

# HLG-MOS Machine Learning Project

Machine Learning (ML) holds a great potential for statistical organisations. It can make the production of statistics more efficient by automating certain processes or assisting humans to carry out the process. It also allows statistical organisations to use new types of data such as social media data and imagery.

Many national statistical offices (NSOs) are investigating 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. While specific business environment may vary depending on country, NSOs face similar type of challenges which can benefit from sharing knowledge and experiences, and collaborating on developing common solutions within the broad official statistical community.

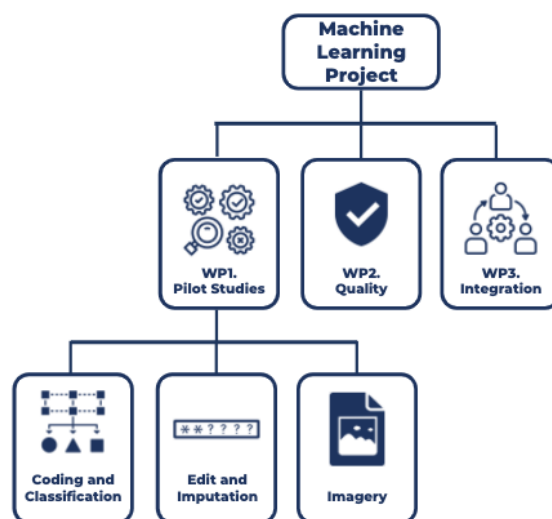
To address this need, [UNECE High-Level Group for the Modernisation of Official Statistics \(HLG-MOS\)](#) launched a Machine Learning Project in 2019. The project aimed to demonstrate the added value of ML, i.e. whether its enables to production of more relevant, timely, accurate and trusted data in an efficient manner. The project also aimed at increasing the capability of NSOs to use ML by identifying and addressing some common challenges encountered when incorporating ML in organisations and their production processes.

The project started in April 2019 with 23 participants from 13 organisations and has grown to over 120 members from 23 countries, 31 national and 4 international organisations. The members either lead, assist or follow numerous studies and other developments. The work of the project is divided into three work packages:

- Work Package (WP) 1. Pilot studies
- Work Package (WP) 2. Quality
- Work Package (WP) 3. Integration challenges

**The project is immensely pleased to share its numerous outputs with the official statistics community!!!**

## Overall structure of UNECE HLG-MOS Machine Learning Project



The Machine Learning Community continues on its journey!

**[CLICK HERE TO ACCESS THE MACHINE LEARNING 2021 GROUP PAGE](#)**

**Machine Learning Project Report:** summary of the project and recommendations on how to advance the use of ML in statistical organisations based on lessons learned and concrete experiences from three work packages (WPs).

#### **WP1 Output**

- Reports and other documents on 19 pilot studies; early developments on the use of ML for data editing; and a generic pipeline for production of official statistics using satellite data and machine learning. - [available at WP1 - Pilot Studies](#). The pilot studies are conducted to assess the added value of ML in three thematic areas: coding and classification, edit and imputation and the use of imagery data. They are conducted on a wide variety of data sources (survey, administrative registers, web-scraped, published official statistics, twitter, satellite images, aerial images) and contexts (proof of concepts, production).
- Theme reports that analyse the approaches and results from the pilot studies on each of the three areas - available at
  - [WP1 - Theme 1 Coding and Classification Report](#)
  - [WP1 - Theme 2 Edit and Imputation Report](#)
  - [WP1 - Theme 3 Imagery Analysis Report](#)
- [Executive Summary report on the pilot studies](#)

**WP2 Output:** A Quality Framework for Statistical Algorithms (QF4SA) provides guidance on the choice of algorithms for the production process. It purposely uses the terminology statistical algorithm as it covers both traditional and modern methods. It proposes five quality dimensions; accuracy, timeliness, cost-effectiveness, explainability and reproducibility - [available at WP2 - Quality](#)

**WP3 Output:** The identification of challenges in moving machine learning solutions from a proof of concept to production, as well as a review of some current practices to address some of the challenges - [available at WP3 - Integration](#)

These reports are accompanied with by other material to assist users in getting into or pursuing the development of ML in their respective contexts:

- ML code used in some of the pilot studies - [available at Studies and Codes](#)
- Two sets of data on which to learn and experiment with ML - [available at Learning and Training](#)
- Links to learning and training material - [available at Learning and Training](#)