

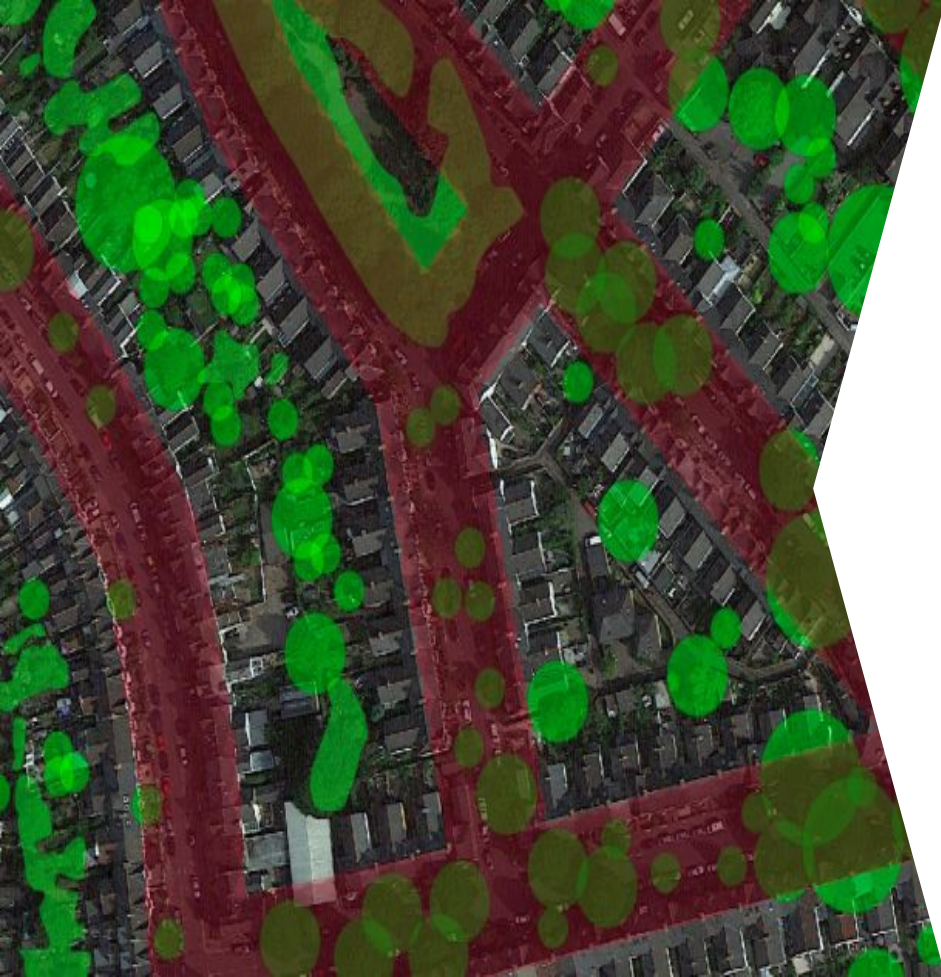


ONS-UNECE ML 2021 Group

Oliver Mahoney,
Data Science Campus
UK Office for National Statistics

ML2021@ons.gov.uk

30 June 2021



Why?

“Any industry with very large amounts of data — so much that humans can’t possibly analyz[s]e or understand it on their own — can utilize AI”

[Gartner 2017](#)

© Satellite imagery copyright Google

Objectives

Platform to facilitate

- Research to modernise official statistics
- Building capacity in machine learning
- Sharing knowledge (data, methods, etc.)

Community driven

- Projects and activities are led by members
- Supported by Group membership
- Decision informed by consensus (i.e. polls)

Public Good

- Open to all official statistical bodies
- Accessible to various levels of expertise.
- Resources available to the wider community.

ML 2021 Group

Work Stream 1

From idea to valid solution
(demonstration)

Activity 1.1

(Brazil)
Classification and aggregation for web-scraped price data

Activity 1.5

(Poland)
Multiple imputation

Activity 1.9

(USA-BLS)
Expenditure estimate

Activity 1.2

(Turkey)
Classification of COICOP

Activity 1.6

(Malaysia)
Estimating rubber production using satellite imagery

Activity 1.10

(Belgium-Flanders)
Transferring knowledge

Activity 1.3

(Denmark)
Multilingual Transformer-based model for classification

Activity 1.8

(Indonesia)
Use of satellite imagery for wealth index

Activity 1.11

(Chile)
Route optimisation

Activity 1.4

(Chile)
Automated coding as shared service

Imagery Theme

C&C Theme

Work Stream 2

From valid solution to production
(operationalization)

Activity 2.1

(IMF)
Automated production tool

Activity 2.2

(Mexico)
Data lake architecture

Activity 2.3

(Mexico)
Workflow

Activity 2.4

(Sweden)
Technical platform for support ML

Work Stream 3

Data ethics and governance

Activity 3.1

(UK)
Data ethics and governance

Work Stream 4

Quality of training data

Activity 4.1

(Finland)
Quality of training data

Work Stream 5

Quality Framework for Statistical Algorithm

Activity 5.1

(Mexico)
Quality Framework for Statistical Algorithm

Important now more than ever

- Automation of ingrained labour intensive and time-consuming processes
- The pandemic has changed what 'fit-for-purpose' means for official statistics
- "... producing historic data, but not relevant data..."

Beyond the method

- There are no ML 'quick wins' in official statistics... Other considerations come into play:
 - Getting past proof of concept
 - Measuring and maintaining quality
 - Explainability and interoperability
 - Considering other governance issues
 - Technology and infrastructure
- Facilitating the transfer of knowledge and developing global capacity
- Cannot forget the wider ML community beyond official statistics- where do public-private partnerships fit in?

Resources

UNECE Wiki spaces - [Public](#) and [Members Only](#)

ML Group within the UN Global Network of Data Officers and Statisticians: yammer.com/unstats/

Coordination team reachable at ML2021@ons.gov.uk

