

1. Linear Regression
   - Training Time: 00.3 seconds
   - MSE: 424.34
   - MAE: 9.83
   - RMSE: 20.60
   - MAPE: 2.28%
   - R-Squared: 1.00
   - Inference Time: 0.00026 seconds
2. XGBoost
   - Training Time: 1.34 seconds
   - MSE: 133.34
   - MAE: 5.58
   - RMSE: 11.55
   - MAPE: 1.23%
   - R-Squared: 1.00
   - Inference Time: 0.00516
3. TFT
   - Training Time: 117.13 seconds
   - MSE: 579568.06
   - MAE: 426.02
   - RMSE: 761.29
   - MAPE: 76.37%
   - R-Squared: -0.23
   - Inference Time: 0.08350 seconds
4. Informer
   - Training Time: 40.74 seconds
   - MSE: 94.52
   - MAE: 7.80
   - RMSE: 9.72
   - MAPE: 2.98%

- R-Squared: -1.38
- Inference Time: 0.51 seconds

5. ARIMA
    - Training Time: 6.06 seconds
    - MSE: 0.12
    - MAE: 0.26
    - RMSE: 0.34
    - MAPE: -1.38%
    - R-Squared: -1.38
    - Inference Time: 0.01470 seconds

C. *Key Findings*

1. **Accuracy**: Informer and TFT outperformed the other models in terms of accuracy, with Informer achieving the lowest MSE and RMSE.
2. **Efficiency:** Linear Regression and XGBoost exhibited the fastest training and inference times but lacked precision in long-term predictions**.**
3. **Interpretability:** ARIMA provided interpretable results but struggled with the complexity of multivariate inputs and temporal patterns.

IV. Conclusion

This study demonstrates the potential of advanced transformer models like TFT and Informer for accurate and efficient gold price forecasting. While traditional models like Linear Regression and ARIMA offer simplicity and speed, they fall short in capturing intricate temporal dependencies. Future work can explore hybrid models and the inclusion of external factors such as economic indicators to further enhance prediction accuracy.

Acknowledgment *(Heading 5)*

We extend our gratitude to Investing.com for providing data access and the developers of Google Colab for facilitating computational resources. Special thanks to the contributors of Python libraries like Selenium and Pandas for enabling seamless data handling.

References

[1] DataCamp, "Time Series Analysis Tutorial". Available:https://www.datacamp.com/tutorial/time-series-analysis-tutorial

[2] PyTorch, "PyTorch Documentation,". Available: https://pytorch.org

[3] Pandas Documentation, "10 Minutes to pandas," Pandas User Guide. Available: https://pandas.pydata.org/docs/user_guide/10min.html

[4] Veri Bilimi Okulu, "Time Series Analysis with Python," Available: https://www.veribilimiokulu.com/python-ile-zaman-serisi-analizi/

[5] Zhou, H., "Informer2020: A PyTorch Implementation of Informer," GitHub, Available: https://github.com/zhouhaoyi/Informer2020