Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Seventy-second plenary session
Geneva, 20 and 21 June 2024
Item 11 (a) of the provisional agenda
Programme of work of the Statistics subprogramme
of the United Nations Economic Commission for Europe:
Reports on the work of the Conference of European Statisticians,
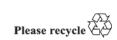**its Bureau and Teams of Specialists**

Report of the Machine Learning for Official Statistics Workshop

Prepared by the Secretariat*

Summary

  The document presents the key outcomes of the Machine Learning for Official Statistics Workshop, which took place in Geneva, Switzerland on 5–7 June 2023. The Conference is invited to take note of its contents, and to provide any guidance as appropriate.

 * This document was submitted late for processing due to resource constraints.

Please recycle

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 in Geneva, Switzerland on 5–7 June 2023.

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 the United Kingdom of Great Britain and Northern Ireland.

3.   In addition to experts from academia, private sector and a non-governmental organization (NGO), the workshop was also attended by representatives from the Bank for International Settlements (BIS), the United Nations Interim Administration Mission in Kosovo[1] (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 Statistics Division (UNSD), the International Labour Organization (ILO) and the United Nations Economic Commission for Europe (UNECE).

4.   The Organizing Committee of the workshop included: M. Reusens (Statistics Flanders, Belgium), R. Piela (Statistics Finland), F. Dumpert (Federal Statistical Office of Germany), D. Nowak (Statistics Poland), J. Karanka (Office for National Statistics, United Kingdom) and R. Becker (UNSD).

5.   The workshop consisted of the following substantive topics in which a total of 24 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. D. 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 overdependence on a few skilled individuals.

- Unclear goals, benefits, and key performance indicators (KPIs), along with impatience for fast results, complicate the ML project.

- Agile project management ("fail fast and chape") and a culture of innovation and experimentation, supported by sandbox environments, encourage creativity and enable iterative development.

- Knowledge-sharing in the community of practice (within the organizations or across different organizations) creates a strong learning environment.

---

[1] References to Kosovo shall be understood to be in the context of the United Nations Security Council resolution 1244 (1999).

- Transparent processes and clear communication of goals help build trust.

- Having strong, enthusiastic, and skilled multidisciplinary 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

Workshop proceeding

Session 1: Machine learning applications

1.   This session was chaired by M. Reusens (Statistics Flanders, Belgium) and J. Karanka (Office for National Statistics, United Kingdom) and it included the following presentations:

- Classifying companies in France using machine learning – T. Faria and T. Seimandi (Insee, France)

- Using Webdata to derive the Economic Activity of Enterprises – M. Mangat (Statistics Austria)

- Clothing Price Index using Web-Scraped Data – L. Christen and A. Aydin (Office for National Statistics, United Kingdom)

- Imputation of occupation in the Occupational Register – J. Malmros (Statistics Sweden)

- Too good to be true? A case of machine learning in the validation process of the R&D statistics – E. Berner and S. Bjørkholt (Statistics Norway)

- Geospatial Bayesian Methods for Hazard-Impact Modelling – H. Patten (University of Oxford)

- Progression patterns in the Swiss social security system based on Machine Learning: methods for evaluating quality and model drift – A. Chalimourda (Swiss Federal Statistics Office)

- ML Poverty: Using Machine Learning to estimate poverty rates in Switzerland at the canton level – Y. Awad (Swiss Federal Statistics Office)

- Creating a modern business index: Machine learning record linkage at scale – I. Breton and J. Sheppard (Office for National Statistics, United Kingdom)

- Time Series Outlier Detection using Metadata and Data Machine Learning in Statistical Production – O. Sirello (BIS)

- Timeliness and Accuracy with Machine Learning Algorithms: Early Estimates of the Industrial Turnover Index – D. Salgado (Statistics Spain)

- Nowcasting TiVA indicators: improving timeliness of trade data – P. Knutsson (OECD).

Session 2: Quality aspects of machine learning in official statistics

2.   This session was chaired by F. Dumpert (Federal Statistical Office of Germany) and R. Becker (UNSD) and it included the following presentations:

- Quality Framework for Statistical Algorithms – I. Choi (UNECE)

- A Quality Concept for the Use of Machine Learning in Official Statistics – F. Dumpert (Federal Statistical Office of Germany)

- Exploring quality dimensions in trustworthy Machine Learning in the context of official statistics: model explainability, accuracy and uncertainty – S. Molladavoudi (Statistics Canada)

- Understanding model quality in the context of trustworthiness and value – E. Barrington (Office for Statistics Regulation, United Kingdom)

- Lessons learned when applying Machine Learning in Official Statistics: Why it helps to be a survey statistician and a data scientist! – P. Daas (Statistics Netherlands)

• Changing Data Sources in the Age of Data Science for Official Statistics – C. De Boom (Statistics Flanders, Belgium).

Session 3: Toward system-wide transformation of statistical production

3.   This session was chaired by R. Piela (Statistics Finland) and D. Nowak (Statistics Poland) and it included the following presentations:

• Facilitators and Blockers of ML Adoption in Official Statistics – J. Karanka (ONS, United Kingdom)

• A Machine Learning Capability Uplift Strategy – C. Clarke (Australian Bureau of Statistics)

• ML training: Who? What? How? and… What for? – C. Bontemps (United Nations Statistical Institute for Asia and the Pacific, ESCAP)

• Balsam: A Collaborative Platform to Support ML and ML-Ops initiatives – J. Engdahl (Statistics Sweden)

• An open-source data science platform to foster innovative and production-ready machine learning systems – R. Avouac (Insee, France).

Soapbox session

4.   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 – H. Hõrak (Statistics Estonia)

• Integrating official statistics into large language model (LLM) services – J. Engdahl (Statistics Sweden)

• A project on ISCO classification with ML – W. Lei (ILO).

———————