

**REPORT OF THE WORKSHOP**

1. The workshop was organized as part of the Conference of European Statisticians' work programme for 2023, within the context of the High-Level Group for the Modernisation of Official Statistics (HLG-MOS) activity. It was held on 5-7 June 2023 in Geneva, Switzerland.
2. The meeting was attended by 54 participants, including representatives of national statistical offices and government agencies of the following 22 countries: Australia, Austria, Belgium, Canada, Cyprus, Ecuador, Estonia, Finland, France, Germany, Ireland, Luxembourg, Netherlands (Kingdom of the), Norway, Oman, Philippines, Portugal, Slovenia, Spain, Sweden, Switzerland and United Kingdom of Great Britain and Northern Ireland.
3. In addition to experts from academia, private sector and non-government organization (NGO), the workshop was also attended by representatives from the Bank for International Settlements (BIS), the United Nations Interim Administration Mission in Kosovo (UNMIK), the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP), the Organisation for Economic Co-operation and Development (OECD), the United Nations Statistical Division, the International Labour Office (ILO) and the United Nations Economic Commission for Europe (UNECE).
4. The Organizing Committee of the workshop included: Michael Reusens (Statistics Flanders, Belgium), Riitta Piela (Statistics Finland), Florian Dumpert (Federal Statistical Office of Germany), Dominika Nowak (Statistics Poland), Joni Karanka (Office for National Statistics, UK) and Ralf Becker (UN Statistical Division).
5. The workshop consisted of the following substantive topics in which a total of twenty-four presentations were made:
  - Session 1: Machine Learning (ML) Applications
  - Session 2: Quality Aspects of ML in Official Statistics
  - Session 3: Toward System-wide Transformation of Statistical Production
6. In addition to the presentations, the workshop featured a keynote address by Prof. Diego Kuonen (Statoo Consulting & Geneva School of Economics and Management, University of Geneva). The meeting also included a 3-hour hands-on lab titled "An Introduction to MLOps with Mlflow," provided by the National Institute of Statistics and Economic Studies (Insee, France), and a 30-minute soapbox session.
7. The workshop included a group discussion on the facilitators and blockers for transforming statistical organizations. The main points identified were:
  - Inadequate infrastructure (hardware and software) and limited financial investment/lack of funding pose significant barriers to the implementation of ML projects.
  - A lack of trust in ML, challenges with access to appropriate data (which is crucial for building ML solution) coupled with administrative burdens and security concerns, hinders the ML initiatives.
  - There is a general lack of knowledge in organizations and this creates over-dependence on a few skilled individuals.
  - Unclear goals, benefits, and key performance indicators (KPIs), along with impatience for fast results, complicates the ML project.
  - Agile project management ("fail fast and cheap") and a culture of innovation and experimentation, supported by sandbox environments, encourages creativity and enables iterative development.

- Knowledge sharing in the community of practice (within the organizations or across different organizations) creates a strong learning environment.
- Transparent processes and clear communication of goals help build trust.
- Having strong, enthusiastic, and skilled multi-disciplinary teams is essential for the successful implementation of ML projects.
- Sharing success stories serves as internal motivation and can create a snowball effect.

8. The timetable, papers, and presentations from the workshop are available on the UNECE website <https://unece.org/statistics/events/ML2023>.

## **Annex I. Workshop Proceeding**

### **Session 1: Machine Learning Applications**

9. This session was chaired by Michael Reusens (Statistics Flanders, Belgium) and Joni Karanka (Office for National Statistics, UK) and it included the following presentations:

- Classifying companies in France using machine learning - Thomas Faria and Tom Seimandi (Insee, France)
- Using Webdata to derive the Economic Activity of Enterprises - Manveer Mangat (Statistics Austria)
- Clothing Price Index using Web-Scraped Data - Laura Christen and Ahmet Aydin (Office for National Statistics, UK)
- Imputation of occupation in the Occupational Register - Jens Malmros (Statistics Sweden)
- Too good to be true? A case of machine learning in the validation process of the R&D statistics - Eva Charlotte Berner and Solveig Bjørkholt (Statistics Norway)
- Geospatial Bayesian Methods for Hazard-Impact Modelling - Hamish Patten (University of Oxford)
- Progression patterns in the Swiss social security system based on Machine Learning: methods for evaluating quality and model drift - Athanassia Chalimourda (Swiss Federal Statistics Office)
- ML Poverty: Using Machine Learning to estimate poverty rates in Switzerland at the canton level - Yara Abu Awad (Swiss Federal Statistics Office)
- Creating a modern business index: Machine learning record linkage at scale - Isabela Breton and Joanne Sheppard (Office for National Statistics, UK)
- Time Series Outlier Detection using Metadata and Data Machine Learning in Statistical Production - Olivier Sirello (BIS)
- Timeliness and Accuracy with Machine Learning Algorithms: Early Estimates of the Industrial Turnover Index - David Salgado (Statistics Spain)
- Nowcasting TiVA indicators: improving timeliness of trade data - Polina Knutsson (OECD)

### **Session 2: Quality Aspects of Machine Learning in Official Statistics**

10. This session was chaired by Florian Dumpert (Federal Statistical Office of Germany) and Ralf Becker (UN Statistical Division) and it included the following presentations:

- Quality Framework for Statistical Algorithms - InKyung Choi (UNECE)
- A Quality Concept for the Use of Machine Learning in Official Statistics - Florian Dumpert (Federal Statistical Office of Germany)
- Exploring quality dimensions in trustworthy Machine Learning in the context of official statistics: model explainability, accuracy and uncertainty - Saeid Molladavoudi (Statistics Canada)
- Understanding model quality in the context of trustworthiness and value - Emily Barrington (Office for Statistics Regulation, UK)
- Lessons learned when applying Machine Learning in Official Statistics: Why it helps to be a survey statistician and a data scientist! - Piet Daas (Statistics Netherlands)
- Changing Data Sources in the Age of Data Science for Official Statistics - Cedric De Boom (Statistics Flanders, Belgium)

### **Session 3: Toward system-wide transformation of statistical production**

11. This session was chaired by Riitta Piela (Statistics Finland) and Dominika Nowak (Statistics Poland) and it included the following presentations:

- Facilitators and Blockers of ML Adoption in Official Statistics - Joni Karanka (ONS, UK)

- A Machine Learning Capability Uplift Strategy - Claire Clarke (Australian Bureau of Statistics)
- ML training: Who? What? How? and... What for? - Christophe Bontemps (UN Statistical Institute for Asia and the Pacific, ESCAP)
- Balsam: A Collaborative Platform to Support ML and ML-Ops initiatives - Jakob Engdahl (Statistics Sweden)
- An open-source data science platform to foster innovative and production-ready machine learning systems - Romain Avouac (Insee, France)

### **Soapbox Session**

12. This 30-minute soapbox session provided an opportunity for participants to share their ideas, projects, and collaborations related to ML and data science in a more casual and informal setting. Following three presentations were made:

- ML applications in official statistics of countries with small populations - Hans Hõrak (Statistics Estonia)
- Integrating official statistics into large language model (LLM)-services - Jakob Engdahl (Statistics Sweden)
- A project on ISCO classification with ML - Weichen Lei (ILO)