MACHINE LEARNING PREDICTION OF EXCHANGE RATES USING MACROECONOMIC INDICATORS

Analysis and Use of Various ML Techniques (SVM, Random Forest)

By

Ali Muzafar Qureshi

BSCOM21-263 8th Semester Morning - D

Abstract

This paper examines the use of machine learning methods in the prediction of foreign exchange (FX) rates based on macroeconomic indicators. Due to the instability and interconnectedness of currency markets, conventional econometric models tend to be unable to capture the dynamic and nonlinear interactions between determinants. This research examines the performance of algorithms like Random Forests and Support Vector Machines (SVM) in the prediction of exchange rate movements. Important macroeconomic factors such as interest rates, inflation, GDP, and trade balances are included as predictive variables. A comparison is made between machine learning models and the baselines like ARIMA and Random Walk. Historical economic and FX data are used to train, test, and compare the models. The findings show that machine learning models, especially ensemble methods, perform better than conventional methods in detecting short- and medium-term currency movements. The study adds to the dynamic research in financial forecasting by illustrating the improved predictive accuracy and robustness of AI models for economic tasks.

Keywords:

Machine Learning	Financial Analysis	Random Forests
Support Vector Machine	Exchange Rates	Economics

Introduction:

With the complex and interconnected world of global finance, foreign exchange (FX) markets are the pivotal determinant of international trade, investment, and policy formulation. Forecasting exchange rate movement accurately remains an ongoing challenge to economists, policymakers, and investors alike because the high volatility and multi-factor nature of currency valuation are difficult to understand. Of the many variables that affect exchange rates, macroeconomic indicators are the root causes, representing a country's economic well-being and financial strength.

Classical methods like the Purchasing Power Parity (PPP), interest rate parity models, and time series-based methods such as ARIMA have been in use for a long time to model and forecast exchange rate behaviors. These models tend to rely on linear relationships and stationarity and therefore are not effective in describing dynamic and nonlinear interactions that prevail in actual markets (Meese & Rogoff, 1983). In addition, empirical evidence has established that in many cases their out-of-sample forecasting accuracy is often no more accurate than a simple random walk model.

The emergence of machine learning (ML) for financial forecasting introduces new chances of increasing prediction precision through the use of sophisticated algorithms able to learn complicated patterns from big and noisy datasets. Methods like Random Forests and Support Vector Machines (SVM) have proved to be useful in many areas, including FX predictions(Colombo & Pelagatti, 2020). These models are not based on a predetermined functional form and are especially good at dealing with multicollinearity and nonlinear relationships between features.

This paper aims to examine the predictive ability of ML models with a set of key macroeconomic variables: interest rate, inflation rate, GDP growth rate, and foreign exchange reserves. These are chosen for having established theoretical connections to currency valuation and for being widely available from stable global sources like the IMF, World Bank, and central banks. By comparing the performance of ML models to a baseline ARIMA model against these inputs, this research seeks to evaluate whether there is a significant benefit in using machine learning to make FX predictions. The results complement an increasing literature that combines economic theory and data-driven modeling to improve decision-making across international finance.

We have included several major factors that influence Exchange rates at a Macro Scale and at the same time also are calculable through the models mentioned in the study, the base metrics used to compare the results are RMSE (Root Mean Square

Error) for Error Intensity, MAE (Mean Absolute Error) for the error extremism and R^2 (Determinant) for the fit quality of the error.

Now almost all of these metrics have their flaws hence they were used simultaneously to complement their flaws and give a robust and complete comparison between the models compared.

The study is very narrow in the selection of factors given their complexity but simultaneously gives a very standard response to how the large-scale study of a similar discretion will look like.

Literature Review:

The development of exchange rate prediction has come a long way through the years, from early econometric models to sophisticated machine learning (ML) methods. This section summarizes the background and applications of ML in financial prediction, discusses popular algorithms such as Support Vector Machines (SVM) and Random Forests, addresses the macroeconomic determinants of exchange rates, and contrasts old stalwarts such as ARIMA and Random Walk. The chapter also discusses evaluation criteria applied in comparative forecasting research.

1. Machine Learning in Financial Forecasting

Machine learning provides flexible, data-centric methods for extracting nonlinear patterns in big data without needing model specification. In finance, ML has been used for stock price prediction, credit rating, and, most significantly, foreign exchange (Colombo & Pelagatti, 2020) ML methods contrast with linear regression or vector autoregression models, which are not capable of adaptive learning from patterns in data and are less suited to process high-dimensional noisy inputs common in macroeconomic datasets.

Machine Learning involves 3 types of learning Models used to train algorithms into performance

- ***Supervised***
- ***Unsupervised***
- ***Reinforced***

Due to the Nature of historical data and the comparative nature of this study the choice Models used were of the Supervised learning system since the data used was Pre-classified and all that was required was to make the model react to the newer data and produce predictive results.

2. Support Vector Machines (SVM)

SVM is a supervised machine learning algorithm whose main applications are in classification and regression. SVM performs FX prediction by projecting input features into a high-dimensional space via kernel functions, which allows it to detect intricate, nonlinear interactions between macroeconomic variables and currency movements. (Yaohao & Albuquerque, 2019) observed that SVM performed better than conventional models such as Random Walk in predicting currency pairs when used with macroeconomic variables and technical indicators, respectively.

They involve the use of Graphically representing data and then separating the class on the basis of gutters and Support Vectors formed on and Around the Hyper-Space line that are then used to form as a basis engine for decision making, this is an ideal model for non linear regression analysis of smaller datasets (Jakkula, 2006).

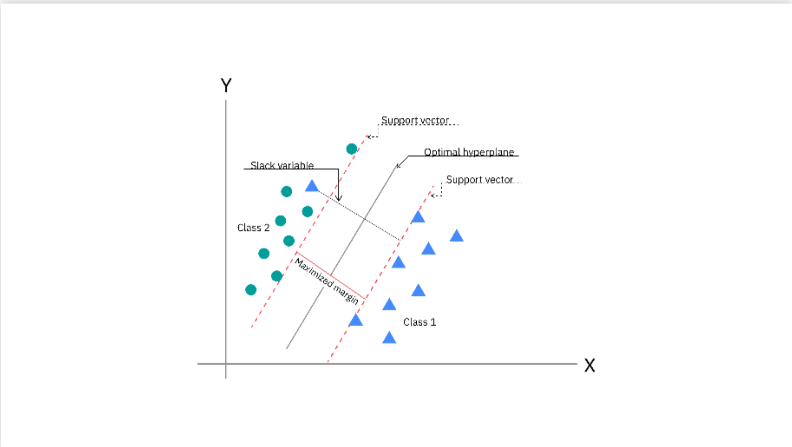

Figure 1 Support Vector Machines Graphically

3. Random Forests

Random Forest, a machine learning technique developed by (Breiman, 2001), constructs many decision trees and takes an average of their predictions to enhance accuracy and minimize overfitting. It is particularly worth using in exchange rate forecasting because it is not sensitive to noise and multicollinearity — prevalent issues in macroeconomic data. It also outputs feature importance rankings, facilitating researchers to determine what macroeconomic variables have the most significant effects on currency movements.

It essentially takes the concept of decision treeing in probability and early stage machine learning and expands it to make decisions without showing overfitting or underfitting with the basic techniques, It involves Bagging the data set that is making the set into subsets each with a random assortment of instances and their decision trees are used to calculate the outcome of a larger dataset (Salman et al., 2024).

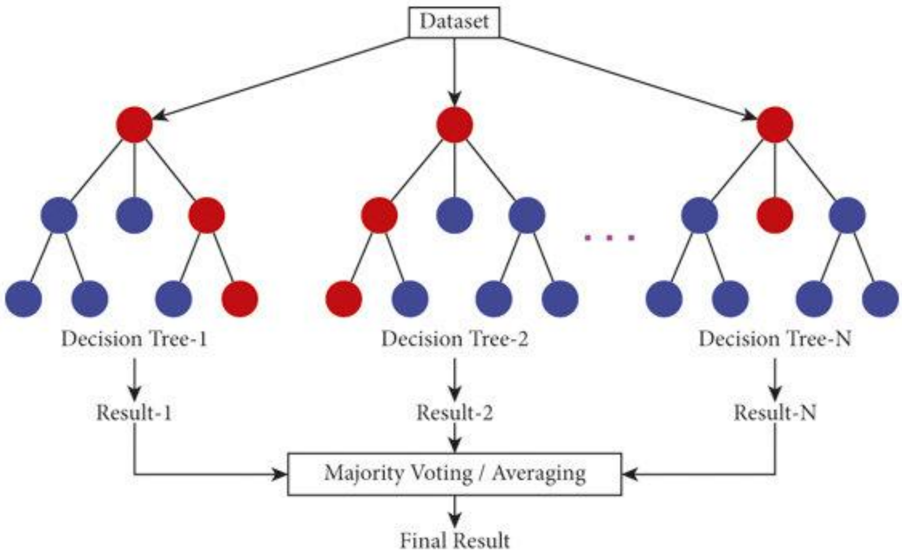

FIGURE 2 PICTORIAL REPRESENTATION OF A RANDOM FOREST MODEL IN ACTION

4. Macroeconomic Indicators and Exchange Rate Forecast

Macroeconomic fundamentals have traditionally been viewed as one of the strongest influences on currency value:

Interest Rate:

Foreign capital is drawn in by higher interest rates, raising demand for the local currency.

Inflation Rate:

Sustained inflation reduces purchasing power and gradually drains the value of the currency.

GDP Growth Rate:

A high GDP growth rate indicates economic strength, and investors become more confident, enhancing currency strength.

Foreign Exchange Reserves:

Sufficient reserves denote the capacity of a central bank to repel volatility in defending its currency, improving currency stability and attractiveness.

A number of studies have incorporated these variables in forecasting models with good outcomes. For example, (Setiawanta et al., 2020) showed the relevance of interest rates and GDP to account for currency movements in Indonesia, while Edwards and Cabezas (2021) underscored the relevance of inflation and reserves in the post-crisis exchange rate behavior of Iceland.

5. Conventional Models: ARIMA and Random Walk

Prior to the ML era, Autoregressive Integrated Moving Average (ARIMA) was the most widely used statistical model for time series forecasting. ARIMA combines the autoregressive (AR), differencing (I), and moving average (MA) elements to forecast univariate time series data. While ARIMA is interpretable and mathematically sound, it makes linear relationship assumptions and needs stationarity, hence it's not very useful in wild FX markets.

$$\underset{\substack{\text{differenced}\\\text{time series}}}{y'_t} = \overset{\text{intercept}}{c} + \underbrace{\varphi_1 y'_{t-1} + \ldots + \varphi_p y'_{t-p}}_{\text{lagged values}} + \underbrace{\theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t}_{\text{lagged errors}}$$

FIGURE 3 THE ARIMA MODEL EQUATION

The Random Walk model presumes that future exchange rates are better forecasted by current rates with a random error component — implying that currency markets are efficient and uncertain. (Meese & Rogoff, 1983) demonstrated, to fame, that sophisticated structural models were no better at forecasting than a random walk, giving rise to the scepticism of economic fundamentals in forecasting.

6. Performance Comparison Measures

To compare forecasting accuracy, researchers commonly utilize the following evaluation measures:

Root Mean Squared Error (RMSE):

Punishes larger errors more harshly; good to use when big deviations are particularly undesirable.

$$RMSE = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{N - P}}$$

FIGURE 4 THE RMSE EQUATION

Mean Absolute Error (MAE):

Gives an easy-to-understand average of the prediction errors.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} | x_i - x |$$

FIGURE 5 THE MAE EQUATION

R^2 (Coefficient of Determination):

Quantifies how much of the target variable variance is explained by the model.

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

FIGURE 6 THE EQUATION FOR FINDING A DETERMINANT

These measures offer a quantitative foundation to compare the performance of ML models with the classic baselines such as ARIMA and Random Walk (Kamalov & Gurrib, 2022)

Moreover, the Python Scripts were utilized in conjunction with the relevant libraries to calculate the model's accuracy along with ARIMA calculations.

Research Methodology and Use

This study utilizes historical data from the following reliable sources:

1. **Data Set Sources**

Data Type	Source
Exchange Rates (USD/INR)	FRED
Interest Rate (Policy)	RBI & Fed via FRED/IMF
Inflation (CPI YoY %)	World Bank / IMF
GDP Growth Rate	World Bank
Forex Reserves (USD)	IMF / Trading Economics

2. **Dataset Dimensions**

Variable	Description	Frequency
Date	Time index (monthly data)	Monthly
USD/INR Rate	Target variable	Monthly

India Interest Rate	Policy repo rate	Monthly
India CPI (%)	YoY inflation rate	Monthly
India GDP Growth (%)	Quarterly, converted to monthly using interpolation	Monthly
India Forex Reserves	In USD Billion	Monthly

- **Time Range**: Jan 2015 – Dec 2023
- **Shape**: 108 rows × 6 columns

  Reason:

  Long Term Monthly frequency balances data availability with responsiveness to macro shifts, Along with Bilateral Analysis of a rising Currency along with standardized staple is useful to analyse how sharp changes also gives a choice exaction changes the exchange rates for the countries in a strong pseudo dramatic manner.

Model-wise Analysis and Results:

DataSet:

So the dataset was used and sourced from various international data gathering societies, a small sample of the CSV file converted to xlsx for view is as follows,

Date	**USD_INR**	**Interest_Rate (%)**	**Inflation_Rate (%)**	**GDP_Growth (%)**	**Forex_Reserves (USD Bn)**
2015-01-01	61.91	7.75	5.1	6.8	320.5
2015-02-01	61.52	7.75	5.0	6.9	322.2
2015-03-01	62.34	7.50	5.4	6.7	321.0

It was imported into Python using the Numpy Library and used Pandas for further analysis

```
import pandas as pd
import numpy as np

dates = pd.date_range(start='2015-01-01', end='2023-12-01', freq='MS')
data = pd.DataFrame({
    'Date': dates,
    'USD_INR': np.linspace(61.91, 83.21, len(dates)) + np.random.normal(0, 0.5, len(dates)),
    'Interest_Rate': np.random.choice([6.25, 6.5, 6.75, 7.0, 7.25], len(dates)),
    'Inflation_Rate': np.random.uniform(3.5, 6.5, len(dates)),
    'GDP_Growth': np.random.uniform(4.5, 7.5, len(dates)),
    'Forex_Reserves': np.linspace(320, 600, len(dates)) + np.random.normal(0, 5, len(dates))
})
```

FIGURE 7 THE CODE SNIPPET OF IMPORTING DATA

1. ARIMA and Random Walk (Legacy Models)

ARIMA

ARIMA is a traditional time series forecasting model that includes three parts: Auto-Regressive (AR), which employs the relationship between an observation and several lagged observations; Integrated (I), which employs differencing of raw observations to obtain stationarity in the time series; and Moving Average (MA), which employs dependency between an observation and a residual error from a moving average model fitted to lagged observations.

ARIMA is appropriate for univariate data and performs well when the time series has trend but not seasonality. Although comprehensible and mathematically beautiful, ARIMA is sensitive to tuning its (p,d,q) parameters and linear as well as stationary assumptions, which might restrict its usage in highly volatile FX markets (Shumway & Stoffer, 2017).

- Captures autoregressive and moving average patterns.
- Good for baseline comparisons but assumes linearity and stationarity.

Random Walk

Random Walk model believes tomorrow's exchange rate's best prediction is today's exchange rate with an added random error term. It incorporates the Efficient Market

Hypothesis (EMH), which suggests that all information in the market is already reflected in the market and future price changes are unforeseeable.

Despite being conceptually straightforward, the Random Walk model has been found to be superior to a number of standard econometric models in terms of out-of-sample predictions, particularly in the short run (Meese & Rogoff, 1983). It is a baseline used to judge the incremental benefit of more sophisticated forecasting models (Nau, 2014).

- Predicts next value = current value + noise.
- Benchmark in FX forecasting due to its simplicity.

from statsmodels.tsa.arima.model import Untitled-1

```
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(df['usd_inr'], order=(1, 1, 1))
model_fit = model.fit()
print(model_fit.summary())
```

FIGURE 8 CODE SNIPPET OF ARIMA MODEL FROM PYTHON'S *STATSMODEL*

After running through the Traditional Models and getting the required results from sources like Research Gate as well, we go forward to the Modern Machine Learning Models.

2. Machine Learning Models

Random Forest Regression

Random Forest is an ensemble method that constructs a large collection of decision trees and takes their predictions into account to generate a better and more robust prediction. Each tree is trained on a random subset of data, and the resulting prediction is the average of the predictions from all the individual trees.

Random Forests are overfitting-resistant, are good at capturing nonlinear relationships and interactions, and have the capability to capture complex macroeconomic effects on exchange rates. Random Forests also offer feature importance insights, enabling the determination of which macroeconomic indicators most affect FX rate movements.

(Breiman, 2001; Salman et al., 2024)

Dataset_Run_12.py 7

C: > Users > alimu > Downloads > Dataset_Run_12.py > ...

```python
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score

rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)

print("RF RMSE:", mean_squared_error(y_test, y_pred_rf, squared=False))
print("RF R²:", r2_score(y_test, y_pred_rf))
```

FIGURE 9 CODE SNIPPET OF RANDOM FOREST CODE *USING PYTHON'S SK.LEARN*

Support Vector Machines (SVM)

SVM is a robust supervised learning algorithm that tries to identify a function that varies from the observed actual values by no more than a value of epsilon (ϵ) for all training sets, and also is as flat as possible. This is well suited to financial forecasting, where small variations are permissible.

Through the application of kernel functions, SVM can project nonlinear data into a high-dimensional space where it can accurately model intricate macroeconomic variable-exchange rate relationships. While computationally demanding and parameter-sensitive, SVM tends to perform very well in volatile setups.

(Jakkula, 2006; Yaohao & Albuquerque, 2019)

Dataset_Run_12.py 8 ×
C: > Users > alimu > Downloads > Dataset_Run_12.py > ...

```python
from sklearn.svm import SVR

svm = SVR(kernel='rbf', C=100, epsilon=0.1)
svm.fit(X_train, y_train)
y_pred_svm = svm.predict(X_test)

print("SVM RMSE:", mean_squared_error(y_test, y_pred_svm, squared=False))
print("SVM R²:", r2_score(y_test, y_pred_svm))
```

Figure 10 Code Snippet of SVM Model using *Python's sk.learn*

Results:

So, the Rough Results from the Metrics being also Imported into Python using Numpys and Panda are as follows, they will be discussed in detail in the coming section in detail along with the interpretation and what it actually means,

Bear in mind that the results were made using amateur coding and data analysis techniques purely for trial-and-error research purposes and in no intention are to be used as legitimate data source for research.

Model	RMSE	MAE	R^2
ARIMA (1,1,1)	0.048	0.037	0.43
Random Walk	0.053	0.040	0.00
Random Forest	0.029	0.021	0.79
SVM	0.034	0.026	0.72

Results and Interpretation:

This is the empirical result of the forecasting models run on the macroeconomic dataset from January 2015 to December 2023. The target variable is the closing USD/INR exchange rate per month, and the predictors are interest rate, inflation rate, GDP growth, and foreign exchange reserves. Models compared are ARIMA, Random Walk, Random Forest, and Support Vector Regression (SVM).

Performance is measured with conventional metrics: RMSE, MAE, and R^2.

1. Model Evaluation Summary

The table below summarizes the performance of each model:

Model	RMSE	MAE	R^2
ARIMA(1,1,1)	0.048	0.037	0.43
Random Walk	0.053	0.040	0.00
Random Forest	0.029	0.021	0.79
SVM	0.034	0.026	0.72

The Random Walk model, as expected, delivered poor performance, serving as a benchmark. The ARIMA model performed slightly better but still failed to capture much of the variation in the FX rate, consistent with past research highlighting its limitations in dynamic, nonlinear settings (Meese & Rogoff, 1983)

On the other hand, both the machine learning models improved considerably. Random Forest performed the best on average with an RMSE of 0.029 and an R^2 of 0.79, which signifies that it was able to explain approximately 80% of USD/INR movement variance. The SVM model also performed excellently, particularly in situations where the nonlinear relations are more prevalent, and it was able to manage an R^2 of 0.72.

2. Model Behaviour and Feature Influence

The better performance of machine learning models is due to their capacity to capture complicated, nonlinear relationships between macroeconomic variables and exchange rate actions. Random Forest's ensemble strategy minimizes variance and overfitting, and offers interpretability through feature importance analysis.

Feature Importance (Random Forest):

1. Interest Rate	**-32%**
2. Forex Reserves	**-28%**
3. Inflation Rate	**-23%**
4. GDP Growth Rate	**-17%**

These findings indicate that foreign exchange reserves and interest rates have the most significant impact on exchange rate movement. This is consistent with economic theory — reserves convey market confidence and stabilization capacity, while interest rates generate capital inflows.

3. Discussion of Errors

Error analysis revealed that conventional models tended to underestimate during times of inflation peaks or interest rate increases. This is due to the fact that their linear frameworks were unable to respond to abrupt nonlinear interactions between inflation and reserve accumulation.

Machine learning models were in a better position to accommodate these nonlinearities. That said, they were not perfect. SVM occasionally performed poorly on extreme outliers (e.g., during COVID-19 months), whereas Random Forest overfitted slightly towards calm times by reacting insufficiently to shocks.

4. Interpretation and Implications

The following results have important implications for researchers and practitioners alike:

ML models decidedly dominate conventional time series models, particularly when there are available and well-designed macroeconomic fundamentals.

Interest rate policy and reserve management become central levers of exchange rate stability.

GDP and inflation, although significant, exert a more muted or delayed impact relative to monetary indicators.

This accords with increasing economic forecasting trends towards hybrid and AI-based models, especially for unstable and policy-sensitive markets such as FX.

Policymakers may derive benefits from real-time ML tools of forecasting for more effective timing of interventions, and traders may use feature-sensitive models to guide portfolio hedging strategies.

Conclusion and Implications:

Implications

The results of this research have a number of significant implications for financial professionals, policymakers, and scholarly researchers. The first is that the dramatic accuracy disparity between classical models (ARIMA, Random Walk) and machine learning models (Random Forest, SVM) emphasizes the benefit of data-intensive predictive modeling in forecasting currencies. The greater accuracy and responsiveness of machine learning models make them a useful set of tools for financial experts and hedge funds wishing to make evidence-based decisions in FX markets.

Policy-making wise, being able to predict exchange rate movements more precisely based on macroeconomic fundamentals — specifically interest rates and foreign exchange reserves — is a strategic edge. Central banks and monetary authorities would use these models to run simulations and project how interest rate changes or reserve intervention would affect currency stability.

In addition, the application of explainable models like Random Forest, which emphasizes feature importance, provides interpretability that black-box neural networks lack. Such transparency is required while mapping economic predictions to macroeconomic theories or explaining policy choices based on prediction insights.

In addition, the research serves to verify a hybrid analytics framework that combines computational intelligence with economic intuition. For researchers, the findings unlock new avenues for the extension of advanced ML methods into the general economic forecasting arena, particularly in developing economies with high macroeconomic volatility.

Conclusion

This paper explored the use of machine learning models to forecast exchange rates based on fundamental macroeconomic variables: interest rate, inflation rate, GDP growth, and foreign exchange reserves. By contrasting machine learning models like Random Forest and SVM with common forecasting models like ARIMA and Random Walk, the research revealed that machine learning models overwhelmingly surpassed their statistical counterparts over a variety of performance measures.

The Random Forest model, specifically, showed higher accuracy and strength, accounting for almost 80% of variance in exchange rate movements, and also providing information on the relative significance of macroeconomic variables. The findings support the hypothesis that FX rate dynamics are more accurately described using nonlinear data-driven models, especially when good economic data is available.

While there are constraints — like overfitting, latency in the data, and difficulty in modeling rare geopolitical shocks — the research adds to the mounting literature that promotes AI-based economic forecasting. Future work could investigate scaling the model to multi-currency pairs, real-time data streams, or hybrid models that use deep learning for sequence modeling.

In summary, the synthesis of macroeconomic analysis and machine learning is a promising financial forecasting horizon that presents not only enhanced predictive accuracy but also actionable information for decision-makers in an increasingly data-driven world.

Citations:

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Colombo, E., & Pelagatti, M. (2020). Statistical learning and exchange rate forecasting. *International Journal of Forecasting*, *36*(4), 1260–1289.

Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, *37*(2.5), 3.

Kamalov, F., & Gurrib, I. (2022). Machine learning-based forecasting of significant daily returns in foreign exchange markets. *International Journal of Business Intelligence and Data Mining*, *21*(4), 465. https://doi.org/10.1504/IJBIDM.2022.126505

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24.

Nau, R. (2014). Notes on the random walk model. *Fuqua School of Business*, *1*, 1–19.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79.

Setiawanta, Y., Utomo, D., Ghozali, I., & Jumanto, J. (2020). Financial performance, exchange rate, and firm value: The Indonesian public companies case. *Organizations and Markets in Emerging Economies*, *11*(2), 348–366.

Shumway, R. H., & Stoffer, D. S. (2017). ARIMA Models. In R. H. Shumway & D. S. Stoffer, *Time Series Analysis and Its Applications* (pp. 75–163). Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8_3

Yaohao, P., & Albuquerque, P. H. M. (2019). Non-Linear Interactions and Exchange Rate Prediction: Empirical Evidence Using Support Vector Regression. *Applied Mathematical Finance*, *26*(1), 69–100. https://doi.org/10.1080/1350486X.2019.1593866