



UNECE HLG-MOS

Machine Learning for Official Statistics

Project outputs are available on [this wiki page](#)

- When you join WebEx, you will be asked to provide "name" and "email", **please use country/organisation, first name and last name as "name"** (e.g. UNECE_InKyung Choi). If you want to change the name after log-in, please check [this instruction](#)
- For smooth proceeding, **please mute microphone** (except when you speak) **and turn off camera**.
- If you don't hear any sound, the sound quality is bad, or others can't hear you, please double check your device connection. See detailed instructions at [Webex Audio Troubleshooting](#).
- If you have any problem or question about connection during the webinar, please send a chat message to UNECE

Day 1 - November 16

Time (CET)	Presentation	Speakers	Documents
13:30 - 13:40	Welcome and introduction to ML project & Pilot studies	Stéphane Dufour (Statistics Canada and Co-chair of HLG-MOS Executive Board) Claude Julien (UNECE) Eric Deeben (UK, ONS Data Science Campus)	Slides
Work Package 1. Pilot Study: Theme - Coding & Classification			
13:40 - 13:55	Occupation and Economic activity coding using natural language processing	Jael Pérez Sánchez (INEGI, Mexico)	Presentation
13:55 - 14:10	Demonstration with ML code and data	Krystyna Piatkowska & Marta Kruczek-Szepel (Statistics Poland)	Slides User' experiences
14:10 - 14:20	Questions and answers		
14:20 - 14:40	Coding and Classification theme report	Claus Sthamer (UK, ONS Data Science Campus)	Presentation Theme Report Wiki
14:40 - 15:00	Questions, answers and discussion		
15:00 - 15:10	Break		
Work Package 1. Pilot Study: Theme - Editing and Imputation			
15:10 - 15:25	Imputation of the variable "Attained Level of Education" in Base Register of Individuals	Fabrizio De Fausti (Istat, Italy)	Presentation
15:25 - 15:40	Machine Learning for Data Editing Cleaning in NSI : Some ideas and hints	Fabiana Rocci (Istat, Italy)	Presentation
15:40 - 15:50	Questions and answers		
15:50 - 16:10	Editing and Imputation theme report	Florian Dumpert (Destatis, Germany)	Presentation Theme Report Wiki
16:10 - 16:30	Questions, answers and discussion		
16:30 - 16:40	Break		
Work Package 2. Quality			
16:40 - 17:10	Quality Framework for Statistical Algorithms (QF4SA)	Wesley Yung (Statistics Canada)	Presentation
	Explainability	Joep Berger (Statistics Netherlands)	Presentation
17:10 - 17:30	Questions, answers and discussion		

Day 2 - November 17

Time (CET)	Presentation	Speaker	Documents
Work Package 1. Pilot Study: Imagery			
13:30 - 13:45	Generic Pipeline for Production of Official Statistics Using Satellite Data and Machine Learning	InKyung Choi (UNECE)	Presentation
13:45 - 14:00	Address Register Automated Image Recognition (AIR) model	Daniel Merkas & James Farnell (Australian Bureau of Statistics)	Presentation
14:00 - 14:10	Questions and answers		
14:10 - 14:30	Imagery theme report	Abel Coronado and Jimena Juárez Carrillo (INEGI, Mexico)	Presentation Theme Report Wiki
14:30 - 14:50	Questions, answers and discussion		
14:50 - 15:05	Pilot studies - summary	Eric Deeben (ONS Data Science Campus, UK)	Presentation
15:05 - 15:15	Break		
Work Package 3. Integration of Machine Learning			
15:15 - 15:40	Integration of Machine Learning	Alex Measure (Bureau of Labor Statistics, USA)	Presentation
15:40 - 16:00	Questions, answers and discussion		
Project output, conclusion and future			
16:00 - 16:15	Machine Learning Project Outputs	InKyung Choi & Claude Julien (UNECE)	Slides HLG-MOS Machine Learning Project
16:15 - 16:30	Conclusion	Claude Julien (UNECE)	Slides
16:30 - 16:45	Break		
16:45 - 17:30	Future directions	Eric Deeben (ONS Data Science Campus, UK)	Slides Poll results

ABOUT THE WEBINAR

With rapidly growing interest in the use of machine learning for official statistics but with limited experience with concrete applications, there was a great need for a common platform where experts in national statistics offices to test their ideas, exchange experiences and collaborate on developments. National statistics offices work on similar type of problems and operate with similar business constraints, hence can benefit from developing shared understanding. To address this need, the [High-Level Group on Modernisation of Official Statistics \(HLG-MOS\)](#) launched the Machine Learning project in early 2019 with aims to:

- Investigate and demonstrate the value added of ML in the production of official statistics, where "value added" is increase in relevance, better overall quality or reduction in costs;
- Advance the capability of ML to add value to the production of official statistics;
- Advance the capability of national statistical organisations to use ML in the production of official statistics;
- Enhance collaboration between statistical organisations in the development and application of ML.

Following these objectives, the project team identified three main areas to advance the use of ML in statistical organisations:

- Work package 1 – Pilot Studies (demonstration of value added)
 - Coding and Classification
 - Edit and Imputation
 - Imagery
- Work package 2 – Quality
- Work package 3 – Integration of ML into organisation

This webinar is the first public event where the outputs of the project will be communicated. This includes study reports, shared code and data, analysis of value added, recommended ML practices, quality framework elements and examples of organisational practices to address integration challenges.

The project will officially close at the end of the year (2020). Since it was launched in March 2019, the number of participants and other collaborators has grown from 20 to over 120. Given this strong interest, the project will evolve into a group to continue the advancement of ML in the production of official statistics.