Forecasting Currency Exchange Trend

My Case Study is United States Dollar (USD) vs Canadian Dollar (CAD)

Master's Thesis

By: *Akubue Simon C*

On: *18-Dec-2024*

Motivation

Majority of the Forex traders lose their money due to inadequate knowledge on the dynamics of the market forces. Most of the times, market forces is determined by the Macroeconomics news from the countries around the world. Mitigating against this high rate of losses formed the basis of my research dream on "**Forecasting Forex Exchange Trend**"

Content

Pre-Amble

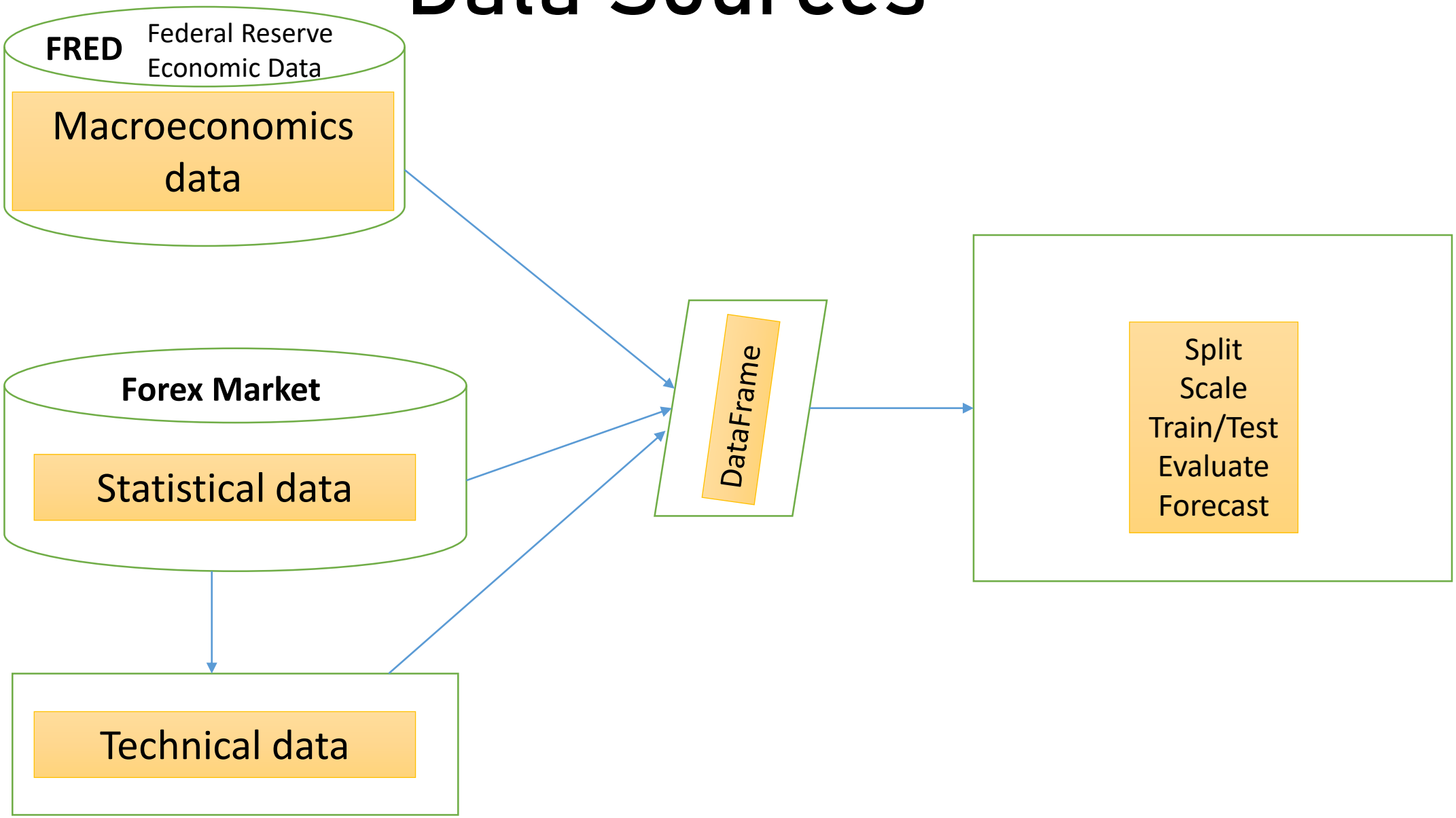
The foreign exchange (forex) market stands as one of the most **dynamic** and influential financial markets globally, facilitating the exchange of currencies and shaping international trade and investment. Among the myriad currency pairs traded in this market, the USD/CAD pair occupies a significant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

Last five years data on USD/CAD

https://markets.ft.com/data/currencies/tearsheet/summary?s=USDCAD

Last fifty years data on USD/CAD
https://tradingeconomics.com/canada/currency

Data Sources



Data Source {*Statistical*}

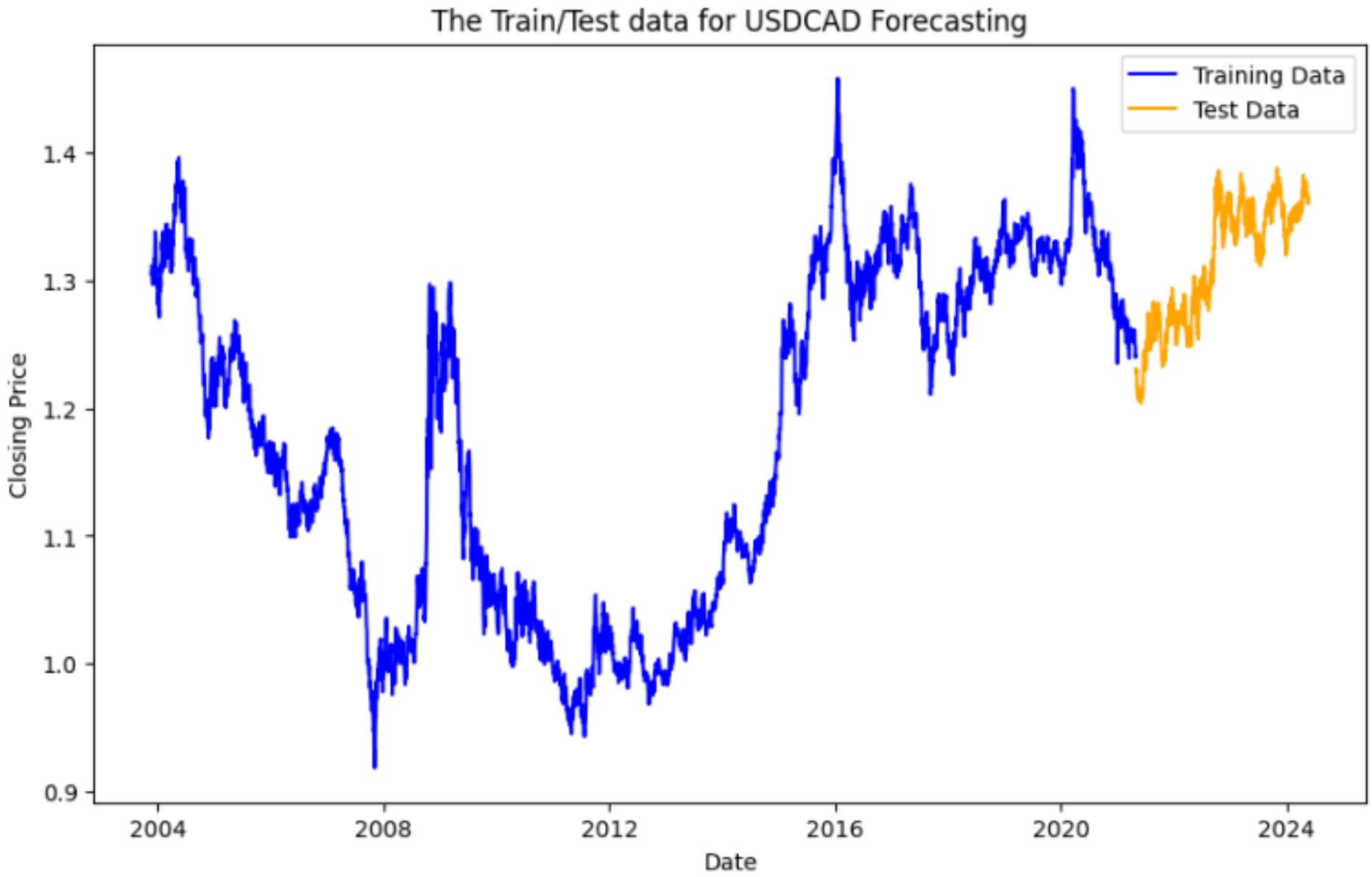
My Data Set is taken from either of these two sources:

(a) Tick Data Suite (TDS) ➔ From year 2003 to May-2024 {with over 5,500 records}

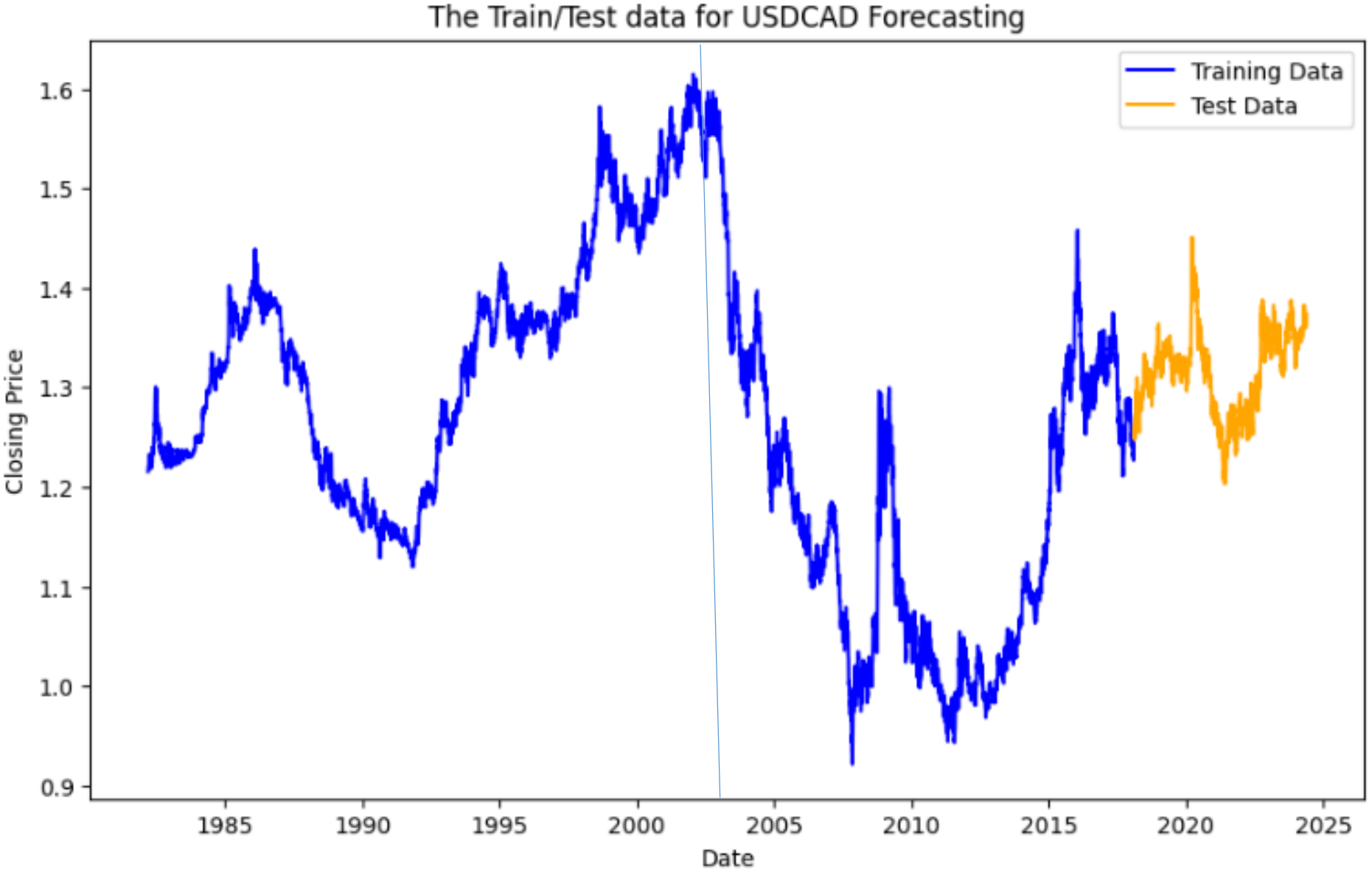
(b) Investing.Com ➔ From 1982 to May-2024 {with over 11,100 records}

The data in either of these two cases is maintained on a daily bases.

Raw data from TDS



Raw data from Investing.Com



Data Source {*Macroeconomics*}

The following Macroeconomics data played a strong role in my model. They were sourced from the Federal Reserve Economic Data (FRED)'s(https://fred.stlouisfed.org/) database through an API that was exposed as XMLs.

Consumer Price Index (CPI) of USA
The Interest Rates of USA
Imports of USA
Exports of USA
Un-Employment Rates of USA

Consumer Price Index (CPI) of Canada
The Interest Rates of Canada
Imports of Canada
Exports of Canada
Un-Employment Rates of Canada

Data Source {*Technical*}

The following technical indicators were integrated:

(a) RSI – Relative Strength Index

This is a momentum indicator that measures the magnitude of recent price changes to analyze overbought or oversold conditions

(b) SMA - Simple Moving Average

This is the average price over the specified period. It is calculated by adding up the last "X" period's closing prices and then dividing that number by of Xs

Data Source {.....*Technical*}

(c) EMA – Exponential Moving Average

This is a type of moving average that gives more weight to recent data points compared to older data points. It is calculated by applying a smoothing factor to the previous EMA value and the current price.

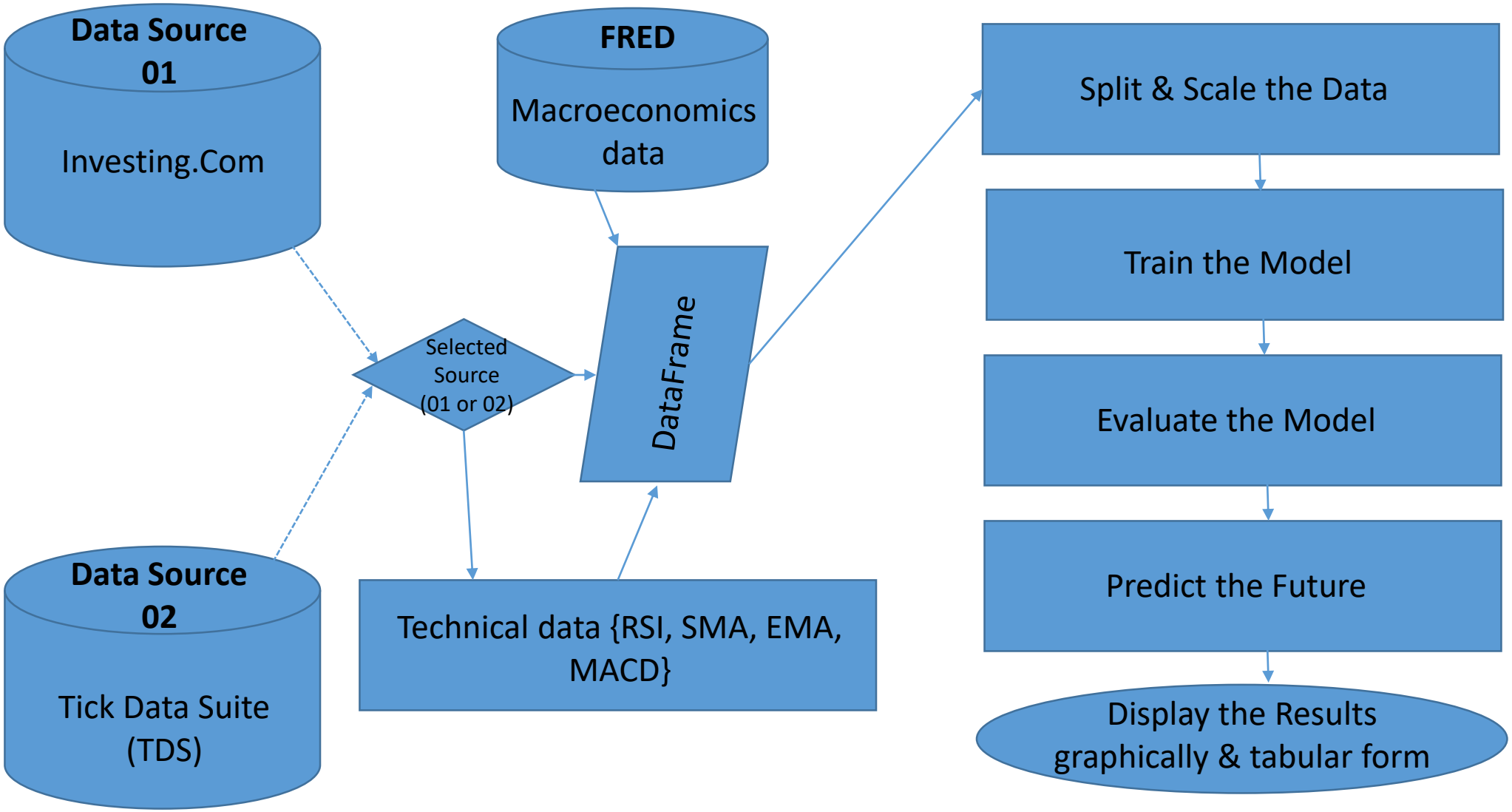
(d) MACD - Moving Average Convergence Divergence

This is a popular technical analysis indicator used to identify trend reversals and momentum changes in financial markets, including stocks, currencies, and commodities. It consists of three components: the MACD line, the signal line, and the histogram.

Algorithms

I am comparing many models like:
I am considered the following models for my forecasting:
XGBoost model
Lasso Regression model
Ridge Regression model
Linear Regression model
AdaBoost Regression model
Random Forest Regressor model
Bayesian Ridge Regression model
Support Vector Regression (SVR) model

The Model's Flow Diagram



Statistical Data

Date	Open	High	Low	Price
2024-11-22	1.3968	1.4020	1.3955	1.3980
2024-11-21	1.3975	1.3978	1.3928	1.3970
2024-11-20	1.3955	1.4014	1.3948	1.3972
2024-11-19	1.4012	1.4036	1.3952	1.3953
2024-11-18	1.4090	1.4102	1.4001	1.4014
2024-11-15	1.4054	1.4105	1.4031	1.4088
2024-11-14	1.3995	1.4064	1.3985	1.4057

Models Comparison {Data from Investing.Com}

90% Train – 10% Test Splitting, with FRED's Macroeconomics data

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.

Models Comparison {Data from TDS}

85% Train – 15% Test Splitting, with FRED's Macroeconomics data

ModelName	MAE	MSE	RMSE	R^2_Train	R^2_Test	Comment
Linear Regression	0.010249	0.000165	0.012847	0.988130	0.908703	Good Model.
Random Forest Regressor	0.015146	0.000338	0.018380	0.999406	0.813114	Bad Model.
XGBoost Regression	0.011689	0.000205	0.014317	0.990652	0.886613	Bad Model.
Ridge Regression	0.010196	0.000163	0.012766	0.988119	0.909844	Good Model.
Lasso Regression	0.009331	0.000134	0.011580	0.987471	0.925820	Good Model.
Support Vector Regression	0.149788	0.024304	0.155897	0.114638	-12.444989	Very Bad Model.
Bayesian Ridge Regression	0.010204	0.000163	0.012777	0.988122	0.909682	Good Model.
AdaBoost Regression	0.012066	0.000240	0.015481	0.986062	0.867426	Bad Model.

Models Comparison {Data from Investing.Com}

90% Train – 10% Test Splitting, with World Bank's Macroeconomics data

ModelName	MAE	MSE	RMSE	R^2_Train	R^2_Test	Comment
Linear Regression	0.010944	0.000185	0.013587	0.993928	0.922097	Good Model.
Random Forest Regressor	0.013020	0.000258	0.016058	0.999656	0.891178	Bad Model.
XGBoost Regression	0.009235	0.000137	0.011697	0.994549	0.942260	Good Model.
Ridge Regression	0.010756	0.000178	0.013357	0.993924	0.924707	Good Model.
Lasso Regression	0.009143	0.000131	0.011461	0.993451	0.944572	Good Model.
Support Vector Regression	0.061156	0.005408	0.073537	0.105346	-1.282029	Very Bad Model.
Bayesian Ridge Regression	0.010822	0.000181	0.013438	0.993926	0.923797	Good Model.
AdaBoost Regression	0.012639	0.000250	0.015819	0.990293	0.894394	Good Model.

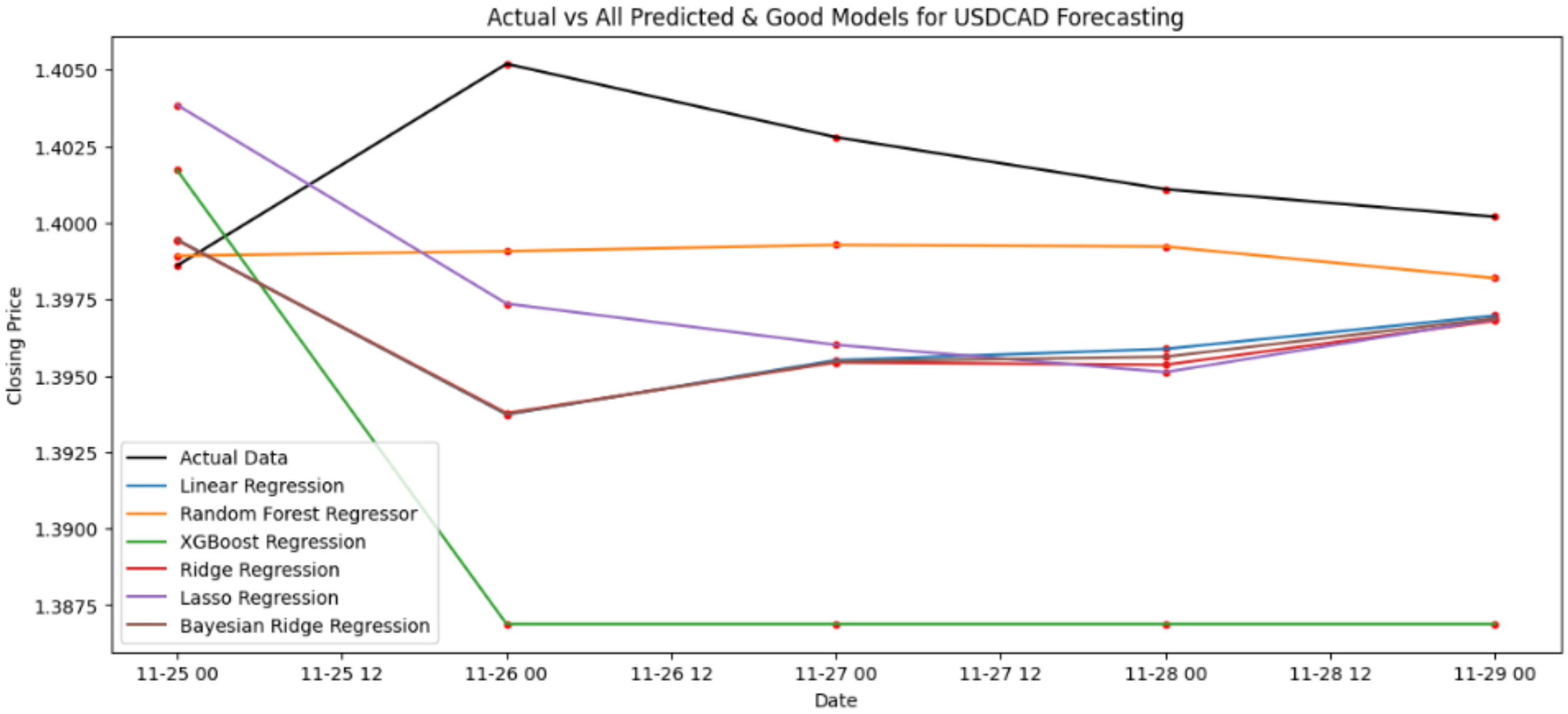
Models Comparison {Data from Investing.Com}

90% Train – 10% Test Splitting, **without** FRED's Macroeconomics data

ModelName	MAE	MSE	RMSE	R^2_Train	R^2_Test	Comment
Linear Regression	0.008967	0.000128	0.011329	0.993736	0.945837	Good Model.
Random Forest Regressor	0.009715	0.000150	0.012264	0.999458	0.936531	Good Model.
XGBoost Regression	0.009235	0.000138	0.011731	0.994462	0.941927	Good Model.
Ridge Regression	0.008935	0.000127	0.011286	0.993732	0.946244	Good Model.
Lasso Regression	0.009040	0.000128	0.011334	0.993433	0.945792	Good Model.
Support Vector Regression	0.052410	0.004011	0.063332	0.135900	-0.692631	Very Bad Model.
Bayesian Ridge Regression	0.008958	0.000128	0.011318	0.993736	0.945946	Good Model.
AdaBoost Regression	0.011575	0.000216	0.014695	0.989649	0.908873	Good Model.

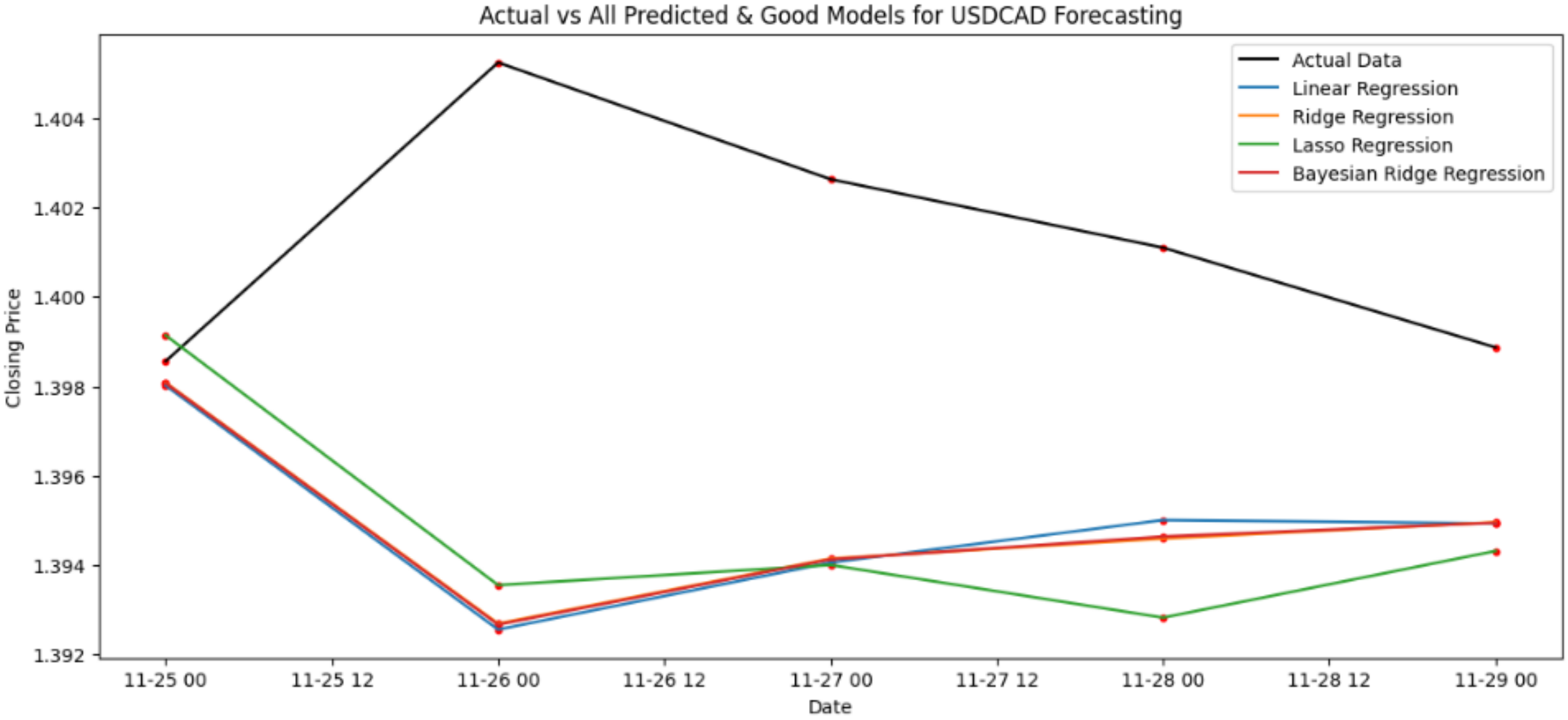
Visualization {Data from Investing.Com}

90% Train – 10% Test Splitting, with FRED's Macroeconomics data



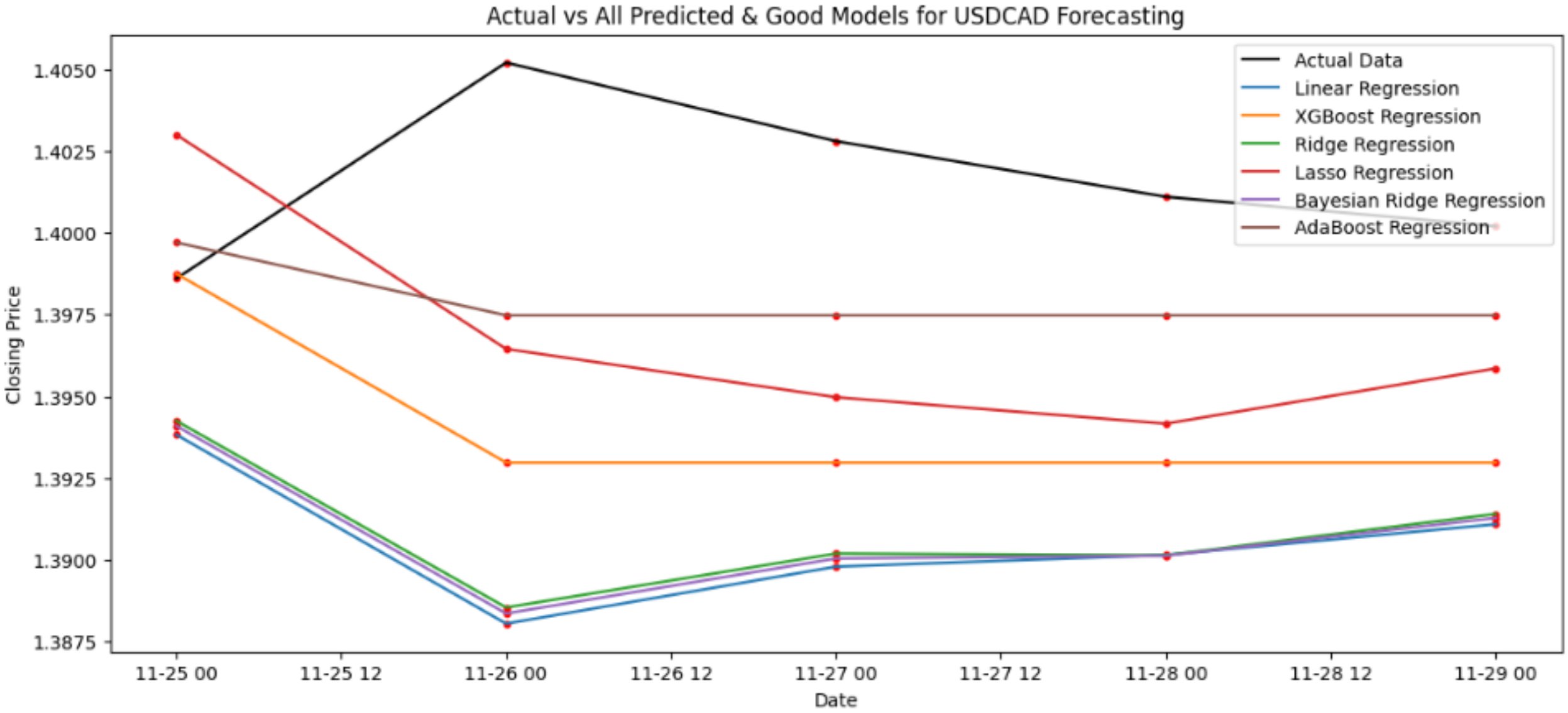
Visualization {Data from TDS}

85% Train – 15% Test Splitting, with FRED's Macroeconomics data



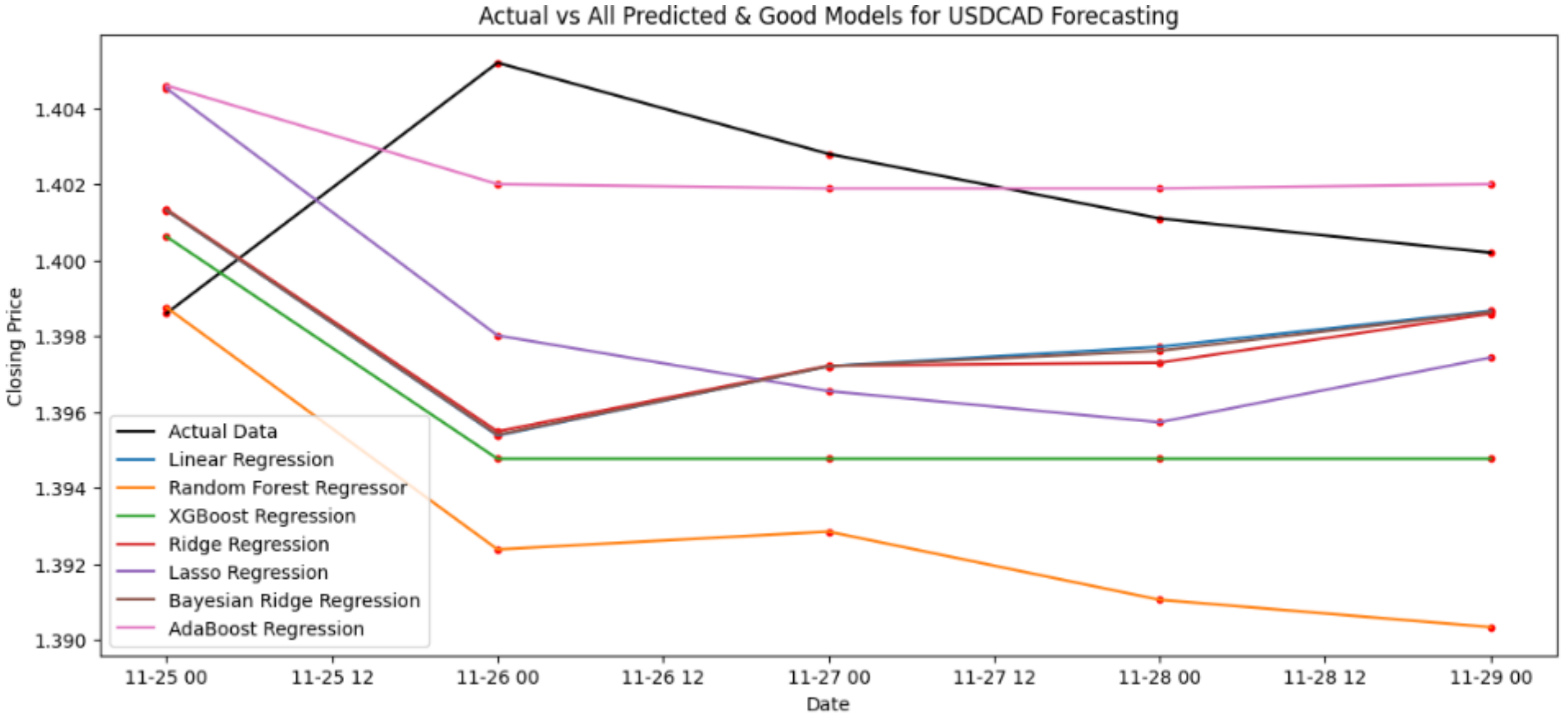
Visualization {Data from Investing.Com}

90% Train – 10% Test Splitting, **with** World Bank's Macroeconomics data

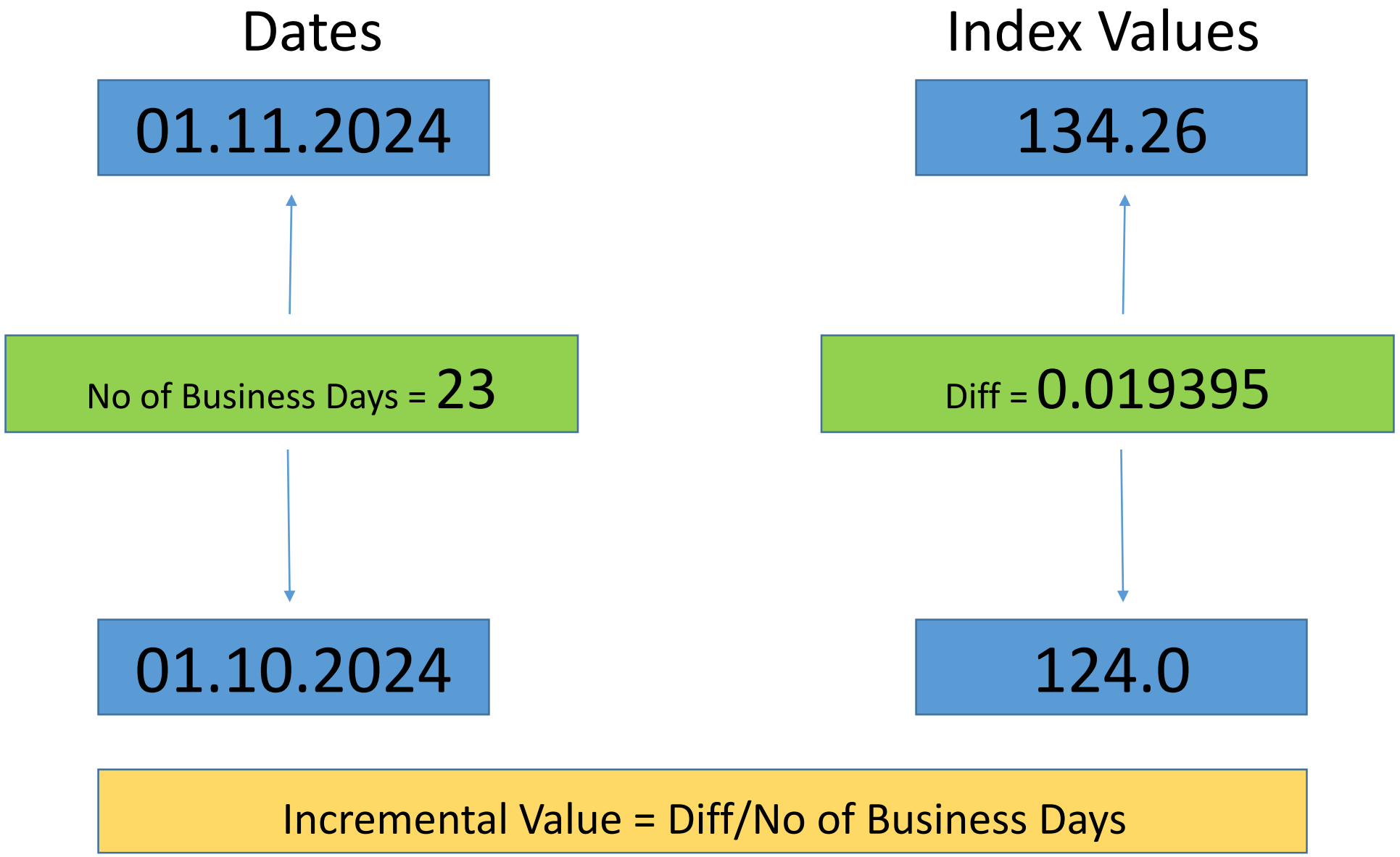


Visualization {Data from Investing.Com}

90% Train – 10% Test Splitting, **without** Macroeconomics data



Calibration approach



Visualization {Data from Investing.Com}

90% Train – 10% Test Splitting, with FRED's Calibrated Macroeconomics data values

ModelName	MAE	MSE	RMSE	R^2_Train	R^2_Test	Comment
Linear Regression	0.010821	0.000182	0.013490	0.994165	0.926100	Good Model.
Random Forest Regressor	0.011777	0.000216	0.014706	0.999775	0.912174	Good Model.
XGBoost Regression	0.009628	0.000146	0.012080	0.994868	0.940739	Good Model.
Ridge Regression	0.010766	0.000180	0.013425	0.994159	0.926814	Good Model.
Lasso Regression	0.009147	0.000132	0.011471	0.993642	0.946566	Good Model.
Support Vector Regression	0.072229	0.007031	0.083849	0.112893	-1.854950	Very Bad Model.
Bayesian Ridge Regression	0.010794	0.000181	0.013459	0.994163	0.926442	Good Model.
AdaBoost Regression	0.011812	0.000228	0.015105	0.990925	0.907353	Good Model.

Visualization {Models with Data Comparism}

Comparing the data from different sources.

Models	Metrics	Investing/FRED	TDS/FRED	Investing/WorldBank	Investing Only
Linear Regression	MAE	0.009694	0.010249	0.010944	0.008967
	RMSE	0.012171	0.012847	0.013587	0.011329
	R^2_Train	0.994074	0.988130	0.993928	0.993736
	R^2_Test	0.939845	0.908703	0.922097	0.945837
Random Forest Regressor	MAE	0.011429	0.015146	0.013020	0.009715
	RMSE	0.014234	0.018380	0.016058	0.012264
	R^2_Train	0.999713	0.999406	0.999656	0.999458
	R^2_Test	0.917730	0.813114	0.891178	0.936531
XGBoost Regression	MAE	0.009695	0.011689	0.009235	0.009235
	RMSE	0.012126	0.014317	0.011697	0.011731
	R^2_Train	0.994891	0.990652	0.994549	0.994462
	R^2_Test	0.940292	0.886613	0.942260	0.941927
Ridge Regression	MAE	0.009634	0.010196	0.010756	0.008935
	RMSE	0.012092	0.012766	0.013357	0.011286
	R^2_Train	0.994068	0.988119	0.993924	0.993732
	R^2_Test	0.940625	0.909844	0.924707	0.946244

Visualization {Models with Data Comparism}

Comparing the data from different sources.

Models	Metrics	Investing/FRED	TDS/FRED	Investing/WorldBank	Investing Only
Lasso Regression	MAE	0.009127	0.009331	0.009143	0.009040
	RMSE	0.011449	0.011580	0.011461	0.011334
	R^2_Train	0.993642	0.987471	0.993451	0.993433
	R^2_Test	0.946768	0.925820	0.944572	0.945792
Support Vector Regression	MAE	0.072223	0.149788	0.061156	0.052410
	RMSE	0.083837	0.155897	0.073537	0.063332
	R^2_Train	0.112613	0.114638	0.105346	0.135900
	R^2_Test	-1.854152	-12.444989	-1.282029	-0.692631
Bayesian Ridge Regression	MAE	0.009662	0.010204	0.010822	0.008958
	RMSE	0.012130	0.012777	0.013438	0.011318
	R^2_Train	0.994073	0.988122	0.993926	0.993736
	R^2_Test	0.940252	0.909682	0.923797	0.945946
AdaBoost Regression	MAE	0.013612	0.012066	0.012639	0.011575
	RMSE	0.017028	0.015481	0.015819	0.014695
	R^2_Train	0.990580	0.986062	0.990293	0.989649
	R^2_Test	0.882256	0.867426	0.894394	0.908873

Result

From the above analysis, on reviewing the MAE and MSE, RMSE, and R^2 values from all the models considered, it is apparently clear that "**Linear Regression**", "**Random Forest Regressor (RFR)**", "**Ridge Regression**" and "**Bayesian Ridge Regression**" displayed wonderful results. However, I will be relying more on "**Random Forest Regressor (RFR)**" due to its robustness about being exposed to many hyperparameters for tuning.

Random Forest Regressor (RFR)

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's an extension of the Random Forest algorithm for classification.

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.

Future Plans

I will also explore alternative source where I can access the Macroeconomics indices maintained at more frequent basis; possibly weekly, or even daily levels.

I plan to improve on my model by adding more Technical features.

I will explore a way of automatically accessing the Statistical data from a sources like Investing.Com or Tick Data Suite(TDS) whose data is as close as the live data.

Thank you