

Evaluation of Consumer Confidence Indicators Using Social Media and Administrative Data



STATISTICS LITHUANIA
STATE DATA AGENCY



Vilnius
University

Akvilė Vitkauskaitė, Andrius Čiginas

State Data Agency (Statistics Lithuania); Vilnius University

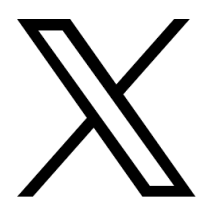
Akvile.Vitkauskaite@stat.gov.lt, Andrius.Ciginas@mif.vu.lt

Introduction

Consumer confidence indicator (CCI) is an important economic measure estimated using a sample survey. The administration of the CCI lies under the purview of Eurostat, the statistical office of the European Union. The State Data Agency (Statistics Lithuania) is responsible for conducting the corresponding survey in Lithuania every month. Social media platforms and administrative registers are alternative data sources that can improve the estimation. This study examines the relationships between traditional survey-based indicators and consumer sentiment expressed on social media platforms. It also explores the potential of using administrative data as auxiliary variables to enhance the forecasting accuracy of CCI.

Social Media Data Source

- The Social Media Index (SMI) is constructed using Twitter (X) data applying sentiment analysis to measure monthly public sentiment.



- SMI provides real-time insights into consumer behavior.
- Data from platform X is legally and easily accessible, but it has limited popularity in Lithuania.

Administrative Data

Objective economic indicators for enhanced forecasting:

- Inflation rate
- Income statistics
- Unemployment rate

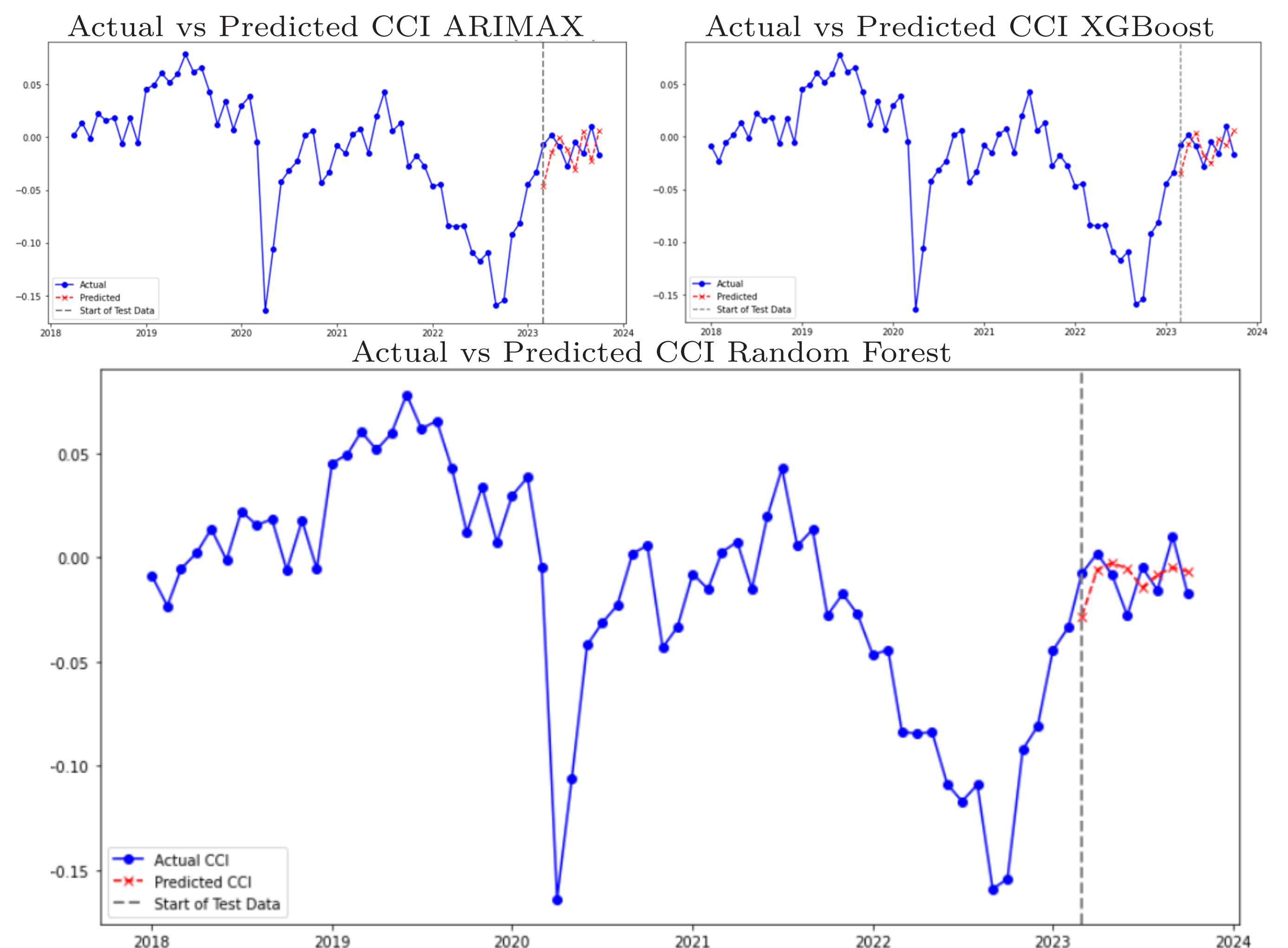
Forecasting Accuracy

Model	MAE	MSE	RMSE
ARIMAX	0.0226	0.0006	0.0244
XGBoost	0.0167	0.0003	0.0177
RF	0.0124	0.0002	0.0137

Methodology and Analysis

We explored different models to forecast CCI, integrating a variety of auxiliary variables. Here, we show the results of 3 models: ARIMAX, XGBoost, and Random Forest. All models used a rolling forecast approach, utilizing the most recent data to make predictions.

- The ARIMAX model incorporated differenced SMI, 'Average wage', 'Pension', 'Inequality indicator', 'Inflation rate' and 'Unemployment rate' series, where the first three series are lagged.
- XGBoost predicted effectively using 'Inflation rate', lagged SMI, 'Average wage', 'Pension', 'Unemployment rate', and CCI.
- Meanwhile, The Random Forest approach utilized SMI, and historical CCI.



Summary

Overall, our research suggests that combining survey data with alternative sources like SMI and key economic indicators can improve CCI nowcasting accuracy. The Random Forest model outperforms other models based on forecasting accuracy metrics.