Ekonomi, Politika & Finans Araştırmaları Dergisi, 2024, 9(4): 628-655
Journal of Research in Economics, Politics & Finance, 2024, 9(4): 628-655
Araştırma Makalesi / Research Article, https://doi.org/10.30784/epfad.1495757

ARTIFICIAL INTELLIGENCE-ASSISTED MACHINE LEARNING METHODS FOR FORECASTING GREEN BOND INDEX: A COMPARATIVE ANALYSIS

Yeşil Tahvil Endeksinin Tahmini için Yapay Zeka Destekli Makine Öğrenme Yöntemleri: Karşılaştırmalı Bir Analiz

Yunus Emre Gür[*], Ahmed İhsan ŞİMŞEK[**] & Emre BULUT[***]

Abstract

The main objective of this study is to contribute to the literature by forecasting green bond index with different machine learning models supported by artificial intelligence. The data from 1 June 2021 to 29 April 2024, collected from many sources, was separated into training and test sets, and standard preparation was conducted for each. The model's dependent variable is the Global S&P Green Bond Index, which monitors the performance of green bonds in global financial markets and serves as a comprehensive benchmark for the study. To evaluate and compare the performance of the trained machine learning models (Random Forest, Linear Regression, Rational Quadratic Gaussian Process Regression (GPR), XGBoost, MLP, and Linear SVM), RMSE, MSE, MAE, MAPE, and R^2 were used as evaluation metrics and the best performing model was Rational Quadratic GPR. The concluding segment of the SHAP analysis reveals the primary factors influencing the model's forecasts. It is evident that the model assigns considerable importance to macroeconomic indicators, including the DXY (US Dollar Index), XAU (Gold Spot Price), and MSCI (Morgan Stanley Capital International). This work is expected to enhance the literature, as studies directly comparable to this research are limited in this field.

Keywords:
Green Bonds, Machine Learning, Rational Quadratic Gaussian Process Regression, SHAP Analysis, Nonlinear Relationships.

JEL Codes:
C45, C53, G12, Q56.

Öz

Bu çalışmanın temel amacı, yeşil tahvil endeks değerlerini yapay zeka destekli farklı makine öğrenmesi modelleri ile tahmin ederek literatüre katkıda bulunmaktır. Çeşitli kaynaklardan bir araya getirilen, 1 Haziran 2021 ile 29 Nisan 2024 tarihlerini kapsayan veriler, eğitim ve test kümelerine ayrılmış ve her biri için standart ön işlemler gerçekleştirilmiştir. Modelin bağımlı değişkeni, küresel finans piyasalarındaki yeşil tahvillerin performansını izleyen ve çalışma için kapsamlı bir ölçüt görevi gören Küresel S&P Yeşil Tahvil Endeksi'dir. Eğitilen makine öğrenmesi modellerinin (Random Forest, Doğrusal Regresyon, Rasyonel Kuadratik Gauss Süreci Regresyonu (GPR), XGBoost, MLP ve Doğrusal DVM) performansını değerlendirmek ve karşılaştırmak için değerlendirme ölçütleri olarak RMSE, MSE, MAE, MAPE ve R^2 kullanılmış ve en iyi performans gösteren model Rasyonel Kuadratik GPR modeli olmuştur. SHAP analizinin son bölümü modelin tahminlerini etkileyen başlıca faktörleri ortaya koymaktadır. Modelin DXY (ABD Doları Endeksi), XAU (Spot Altın Fiyatı) ve MSCI (Morgan Stanley Capital International) gibi makroekonomik göstergelere büyük önem verdiği görülmektedir. Bu çalışmanın, literatürde doğrudan karşılaştırılabilir benzer çalışmaların sınırlı olması nedeniyle alana önemli bir katkı sağlayacağı düşünülmektedir.

Anahtar Kelimeler:
Yeşil Tahviller, Makine Öğrenmesi, Rasyonel Kuadratik Gauss Süreci Regresyonu, SHAP Analizi, Doğrusal Olmayan İlişkiler.

JEL Kodları:
C45, C53, G12, Q56.

[*] Assist. Prof. Dr., Fırat University, Faculty of Economics and Administrative Sciences, Türkiye, yegur@firat.edu.tr
[**] Assist. Prof. Dr., Fırat University, Faculty of Economics and Administrative Sciences, Türkiye, aisimsek@firat.edu.tr
[***] Res. Assist. Dr., Fırat University, Faculty of Economics and Administrative Sciences,, Türkiye, ebulut@firat.edu.tr

Received Date (Makale Geliş Tarihi): 04.06.2024 Accepted Date (Makale Kabul Tarihi): 27.12.2024

1. Introduction

Green projects are often funded via green crowdfunding which are connected to sustainable agriculture, forestry, waste management, renewable energy, and water conservation among others. Green initiatives have three core objectives - to minimize ecological footprint, protect natural assets, and stop causing environmental harm. This aligns with the triple bottom line of sustainability—the process by which companies manage their financial, social, and environmental risks, obligations, and opportunities, while also increasing long-term shareholder value; all three must be met, as well as maintained and improved upon, in order for corporations to ensure that they meet their needs for future generations to come. Climate initiatives often need additional upfront capital seem riskier and have slower returns than conventional choices, which makes green projects less attractive to private-sector investors, causing very limited investments. To this end, it is absolutely essential that private sector investments in green schemes are promoted and encouraged. Investors may lend money to green projects via financial products like interest-free green bonds and green sukuk (green bonds).

Introduced in 2007, green bonds began attracting financial market interest as environmental and sustainable investing gained prominence. The issuance of green bonds has expanded dramatically in recent years as seen in Figure 1.

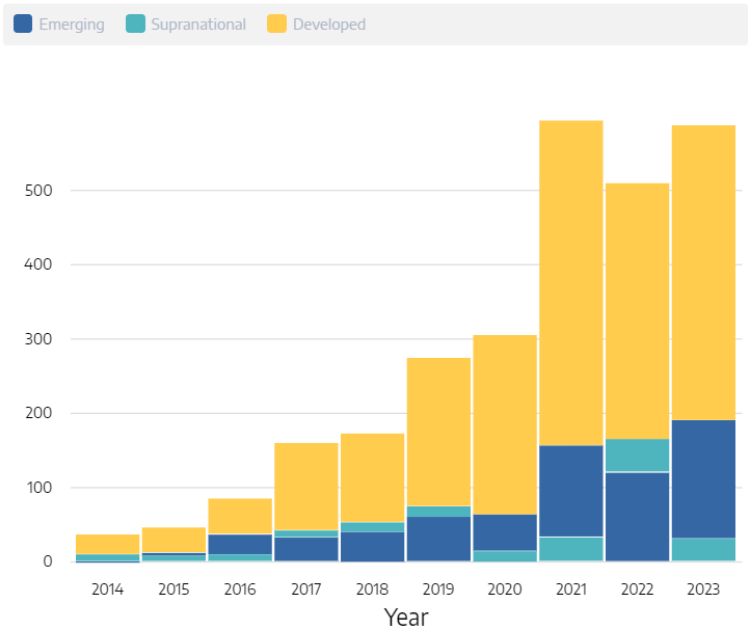

Figure 1. Amount of Issued Green Bonds
Source: www.climatebonds.net

The issue of green bonds entails soliciting funds from bond investors to finance initiatives like energy efficiency and renewable energy. Accurately estimating and projecting the market dynamics of green bonds is crucial for investors and regulators, as their popularity grows. The variables that impact the price of green bonds have been the subject of several studies that have highlighted the importance of these aspects (Tu et al., 2020; Dorfleitner et al., 2022; Verma and Bansal, 2023; Marín-Rodríguez et al., 2022; Çetin, 2022; Wei et al., 2024).

In this context, predicting the price of green bonds will continually relate to broader economic dimensions (both micro and macro), as well as commodity and financial markets and alternative investment prices. This paper's major goal is to answer the research question, which is to estimate the value of green bonds using both conventional (Linear models) and artificial intelligence-based forecasting techniques (Non-linear models), both of which have the ability to affect the green bond index. To represent stock markets across various regions, the Morgan Stanley Capital International Index (MSCI) is utilized, while the Global S&P Green Bond Index (HGGB) is employed to capture the performance of green bonds in the global financial market. Furthermore, our analysis includes a composite of traditionally perceived low-risk assets, such as gold (XAU), the US Dollar Index (DXY), and West Texas Intermediate (WTI). In addition to these, Bitcoin (BTC) is examined, the leading cryptocurrency by market capitalization, which presents a notably distinct risk profile compared to the other assets. Within the scope of the research, random forest, rational quadratic Gaussian Process Regression (GPR), XGBoost, MLP (non-linear), linear SVM, and linear regression (linear) methods were used. The data analyzed covers from June 1, 2021 until April 29, 2024.

The Global S&P Green Bond Index was chosen as the dependent variable in this study because of its extensive representation of green bonds in the global financial markets. This index provides a benchmark for the performance of bonds issued to finance ecologically sustainable projects, reflecting the fast growth of the relatively new green bond category. The index is well-regarded and encompasses bonds from many industries and geographical regions, rendering it an appropriate selection for reflecting global trends and dynamics in sustainable financing. Moreover, its broad scope enables the evaluation of the general trends and predictability of green bonds in relation to macroeconomic issues, which corresponds with the study's aim of comprehending how various variables affect the pricing of green bonds. This index facilitates an evaluation of global green bond market dynamics, enhancing the study's results' robustness and usefulness.

Despite being managed by a Canadian company, the Global S&P Green Bond Index tracks green bonds issued in various regions, including the USA, Germany, China, and others, which have the highest climate bond values. This index represents the global green bond market holistically by including bonds from many issuers from many industries and locations. This index was chosen over country-specific indexes because it provides a holistic perspective of the global green bond market, which is important for studying worldwide green bond patterns and macroeconomic issues. The Global S&P Green Bond Index was chosen for this research because it represents a wide range of countries rather than just one.

With the use of artificial intelligence-supported approaches, which are presently in short supply, this study seeks to add to the body of literature by shedding light on future research on the prediction of the green bond index. In addition to reducing investors' risk perception of green bonds, a relatively new financial instrument, and increasing their profitability, increasing the demand for green bonds and reducing the cost of financing green projects can be achieved by accurately estimating the green bond index.

There are five parts to the research. The introduction is covered in the first portion, and the second part is a review of the literature on green bonds and the variables influencing them. This is fully detailed in the third part of the study on the methods. The fourth part is the conclusion of the study. Lastly, Part 5 concludes and suggests recommendations.

2. Literature Review

Green bonds can also help bridge the financing gap of climate and ecologically favorable investments for those who will not invest directly in climate solutions. The World Bank issued the world's first green bonds (labeled climate bonds) in 2007 with the aim of providing investment instruments that incorporate environmental, social, and governance criteria, more specifically criteria designed using these markers as the lifespan of underlying investments. As far back as 2007, The European Investment Bank launched the first green bond for energy efficiency and renewable energy projects in the amount of 600 million EUR (Ehlers and Packer, 2017). The issuance of green bonds, an innovative financial instrument, was a major breakthrough in the fight against climate change and the promotion of sustainable development goals.

The price anticipation of green bonds based on commodity is a complicated and intricate one and how these factors play around each other are very different. Indeed, research has shown it to be possible to forecast the prices of various asset classes, including the green bonds, using the commodity prices, for example (Chen et al., 2010). Commodity prices and bond prices contain information about the relationship between investment opportunities in the emergence of market volatility hedges (Hong and Yogo 2012). Moreover, some researchers also investigate on the price prediction of corporate green bonds using AI models like artificial neural networks. And this is represented by the interest in new prediction methods (Çetin, 2022). Green bonds and the support from the house bond research have been demonstrated in a variety of studies, which highlight the diversification benefits of green bonds for investors as a portfolio investment alternative, and associations between green bonds and conventional asset classes (Abakah et al., 2022). Research on the Predictability of Commodity-Prices by Different Branches of the US Government Bond Yield Curve: There is also evidence that there may be a connection between bond yields and commodity price movements (Idilbi-Bayaa and Qadan, 2021). In fact, advanced models like copulas and transfer entropy have been used to study the dynamic link between green bonds with conventional asset classes, especially commodities indexes (Hung, 2021). Research suggests the importance of including commodity prices in forecasting models (Black et al., 2014). In a separate study, Broadstock and Cheng (2019) investigated how the co-movement between black and green bond prices changes over time by allowing for time-varying volatility in the financial market, unpredictability in the financial market, the business cycle, the crude oil price, and the news associated with the green bond. In their study, it is found that macro-economic variables and micro variables are equally influential on the green bond price. When market conditions are unfavorable, this means that the market is interconnected and defines a feedback link between oil prices and green bond prices (Marín-Rodríguez et al., 2022).

Like commodity prices, exchange rates have a profound effect on financial markets. In other words, it is important to analyze and understand the impact of foreign exchange rates on the prices of green bonds to manage investment strategies. The relationship between exchange rates and bond yields is analyzed by Tsui and Zhang (2021). One of the methods for predicting green bond prices is presented by Jia (2021) who used deep learning algorithms to forecast exchange rates and financial indices such as the CSI 300 Index. It has also been documented that commodity prices do exert an impact on exchange rates, implying the changes in commodity prices are widespread and have profound implications on financial markets (Zou et al., 2017, Ghoshray and Pundit, 2020). Wei et al. (2024) and Tu et al. (2020) examine the effect of exchange rates, commodity price levels, market volatility, and global economic indicators on the functionality of

the green bond market by examining Commodities prices, currency rates, and especially the price of WTI crude oil are the three key foundation stones for green bond market trends.

In literature, a large number of studies have examined the correlation of green bonds with other financial markets and have demonstrated the forecast performance of models based on these correlations. More recently, there has also research on price spillovers and co-movement effects between the green bond market and other financial markets (e.g., Reboredo, 2018 and Dutta et al., 2021 showing that price shocks in green bonds are affected by foreign exchange markets). Furthermore, in their analyses of green bonds and financial markets, Reboredo and Ugolini (2020) and Naeem et al. (2021) highlight the robust relationship between green bonds and the USD index and bond index.

It was also mentioned in the literature how gold price affects green bond price forecasting. Gao and Zhang (2023) also show that the gold price is significantly associated with green bonds, highlighting an important role played by gold in the volatility of the green bond price. In addition, it is controlled for cointegration between gold prices and other asset classes and therefore argued that bond risk premiums can be estimated from commodity prices, especially gold prices (Bouri et al., 2021a; Bouri et al., 2021b).

The price of green bonds is now being discussed in line with the cryptosystem markets of that era. Huynh et al. (2020) found that green bonds are necessary for climate hedging. Yadav et al. (2022) conducted research on cross-market linkages for green bonds with crypto and other markets. In their study, it is found that the green bond market provides protection against risk. Since the analysis of green bond price prediction is a very complex task, information on stock exchanges, stocks, and indices is very important. Different studies have tried to clarify the relationship between these factors. Accordingly, Xi and Jing (2021) investigate the extent to which returns of listed firms change after green bond issuance and find that investor demand for green bonds has an undeniable impact on security value. Consistently, Zhou and Cui (2019) find that green bond issuance has a positive impact on the stock price of firms with better profitability and operational performance. According to Lebelle et al. (2020), for firms in 28 countries between 2007 and 2017, stock prices increased following green bond issuance announcements. Moreover, Chai et al. (2022) reported that green bonds have a positive correlation with stock returns.

Green bond price forecasting using Artificial Intelligence (AI) has attracted much attention. Cetin (2022) used Artificial Neural Network models to build a framework and identify predictors based on green bonds to forecast corporate green bond prices. This resulted in better forecasting accuracy facilitated by the application of artificial intelligence in modeling green bond price dynamics. Wang et al. (2022a) developed a CEEMDAN-LSTM-based model for forecasting the green bond index. Furthermore, Artificial Neural Networks with time series algorithms and Machine Learning techniques have been applied to forecast asset prices in non-stationary financial time series (Dutta et al., 2020). More specifically to financial assets, tree-based ensemble machine learning models such as Random Forest and XGBoost have been applied to green bond forecasting where their models have achieved better results than tree-based ensemble machine learning models (Ampomah et al., 2020). These models demonstrate how artificial intelligence can predict the prices and values of financial assets. Models such as backpropagation neural networks and extreme learning machine neural networks have obtained success in predicting bond dynamics with stock prices (Maneejuk, 2023). AI has also been used to predict the prices of green securities and machine learning methods as well, such as Random Forests are among the forecast

methods used in different financial applications (Sadorsky, 2021). Dorfleitner et al. (2022) stressed the pricing of green bonds is mostly dependent on the influence of extrinsic variables, and the extent of on environmentally friendly efforts in accordance with which models can be constructed using AI-assisted methods to predict green bond prices. Together, these papers have underscored the importance of AI-supported approaches in the prediction of green bond prices.

This study improves the literature by offering a novel perspective through the integration of AI-assisted methodologies into current green bond pricing forecasting models. This study aims to determine the most effective model by comparing several machine learning models, in contrast to prior research that primarily examined the link between commodity prices, currency rates, macroeconomic indicators, and the green bond index. This study further identifies the macroeconomic elements to which the model assigns more significance using SHAP analysis, a topic rarely addressed in the literature, and emphasizes the importance of variables such as DXY, gold prices, and MSCI. In these aspects, our work diverges from the extant literature both methodologically and practically.

3. Methodology

This study uses machine learning and the fundamentals of artificial intelligence in predicting the daily green bond index. This paper contains a Section with respect to Dataset, its preprocessing steps, hardware and software features of the computer for the analysis, machine learning models for the prediction process, performance evaluation metrics, and the proposed approach.

3.1. Dataset and Preprocessing

This tool-based analysis forecasts the daily green bond index based on data from 1 June 2021 to 29 April 2024. The analysis begins on 1 June 2021, which corresponds to the inception date of the index used in this study. In a highly volatile market, determining a precise starting point can be challenging. To address this, June was chosen as it represents a midpoint in the year, offering a balanced reference for analysis. Moreover, as shown in Graphic 1, 2021 witnessed a significant rise in green bond transactions, marking a pivotal period for sustainable finance. The dataset consists of 710 daily observations of the MSCI, Gold Spot Price, WTI, US Dollar Index (DXY), Global X S&P Green Bond Index ETF, and Bitcoin Spot Price (BTC/USD). In Table 1, the sources of the variables that make up the data collection are shown.

Table 1. Sources of the Variables

Variables	Curency	Type	Source
MSCI	USD	Input	investing.com
Gold Spot Price	USD	Input	investing.com
WTI	USD	Input	investing.com
Dollar Index (DXY)	USD	Input	investing.com
Bitcoin Spot Price	USD	Input	investing.com
Global X S&P Green Bond Index ETF	CAD	Output	finance.yahoo.com

Devereux (2009) argues that although there is volatility in the CAD/USD parity, these fluctuations do not affect the stability of the CAD/USD parity in the medium and long run and

therefore the relative impact of the parity can be neglected. For this reason, the Global X S&P Green Bond Index ETF variable is denominated in CAD while the other variables are denominated in USD.

The data set obtained above was then imported from an Excel file using Python's Pandas package. Also the dropna() function from Python's Pandas package was used to remove any NaN (missing) values from the dataset during data preprocessing. Next, the data frame was divided into the independent variable (X) and dependent variable (Y). The data were standardized using min-max scaling as part of a preprocessing stage. This improves the performance of the models by converting each feature to a value between 0 and 1. In Equation 1, this scaling is shown.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Following the data normalization procedure, the training and test sets of the data set were separated for the regression process. Using the "train test split" command, the data was divided such that the training set included 80% of the data and the test set comprised 20% (568 training data and 142 test data sets were obtained). In order to ensure consistency in this process and to allow comparison of model performances by splitting the data sets in the same way each time, the 'random_state' parameter was also used during data partitioning. Random_state acts as a random number seed, ensuring that the same data partitioning is repeated in each run and the results are comparable. This ensures consistency between different trials (Bisht and Bisht, 2022).

3.2. Software and Hardware

Sign with green papers shall be predicted in this study. The data for the prediction is obtained from published daily through the report database. Eight models are employed using the scikit-learn library, a popular Python library for the implementation of various machine-learning models. The scikit library is preferred in the sub-section because, in comparison with the other libraries, using the scikit library is the easiest way to implement random forest, linear regression, rational quadratic GPR, XGBoost, MLP, and linear SVM models (Hao and Ho, 2019). For the rational quadratic GPR model, the authors used the GaussianProcessRegressor class from the scikit library. The scikit library is also used for the implementation of random forest and linear regression. Based on the explanations given for each model. The library recommended for the XGBoost model is the xgboost library, while the SVM module of the sci-kit-learn library was used for the linear SVM model. For multilayer perceptron models, the tensorflow or Keras libraries were used. These libraries are powerful and are widely used in making models for data analysis and machine learning (Gevorkyan et al., 2019). The pandas and numpy libraries were also used for making models for data manipulation and computation as well as for other general mathematical calculations. Moreover, the matplotlib and Seaborn libraries were used for data visualization during data processing and model evaluation. Using these libraries together, various machine-learning projects can be completed from start to finish.

The computer hardware utilized for all regression procedures is characterized by the following specifications. The processor is represented by a high-performance 8-core Intel Core i7-10700K. In comparison to lower generations of Intel processors, it facilitates higher performance when dealing with larger datasets and conducting more complex mathematical operations. The amount of RAM stands at 32 GB is DDR4, which appears to be sufficient to avoid

capacity shortages during datawork and model training. The graphics card enacting NVIDIA GeForce RTX 3080 type also belongs to the category of powerful GPUs. It is primarily tasked with enhancing computational power, a feature that is valuable when dealing with machine learning models. The amount of storage is equal to 1TB which is provided by means of an NVMe SSD.

3.3. Machine Learning Models

3.3.1. Random Forest Model

Random Forest was introduced as a machine-learning technique by Breiman in 2017 (Chen et al., 2020). This is known as Bagging, a random forest model method that randomly selects more features through a forest (or collection) of decision trees to improve node partitioning and then improves model performance without overfitting (Ren et al., 2020). This algorithm has been applied in a wide variety of fields as it is very cost-effective in tasks related to prior pattern recognition. The random forest has also been used in medical prognosis to predict response to treatment, survival rates, and clinical behavior in complex diseases such as mid-to-late stage cervical cancer (Liu et al., 2021) and anti-LGI1 encephalitis (Li et al., 2022). Furthermore, the random forest model has been used in environmental research on soil moisture and hydrology (Wang et al., 2022b) and in taste and odor prediction in drinking water reservoirs (Kehoe et al., 2015). In addition to these, random forest models have been implemented in green bond market studies in order to forecast price movements, performance scores, and volatilities (Xia et al., 2022; D'Amato et al., 2022; Soltani et al., 2024).

Random forest methods are known for their application to complex data features and their strong predictive performance. This is because it is very simple, versatile, and can easily work with high-dimensional data (Muñoz et al., 2018). The random forest model has undergone other modifications and extensions. For example, Shahhosseini (2021) considered weighted random forest models for classification tasks. Also, for forecasting tasks, Sun et al. (2020) propose randomized Shapley forests. Random forest is a very flexible and easy-to-use algorithm that can be used in most of the highly competitive solutions. Batch learning combined with decision trees is a very powerful support for various prediction problems.

3.3.2. Linear Regression Model

Usually, the Linear Regression Model is used for linear regression. A statistical technique, multivariate analysis aims to understand and predict the relationship between a dependent variable (Y) and multiple independent variables (x1, x2,..., xk). The goal of linear regression is to estimate the value of each regression coefficient, considering the sign, size, and statistical inference of each predictor variable. The forecast target of linear regression estimates how well the explanatory variables can predict the explicative variable (Harrell, 2015). Linear Regression Prediction Models formula is give in Equation 2 (Pentoś et al., 2022):

$$Y_t = X_t\beta + \varepsilon_t \quad (2)$$

This is given by the equation Y_t, the expected value at time t. The vector β = (β0, β1, ..., βk) relates the independent and dependent variables. The vector X_t= (1, x1t, x2t,..., xjt) is a j-dimensional vector containing the j-dependent variables at time t. The stochastic error term at

time t is denoted by ε_t, where t=1, ..., N and the error terms are chosen from a Gaussian distribution.

3.3.3. Rational Quatratic GPR Model

The GPR approach was first suggested by Rasmussen (2003). GPR has the advantage over other techniques that allow easy combinations of multiple ML tasks such as parameter estimation. Thus the regression simplified and its control on the output diminished, which shall make the whole process of regression explicable. The flexible kernel function of GPR is able to provide a confidence interval for prediction whereas GPR works very well even with a small training data set. A limitation of the prediction method is that its cost of computing grows cubically with |x|, which is impractical for all but small data sets (Su et al., 2019). Recent advances from the machine learning community have focused on the development of non-parametric methods to handle very nonlinear problems; one of the most well-known is the Gaussian process (GP). The system includes some stochastic variables, and it is considered that the distribution of all the input and output data is Gaussian. The GP assumes probability distributions for any feasible function that are consistent with the training dataset. Because of this, a GP's variable count is infinite and increases with the size of the training dataset. GPR is a mathematical model made up of a kernel function t(x, x') and a mean function n(x). (Zazoum, 2022).

$$F(x) = GPR(n(x), t(x, x') \tag{3}$$

The measure of central tendency for the variable F is represented by the sign n (x). The values of the test input "x" and test output "Y" are linked in the following way.

$$Y = F(x) + \varepsilon \tag{4}$$

The independent noise term is represented by the symbol ε. It is covered by a distribution with zero mean and σ_m variance. It has the following definition:

$$\varepsilon = D(0, \sigma_m^2) \tag{5}$$

The sample from the dataset follows a marginal probability, which is defined as:

$$H(y|f) = D(y|f, \sigma_m^2 J) \tag{6}$$

$$Y = [Y_1, Y_2, Y_3, \dots\dots\dots, Y_n]^T \tag{7}$$

$$f = [f(x_1), f(x_2), f(x_3), \dots\dots\dots, f(x_m)]^T \tag{8}$$

The forecast dataset conforms to a specific distribution, which is defined by:

$$H(Y_s|x, x', Y) = D(\mu_s, \sigma_s^2) \tag{9}$$

$$\mu_s = k_{sM}(k_m + \sigma_m^2 J)^{-1} Y \tag{10}$$

$$\sigma_s^2 = k_{ss} - k_{sM}(k_m + \sigma_m^2 J)^{-1} k_{Ms} \tag{11}$$

μ_s and σ_s^2 are the posterior mean and posterior variance respectively (in the case of GP). For the Input Layer, x and J are square matrices of size (M x M) and k_{sM} is the covariance matrix between training and test data, respectively.

3.3.4. SVM Model

Cortes and Vapnik introduced the support vector machine (SVM) in 1995 as supervised learning. SVM has a number of advantages, such as the ability to handle small sample sizes, the generalization of dimensional spaces well and flexibility etc. It also can recover from the shortcomings of local optimal solutions. The kernel of statistical learning theory, central of learning theory SVM, The SVM considers three main tasks: density-probability estimation, regression prediction and pattern recognition. The fundamental principle of SVM is to establish a continuous functional linkage between input and output views, relying on a few training data points. The objective is to make sure that the regression function can be as smooth as possible while minimising the error between the predicted regression value and the actual output value (Liu et al., 2020). Successive states of the SVM approach are defined in Equation 12-18.

$$f\ (x) = \ \omega^T\ (x) + b \tag{12}$$

$$L(f\ (x), y, \varepsilon) = f(x) = \begin{cases} 0 & |y - \ f\ (x)| \le \varepsilon \\ |\ y - f\ (x)| - \varepsilon & |y - \ f\ (x)| > \varepsilon \end{cases} \tag{13}$$

$$\begin{cases} Min. \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \\ sub.t. \begin{cases} y_i - \omega\Phi(x_i) - b \le \varepsilon + \xi_i \\ -y_i + \omega\Phi(x_i) + b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \end{cases} \tag{14}$$

$$\omega^* = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\,\Phi(x_i) \tag{15}$$

$$b^* = \frac{1}{N_{nsv}}\left\{\sum_{0<\alpha_i<C}\left[y_i - \sum_{x_i \in SV}(\alpha_i - \alpha_i^*)K(x_i, x_j) - \varepsilon\right] + \sum_{0<\alpha_i<C}\left[y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)K(x_i, x_j) + \varepsilon\right]\right\} \tag{16}$$

$$K(x_i, x_j) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{17}$$

$$f(x) = \ \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b^* \tag{18}$$

The SVM calculates a linear regression function denoted as f(x) in a high dimension space. That is the sample vector with x being any real number. Mapping of that function is done in non-linear fashion. This is of great benefit for the optimization problem, as it is used the linear insensitivity loss function L(f (x), y, ε). The loss function is represented by Equation 14. Equation 14 represents the input vector as $\boldsymbol{x_i}$ and the output value as $\boldsymbol{y_i}$. The variables in question are associated with a certain serial number, denoted by the symbol i. The set of real numbers, represented by the letters R, includes the variables $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$. The dimension of the input vector is d. In this case, the cardinality of the items in an input vector is indicated by the variable d. The number of training samples is indicated by n. The degree of accuracy in regression analysis is indicated by the symbol ε. The variable C represents a penalty factor that quantifies the severity of the penalty imposed on a data sample when its mistake surpasses the threshold value ε. The slack variables $\boldsymbol{\xi_i}$ and $\boldsymbol{\xi_i^*}$ are used to apply penalties on the complexity of the fitting parameters. In order to ascertain the estimation of variables a and b, it is imperative to address the optimization problem as delineated in Equations 15 and 16. The variable $\boldsymbol{N_{nsv}}$ represents the number of support vectors that have been explicitly identified. The Lagrange multipliers, represented by $\boldsymbol{\alpha_i}$ and $\boldsymbol{\alpha_i^*}$, must satisfy the condition of being greater than or equal to zero. Equation 17 in this particular case employs the kernel function, represented as $\boldsymbol{K(x_i, x_j)}$. The Gaussian kernel function, known

for its exceptional capacity to generalize, is selected. Equation 18 denotes the ultimate regression function. The SVM model employed a kernel function.

3.3.5. XGBoost Model

The machine learning technique XGBoost is helpful in predicting prices. This technique has been used in a wide range of fields, such as stock price prediction (Zheng et al., 2017; Yue et al., 2021), home price prediction (Zaki et al., 2022; Sharma et al., 2024), power price prediction (Wu et al., 2022), gold price prediction (Jabeur et al., 2024), and stock market collapse forecasting (Zhu et al., 2022). Additionally, according to Simsek (2024), XGBoost has also been used in the process of enhancing the performance of models that anticipate stock prices. Traditional Boosting Tree approaches only rely on first derivatives. The distributed training of the n^{th} tree becomes more difficult when the residual from the preceding n-1 trees is included. Through the use of second-order Taylor expansion of the loss function, CPU multithreaded processing is optimized in XGBoost. XGBoost uses a variety of methods to reduce overfitting (Li et al., 2019).

The method outlined in Equations 19-26 is used to solve the XGBoost algorithm. Equation 19 represents the collection of regression trees, denoted as F. The variable f_k reflects the number of learners who are weak, whereas K represents the total number of learners who are weak. Equation 20 provides the goal function. The parameter $l\left(y_i, \hat{y}_i{}^{(t)}\right)$ in Equation 20 includes a range of loss functions that are used to address specific problems. Equation 20 is often used to measure the discrepancy between the actual value (y_i) and the predicted value ($\hat{y}_i{}^{(t)}$), as well as the overall intricacy of the model, denoted by $\sum_{k=1}^{t}\Omega(f_k)$. To evaluate the major component, substitute the expected value ($\hat{y}_i{}^{(t)}$) for the sample i^{th} in the repeated cycle t^{th}. The calculation is performed by using the subsequent approximation of the Taylor series at the expected value of y from the previous iterations, referred to as ($\hat{y}_i{}^{(t-1)}$), as seen in Equation 19. In Equation 22, the variables g_i and h_i represent the first and subsequent derivatives of the loss function $l\left(y_i, \hat{y}_i{}^{(t)}\right)$. Based on the information provided above, now it can be calculated the derivative by substituting the corresponding formulas from Equations 22, 23, and 24 into Equation 20. Equations 25 and 26 may be used to formulate solutions for a problem. The numerical representation of the outcomes of the loss function is denoted by the variable obj*. A lower score indicates that the assessed tree structure is more idealistic. The variable w_j^* offers the optimal answer for the weighting variables in the specific case under evaluation.

$$\hat{y}_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \tag{19}$$

$$\min L^{(t)}(y_i, \hat{y}_i{}^{(t)}) = \min\left(\sum_{i=1}^{n} \iota\left(y_i, \hat{y}_i{}^{(t)}\right) + \sum_{k=1}^{t}\Omega(f_k)\right) \tag{20}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda w^2 \tag{21}$$

$$\min L^{(t)} = min\left(\sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t)\right) \tag{22}$$

$$g_i = \partial_{\hat{y}_i{}^{(t-1)}} l\left(y_i, y_i^{t-1}\right) \tag{23}$$

$$h_i = \partial^2_{y_i^{t-1}} l(y_i, y_i^{t-1}) \tag{24}$$

$$fw_j^* = -\frac{\sum g_i}{\sum h_i + \lambda} \tag{25}$$

$$obj^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{(\sum g_i)^2}{\sum h_i + \lambda} + \gamma . T \tag{26}$$

The usefulness and superiority of XGBoost in comparison to other machine learning algorithms has been shown by the fact that it has been demonstrated to be a flexible and strong tool for price prediction across a variety of distinct areas.

3.3.6. Multilayer Perception (MLP) Model

MLP is a simple type of artificial neural network with only one or many continuous layers. At a minimum, a multilayer perceptron should contain an input layer, a hidden layer and an output layer. MLP architecture usually has several hidden layers, which can be used to solve difficult problems (such as approximation) because they provide approximate solutions. By its nature, the Association Representation is a directed graph that maps a set of input vectors to a set of output vectors and reflects the MLP concept. It is a layer of node models that are connected in some way to the members of the next layer. Synaptic connections are also known as synapses or links (Pinkus, 1999; Gao et al., 2020). MLP models are used to model complex situations and predict test scores in multidimensional data (Wilamowski, 2009).

After calculating the weighted sum of inputs with the bias unit, on each unit, or neuron, an activation function is applied in an MLP. The outcome is then produced. The mathematical representation of the result produced by the neuron, is given by Equation 27.

$$a_j = f(\sum_{i=1}^{n} w_{ij}x_i + b_j) \tag{27}$$

Here, a_j is the activation or output of the neuron j, "f" is the activation function which can be sigmoid, tanh, ReLU etc where w_{ij} is the weight between the input i and the neuron j. Additionally, b_j is the input value "i" and x_i is the bias value of neuron j. In MLP, each layer receives signals from all neurons in the previous layer and multiplies these signals by weights. The output of a neuron can also be given as input to the next layer. This process continues until the output layer is reached. For the final activation function in the output layer, it may depend on whether the problem is a regression or classification problem. This allows the network to calculate the errors in the outputs that the network penalizes and then adjust the weights of the network to minimize the errors (Boughrara et al., 2016). This is a mathematical update of the weights as in Equation 28:

$$w_{ij}^{new} = w_{ij}^{old} - n\frac{aE}{aw_{ij}} \tag{28}$$

where 'w_{ij}^{new}' is the new weight value, 'w_{ij}^{old}' is the existing weight value, 'n' is the learning rate, 'E' is the error function and '$\frac{aE}{aw_{ij}}$' is the partial derivative of the error function for the weight. As mentioned above, the hyperparameter configurations for all the methods analyzed in this study were made with the Random Search algorithm and are shown in Table 2.

Table 2: Determined Hyperparameters of the Models as a Result of Optimization with Random Search Algorithm

Random Forest	**Linear Regression**	**Rational Quadratic GPR**
n_estimators: 100	Preset: Linear	Basis Function: Constant
max depth: 40	Terms: Linear	Kernel Function: Rational Quadratic
min samples split: 2	Robust option: Off	Use Isotopric Kernel: Yes
min samples leaf: 1	Fit_intercept: True	Kernel Scale: 1.0
max features: log2	Random state: 42	Signal Standard Deviation: 1.0
bootstrap: False		Sigma: 1.0
Random state: 42		Optimize Numeric parameters: Yes
criterion: gini		n_restarts_optimizer: 0
		Alpha: 1e-10
		Random state: 42
Linear SVM	**XGBoost**	**MLP**
Preset: Linear	Subsample: 0.7	Activation: ReLU
Epsilon: 0.1	n_estimators: 300	Alpha: 0.01727
C: 100	max_depth: 7	Hidden layer sizes: (50,50)
PCA: Disabled	learning rate: 0.05	Learning rate: Adaptive
Max_iter: 1000	colsample_bytree: 1.0	Learning_rate_init: 0.00073
Tol: 1e-3	min_child_weight: 1	Max_iter: 1000
Shrinking: True	gamma: 0	Optimizer: Adam
Cache_size: 200	reg_alpha: 0	Loss Function: Mean Squared Error (MSE)
Verbose: False	reg_lambda: 1	Batch_size: 32
Random state: 42	scale_pos_weight: 1	Beta_1: 0.9
	Random state: 42	Beta_2: 0.999
		Epsilon: 1e-8
		Early stopping: False
		Tol: 1e-4 Random state: 42

3.3.7. SHAP (Shapley Additive Explanation) Approach

Machine learning has demonstrated significant potential in predicting time series data, yet the lack of interpretability in its predictions often hinders practical applications. To address this challenge, Lundberg and Lee (2017) introduced the SHAP (SHapley Additive exPlanations) method, which is designed to explain the predictions of complex machine learning models, including LightGBM, NGBoost, CatBoost, XGBoost, and Scikit-learn tree-based algorithms.

SHAP builds on the foundational concepts of game theory proposed by Shapley (1953), offering a systematic approach to assess the contribution of each input feature to a model's prediction. By calculating Shapley values, the method enables users to determine how much each variable influences the outcome for a specific input instance. This level of transparency helps bridge the gap between the black-box nature of machine learning models and the need for explainable, actionable results. The Shapley value is calculated as follows (Jabeur et al., 2024):

$$\hat{\phi}_j = \frac{1}{K} \sum_{k=1}^{K} (\hat{g}\left(x_{+j}^{m}\right) - \hat{g}\left(x_{-j}^{m}\right)) \quad (29)$$

where $\hat{g}\left(x_{+j}^{m}\right)$ represents the forecast for particular input (x), but with a stochastic number of feature values.

Lundberg et al. (2018) developed TreeSHAP, a specialized approach for interpreting gradient boosting models such as XGBoost. TreeSHAP improves upon conventional methods like feature importance metrics and partial dependence plots by offering a more precise and consistent explanation of feature contributions within the model. Jabeur et al., (2024) propose that TreeSHAP interaction values can be determined in the following manner:

$$\emptyset_{i,j} = \sum_{S \subseteq N\{i,j\}} \frac{|S|\,!\,(M - |S| - 2)\,!}{2\,(M-1)\,!}\delta_{ij}(S) \tag{30}$$

While i $\neq j, \delta_{ij}(S) = f_x(S \cup \{i,j\} - f_x(S \cup \{i\}$ - $f_x(S \cup \{j\}$ + $f_x(S)$, M represents the quantity of features, whereas S represents all potential feature subsets. SHAP values enhance our comprehension of tree models by including feature significance, feature dependency visualizations, local explanations, and summary plots.

3.4. Performance Evaluation

Different evaluation measures such as MSE, MAE, MAPE, RMSE, and R^2 were used to observe how well the machine learning models performed the obtained results were used evaluate the prediction accuracy and efficiency of green bond prices models. The mathematical calculations of these metrics are shown in Equation 31-35 respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{31}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{N} = \frac{\sum_{i=1}^{n} e_j}{N} \tag{32}$$

$$MAPE = \frac{\sum_{t=1}^{n}\frac{u_i}{Y_i}}{N} \times 100 \tag{33}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{34}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu)^2} \tag{35}$$

where y_i means the true rating in testing data set at time i; $\hat{y}_i$ means the prediction rating at time i; e refers to the error of the model; u means the difference between the actual value and the predicted value; N is the quantity of rating prediction pairings between the test data and the predicted results.

3.5. Proposed Approach

This research compares the performance of many artificial intelligence and machine learning models to determine which forecasting model is the most successful in predicting green bond values. This approach is designed to enhance forecasting accuracy and optimize financial analysis processes. The steps of data collection, preprocessing, and model training were meticulously planned and executed. Furthermore, a variety of indicators are used to evaluate the models' performance. An explanation of the general organization and functioning of the suggested method is given in this context via a diagram, upon which a thorough analysis will be based. Figure 2 depicts the suggested approach's visualization.

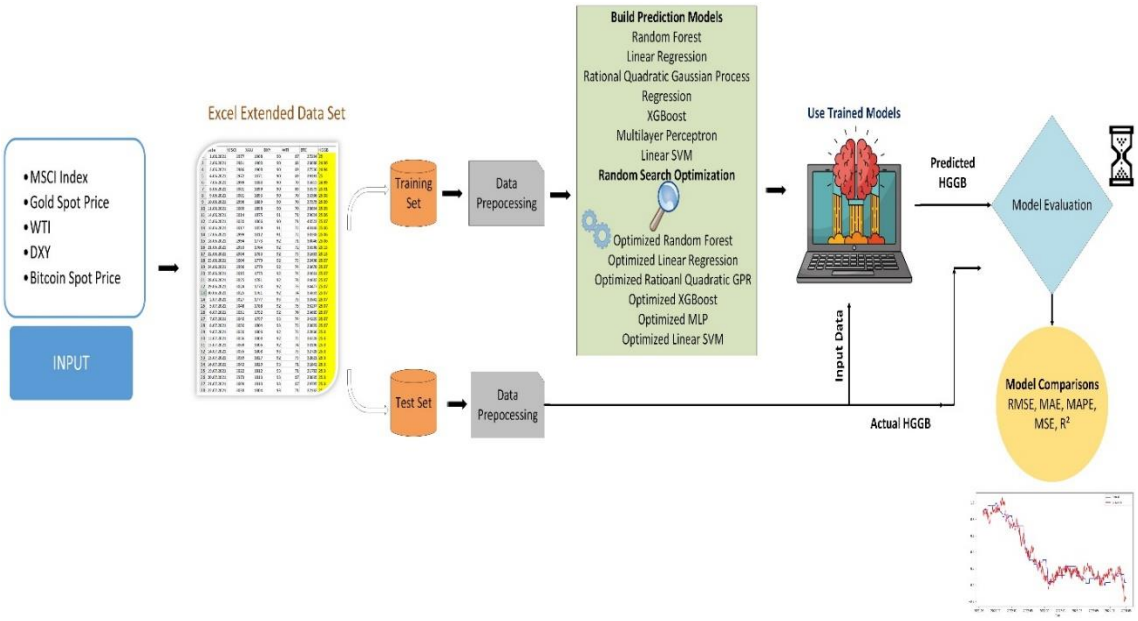

Figure 2. Proposed Approach

The approach outlined in Figure 2 follows a multi-stage and systematic process for forecasting green bond index. Initially, an extensive dataset is constructed using independent variables such as the MSCI, Gold Spot Price, WTI, DXY, and Bitcoin Spot Price. After that, this dataset is split into training and test sets, and each set goes through a series of data pretreatment procedures to be ready for modeling. During the modeling phase, various machine learning and artificial intelligence methods are employed. These models include Random Forest, Linear Regression, Rational Quadratic GPR, XGBoost, MLP, and Linear SVM. In addition, the models apply random search optimization, which helps increase the performance of each model. The training data is used in the training of the models which are then applied to the test one generating predictions. As such, the predictions made are compared to the actual HGGB values. In addition, measured performance against RMSE, MAE, MAPE, MSE, and R2 are as follows: The models will also be compared: the results from the two models will be compared to evaluate how well the models are performing relative to each other and how accurately the models are predicting green bond prices. Lastly, the final end goal of the quantitate analysis is to streamline the process of financial analysis and identify the best model for the future value that can be expected.

4. Findings

In this section, first, it is presented a correlation analysis between the variables used in the study. The results of the correlation analysis are shown in Figure 3. Following the correlation analysis, it is presented a discussion of the results of the AI-based machine learning techniques and their performance in green bond forecasting. Five different evaluation metrics are used to measure the accuracy and efficiency of the models: Random Forest, Linear Regression, Rational Quadratic GPR, XGBoost, MLP, and Linear SVM. In the first step, the models are trained using data from the training set. After comparing the training performances, the results are measured using R^2, MAE, RMSE, MAPE, and MSE for comparison purposes. The prediction plot of the

training results is shown in Figure 3, while Table 3 shows the metric measurements of the training results.

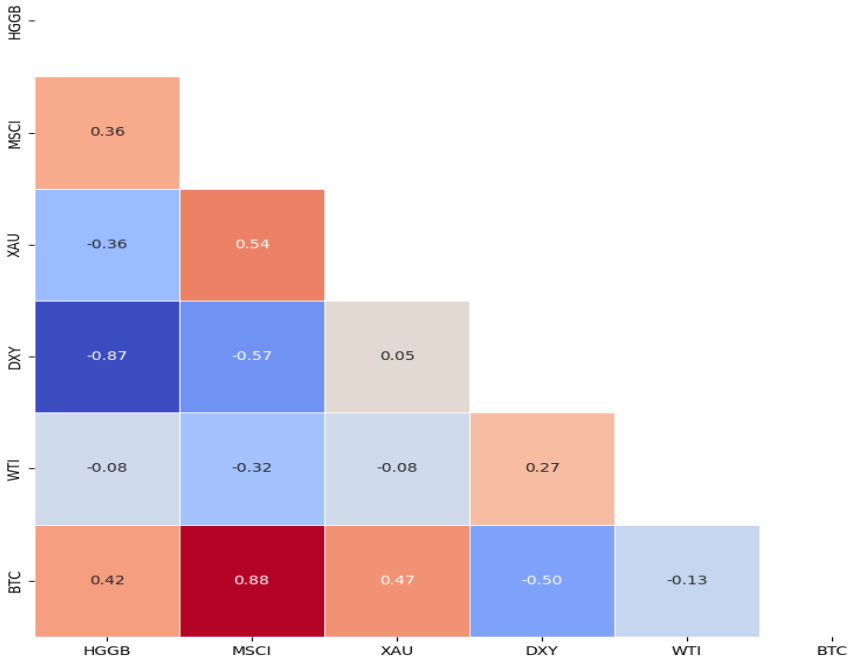

Figure 3. Correlation Matrix (Upper Triangle)

The correlation matrix in Figure 3 shows the relationship between HGGB (the dependent variable) and other independent variables, as well as the relationships between the independent variables themselves. Looking at the correlations of HGGB with other variables, it is seen that the highest positive correlation is with BTC (Bitcoin) at 42%, indicating that HGGB has a moderately positive relationship with Bitcoin prices. A positive correlation of 36% was also observed with the MSCI (international stock index). On the other hand, its correlations with XAU (gold price) and DXY (US dollar index) are negative by 0.36% and 0.87%, respectively. This suggests that the HGGB has a particularly strong negative correlation with the DXY, indicating that as the dollar index increases, the HGGB tends to decrease in a statistically linear relationship. The strong negative correlation between HGGB and DXY (−87%) aligns with the SHAP analysis results, which identify DXY as the most influential variable in predicting HGGB fluctuations. This correlation reinforces DXY's dominant role in the model's predictions. Among the other independent variables, the strong positive correlation between MSCI and BTC (88%) is noteworthy, suggesting that the price movements of the two assets are closely correlated. WTI (crude oil prices), on the other hand, is generally weakly correlated with other variables. In particular, its relationship with HGGB is very weak, with a negative correlation of 8%.

Figure 4 shows the prediction graphs of the training results of the models used in the prediction of the green bond index. In the Random Forest, Rational Quadratic GPR, and XGBoost models, the lines between the actual (blue) and predicted (red) values are closely intertwined, indicating that these models produce highly accurate predictions on the training set. The tight match between the anticipated and real values indicates a high level of model performance.

Conversely, although the Linear Regression, MLP, and Linear SVM models also produced reasonably accurate predictions, there are notable discrepancies between the actual and predicted values in certain periods. This is primarily due to their limited capacity to capture non-linear relationships adequately. The models with the highest overall performance were the Random Forest, Rational Quadratic GPR, and XGBoost models; the models with the worst performance were the Linear Regression, MLP, and Linear SVM models. These findings indicate which models are more effective on the training set.

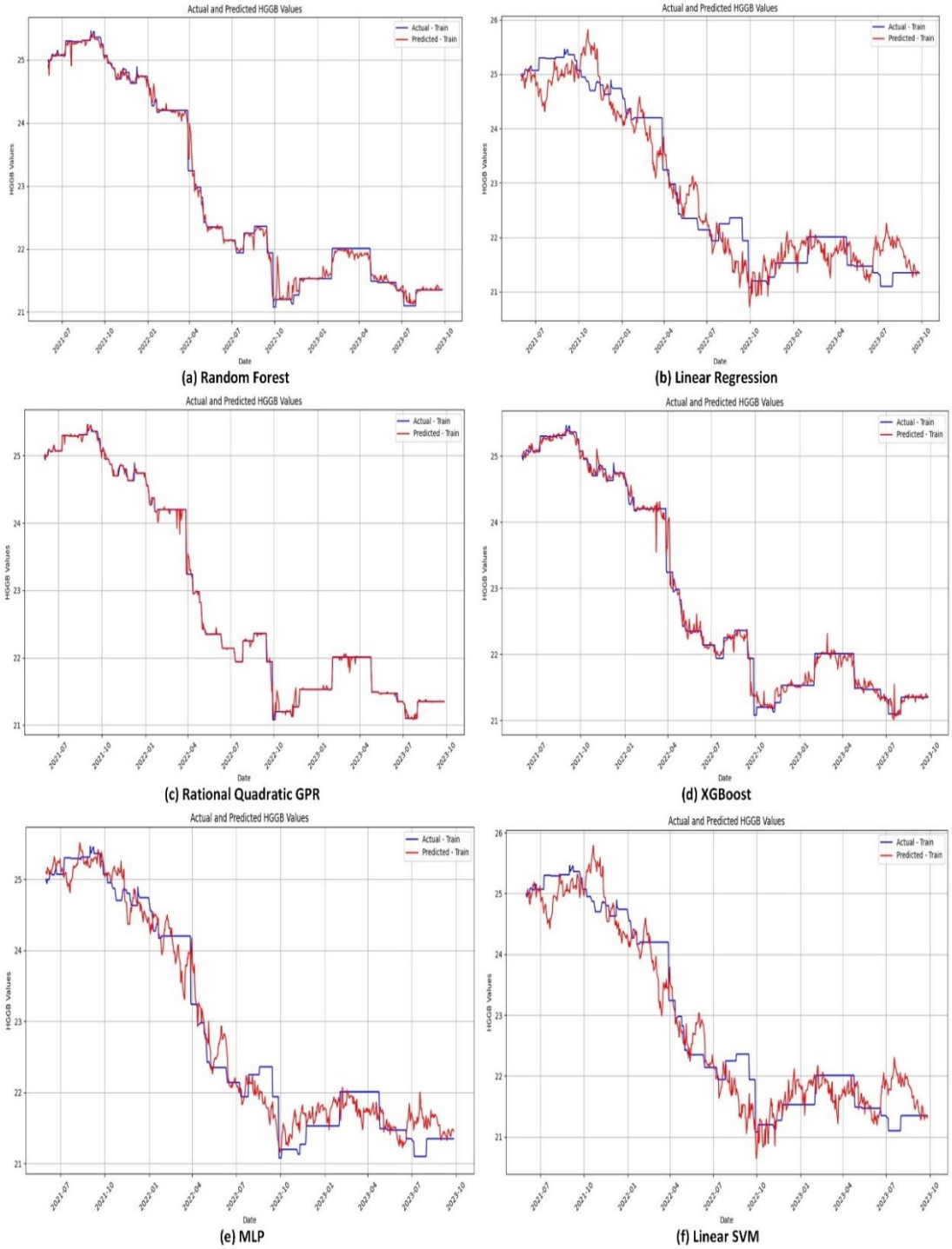

Figure 4. Prediction Graphs of the Training Set of the Models

Table 3. Metric Measurements of Training Results of the Models

	Random Forest	Linear Regression	Rational Quadratic GPR	XGBoost	MLP	Linear SVM
RMSE	0.0125	0.0942	0.0057	0.0076	0.0721	0.0945
MSE	0.0010	0.0088	0.0001	0.0006	0.0052	0.0089
MAE	0.0093	0.0736	0.0021	0.0018	0.0572	0.0749
MAPE	0.0029	0.0145	0.0004	0.0006	0.0113	0.0147
R^2	0.9969	0.9184	0.9999	0.9998	0.9520	0.9175

Additionally, the metric measurements of the training results of the models, as presented in Table 3, provide a detailed view of each model's performance. With an RMSE of 0.0057, MSE of 0.0001, MAE of 0.0021, MAPE of 0.0004, and R^2 of 0.9999, the Rational Quadratic GPR model performs the best, showing that the predicted values almost exactly match the actual values. Similarly, with an RMSE of 0.0076, MSE of 0.0006, MAE of 0.0018, MAPE of 0.0006, and R^2 of 0.9998, the XGBoost model did very well. With an RMSE of 0.0125, MSE of 0.0010, MAE of 0.0093, MAPE of 0.0029, and R^2 of 0.9969, the Random Forest model likewise showed great performance. In contrast, the MLP model demonstrated strong performance with an RMSE of 0.0721, MSE of 0.0052, MAE of 0.0572, MAPE of 0.0113, and R^2 of 0.9520. The Linear Regression model performed quite well, although it struggled to capture non-relativistic correlations with RMSE = 0.0942, MSE = 0.0088, MAE = 0.0736, MAPE = 0.0145, and R^2 = 0.9181. The Linear SVM model showed an average performance with RMSE = 0.0945, MSE = 0.0089, MAE = 0.0749, MAPE = 0.0147 and R^2 = 0.9175. The Rational Quadratic GPR model was the most effective model in predicting the green bond index. In general, the numerical results revealed that the XGBoost and Random Forest models also performed well. Figure 5 below reports the prediction graphs of the models on the test set.

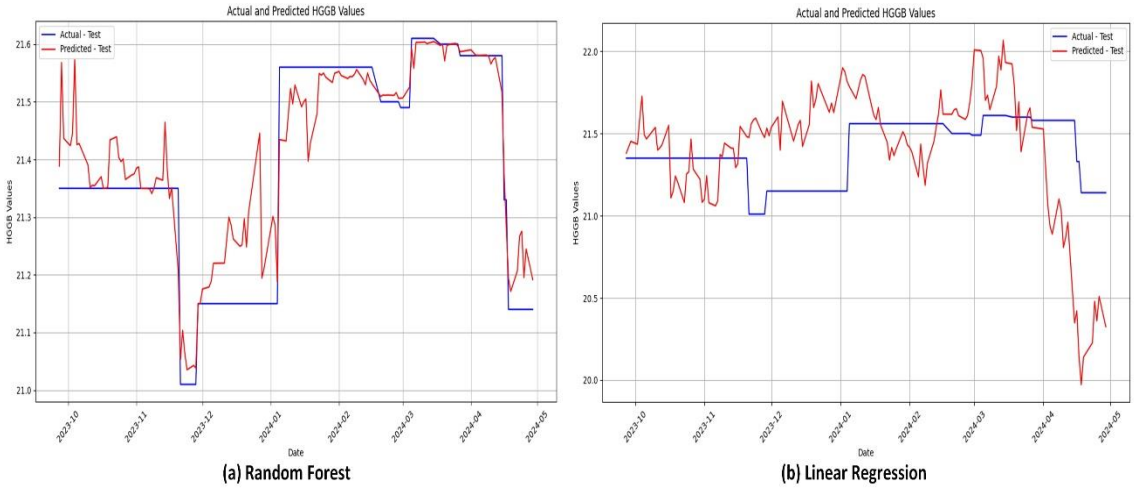

Figure 5. Prediction Graphs of the Test Set of the Models

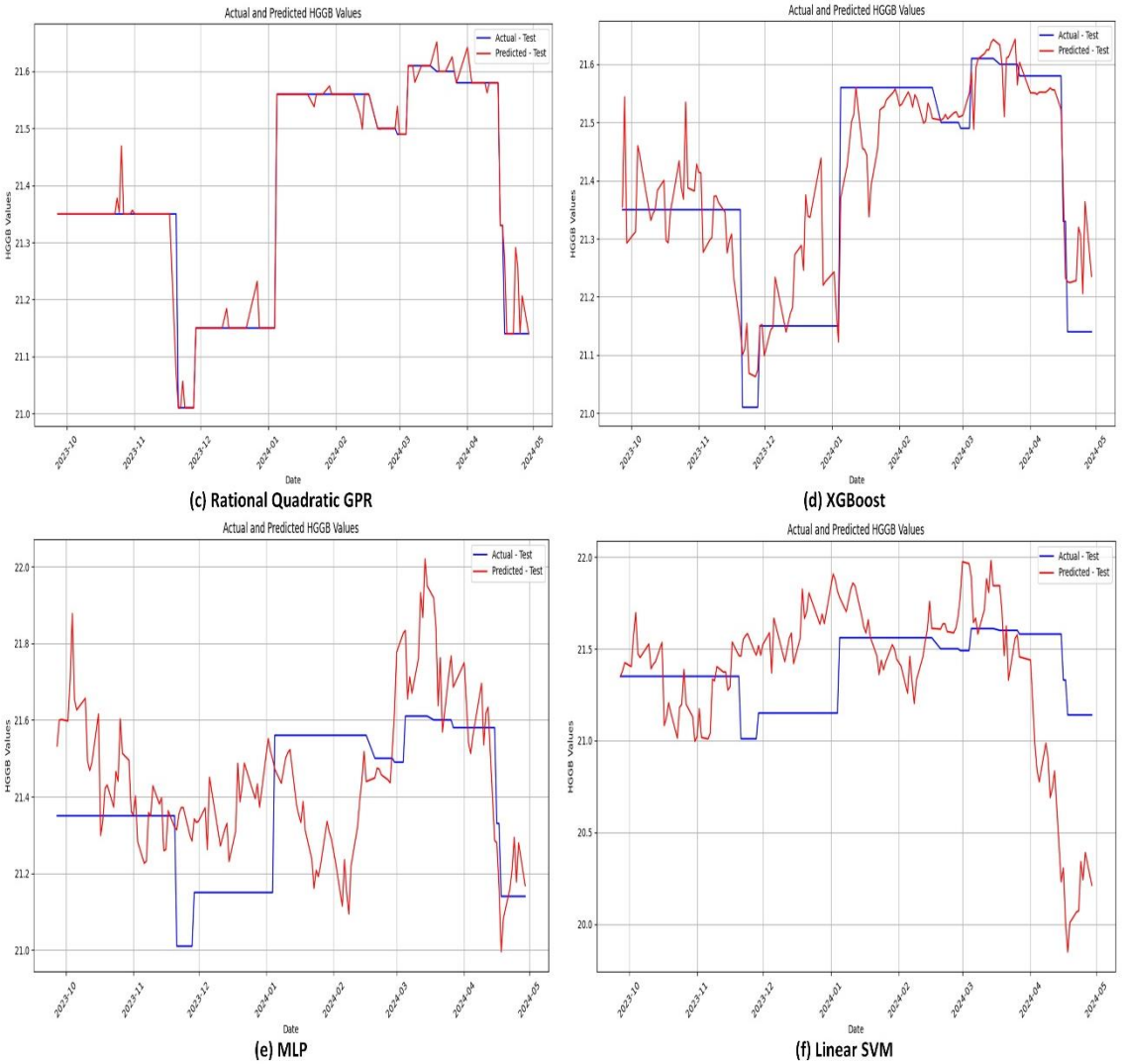

Figure 5. Continued

The graphs presented in Figure 5 illustrate test outcomes for the various models used to predict green bond index. Specifically, these graphs plot the changes in realized and predicted over time. It was determined that the most performing model was the Rational Quadratic GPR model, as the lines connecting realized and anticipated values show a very accurate forecast. In addition, the Random Forest and XGBoost models had very close to reality-predicted values on the test data. The MLP model, which was used to find non-linear relationships in data, presents very accurate forecasts. Finally, the Linear Regression and Linear SVM models showed that the predicted values were significantly different because the models were unable to find non-linear relationships properly. These experiments once again revealed that the most suitable model for forecasting green bond index is Rational Quadratic GPR, followed by other two models from the previous tests, XGBoost and Random Forest. Furthermore, the MLP model is one of the most performing models as well, while the Linear Regression and Linear SVM provide the weakest results. It should be noted that the results offered in Table 4 are used to arrive at these conclusions.

Table 4. Metric Measurements of Test Results of the Models

	Random Forest	Linear Regression	Rational Quadratic GPR	XGBoost	MLP	Linear SVM
RMSE	0.0305	0.1019	0.0260	0.0324	0.0781	0.1026
MSE	0.0010	0.0103	0.0006	0.0011	0.0061	0.0105
MAE	0.0178	0.0818	0.0181	0.0187	0.0645	0.0837
MAPE	0.0035	0.0160	0.0035	0.0037	0.0127	0.0163
R^2	0.9897	0.9181	0.9946	0.9885	0.9510	0.9172

A thorough understanding of each model's performance may be obtained from the metric measures of the test results of the models shown in Table 4. Among the models, the Rational Quadratic GPR model performs the best, with R^2 of 0.9946, RMSE of 0.0260, MSE of 0.0006, MAE of 0.0181, and MAPE of 0.0035. This demonstrates very high prediction accuracy by demonstrating a very little discrepancy between the actual and projected numbers. The XGBoost model also performed extremely well, with an R^2 of 0.9885, RMSE of 0.0324, MSE of 0.0011, MAE of 0.0187 and MAPE of 0.0037. The Random Forest model performed well with an RMSE of 0.0305, MSE of 0.0010, MAE of 0.0178, MAPE of 0.0035, and R^2 of 0.9897. MLP model — RMSE of 0.0781, MSE of 0.0061, MAE of 0.0645, MAPE of 0.0127 and 0.9510 R^2, which is not as good as two other models but still a good result. The average Linear Regression and Linear SVM models (respectively) achieved RMSE values of 0.1019 and 0.1026, MSE values of 0.0103 and 0.0105, MAE values of 0.0818 and 0.0837, MAPE values of 0.0160 and 0.0163, and R^2 values of 0.9181 and 0.9172. These results illustrated that the Rational Quadratic GPR model is the most suitable to establish predictive models for the green bond index, with Random Forest, and XGBoost being the second and third effective models respectively, and the MLP method is effective but slightly worse than the two above models, Lasso regression, linear regression, linear SVM presented mediocre performance.

In this study, time series cross-validation is used to evaluate the model while preserving the temporal structure of the data. n_splits=5 is chosen and the dataset is split into five sequential folds, where each training set progressively contains past observations and the corresponding validation set consists of future data points immediately following the training period. This approach avoids data leakage by maintaining the chronological order of the data and simulating real-world forecasting scenarios where future observations are not available during model training (Montaño et al., 2020). As a result, time series cross-validation provides a reliable assessment of the model's generalization performance over time and its predictive ability in unseen future periods (Blossier et al., 2017). The scores, obtained from time series cross-validation results in Table 5, are shown which again indicates the generalization competencies of our models.

Table 5. Metric Measurements of Time Series Cross-Validation Results of the Models

	Random Forest	Linear Regression	Rational Quadratic GPR	XGBoost	MLP	Linear SVM
RMSE	0.0499	0.0985	0.0442	0.0704	0.0736	0.0985
MSE	0.0026	0.0097	0.0021	0.0051	0.0055	0.0097
MAE	0.0329	0.0750	0.0267	0.0542	0.0580	0.0745
MAPE	0.0065	0.0148	0.0052	0.0107	0.0114	0.0148
R^2	0.9745	0.9072	0.9798	0.9507	0.9475	0.9074

Table 5 presents the results of the time series cross-validation for six models, evaluated using multiple performance metrics: RMSE, MSE, MAE, MAPE, and R^2. The Rational Quadratic GPR achieved the best performance across all metrics, with the lowest RMSE (0.0442), MSE (0.0021), MAE (0.0267), and MAPE (0.0052), along with the highest R^2 (0.9798), indicating its superior generalization capability and accuracy in capturing the patterns within the data. The Random Forest model also demonstrated strong performance, achieving an R^2 of 0.9745 and competitive error values, with an RMSE of 0.0499 and MAPE of 0.0065, positioning it as the second-best performer. In contrast, the Linear Regression and Linear SVM models yielded the highest error values, such as RMSE of 0.0985 and MAPE of 0.0148, with relatively lower R^2 values of 0.9072 and 0.9074, respectively, indicating their limited ability to handle the complexity of the dataset. The XGBoost and MLP models achieved intermediate performance, with R^2 values of 0.9507 and 0.9475, respectively, and moderate error metrics. These results demonstrate that nonlinear and ensemble-based models, such as Rational Quadratic GPR and Random Forest, are better suited for the underlying time series data due to their capacity to capture complex relationships, while linear models show relatively lower performance in comparison.

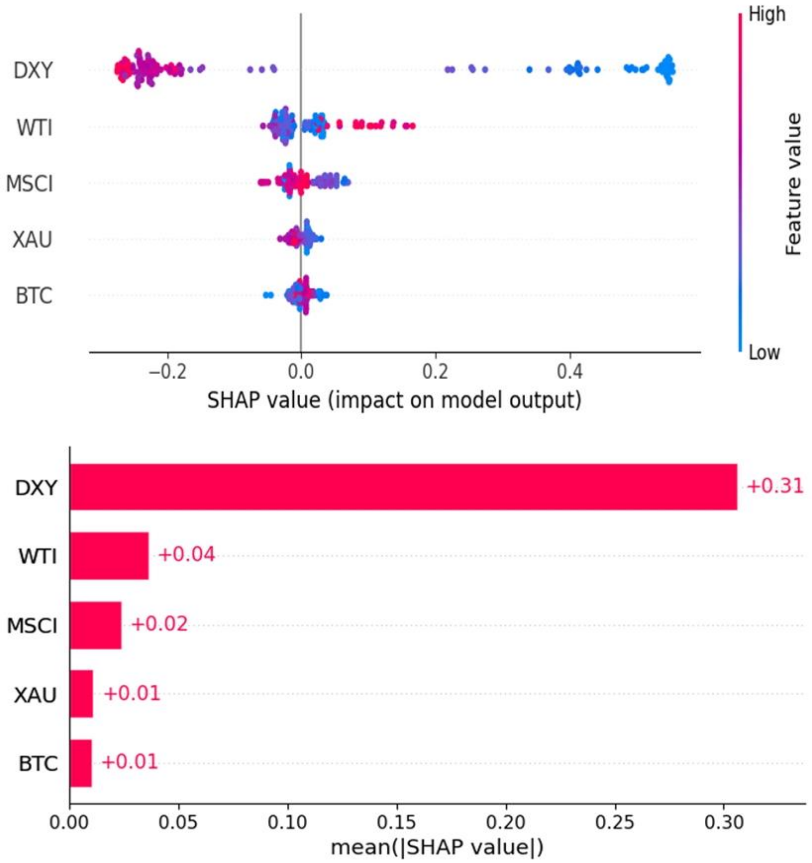

Figure 6. SHAP Analysis Results of Rational Quadratic Gaussian Process Regression Model

Figure 6 presents the results of the SHAP analysis, illustrating the feature importance assigned by the Rational Quadratic GP regression model and the impact of each independent variable on the model's predictions. In the SHAP summary plot (top), the horizontal axis represents the SHAP value, indicating the marginal impact of each feature on the model output,

while the colour indicates the feature value (red for high values and blue for low values). In the SHAP bar chart (below), the features are ranked by their average absolute SHAP value, highlighting their relative importance in the model. The DXY (US Dollar Index) stands out as the most influential determinant in the model, with the largest average absolute SHAP value. The wide dispersion of SHAP values for the DXY indicates its complex and non-linear relationship with the model's predictions. Importantly, high DXY values (red dots) tend to have a negative impact on model performance, while low DXY values (blue dots) generally have a positive impact. WTI (crude oil price) is the second most important variable but with a more limited impact compared to DXY. In contrast to DXY, the SHAP values for WTI are more concentrated around zero. However, high WTI values (red dots) tend to have a positive effect, while low WTI values (blue dots) tend to reduce model performance. The narrower range of SHAP values for WTI suggests a more consistent but weaker impact on predictions. The MSCI index shows a relatively small impact, with SHAP values clustered around zero. Both high and low values of the MSCI (red and blue dots) do not show a consistent direction of influence, indicating its limited contribution to the model. Similarly, the gold price (XAU) and bitcoin price (BTC) are the least influential variables, with SHAP values largely clustered around zero. In summary, a high WTI helps model forecasts, while a low WTI has a negative impact. High DXY values negatively affect model forecasts, while low DXY values have a positive effect. The MSCI index has a minimal and inconsistent influence, while XAU shows a distinct pattern where high values negatively impact performance, and low values positively influence it. BTC remains the least influential variable, with negligible contribution to model predictions.

5. Conclusion, Discussion and Recommendations

This study utilizes various artificial intelligence-supported machine learning methods to forecast the green bond index. Green bonds are crucial in financing sustainable and environmentally friendly projects, making accurate price predictions highly significant for investors and policymakers. By using many machine learning models to better capture the oscillations and nonlinearities in the green bond market, this study advances the discipline.

This is distinctive to the study in that it explains the working mechanism of the Rational Quadratic GPR model because of SHAP analysis and exemplifies higher results compared to the rest of the studies resembling the same kind. In the training and testing stages, the trained Rational Quadratic GPR model was found to be the best model among the models used in this study. Based on the cross-validation results, the Rational Quadratic GPR model has the lowest error rates along with the highest accuracy in RMSE, MSE, MAE, MAPE, and R^2 metrics. This model found success because it able the nonlinear, bond index. Our model accurately captured extreme events in the green bond market, including bursts of energy price shocks and COVID-19-induced economic turbulence over the period of the dataset.

The SHAP analysis shows that economic, macro, and financial characteristics such as DXY (US Dollar Index) and WTI (West Texas Oil) are important in influencing the model forecasts. This analysis has explained the features that have more weight and will affect the forecast results. A high DXY or WTI helps model forecasts, while a low DXY or WTI has a negative impact. Similarly, higher values of the MSCI increase forecasts, while lower values decrease them. BTC (Bitcoin Spot Price) and XAU (Gold Spot Price) affect the forecasts in a more balanced way and with a smaller effect area than the others. These findings are important in explaining how the

model makes predictions and which factors are more important for predicting the value movement of the green bond.

The results of the study are in line with prior findings in the literature. Studies by Wang et al. (2022a) and Çetin (2022) show how good artificial intelligence models predict the values of green bonds. As the predictions from these experiments have demonstrated, the rational quadratic GPR model is better as it can incorporate nonlinear interactions. In addition, the findings are consistent with Reboredo (2018) and Abakah et al. (2022), which delve into relationships between the green bond market and other related financial assets dynamism. However, the research does have some limitations. For one, assessments run over a larger time frame might offer a better evaluation of the models' performance as the data set covers only limited time. Next up, other methods could again improve the hyperparameterization of the models. Wider and diversified datasets need to be used in the future for the evaluation of the models as well as newer optimization techniques. Additionally, a deeper dive into other factors that can drive the value of green bonds could create a model more precise.

In conclusion, this research aims to exhibit the potential of machine learning-based tools with AI support in predicting the value of green bonds, and provides prospective areas of research using this research. Importantly, the Rational Quadratic GPR model was competent at predicting fluctuations and outliers in the green bond market, while identifying nonlinear relationships, one of the study's principal findings. This indicates that this model can be used for a general financial analysis and investment selection initiative.

This study provides a basis for comprehending the dynamics of green bond markets via advanced machine learning approaches; yet, several opportunities for additional investigation remained. Future studies should examine the effects of currency volatility by standardizing all data into a uniform currency, such as USD. Although our research indicates negligible effects of CAD/USD fluctuations over the study period, precise currency adjustments may heighten sensitivity. Furthermore, expanding the dataset to encompass extended timeframes and incorporating a broader array of green bond indexes will enhance the validity of the results, particularly during times of economic instability. Ultimately, investigating supplementary factors like interest rates or ESG measures, together with employing sophisticated machine learning techniques such as ensemble learning, may enhance predictive accuracy and broaden the usefulness of these models in sustainable finance. By focusing on these aspects, subsequent research might enhance the comprehension of green bond dynamics and refine financial forecasting instruments.

Declaration of Research and Publication Ethics
This study which does not require ethics committee approval and/or legal/specific permission complies with the research and publication ethics.

Researcher's Contribution Rate Statement
The authors declare that they have contributed equally to the article.

Declaration of Researcher's Conflict of Interest
There is no potential conflicts of interest in this study.

References

Abakah, E.J.A., Tiwari, A.K., Sharma, A. and Mwamtambulo, D.J. (2022). Extreme connectedness between green bonds, government bonds, corporate bonds and other asset classes: Insights for portfolio investors. *Journal of Risk and Financial Management*, 15(10), 477. https://doi.org/10.3390/jrfm15100477

Ampomah, E.K., Qin, Z. and Nyame, G. (2020). Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. *Information*, 11(6), 332. https://doi.org/10.3390/info11060332

Bisht, R.K. and Bisht, I.P. (2022). Investigation of the role of test size, random state, and dataset in the accuracy of classification algorithms. In H. Sharma, V. Shrivastava, K.K. Bharti and L. Wang (Eds.), *Communication and Intelligent Systems* (pp. 715-726). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2100-3_55

Black, A.J., Klinkowska, O., McMillan, D.G. and McMillan, F.J. (2014). Forecasting stock returns: Do commodity prices help? *Journal of Forecasting*, 33(8), 627-639. https://doi.org/10.1002/for.2314

Blossier, B., Bryan, K.R., Daly, C.J. and Winter, C. (2017). Shore and bar cross-shore migration, rotation, and breathing processes at an embayed beach. *Journal of Geophysical Research: Earth Surface*, 122(10), 1745–1770. https://doi.org/10.1002/2017JF004227

Boughrara, H., Chtourou, M., Ben Amar, C. and Chen, L. (2016). Facial expression recognition based on a mlp neural network using constructive training algorithm. *Multimedia Tools and Applications*, 75, 709-731. https://doi.org/10.1007/s11042-014-2322-6

Bouri, E., Çepni, O., Gabauer, D. and Gupta, R. (2021a). Return connectedness across asset classes around the COVİD-19 outbreak. *International Review of Financial Analysis*, 73, 101646. https://doi.org/10.1016/j.irfa.2020.101646

Bouri, E., Demirer, R., Gupta, R. and Wohar, M.E. (2021b). Gold, platinum and the predictability of bond risk premia. *Finance Research Letters*, 38, 101490. https://doi.org/10.1016/j.frl.2020.101490

Breiman, L. (2017). *Classification and regression trees.* London: Routledge.

Broadstock, D.C. and Cheng, L.T.W. (2019). Time-varying relation between black and green bond price benchmarks: Macroeconomic determinants for the first decade. *Finance Research Letters*, 29, 17-22. https://doi.org/10.1016/j.frl.2019.02.006

Chai, S., Chu, W., Zhang, Z., Li, Z. and Abedin, M.Z. (2022). Dynamic nonlinear connectedness between the green bonds, clean energy, and stock price: The impact of the COVİD-19 pandemic. *Annals of Operations Research*, 1-28. https://doi.org/10.1007/s10479-021-04452-y

Chen, S., Tao, F., Pan, C., Hu, X., Ma, H., Li, C., … and Wang, Y. (2020). Modeling quality changes in pacific white shrimp (litopenaeus vannamei) during storage: Comparison of the arrhenius model and random forest model. *Journal of Food Processing and Preservation*, 45(1), e14999. https://doi.org/10.1111/jfpp.14999

Chen, Y., Rogoff, K. and Rossi, B. (2010). Can exchange rates forecast commodity prices? *Quarterly Journal of Economics*, 125(3), 1145-1194. https://doi.org/10.1162/qjec.2010.125.3.1145

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297. https://doi.org/10.1007/BF00994018

Çetin, D.T. (2022). Green bonds in climate finance and forecasting of corporate green bond index value with artificial intelligence. *Journal of Research in Business*, 7(1), 138-157. https://doi.org/10.54452/jrb.992368

D'Amato, V., D'Ecclesia, R. and Levantesi, S. (2022). ESG score prediction through random forest algorithm. *Computational Management Science*, 19(2), 347-373. https://doi.org/10.1007/s10287-021-00419-3

Devereux, M.B. (2008). Much appreciated? The rise of the Canadian Dollar, 2002-2008. *Review of Economic Analysis*, 1(1), 1-33. https://doi.org/10.15353/rea.v1i1.1477

Dorfleitner, G., Utz, S. and Zhang, R. (2022). The pricing of green bonds: External reviews and the shades of green. *Review of Managerial Science*, 16, 797-834. https://doi.org/10.1007/s11846-021-00458-9

Dutta, A., Bouri, E. and Noor, H. (2021). Climate bond, stock, gold, and oil markets: Dynamic correlations and hedging analyses during the COVİD-19 outbreak. *Resources Policy*, 74, 102265. https://doi.org/10.1016/j.resourpol.2021.102265

Dutta, A., Kumar, S.S. and Basu, M. (2020). A gated recurrent unit approach to bitcoin price prediction. *Journal of Risk and Financial Management*, 13(2), 23. https://doi.org/10.3390/jrfm13020023

Ehlers, T. and Packer, F. (2017). Green Bond Finance and Certification. *BIS Quarterly Review*, September 2017, 89–104. Retrieved from https://www.bis.org/quarterlyreviews/index.htm?m=158

Gao, P., Zhang, R. and Yang, X. (2020). The application of stock index price prediction with neural network. *Mathematical and Computational Applications*, 25(3), 53. https://doi.org/10.3390/mca25030053

Gao, Y. and Zhang, J. (2023). Investigating financialization perspective of oil prices, green bonds, and stock market movement in COVİD-19: Empirical study of E7 economies. *Environmental Science and Pollution Research*, 30(23), 64111-64122. https://doi.org/10.1007/s11356-023-26808-6

Gevorkyan, M.N., Demidova, A.V., Demidova, T.S. and Sobolev, A.A. (2019). Review and comparative analysis of machine learning libraries for machine learning. *Discrete and Continuous Models and Applied Computational Science*, 27(4), 305-315. https://doi.org/10.22363/2658-4670-2019-27-4-305-315

Ghoshray, A. and Pundit, M. (2021). Economic growth in China and its impact on international commodity prices. *International Journal of Finance & Economics*, 26(2), 2776-2789. https://doi.org/10.1002/ijfe.1933

Hao, J. and Ho, T.K. (2019). Machine learning made easy: A review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics*, 44(3), 348-361. https://doi.org/10.3102/107699861983224

Harrell, F.E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis.* Berlin: Springer.

Hong, H. and Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, 105(3), 473-490. https://doi.org/10.1016/j.jfineco.2012.04.005

Hung, N.T. (2021). Green bonds and asset classes: New evidence from time-varying copula and transfer entropy models. *Global Business Review*. Advance online publication. https://doi.org/10.1177/09721509211034095

Huynh, T.L.D., Hille, E. and Nasir, M. (2020). Diversification in the age of the 4th industrial revolution: The role of artificial intelligence, green bonds and cryptocurrencies. *Technological Forecasting and Social Change*, 159, 120188. https://doi.org/10.1016/j.techfore.2020.120188

Idilbi-Bayaa, Y. and Qadan, M. (2021). Forecasting commodity prices using the term structure. *Journal of Risk and Financial Management*, 14(12), 585. https://doi.org/10.3390/jrfm14120585

Jabeur, S.B., Mefteh-Wali, S. and Viviani, J. (2024). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 334(1-3), 679-699. https://doi.org/10.1007/s10479-021-04187-w

Jia, H. (2021). Deep learning algorithm-based financial prediction models. *Complexity*, 2021, 560886. https://doi.org/10.1155/2021/5560886

Kehoe, M., Chun, K.P. and Baulch, H.M. (2015). Who smells? Forecasting taste and odor in a drinking water reservoir. *Environmental Science & Technology*, 49(18), 10984-10992. https://doi.org/10.1021/acs.est.5b00979

Lebelle, M., Jarjir, S.L. and Sassi, S. (2020). Corporate green bond issuances: An international evidence. *Journal of Risk and Financial Management*, 13(2), 25. https://doi.org/10.3390/jrfm13020025

Li, G., Liu, X., Wang, M., Yu, T., Ren, J. and Wang, Q. (2022). Predicting the functional outcomes of anti-LGI1 encephalitis using a random forest model. *Acta Neurologica Scandinavica*, 146(2), 137-143. https://doi.org/10.1111/ane.13619

Li, W., Yin, Y., Quan, X. and Zhang, H. (2019). Gene expression value prediction based on XGBoost algorithm. *Frontiers in Genetics*, 10, 1077. https://doi.org/10.3389/fgene.2019.01077

Liu, D., Zhang, X., Zheng, T., Shi, Q., Cui, Y., Wang, Y., … and Liu, L. (2021). Optimisation and evaluation of the random forest model in the efficacy prediction of chemoradiotherapy for advanced cervical cancer based on radiomics signature from high-resolution T2 weighted images. *Archives of Gynecology and Obstetrics*, 303(3), 811-820. https://doi.org/10.1007/s00404-020-05908-5

Liu, Y., Chen, H., Zhang, L., Wu, X. and Wang, X.J. (2020). Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *Journal of Cleaner Production*, 272, 122542. https://doi.org/10.1016/j.jclepro.2020.122542

Lundberg, S.M. and Lee, S-I. (2017). *A unified approach to interpreting model predictions.* Paper presented at the 31st Conference on Neural Information Processing System. Long Beach, CA, USA. Retrieved from https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Lundberg, S.M., Erion, G.G. and Lee, S.I. (2018). Consistent individualized feature attribution for tree ensembles. *ArXiv*, abs/1802.03888. https://doi.org/10.48550/arxiv.1802.03888

Maneejuk, P., Zou, B. and Yamaka, W. (2023). Predicting Chinese stock prices using convertible bond: An evidence-based neural network approach. *Asian Journal of Economics and Banking*, 7(3), 294-309. https://doi.org/10.1108/ajeb-08-2023-0080

Marín-Rodríguez, N.J., González-Ruiz, J.D. and Botero, S.B. (2022). Dynamic relationships among green bonds, CO_2 emissions, and oil prices. *Frontiers in Environmental Science*, 10, 992726. https://doi.org/10.3389/fenvs.2022.992726

Montaño, J., Coco, G., Antolínez, J.A.Á., Beuzen, T., Bryan, K.R., Cagigal, L., … and Vos, K. (2020). Blind testing of shoreline evolution models. *Scientific Reports*, 10(1), 2137. https://doi.org/10.1038/s41598-020-59018-y

Muñoz, P., Orellana-Alvear, J., Willems, P. and Célleri, R. (2018). Flash-flood forecasting in an andean mountain catchment-Development of a step-wise methodology based on the random forest algorithm. *Water*, 10(11), 1519. https://doi.org/10.3390/w10111519

Naeem, M.A., Mbarki, I., Alharthi, M., Omri, A. and Shahzad, S.J.H. (2021). Did COVİD-19 impact the connectedness between green bonds and other financial markets? Evidence from time-frequency domain with portfolio implications. *Frontiers in Environmental Science*, 9, 657533. https://doi.org/10.3389/fenvs.2021.657533

Pentoś, K., Mbah, J.T., Pieczarka, K., Niedbała, G. and Wojciechowski, T. (2022). Evaluation of multiple linear regression and machine learning approaches to predict soil compaction and shear stress based on electrical parameters. *Applied Sciences*, 12(17), 8791. https://doi.org/10.3390/app12178791

Pinkus, A. (1999), Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8, 143–195. https://doi.org/10.1017/S0962492900002919

Rasmussen, C.E. (2003). Gaussian processes in machine learning. In O. Bousquet, U. von Luxburg and G. Rätsch (Eds.), *Summer school on machine learning* (pp. 63-71). Berlin: Springer.

Reboredo, J.C. (2018). Green bond and financial markets: Co-movement, diversification and price spillover effects. *Energy Economics*, 74, 38-50. https://doi.org/10.1016/j.eneco.2018.05.030

Reboredo, J.C. and Ugolini, A. (2020). Price connectedness between green bond and financial markets. *Economic Modelling*, 88, 25-38. https://doi.org/10.1016/j.econmod.2019.09.004

Ren, Z., Yang, K. and Dong, W. (2020). Spatial analysis and risk assessment model research of arthritis based on risk factors: China, 2011, 2013 and 2015. *IEEE Access*, 8, 206406-206417. https://doi.org/10.1109/access.2020.3037912

Sadorsky, P. (2021). A random forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management*, 14(2), 48. https://doi.org/10.3390/jrfm14020048

Shahhosseini, M. (2021). Improved weighted random forest for classification problems. In T. Allahviranloo, S. Salahshour and N. Arica (Eds.), *Progress in intelligent decision science* (pp. 42-56). Papers presented at the 4th International Online Conference on Intelligent Decision Science (IDS 2020), Istanbul, Türkiye. https://doi.org/10.1007/978-3-030-66501-2_4

Shapley, L.S. (1953). *A value for n-person games. Contribution to the Theory of Games, 2.* Princeton: Princeton University Press

Sharma, H., Harsora, H. and Ogunleye, B. (2024). An optimal house price prediction algorithm: Xgboost. *Analytics*, 3(1), 30-45. https://doi.org/10.3390/analytics3010003

Simsek, A.I. (2024). Improving the performance of stock price prediction: A comparative study of random forest, XGBoost, and stacked generalization approaches. In N. Geada, R. Sood, and A. Sidana (Eds.), *Revolutionizing the global stock market: Harnessing blockchain for enhanced adaptability* (pp. 83-99). https://doi.org/10.4018/979-8-3693-1758-7.ch005

Soltani, H., Taleb, J., Ben Hamadou, F. and Boujelbène-Abbes, M. (2024). Using machine learning to forecast clean energy, commodities, green bonds and ESG index prices: How important is financial stress? *EuroMed Journal of Business,* Advance online publication. https://doi.org/10.1108/EMJB-12-2023-0341

Su, M., Zhang, Z., Zhu, Y., Zha, D. and Wen, W. (2019). Data driven natural gas spot price prediction models using machine learning methods. *Energies*, 12(9), 1680. https://doi.org/10.3390/en12091680

Sun, J., Yu, H., Zhong, G., Dong, J., Zhang, S. and Yu, H. (2020). Random shapley forests: Cooperative game-based random forests with consistency. *IEEE Transactions on Cybernetics*, 52(1), 205-214. doi: 10.1109/TCYB.2020.2972956

Tsui, A.K. and Zhang, Z. (2021). Trading macro-cycles of foreign exchange markets using hybrid models. *Sustainability*, 13(17), 9820. https://doi.org/10.3390/su13179820

Tu, C.A., Rasoulinezhad, E. and Sarker, T. (2020). Investigating solutions for the development of a green bond market: Evidence from analytic hierarchy process. *Finance Research Letters*, 34, 101457. https://doi.org/10.1016/j.frl.2020.101457

Verma, R.K. and Bansal, R. (2023). Stock market reaction on green-bond issue: Evidence from Indian green-bond issuers. *Vision*, 27(2), 264-272. https://doi.org/10.1177/09722629211022523

Wang, J., Tang, J. and Guo, K. (2022a). Green bond index prediction based on CEEMDAN-LSTM. *Frontiers in Energy Research*, 9, 793413. https://doi.org/10.3389/fenrg.2021.793413

Wang, Z., Dong, W. and Yang, K. (2022b). Spatiotemporal analysis and risk assessment model research of diabetes among people over 45 years old in china. *International Journal of Environmental Research and Public Health*, 19(16), 9861. https://doi.org/10.3390/ijerph19169861

Wei, P., Zhou, J., Ren, X. and Taghizadeh-Hesary, F. (2024). The heterogeneous role of economic and financial uncertainty in green bond market efficiency. *Review of Accounting and Finance*, 23(1), 130-155. https://doi.org/10.1108/RAF-07-2023-0202

Wilamowski, B.M. (2009). Neural network architectures and learning algorithms. *IEEE Industrial Electronics Magazine*, 3(4), 56-63. Retrieved from https://ieeexplore.ieee.org/

Wu, K., Chai, Y., Zhang, X. and Zhao, X. (2022). Research on power price forecasting based on pso-xgboost. *Electronics*, 11(22), 3763. https://doi.org/10.3390/electronics11223763

Xi, B. and Jing, H. (2021). Research on the impact of green bond issuance on the stock price of listed companies. *Kybernetes*, 51(4), 1478-1497. https://doi.org/10.1108/k-12-2020-0900

Xia, Y., Ren, H., Li, Y., Xia, J., He, L. and Liu, N. (2022). Forecasting green bond volatility via novel heterogeneous ensemble approaches. *Expert Systems with Applications*, 204, 117580. https://doi.org/10.1016/j.eswa.2022.117580

Yadav, M.P., Tandon, P., Singh, A.B., Shore, A. and Gaur, P. (2022). Exploring time and frequency linkages of green bond with renewable energy and crypto market. *Annals of Operations Research*. Advance online publication. https://doi.org/10.1007/s10479-022-05074-8

Yue, Y., Wu, Y.C., Wang, P. and Xu, J. (2021). Stock price prediction based on XGBoost and LightGBM. *E3S Web of Conferences*, 275, 01040. https://doi.org/10.1051/e3sconf/202127501040

Zaki, J., Nayyar, A., Dalal, S. and Ali, Z.H. (2022). House price prediction using hedonic pricing model and machine learning techniques. *Concurrency and Computation: Practice and Experience*, 34(27), e7342. https://doi.org/10.1002/cpe.7342

Zazoum, B. (2022). Solar photovoltaic power prediction using different machine learning methods. *Energy Reports*, 8, 19-25. https://doi.org/10.1016/j.egyr.2021.11.183

Zheng, H., Yuan, J. and Chen, L. (2017). Short-Term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies*, 10(8), 1168-1188. https://doi.org/10.3390/en10081168

Zhou, X. and Cui, Y. (2019). Green bonds, corporate performance, and corporate social responsibility. *Sustainability*, 11(23), 6881. https://doi.org/10.3390/su11236881

Zhu, Y., Duan, S., Fu, Z. and Liu, Z. (2022). Stock price crash warning in the chinese security market using a machine learning-based method and financial indicators. *Systems*, 10(4), 108. https://doi.org/10.3390/systems10040108

Zou, L., Zheng, B. and Li, X. (2017). The commodity price and exchange rate dynamics. *Theoretical Economics Letters*, 7(6), 1770-1793. https://doi.org/10.4236/tel.2017.76120