

PredictFX: Exchange Rate Prediction

End-to-End MLOps Pipeline (Maturity Level 2 to 3) for Monthly USD/CAD Forecasting

Team Members: 1) Harshavardhan Gadila 2) Aditya Rajpurohit

Demo site <https://harshavardhangadila.com> — plots & live inference- data.

Public repo : <https://github.com/Harshav88/Team-Singularity>

Abstract

USD/CAD reflects oil prices, rate-differentials, and cross-border trade, making it an excellent test-bed for macro-aware time-series forecasting. We assemble sixteen monthly indicators from FRED (eight U.S., eight Canadian) and daily FX closes from Yahoo Finance over 2011-01 → 2025-04. After forward-filling and magnitude scaling, we benchmark six neural architectures: Multilayer Perceptron (MLP), TabNet, plain LSTM, Temporal Convolutional Network (TCN), Seq2Seq LSTM, and an Attention-enhanced LSTM. An 80 / 20 chronological split preserves temporal integrity; the target column is zero-masked at inference to avoid leakage. Best test-set errors are **MAE 0.028 ¢ / RMSE 0.044 ¢** (MLP) and **0.030 ¢ / 0.047 ¢** (Seq2Seq). We wrap the workflow in a **Vertex AI scheduled pipeline** that (i) polls FRED each night, (ii) retrains when fresh data arrive, (iii) registers models in MLflow, and (iv) pushes the champion image to Cloud Run. A Gradio front-end served on <harshavardhangadila.com> displays actual-vs-predicted curves, attention heat-maps, and TabNet feature masks. TensorBoard logs, experiment metadata, and CI/CD logs (GitHub Actions + Cloud Build) satisfy **MLOps maturity level 2**—automated, repeatable training with manual but low-friction deployment. The project demonstrates that lightweight deep learning plus disciplined MLOps delivers accurate, explainable, and continuously-updated FX forecasts suitable for treasury risk management and macro research.

1. Introduction

Foreign-exchange volatility affects commodity exporters, importers, and retail remitters alike. While the classic random-walk often beats naïve econometric models, practitioners know that interest-rate parity and inflation gaps influence USD/CAD on monthly horizons. We posit that non-linear deep learners, supplied with curated macro features, outperform both statistical baselines and price-only LSTMs. Our deliverable is a **production-grade MLOps pipeline**—not merely a notebook—showing reproducible training, automated retraining, continuous evaluation, and a public demo site.

2. Related

- **Econometric roots** – Meese & Rogoff (1983) established the random-walk dominance; subsequent cointegration models add CPI and rate spread.
- **Deep learning on FX ticks** – Bao et al. (2017) used CNN-LSTM on high-frequency EUR/USD; Zhang & Yan (2020) introduced transformer-based sequence tagging.
- **Macro-aware nets** – Few studies fuse low-frequency fundamentals with deep nets. Kwon (2023) applied TabNet to KRW/USD but without MLOps rigor. Our work differs by (i) integrating six architectures in a uniform pipeline, (ii) publishing all artifacts, and (iii) delivering Vertex-AI automated training.

3. Data

Category	U.S. series (FRED ID)	Canada series (FRED ID)
Policy rates	FEDFUNDS	IRSTCB01CAM156N
Inflation	CPIAUCSL	CANCPIALLMINMEI

Category	U.S. series (FRED ID)	Canada series (FRED ID)
Unemployment	UNRATE	LRUNTTTTTCAM156S
10-yr yield	DGS10	IRLTLT01CAM156N
Exports (\$ B)	EXPGS	XTEXVA01CAQ667S
Imports (\$ B)	IMPGS	XTIMVA01CAQ667S
House-price index	CSUSHPINSA	QCAR628BIS
Retail spend (\$ B)	PCEC	NCPHISAXDCCAQ

Forex data: Yahoo CAD=X daily close → monthly last-value.
 Pre-processing: forward-fill, convert raw \$ to billions, MinMax scaling.
 Split: 2011-02 → 2022-04 train, 2022-05 → 2025-04 test (172 rows total).

4. Methods

4.1 Feature engineering

1. **Static matrix** for TabNet/MLP: one row = one month.
2. **Sliding windows** (length 6) for sequence models, target = t + 1.
3. **Leakage guard**: drop/zero FX column during test inference.

4.2 Model

Model	Config	Motivation
TabNet	3 decision steps, N_d = 8	interpretable masks for tabular data
MLP	128 → 64 → 32 dense, ReLU	fast, strong baseline
Plain LSTM	64 units, return_seq False	compare to price-only nets
TCN	4 blocks, kernel 3, dilation {1,2,4,8} long receptive field, parallel	
Seq2Seq LSTM	Enc 128, Dec 64, RepeatVector 3	multi-step forecast
Attn-LSTM	LSTM 64 + additive attention	accuracy + explainability

All optimise **MSE**, Adam, EarlyStopping(patience 10) prevents over-fit. Hyper-parameter choices, loss curves.

5 Experiments

5.1 Metrics

Model	MAE (¢) RMSE (¢) Params		
MLP	0.028	0.044	12 k
Seq2Seq LSTM	0.030	0.047	96 k
Attn-LSTM	0.061	0.066	38 k
TCN	0.072	0.083	44 k
LSTM	0.089	0.102	26 k

Model	MAE (¢)	RMSE (¢)	Params
TabNet	0.140	0.155	21 k

5.2 Visualisation

TabNet – TabNet model: USD CAD Prediction vs Actual

Seq2Seq_LSTM – Seq2Seq LSTM: USD CAD Prediction vs Actual

MLP.png – Multilayer Perceptron: USD CAD Prediction vs Actual

Attention_LSTM – Attention-Enhanced LSTM: USD CAD Prediction vs Actual

Macro Economic Indicators - plots for various indicators USA vs CANADA

5.3 Ablation study

Variant	MAE	Observation
Window 3 (Seq2Seq)	+0.012	shorter context hurts
Drop CPI features	+0.018	inflation gap critical
Drop export/import	+0.004	minor impact

5.4 Comparison with published baselines

Random-walk MAE on same test window = 0.089 ; our best model beats it by **68 %**.

6 MLOps Implementation

6.1 Pipeline architecture

FRED/Yahoo API Data → Vertex AI Jupyter Notebook (Cloud Run) -> DropBox to store plots -> Frontend React website to display(<https://harshavardhangadila.com/>)

{ Vertex AI Pipeline – Scheduler } to Run on 1st of every month. The above flow continuous i.e., Automated training and predictions.

6.2 MLOps maturity level in between 2 to 3

- **Automated, reproducible training**
- **Centralised experiment tracking** (MLflow)
- **Manual but one-click deployment** (Cloud Run revision)

7. Front-End & UX

https (SSL) secured website to avoid malicious/wrong plots/Graphs injection (<https://harshavardhangadila.com/>) developed using ReactJS and deployed on cloud.

8. Conclusion

Deep learning, when carefully engineered macro inputs, can beat naïve and econometric FX baselines. A simple MLP suffices for near-term accuracy, but attention and Seq2Seq nets offer interpretability and multi-step power. The Vertex-AI pipeline ensures models retrain automatically as new macro prints arrive, maintaining relevance without manual ops effort. Future work targets Informer and Temporal Fusion Transformer for even longer contexts, real-time data ingestion (daily), and drift-triggered auto- rollbacks (MLOps Level 3 to 4).