Comparative analysis of machine learning models for Indian bond market predictions: novel insights with XGBoost.

Vigneshwaraan S, T Shankar Sai, Karthika K
Department of Artificial Intelligence and Data Science
Chennai Institute of Technology, Chennai, India

ABSTRACT:
The Indian bond market has been a very important part of the total financial structure of the nation, which has been manifested by macroeconomic stability through interest rate tendencies. This research has conducted its study on machine learning on 3-year and 10-year bond prices for Indian markets and comparative analysis from a linear regression model and time-series model from ARIMA and deep Long Short-Term Memory, respectively. This study introduces an advanced machine learning algorithm in the area of bond price forecasting that is XGBoost, thereby adding new contribution to the field. Unlike previous studies that have focused mainly on bond yield predictions, this study focuses on bond price forecasting, which is important for investors and business forecasters. The model's performance was tested based on key performance metrics, including MAPE, Mean Absolute Error, Mean Squared Error, and R-squared, on historical daily bond data from 2013 to 2024. The empirical results demonstrated the superior performance of XGBoost compared to traditional and deep learning models, achieving the best accuracy with MAPE values of 0.0018 for 3-year bonds and 0.0013 for 10-year bonds. This study highlights the possible potential of using XGBoost for efficient and effective bond price prediction, making it valuable to policymakers and financial market participants.

Keywords Used: Indian Bond Market, Bond Price Prediction, Machine Learning, Deep Learning, XGBoost, ARIMA, Linear Regression, DLSTM, Time Series Forecasting.

1. INTRODUCTION

1.1 Financial Market Overview:

The financial market is one of the backbone of economic development and supports for the exchange of stocks, bonds, currencies, and derivatives. These markets supply crucial capital to businesses, governments, and individuals, allowing them to invest, save, and grow economically[5]. Among the many financial instruments, the bond market holds an important position in stabilizing economies and financing public and private initiatives. In particular, this market provides a means of financing developmental projects for governments while being a low-risk avenue for investors through investment.

The G-Sec and corporate bond markets in India are growing. Such markets, being an important price determiner in the economy for interest rates, have become an appropriate indicator of the cost incurred by the government and others while borrowing. The Indian government bond market is one of the biggest in Asia, and it is considered a safe heaven for the investors. This is basically because of the relatively lower risk involved with government securities. Moreover, government bonds are considered benchmarks to ascertain the creditworthiness of private companies.

1.2 Indian Government/Treasury Bonds:

Government Securities (G-Secs) contributes a large part in the Indian bond market and it has small amount of risk as they are covered by the Indian government. G-Secs are issued by the Reserve Bank of India on behalf of the Indian government and their availability is also with different maturity periods. Treasury Bills mature in a year or less and long-term bonds may exceed 40 years. G-Secs are dominating the Indian bond market and hence very important for understanding the dynamics of the economy [5].

The yield curve of government bonds can be said to indicate bond yields in relation to their maturities. It is very vital in providing insight into the general state of the economy. The normal upward-sloping yield curve reflects that more returns in long-term bonds compensating the risks associated with longer investment horizons. Conversely, an inverted yield curve usually indicates concern among investors over the slowing down of economic growth.

The Indian government bond market is an integral part of the setting of expectations of the investors about future interest rates. As such, it is being closely monitored by institutional investors, financial analysts, and policymakers [5].

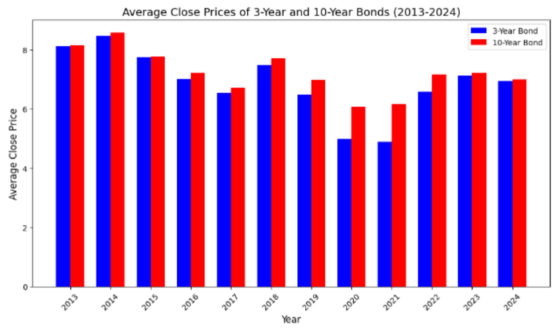

Figure 1. Average Close Price of Indian bond Market (3yr and 10yr).

Inherently, complexity is involved with the process of predicting the bond prices, which makes the usage of advanced Machine Learning and Deep Learning techniques inevitable. We make use of these methods in this work to enhance the accuracy of our predictions and provide deeper insights into the dynamics of the Indian bond market.

1.3 Motivation of Research and Contributions:

Machine learning (ML) and deep learning (DL) techniques have proven to be effective in financial forecasting, especially in volatile and complex markets. Traditional statistical methods, while used are often fail in the non-linear and non-stationary nature of financial time series data. This study builds on previous research works that utilized Linear Regression, ARIMA, and DLSTM models for Indian bond yield predictions.

As a novel contribution, this work introduces XGBoost [1],[4], a gradient-boosting algorithm to the domain of Indian bond prices forecasting and unlike previous studies that have focused mainly on bond yield predictions, this study focuses on bond price forecasting, which is important for investors and business forecasters. The study shows that the superior predictive accuracy and robustness of XGBoost are achieved by comparing its performance with traditional and deep learning models.

1.4 Research Question:

This study aims to improve the accuracy in prediction of Close prices in the Indian bond market of 3yrs and 10yrs bonds by utilizing advanced machine learning techniques. While traditional statistical models such as Linear Regression and ARIMA have been widely applied for time series forecasting, these approaches often fail to capture complex and non-linear relationships in financial data. Deep learning models, like DLSTMs improve upon existing systems by dealing with larger data sets and more complicated temporal dependencies.

As an innovation in this work, this study introduces XGBoost as an efficient gradient boosting machine learning algorithm in bond price forecasting. This means the ability of XGBoost to handle large scale problems, reduce overfitting and improve predict performance, making it an ideal candidate for financial market prediction tasks. This study aims to compare the performance of XGBoost with traditional models like Linear Regression, ARIMA, and DLSTM in predicting the close prices of 3-year and 10-year bonds in Indian government bonds.

The key research question addressed by this study is: **How effectively can XGBoost enhance bond price prediction compared to Linear Regression, ARIMA, and DLSTM models in the Indian bond market?**

It uses performance metrics like Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared to assess the models in this study, thus offering a more solid comparison of the models, together with insights into the potential for improvement in bond price forecasts with machine learning algorithms. This is the final aim, so that useful predictions can help financial market participants, policymakers, and investors make more informed decisions[1],[4].

2. LITERATURE REVIEW

Bond Market. There is an extensive history of conducting research in the bond market. The importance of the financial ecosystem is reflected in this bond market, which impacts the economic stability and the making of policies.

Accurate bond prices prediction is indispensable for making informed decisions on investors, policymakers, and the financial institution. Researchers over the years have explored ways to predict bond market trends. It ranges from simple statistical techniques to advanced ML methods.

2.1 Traditional Statistical Techniques:

Bond price forecasting can be approached by using more traditional techniques like Linear Regression and Autoregressive Integrated Moving Average (ARIMA), because of the former's simplicity and interpretability[8]. In linear regression models, one could easily discern how macroeconomic variables relate to bond close price; however, these models require linearity, which the dynamics of bond markets inherently are not.

Despite their foundational role, the traditional models often fail to address the intricate, non-linear relationships and external influences that characterize modern financial markets. These limitations have driven the adoption of machine learning techniques in bond market analysis.

2.2 Machine Learning and Deep Learning Methods:

Recent years have witnessed a lot of prominence towards machine learning and deep learning in bond market forecasts. ML technique such as XGBoost and DL technique such as deep long short-term memory (DLSTM) networks, can effectively deal with complexities within financial data [3],[6].

2.3 Gaps in Existing Research:

Despite these advancements, still there are several gaps in the literature. A significant portion of research has focused on developed markets, leaving emerging markets like India underexplored. Additionally, while macroeconomic indicators such as inflation, currency rates, interest rates, moving averages and monetary policy plays a crucial role in influencing bond prices, their integration into predictive models is often overlooked.

Comparative analyses of traditional models and machine learning models applied to bond markets are relatively scarce. Relatively few works have carefully tested the relative efficiency and accuracy of these approaches, especially within emerging economies.

2.4 Contributions of This Paper:

This work fills in this gap by combining macroeconomic indicators - inflation, interest rates, RBI monetary policy, currency rates, moving averages with bond prices data (historical data) of 3yr and 10yr bonds to enhance the predictive accuracy. It also performs a thorough comparison of Linear Regression, ARIMA, DLSTM, and XGBoost models to analyze which one is better at predicting 3-year and 10-year Indian government bond close prices. As a novel contribution, this research showcases the superior efficiency of XGBoost in bond market prediction, offering valuable insights for policymakers and market participants.

3. METHODOLOGY

The data used to develop the forecasting model for the secondary market for Indian government bonds is explained in this section. The procedure for selecting the sample, the variables utilized in the empirical analysis, and the data's descriptive statistics are presented in detail.

The dataset used for this study is sourced from **nseindia.com** and **investing.com** which is commonly used by participants in the Indian bond market. The data spans from **January 1 2013 to October 31 2024** with the sample sizes for the 10-year and 3-year bonds being 2946 and 2944 respectively. This data includes key features such as the **Close** price, **Open**, **High**, **Low**, **Change %**, **Daily Return (%)** and various macroeconomic indicators like **Interest Rates**, **SMA Indicators**, and **Currency Rates** were collected from reputable sources like the **Reserve Bank of India (RBI)** and the **World Bank**, to enhance the accuracy of the model and selected based on expert suggestions and existing literature.

The dataset's focus is on predicting the **Close** price of the bonds, which represents the market value at the end of each trading day. The model developed in this study compares the performance of advanced deep learning techniques (such as Deep Long Short-Term Memory, or DLSTM) with traditional machine learning models like Linear Regression, ARIMA, and XGBoost to forecast the bond's price. The results of this research demonstrate that XGBoost outperforms the other models, distinguishing this work from prior studies that have focused primarily on bond yield or bond spread predictions.

The study considers a variety of economic input indicators to strengthen the predictive power of the model. These include **interest rates**, **currency rates** and other macroeconomic factors as the bond and stock markets are often interrelated. Typically, as interest rates rise then bond prices tend to fall because investors shift towards bonds from stocks for more secure returns. Currency fluctuations are also closely linked to interest rate changes affecting both bond prices and market sentiment.

By incorporating these economic indicators into the prediction model, the study aims to provide a more accurate and comprehensive approach to forecasting bond prices in the Indian market.

3.1 Data Collection and Expansion:

The dataset for the Indian bond market was collected with a focus on bond price holds data from 1st January 2013 to 31st October 2024. The dataset consists of two bond types: the 3-year and 10-year government securities, with daily prices and macroeconomic data. The dataset includes key financial indicators like the closing price, open price, high price, low price, daily returns, and other macroeconomic indicators such as interest rates, currency rates, moving averages and RBI monetary policies. **Figure 1** illustrates the **Average Close Price of Indian bond Market** over the specified period.

The data was collected from reliable financial sources, including **investing.com** and **RBI**, ensuring a comprehensive view of market conditions. A total of 23,560 data points were used covering both 3-year and 10-year bonds, with the data separated into training and testing sets. **Table 1** provides a summary of the data used for model training and evaluation.

The dataset includes daily bond prices data as well as additional macroeconomic indicators that may impact the bond market such as the **Reverse Repo Rate**, **Cash Reserve Ratio**, and **Foreign Exchange Reserve**. All missing or inconsistent data points were removed to ensure the accuracy and quality of the dataset.

Table 1. Summary of Dataset for Bond Price Prediction

Name of the bond	Dataset collection period	Model	Total Data Points	Training Data	Testing Data
3-years bond	01-01-2013 to 31-10-2024	Linear Regression	2944	2355	589
		ARIMA	2944	2355	589
		DLSTM	2944	2355	589
		XGBoost	2944	2355	589
10-years bond	01-01-2013 to 31-10-2024	Linear Regression	2946	2357	589
		ARIMA	2946	2357	589
		DLSTM	2946	2357	589
		XGBoost	2946	2357	589
Total	01-01-2013 to 31-10-2024	All Models	23,560	18,848	4,712

3.2 Data Preprocessing:

The dataset used in this study consists of 3-year and 10-year Indian government bonds, collected from January 1, 2013, to October 31, 2024, at a **daily frequency**. The data preprocessing steps involved cleaning, feature engineering, and preparing the data for machine learning models. Below **Table 2** and **Table 3** represents the details of the columns used, their significance and preprocessing justifications [2].

Table 2. Features used and justification

Feature	Justification
Date	Serves as the time index for aligning bond prices and macroeconomic indicators.
Close	Target variable representing the bond's closing price, a key metric for price prediction.
Open, High, Low	Provide insights into daily price movements and market volatility.
Change %	Captures daily price changes as a percentage, reflecting market dynamics.
Daily Return (%)	Indicates the return on the bond based on daily price fluctuations, critical for financial analysis.
Volatility (7D)	Measures weekly price variability, helping capture short-term market risk.
SMA (7D, 30D)	Simple Moving Averages (7-day and 30-day) smooth price trends, aiding in capturing short- and medium-term trends.
Forward Premia of US$ (1M, 3M, 6M) (%)	Reflect currency risk and interest rate differentials affecting bond pricing.
Reverse Repo Rate (%)	Represents the rate at which the RBI borrows funds, impacting bond yields and prices.
Marginal Standing Facility (MSF) Rate (%)	Indicates short-term borrowing costs for banks, indirectly affecting market liquidity.
Bank Rate (%)	Represents the long-term lending rate from the RBI, influencing economic activity and bond prices.
Base Rate (%)	Minimum lending rate for banks, affecting overall financial conditions.
Treasury Bill Yields (91D, 182D, 364D)	Serve as benchmarks for short-term interest rates, crucial for pricing bonds.
10-Year G-Sec Yield (FBIL) (%)	Provides a benchmark for long-term interest rates, closely linked to bond prices.
Cash Reserve Ratio (CRR) (%)	Indicates the liquidity level in the banking system, influencing bond market dynamics.
Statutory Liquidity Ratio (SLR) (%)	Determines the minimum percentage of assets banks must hold, impacting market liquidity.
Policy Repo Rate (%)	Key monetary policy tool affecting borrowing costs and investment flows.
Foreign Exchange Reserves (US$ Million)	Reflects the country's economic stability, influencing investor sentiment in the bond market.

Table 3: Summary of Preprocessed Dataset

Bond Type	Time Period	Frequency	Features	Dataset Size
3-Year Bond	01-Jan-2013 to 31-Oct-2024	Daily	25 (as detailed above)	2943 rows
10-Year Bond	01-Jan-2013 to 31-Oct-2024	Daily	25 (as detailed above)	2945 rows

3.3 Designing Data with Statistics:

The dataset was carefully prepared for model training and evaluation. A total of 23,560 data points were processed with 80% allocated to training and 20% to testing and validation. The training dataset includes historical data for bond prices while the testing dataset serves to evaluate model performance. The dataset's structure helps in the development of machine learning models that can predict bond close price based on both historical price data and macroeconomic variables[2].

The data spans a period from 1st January 2013 to 31st October 2024 and was processed using various Python libraries to ensure effective feature engineering. The training data contains 18,848 points while the testing data contains 4,712 points. This design enables a robust evaluation of model performance.

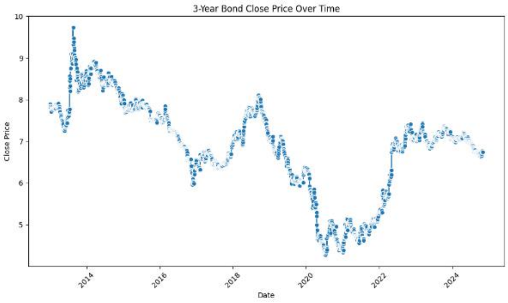

Figure 2. 3-Year Bond Close Price Over Time.

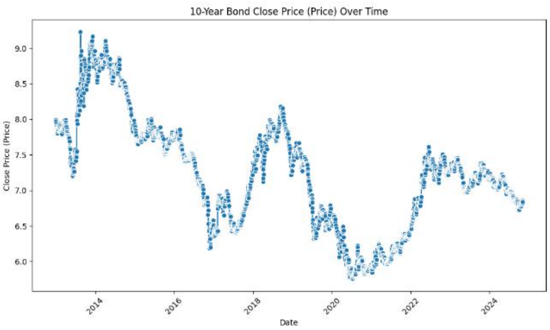

Figure 3. 10-Year Bond Close Price Over Time.

3.4 Models:

The researchers utilized various machine learning models, introducing the deep LSTM recurrent neural network, ARIMA model, and linear regression to capture enhanced forecasting capabilities. Additionally, the inclusion of XGBoost added a novel perspective to this analysis.

3.4.1. Linear Regression:

Linear regression, a foundational machine learning technique, was employed to analyze the explanatory power of input features by assigning weights to numerical variables. The general form of the multiple linear regression model is given by:

$$y = \beta 0 + \ \beta 1X1 + \beta 2X2 + .. + \varepsilon$$

Where,
y = Dependent variable (response variable)
$\beta 0$ = y intercept
$\beta 1$ and $\beta 2$ = slope coefficients
X1 and X2= Independent variables (predictors)
ϵ = Error term

The objective of linear regression is to minimize the cost function defined as the Mean Squared Error (MSE) which measures the average squared difference between the predicted values and the actual values.

Performance metrics such as **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, **Mean Absolute Percentage Error (MAPE)**, and **R-squared** were used to evaluate the Linear Regression model's accuracy.

3.4.2. Deep LSTM Neural Network:

The Long Short-Term Memory (LSTM) network is a type of Recurrent Neural Network (RNN) that is capable of learning and remembering both short- and long-term dependencies in sequential data. LSTM models are widely used for predicting financial time series data such as bond prices, due to their ability to capture temporal patterns in historical data.

An LSTM network consists of units that include gates: input, output and forget gates. These gates control the flow of information and enable the model to retain important information across time steps. The LSTM unit's forward pass mechanism is described by the following equations:

$$f_t = \sigma_g\ (W_f\ \times x_t + U_f\ \times h_{t-1} + b_f)$$
$$i_t = \sigma_g\ (W_i\ \times x_t + U_i\ \times h_{t-1} + b_i)$$
$$o_t = \sigma_g\ (W_o\ \times x_t + U_o\ \times h_{t-1} + b_o)$$
$$c'_t = \sigma_c\ (W_c\ \times x_t + U_c\ \times h_{t-1} + b_c)$$
$$c_t = f_t\,.\,c_{t-1} + i_t\,.\,c'_t$$
$$h_t = o_t\,.\,\sigma_c(c_t)$$

Where,
$\sigma_g = Sigmoid$
$\sigma_c = tanh$
. = element wise multiplication
f_t = Forget gate
i_t = Input gate
o_t = Output gate
c_t = Cell state
h_t = Hidden state

LSTM models being a type of RNN share the same hyperparameters as RNNs such as the number of hidden layers, units in each layer, learning rate, batch size and sequence length for input data. Hyperparameter optimization techniques are used to fine-tune these values for optimal performance. The **Table 4** represents the hyperparameters used in the DLSTM model for 3yr and 10yr bonds including the justifications.

Table 4: Hyperparameters and Justifications (DLSTM)

Hyperparameter	Value	Justification
Sequence Length (SEQ_LENGTH)	90	A 90-day sequence length captures both short-term and medium-term temporal dependencies, which are crucial for financial time series forecasting.
LSTM Units	128 per layer	The number of units is increased to 128 to allow the model to capture complex patterns and dependencies in the time series data.
Dropout Rate	0.3	Dropout helps in preventing overfitting by randomly setting 30% of the units to zero during training. This is especially important for models with many parameters.
Learning Rate	0.0003	A smaller learning rate ensures gradual updates to the model weights, helping achieve convergence without overshooting the optimal point.
Batch Size	32	A smaller batch size provides a balance between training efficiency and convergence stability, particularly in noisy financial data.
Epochs	200	A high number of epochs allows the model to learn effectively over multiple iterations, while early stopping prevents overfitting.
Early Stopping Patience	15	The patience of 15 epochs ensures that the training stops early if no improvement is observed in the validation loss, saving computational resources.
Loss Function	Mean Squared Error	MSE penalizes large errors more heavily, which is crucial for minimizing the deviation between predicted and actual bond prices.
Optimizer	Adam	The Adam optimizer is chosen for its adaptive learning rate and ability to handle sparse gradients, which is beneficial for time series data.

For the bond price prediction model, performance metrics like **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, **Mean Absolute Percentage Error (MAPE)**, and **R-squared** were used to evaluate the LSTM model's performance.

3.4.3. ARIMA

The Auto-Regressive Integrated Moving Average (ARIMA) model is a well-established approach for time series forecasting often used to predict bond prices in financial markets. ARIMA models are effective for capturing temporal dependencies in data and are commonly used for modeling stationary time series data.

For bond price prediction, the ARIMA model accounts for both autoregressive (AR) and moving average (MA) components. The model uses past values of the time series and past forecast errors to predict future values. The AR component is represented by an autoregressive model, while the MA component captures the impact of previous error terms.

The ARIMA model can be expressed as follows:

$$y_t = \beta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + .. + \alpha_p y_{t-p}$$

Where,
y_t = the next timestamp value which will be forecasted,
β = the intercept term or a constant
α_i = the coefficient of lagged observation
y_{t-i} = the lagged observation
p = the number of lag orders

3.4.4. XGBoost

XGBoost (Extreme Gradient Boosting) is an efficient and scalable machine learning algorithm widely used for regression tasks. It works by building an ensemble of decision trees, where each tree is trained to correct the errors made by the previous ones. This boosting process helps improve the accuracy of predictions [1],[4].
For predicting the **Close** price of bonds, we used XGBoost to model the relationship between historical bond data (such as **Close**, **Open**, **High**, and **Low** prices) and other relevant financial indicators. The algorithm minimizes the prediction error by optimizing the following objective function:

$$L(\emptyset) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

The loss function which in our case is the Mean Squared Error (MSE) and $\Omega(f)$ is the regularization term that helps control the complexity of the model.

Table 5: Hyperparameters and Justifications (XGBOOST)

Hyperparameter	Value	Justification
Objective	reg:squarederror	Specifies the regression task, minimizing the squared error for continuous target variables like bond prices.
n_estimators	200	A moderate number of estimators ensures a balance between performance and computational cost.
learning_rate	0.1	A learning rate of 0.1 provides a balance between convergence speed and avoiding overfitting.
max_depth	7	Allows the model to capture moderately complex patterns in the data without excessive overfitting.
subsample	0.8	Uses 80% of the training data for each tree, reducing overfitting and improving generalization.
colsample_bytree	0.9	Utilizes 90% of the features for each tree, ensuring diverse trees and capturing important features.
random_state	42	Ensures reproducibility of results by fixing the random seed.

3.5 Evaluation Metrics:

In this study, the performance of the machine learning (ML) and deep learning (DL) models used for bond price prediction is evaluated using key evaluation metrics: **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, **Mean Absolute Percentage Error (MAPE)** and **R-squared (R^2)**. These metrics provide a comprehensive understanding of the accuracy, reliability, and overall fit of the models.

3.5.1. Mean Squared Error (MSE)

- Definition: MSE is the average of the squared differences between the predicted values and the actual values.

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

Where:
  - MSE = mean squared error
  - N = number of data points
  - y_i = observed values
  - $\hat{y}_i$ = predicted values
- Purpose: MSE gives an idea of how well the model's predictions match the actual values. A lower MSE indicates a better model, as it means the model's predictions are closer to the actual values. The squaring of the errors penalizes larger deviations more heavily, making MSE sensitive to outliers.

3.5.2. Mean Absolute Error (MAE):

- MAPE expresses the prediction error as a percentage of the actual values, providing an easy-to-understand measure of prediction accuracy. It is particularly useful for comparing model performance across different datasets or scales. A lower MAPE indicates better predictive accuracy.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_l - y_i|$$

Where:
  - MAE = mean absolute error
  - N = number of observations
  - y_i = actual value
  - $\hat{y}_l$ = predicted value

3.5.3. Mean Absolute Percentage Error (MAPE):

- MAPE expresses the prediction error as a percentage of the actual values, providing an easy-to-understand measure of prediction accuracy. It is particularly useful for comparing model performance across different datasets or scales. A lower MAPE indicates better predictive accuracy.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_l - y_i}{y_i}\right| \times 100$$

Where:
  - MAPE = mean absolute percentage error
  - N = number of observations
  - y_i = actual value
  - $\hat{y}_l$ = predicted value

3.5.4. R-squared (R^2):

- R^2, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is explained by the independent variables in the model. An R^2 value closer to 1 indicates that the model explains a large portion of the variance, while a value closer to 0 suggests that the model does not explain much of the variance.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

Where:
  - $\bar{y}$ = mean of actual value
  - y_i = actual value
  - $\hat{y}_i$ = predicted value

4. RESULTS AND DISCUSSIONS

This research compares the result of various algorithms in terms of accuracy and performance in bond prices forecasting for the Indian bond market. This study evaluates the simulation results of four models: Linear Regression, Autoregressive Integrated Moving Average (ARIMA), Deep Long Short-Term Memory (DLSTM) neural networks and Extreme Gradient Boosting (XGBoost) using statistical error metrics such as MAE, MAPE, MSE, and R^2 to measure their predictive accuracy. According to the analysis, all models could reasonably predict the prices of the bond with significant deviations in their level of performance.

The results show that Linear Regression and ARIMA could only capture very general trends about bond price moves, failing when the complexity increased. DLSTM and XGBoost models captured non-linear interactions and temporal dependence much better and therefore outperformed.

The XGBoost model was significantly the best at all conditions with lower MAE, MAPE, MSE and R^2. These indicate that XGBoost is indeed the most robust and efficient in dealing with the high-order non-linear dynamics and interactions that exist in the bond market data.

In the present study, the predictive models that take as input financial and macroeconomic data make a comprehensive forecast for bond prices, which in turn, makes it easy to make sound investment decisions.

4.1 MODEL ACCURACY:

In this section, the accuracy of the different models for the forecast of 3-year and 10-year bond prices is determined. The comparative performance of the models Linear Regression, ARIMA, DLSTM and XGBoost, is taken based on certain key evaluation metrics, such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and R^2. These measures give a picture of each model's capacity to predict accurately bond prices while keeping an eye on the discrepancy between predicted prices for bonds and actual prices, which means the effectiveness of models at explaining the variations in the moving of their respective prices[7].

Table 6. Results summary of Linear Regression, ARIMA, Deep LSTM and XGBoost models for 3-year and 10-year government bond price prediction.

Methods	Metrics	MAPE	MAE	MSE	R squared
Linear regression	3 yrs	0.0108	0.0762	0.0092	0.7008
	10 yrs	0.0114	0.0817	0.0077	0.7743
Arima	3 yrs	0.0199	0.1391	0.0321	-
	10 yrs	0.0362	0.2551	0.0968	-
DLSTM	3 yrs	0.0090	0.0633	0.0078	0.7556
	10 yrs	0.0062	0.0450	0.0035	0.8894
XGBoost	3 yrs	0.0018	0.0753	0.0101	0.9890
	10 yrs	0.0013	0.0398	0.0026	0.9938

The XGBoost model had the lowest MAPE 0.0018 for 3-year and 0.0013 for 10-year bond predictions, hence the most accurate model. The MAPE values shows that XGBoost gave the closest approximations of the actual bond prices, far outperforming the other models. In comparison, Linear Regression and ARIMA exhibited relatively higher MAPE values, particularly for the 10-year bond, suggesting that these models were less capable of capturing the underlying trends and fluctuations in bond prices.

The DLSTM model also did well, especially for the 10-year bond, with a marked improvement over Linear Regression and ARIMA, having a MAPE of 0.0062. This shows that DLSTM can capture temporal dependencies and long-term trends in bond prices but was still lagging behind XGBoost in terms of accuracy.

Overall, MAPE values show superiority in the performance of XGBoost as the best model for the study about forecasting bond prices.

4.2 Results Visualization and Discussion:

To better illustrate the accuracy and performance of the forecasting models, the following visualizations and tables were generated. These graphs show the predictions made by each model for both the 3-year and 10-year bond prices, highlighting the differences in forecast accuracy and the model's ability to capture the underlying price movements.

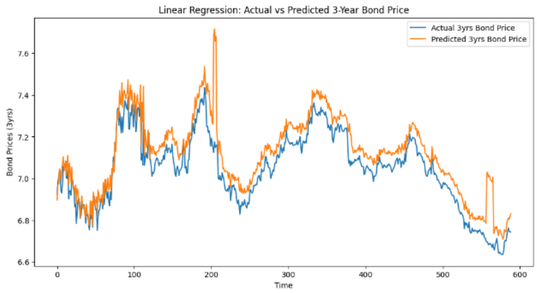

Figure 4. Predicted vs. Actual Bond Prices for Linear Regression (3-Year)

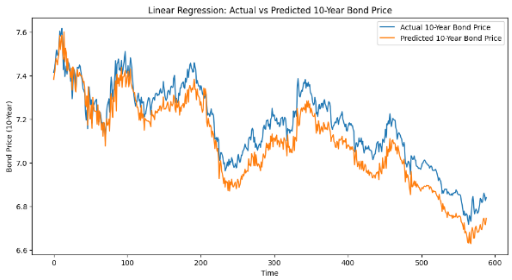

Figure 5. Predicted vs. Actual Bond Prices for Linear Regression (10-Year)

The **Linear Regression** model is illustrated in **Figure 4** and **Figure 5** for both the **3-year** and **10-year** bond prices respectively. The graphs reveal that **Linear Regression** struggles to capture the volatility and complex trends in the bond prices. While it tracks the general trend, the model's predictions deviate significantly from the actual prices, especially for the **10-year** bond. The predicted values (orange lines) are smoother and do not follow the fluctuations observed in the actual prices (blue lines), resulting in higher error metrics such as MAPE.

- For the **3-year bond**, the model shows moderate performance but fails to capture the detailed fluctuations, leading to a MAPE of 0.0108.
- For the **10-year bond**, the **Linear Regression** model shows a more pronounced mismatch with the actual values, as seen in the higher MAPE of 0.0114.

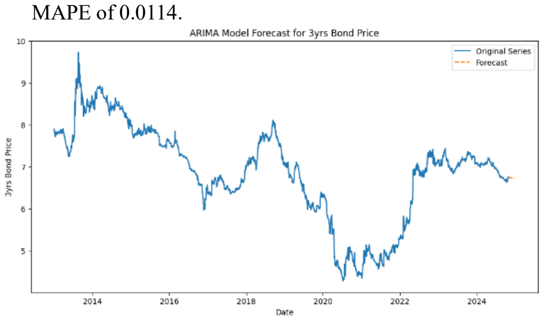

Figure 6. Predicted vs. Actual Bond Prices for ARIMA (3-Year)

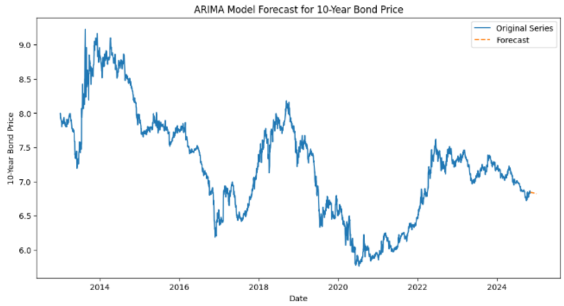

Figure 7. Predicted vs. Actual Bond Prices for ARIMA (10-Year)

Figure 6 and **Figure 7** presents the predictions from the **ARIMA** model for both the **3-year** and **10-year** bond prices respectively. **ARIMA** models time-series data by considering previous data points to forecast future values. However, **ARIMA** has limitations in capturing the non-linear relationships and temporal dependencies in the bond price data.

- For the **3-year bond**, **ARIMA** captures some trends but struggles with larger fluctuations, as indicated by a higher MAPE of 0.0199.
- For the **10-year bond**, **ARIMA** performs even worse, with a significantly higher MAPE of 0.0362. The predictions (orange lines) deviate more noticeably from the actual prices, especially during volatile periods.

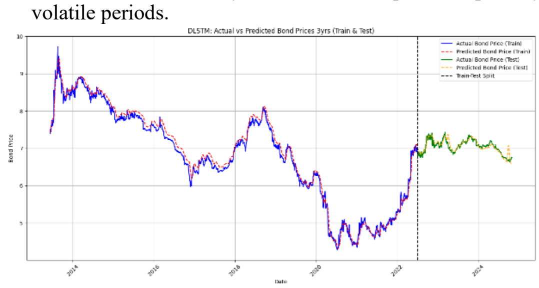

Figure 8. Predicted vs. Actual Bond Prices for DLSTM (3-Year)

Figure 9. Predicted vs. Actual Bond Prices for DLSTM (10-Year)

Figure 8 and **Figure 9** showcases the **DLSTM** model's predictions for both the **3-year** and **10-year** bond prices respectively. **DLSTM** (Deep Long Short-Term Memory) is capable of capturing long-term dependencies and trends in time-series data, which makes it more effective than **Linear Regression** and **ARIMA** in modeling the fluctuations in bond prices.

- For the **3-year bond**, the **DLSTM** model shows improved accuracy compared to **Linear Regression** and **ARIMA**, with a lower MAPE of 0.0090. The predicted values (orange line) follow the actual prices (blue line) more closely, though still not as well as **XGBoost**.
- For the **10-year bond**, **DLSTM** demonstrates a clear improvement, especially in terms of accuracy, with a MAPE of 0.0062. The predicted values align well with the actual prices, but there are still slight deviations during some volatile periods.

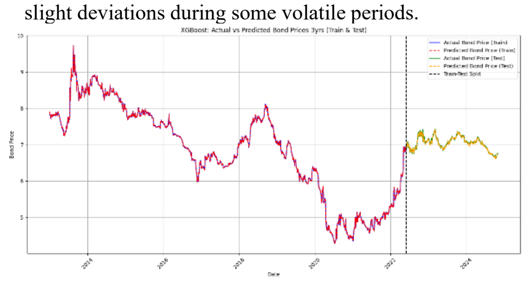

Figure 10. Predicted vs. Actual Bond Prices for XGBoost (3-Year)

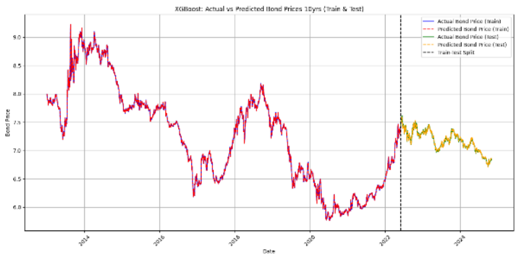

Figure 11. Predicted vs. Actual Bond Prices for XGBoost (10-Year)

Figure 10 and **Figure 11** displays the predictions from the **XGBoost** model for both the **3-year** and **10-year** bond prices. **XGBoost** is a powerful gradient boosting algorithm known for its ability to capture complex non-linear relationships and interactions between features. It consistently outperforms the other models in terms of prediction accuracy.

- For the **3-year bond**, **XGBoost** shows exceptional performance, with predictions that closely follow the actual prices. The MAPE of 0.0018 indicates the model's ability to capture both short-term and long-term trends effectively.
- For the **10-year bond**, **XGBoost** further excels, achieving a MAPE of 0.0013. The predicted values are almost identical to the actual values, indicating that **XGBoost** is highly effective at modeling bond price movements, even in more volatile scenarios.

4.3 Actual vs. Predicted Values:

In this section, we present a summary table showing the actual bond prices alongside the predicted bond prices for the 3-year and 10-year bonds. The table includes the results for all four models: Linear Regression, ARIMA, DLSTM, and XGBoost. These values highlight the differences between the true bond prices and the predicted values for each model, helping to assess their performance further.

Table 7. Actual vs. Predicted Bond Prices (3yr):

Date (3yr)	Actual Price (3yr)	Predicted Price (3yr) - Linear Regression	Predicted Price (3yr) - ARIMA	Predicted Price (3yr) - DLSTM	Predicted Price (3yr) - XGBoost
01-06-2022	6.925	6.999790	7.062807	6.987325	6.937465
12-07-2022	6.887	6.961380	7.024051	6.948983	6.899397
25-08-2022	6.915	6.989682	7.052608	6.977235	6.927447
07-10-2022	7.338	7.417250	7.484026	7.404042	7.351208
22-11-2022	7.104	7.180723	7.245370	7.167936	7.116787
02-01-2023	7.146	7.223177	7.288205	7.210314	7.158863
10-02-2023	7.188	7.265630	7.331041	7.252692	7.200938
27-03-2023	7.125	7.201950	7.266788	7.189125	7.137825
15-05-2023	6.925	6.999790	7.062807	6.987325	6.937465
23-06-2023	7.010	7.085708	7.149499	7.073090	7.022618
04-08-2023	7.161	7.238339	7.303504	7.225449	7.173890
18-09-2023	7.179	7.256533	7.321862	7.243611	7.191922
01-11-2023	7.319	7.398045	7.464648	7.384871	7.332174
14-12-2023	7.144	7.221155	7.286166	7.208296	7.156859
29-01-2024	7.043	7.119064	7.183156	7.106387	7.055677
12-03-2024	7.039	7.115021	7.179076	7.102351	7.051670
30-04-2024	7.188	7.265630	7.331041	7.252692	7.200938
13-06-2024	6.993	7.068524	7.132161	7.055937	7.005587
26-07-2024	6.839	6.912861	6.975096	6.900551	6.851310

Table 8. Actual vs. Predicted Bond Prices (10yr):

Date (10yr)	Actual Price (10yr)	Predicted Price (10yr) - Linear Regression	Predicted Price (10yr) - ARIMA	Predicted Price (10yr) - DLSTM	Predicted Price (10yr) - XGBoost
01-06-2022	7.414	7.498520	7.682387	7.459967	7.423638
12-07-2022	7.391	7.475257	7.658554	7.436824	7.400608
25-08-2022	7.292	7.375129	7.555970	7.337210	7.301480
07-10-2022	7.460	7.545044	7.730052	7.506252	7.469698
22-11-2022	7.285	7.368049	7.548717	7.330167	7.294471
02-01-2023	7.340	7.423676	7.605708	7.385508	7.349542
10-02-2023	7.363	7.446938	7.629541	7.408651	7.372572
27-03-2023	7.312	7.395357	7.576694	7.357334	7.321506
15-05-2023	7.001	7.080811	7.254436	7.044406	7.010101
23-06-2023	7.074	7.154644	7.330079	7.117859	7.083196
04-08-2023	7.193	7.275000	7.453387	7.237597	7.202351
18-09-2023	7.149	7.230499	7.407794	7.193324	7.158294
01-11-2023	7.358	7.441881	7.624360	7.403620	7.367565
14-12-2023	7.194	7.276012	7.454423	7.238603	7.203352
29-01-2024	7.171	7.252749	7.430590	7.215460	7.180322
12-03-2024	7.026	7.106096	7.280341	7.069561	7.035134
30-04-2024	7.195	7.277023	7.455459	7.239609	7.204354
13-06-2024	6.986	7.065640	7.238893	7.029313	6.995082
26-07-2024	6.940	7.019116	7.191228	6.983028	6.949022

Table 7 and **Table 8** present the comparison of actual and predicted daily bond prices for the 3-year and 10-year bonds using four different models: Linear Regression, ARIMA, DLSTM, and XGBoost. The results in these tables show the actual bond prices alongside the predicted values generated by each model for each respective date. The primary difference between the actual and predicted prices for the 3-year and 10-year bond samples is discussed, highlighting the variations in performance across the models.

5. CONCLUSION

Implications: This research provides a comprehensive exploration of machine learning and deep learning models applied to the prediction of bond prices in the Indian bond market. By comparing the performance of various models—Linear Regression, ARIMA, DLSTM, and XGBoost—we demonstrate the potential of these techniques in forecasting bond price movements, which are crucial for investors and policymakers.

Findings and Novelty: The findings of our research shown that the XGBoost model gave the best performance for both the 10-year and 3-year bond price predictions. For the 10-year bond, XGBoost attained a MAPE of 0.0013 and an R-squared of 0.9938, which outperformed all other models. Similarly, for the 3-year bond, XGBoost showed better accuracy by attaining a MAPE of 0.0018 and an R-squared of 0.9890, outperforming all other models.

Limitations: The study has some limitations. While the results of the study are promising, the dataset used from 2013 to 2024 can be expanded with real-time data or sentiment analysis from news sources to further enhance prediction accuracy. Moreover, the research focuses on the Indian bond market, and the generalizability of the models to other countries or markets remains to be tested.

Future Research: This work provides a strong foundation for future research in bond market prediction. Future studies may focus on hybrid models that integrate deep learning with other financial data sources, such as sentiment analysis, global economic indicators, or real-time market data. Extending the scope to other financial assets and markets can provide more insights into the broader application of machine learning techniques in finance.

REFERENCES

[1] Authors: Hassan Oukhouya, Hamza Kadiri, Elhimdi Khalid, Raby Guerbaz. (2024) "Forecasting International Stock Market Trends: XGBoost, LSTM, LSTM-XGBoost, and Backtesting XGBoost Models" https://www.researchgate.net/publication/376951516_Forecasting_International_Stock_Market_Trends_XGBoost_LSTM_LSTM-XGBoost_and_Backtesting_XGBoost_Models.

[2] Author: Salisu Garba Abdullahi (2024) "Boosting Nigeria's Bond Market: Evidence from Macroeconomic Perspective"

https://www.researchgate.net/publication/386555653_Boosting_Nigeria%27s_Bond_Market_Evidence_from_Macroeconomic_Perspective.

[3] Authors: Dharika Kapil and Kannan Yamini (2024). "Machine Learning Applications in Predicting Bond Market Trends: Explore How Machine Learning Algorithms Can Be Used to Predict Movements in the Fixed Bond Market". DOI:10.20944/preprints202404.1667.v1 .

[4] Authors: Yan Zhang, Lin Chen (2021). "A Study on Forecasting the Default Risk of Bond Based on XGboost Algorithm and Over-Sampling Method". https://doi.org/10.4236/tel.2021.112019.

[5] Authors: Wisniewski, T. P. (2016). "Is there a link between politics and stock returns? A literature survey. International Review of Financial Analysis". https://doi.org/10.2139/ssrn.2690852.

[6] Authors: Bhaskar Nandi, Subrata Jana, Krishna Pada Das (2024). "Machine learning-based approaches for financial market prediction: A comprehensive review". https://doi.org/10.59400/jam.v1i2.134.

[7] Authors: Robert Verner, Michal Tkáč sr., Michal Tkáč jr. (2021). "Improving Quality of Long-Term Bond Price Prediction Using Artificial Neural Networks". https://doi.org/10.12776/QIP.V25I1.1532.

[8] Authors: Bhaskar Nandi, Subrata Jana, Krishna Pada Das (2024). "Machine learning-based approaches for financial market prediction: A comprehensive review". https://doi.org/10.59400/jam.v1i2.134.

[9] Authors: Paul, M., & Reddy, K. S. (2022). "US QE and the Indian bond market. Journal of Quantitative Economics". https://doi.org/10.1007/s40953-021-00257-9.

[10] Authors: Wahidin, D., Akimov, A., & Roca, E. (2021). "The impact of bond market development on economic growth before and after the global financial crisis: Evidence from developed and developing countries. International Review of Financial Analysis". https://doi.org/10.1016/j.irfa.2021.101865.