Cureus Journal of Computer Science

Open Access Original Article

Predictive Models for Foreign Exchange Reserve Dynamics of India

Rohit D. Poul [1] , Ankush D. Sawarkar [1] , Aditya S. Baheti [1] , Ubaid Ahmed Shaikh [1] , Anurag Agrahari [2] , Atul Halmare [3]

1. Department of Information Technology, Shri Guru Gobind Singhji Institute of Engineering and Technology (SGGSIE&T), Nanded, IND 2. Department of Computer Science and Engineering, Visvesvaraya National Institute of Technology (VNIT), Nagpur, IND 3. Department of Information Technology, Jaywantrao Sawant College of Engineering, Pune, IND

Corresponding author: Rohit D. Poul, 2021bit063@sggs.ac.in

Review began 10/29/2024
Review ended 11/12/2024
Published 11/24/2024

DOI: https://doi.org/10.7759/s44389-024-01222-2

Abstract

In this research paper, we undertake an in-depth examination of the predictive capacity of various regression models in forecasting the foreign exchange reserves of India. Our study focuses on three distinct methodologies: multiple regression, polynomial regression, and ridge regression. The system architecture for this study incorporates a structured pipeline for data preprocessing, feature selection, and model evaluation, ensuring consistent handling and analysis of economic indicators that impact foreign reserves. Through meticulous analysis, we assess the accuracy of each model to provide extensive insights into their predictive efficacy. The multiple regression model, with a coefficient of determination (R^2) of 0.9317, emerges as a reliable tool for explaining the fluctuations in Indian foreign exchange reserves. Its ability to acquire the nuances of multiple predictors contributes significantly to its predictive power. Expanding upon this foundation, polynomial regression demonstrates remarkable precision, boasting an impressive R^2 value of 0.9963. This polynomial approach, characterized by its flexibility in fitting complex data patterns, showcases its superiority in capturing the intricacies of the reserve dynamics. Moreover, the ridge regression model, with an R^2 value of 0.9934, reinforces the vigorousness of our analysis. By mitigating multicollinearity and overfitting, ridge regression offers a balanced compromise between bias and variance, enhancing the overall predictive performance. These findings underscore the prominence of employing diverse regression techniques in forecasting Indian foreign exchange reserves, thereby furnishing policymakers and financial analysts with valuable insights to inform strategic decision-making processes. Our research extends beyond assessing mere numerical accuracy; we inquire into interpretability and robustness of each model, contemplating factors such as model complexity and generalizability. Such insights fortify our understanding of the predictive mechanisms underpinning foreign exchange reserve dynamics, facilitating more informed decision-making in financial policy formulation and risk management.

Categories: Data Engineering, Data Mining, Machine Learning (ML)
Keywords: rbi, machine learning, foreign reserves, polynomial regression, economic forecasting

Introduction

Foreign exchange reserves serve as a pivotal tool for central banks, ensuring sufficient liquidity and managing unexpected capital flow shifts. By collecting historical data on relevant factors and corresponding reserves over a defined timeframe, a regression model can be constructed to elucidate their relationship, facilitating future reserve level predictions based on variable fluctuations. This endeavor holds significant implications for policymakers, investors, and financial analysts, offering insights paramount for maintaining financial solidity, exchange rate equilibrium, and overall economic well-being [1]. Informed decisions regarding monetary policy, exchange rate interventions, and international trade negotiations rely on reliable reserve forecasts. Addressing the core issue of forecasting India's foreign currency holdings, this study employs multiple regression analysis on data furnished by the Reserve Bank of India (RBI). By examining various economic indicators, the research endeavors to accurately anticipate India's future foreign investment landscape. The prominence of such forecasts lies in their profound significance on financial stability, monetary policy efficacy, exchange rate dynamics, and overall economic resilience. A nation's foreign exchange reserves play a pivotal role in navigating external shocks, managing balance sheets, and exerting influence in currency markets to maintain stability [2].

Precise forecasts of future foreign exchange rates empower policymakers to craft astute strategies concerning monetary policy adaptations, interventions in exchange rates, and the formulation of international trade policies. For instance, if the predictive model signals a forthcoming decline in foreign exchange reserves, policymakers may contemplate methods to entice foreign investments or tweak interest rates to maintain currency stability. Consequently, through the functioning of regression analysis and data sourced from the RBI, this research endeavors to equip policymakers with informed insights for navigating India's foreign exchange reserve landscape adeptly [3].

How to cite this article
Poul R D, Sawarkar A D, Baheti A S, et al. (November 24, 2024) Predictive Models for Foreign Exchange Reserve Dynamics of India. Cureus J Comput Sci 1 : es44389-024-01222-2. DOI https://doi.org/10.7759/s44389-024-01222-2

The study's aim is to construct a comprehensive predictive framework for India's future foreign reserves, incorporating various industrial constraints like GDP growth, inflation rates, trade balances, and interest rates. Through this intricate model, we seek to discern the distinct influence of each factor on reserve levels, illuminating the key drivers of fluctuations. By analyzing historical data from the RBI, we will assess the model's precision and dependability. Ultimately, the insights garnered from our analysis will be leveraged to offer practical recommendations to policymakers, investors, and financial institutions, with the overarching objective of fortifying economic stability, facilitating well-informed monetary decisions, and optimizing investment strategies [3].

The paper is structured into seven divisions, each serving a distinct purpose in the exploration of India's foreign exchange reserves forecasting. Beginning with the introduction, Division 1 provides crucial context regarding the value of accurate forecasting for India's economic landscape, while also delineating the study's objectives and methodology. Division 2 expounds into a comprehensive literature review, meticulously evaluating previous research on foreign investment forecasting, methodologies employed, and the array of influencing factors and policy determinants. Following this, Division 3 meticulously lays out the way for data collection, variable selection, and statistical analysis, essential for constructing robust regression models. Division 4 details the experimental design, elucidating the methods, software tools, and parameter settings utilized for the regression analysis, ensuring methodological rigor and reproducibility. In Division 5, findings are meticulously analyzed and discussed, offering perception into the implications of the study. The concluding section, Division 6, encapsulates the key findings, discusses their implications, and suggests capacity avenues for upcoming research. Finally, Division 7 houses the references section, providing a comprehensive list of all sources cited throughout the paper, thus enriching its academic credibility and facilitating further exploration of the topic.

Materials And Methods

Authors	Focus	Key Findings	Ref.
Chinn et al.	Examines medium-term determinants of current account balances in industrial and developing countries, focusing on demographics, fiscal policies, and productivity growth, with empirical insights across various economies.	The paper finds that aging populations and demographic shifts significantly impact current account balances, with government deficits linked to deficits, particularly in industrial countries. It also highlights that productivity growth enhances export competitiveness, while developing countries are more sensitive to external factors like global interest rates and commodity prices.	[4]
Kaminsky et al.	Explore the phenomenon of "twin crises," where banking crises and balance-of-payments crises occur concurrently. The paper investigates the causal relationships between these two crises and employs empirical analysis using historical data from various countries.	Identify common factors leading to twin crises, including weak economic fundamentals and fragile financial institutions. They highlight precursor indicators such as real exchange rate misalignment and rising public deficits. Their findings emphasize the importance of monitoring these indicators to inform crisis prevention strategies.	[5]
Chen et al.	The paper analyzes the implementation of monetary policy in China, emphasizing the role of the interbank market and bank lending in shaping monetary transmission mechanisms.	The authors find that the interbank market is crucial for effective monetary policy transmission, influencing liquidity and lending rates, while also highlighting that bank lending practices are significantly affected by both regulatory frameworks and market conditions.	[6]
Jeanne et al.	Present a model to determine the optimal level of international reserves for small open economies aiming to protect against sudden stops in capital flows, deriving a formula for the optimal reserves.	The model indicates that the optimal level of reserves can explain the magnitude of reserves in many emerging market countries, particularly emphasizing that significant anticipated output costs of sudden stops and high levels of risk aversion are crucial for understanding the reserve accumulation in emerging market Asia.	[7]
Kearney et al.	Reassess the existence of an emerging yen block in North and Southeast Asia, examining the influence of the Japanese yen on regional currencies and trade dynamics.	The study finds limited evidence supporting the emergence of a yen block, suggesting that while the yen plays a significant role in regional trade, currency linkages are not as strong as previously thought, with other currencies also impacting regional dynamics.	[8]
Ghosh et al.	Investigate the degree of exchange rate pass-through in India, analyzing how fluctuations in the exchange rate impact domestic prices over time.	The study finds that exchange rate pass-through in India is relatively high but has decreased over the years, indicating that domestic price adjustments to exchange rate changes have become less pronounced, potentially due to improved monetary policy frameworks and greater market competition.	[9]
Razzaque	Razzaque examines the challenges facing Bangladesh's export trade, emphasizing the need for policy reforms to accelerate growth and diversify exports.	The study identifies critical issues such as reliance on a narrow range of products, the need for improved infrastructure, and investment in skill development, suggesting that targeted policy interventions are essential for enhancing competitiveness and achieving sustainable export growth.	[10]

TABLE 1: Summary of work done for predictive models for foreign exchange reserve dynamics of India using machine learning

Foreign investment utilizing regression analysis has been an extensively studied topic in economics and finance. Preceding studies have used various methods and economic indicators to infer changes in the standard of foreign reserves, aiming to provide insight into the determinants of deposit accumulation and expenditure and to create policy notification of decisions. One common approach in the literature is to use multiple regression analysis to model the relationship between foreign investment and independent variables representing different economic baselines (Table *1*).

The study aims to address gaps in understanding foreign investment determinants by leveraging various regression methods and RBI data. Previous research has touched on this topic, but there is a need for a comprehensive investigation using RBI data. This study stands out by employing multiple regression analysis, polynomial regression, and ridge regression tailored to RBI data limitations. These advanced statistical techniques will scrutinize a broad range of economic indicators comprehensively. Additionally, existing literature may overlook interactions or nonlinear associations between monetary variables and reserve levels, leading to an incomplete understanding of dynamics. By exploring these correlations using multiple regression methods, the study aims to provide deeper insights into India's foreign investment determinants. Moreover, while previous examination have identified significant factors, they often lack rigorous evaluation of predictive accuracy. Thus, this study emphasizes assessing model performance to enhance practical utility for policymakers, investors, and stakeholders. Through this, it contributes to a nuanced understanding of India's foreign investment determinants, crucial for financial stability, policymaking, and investment decisions. The multiple linear, polynomial, and ridge regression models developed serve as bridges between existing RBI data models and foreign exchange forecasts. Considering factors like finance charges, inflation, GDP growth, and political stability simultaneously, these models offer a holistic understanding of intricate relationships. With improved accuracy, they enable reliable forecasting, identification of key variables, risk management, and straightforward interpretation, making them invaluable tools for navigating complex economic decisions [1,2].

These data were made obtainable from the official site of RBI. This dataset includes very informative and real-time data collected by RBI. This dataset includes 333 rows and 16 columns (Figure 1).

	Period	Forward Premia of US$ 1-month (%)	Forward Premia of US$ 3-month (%)	Forward Premia of US$ 6-month (%)	Reverse Repo Rate (%)	Marginal Standing Facility (MSF) Rate (%)	Bank Rate (%)	Base Rate (%)	91-Day Treasury Bill (Primary) Yield (%)	182-Day Treasury Bill (Primary) Yield (%)	364-Day Treasury Bill (Primary) Yield (%)	10-Year G-Sec Yield (FBIL) (%)	Cash Reserve Ratio (%)	Statutory Liquidity Ratio (%)	Policy Repo Rate (%)	Foreign Exchange Reserves
0	01-Mar-24	1.51	1.39	1.42	3.35	6.75	6.75	9.1	6.96	7.17	7.12	7.06	4.5	18.0	6.5	625625.83
1	23-Feb-24	1.19	1.51	1.50	3.35	6.75	6.75	9.1	7.02	7.19	7.13	7.09	4.5	18.0	6.5	619072.07
2	16-Feb-24	1.27	1.57	1.55	3.35	6.75	6.75	9.1	7.05	7.18	7.15	7.10	4.5	18.0	6.5	616097.24
3	09-Feb-24	1.36	1.60	1.55	3.35	6.75	6.75	9.1	7.01	7.15	7.11	7.11	4.5	18.0	6.5	617229.53
4	02-Feb-24	1.20	1.57	1.55	3.35	6.75	6.75	9.1	7.04	7.18	7.15	7.06	4.5	18.0	6.5	622469.42

FIGURE 1: Preview of the dataset

Let us see how our data varies and what the factors are that decide the predictions.

Independent variables

• Period: Time interval.

• Forward Premia of US$ 1-month (%): Forward premium for 1-month USD.

• Forward Premia of US$ 3-month (%): Forward premium for 3-month USD.

• Forward Premia of US$ 6-month (%): Forward premium for 6-month USD.

• Reverse Repo Rate (%): Central bank borrowing rate.

• Marginal Standing Facility (MSF) Rate (%): Emergency borrowing rate.

• Bank Rate (%): Interest rate charged by central bank.

• Base Rate (%): Minimum lending rate.

• 91-Day Treasury Bill (Primary) Yield (%): Short-term government debt yield.

• 182-Day Treasury Bill (Primary) Yield (%): Medium-term government debt yield.

• 364-Day Treasury Bill (Primary) Yield (%): Long-term government debt yield.

• 10-Year G-Sec Yield (FBIL) (%): 10-year government bond yield.

• Cash Reserve Ratio (%): Proportion of deposits banks must keep with central bank.

• Statutory Liquidity Ratio (%): Proportion of deposits banks must maintain in liquid assets.

• Policy Repo Rate (%): Central bank lending rate.

Dependent variable

• Total foreign currency holdings maintained by a central bank.

In our model, we conducted thorough data cleaning and discovered a dataset free of missing values. With no need for imputation, our focus shifted to outlier handling and duplicate removal. Employing statistical methods like *Z*-score, we identified and treated outliers, ensuring data integrity. Additionally, duplicate records were meticulously identified and eliminated. This meticulous approach fortified the dataset, laying a robust foundation for our successive multiple linear regression evaluation and predictions [11].

In our feature selection/engineering process, we have chosen to include all available features as we believe each one is important for our analysis. This comprehensive approach ensures we capture the full spectrum of factors influencing foreign investment in India accurately (Figure *2*) [11,12].

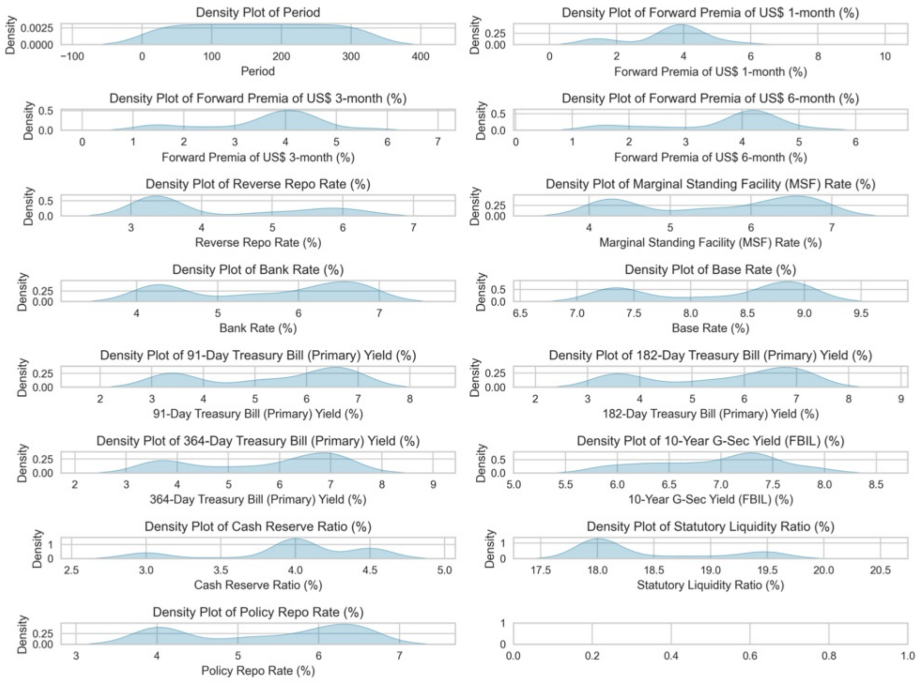

FIGURE 2: Density plot for each feature

We utilized one-hot encoding to preprocess categorical data representing periods containing dates. This categorical feature, characterized by temporal information, required special treatment to ensure compatibility with machine learning algorithms. To achieve this, we initially converted the date-based periods into categorical variables, ensuring uniform representation [13].

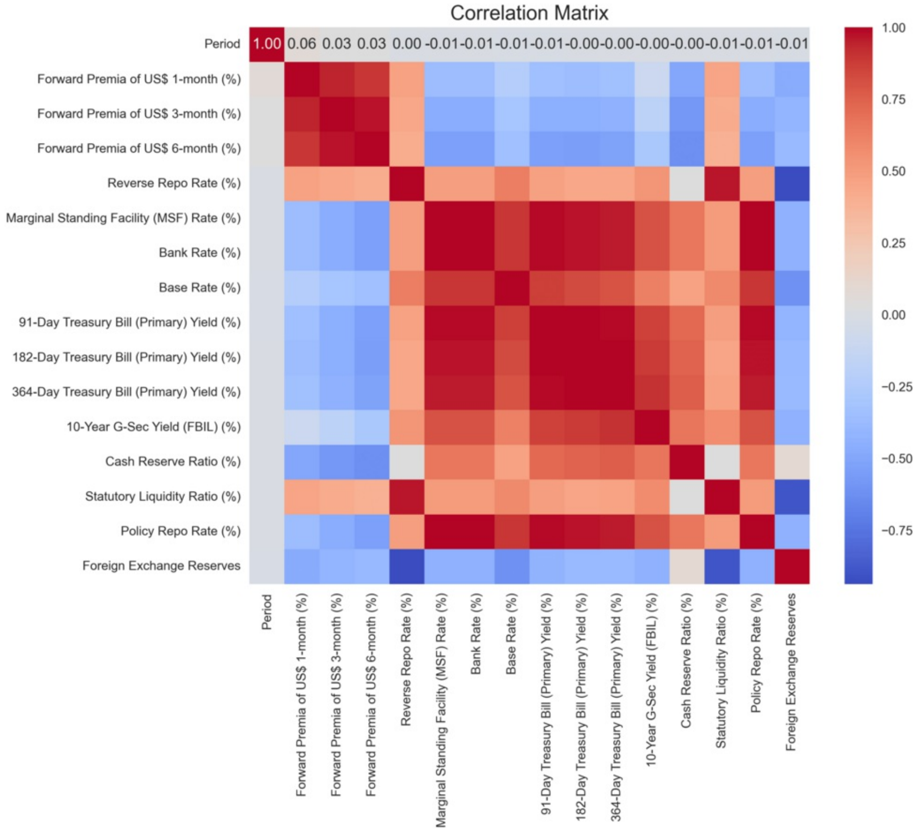

FIGURE 3: Correlation matrix

The heatmap visualization in the picture shows the correlation matrix represented in colors, where every cell is the correlation coefficient between two variables. The colors are brighter in case of strong positive correlations (e.g., red) and darker in case of strong negative correlations (e.g., blue). Having a correlation coefficient near to either +1 or −1 indicates a strong linear relationship while those around 0 mean there is little no correlation between them (Figure *3*).

In our study, we meticulously evaluated various criteria for selecting regression algorithms to ensure the appropriateness of our choices. Understanding the dataset's characteristics, such as size, item count, and the presence of missing or excessive data, guided our decision to employ three distinct regression models: multiple regression, ridge regression, and polynomial regression. This diversified approach enables us to capture both linear and nonlinear relationships between the features and the target variable effectively. Multiple regression serves as a foundational model, allowing us to examine the linear relationships among the predictors and the response variable. Its interpretability and straightforward implementation make it an essential benchmark against which the performance of more complex models can be assessed.

Ridge regression was chosen to address potential issues of multicollinearity, which can arise when predictors are highly correlated. By applying L2 regularization, ridge regression effectively reduces the variance of the estimates without significantly impacting bias, thereby enhancing the model's predictive accuracy. This feature is particularly beneficial in scenarios where we have a large number of predictors relative to the number of observations.
Polynomial regression allows us to model nonlinear relationships by incorporating polynomial terms. This capability is crucial for capturing intricate patterns in the data that a standard linear model may overlook. By comparing the performance of polynomial regression to our other models, we can assess the degree to which nonlinear relationships contribute to predicting the target variable.
We ensured that our chosen regression algorithms adhered to their underlying assumptions and validated crucial aspects such as linearity, homoscedasticity, and independence of errors, particularly for multiple regression.

To evaluate model performance comprehensively, we incorporated metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2). Following rigorous testing, our ridge regression model exhibited strong predictive performance, yielding R^2 scores of

0.9934 and 0.9963 on the testing data, and an R^2 value of 0.9317 on the training data.
In summary, our decision to utilize multiple regression, ridge regression, and polynomial regression stems from the diverse nature of our dataset and the necessity to capture both linear and nonlinear relationships effectively. This comprehensive approach facilitates a robust comparative analysis of the determinants of our target variable, allowing us to assess the strengths and weaknesses of each modeling technique in context [14]. Regression is a statistical technique used to understand the interaction between one dependent variable and two or more independent variables. The theoretical formulas for multiple regression involve the calculation of coefficients, predicted values, and various statistical measures of model performance [15]. Here are the key theoretical formulas for multiple regression, polynomial regression, and ridge regression:

Multiple regression equation

The multiple regression equation describes the association between the dependent variable Y and sovereign variables $_1, _2, \ldots,$ as follows:

$$Y = \beta_0 + \beta_1 X + \beta_2 X_2 + \beta_2 X_3 + \cdots + \beta_2 X_n + \varepsilon \qquad (1)$$

Where in equation (1)

β_0 is the intercept (the value of when all independent variables are zero).

β_1 , β_2 , … , β_n are the coefficients (slopes) of the independent variables.

ε is the error term (residuals), representing the difference between the observed and predicted values.

Polynomial regression equation

Polynomial regression is a regression analysis technique used to model the relationship betwixt the independent variable(s) and the dependent variable in a nonlinear manner. Unlike simple linear regression, where the relationship is assumed to be linear, polynomial regression allows for curved relationships between the variables [16]. The formula for polynomial regression can be expressed as:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_2 X^3 + \ldots + \beta_n X^n + \epsilon \qquad (2)$$

In equation (2), Y represents the dependent variable, X denotes the independent variable, β_0 , β_1 , β_2 ,... are the coefficients corresponding to each term in the polynomial equation, X , X^2 , X^3 , .., X^3 and ϵ represents the error term.

Ridge regression equation

Ridge regression, also known as Tikhonov regularization, is a regression analysis technique used to mitigate multicollinearity and overfitting in linear regression models. It achieves this by imposing a penalty on the size of the coefficients, thereby shrinking them toward zero while still allowing them to pitch in to the model. The formula for ridge regression can be expressed as:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y \qquad (3)$$

The equation (3) represents the vector of coefficients estimated by ridge regression, is the matrix of independent variables, is the vector of observed values for the dependent variable, λ is the regularization parameter, and is the identity matrix.

In ridge regression, the standardization parameter λ controls the degree of shrinkage applied to the coefficients. A larger value of results in greater shrinkage, which reduces the extent of the coefficients and prevents overfitting. The choice of λ is typically determined through procedures such as cross-validation.

When assessing regression models, various evaluation metrics were employed to gauge their performance and accuracy. These metrics offer numerical insights into the effectiveness of the models, aiding in comparing their predictive capabilities and identifying the most suitable one for the given task.

Mean squared error

The MSE calculates the average of the squared differences between predicted and actual ratings. Lower MSE values indicate a closer alignment between predicted and actual values, signifying better model performance [17].

\begin{equation} \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - p_i)^2 \qquad (4) \end{equation}

Mean absolute error

The MAE calculates the average absolute difference between predicted and actual ratings. Similar to the MSE, lower MAE values signify higher accuracy, indicating a smaller disparity between predicted and actual values [18]. This implies that a model with a lower MAE is better at accurately predicting outcomes, as it exhibits less deviation from the observed data points.

\begin{equation} \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - p_i| \qquad (5)\end{equation}

R-squared (R^2) score

The R^2 score procedures the extent to which the variance in app ratings can be attributed to app features. With values between 0 and 1, higher R^2 scores signify a better alignment of the model with the data, indicating a more precise depiction of how app features relate to ratings.

\begin{equation} R^2 = 1 - \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \qquad (6)\end{equation}

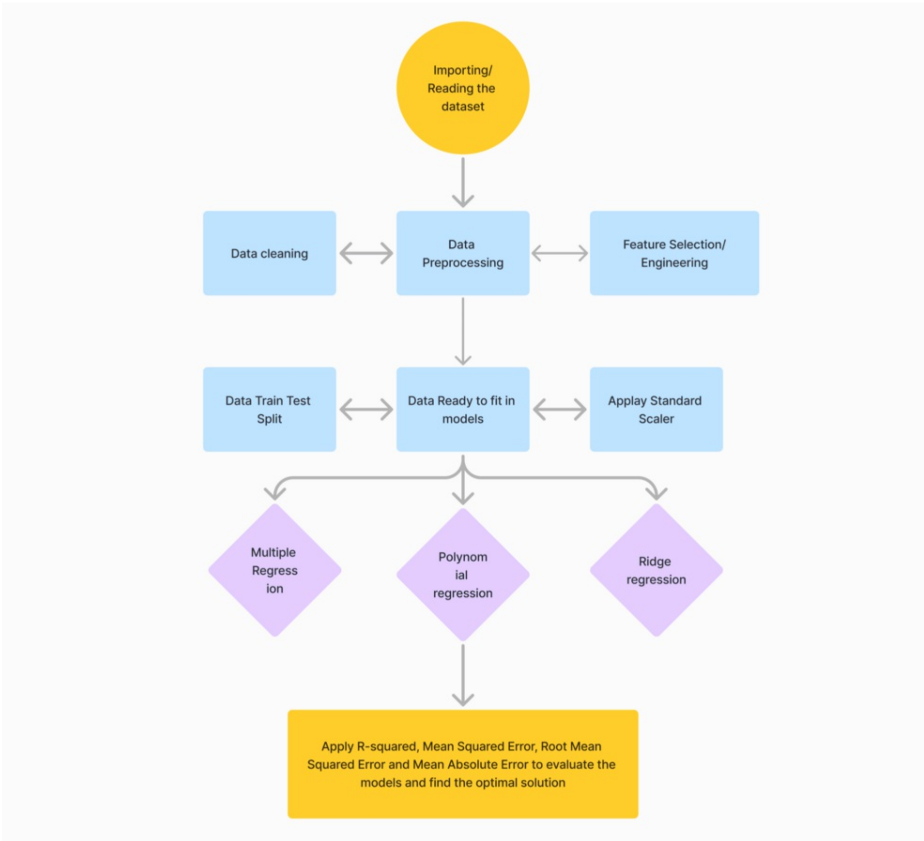

FIGURE 4: Flowchart of working of experiment

To build a regression model for predicting foreign reserves, we start by importing and cleaning the dataset to ensure data quality, as shown in Figure *4*. Next, data preprocessing and feature selection prepare the variables for analysis, focusing on economic indicators relevant to reserve levels. Standardization follows to bring features onto a similar scale, and then we split the data into training and testing sets. We experiment with multiple regression models, including linear, polynomial, and ridge regression, to find the best fit for the data. Finally, model performance is evaluated using metrics like R^2 and RMSE, helping identify the optimal model for accurate predictions (Figure *4*) [19].

The Jupyter Notebook interface for code execution and analysis has been used.

Implementation process

1) Data preparation:

Load the dataset using Pandas.

Perform any necessary data cleaning, such as handling missing values or outliers. Split the dataset into features (X) and target variable (y).

2) Data preprocessing:

If necessary, encode categorical variables employing techniques such as label encoding or one-hot encoding. Impute missing values utilizing methods like mean, median, or mode imputation. Scale the features as needed to ensure uniform scale across all features, employing techniques such as StandardScaler or MinMaxScaler.

3) Model selection:

Polynomial regression was selected as our model of choice for its capability to accurately capture nonlinear relationships in the dataset, outperforming other models in terms of predictive accuracy and fit.

Model training

Split the dataset into tutoring and testing sets using train_test_split from scikit-learn. Fit the regression model to the training data using the fit method.

4) Model evaluation:

Evaluate the trained model's performance using appropriate evaluation metrics such as MSE, MAE, or R^2 score. Plot visualizations like prediction errors or residuals to understand model performance visually.

5) Model interpretation:

Interpret the coefficients or weights of the regression model to understand the relationship between features and the target variable. Identify the most important features contributing to the prediction.

Results

The comparative analysis of the regression model elucidates marked disparities in their performance metrics, particularly the R^2 values (Table 2). This metric, representing the fraction of variance in the dependent variable explicable by the independent variables, underscores the superiority of polynomial regression, which boasts an R^2 value of 0.9963. This is marginally eclipsed by ridge regression's R^2 value of 0.9935, while multiple regression lags significantly with an R^2 of 0.9317.

Metrics	Multiple	Polynomial	Ridge
R^2	0.9317	0.9963	0.9935
Mean absolute error	18823.77	4511.04	5186.64
Mean squared error	553573495.3	29634515.20	7286.50
Root mean squared error	23528.14	5443.75	0.9934

TABLE 2: Evaluation metrics

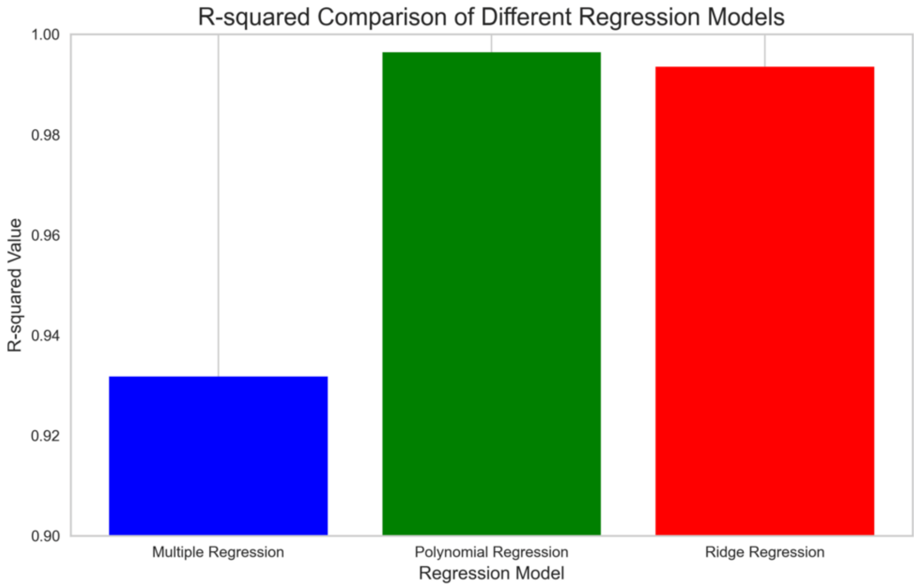

FIGURE 5: Model score comparison

Polynomial regression's preeminence is attributable to its adeptness at encapsulating complex, nonlinear interdependencies within the data, a capability that traditional multiple regression lacks (Figure 5). Ridge regression, while also demonstrating substantial predictive power and effectively mitigating multicollinearity issues, does not match the nuanced precision of polynomial regression. The latter's elevated R^2 value signals an exceptional fit, capturing nearly the entirety of the dependent variable's variance. Additionally, polynomial regression maintains low error metrics across RMSE and MAE, both in training (Figure 6) and testing (Figure 7) phases, thus evidencing its robust generalization and resistance to overfitting. Hence, in light of its unparalleled ability to model intricate relationships and sustain predictive accuracy, polynomial regression emerges as the optimal model for this dataset, adeptly balancing complexity and precision to deliver superior predictive performance.

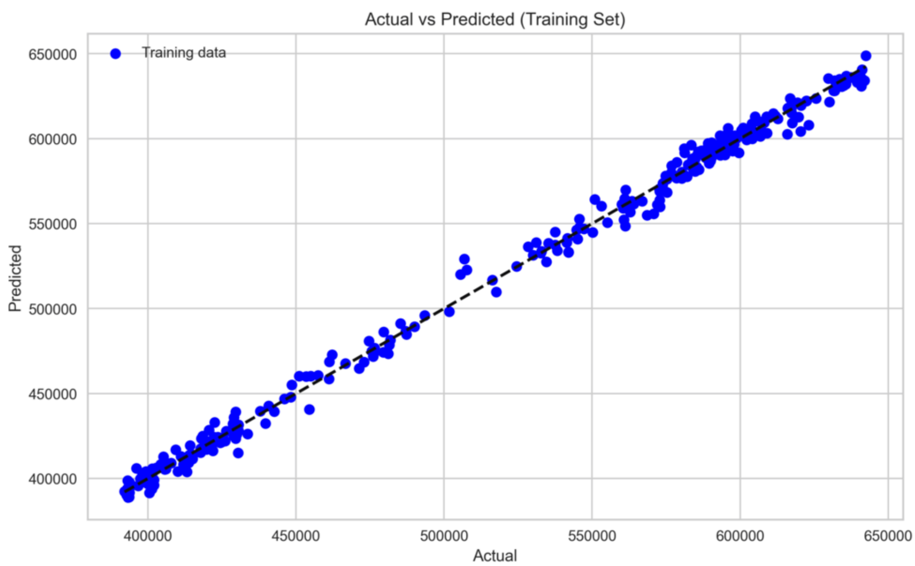

FIGURE 6: Actual vs predicted polynomial regression (training)

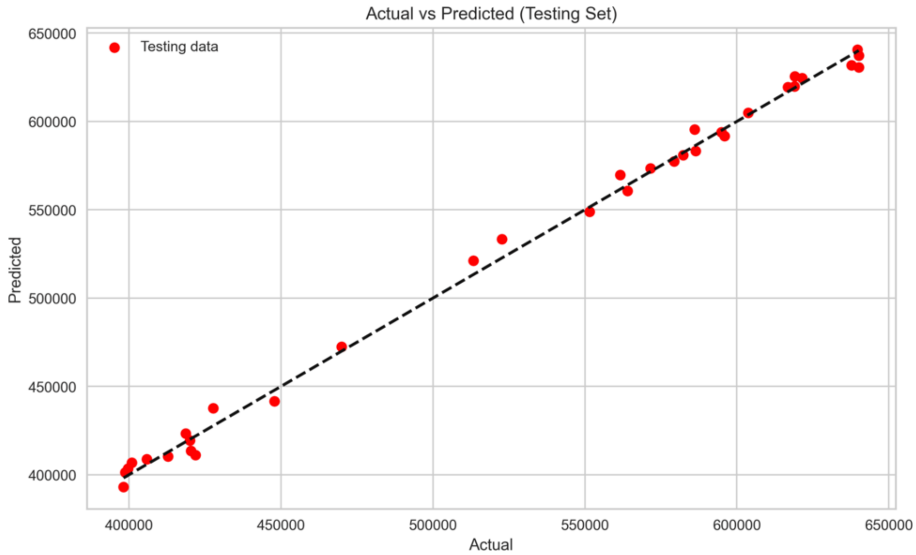

FIGURE 7: Actual vs predicted polynomial regression (testing)

Discussion

The analytical evaluation of regression models deployed for predicting foreign exchange reserves reveals a pronounced superiority of polynomial regression. This model exhibits an R^2 value of 0.9963, indicating it elucidates nearly the entire variance in foreign exchange reserves, substantially outperforming both multiple regression, with its R^2 of 0.9317, and ridge regression, with an R^2 of 0.9935. Such an exceptional R^2 value underscores polynomial regression's proficiency in encapsulating intricate, nonlinear interdependencies within the dataset, a feat that eludes the simpler linear models. Moreover, polynomial regression's error metrics such as MSE, RMSE, and MAE remain significantly low across both training and testing phases. These metrics attest to the model's precision and its robust generalization capability, ensuring minimal prediction errors when applied to new data. This attribute is particularly critical in the realm of economic forecasting, where accurate predictions of foreign exchange reserves are indispensable for strategic economic planning and decision-making. Polynomial regression's adeptness at balancing model complexity with predictive accuracy renders it the quintessential model for this application. Its ability to consistently provide precise forecasts, as evidenced by its superior performance metrics, makes it an invaluable tool for economists and policymakers who demand high fidelity in their predictive models. Thus, in light of these comprehensive advantages, polynomial regression emerges as the preeminent model for forecasting foreign exchange reserves, combining unparalleled accuracy with the capacity to model complex, nonlinear relationships.

Juxtaposing our study's outcomes with existing research elucidates both the consistency and uniqueness of our results. Our polynomial regression model achieved an exceptional 99.63% accuracy in predicting India's foreign exchange reserves. In contrast, our multiple regression model demonstrated a respectable but lower accuracy of 93%. It is crucial to contextualize these findings within the broader spectrum of academic work. Previous studies might have utilized varying methodologies, different datasets, or alternative predictor variables, which could result in differing outcomes. Consistency in the predictive power of regression models across various studies would highlight their robustness and reliability in capturing the intricate relationships governing foreign exchange reserves. On the other hand, any observed discrepancies might be attributable to differences in data quality, model specifications, or changing economic conditions. By comparing our findings with those of prior research, we can identify recurring patterns or distinct trends, thereby enhancing our understanding of the determinants of foreign exchange reserves. This comparative analysis not only contributes to the existing body of knowledge but also underscores the efficacy of polynomial regression in modeling complex economic phenomena, reinforcing its value in scholarly discourse.

Conclusions

Our comprehensive analysis using advanced regression techniques has yielded profound insights into predicting foreign exchange reserves, with polynomial regression emerging as the most effective model. Achieving an exceptional accuracy of 99.63% in forecasting foreign exchange reserves, polynomial regression significantly outperformed multiple regression, which achieved 93% accuracy. These findings underscore the superior capability of polynomial regression in capturing the intricate, nonlinear relationships inherent in economic data. This study not only highlights the predictive power of machine learning regression models but also offers valuable, actionable insights for policymakers, economists, and financial analysts. Despite the promising results, it is crucial to recognize limitations such as data quality and model assumptions, which may affect the transferability of our findings. Future research should aim to mitigate these limitations by incorporating diverse datasets, refining methodologies, and exploring alternative regression techniques. By continually advancing our data-driven approaches and leveraging machine learning innovations, we can enhance the precision of economic forecasts and inform strategic decision-making in a dynamic global economy.

Appendices

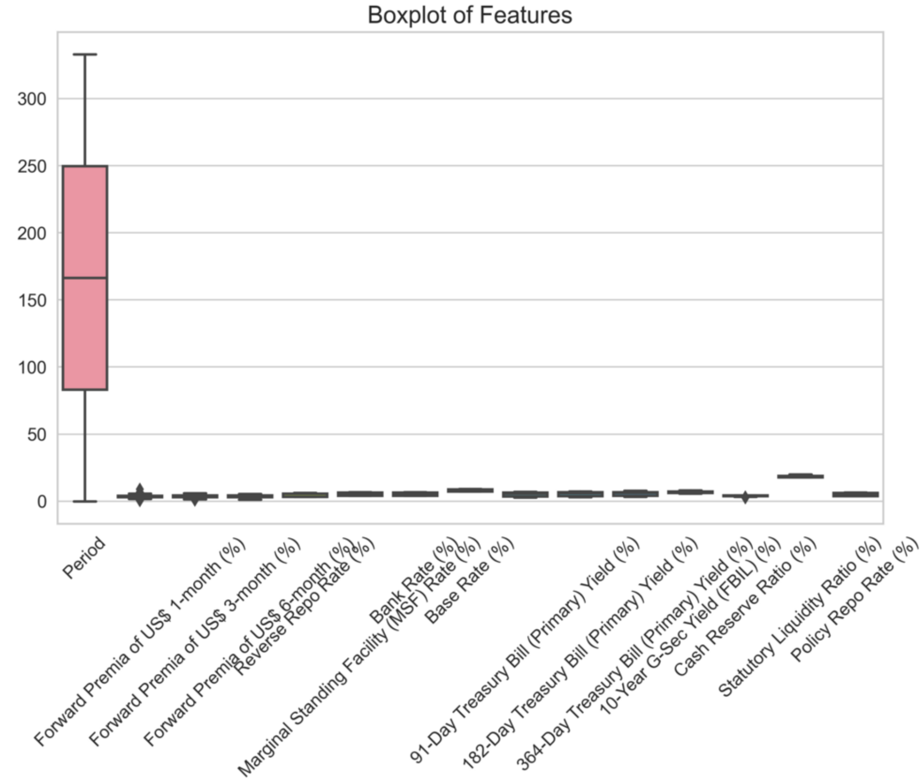

FIGURE 8: Boxplot of features

Boxplot visualizes the distribution of various features. The vertical axis represents the feature values, while the horizontal axis lists the different features being analyzed. The boxplot for the "Period" feature stands out as having a significantly higher value compared to the other features displayed (Figure 8).

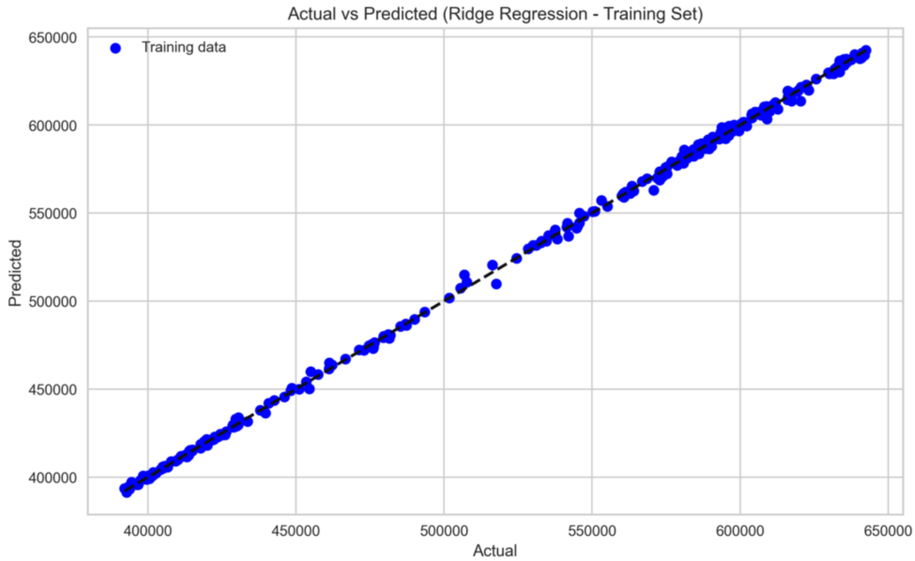

FIGURE 9: Actual vs predicted ridge regression (training)

Scatterplot of actual values vs predicted values using the ridge regression model on the training dataset. The blue dots represent the training data points, and the plot displays a clear linear relationship between the actual and predicted values. This suggests that the ridge regression model is able to accurately predict the values in the training set, as the predicted values closely align with the actual values along the diagonal (Figure 9).

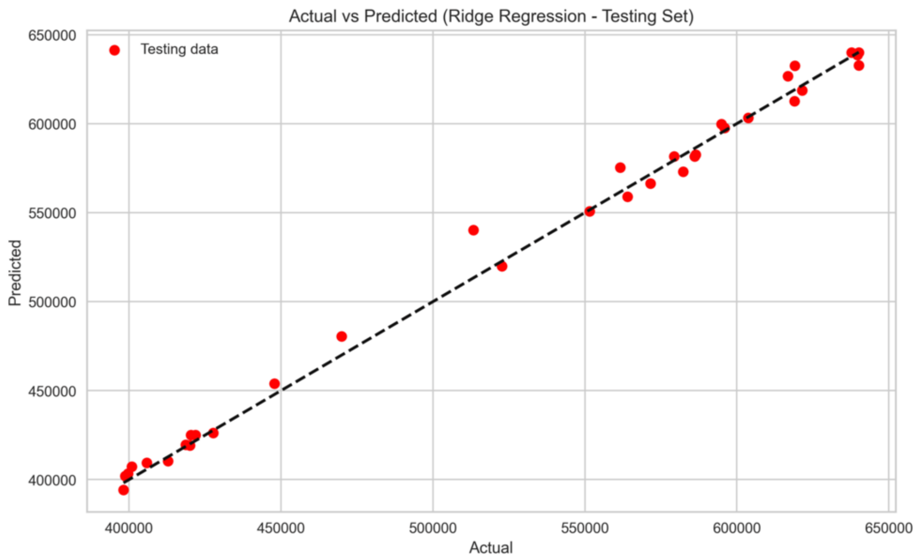

FIGURE 10: Actual vs predicted polynomial ridge regression (testing)

Scatterplot of actual vs predicted values using the ridge regression model on the testing dataset. The red dots represent the testing data points, and the plot displays a linear relationship between the actual and predicted values, though with more scatter compared to the training data plot. This suggests the ridge regression model can reasonably predict the testing set values, but with slightly less accuracy than the training data (Figure 10).

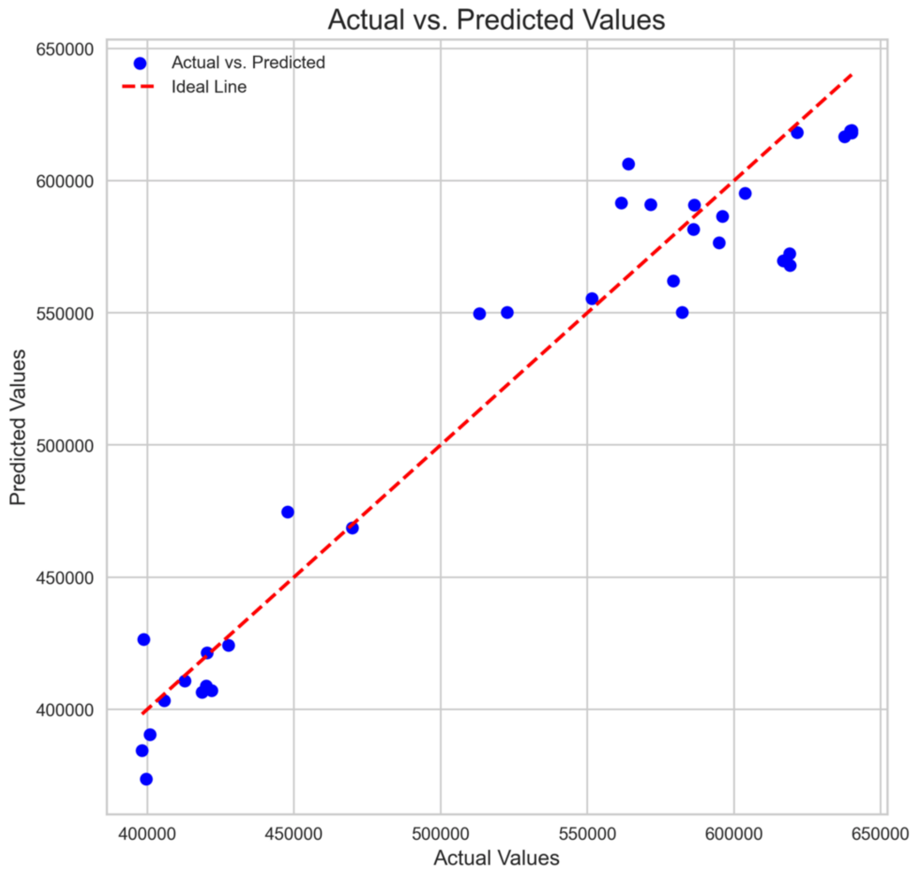

FIGURE 11: Actual vs predicted multiple regression

Scatterplot of actual values vs predicted values using a model. The data points, represented by blue dots, form a clear linear pattern along the ideal line, indicating the predicted values closely align with the actual values. This suggests the model is effectively predicting the target variable based on the given data (Figure *11*).

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Rohit D. Poul, Ankush D. Sawarkar, Aditya S. Baheti, Ubaid Ahmed Shaikh, Anurag Agrahari, Atul Halmare

Acquisition, analysis, or interpretation of data: Rohit D. Poul, Ankush D. Sawarkar, Aditya S. Baheti, Ubaid Ahmed Shaikh, Anurag Agrahari, Atul Halmare

Drafting of the manuscript: Rohit D. Poul, Ankush D. Sawarkar, Ubaid Ahmed Shaikh, Anurag Agrahari, Atul Halmare

Critical review of the manuscript for important intellectual content: Rohit D. Poul, Ankush D. Sawarkar, Aditya S. Baheti, Ubaid Ahmed Shaikh, Anurag Agrahari

Supervision: Ankush D. Sawarkar

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors are thankful to the Director, Shri Guru Gobind Singhji Institute of Engineering and Technology (SGGSIE&T), Nanded, India, for providing the necessary facilities for this work.

References

1. Bosco ED: Central Banks' management of foreign exchange reserves. Open Economies Review. 1998, 9:665-83. 10.1023/a:1008329207204
2. Chanda MM, Bandyopadhyay G, Banerjee N: Analysis and estimation of India's foreign exchange reserves using soft computing techniques. IIMB Management Review. 2020, 32:280-90. 10.1016/j.iimb.2019.10.010
3. Philip B, Bajaj K, Pareek W: Indian rupee: is it unpredictable? A regression model to predict the value of the Indian rupee. Contributions to Environmental Sciences & Innovative Business Technology. Springer Nature Singapore, 2023. 127-43. 10.1007/978-981-99-3366-2_16
4. Chinn MD, Prasad ES: Medium-term determinants of current accounts in industrial and developing countries: an empirical exploration. Journal of International Economics. 2003, 59:47-76. 10.1016/s0022-1996(02)00089-2
5. Kaminsky GL, Reinhart CM: The twin crises: the causes of banking and balance-of-payments problems. American Economic Review. 1999, 89:473-500. 10.1257/aer.89.3.473
6. Chen H, Chen Q, Gerlach S: The implementation of monetary policy in China: the interbank market and bank lending. Global Banking, Financial Markets and Crises. 2013, 14:31-69.
7. Jeanne O, Rancière R: The optimal level of international reserves for emerging market countries: a new formula and some applications. Economic Journal (London, England). 2011, 121:905-30. 10.1111/j.1468-0297.2011.02435.x
8. Kearney C, Muckley C: The role of the Japanese Yen in Asian exchange rate determination. Contemporary Studies in Economic and Financial Analysis. 2005, 86:29-51. 10.1016/S1569-3759(05)86002-4
9. Ghosh A, Rajan R: A selective survey of exchange rate pass-through in Asia: what does the literature tell us?. SSRN Electronic Journal. 2007, 39. 10.2139/ssrn.984063
10. Razzaque MA: Revitalising Bangladesh’s export trade: policy issues for growth acceleration and diversification. ResearchGate. 2017, 10.13140/RG.2.2.17214.38721
11. Rehman A, Belhaouari SB: Unsupervised outlier detection in multidimensional data. Journal of Big Data. 2021, 8:80. 10.1186/s40537-021-00469-z
12. Shen J, Shafiq MO: Short-term stock market price trend prediction using a comprehensive deep learning system. Journal of Big Data. 2020, 7:66. 10.1186/s40537-020-00333-6
13. Potdar K, Pardawala TS, Pai CD: A comparative study of categorical variable encoding techniques for neural network classifiers. International Journal of Computer Applications. 2017, 175:7-9. 10.5120/ijca2017915495
14. Bentéjac C, Csörgő A, Martínez-Muñoz G: A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review. 2020, 54:1937-67. 10.1007/s10462-020-09896-5
15. Fahrmeir L, Kneib T, Lang S, Marx BD: Regression. Springer, 2022. 10.1007/978-3-662-63882-8
16. Bera D, Chatterjee ND, Bera S: Comparative performance of linear regression, polynomial regression and generalized additive model for canopy cover estimation in the dry deciduous forest of West Bengal. Remote Sensing Applications Society and Environment. 2021, 22:100502. 10.1016/j.rsase.2021.100502
17. Koren Y, Bell R, Volinsky C: Matrix factorization techniques for recommender systems. Computer. 2009, 42:30-37. 10.1109/mc.2009.263
18. Shani G, Gunawardana A: Evaluating recommendation systems. Recommender Systems Handbook. Springer, 2011. 257-97. 10.1007/978-0-387-85820-3_8
19. Baheti AS, Sawarkar AD, Shaikh UA, Shrimankar DD: Alcohol quality analysis using machine learning regression technique. Cureus Journal of Computer Science. 2024, 10.7759/15