

ONS-UNECE Machine Learning Group 2022

Programme Launch

Alison Baily

International Programme Manager, ONS

InKyung Choi

Statistician, UNECE

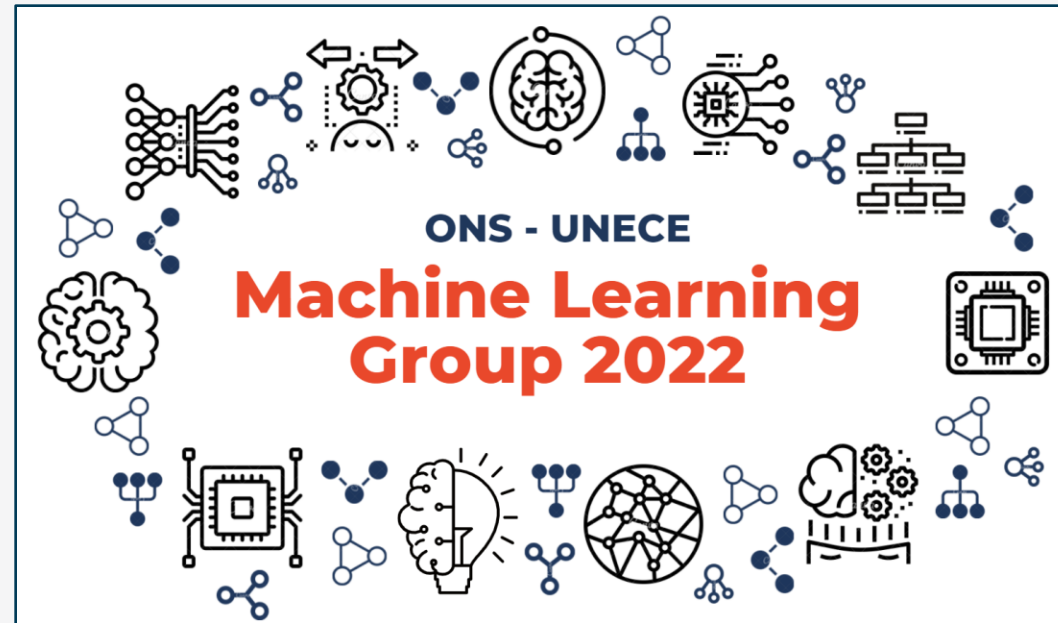
9 February 2022



Welcome to ML 2022!

“enabled me to access a vast repertoire of experience on the use of ML for the production of official statistics”

“helped me understand ML in the context of official statistics and government data science”



“It motivated my team to increase their ML skills”

“I built awareness of different uses for ML applications and how ML applications combine with other tools in the statistical processes.”

“Learning which projects other organisations are successfully doing helps us allocate our limited ML capacity.”

Membership survey, October 2021

Objectives

Platform to facilitate

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

Community driven

- Members design programme and provide content
- Interaction and collaboration is key
- Every contribution is welcome!

Public good

- Open to all official statistical organisations
- Accessible to different levels of expertise
- Resources shared with wider community

What we do

Knowledge Sharing

- Monthly meetings with expert presentations
- External engagement at international conferences
- Regular updates of ML news and opportunities

Research Collaboration

- Research projects explore issues from design to implementation
- Findings shared on public website

Capacity Building

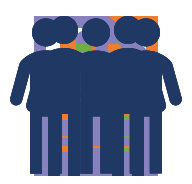
- Coffee and Coding sessions
- Learning and training resources

Item	Topic	Duration
1	Welcome and introduction to ML 2022 - Alison Baily, ONS	00:05
2	Programme Plans – InKyung Choi, UNECE, Alison Baily	00:10
3	Activity Proposals – Part I	00:30
	Coding and Classification, InKyung Choi	
	Area modelling using AIS data, Joseph Crispell, UK FCDO Data Science Hub	
	Model assisted estimation – Sorcha O’Callaghan, Central Statistics Office, Ireland	
	Web scraped data - Michael Reusens, Statistics Flanders	
4	Introduction to ML Group Resources – InKyung Choi	00:05
5	Activity proposals – Part II	00:30
	Convolutional neural networks for learning target variables extracting image features from EO, Joep Burger, CBS Netherlands	
	Crop identification using satellite imagery and ML, Nathalie Uwamahoro, National Institute of Statistics Rwanda	
	Model retraining – Alison Baily	
	Assessing quality of ML algorithms - Marco Puts, CBS Netherlands	
	Creating an open base platform for process and analysis, Jakob Engdahl, Statistics Sweden	
6	Next steps and AOB	00:10

Programme Plans

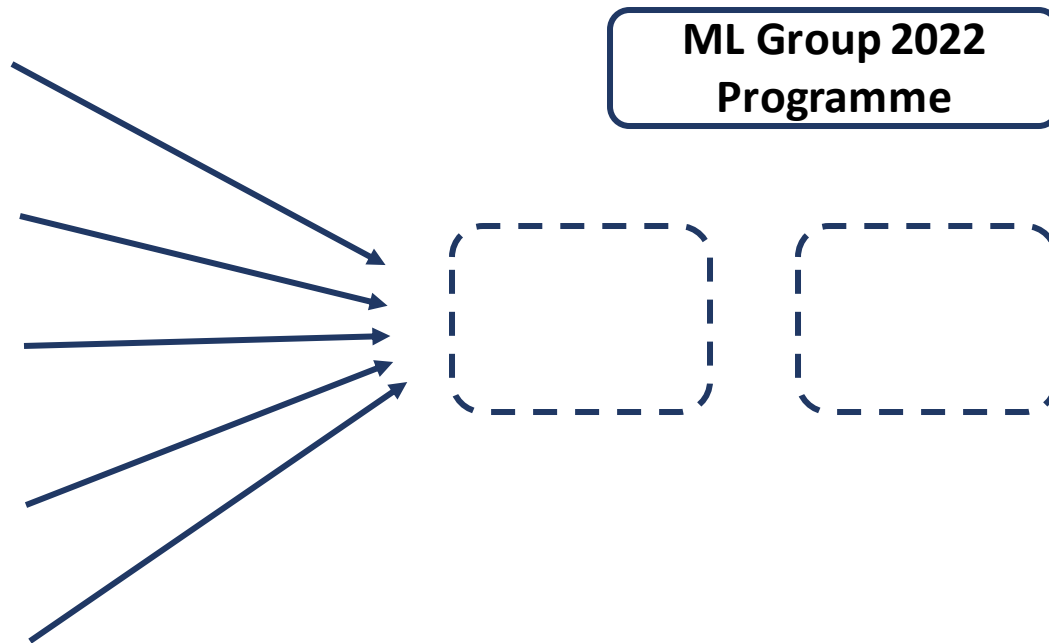
InKyung Choi, UNECE

Alison Baily, ONS



ML 2022 Programme

How it developed.....



ML 2022 members

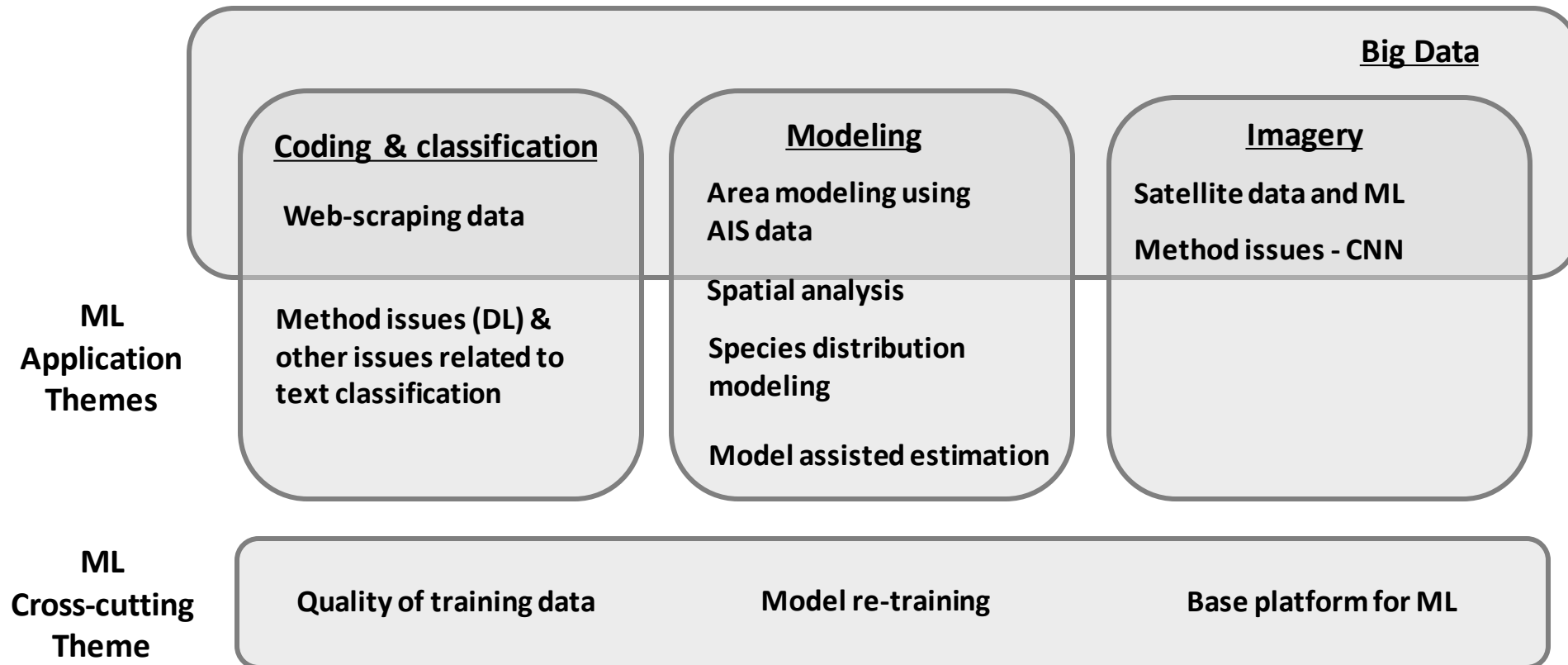
Timeline

Dec 2021 – Jan 2022

Jan 2022-Feb 2022

Feb 2022

ML 2022 Programme



What is your role?



Public

- Public Website
- Final report + webinar
- Coffee and Coding Sessions



All members

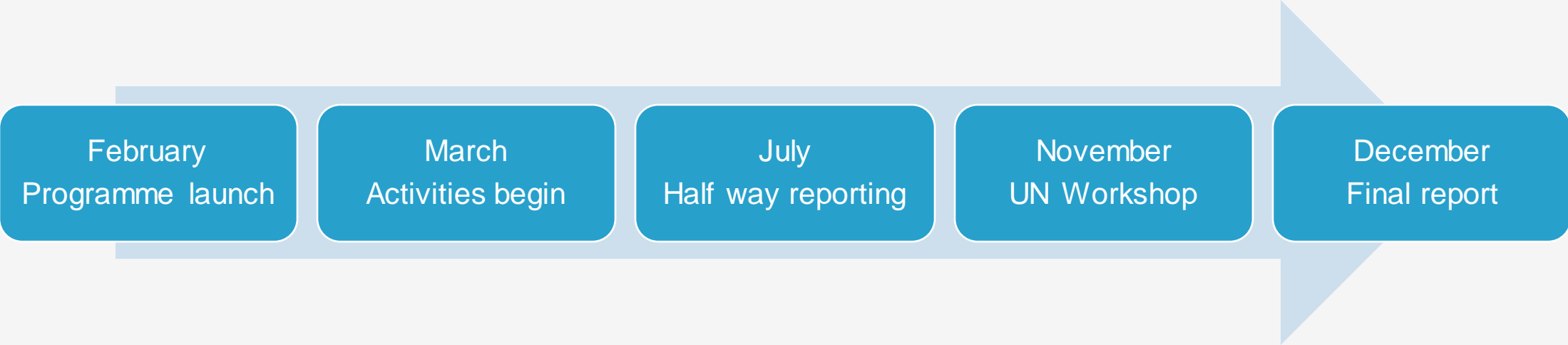
- Monthly meeting
- Newsletter
- Members website
- Catalogue
- Contribute input where possible



Themes

- Research projects
- Study groups
- External presentations
- Regular collaboration

Our plans for 2022



Next steps

1. Decide how you want to be involved in the group this year
2. Let us know by filling out the online poll by Monday 21 February
3. Theme groups – set-up meetings to take place end February/early March
4. Next all group meeting – March 2, 1300-1430 gmt / 1400-1530 CET

Machine Learning page on the UNECE website

Final report 2021

Work Stream 1 (WS1) - From Idea to Valid Solution

The pilot studies are conducted to assess the added value of ML in various thematic areas: coding a classification, edit and imputation the use of imagery data, modeling and route optimization. A study on the replication experience highlighted that benefits of sharing these ML projects.

Theme	Paper
Coding and Classification	Brazil - Apply ML techniques to classification and aggregation web scraped products Turkey - Using Big Data Tools and Machine Learning Techniques to Assign Classification Categories Chile - Coding and Classification: Automated coding of classifiers as a shared report (coming soon) Poland - Using ML classify unstructured information hidden in the text descriptions: estate advertisements - Full report (coming soon) UK - Automated coding of Standard Industrial and Occupational Classifications with github repo
Edit and Imputation	Poland - Multiple imputation through machine learning in a survey of sport clubs
Imagery Analysis	Malaysia - Estimating Malaysia Rubber Plantation Area Productivity Using Satellite Imagery and Machine - Full report (coming soon) Indonesia - Feasibility study of Satellite Imagery Analysis for Wealth Index Development in Indonesia
Modeling	US (BLS) - State level expenditure estimates based on ML techniques
Route Optimisation	Chile - Route Optimisation through genetic algorithm
Replication	Belgium (Flanders) - Replicating successful data science projects across NSOs

ONS-UNECE MACHINE LEARNING 2021 GROUP
By Alison Bailey (UK ONS) and Inkyung Cho (UNECE)

BACKGROUND

Machine Learning (ML) holds great potential for the production of official statistics. It can increase efficiency by automating certain tasks or by assisting manual processes. It also allows statistical organisations to use new types of data such as social media and imagery. Many national and international statistical organisations are exploring how ML can be used to increase the relevance and quality of official statistics in an environment of growing demands for trusted information, rapidly developing and accessible technologies, and numerous competitors. ML is still a relatively new method for official statistics and many statistical organisations are in the process of testing its utility for their data and environment. While local conditions vary, statistical organisations in different countries face similar types of challenges and they can learn from sharing knowledge and experiences with each other, and from working together to develop common solutions.

How international collaboration is advancing machine learning in official statistics

Alison Bailey | January 10, 2022
Categories: International, News

Related links

- Submit an activity proposal for ML 2022
- Public Web - ML 2021
- Leading international collaboration in machine learning for official statistics

Recent Posts

- Visualising in pairs - growing data visualisation skills through mentoring
- How effective mentoring relationships are growing data science skills and capacity
- How international collaboration is advancing machine learning in official statistics
- Boost your data science and visualisation skills in 2022
- Making an international impact with data science mentoring

Data Science for Public Good

Home About us Projects Learning and development Team Careers Contact us

How technologies and data sources have tremendous potential to improve statistical production. They offer a way to generate statistics in a more timely, accurate and cost-efficient manner. Yet, keeping up with the pace of change is challenging, especially for National Statistical Organisations (NSOs) that must innovate with care to maintain a "gold standard" in their outputs. International cooperation between NSOs and other official statistical bodies is one way to help accelerate change in a responsible way.

In 2021, the Office for National Statistics (ONS) and the United Nations Economic Commission for Europe (UNECE) Machine Learning Group (ML 2021) demonstrated the benefits of international cooperation for technological advancement in the development of official statistics. A key objective of the group was to help accelerate change in a responsible way.

Blog on the UK Data Science Campus website

Model retraining

Model retraining

What we achieved in 2021

- Simulation of data drift and feature drift
- Guidance note on the simulation process
- Literature review identifying types of drift
- Importance of data quality monitoring
- ML Ops pipeline
- NSO approaches to model retraining – discussion panel



Reference: Workstream 4 report, Riitta Piela, Statistics Finland, on the ML 2021 website

Model retraining

Plans for 2022

- Research and study group
 - Simulation of model drift
 - Testing of drift detection methods and tools
 - Monitoring results of model for predicting indicator codes
 - MLOps (IT infrastructure and ML-process)
 - Sharing knowledge and experience of model retraining

Next steps

1. Decide how you want to be involved in the group this year
2. Let us know by filling out the online poll by Monday 21 February
3. Theme groups – set-up meetings to take place end February/early March
4. Next all group meeting – March 2, 1300-1430 gmt / 1400-1530 CET

Machine Learning page on the UNECE website

Final report 2021

Work Stream 1 (WS1) - From Idea to Valid Solution

The pilot studies are conducted to assess the added value of ML in various thematic areas: coding a classification, edit and imputation the use of imagery data, modeling and route optimization. A study on the replication experience highlighted that benefits of sharing these ML projects.

Theme	Paper
Coding and Classification	Brazil - Apply ML techniques to classification and aggregation web scraped products Turkey - Using Big Data Tools and Machine Learning Techniques to Assign Classification Categories Chile - Coding and Classification: Automated coding of classifiers as a shared report (coming soon) Poland - Using ML classify unstructured information hidden in the text descriptions: estate advertisements - Full report (coming soon) UK - Automated coding of Standard Industrial and Occupational Classifications with github repo
Edit and Imputation	Poland - Multiple imputation through machine learning in a survey of sport clubs
Imagery Analysis	Malaysia - Estimating Malaysia Rubber Plantation Area Productivity Using Satellite Imagery and Machine - Full report (coming soon) Indonesia - Feasibility study of Satellite Imagery Analysis for Wealth Index Development in Indonesia
Modeling	US (BLS) - State level expenditure estimates based on ML techniques
Route Optimisation	Chile - Route Optimisation through genetic algorithm
Replication	Belgium (Flanders) - Replicating successful data science projects across NSOs

ONS-UNECE MACHINE LEARNING 2021 GROUP
By Alison Bailey (UK ONS) and Inkyung Cho (UNECE)

BACKGROUND

Machine Learning (ML) holds great potential for the production of official statistics. It can increase efficiency by automating certain tasks or by assisting manual processes. It also allows statistical organisations to use new types of data such as social media and imagery. Many national and international statistical organisations are exploring how ML can be used to increase the relevance and quality of official statistics in an environment of growing demands for trusted information, rapidly developing and accessible technologies, and numerous competitors. ML is still a relatively new method for official statistics and many statistical organisations are in the process of testing its utility for their data and environment. While local conditions vary, statistical organisations in different countries face similar types of challenges and they can learn from sharing knowledge and experiences with each other, and from working together to develop common solutions.

How international collaboration is advancing machine learning in official statistics

Alison Bailey | January 10, 2022
Categories: International, News

Related links

- Submit an activity proposal for ML 2022
- Public Web - ML 2021
- Leading international collaboration in machine learning for official statistics

Recent Posts

- Visualising in pairs - growing data visualisation skills through mentoring
- How effective mentoring relationships are growing data science skills and capacity
- How international collaboration is advancing machine learning in official statistics
- Boost your data science and visualisation skills in 2022
- Making an international impact with data science mentoring

Data Science for Public Good

Home About us Projects Learning and development Team Careers Contact us

How technologies and data sources have tremendous potential to improve statistical production. They offer a way to generate statistics in a more timely, accurate and cost-efficient manner. Yet, keeping up with the pace of change is challenging, especially for National Statistical Organisations (NSOs) that must innovate with care to maintain a "gold standard" in their outputs. International cooperation between NSOs and other official statistical bodies is one way to help accelerate change in a responsible way.

In 2021, the Office for National Statistics (ONS) and the United Nations Economic Commission for Europe (UNECE) Machine Learning Group (ML 2021) demonstrated the benefits of international cooperation for technological advancement in the production of official statistics.

Blog on the UK Data Science Campus website