Examine accuracy of SHAP explanation on Gold Price prediction using XGBoost

Yanyu Chen: yc3823@nyu.edu

Abstract

I chose Track 2 as my project Track. This is the GitHub Repository Link of my project.

1 Research Paper Summary

1.1 "Forecasting gold price with the XGBoost algorithm and SHAP interaction values"

In this article [1], Jabeur and his partners applied 6 different machine learning models to predict the gold price and compare their performance. Then, they found that the prediction accuracy of the XGBoost was the best. Meantime, since the investment decision on gold is not only made by the model, but also by the human investor, who usually make the final decision. Jabeur used the Shapley additive explanations (SHAP) to help humans interpret predictions of the XGBoost model. In specific, they used the depends plots to interpret feature attributions.

In general, with XGBoost forecasts quantitative gold price, and SHAP helps investors and Policy makers interpret results to do quantitative decisions. Jabeur and his teammates concluded this strategy can increase the gold price forecasting performance.

However, Jabeur et. al. also admit that this study has limitations similar to other price forecast studies. The real environment of gold market is much more complicated and volatile. Especially when some unexpected geo-political decisions are made worldwide. Hence, they suggest some future work like adding more feature variable can be done. Also, applying this strategy of XGBoost and SHAP explanation to other commodities' forecast can be tried too.

1.2 "Problems with Shapley-value-based explanations as feature importance measures"

Kumar et. al. [2] elucidates from two perspectives to explain how the Shapley values generate issues with feature importance explanation.

The first is mathematical perspective, and a main mathematical issue with SHAP values explanation is that SHAP is not globally applicable, in other words, it's limited for non-additive models.

Secondly, the SHAP values may become not useful in explaining black-box model due to the human reason. In specific, there is no general standard norm to explain the feature importance relationships visualized by the help of Shapley values. Hence, the different subjective human interpretation may influence the objective evaluation of a model's prediction ability.

Hence, Kumar concludes that SHAP is ill-suited as general solution to quantifying feature importance. To further investigate how to improve SHAP, more case studies involving human interpretation shall be done.

1.3 "Interpreting financial time series with SHAP values"

Mokhtari et al. [3] investigate how non-linear models predict text commentaries on financial time series data, and use SHAP values to explain the predication results. Mokhtari also used the SHAP values to assess the usefulness of additional datasets. In specific, probing how the predication results change with new features added. The commentary dataset they used is provided by their industrial partner in retailing.

In the end, Mokhtari found out that new dataset doesn't improve the model learning greatly, but SHAP values itself can help improving model prediction accuracy. In particular, using the SHAP values as transformation of original dataset to do model training in addition to explaining.

1.4 "Synthetic Data"

Barr introduce a series of code to generate nonlinear complex synthetic data for Black-box machine learning models [4]. And I will reference to his code to create synthetic dataset for my project.

References

[1] S.B.Jabeur. S.Mefteh-Wali. and J.L. Viviani. "Forecasting gold price with the XGBoost algorithm and SHAP interaction values,". *Springer, Annals of Operations Research*, 2021, Doi.

[2] I. E. Kumar. S. Venkatasubramanian. C. Scheidegger. and S. Friedler. "Problems with Shapley-value-based explanations as feature importance measures". *MLR, International Conference on Machine Learning, page 5491-5500*, 2020.

[3] K.Mokhtari. B.P.Higdon.and B.Ayşe. "Interpreting financial time series with SHAP values". *CASCON '19, Proceedings of the 29th Annual International Conference Page 166-172*, 2019, Doi.

[4] Brian Barr, Ke Xu, Claudio Silva, Enrico Bertini, Robert Reilly, C. Bayan Bruss, and Jason D. Wittenbach. "Towards ground truth explainability on tabular data". In *2020 ICML Workshop on Human Interpretability in Machine Learning (WHI 2020)*, pages 362–367, 2020 Git-hub.

2 Project Description

My main goal is to create an ideal synthetic dataset of gold prices and some influencing features, then train the XGBoost model to obtain predictions results for further SHAP explanation applications. In reference to the works of Kumar and Mokhtari, additional details of SHAP explanations' accuracy of results will be probed in my project.

Indeed, this plan sounds very similar to the one done by Jabeur et. al.. However, I plan to add different features to my synthetic dataset to see whether the prediction accuracy can be improved, and the SHAP values visualization will be included to ensure interpretability for normal people. In specific, with reference to Jabeur's discussion, I plan to add some political features, like the occurrence frequency of wars and diplomatic conflict each month. And I plan to use the dependency plots and force-plot to help interpret the result. This dataset will be non-linear time series data. For the examination of SHAP explanations' accuracy, the specific two details, or hypotheses, I want to continue probing are the following.

First, previously we see that Kumar elucidates the SHAP value is not globally applicable, and it is ill-suited for quantifying feature importance due to the lack of standard norm for the explanation of black-box model results. Hence, I plan to explain the results of predictions with SHAP value visualization, and by presenting them to people to see how the verbal difference leads to biased or different decision-making in gold investment intention.

Second, Mokhtari concludes in his research that the usefulness of additional datasets is limited, but using the SHAP values as training data can improve the prediction accuracy of his nonlinear model. Hence, I might try to use the SHAP values of the gold dataset to train XGBoost and probing results. By doing so, an improved strategy of forecasting commodity prices might be made upon Jabeur's work.

In the end, to practice the strategy combo in the real world, I will pick one specific commodity to investigate. But I have decided which one to choose, I will explain it in the next stage.