Nowcasting Socio-Economic Trends: An LSTM-based Approach Leveraging Google Trends, GDELT, and Eurostat Data

Cristina Jiménez, Ayah Dahmani, Ricardo González Otal and Juan Miguel López Piñero

1. ABSTRACT

This project aims to develop nowcasting models using advanced machine learning techniques, such as random forests, extreme gradient boosting, stacked ensembles, and neural networks. These models will be employed to predict, in real-time, a set of socio-economic variables in relation to labor market integration, providing valuable insights for decision-making processes. Moreover, this project will serve as a tool for new graduates from the Data Science program 2023-2024 of The Bridge School, allowing them to apply and showcase their skills in machine learning and data analysis.

2. INTRODUCTION

In today's rapidly evolving world, the ability to predict and understand socio-economic trends in real-time has become increasingly valuable. This project, created by the company Rambee and assigned to a team of four graduates from The Bridge School, aims to develop a nowcasting model based on machine learning techniques to forecast a specific socio-economic variable using big data obtained from Google Trends, the Global Database on Events, Language, and Tone (GDELT), and Eurostat.

Nowcasting,("now" and "forecasting"), refers to the process of predicting the present or the very near future by analyzing current and historical data. This approach is particularly useful in situations where traditional forecasting methods may not be as effective due to the lack of timely data or the need for more immediate insights.

The main goal of this project is to develop and validate an LSTM neural network nowcasting model for sequential data analysis. The model will utilize three key data sources: Google Trends for search volume trends across countries, GDELT for sentiment indicators and topic popularity from news articles, and Eurostat for official socio-economic statistics from the European Union.

By combining these diverse data sources, the model aims to provide insights into consumer behavior, sentiment trends, and socio-economic indicators for forecasting purposes.

3. AIMS AND OBJECTIVES

The primary tasks involved in this project are as follows:

1. Extract information on search volume queries from Google Trends for a predefined set of categories, with weekly frequency, across different countries.
2. Retrieve sentiment indicators and topic popularity rates from GDELT for a series of relevant socio-economic themes, in the form of "Article Tone" and "Topic Popularity Rate" data.
3. Obtain relevant socio-economic data from Eurostat for the variable of interest.
4. Construct and validate an LSTM-based nowcasting model for a specific socio-economic variable of interest, using the extracted data from Google Trends, GDELT, and Eurostat as predictors.

The final nowcasting model will enable a deeper understanding of current trends and facilitate proactive decision-making processes.

4. PYTRENDS API