Literature Review.

Classification and Predictive Analytics

on the Ontario interest rates.

Nara Galliamova

501143606

Ashok Bhowmick, Ceni Babaoglu, Sedef Akinli Kocak,

Uzair Ahmad, Zekiye Erdem

February 23, 2026

Table of Contents

Literature Review

The research on interest rate forecasts traditionally is looking at regression-based approaches, predicting continuous rate levels ((Stock, J. H., & Watson, M. W., 2002). Nevertheless, in financial and macroeconomic modeling, directional forecasting is gaining more attention by predicting if the rates of interest will increase, decrease or remain stable.

By reviewing different studies in recent years, XGBoost has become a scalable tree-boosting system that can work on nonlinear relationships and complex feature interaction. XGBoost demonstrates strong predictive ability when most applications are considering financial market returns rather than fiscal tax interest rates of the provincial governments. Furthermore, not many studies have strict time-series validation (Chen, T., & Guestrin, C., 2016).

Also, in a 2024 study compared various machine learning models to predict interest rate and discovered that three-based methods are performing significantly better than nonlinear environments. It is important to mention that this study is focused on continuous forecasting, rather than directional classification (Salem, A. A. M., & Albourawi, A. J. A., 2024).

Moreover, most studies use regressions instead of classification, do not consider fiscal tax interest dynamic, and often use cross-validation, which is not suitable for times-series data. Many studies also focus mainly on national interest rates and, therefore, there is a gap in provincial-level directional forecasting. Applied machine learning spotlights walk-forward or rolling-origin validation for sequential data. Time-series modeling needs validation strategies to avoid the artificial inflation of accuracy. Standard k-fold cross-validation violates temporal ordering and may cause bias. Walk-forward validation preserves chronological structure and reduces optimistic bias (Makika, H., Romano, J. M. T., & Ballini, R., 2023).

This project will look into two scenarios, comparing the effects of historical-rate only models with historical and macroeconomic models. Classification problems of fiscal interest rate will be addressed, including walk-forward evaluation and assessing economic signal contributions. Even though current studies contribute to the feasibility of machine learning, there is still a need for more research on time-based provincial focus, directional classification, and incremental feature testing. Time-series forecasting requires rolling-origin evaluation to preserve temporal structure (Hyndman, R. J., & Athanasopoulos, G., 2021).

Data Description

The main dataset is taken from Ontario Data Catalogue, containing quarterly tax interest rates that are set by the Ministry of Finance (Ontario Ministry of Finance, 2026). The research is focusing on underpayment interest rates. Time span is 2000Q1-2026Q1, which is approximately 105 quarters. Frequency is quarterly, unit of analysis is one calendar quarter. The dataset is a public government resource that is available to the public. Core variables are year, quarter (Q1-Q4), underpayment_rate (numeric, percentage), and overpayment_rate. The predictive decision is supporting fiscal planning by sending signals to the Ministry about the rate adjustments.

Target variable defined as: $\Delta r_t=r_t-r_{t-1}$, where Δr_t is a change in interest rate with a previous quarter. r_t = Underpayment Interest Rate in quarter t. The objective is to predict direction t, using only available information at a time t-1.

The classification label is defined as:

Increase if $\Delta r_t > 0$

Decrease if $\Delta r_t < 0$

Stable if $\Delta r_t = 0$

Class labels: Increase, Decrease, Stable.

External Indicators Collected from Statistics Canada, Bank of Canada. Macroeconomic indicators are combined using year-quarter keys. The indicators are lagged by one quarter to avert information leakage.

Macroeconomic variables are CPI (lagged 1 quarter), GDP growth (lagged 1 quarter), unemployment rate (lagged 1 quarter), policy rate (lagged 1 quarter). Those are commonly used in financial forecasting models (Stock, J. H., & Watson, M. W., 2002). All variables are converted to quarterly frequency. Descriptive statistics include frequency distribution of target class, maximum, minimum, mean, median, and standard deviation. Those statistics are ensuring that the dataset is suitable for time-series classification modeling.

Preprocessing Steps: sorting chronologically, removing missing values, creating delta_rate and classification labels. Generating historic lag features (lag1_rate, lag2_rate), creating rolling statistics (4-quarter rolling mean and standard deviation) and lagged macroeconomic indicators by one quarter (lagged policy rate, lagged CPI inflation, lagged unemployment rate, lagged GDP indicators). Lastly, removing missing observations from lag creation. Outliers should be assessed using boxplots and interquartile range criteria, correlation matrices can be generated to assess multicollinearity.

Constraints are limited quarterly sample size, about a hundred of observations. Class imbalance possibility, and structural breaks, such as the 2008 crisis, COVID-19.

Project Approach

Using quarterly data from the Ontario Ministry of Finance from 2000Q1 to 2026Q1, the research is formulating a three-class classification scenario, based on the first difference of the rate. The study is evaluating predictive feasibility using walk-forward validation and Macro-F1 score. Statistics Canada and Bank of Canada lagged macroeconomic indicators, along with historic rates are incorporated in the project.

The study is following a clear plan for a time-series classification plan to predict the quarterly direction of the underpayment tax interest rate.

First of all, the target variable is defined as a tree-class categorical outcome, based on the sign of the difference of underpayment rate. Second of all, lag features engineering, combining lagged macroeconomic variables and then applying walk-forward validation. The process includes training the model on first k quarters, predicting next quarter. Then expanding the training window and repeating the process.

Traditional random test-train splits are not suitable for time-series data. Afterwards, it is possible to see the difference between the baseline and machine learning models. Lastly, evaluate Macro-F1 and interpret the feature importance. Averages of Macro-F1 scores are great for imbalanced classification problems (Sokolova, M., & Lapalme, G., 2009)

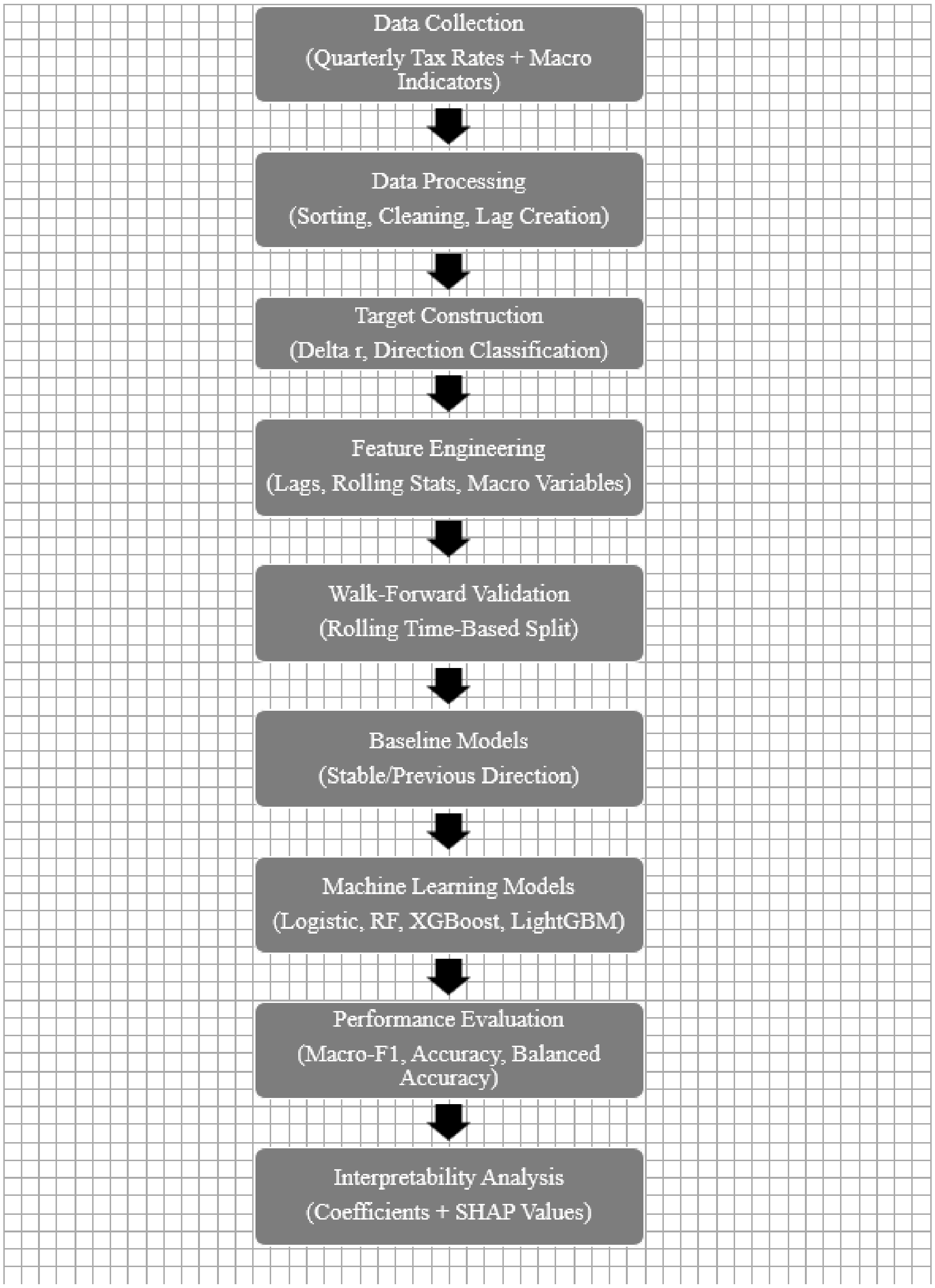
Baseline models always predict stable direction, as well as previous quarter direction. Machine learning models have to outperform those two baselines. Machine learning models are performed using logistic regression with Random Forest, XGBoost, and LightGBM. (Chen, T., & Guestrin, C., 2016).

Baseline 1: y^t=yt-1. Baseline 2: y^t=Stable.

Walk-forward validation is done with the primary metrics of Macro-F1. Secondary metrics include accuracy, balanced accuracy, confusion matrix, precision and recall per class. Focusing on time periods of initial time window 2000Q1-2015Q4 and then a rolling test window 2026Q1-2026Q1, using expanding window strategy. Feature importance analysis is done using tree-based importance metrics, as well as SHAP, Shapley Additive Explanations. Macro-F1 is selected due to potential class imbalance.

Time-series classification needs chronological validation. Feature interpretability is important in public policy forecasts so therefore, walk-forward validation is eliminating positive bias.

Visual Representation



GitHub Repository Link

https://github.com/ggalliam/Classification-and-Predictive-Analytics

References

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system.* Association for Computing Machinery. https://dl.acm.org/doi/10.1145/2939672.2939785

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice.* OTexts. https://otexts.com/fpp3/

Makika, H., Romano, J. M. T., & Ballini, R. (2023). *Deep Learning Techniques For Time Series Forecasting: An Application For Exchange Rates.*

Ontario Ministry of Finance. (2026). *Tax interest rates*. Ontario Data Catalogue. https://data.ontario.ca/dataset/tax-interest-rates

Salem, A. A. M., & Albourawi, A. J. A. (2024). *Predictive models for interest rate forecasting using machine learning: A comparative analysis and practical application.* Brilliance. https://doi.org/10.47709/brilliance.v4i2.4983

Sokolova, M., & Lapalme, G. (2009). *A systematic analysis of performance measures for classification tasks*. ScienceDirect. https://doi.org/10.1016/j.ipm.2009.03.002

Stock, J. H., & Watson, M. W. (2002). *Macroeconomic forecasting using diffusion indexes.* American Statistical Association Journal of Business & Economic Statistics. https://www.tandfonline.com/doi/abs/10.1198/073500102317351921