High-Level Group for the Modernisation of Official Statistics

The HLG-MOS oversees and manages:

⭐ Collaboration groups: Supporting Standards, Capabilities and Communication, Applying Data Science and Modern Methods, Blue Skies and Machine Learning Community

⭐ Modernisation Projects 2023: Data Governance for Interoperability Framework 2, Cloud for Official Statistics and ModernStats Carpentries

⭐ Specialised topics: Dissemination and Communication, Data Collection, Data Editing and Confidentiality

⭐ Modernisation Expert group meetings, workshops and seminars

What is the HLG-MOS
HLG-MOS Outputs
Modernstats Updates

Generic Statistical Business Process Model (GSBPM) describes the core business processes undertaken by statistical organisations to produce statistical outputs. It is used by more than 50 organisations world wide. Read more here

Generic Statistical Information Model (GSIM) describes the core pieces of information needed by statistical organisations to produce statistical outputs. Read more here

Common Statistical Data Architecture (CSDA) is reference architecture and guidance for the modernisation of their processes and systems. Read more here

Strategic Communication Framework is a guide for Statistical Organisations to the development and implementation of a communication strategy. Read more here

Generic Activity Model for Statistical Organisations (GAMSO) extends and complements the GSBPM by describing overarching activities and processes to support the production of official statistical production. Read more here
Common Statistical Production Architecture (CSPA) helps statistical organisation create interoperable tools to share within and between statistical organisations. Read more here

Generic Statistical Data Editing Model (GSDEM) is intended as a reference for all official statisticians whose activities include data editing. Read more here

Machine Learning for Official Statistics the 2019-2020 project continued as a collaboration community in 2021. Read more here

Blog on Recent highlights and updates

Blog Posts

- Blog: New Group on Applying Data Science and Modern Methods created by Wai Kit Si Tou
  04 May, 2022
  High-Level Group for the Modernisation of Official Statistics

- Blog: Machine learning for official statistics created by InKyung Choi
  05 Apr, 2022
  High-Level Group for the Modernisation of Official Statistics

- Blog: Use of ModernStats Models: Progress and Way Forward created by Wai Kit Si Tou
  20 Oct, 2021
  High-Level Group for the Modernisation of Official Statistics

- Blog: New framework sets out a strategic approach to statistical communication created by Taeke Gjaltema
  22 Jul, 2021
  High-Level Group for the Modernisation of Official Statistics

- Blog: Despite the Covid19 pandemic, the involvement in the ModernStats Community increased from 2019 to 2020 update: 2021 involvement created by Taeke Gjaltema
  27 Jan, 2021
  High-Level Group for the Modernisation of Official Statistics
We are glad to announce the establishment of the new group on Applying Data Science and Modern Methods in 2022.

In view of the increasing importance of new data sources and methods for the compilation of official statistics, this new group aims to identify data science initiatives and new methods needed for modernising existing business processes.

Examples of potential activities cover: (i) developing supporting materials to help implement case studies and good practices, (ii) organising workshops and trainings to promote and ensure consistent use of the HLG-MOS supported data science initiatives and modern methods, and (iii) managing the periodic reviews of the new data science initiatives and methods to measure their impacts.

The group has already started conducting a market landscape analysis to take stock of existing work in the field of data science and modern methods. It will help identify specific needs, challenges, and opportunities that support modernising statistical production and services.

Some potential areas of work that the group is considering include (i) use cases on data collection methods, (ii) data platform that facilitates the access to and integration of data from different sources, (iii) responsible AI, (iv) modelling methods for different type of data, and (v) automated dissemination, with user-centric services.

We look forward to welcoming you to participate in the group and other activities.

Please contact Wai Kit Si Tou if you are interested in joining the group.

Machine learning for official statistics

Machine learning for official statistics: UNECE help statistical organisations harness the power of machine learning

The new UNECE publication Machine Learning for Official Statistics helps national and international statistical organizations harness the power of machine learning to modernize the production of official statistics.

Machine learning in the context of modernisation for statistical organizations

Statistical organizations produce crucial indicators that portray various aspects of the economy and society that we live in. These include measures such as the gross domestic product, the inflation rate, the population growth, and the unemployment rate on which government and business alike depend when making important decisions. While these may be just a few digit numbers from the end-user side, statistical organizations employ a series of carefully designed and executed processes to distil this key information from the vast amount of raw data.
With increasing challenges arising from new data sources, technological developments and competitions with private companies, statistical organizations have been striving to modernise every part of this production process to provide more relevant and detailed official statistics in a more timely and accessible manner. They utilise the computer-assisted interview and web scraping tool to collect data more efficiently, and build infrastructure for data and IT tools to manage them across the organizations more easily.

Yet, one area that is difficult to modernise is the processes that require “human-like” decision-making, such as reading a textual description to assign a matching classification code or looking at the image to identify what it represents. Traditionally, this has been done either manually or through a complex rule-based system, both of which are costly, time-consuming and hard to manage. This is particularly daunting when statistical organizations try to use big data sources (e.g., price information web-scrapped from online stores) as the cost of resources needed to process such a large amount of data in the traditional manual way is simply too prohibitive.

Machines learning holds great potential for statistical organizations

The recent developments in machine learning technique are pushing the boundary of tasks considered for humans and machines - machines can now draw a painting in the style of an old master and write an article just like humans.

How does this technology work? In one of the most popular approaches called “supervised learning”, machines are first trained on the data that humans labelled, for example, images labelled as “urban” or “rural”. With this data, they figure out patterns associated with labels by adaptively improving their internal logic that maps from the input (image) to output (label). In this way, machines can determine whether an area shown in an image is urban or rural without us providing all possible rules explicitly.

As the machine learning technique can carry out tasks that we used to solely rely on manual works, it holds a great potential to increase the efficiency of statistical organizations, just like the use of machinery powered by steam engines made a huge leap in the productivity in the manufacturer industry few centuries ago. Also, their capability to process various types of data such as text, image and video offers statistical organizations to take advantage of new data sources to produce new statistics that could meet the evolving needs of society.

Challenges in using machine learning for official statistics

Like with any innovation, however, the journey of integrating machine learning in the organization abounds with challenges and setbacks. The technology itself is still relatively new and requires a different skill set that many statistical organizations do not possess; hence it needs to be built inside or acquired from outside.

The real difficulty, however, starts when the machine learning solution needs to move to production, meaning that it is connected to existing processes seamlessly and used for the regular business, beyond an “experiment” stage. Unfortunately, even after successful pilot studies, many machine learning solutions end up being left on the shelf. The difficulty is experienced widely across sectors and domains. It is said that over 80% of machine learning projects never make it to production. Moving machine learning into production requires changes in infrastructure, culture, organizational structure or business processes, none of which is a small task with a lasting effect.

UNECE supports statistical organizations in advancing the use of machine learning
Based on the two international initiatives, the UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) Machine Learning Project (2019-20) and the United Kingdom Office of National Statistics (ONS) – UNECE Machine Learning Group 2021, the publication Machine Learning for Official Statistics aims to help statistical organizations navigate the difficult journey of advancing the use of this new technology. It presents the practical applications of machine learning in three working areas within statistical organizations and discusses their value-added, challenges and lessons learned. The publication also includes a quality framework that could help guide the choice of methods, demonstrates key steps for moving machine learning from the experimental stage to the production stage, and key messages to facilitate the use of machine learning in the statistical organizations.

The machine learning field is fast evolving with new methods, platforms and approaches coming out every month. To keep up with the pace of change and avoid duplication of efforts, there is a great need for knowledge sharing and collaboration within the official statistics community. UNECE continues its engagement in the international initiative this year, through Machine Learning Group 2022 with the ONS, to support statistical organizations to harness the power of machine learning.