

INTEGRATING SOCIAL MEDIA AND ADMINISTRATIVE DATA FOR THE REAL-TIME PREDICTION OF THE CONSUMER CONFIDENCE INDICATOR

Akvilė Vitkauskaitė^{1,2}, Andrius Čiginas^{1,2}

¹State Data Agency (Statistics Lithuania), Lithuania

²Vilnius University, Lithuania



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Consumer Confidence Indicator

The Consumer Confidence Index (CCI) measures public sentiment about the economy through a survey of four questions regarding household finances, economic outlook, and spending plans over the past and coming year.





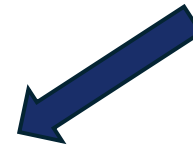
EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Key Objectives

- **Timely CCI Estimates:** Estimate the current month's CCI faster than the traditional survey method.
- **Integrating Data Sources:** Provide early estimates of CCI by combining key economic indicators with social media sentiment and Google Trends data.



Data Sources



STATISTICS
LITHUANIA
STATE DATA
AGENCY

Administrative Data

- Inflation rate
- Income statistics
- Unemployment rate



Social Media Data (X)

- Tweets collected weekly via official API
- Keywords related to economic sentiment
- Sentiment analysis using NLP tools (TextBlob, Vader, Afinn, Transformers, Flair)



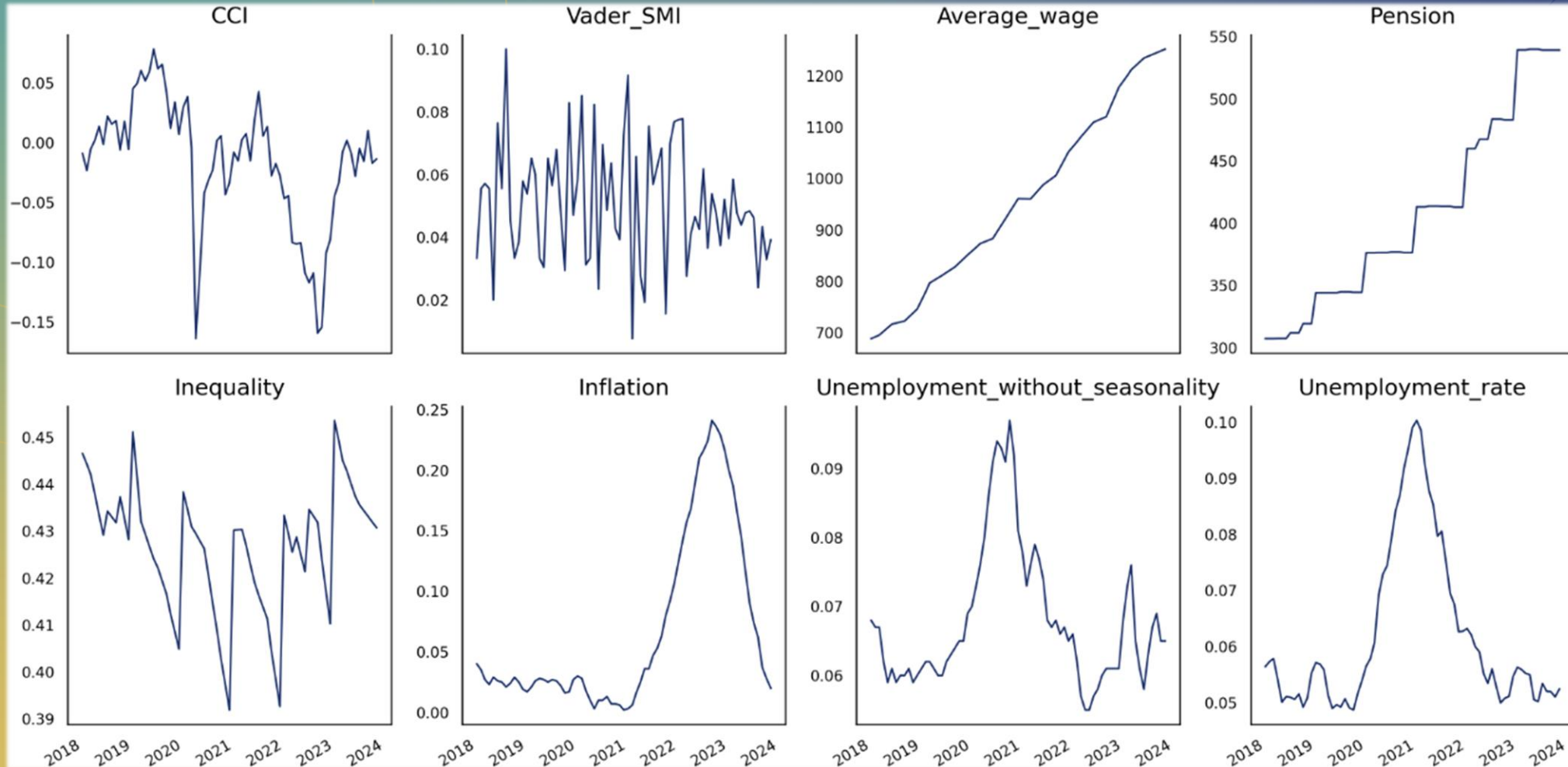
Google Trends Data

- Manual collection and compilation into a monthly dataset
- Keywords: economics, **purchases, work costs, buy, unemployed, inflation**

CCI, sentiment indices and key economic indicators



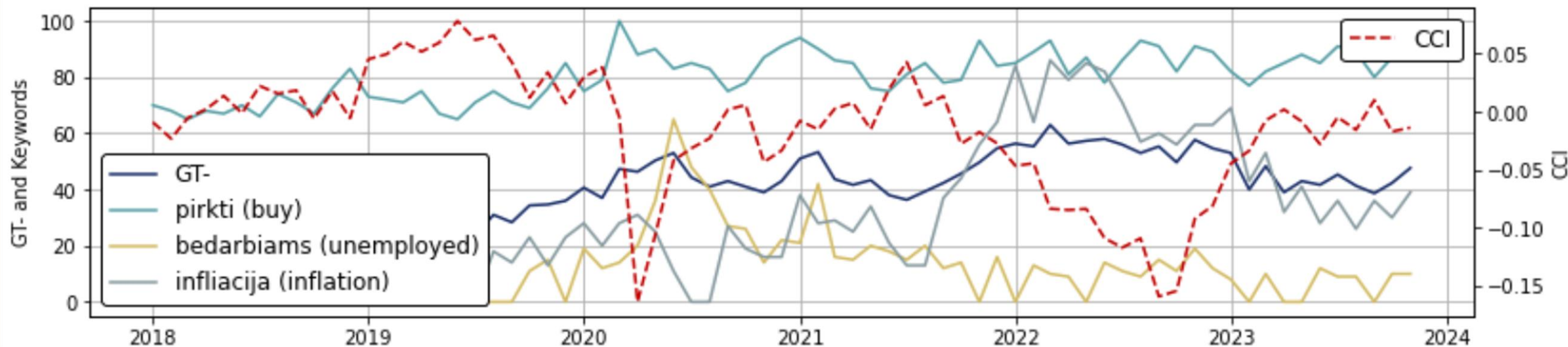
EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL



GT+, GT- and CCI



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL





Metodology

Predictive Models

Statistical Models

- SARIMAX
- VECM

Machine Learning Models

- Random Forest
- XGBoost

Model Selection and Optimization

Statistical Models

Exogenous variables were selected based on Johansen's cointegration test and Spearman's rank correlation.

Machine Learning Models

- Selection is based on feature importance
- Incorporation of rolling averages

Rolling Forecast Method

Predicts the next step and updates the model with the latest information.

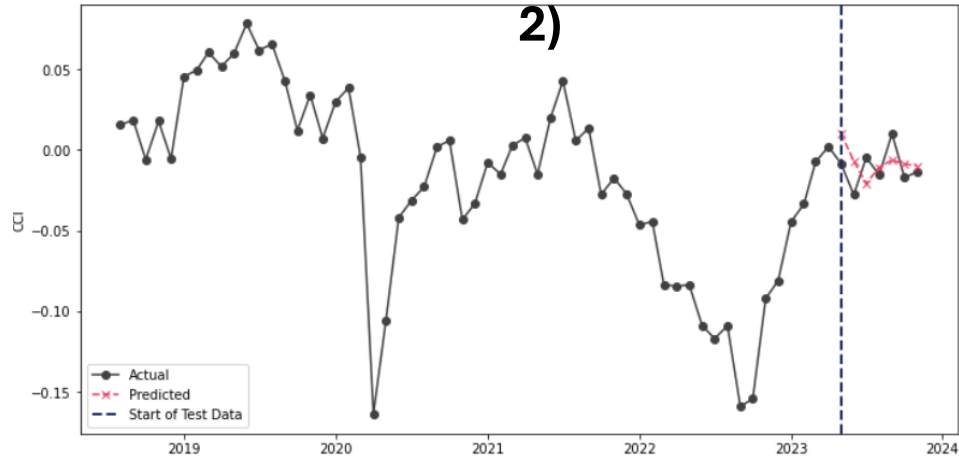
Evaluation Metrics

MAE, MSE, RMSE (*for all*)
AIC (*only for SARIMAX*).

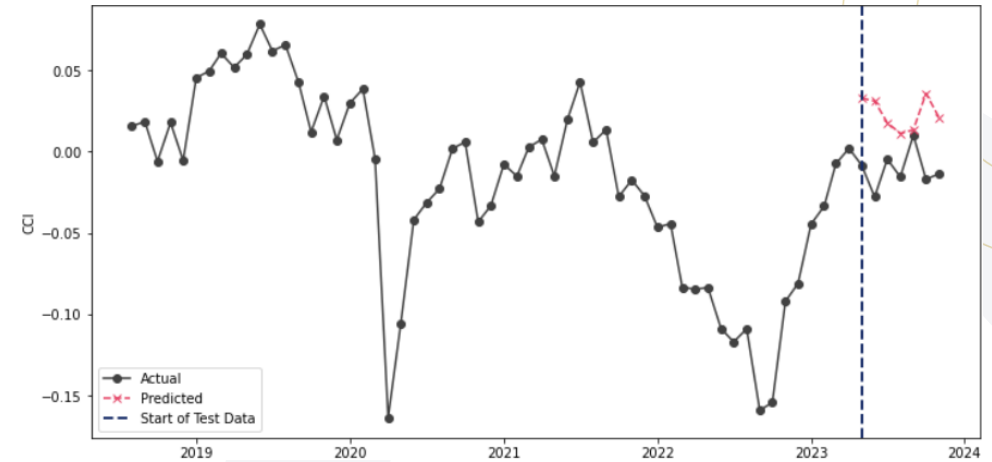


Final models for CCI forecasting

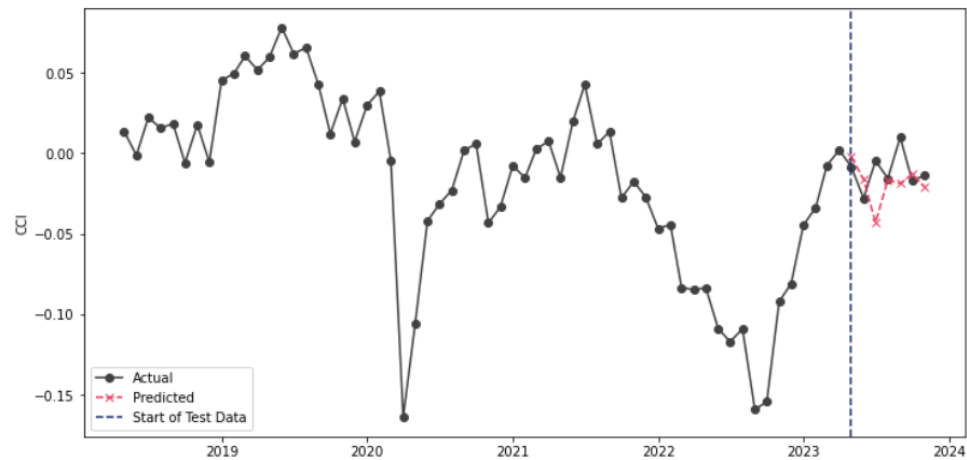
SARIMAX(5,1,6,1,0,1,1 2)



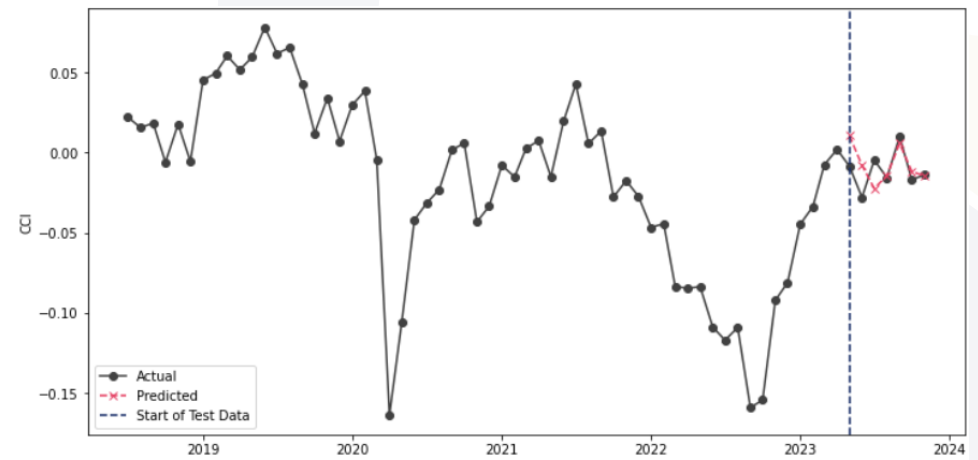
VECM



Random Forest



XGBoost





Results

Model name	Exogenous variables (features)	MAE
SARIMAX	Vader_SMI_diff_1, CCI_diff_1, GT-_diff_1, Inflation_diff_1 (3 months lag), Unemployment_without_seasonality_diff_1 (current, 6 months lag)	0.0127
VECM	Vader_SMI_diff_1, Inflation_diff_1, GT-_diff_1 (all up to 1 month lag)	0.0343
Random Forest	CCI (1 month lag; 2 periods), GT (current and 1 month lag; 2 periods), GT- (current and 1 month lag; 4 periods)	0.0139
XGBoost	GT- (current and 1 month lag; 2 periods), GT+ (1 month lag; 6 periods), Average wage (current; 4 periods), Inflation (3 months lag; 4 periods), Unemployment_without_seasonality (5 months lag; 6 periods)	0.0098



Literature

- **Enhancing CCI Forecasting with SMI from Facebook and X Data Integration**

Van den Brakel, J. A., Söhler, E., Daas, P. J. H., & Buelens, B. (2017). Social media as a data source for official statistics; the Dutch Consumer Confidence Index. *Survey Methodology*, 43(2), 183-210.

- **SMI Indicator from X Data**

Bruno, M., Catanese, E., Iannaccone, R., Righi, A., Scannapieco, M., Testa, P., Valentino, L., Zardetto, D., & Zurlo, D. (2024, May 2). Social mood on economy index: A daily measure of the Italian sentiment on the economy based on X (ex Twitter) data. *Istat*.

- **Enhancing Sri Lanka's CCI Forecasting with Google Trends (Motorcycles and Birth Control) Data**

Zhang, H. (2017). Incorporating Google Trends data in predicting consumer confidence in Sri Lanka. Available at SSRN 3085010.



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

**Thank you for your
attention!**

**I welcome any questions
you may have.**



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat 

The conference is partly
financed by the European Union