

FORECASTING SUBSCRIPTION CHURN USING ECONOMIC SENTIMENT IN MATLAB

1. Objective

This project aimed to develop a forecasting model to predict monthly churn among subscribers of a mid-sized SaaS company. The unique approach was to integrate external economic sentiment indicators with internal usage and engagement data, leveraging MATLAB's econometrics and time series capabilities.

2. Client and Use Case

Client:

A B2B SaaS company offering project management software in the U.S. and UK.

Use Case:

The client wanted to understand how broader economic conditions (like recession fears) influence customer churn, especially for monthly contracts, and wanted to include these factors in its revenue planning and customer retention strategies.

3. Dataset Overview

Time Period:

January 2020 – December 2023 (48 months)

Unit of Analysis:

Monthly churn rate (percentage of users unsubscribed out of the active base)

Variables Used:

- **Dependent Variable:**
Monthly churn rate (%)
- **Internal Predictors:**
 - Average session time per user
 - Support tickets per user
 - Product usage frequency index
- **External Predictors (from economic sentiment databases):**
 - University of Michigan Consumer Sentiment Index (U.S.)

- GfK Consumer Confidence Index (UK)
- Google Trends: "Cancel Subscription", "Software Alternatives"

Data Sources:

- Company's internal analytics (Redshift export)
- Economic indicators via FRED and GfK APIs
- Google Trends API via PyMatBridge for import into MATLAB

4. Methodology

4.1 Data Preparation

- All monthly data was synchronized and normalized.
- Log-differenced all continuous predictors to ensure stationarity.
- MATLAB's fillmissing and normalize functions were used to handle missing or skewed data.

4.2 Forecasting Model in MATLAB

Model Type:

Multiple Time Series Regression with ARIMA residual correction

$$\text{Churn}_t = \beta_0 + \sum \beta_k X_{kt} + \varepsilon_t, \quad \varepsilon_t \sim \text{ARIMA}(1,0,1)$$

- Developed using regARIMA() in MATLAB
- Stepwise feature selection performed with stepwiselm()
- Evaluated forecasting accuracy using rolling-origin validation (12 steps ahead)
- Compared against baseline naive and AR-only models

5. Key Findings

- **Consumer Sentiment Index** (U.S.) was a **significant predictor** ($p < 0.01$); lower sentiment scores predicted higher churn.
- **Google Trends search volume** for subscription-related terms showed a **1-month lead** effect on churn spikes.
- **Support tickets per user** was the strongest internal predictor of churn.

Forecast Accuracy:

- RMSE = 1.84%, MAPE = 5.7%
- Improved forecasting accuracy by **22% over baseline ARIMA** without sentiment variables.

6. Visualizations and Interpretation

- **Time Series Plots:** Overlaid churn with sentiment index
- **Forecast Envelopes:** 80% and 95% confidence bands
- **Contribution Charts:** Variance explained by each predictor
- **Scenario Simulation:** Allowed client to input hypothetical economic sentiment scores and forecast churn rates

All visualizations were built using plot, area, bar, and forecast plotting utilities in MATLAB.

7. Deliverables

- 22-page forecasting report with explanation of model structure, assumptions, and recommendations
- Dynamic .mlapp tool to test sentiment-based churn scenarios
- Monthly forecasting script with user-defined input capability
- Training session conducted with client's in-house analyst team

8. Strategic Value Delivered

- Enabled the finance team to integrate churn forecasts into MRR/ARR projections
- Provided early-warning signals tied to external sentiment for customer success teams
- Influenced the creation of a marketing “win-back” campaign timed with low sentiment periods

9. Tools and Technology Stack

- MATLAB R2023a
- Econometrics Toolbox

- Signal Processing Toolbox
- Custom API connectors for FRED and Google Trends
- PyMatBridge (to import Google Trends into MATLAB)

Statssy