CHAPTER I

INTRODUCTION

1.1. Background

Gold has long been recognized as one of the most widely used investment instruments and serves as a safe-haven asset due to its ability to preserve value during periods of economic uncertainty and inflation. Nevertheless, gold prices do not follow a linear pattern. Instead, they are influenced by a wide range of factors, including global gold prices, exchange rates, inflation, monetary policy, and global commodity prices. The interaction among these factors results in highly dynamic and complex price movements, making gold price forecasting a significant challenge in time series analysis [1], [2]. Accurate forecasting is valuable not only for helping investors formulate effective investment strategies and manage risk, but also for supporting more informed economic decision-making [3].

Recent advances in machine learning have provided an alternative approach for modeling time series data characterized by non-linear relationships. One of the most widely adopted algorithms is Support Vector Regression (SVR), an extension of Support Vector Machine designed for regression tasks. By employing kernel functions to capture complex relationships, SVR is capable of achieving strong generalization performance [7]–[9]. In addition, Random Forest has been widely applied to forecasting tasks because of its ability to capture non-linear relationships, reduce the risk of overfitting, and maintain robustness against noise through its ensemble learning approach [10], [11]. Given the distinct characteristics of these two algorithms, comparing their predictive performance is particularly relevant for modeling the highly volatile behavior of gold prices.

In addition to algorithm selection, model performance is also influenced by the hyperparameter tuning process. Suboptimal hyperparameter settings may limit the model's ability to produce accurate predictions. Therefore, this study employs Bayesian Optimization as the hyperparameter optimization technique to identify the optimal model configuration. Compared with conventional approaches such as grid search and random search, Bayesian Optimization is more efficient because it employs a probabilistic strategy to identify promising hyperparameter

combinations based on the outcomes of previous evaluations [12]–[14].

Previous studies have demonstrated that both Support Vector Regression and Random Forest are capable of producing accurate predictions of gold prices, achieving strong predictive performance across various forecasting scenarios [10], [15]–[17]. Nevertheless, most previous studies have relied on a univariate approach that uses only historical gold price data, without fully accounting for the influence of macroeconomic variables and global factors. In addition, the use of temporal feature engineering techniques, such as lag features and moving averages, remains limited, while hyperparameter optimization is still commonly performed using conventional methods. Comparative studies evaluating the performance of Support Vector Regression and Random Forest on Indonesian gold price forecasting using a multivariate time series approach are also relatively scarce.

On the other hand, more complex machine learning models do not necessarily deliver better predictive performance than simpler approaches. For time series data with strong trend patterns, the Naive Forecast method can produce competitive forecasting results. Therefore, Naive Forecast is employed as the baseline model to verify whether the proposed machine learning algorithms provide meaningful performance improvements over a simple forecasting approach. Using a baseline model also enables a more objective evaluation, allowing the added value of the developed model to be demonstrated empirically.

Based on these considerations, this study proposes a multivariate time series model for gold price forecasting by integrating global gold prices, the USD/IDR exchange rate, inflation, and world oil prices as predictor variables. The proposed framework also incorporates temporal feature engineering techniques, including lag features, moving averages, volatility, and a day index, while employing Bayesian Optimization to optimize the hyperparameters of the Support Vector Regression and Random Forest models. The predictive performance of both algorithms is then compared with that of the Naive Forecast baseline using MAE, RMSE, and MAPE, with the objective of identifying the most suitable model for forecasting gold prices in Indonesia.

1.2. Formulation of the problem

Based on the background presented above, the research problem addressed in this study is formulated as follows:

2. How do the predictive performances of Support Vector Regression (SVR) and Random Forest compare in forecasting gold prices using a multivariate time series approach, and to what extent do these algorithms outperform the Naive Forecast baseline?

2.1. Research Purpose

The objectives of this study are as follows:

1. To analyze and compare the predictive performance of Support Vector Regression (SVR) and Random Forest for forecasting gold prices using a multivariate time series approach.
2. To evaluate the extent to which Support Vector Regression (SVR) and Random Forest improve forecasting performance relative to the Naive Forecast baseline.

2.2. Scope of Problem

To ensure that the study remains focused and aligned with its objectives, the following research scope and limitations are defined::

1. This study uses multivariate time series data covering the period from January 2018 to April 2026.
2. The primary variable used in this study is the domestic gold price, measured in Indonesian Rupiah (IDR) per gram, which serves as the prediction target.
3. The predictor variables include global gold prices (USD/oz), the USD/IDR exchange rate, Indonesia's inflation rate, and global crude oil prices, representing macroeconomic and global factors that influence gold price movements.
4. This study does not include statistical significance testing, such as the Diebold–Mariano test. Consequently, differences in predictive performance among the algorithms are evaluated solely based on

forecasting error metrics.

2.3. Benefit of Research

The expected contributions of this study are as follows:

1 Academic Benefits

This study is expected to contribute to the advancement of knowledge in the fields of data mining and machine learning, particularly in the application of multivariate time series forecasting models. It also provides insights into the comparative performance of Support Vector Regression (SVR) and Random Forest in modeling the non-linear and volatile behavior of gold prices. Furthermore, this study contributes to the evaluation of the effectiveness of the Naive Forecast method as a baseline model and the application of Bayesian Optimization to improve forecasting performance.

2 Practical Benefits

From a practical perspective, this study is expected to provide more reliable insights into gold price movements in Indonesia based on historical data and macroeconomic variables. The forecasting results may serve as a valuable reference for supporting data-driven investment decisions and more effective risk management. In addition, this study can serve as a foundation for the development of machine learning–based gold price forecasting systems capable of providing quantitative price estimates for future periods.