Forecasting Currency Exchange Trend on USD/CAD

Simon Akubue

Master’s Thesis

MEng Big Data Analytics

2024

Master's Thesis

Simon Akubue

Forecasting Currency Exchange Trend on USD/CAD

Arcada University of Applied Sciences: MEng in Big Data Analytics, 2024.

Commissioned by:

N/A

Abstract:

Currency exchange rate forecast has been, and remains a challenging tasks. With unpredictable natural disasters, political instabilities, government policies, and many other factors, it becomes difficult to correctly forecast the currency exchange rate. Many researchers in the past have done great works on forecasting the exchange rate of the United State Dollar (USD)/Canadian Dollar (CAD) using statistical approach. Even the Fundamental approach of relying on macroeconomic factors of the two countries, such as GDP ratio, Import/Export, government revenue, etc. were considered at various points. But (while forecasting the USD/CAD exchange rate), none of the previous methods considered deeply the underlying market trends which forms the basics of Technical analysis. We have included various machine learning models and time-sensitive indicators that directly aligns with the USD/CAD exchange rate movement so as to improve on this issue. These features will create a new dimension for researchers to predict and forecast the USD/CAD exchange rate. We have considered various types of models for predicting and forecasting the USD/CAD exchange rate, and realized that among all our models, Time Series models provides the best accuracy. While building our model, I relied of daily forex statistical data maintained by Tick Data Suite (TDS) and Investing.Com. All our macroeconomics data (Consumer Price Indices (CPI), Interest Rates, Imports, Exports, and Un-Employment Rates) were automatically downloaded from the Federal Reserve Economic Data (FRED)'s portal via an API. Due to huge data size available to me, I used 90% of my data for training my models, and the remaining 10% for my testing. At the end, we were able to come up with good results from our models, of which Linear Regression, Bayesian Ridge Regression, and Random Forest Regressor (RFR) outperformed others based on accuracy.

Keywords:

Modelling, Forex, USD/CAD, Random Forest Regressor

Contents

Figures

Tables

1 Introduction

The foreign exchange (forex) market stands as one of the most dynamic and influential financial markets globally, facilitating the exchange of currencies and shaping international trade and investment. Among the myriad currency pairs traded in this market, the USD/CAD pair occupies a significant position due to its prominence in North American trade and its influence on the economies of the United States and Canada.

The USD/CAD exchange rate reflects the relative strength of the US dollar (USD) against the Canadian dollar (CAD) and is subject to various factors, including macroeconomic indicators, geopolitical events, central bank policies, and market sentiment. Understanding and accurately forecasting movements in the USD/CAD exchange rate are crucial for stakeholders ranging from multinational corporations and financial institutions to individual traders and policymakers.

Given the complexities and uncertainties inherent in the forex market, the development of reliable forecasting models for the USD/CAD exchange rate becomes imperative. Such models provide valuable insights into future exchange rate movements, enabling market participants to make informed decisions and manage risks effectively.

This thesis aims to contribute to the field of forex forecasting by proposing a robust and accurate model for predicting the USD/CAD exchange rate. Leveraging a combination of econometric techniques, statistical methods, and machine learning algorithms, the proposed model seeks to capture the underlying relationships and dynamics driving USD/CAD exchange rate fluctuations.

By incorporating a diverse set of predictors, including macroeconomic indicators, technical analysis indicators, and sentiment analysis metrics, the model endeavors to provide a comprehensive and holistic approach to USD/CAD forecasting. Additionally, the utilization of advanced modeling techniques, such as time series analysis, regression analysis, and machine learning algorithms, enables the model to adapt to changing market conditions and improve predictive performance over time.

The thesis will proceed by reviewing relevant literature on forex forecasting methodologies and exploring the key factors influencing the USD/CAD exchange rate. Subsequently, it will detail the methodology employed in developing the forecasting model, including data collection, feature selection, model training, and evaluation. Finally, the thesis will present empirical results, discuss implications for stakeholders, and offer recommendations for future research in the field of forex forecasting.

Through this research endeavor, we aim to contribute to the advancement of forex forecasting techniques and provide valuable insights into the dynamics of the USD/CAD exchange rate, ultimately assisting market participants in making more informed decisions and mitigating risks in the forex market.

1.1 Literature review

Over the years, many people has conducted research on how to arrive at a methodology for predicting foreign exchange values accurately. The exchange rate values which do not aligned with macroeconomic indices could have contributed to this global imbalances. Despite the number of research on exchange rate risks, no consensus has emerged regarding the overall global effects of monetary variables on exchange rate volatility. The foreign exchange market is the largest liquid market consisting of a global network of buyers and sellers of currencies. This market trades more than $3 trillion daily, surpassing any other financial market.

The Bretton Woods' agreement established a system through which a fixed currency exchange rate could be created using gold as the universal standard. This agreement which involved representatives from 44 nations (including United States, Canada, Australia, Western European countries, and others) brought about the creation of the International Monetary Fund (IMF) and the World Bank. The whole idea of this agreement is to regulate Foreign Exchange Rates using Gold as a standard. Ever since the collapse of this system, understanding the effect of exchange rate policy and currency movements has been the dominant area in international financial research as the value of a currency fluctuations affects households and businesses. Exchange rate instability increases the uncertainty for the participants of foreign exchange markets, and this affects the flow of international trades. Stable exchange rates help firms evaluate the performance of investments, promotes financing, and management of operational risks. The correlation among exchange rate instability and international trade depends upon how much risk a firm is willing to accept. Empirical studies has examined the stability between trade and exchange rate volatility using econometric models. However, neither empirical nor theoretical literature provides sufficient evidence on the trade effects.

Empirical Studies on USD/CAD Forecasting:

So many research efforts has been applied though empirical means so as to come up with a standard technique of predicting the exchange rates. Empirical approaches like Time series analysis, econometric models, and machine learning algorithms have been employed to predict short-term and long-term trends in the currency pair. Results from some of these empirical approaches were convincing while others were inconsistent due to factor like natural disasters, and political instabilities.

Role of Macroeconomic Factors:

Macroeconomic factors such as interest rate differentials, GDP growth, inflation rates, and Consumer Price Index plays crucial roles in determining the exchange rate of any currency. Central bank policies influence investor sentiment and market expectations, and thus, leads to fluctuations in the currency rates. Incorporating these macroeconomic factors into forecasting models is essential for improving the accuracy of our forecasts.

Technical Analysis:

In addition to fundamental analysis, technical analysis techniques and market sentiment indicators are widely used by traders and analysts to forecast the exchange rates. Chart patterns, relative strength indicator, moving averages, and momentum oscillators helps in identifying the support and resistance levels of the market. It also suggests the trend direction, and potential reversal points.

Conclusion:

In conclusion, forecasting currency exchange rate is a complex and challenging task that requires careful consideration of macroeconomic factors, technical analysis tools, and market sentiment analysis. While existing research provides valuable insights into forecasting methodologies like technical analysis, further research is needed to incorporate other factors like macroeconomics indices so as to improve prediction accuracy of the foreign exchange rates.

This literature review sets the stage for the subsequent research in this thesis, which aims to develop and evaluate forecasting models for predicting USD/CAD exchange rate trends using a combination of statistical data, macroeconomics data, and technical analysis techniques.

1.2 Research question

How do I mitigate against losing money in Foreign Exchange market (FOREX) by being able to predict the price direction accurately most of the times?

Ever since the world economy got heavily dependent on international trade, buying goods and services from a country requires an individual or an organization to buy them in the accepted local currency of that country. For example, if an organization purchases goods from Canada, that entity must pay in Canadian Dollar (CAD) despite having United States Dollars (USD). The currency exchange rate plays a vital role in this transaction. That organization can exchange US Dollars with Canadian Dollar depending on the exchange rates set by the central bank of Canada. So the Exchange rate means the value of one nation's currency in terms of another. As exchange rates became a crucial element in international trade, predicting currency exchange rates became a demanding and a challenging process for businesses and individuals involved in the FOREX (foreign exchange market).

In the earlier days, economists tried to evaluate their mathematical exchange-rate models using the horse race approach, where they saw which model performed better in predicting the actual values of the exchange rate. Machine learning techniques have added a new dimension by making devices self-learner. Machine learning algorithms are capable of doing complex calculations faster and capable of making decisions more accurately. For that reason, predicting a currency exchange rate has shifted from manual accounting to machine learning algorithms, which have proven much more efficient and accurate than previous approaches. Over the last thirty years, the unanimity on the determinants of currency exchange rate movements has further broken down. The actual reason for the currency exchange movements can be explained partly by the world economy and the development of new theories of exchange rate determination.

The work done in this paper has focused on the perspectives of the world economy in the exchange rate movements of two big economies by taking into account the driving forces of the currency Supply versus Demand in the global market. Though the USD/CAD exchange rate prediction was made before using time series models, machine learning, and neural network models, much more inaccuracy still needs to be addressed in most previous research works. In the existing studies on the USD/CAD exchange rate, there was no inclusion of the underlying Technical features as well as macroeconomics indices, which are the dependent variables that will help predict the trend better. All the existing research just used the USD/CAD exchange rate pattern to train the models but ignored the factors that can directly suggest the exchange rate. None of them trained their models using the underlying features. The challenge is to improve the results even by a small margin to mitigate investment risk, and promote higher returns while maintaining an investment portfolio.

Due to lack of in-depth knowledge and proper analysis, many investors has lost so much in Forex. This triggered my interest in exploring ways of (fully or partially) automating this predictive analysis. I am using USD/CAD as a case model for other forex pairs. My approach to this study of forecasting the USD/CAD exchange rates stands on the combination of Statistical data, Technical features, and Macroeconomic indices, with deep learning, and machine learning algorithms. Every country has macroeconomics policies that drives her economy. Previously, researchers have trained their models by relying heavily on the exchange rate pattern only. Those approaches were much more statistical than considering the underlying reasons behind the fluctuation of the USD/CAD exchange rate. In this research work, I will be considering few macroeconomics indices of United States (USA) and that of Canada (CAN) that directly have large impact of their exchange rates. I will also consider statistical data as well as some established technical features like RSI, SMA, EMA, and MACD. I am convinced that pulling energy from these three reliable techniques will give me better forecasting compared to the existing works. This novel empirical approach has created a new dimension for the researchers to factor in their insights and develop models for better prediction and reduce investors' risk in the FOREX market. Reducing investors' risk in the FOREX market will significantly increase investment. For a fast growing country such as Canada, it is essential to increase investments in order to improve its economy. Therefore the main concept of this research can be described by these points:

(i) Previously, researchers have forecasted the USD/CAD exchange rate by using the pattern of the market rate only. Those approaches were much more statistical than considering the underlying factors behind the change in the USD/CAD exchange rate.

(ii) To address this issue. First, we have studied the reasons behind the change in the USD/CAD exchange rate from an economic perspective.

(iii) Secondly, we identified the models that can be used for forecasting the USD/CAD exchange rate.

(iv) Then, we applied these interesting techniques to machine learning models in our research. We have trained our models using reliable macroeconomics indices, statistical data, and technical features, thus, found a better result than other existing research.

(v) Finally, we have proposed a pipeline for improving the result further.

1.3 Related research works

Before starting the implementation of our research work, we explored a lot of related research works in the field of machine learning, neural networks, and economics historical data. Some notable research works related to our USD/CAD currency exchange rate forecasting have been mentioned below:

Researchers like *Saeed et al. (2012)* has analyzed the determinants of the exchange rate between USD and PKR within the framework of the monetary approach. They have taken stock of money, foreign exchange reserves, and total debts of Pakistan relative to the USA as the dummy variables and took them as the determinants of the exchange rate. They have applied the ARDL approach to the co-integration and error correction model. The results confirmed that stock of money, debt, and foreign exchange reserves are significant determinants of the exchange rate. *Yao & Tan (2000)* have provided empirical evidence that a neural network model is applicable for predicting the exchange rate. The authors have fed time-series data, moving averages into the model to determine the underlying rules of exchange rate movements between different currencies. Their work further showed that foreign currency's dynamic supply and demand factors made it difficult to predict the foreign exchange rates effectively. This aligns with the fundamentals of economics, which are that prices increase when the demand is greater than the supply and decrease when the demand is less than the supply. Since FOREX is a currency market, this fundamental aspect of economics stands true.

Using data mining techniques, *Carbureanu (2011)* have tried to predict the Romanian LEU/Euro. In that study, they mentioned that political, economic, and social events in a given timeline influence the supply and demand of a currency and have an instant effect on the currency exchange rate. Since there is no control over political and social events, our study looked into the economic factors that strongly impact currency exchange rates.

Ramasamy & Abar (2015) have used the yearly currency rates of some countries together with their macroeconomic variables. Macroeconomics variables like interest rates was considered in verifying the impacts on the exchange rates. They used the bootstrapping technique to increase the sample size and run regressions to study the effect. According to their study, macroeconomic variables such as tax rate, the balance of payments, inflation rate, interest rates, and other factors randomly influence the exchange rates. However, these macroeconomic variables might be unstable depending on the state of the economy within a country.

Twin (2022) has stated a few macroeconomic factors that influence the currency exchange rates: inflation rate, interest rate, current account deficit, public debt, terms of trade, and strong economic performance. According to the exchange rate regime, the floating exchange rate is influenced by the market's driving forces of supply and demand, where the fixed/pegged exchange rate is determined and maintained by the government or central bank *Natto (2010*).

Zwanger (2008) has conducted a study to outline the effects of modern exchange-rate theory on the exchange rate movements of Chili and the United States. Since the Chilean Peso (CLP) is pegged to the USD, they have considered the independent variables of the monetary policy interest rate, money supply, and inflation rates. However, they have mentioned that these independent variables might lose their explanatory ability when economic conditions change or in the case of switching in the foreign exchange rate policy dictated by the central bank.
A study by *Cushman (1988)* has tested the risk effects of real exchange rates on U.S. bilateral trade flows during the floating period using a few of the new and previously used risk measures. They have mentioned that other factors affect the calculation of exchange rates in third-world countries. According to the law of the fixed price model, the prices of goods in different countries which are traded internationally are identical in the perfect market.

Studies by *Akhtar & Hilton (1984)* have found that the uncertainty of exchange rate significantly impacts the imports and exports in Germany and the USA. They have established a negative relationship between the volatility in the exchange rate and the volume of international trade.

In their study, *Broil & Eckwert (1999)* have mentioned that developing countries have insignificant access to international capital. Thus, the domestic inflation rate is connected with fixed/pegged exchange rates. A study by *Kemal (2005)* has revealed that in the case of Canada, the exchange rate volatility is positively related to so much of imports with little exports. Furthermore, they have concluded that currency devaluation occurs by balancing the trade deficit.

Bouraoui & Phisuthtiwatcharavong (2015) conducted a study on Thailand THB/USD exchange rate where they explained how the Thai central bank intervenes in response to certain concerns and shocks in the managed floating regime. In this scenario, the study outlines six important factors that influence the THB/USD exchange rate: interest rates differential, manufacturing production index, terms of trade, monetary base, government debt, and international reserves.

Refenes et al. (1994) conducted a study where they have proved neural networks could outperform the statistical forecasting techniques when the non-linearity approaches are applied in the dataset of stock indices. They have shown that using sensitivity analysis and neural networks can provide a rational explanation of their predictive behavior and model their environment more convincingly than regression models.

Rehman et al. (2014) used CGP and Recurrent Neural Network to predict the exchange rates between AUD and three other currencies. An approach of Recurrent Neuro-Evolution was taken to forecast the currency exchange rate. They have observed that the computational method outperformed other statistical methods due to the flexibility and ability to select the best feature in real-time, and effectively recognize the patterns.

Islam & Hossain (2020) predicted the exchange rates of major currency pairs using the GRU-LSTM hybrid network. They tested the results with the standalone GRU and standalone LSTM models and found that the hybrid model outperformed the standalone models. This provides us with the idea of using a hybrid model to predict the USD/CAD exchange rate.

Pandey et al. (2020) have reviewed the neural network and statistical models to predict the exchange rate and also proposed a machine that identifies the shortcomings of both the neural network and statistical models. They have found that multilayer neural networks had

BAYESIAN learning predictive accuracy performed better than neural networks using backpropagation learning.

Rout et al. (2014) forecasted the exchange rates using an adaptive autoregressive moving average (ARMA) model with differential evolution-based training. They have compared the ARMA-DE model with other competitive models and found that it outperformed other models for the long and short time predictions. The performance was measured based on the model's training time and accuracy. *Panda et al. (2021)* also used convolutional neural networks for multi-currency exchange rate prediction. They have proposed a model that can develop multivariate exchange rate information and use those features better. They have used the adaptive learning rate (ADAM) optimization technique to provide optimal weight for their proposed model.

Majhi et al. (2009) used low-complexity artificial neural network models for efficient exchange rate prediction. This study has developed two ANN models: functional link ANN (FLANN) and cascaded functional link ANN (CFLANN). The models involve nonlinear inputs and a simple ANN structure with one or two neurons. They have observed that CFLANN works better than FLANN having the least error.

Refenes et al. (1993) applied a multilayer perceptron network to predict the currency exchange rate and also discussed the convergence issues related to network architecture.

Galeshchuk & Mukherjee (2017) stated that time series models and shallow neural networks result in acceptable estimates in the future prices for exchange rates but perform poorly at predicting the direction of change. On the other hand, machine learning classifiers trained on input features curated based on domain knowledge produce better results.

Damrongsakmethee & Neagoe (2020) implemented a deep learning model with Long Short Memory (DLSTM) to predict the currency exchange rate of USD/THB and took financial inputs such as interest rate, gross domestic product rate (GDP), balance account, inflation rate, and balance of trade also a finite set of previous exchange rates. *Wang et al. (2021)* have used the model of CNN-TLSTM for USD/CNY exchange rate prediction. So going forward with neural networks and deep learning models has been a credible approach for us.

Baffour et al. (2019) integrated the GJR (Glosten Jagannathan and Runkle) model with Artificial Neural Networks (ANN) model for forecasting the volatility of five currency exchange rate. They have found a significant improvement by using the ANN-GJR hybrid model rather than using benchmark models. These findings influence us to use the CNN-LSTM hybrid model for forecasting the USD/CAD exchange rate.

In their research, *Roy et al. (2020)* compared and analyzed the Random Forest, Gradient Boosting Model and Deep Neural Network for predicting the stock of South Korean companies. In their research, Random Forest performed better than other models, and possible reasons for Deep Neural Networks' lower performance can be the relatively small data set and difficulties in data augmentation.

Bose et al. (2021) proposed a hybrid model by cascading Multivariate Adaptive Regression Splines (MARS) and Deep Neural Network (DNN) for predicting the closing price of the stock. They have found an accuracy of 92% while predicting the closing price of the exchange rate.

Our study has focused on the connection between the USD/CAD exchange rate in connection to the underlying Technical and Macroeconomics factors of exchange rate movement. This has proved the concept using different models. Researchers are yet to follow our approach for forecasting the exchange rate of USD/CAD in a dual floating currency setup. While doing the entire research project, we have followed some modeling, evaluating, statistical, and research methods for implementing our neural network models from here.

2 Research methodology

Forecasting the foreign exchange rates is a challenging task due to its complex dynamics. Different models and techniques can be used for exchange rate forecasting, and each with its advantages and limitations. Below is a list of some commonly used models and techniques that I will consider for forecasting the forex market:

1. **ARIMA (Auto Regressive Integrated Moving Average)**:
ARIMA models are widely used for time series forecasting. It makes use of lagged moving average in its analysis and prediction. It is widely used in forecasting time-series data.

2. **Linear Regression**:
This a statistical practice of calculating a straight line that specifies a mathematical relationship between two variables. Linear regression is defined as an algorithm that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events.

3. **Random Forest Regressor (RFR)**:
This is a machine learning algorithm that works with decision trees. It creates levels of trees, process and analyze data at each tree, and then combines the outputs of these multiple decision trees so as to come up with the final result. The accuracy level of RFR, and its flexibility with hyper parameters made it highly adopted for regression and classification models.

4. **Gradient Boosting Machines (GBM)**:
Gradient Boosting Machine transforms a weak learner into a strong one. Every new tree is fitted on the modified version of the existing data. Thus its logic is centered on combining the next model that is good with previous models, so as to reduce the overall forecasting error. The key idea is to set the target outcomes for this next model in order to minimize the error.

5. **Extreme Gradient Boosting Machines (XGBoost)**:
XGBoost, is decision tree that is scalable with gradient-boosted tree. XGBoost ranks high among the machine learning libraries recommended for classification and regression problems. Its design makes it highly flexible, and efficient.

6. **Recurrent Neural Networks (RNNs)**:

RNN is a classified as a deep learning model. It is specially trained for sequential data like sentences, words, and time-series. It takes sequential data as its input and as well return sequential data as its output. It handles complex syntax rules and semantics properly.

7. **Ridge Regression**:

This is a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated. It has been used in many fields including econometrics, chemistry, and engineering. It includes regularization, which helps prevent overfitting by accepting multiple features.

8. **LASSO Regression**:

LASSO is a linear regression model that conglomerates both regularization and variable selections so as to arrive at the best prediction accuracy. It also provides clear and resulting statistical model that is simple to interpret. It avoids overfitting by applying penalty.

9. **Vector Auto Regression (VAR)**:

VAR model is a workhouse multivariate time series model that relates current observations of a variable with past observations of itself and past observations of other variables in the system.

10. **Long Short-Term Memory (LSTM)**:

LSTMs use a cell state to store information about past inputs. This cell state is updated at each step of the network, and the network uses it to make predictions about the current input. The cell state is updated using a series of gates that control how much information is allowed to flow into and out of the cell.

11. **Support Vector Regression (SVR) Model**:

SVR is an ML algorithm specially used in addressing regression analysis problems. SVR finds functions that estimates the relationship between a continuous target variable and the input variables and as well minimizing the chances of prediction errors.

12. **Bayesian Ridge Regression Model**:

Unlike traditional linear regression, which uses point estimates for parameters, Bayesian Ridge Regression treats parameters as random variables and calculates the posterior distributions given in the data. It allows a natural mechanism to survive insufficient data or poorly distributed data by formulating linear regression using probability distributors rather than point estimates. The output or response is assumed, and drawn from a probability distribution rather than estimated as a single value.

13. **AdaBoost Regression Model**:
The idea behind AdaBoost is to combine the outputs of multiple weak learners to create a strong predictor. A weak learner is a model that performs slightly better than random guessing. The principle behind Ada boosting algorithms is that we first build a model on the training dataset and then build a second model to rectify the errors present in the first model. This procedure is continued until and unless the errors are minimized and the dataset is predicted correctly.

Technical Analysis Indicators:
Various technical indicators, such as Relative Strength Index (RSI), Moving Averages, and Moving Average Convergence Divergence (MACD), can be used for short-term exchange rate forecasting based on historical price and volume of data.

It's essential to note that no single model or technique is universally best for forecasting the forex market. The effectiveness of a forecasting model depends on factors such as the characteristics of the forex data, the forecasting horizon, the availability of relevant features, and the model's ability to adapt to changing market conditions. It is often advisable to use a combination of models and techniques and to continuously evaluate and update the forecasting models based on their performance and the evolving market dynamics.

2.1 Performance metrics

We have forecasted the USD/CAD exchange rate by using various models, and the outputs of those models are continuous values. To make sure how good or bad our models are, we have calculated the error of our model. The badness of a model is how much error is generated by that model while predicting the USD/CAD exchange rate. For measuring the error of our models, we have used Root Mean Square Error (RMSE) which is used widely as a performance metric for regression analysis, forecasting, climatology, and so on. RMSE is known as the Standard deviation of residuals. RMSE penalizes the error more than the Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) by squaring the error. For that reason, we have selected RMSE as our primary performance measurement metric, as well as Mean Absolute Error (MAE). Finally, to check how well our models fits the data, we have calculated the R-squire value (R^2). Equations (1), (2), (3), and (4) shows the formulas used in calculating MSE, RMSE, MAE, and R^2 respectively.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{1}$$

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

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|y_i - x_i\right|}{n} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}_i\right)^2} \tag{4}$$

Where y_i is the actual value of the USD/CAD exchange rate, $\hat{y}_i$ is the predicted value of the USD/CAD exchange rate, and N denotes the total number of predictions or actual values. The model is better when the RMSE and MAPE are lower, and the model is bad when the RMSE and MAPE are high. The R^2 takes values from negative Infinity to 1. The closer the value of R^2 to 1, the better the model is.

2.2 Datasets

In the context of modeling, a dataset is a collection of data that is used for training, testing and validating a predictive model. Below are some key aspects of datasets with respect to modeling:

Features:
A dataset is made up of one or more features (also known as independent variables). Features can be numeric, categorical, or textual, and they capture the characteristics of the data that are relevant for modeling.

Target Variable:
The target is a dependent variable. Based on the features supplied, this is what the model aims to forecast. The target variable represents the outcome or response of interest, and the model learns to make predictions about it based on the input features.

Observations:
Every instance or row in the dataset is a representation of a single observation or sample. Observations may correspond to individual event, data points, entities, or transactions depending on the specific application.

Training, Testing, and Validation Sets:
Datasets are typically divided into subsets for the purpose of training the model, testing the model, and validating the model. The training set is used to train the model, the validation set is used to tune model hyperparameters and evaluate model performance during training, and the testing set is used in assessing and evaluating the final performance of the trained model on the futuristic data.

Data Quality:
Ensuring the quality and integrity of the dataset is crucial for building reliable models. This includes handling missing values, checking for data consistency and correctness, and performing data preprocessing and normalization based on need.

Size and Complexity:
The size and complexity of the dataset can vary depending on the specific modeling task and the amount of available data. Larger datasets may provide more representative samples of the underlying population and allow for more robust model training, while smaller datasets may require careful handling to avoid overfitting.

Overall, datasets play a central role in the modeling process, serving as the foundation upon which predictive models are built, trained, and evaluated. The choice of dataset, along with appropriate data preprocessing and feature engineering techniques, significantly impacts the performance and accuracy of the resulting models. On this research work, we sourced our data from three major categories as explained in the following sub-sections.

2.2.1 Statistical data

Statistical data in the Foreign Exchange market is maintained at various degrees which include minutes, hourly, daily, weekly, and even monthly. On my research work, we relied on the daily data. My data were fetched from two different sources viz: Tick Data Suite (TDS) and Investing.Com[1].

Both the data from Tick Data Suite (TDS) and Investing.Com are downloaded manually, and uploaded to my GitHub folder before usage. However the data from Investing.Com is as twice bigger than that from Tick Data Suite (TDS), thus giving us the leverage of larger data size for Training and Testing, thereby coming out with better results. While TDS maintained USD/CAD data from year 2003 till date (with over 5,500 records), Investing.Com maintained USD/CAD data from 1982 till date (with over 11,100 records).

2.2.2 Technical data

RSI (Relative Strength Index):

This is a popular and regular momentum oscillator used in measuring the speed and changes in price movements. It fluctuates between 0 and 100. RSI is typically used in identifying the overbought and the oversold levels in a market condition. It is intended to graphically present the current, and historical strength or the weakness of the market based on the closing prices of a trading period considered. It is calculated as shown in equation (5):

$$\mathrm{RSI} = 100 - \left[\frac{100}{1 + \frac{n_{up}}{n_{down}}} \right] \tag{5}$$

Where n_{up} is the average of n-day up(gain) closes, and n_{down} is the average of n-day down(loss) closes.

[1] https://www.investing.com/currencies/usd-cad-historical-data

SMA (Simple Moving Average):

This is the average price over the specified period. It is one of the regularly used technical analytical indicator that helps in smoothing out the price data by creating a continuously updated average price over the specific period in consideration. SMA is calculated by adding up the prices of an instrument (or asset) for over a specified periods and then, dividing that total by the number of periods. That is to say, SMA is calculated by summing up the last "Y" period's closing prices, and then, dividing that sum by Y. It is calculated as shown in equation (6) below.

$$\text{SMA} = \frac{\text{Sum of prices over the last } n \text{ periods}}{n} \tag{6}$$

Where n = is the number of periods (e.g., days, hours, weeks) over which you want to calculate the average.

Exponential Moving Average (EMA):

This is a moving average type that gives more weight to recent data points than the older data points. It is calculated by applying a smoothing factor to the previous EMA value and the current price. EMA places a greater significance and weight on the recent data points. It is calculated as shown in equation (7):

$$\text{EMA}_t = \text{EMA}_{t-1} + \alpha * \left(\text{Close}_t - \text{EMA}_{t-1}\right) \tag{7}$$

Where: EMA_t is the Exponential Moving Average at time t

EMA_{t-1} is the Exponential Moving Average at the previous time period (t - 1)

α is the smoothing factor, which is calculated based on the number of period n used in the calculation. It is typically expressed as 2/(n + 1)

Close_t is the closing price of the asset at time t

MACD (Moving Average Convergence Divergence):

MACD is a popular technical analytical indicator that is used in identifying trend reversals and momentum. It is composed of three components: the signal line, the MACD line, and the histogram. The signal line is a nine-period EMA of the MACD line; whereas the MACD line is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA. Therefore, it is calculated as shown in equations (8), (9), and (10):

$$\text{MACD Line} = \text{EMA12} - \text{EMA26} \tag{8}$$

$$\text{Signal Line} = \text{9-period EMA of MACD Line} \tag{9}$$

$$\text{MACD Histogram} = \text{MACD Line} - \text{Signal Line} \tag{10}$$

Each time the MACD line crosses above the Signal line, it is interpreted as a bullish signal, meaning that the trend is potentially going upwards, and thus, suggests a market BUY opportunity.

Each time the MACD line crosses below the Signal line, it is interpreted as a bearish signal, meaning that the trend is potentially heading downwards, and thus, suggests a market SELL opportunity.

2.2.3 Macroeconomics data

Macroeconomic indices refer to the economic indicators and variables that influence the overall performance of an economy. These factors provide insights into the health and direction of an economy and play a crucial role in shaping economic policies, investment decisions, and business strategies. Below are some of the common macroeconomic factors which have great impacts on United States' (USA) and Canadian (CAN) economy.

Gross Domestic Product (GDP):

GDP is as a key indicator of economic growth of any country, and is regularly used in assessing the overall health of an economy. GDP measures the collective value of all produced goods, and services rendered within a nation's borders over a specific period of time. While GDP Growth measures the marginal increase/decrease of the country's economic performance, GDP Ratio measures the strength/ability of a country being able to settle its debts.

Inflation Rate:

Inflation is measured as the rate at which the prices of goods and services rises over a speculated range of time interval. It measures the rate of change of the prices of consumables goods.

In most cases, prices of goods and services rise over time, and in few occasions, prices can also fall. This is a situation called deflation.

Interest Rates:

An interest rate tells you how high the cost of borrowing is, or how high the rewards are for saving is. So, if you're a borrower, the interest rate is the amount you are charged for borrowing money, shown as a percentage of the total amount of the loan.

Unemployment Rate:

The unemployment rate measures the percentage of workers in the labor market who do not currently have a job, and are actively looking for work.

Consumer Price Index (CPI):

The Consumer Price Index is measured as the, average change over a range of time, in the prices of consumer goods and services. It is a strong, key indicator of inflation. It also reflects to the changes in the normal cost of living. Central banks, policymakers, and businesses, closely monitor Consumer Price Index so as to assess price stability, and possibly make monetary policy decisions.

Imports and Exports:

Imports and exports refer to the goods and services a country buys from and sells to other countries, respectively. The balance between imports and exports affects a country's trade balance, and overall economic growth. Trade deficits (when imports exceed exports) or surpluses (when exports exceed imports) can impact exchange rates, domestic industries, and employment levels.

Government Revenue:

Government revenue includes all income generated by the government through taxes, fees, tariffs, and other sources. It is used in funding public expenditures, such as infrastructure projects, social welfare programs, defense, and public services. Government revenue influences fiscal policy decisions, budget allocation, and economic development initiatives.

These macroeconomic factors are interconnected and play critical roles in shaping economic policies, influencing business decisions, and affecting the overall health of an economy. They can have complex and far-reaching effects on economies, financial markets, and individual businesses. Analyzing trends in CPI, imports, exports, and interest rates provides valuable insights into inflationary pressures, international trade dynamics, fiscal sustainability, and economic growth prospects. Policymakers, investors, businesses, and analysts closely monitor these factors to assess economic conditions and make informed decisions.

The Macroeconomics data were sourced from the Federal Reserve Economic Data (FRED)'s[2] database through an API. We are considering nine distinct macroeconomic factors as our features for training the models. Being the fact that USD/CAD is a currency combination pair for two countries, we will consider each of these nine indices for the two countries involved. Below are the macroeconomic factors:

1. Consumer Price Indices (CPI) of USA
2. The Interest Rates of USA
3. Imports of USA
4. Exports of USA
5. Un-Employment Rates of USA
6. Consumer Price Indices (CPI) of Canada
7. The Interest Rates of Canada
8. Imports of Canada
9. Exports of Canada
10. Un-Employment Rates of Canada

We combined these ten macroeconomic factors with the technical tools to form a rich and comprehensive feature as an input to our model training and testing on USD/CAD exchange rate forecasting.

[2] https://fred.stlouisfed.org/

2.3 Modelling

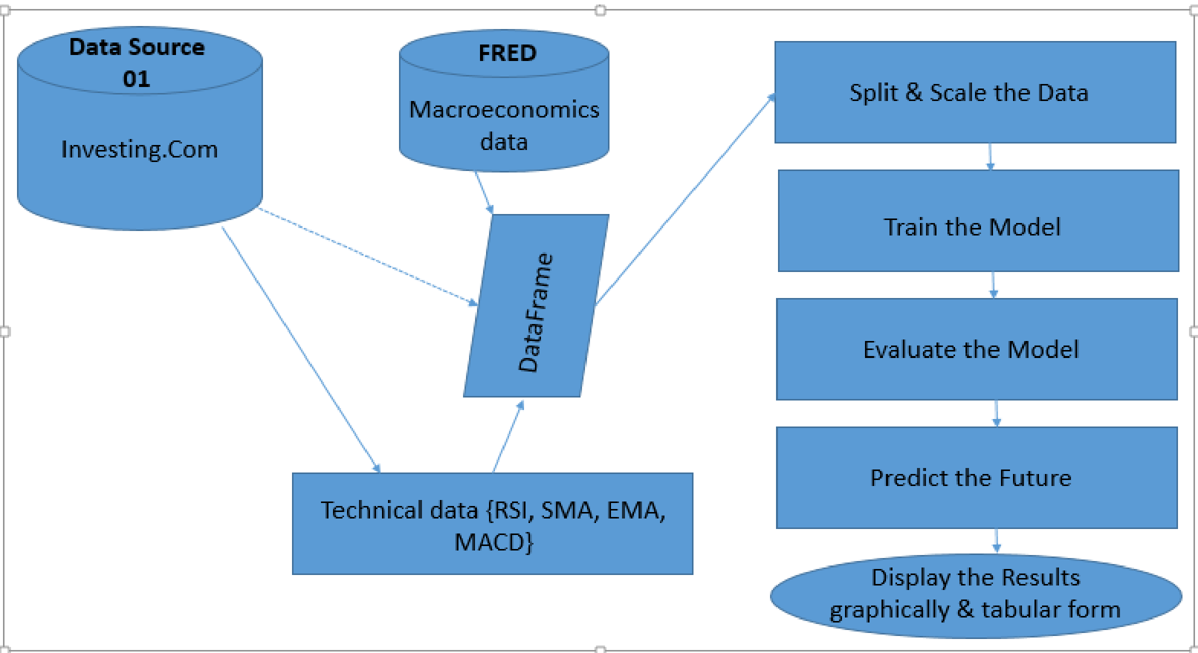

Figure 1. The Model's Flow diagram

The statistical data was supplied from Investing.Com. These are daily movements that tracks the changes in rates of USD against CAD. The basic information stored are **Open**, **High**, **Low**, and **Price**.

Open represents the exchange rate value at the beginning of each day.

High represents the highest rate recorded for the day.

Low represents the lowest rate recorded or the day.

Price represents the exchange rate at the end of the day. For a business day that is still active, Price represents the rate at any moment of that day.

The Macroeconomics data were all fetched from the Federal Reserve Economic Data (FRED)'s server through an API. For the purpose of this research, we concentrated on nine features which impacts the economy most, and thus affects the exchange rates. These features has been listed above in section 2.2.3 of this work.

The Technical data like RSI, SMA, EMA, and MACD were all calculated from the downloaded Statistical data using its formulas as was enumerated in section 2.2.2.

2.3.1 Models considered

Table 1. Models explored and compared

ModelName	MAE	MSE	RMSE	R^2_Train	R^2_Test	Comment
Linear Regression	0.009694	0.000148	0.012171	0.994074	0.939845	Good Model.
Random Forest Regressor	0.011429	0.000203	0.014234	0.999713	0.917730	Good Model.
XGBoost Regression	0.009695	0.000147	0.012126	0.994891	0.940292	Good Model.
Ridge Regression	0.009634	0.000146	0.012092	0.994068	0.940625	Good Model.
Lasso Regression	0.009127	0.000131	0.011449	0.993642	0.946768	Good Model.
Support Vector Regression	0.072223	0.007029	0.083837	0.112613	-1.854152	Very Bad Model.
Bayesian Ridge Regression	0.009662	0.000147	0.012130	0.994073	0.940252	Good Model.
AdaBoost Regression	0.013612	0.000290	0.017028	0.990580	0.882256	Bad Model.

As shown in Table 1 above, eight different Time series models were explored so as to get the best model suitable for predicting USD/CAD rate change. Several metrics were also applied in analysing the models. The closer MAE (Mean Absolute Error) is to zero, the more precise the model is. The closer MSE (Mean Square Error) is to zero, the better the model. The closer RMSE (Root Mean Square Error) is to zero, the more accurate the model prediction is. The R^2 error is a metric that measures the proportion of the variance in the target variable. The closer the R^2 is to 1.00, the better the Train/Test data set will be fit for the model. Under the Comment column, we have values as “Good Model”, “Bad Model”, and “Very Bad Model”. A model is classified as a Good Model if the values both R^2_Train and R^2_Test are greater than 0.90. A model is classified as a Bad Model if R^2_Train and R^2_Test have positive values, but one of them is less than 0.90. A model is classified as a Very Bad Model if the value of either R^2_Train or R^2_Test is less than zero.

Thus, Linear regression, Random Forest Regressor (RFR), Ridge Regression, and Bayesian Ridge Regression models performed well based on the result above.

2.3.2 Modelling approach applied

As shown in Figure 1 above, the decision of the data source is made by the user. After which the data is downloaded into a DataFrame, and merged with the Microeconomics data using the **Date** field as a common column between the two DataFrames. For the purpose of this research work, I relied on the statistical data from Investing.Com as it is more than twice the data from

Tick Data Suite (TDS). Due to the large size of the data from Investing.Com, I obtains a better accuracy for my modelling and forecasting results. At splitting stage, 10% of the data is reserved for testing, while the remaining 90% is used for training the model. The model is evaluated using different metrics as enumerated in section 2.1, and after which, forecasting follows.

2.3.3 Adopted pipeline for predicting USD/CAD exchange rate

Among the top four in the models explored, irrespective of the exceptional performances from Linear Regression and Bayesian Ridge Regression, Random Forest Regressor(RFR) is the most robust as it accepts unlimited number of hyperparameters. It will fit in well with many ranges of time-series data. RFR accepts Hybrid modelling approach such that Statistical data, unlimited Technical features, and unlimited Macroeconomics data can be combined together to come up with a high quality output. Random Forest Regressor mitigates the risk of overfitting, which is crucial when dealing with financial data where noise and outliers are common. By aggregating predictions from multiple decision trees trained on different subsets of the data, Random Forest Regressor reduces variance and generalizes well to unseen data, thus enhancing robustness and reducing the risk of overfitting. The dynamic of its operation makes it most acceptable. It adopts the Kalman filters approach, and operates in such a way that forecasting is iterated. The result of every forecast (n-1) forms part of the data that will be used in forecasting the next value. If for instance, if the RFR model is to predict the exchange rate value for next five days, day one is first predicted, and this rate value from day one will form part of the input for establishing the value or the next day, and so on. Take a look at SMA (Simple Moving Average) which is calculated as sum of the Closed Priced for n days divided by n. In this SMA instance, the predicted value is appended, while the oldest value in the series (based on Date) is popped off.

The time of training, validation, and testing data for Random Forest has been kept almost as the same as the time of our baseline models so that Random Forest can predict the fluctuations of the USD/CAD exchange rate for the same time as our baseline models have predicted the USD/CAD exchange rate. All hyperparameters have been set as default. After training Random Forest using the new dataset, this Random Forest model can predict the USD/CAD exchange rate fluctuations. These predicted fluctuations is a kind of noise generated from the fluctuations of features. If we can deduct some fluctuations from our baseline model predictions, the predictions will be closer to the actual values. First, the predictions of USD/CAD exchange rates provided by the baseline models have been calculated. Also, the fluctuations in the USD/CAD exchange rate predicted by Random Forest for the same time used for USD/CAD exchange rates prediction by our baseline models have been calculated. The baseline models provided the USD/CAD exchange rates for a particular time, and the Random forest provided the predicted fluctuations in the USD/CAD exchange rate for that same time.

Finally, the fluctuations have been deducted from the predicted USD/CAD exchange rates by our baseline models. The predicted fluctuations by Random Forest can be positive or negative. Suppose a single value of the USD/CAD exchange rate, predicted by our baseline models, is greater than the actual USD/CAD exchange rate. In that case, the fluctuation value predicted by Random Forest for that single value of the USD/CAD exchange rate should be positive. In that case, subtracting the fluctuation from the predicted USD/CAD exchange rate will drive the predicted USD/CAD exchange rate closer to the actual USD/CAD exchange rate. Suppose the predicted USD/CAD exchange rate is less than the actual USD/CAD exchange rate. In that case, the predicted fluctuation value should be negative so that it becomes positive while doing subtraction. As a result, the predicted USD/CAD exchange rate gets closer to the actual exchange rate.

Error is generated when there is a gap between the predicted and actual values. The main concept of our proposed pipeline is to make the predicted USD/CAD exchange rate closer to the actual exchange rate so that the error minimizes. But the problem is, let’s say, for example, for the first ten values, the predicted values are greater than the actual values. The predicted values for the next five values are less than the actual ones. Now, if we subtract a small fixed value from all the predicted values, the first ten predicted values will get closer to the actual values because those values were greater than the actual values. But the next five values will generate more error than the previous because those five values were less than the actual values, and subtracting a small value will create more distance from the actual values. Adding a small value will make the last five prediction values closer to the actual values, but again it will generate more distance for the first ten predicted values. So it is clear that adding or subtracting a small discrete value from the prediction value will not ensure that all the predicted values will get closer to the actual value. So it cannot ensure that the overall error will be reduced. Moreover, the gap between the predicted value and actual value is not equal and unidirectional everywhere, meaning predicted values are not either greater or less than actual values in all the situations. So choosing the small discrete value is also very critical. To address this issue, the small value we want to add or subtract will not be a fixed discrete value. Rather than considering a single small discrete value, we will consider different small values for different data points. Those different small values will be the predictions of Random Forest that we have introduced as predicted fluctuations. As we have mentioned earlier, the Random Forest model has been trained by using the features' fluctuations, and the Random Forest's prediction is the fluctuation in the USD/CAD exchange rate. That means the predictions of Random Forest describe how much the price of USD/CAD can fluctuate. So the USD/CAD exchange rate fluctuation will differ for every data point. Those fluctuation values can be both positive and negative. Now, if we deduct the fluctuation values from the predicted values, there is a chance that the predicted values will get closer to the actual values. Because now we are not using a small discrete value for all the data points. Each small value (fluctuation) is dedicated to a single data point as specific small values are predicted for specific data points by Random Forest. The unidirectional problem has been solved because the predicted small values (fluctuations) of Random Forest can be both positive and negative. So if the predicted small value (fluctuation) is positive, the predicted USD/CAD exchange rate will be reduced after subtracting. Again, if the predicted small value (fluctuation) is negative, then the USD/CAD exchange rate will be increased after subtracting. However, it is impossible to predict each fluctuation value correctly. In our experiment, we found an improvement in RMSE for all our models while applying this proposed pipeline. This indicates that most of the fluctuation predictions were correct. Therefore, after subtracting the fluctuations from the predicted USD/CAD exchange rates, the predicted USD/CAD exchange rates got closer to the actual price.

Random Forest Regressor (RFR) is a supervised learning algorithm belonging to the ensemble learning family, specifically the random forest method, used for regression tasks. It is an extension of the Random Forest algorithm for classification. Here is how Random Forest Regressor works:

Ensemble of Decision Trees:

This is a machine learning technique that combines multiple decision trees to make better predictions or classifications. It is like having a group of experts who collaborate to give you the most accurate answer. Each decision tree in the ensemble works like a simple “if-then” rule system. Every decision tree is randomly trained on a subset of the training data, and features.

Decision Tree Construction:

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes. Each decision tree in the ensemble is constructed by recursively partitioning the feature space into regions, where each region corresponds to a leaf node in the tree.

Bootstrap Aggregating (Bagging):

Bagging is the ensemble learning method which is regularly used in reducing variance within a noisy data set. In bagging, sample data in the training set is randomly selected with replacement. This means that a single data point can be chosen many times. To reduce overfitting and improve performance, Random Forest Regressor applies this bagging technique when several decision trees are to be trained on different subsets of the training data.

Prediction:

During prediction, Random Forest Regressor aggregates the predictions of all decision trees in the ensemble to make the final prediction. For regression tasks, the predictions are typically averaged across the ensemble.

Hyperparameters:

Random Forest Regressor has so many parameters that can be tuned to optimize its performance. These are parameters like the maximum depth of each tree, the number of trees in the ensemble, the maximum number of features to consider when looking for the best split, and the minimum number of samples required to split a node.

Random Forest Regressor is known for its robustness, scalability, and ability to handle high-dimensional data with a large number of features. It often performs well across a wide range of regression problems and is less prone to overfitting compared to individual decision trees.

3 Discussion and Results

3.1 Discussion

We used three sets of data in this research. These are the **Statistical** data (also called the OHCL). These data are real representation of what happened in foreign exchange (FOREX) market. Our statistical data spans from the year 1982 till November 2024 with over 11,100 records. Our second set of data are the **Technical** data. In FOREX, we have many proven technical indicators that assists investors in determining the direction of the market based on certain behavioral patterns. These technical indicators are not limited to RSI, EMA, SMA, MACD, Bollinger Bands, ATR, and Stochastic Oscillator. The values for any of these technical indicators are derived from the Statistical (OHCL) data. Thus, Technical data is a direct derivate from statistical data. The last set of data used in this research is the **Macroeconomics** data. Macroeconomics data are information from the several indices that directly affects the economic health of a country as well as living standard of individuals in that country. These indices are not limited to GDP Growth, GDP Ratio, Imports, Exports, Interest rates, Inflation rates, Consumer Price Index, and Government Revenue.

Our Technical data was downloaded from Investing.Com, while our Macroeconomics data is downloaded from FRED (Federal Reserve Economic Data) through an API. Data from the three sources were merged together using the Date column as the common key among them. Many records for the early 80's were removed. The reason for their removal is that some of the Macroeconomics indices considered in this research do not have data as at that time. Instead of having them as zeros, we removed the affected rows.

The data was Splitted, Scaled, Trained, and Tested on several time-series models which includes Linear Regression, Random Forest Regressor, XGBoost Regression, Ridge Regression, Lasso Regression, Support Vector Regression, Bayesian Ridge Regression, and AdaBoost Regression model. Some of these models have exposed hyperparameters for tuning and optimization. Table 2 below lists these models with their optimized hyperparameters. While scaling my data, I choose MinMaxScaler with range(0, 1). I choose this scaler because it is the best scaler for a time-series data like Forex OHCL (Open, High, Close, Low) data. MinMaxScaler works well on bounded data like the prices that falls within certain range of value, and thus, the relative difference of the original data is preserved. Since all the values of OHCL are positive values, the range (0, 1) is recommended against (-1, 1).

Table 2. Explored models with ptimized parameters

	Model	**Optimized Parameters**
1	Linear Regression	None
2	Random Forest Regressor	max_features='log2'
3	XGBoost Regression	max_depth = 2
4	Ridge Regression	alpha = 0.01
5	Lasso Regression	alpha = 0.00005
6	Support Vector Regression	C = 0.00001
7	Bayesian Ridge Regression	None
8	AdaBoost Regression	n_estimators = 101

For the sake of comparison, the model was tested on another source of Technical data named Tick Data Suite (TDS). This data fromTDS spans the year from 2003 till November 2024. It is approximately half the size of the technical data from Investing.Com. The models were also tested on another source of Macroeconomics data which is the World Bank's database. The macroeconomics data from the World Bank is maintained and updated on yearly basis, while that from FRED is maintained and updated in monthly basis (with some of them quarterly basis). The result from the initial two sources remains a better result. The results of my comparison with different data sources and methodologies are displayed in Table 3 below.

Table 3. Comparing Models with Data sources explored

Models	Metrics	Investing/FRED	TDS/FRED	Investing/WorldBank
Linear Regression	MAE	0.009694	0.010249	0.010944
	RMSE	0.012171	0.012847	0.013587
	R^2_Train	0.994074	0.988130	0.993928
	R^2_Test	0.939845	0.908703	0.922097
Random Forest Regressor	MAE	0.011429	0.015146	0.013020
	RMSE	0.014234	0.018380	0.016058
	R^2_Train	0.999713	0.999406	0.999656
	R^2_Test	0.917730	0.813114	0.891178
XGBoost Regression	MAE	0.009695	0.011689	0.009235
	RMSE	0.012126	0.014317	0.011697
	R^2_Train	0.994891	0.990652	0.994549
	R^2_Test	0.940292	0.886613	0.942260
Ridge Regression	MAE	0.009634	0.010196	0.010756
	RMSE	0.012092	0.012766	0.013357
	R^2_Train	0.994068	0.988119	0.993924
	R^2_Test	0.940625	0.909844	0.924707

Models	Metrics	Investing/FRED	TDS/FRED	Investing/WorldBank
Lasso Regression	MAE	0.009127	0.009331	0.009143
	RMSE	0.011449	0.011580	0.011461
	R^2_Train	0.993642	0.987471	0.993451
	R^2_Test	0.946768	0.925820	0.944572
Support Vector Regression	MAE	0.072223	0.149788	0.061156
	RMSE	0.083837	0.155897	0.073537
	R^2_Train	0.112613	0.114638	0.105346
	R^2_Test	-1.854152	-12.444989	-1.282029
Bayesian Ridge Regression	MAE	0.009662	0.010204	0.010822
	RMSE	0.012130	0.012777	0.013438
	R^2_Train	0.994073	0.988122	0.993926
	R^2_Test	0.940252	0.909682	0.923797
AdaBoost Regression	MAE	0.013612	0.012066	0.012639
	RMSE	0.017028	0.015481	0.015819
	R^2_Train	0.990580	0.986062	0.990293
	R^2_Test	0.882256	0.867426	0.894394

We leveraged on Python libraries to achieve this research work. Few of the mostly used libraries are numpy, pandas, fredapi, requests, matplotlib, datetime, and scikit-learn

3.2 Result

We have predicted the USD/CAD exchange rate using all our models and calculated the error of our models (Table 1) by using RMSE and MAE. Most of the researchers uses the RMSE for the evaluation of their models. For that reason, the RMSE has been considered the main evaluation metric. To check the fitting of our data against the models, we considered the R^2 score as a metric. Then we plotted the actual value of the USD/CAD exchange rate price versus the predicted price (Figures 4). After that, we applied our proposed pipeline to our models to check the effectiveness of the proposed pipeline. Finally, we have compared our results with other research.

Figure 2. Last fifty years data on USD/CAD[3]

[3] https://tradingeconomics.com/Canada/currency

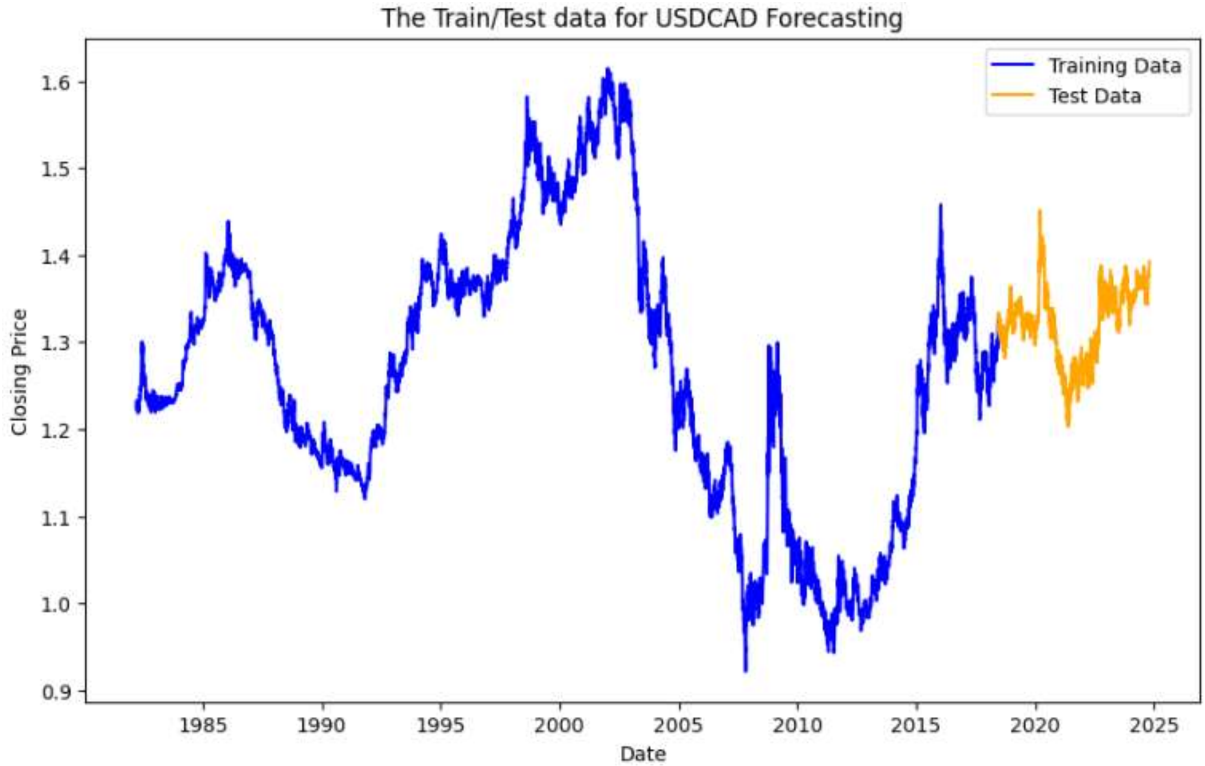

Figure 3. Train-Test Data (Data downloaded as from 1982)

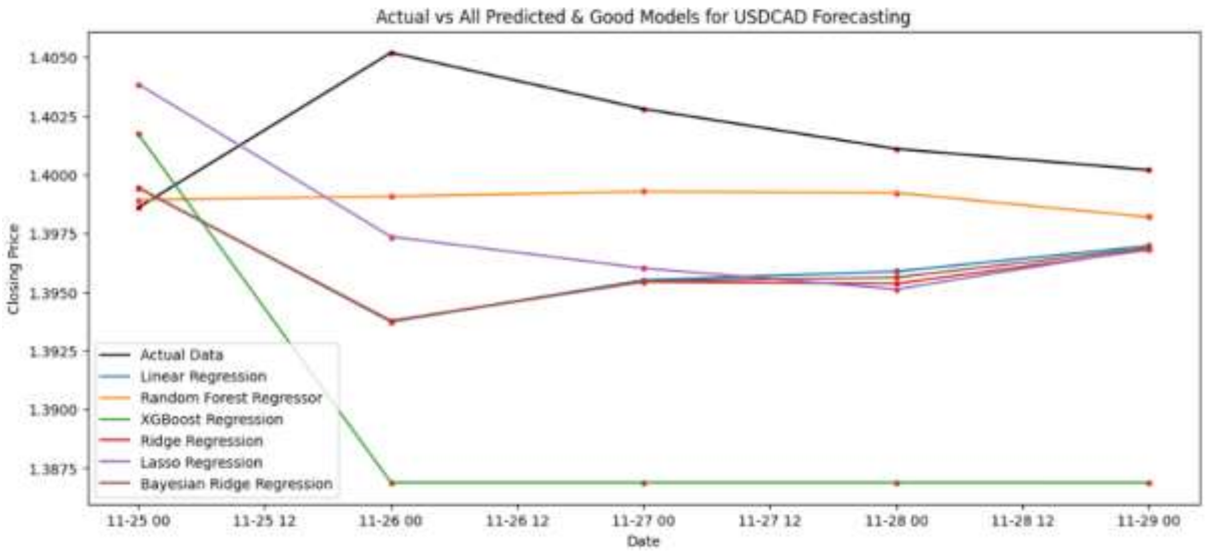

Figure 4. Actual vs All Predicted Models using macroeconomics data

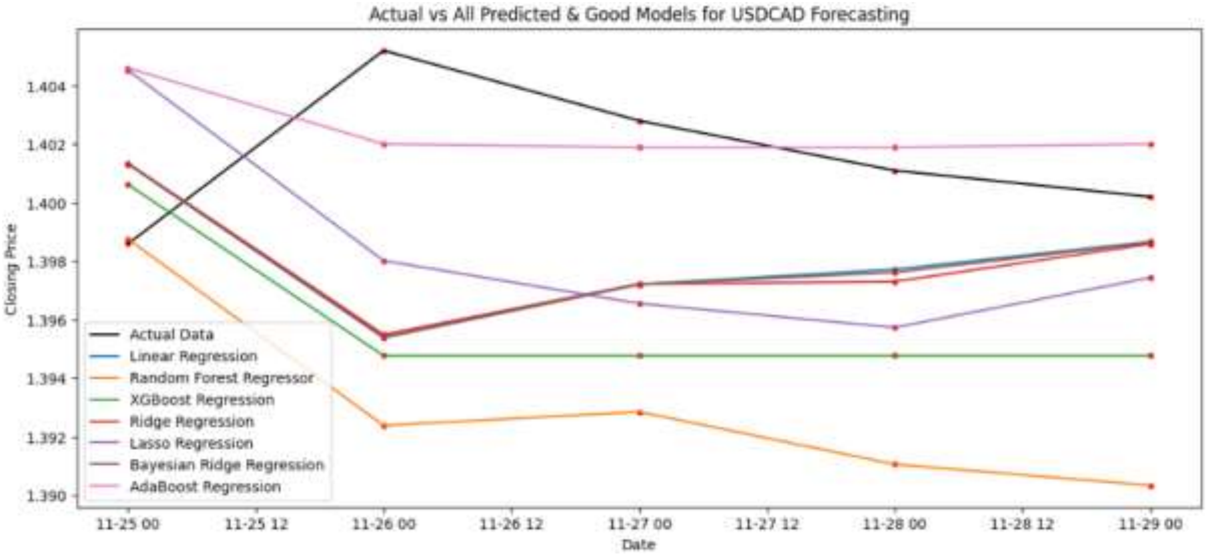

Figure 5. Actual vs All Predicted Models without macroeconomics data

The models performed well both with, and without macroeconomics data. It was expected that the models with macroeconomics data should outperform those without macroeconomics data. This expectation was not met due to the fact that our macroeconomics data source Federal Reserve Economic Data (FRED) maintains data in monthly (and some) quarterly basis. Thus we have less than five hundred records from FRED that is matching over eleven thousand statistical OHCL records. In the future, if we are able to have the macroeconomics data maintained on weekly or daily basis, our models with the macroeconomics data should be much better than that without the data.

4 Conclusion and Future proposals

Forecasting the exchange rate of USD/CAD more accurately can make the FOREX market a secure and reliable place for investors. Previously many researchers tried to forecast the exchange rate of USD/CAD by using time series models, machine learning models, and deep learning models but none of the research included factors that can directly affect the exchange rate of USD/CAD. To introduce a new scope for the researchers, we have included several Technical analytical metrics with ten most important macroeconomic factors in our dataset. These macroeconomic factors are directly correlated with the USD/CAD exchange rate. In our research, we got a very good RMSE of 0.012171(Linear Regression), 0.012130(Bayesian Ridge Regression), and 0.014234(Random Forest Regressor). In the future, research can be conducted using advanced models with these macroeconomic features to determine if the predictions are even more accurate on the other Time-Series models. This study further triggered the need of brainstorming on the existing exchange rate forecast methodologies, and introducing more effective Technical indices and Macroeconomic factors.

As it is, our Macroeconomics feature were fetched from the Federal Reserve Economic Data (FRED)'s database through an API. These data where maintained in Monthly and Quarterly basis. Future research may help us by, gaining access to these macroeconomics data where it is maintained at a more frequent intervals like weekly or even daily basis.

References

Akhtar, M. A., & Hilton, R. S. (1984*)*. Effects of exchange rate uncertainty on German and US trade. *Federal Reserve Bank of New York Quarterly Review*, 9(1), 7–16. https://www.newyorkfed.org/medialibrary/media/research/quarterly_review/1984v9/v9n1article2.pdf

Akubue S. (2024). Forecasting Currency Exchange Trend on USD/CAD. https://github.com/Akubue4U/PyProjects/blob/0c7d22c02b7c549208e27e3950603332c27bca8b/MyBDAThesis/ForecastingCurrencyExchangeTrendOnUSDCAD.pdf

Baffour, A. A., Feng, J., & Taylor, E. K. (2019). A hybrid artificial neural network-GJR modeling approach to forecasting currency exchange rate volatility. *Neurocomputing*, 365, 285–301. https://doi.org/10.1016/j.neucom.2019.07.088

Biswas, A., Uday, I. A., Rahat, K. M., Akter, M. S., & Mahdy, M. R. C. (2023). Forecasting the United State Dollar(USD)/Bangladeshi Taka (BDT) exchange rate with deep learning models: Inclusion of macroeconomic factors influencing the currency exchange rates. *Plos One*, (1st ed., pp. 8-15, 18). https://doi.org/10.1371/journal.pone.0279602

Bose, A., Hsu, C. H., Roy, S. S., Lee, K. C., Mohammadi-Ivatloo, B., & Abimannan, S. (2021). Forecasting stock price by hybrid model of cascading Multivariate Adaptive Regression Splines and Deep Neural Network. *International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC)*, 95. https://doi.org/10.1016/j.compeleceng.2021.107405

Bouraoui, T., & Phisuthtiwatcharavong, A. (2015). On the determinants of the THB/USD exchange rate. *Procedia economics and finance*, 30, 137–145. https://doi.org/10.1016/S2212-5671(15)01277-0

Broil, U., & Eckwert, B. (1999). Exchange rate volatility and international trade. *Southern Economic Journal*, 66(1), 178–185. https://doi.org/10.2307/1060843

Carbureanu, M. (2011). The analysis of currency exchange rate evolution using a data mining technique. *Economic Sciences Series*, 63(3), 105–112. https://upg-bulletin-se.ro/old_site/archive/2011-3/11.%20Carbureanu.pdf

Cushman, D. O. (1988). US bilateral trade flows and exchange risk during the floating period. *Journal of International Economics*, 24(3-4), 317–330. https://doi.org/10.1016/0022-1996(88)90040-2

Damrongsakmethee, T., Neagoe, V. E. (2020). A Deep Learning Model with Long Short-Term Memory (DLSTM) for Prediction of Currency Exchange Rates. *Springer Link* (pp. 484–498). https://link.springer.com/chapter/10.1007/978-3-030-51971-1_40

Galeshchuk, S., Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management*, 24(4), 100–110. https://onlinelibrary.wiley.com/doi/10.1002/isaf.1404

Islam, M. S., & Hossain, E. (2020). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters*, 3, 100009. https://doi.org/10.1016/j.socl.2020.100009

Kemal, M. A. (2005). Exchange rate instability and trade. The case of Pakistan. *ResearchGate* https://www.researchgate.net/publication/237293233_Exchange_Rate_Instability_and_Trade_The_Case_of_Pakistan

Majhi, R., Panda, G., & Sahoo, G. (2009). Efficient prediction of exchange rates with low complexity artificial neural network models. *Expert systems with applications*, 36(1), 181–189. https://www.sciencedirect.com/science/article/abs/pii/S0957417407004356?via%3Dihub

Natto, K. I. (2010). Valuation Models for Floating Currency Economies and Pegged Currency Economies. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1545710

Nayak, S. C., Nayak, S. K., Panda, S. K., & Kumar, C. S. (2021). Cognitive Intelligence-Based Framework for Financial Forecasting. *CRC Press* (pp .205-222) https://www.researchgate.net/publication/350119284_Cognitive_Intelligence-Based_Framework_for_Financial_Forecasting

Panda, M. M., Panda, S. N., & Pattnaik, P. K. (2021). Multi currency exchange rate prediction using convolutional neural network. *Materials Today Proceedings*. https://www.researchgate.net/publication/348570098_Multi_currency_exchange_rate_prediction_using_convolutional_neural_network

Pandey, T. N., Jagadev, A. K., Dehuri, S., & Cho, S. B. (2020). A novel committee machine and reviews of neural network and statistical models for currency exchange rate prediction: An experimental analysis. *Journal of King Saud University-Computer and Information Sciences*, 32(9), 987–999. https://www.sciencedirect.com/science/article/pii/S1319157817303816?via%3Dihub

Ramasamy, R., & Abar, S. K. (2015). Influence of macroeconomic variables on exchange rates. *Journal of Economics, Business and Management*, 3(2), 276–281. https://www.joebm.com/index.php?m=content&c=index&a=show&catid=40&id=478

Refenes, A. N., Azema-Barac, M., Chen, L., & Karoussos, S. (1993). Currency exchange rate prediction and neural network design strategies. *Neural Computing & Applications*, 1(1), 46–58. https://link.springer.com/article/10.1007/BF01411374

Refenes, A. N., Zapranis, A., & Francis, G. (1994). Stock performance modeling using neural networks: A comparative study with regression models. *Neural networks*, 7(2), 375–388. https://www.sciencedirect.com/science/article/abs/pii/0893608094900302?via%3Dihub

Rehman, M., Khan, G. M., & Mahmud, S. A. (2014). Foreign currency exchange rates prediction using cgp and recurrent neural network. *IERI Procedia*, 10 239–244. https://www.sciencedirect.com/science/article/pii/S2212667814001312?via%3Dihub

Rout, M., Majhi, B., Majhi, R., & Panda, G. (2014*)*. Forecasting of currency exchange rates using an adaptive ARMA model with differential evolution based training. *Journal of King Saud University-Computer and Information Sciences*, 26(1), 7–18. https://www.sciencedirect.com/science/article/pii/S1319157813000037?via%3Dihub

Roy, S. S., Chopra, R., Lee, K. C., Spampinato, C., & Mohammadi-ivatlood, B. (2020). Random forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies. *International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC)*, 33(1), 62–71. https://www.inderscience.com/offers.php?id=104715

Saeed, A., Awan, R. U., Sial, M. H., & Sher, F. (2012). An econometric analysis of determinants of exchange rate in Pakistan. *International Journal of Business and Social Science*, 3(6), 184–196. https://ijbssnet.com/journals/Vol_3_No_6_Special_Issue_March_2012/23.pdf

Samui, P., Roy, S. S., & Balas, V. E. (2017). Handbook of neural computation. *Computers and Electrical*, 95. https://www.sciencedirect.com/science/article/abs/pii/S0045790621003700?via%3Dihub

Twin, A. (2022). 6 factors that influence exchange rates. *Investopedia* https://www.investopedia.com/trading/factors-influence-exchange-rates/

Wang, J., Wang, X., Li, J., & Wang, H. (2021). A Prediction Model of CNN-TLSTM for USD/CNY Exchange Rate Prediction. *IEEE*, 9, 73346–73354. https://ieeexplore.ieee.org/document/9431182

Yao, J., & Tan, C. L. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1-4), 79–98. https://www.sciencedirect.com/science/article/abs/pii/S0925231200003003?via%3Dihub

Zwanger, S. (2008). Determinants of exchange rates: the case of the chilean peso. *University of North Carolina Wilmington*. https://core.ac.uk/download/pdf/149230671.pdf