UK Statistics Authority

Machine Learning Project

Face-to-Face Sprint

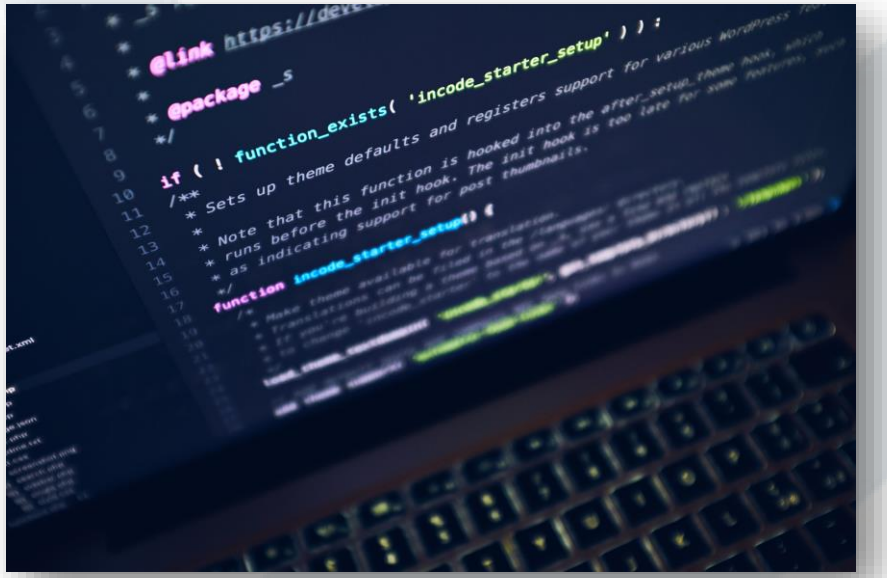
13-15 May 2019

Eric Deeben and Oliver Mahoney

The Sprint

The Machine Learning (ML) Project was proposed and approved during the HLG-MOS November 2018 workshop.

Based on a paper by the Blue Skies Thinking Network, the project aims to inform policymakers about the possibilities to use ML in the production of official statistics and to demystify official statisticians unfamiliar with it.



Work Package Outputs

Practical Tools

- Created or adapted from existing projects.
- Newly created and open-source tools.

Papers

- Technical guidance aligned to the released tools
- User guidelines for assessing quality of ML applications.
- ‘Lessons Learned’ by NSO’s from their ML applications.

WP1 Pilot: Classification and Coding

What we discussed:

- Coding activities are prime candidates for ML application.
- The ‘Web & Sentiment’ pilot theme having the same technique as this pilot. Therefore they have been merged.

What we have planned:

Methods and code from a U.S. Bureau Labour Statistics project will be applied to two proposals from Serbia and Poland.

A proposal from Belgium will also explore Twitter data separately.

WP1 Pilot: Editing and Imputation

What we discussed:

- A shared demand for an automated editing process.
- Limited demonstration of ML adding value to editing currently.

What we have planned:

Apply ML techniques to the ONS Living Cost and Food (LCF) survey to automate the Editing and Imputation process on income data and other data blocks. This can then be used by other NSOs to solve their E&I requirements.

WP1 Pilot: Imagery

What we discussed:

- It’s already proven that ML is needed to effectively exploit imagery data.
- A shared potential area of interest may be the use of satellite data for population measurement.

What we have planned:

Create a tool and methodology that can use satellite data to map urban and rural areas in Mexico. Assess the quality of the tool by comparing it with original field data.

WP2 Quality

Identify quality and performance indicators in two contexts:

- ML applied to carry out traditional processes and data.
- ML applied on non-traditional data sources.

Consider quality features (definitions, dimensions, indicators) from the official statistics and ML communities. Apply some of these indications in WP1 pilot studies.

Making sure to remain focused on ML issues rather than broader issues from the various data sources.

WP3 Lessons Learned

Intention to combine the experiences of organisations who have implemented or are close to implementing ML techniques; with the experiences of organisations who will make advancements through the WP1 pilot studies.

These would form into ‘lessons learned’ on topics such as: facilitators, obstacles, importance and costs of creating and maintaining learning datasets, role of manual operations, etc.

Going Forward

This month the pilot groups agreed on the deliverables for the November workshop and align them to the initial set UNECE HLG-MOS objectives.

A delivery time line will be produced for each pilot, ending in October when leads finalise the presentation of the practical tools and papers.

Q & A

Or contact us .....

UNECE Machine Learning project team