Survey on the use of Machine Learning within the Official Statistics - Focus on the institutional level

Fields marked with * are mandatory.

1 Introduction

The evolution of Machine Learning (ML) has introduced new techniques and technologies that open exciting possibilities for Official Statistical Production. However, much like past advancements—such as data warehousing and web scraping—the challenges encountered in producing official statistics differ significantly from those faced by most industries. This underscores the importance of fostering a community that can share perspectives and address the unique trials posed by ML in this domain.

The primary goal of this public consultation and questionnaire is to gather insights from members of the European Statistical System (ESS). By capturing these perspectives, the consultation seeks to identify key challenges, explore opportunities, and assess organizational readiness for integrating ML into statistical processes.

As part of this initiative, [AIML4OS] Work Package 4 (WP-4) will conduct two surveys. The first, currently underway, is designed to engage National Statistical Institutes (NSIs). It aims to provide a high-level overview of how ML is being implemented and utilized within NSIs. The survey addresses several key areas, including the general use of ML, the human resources allocated and required for its application, the challenges being faced, and organizational as well as technical limitations.

Survey Scope

The survey is primarily composed of multiple-choice, non-technical questions. Its primary objective is to capture a holistic view of each institution, ensuring that **one cohesive organizational response** per National Statistical Institute (NSI) is provided. The figure below demonstrates the concepts survey 1 is covering:

The survey includes the following sections:

1. **Introduction Section** provides participants with guidelines for answering the survey.
2. **Preliminary Section** characterizes the institution.
3. **Organizational Aspects** addresses ML architecture, integration, ownership, and responsibilities within the NSI.
4. **Quality Standards** assesses the existence of quality standards for ML processes.
5. **Human Resources** focuses on ML skill development strategies and staff competencies.
6. **Current Issues and Challenges** evaluates organizational and technical limitations affecting ML initiatives.

By addressing these areas, the survey aims to provide insights into the state of ML integration, workforce readiness, and key challenges faced by NSIs.

Survey Instructions

1. **Respondent Requirements**

- The survey must be completed by a **Management-level representative** with a **comprehensive understanding** of ML within the NSI.
- Ensure only **one consolidated response per NSI**. If input is required from multiple individuals, responses must be **collated and unified**.

2. **Timeline and Progress**
   - The survey is open from the **1st of April to the 30th June 2025**.
   - It **does not need to be completed in one sitting**—responses can be saved as a draft and finalized later.
3. **Support and Resources**
   - A **help page** is available on the right pane in **English, German, and French** to assist with questions about the survey format.
   - A link is provided for **additional project and WP4 context**.
   - For further assistance, contact **Aiml4os-Wp4@destatis.de**.
4. **Navigation Guidelines**
   - Do **not use browser navigation**. Use the **Previous** and **Next** buttons to navigate.
   - You can only proceed to the next section after completing all **mandatory questions**.
5. **Question Types**
   - **Circles** indicate **single-select questions**.
   - **Squares** indicate **multiple-select questions**.

Follow these instructions to ensure smooth completion of the survey.

2 Preliminary

* 2.1 Which National Statistical Institute (NSI) within the European Statistical System (ESS) do you belong to?

- ○ Albania: Institute of Statistics (INSTAT)
- ○ Andorra: Departament d'Estadística
- ○ Austria: Statistik Austria (STAT)
- ○ Belarus: Национальный статистический комитет Республики Беларусь (Belstat)
- ○ Belgium: Statbel
- ○ Bosnia and Herzegovina: Agencija za statistiku Bosne i Hercegovine (BHAS)
- ○ Bulgaria: Национален статистически институт (NSI)
- ○ Croatia: Državni zavod za statistiku (DZS)
- ○ Cyprus: Στατιστική Υπηρεσία Κύπρου (CYSTAT)
- ○ Czech Republic: Český statistický úřad (ČSÚ)
- ○ Denmark: Danmarks Statistik
- ○ Estonia: Statistikaamet
- ○ Finland: Tilastokeskus (Statistics Finland)
- ○ France: Institut National de la Statistique et des Études Économiques (INSEE)
- ○ Germany: Statistisches Bundesamt (Destatis)
- ○ Greece: Ελληνική Στατιστική Αρχή (ELSTAT)
- ○ Hungary: Központi Statisztikai Hivatal (KSH)
- ○ Iceland: Hagstofa Íslands
- ○ Ireland: Central Statistics Office (CSO)
- ○ Italy: Istituto Nazionale di Statistica (ISTAT)
- ○

Latvia: Centrālā statistikas pārvalde (CSP)
- Liechtenstein: Amt für Statistik
- Lithuania: Lietuvos statistikos departamentas
- Luxembourg: Institut national de la statistique et des études économiques (STATEC)
- Malta: National Statistics Office (NSO)
- Moldova: Biroul Național de Statistică
- Monaco: Institut Monégasque de Statistique et des Études Économiques (IMSEE)
- Montenegro: Zavod za statistiku Crne Gore (MONSTAT)
- Netherlands: Centraal Bureau voor de Statistiek (CBS)
- North Macedonia: Државен завод за статистика (State Statistical Office)
- Norway: Statistisk sentralbyrå (SSB)
- Poland: Główny Urząd Statystyczny (GUS)
- Portugal: Instituto Nacional de Estatística (INE)
- Romania: Institutul Național de Statistică (INS)
- Serbia: Republički zavod za statistiku
- Slovakia: Štatistický úrad Slovenskej republiky (ŠÚ SR)
- Slovenia: Statistični urad Republike Slovenije (SURS)
- Spain: Instituto Nacional de Estadística (INE)
- Sweden: Statistiska centralbyrån (SCB)
- Switzerland: Bundesamt für Statistik (BFS)
- Turkey: Türkiye İstatistik Kurumu (TÜİK)
- United Kingdom: Office for National Statistics (ONS)
- Other

* 2.2 Please provide the name of your organization.

* 2.3 Which option best describes your role in your organization?
- Statistician/Data Scientist
- Analyst/Subject matter specialist
- Manager/Policy Maker
- Software Engineer/ Information Technology Specialist

* 2.4 Are you using Machine Learning in your institution?
- Yes
- No

2.5 If you are not using ML in your institution, can you explain us why?

* 2.6 How do you define ML in your organization? In particular would you agree with this following definition:

"*Machine Learning often (but not always!) involves:*
- non-parametric statistical methods,

- recognizing patterns and relationships,
- a focus – at least in so-called supervised learning – on prediction (and less on explanation),
- methods which are used to answer specific questions without being explicitly given the solution,
- thus are generally data-driven rather than model-driven
- and are characterized by the fact that the solution space (meaning the hypothesis space) is often so large that it (approximately) contains all patterns and relationships"

- ○ Agree completely: This definition accurately captures all key aspects of ML in our organization.
- ○ Mostly agree with minor reservations: The definition is largely correct but overlooks some important aspects to us.
- ○ Partially agree: While some points are valid, this definition is too narrow and doesn't fully represent how we view ML.
- ○ Disagree: Our organization defines ML differently.
- ○ Strongly disagree: This definition is misleading and doesn't align with our understanding of ML at all.
- ○ Neutral: We don't have a formal definition of ML in our organization, so I can't compare it to this one.

3 Organizational aspects

* 3.1 How has Machine Learning evolved and been integrated into your organization over time? Please select all that apply.

- ☐ Through a dedicated digital strategy across the organization that supported the integration of ML into statistical processes
- ☐ By identifying and implementing several high-impact use cases where ML showed significant potential
- ☐ Gradually, through bottom-up initiatives from data scientists and analysts
- ☐ In response to specific business challenges

* 3.2 What is your organization's architecture to support its ambitions towards Machine Learning? Please select the most relevant option(s).

- ☐ Centralized, Strategic Approach - aligned with our overall strategic goals
- ☐ Demand-Driven, Collaborative Approach - Each business unit within our organization is encouraged to propose potential ML applications based on their specific needs and challenges
- ☐ Data-Driven Decision-Making - We perform regular audits to identify areas where large datasets are underutilized or where manual processes can be optimized with ML
- ☐ Innovation Hub and Pilot Projects - Our organization has an innovation hub responsible for exploring ML opportunities. We encourage experimentation by running pilot projects
- ☐ Hybrid Approach with Expert Input - combining both top-down strategic direction and bottom-up input from individual business units

4 Quality Standards

* 4.1 Does your organization have regulations, guidelines, instructions, frameworks or standards that are relevant and ML-specific for the use of ML? Examples could include: special compliance requirements (e.g. data protection, ethics, information security), a special quality framework, special methodological requirements, definition of special roles and responsibilities, etc.

- ○ Yes
- ○ No

* 4.2 Which ones?

* 4.3 Does your organization plan to introduce regulations, guidelines, instructions, frameworks or standards that are relevant and ML-specific for the use of ML in the foreseeable future?

- Yes
- No
- Currently under discussion

4.4 Which ones?

* 4.5 How do you make the application of standards/frameworks visible (internally and/or externally)? Please select the most relevant option(s).

- On the National Statistical Institution website
- In reports
- Through special events or workshops
- In internal newsletters or communication channels
- Via social media posts or blog articles
- In employee training materials
- The communication strategy is currently being developed

* 4.6 Was the introduction of standards/frameworks a bottom-up process (inspired by the needs from daily activities) or a top-down process (management's instruction to issue a guideline, for example)?

- Bottom-up process
- Top-down process
- Both

5 Human Resources

5.1 How do you evaluate the level of competencies and human resources required to support the implementation of machine learning in your organization across the following areas?

	Need to expand /improve	Sufficient	Advanced
* Data Engineering			
* Data Management			
* Data Visualisation			
* Statistical and Mathematical Foundations			
* Programming and Scripting Skills			
* Statistical Domain Expertise (within the statistical business units)			

* Organizational	○	○	○
* Machine Learning Algorithms and Techniques	○	○	○
* Model deployment/MLOps	○	○	○

Indication: In the context of public statistical institutes, a "business unit" refers to a specialized division focused on the design, implementation, and publication of specific statistical products.

* 5.2 Within the domain "Data Engineering": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - ☐ Data Collection and Integration
  - ☐ Data Cleaning and Preparation
  - ☐ Data Storage Management
  - ☐ Data Pipeline Development ETL (Extract, Transform, Load)
  - ☐ Processes Big Data Handling (e.g., Hadoop, Spark)

* 5.3 Within the domain "Data Management": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - ☐ Cloud Data Management (AWS, Azure, GCP)
  - ☐ Data Lifecycle Management
  - ☐ Metadata Management
  - ☐ Data Governance and Compliance

* 5.4 Within the domain "Data Visualisation": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - ☐ Expertise in building interactive visualizations
  - ☐ Competency in developing real-time visual dashboards
  - ☐ Display data patterns or clustering of high-dimensional data
  - ☐ Ability to create custom visualizations tailored to specific ML problems

* 5.5 Within the domain "Statistical and Mathematical Foundation": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - ☐ Probability Theory
  - ☐ Descriptive Statistics
  - ☐ Inferential Statistics
  - ☐ Linear Algebra
  - ☐ Optimization Methods
  - ☐ Bayesian Statistics
  - ☐ Hypothesis Testing for validating assumptions and determining the significance of model results
  - ☐ Matrix Operations
  - ☐ Stochastic Processes, useful for modelling time series, random processes, and reinforcement learning
  - ☐ Imputation
  - ☐ Data Editing
  - ☐ Calibration

* 5.6 Within the domain "Organizational": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - [ ] ML strategy alignment with business goals
  - [ ] Cross-functional collaboration and team structure
  - [ ] ML workflow, process management, and automation
  - [ ] Resource allocation and agile project management
  - [ ] Knowledge sharing, documentation, and change management

* 5.7 Within the domain "Programming and scripting skills": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - [ ] Knowledge of Database querying languages like SQL
  - [ ] Familiarity with ML Libraries (e.g., TensorFlow, PyTorch, Scikit-learn)
  - [ ] Scripting for Automation (e.g. packaging R/Python libraries)
  - [ ] Version Control (e.g., Git)
  - [ ] Shell Scripting Cloud Computing (AWS, GCP, Azure)

* 5.8 Within the domain "Statistical Domain Expertise (within the statistical business units)": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - [ ] Domain-Specific Knowledge
  - [ ] Data Exploration and Profiling capabilities
  - [ ] Feature Engineering Awareness
  - [ ] Ability to collaborate with Data Scientists and Stakeholders

* 5.9 Within the domain "Machine Learning Algorithms and Techniques": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - [ ] Supervised Learning
  - [ ] Unsupervised Learning
  - [ ] Reinforcement Learning
  - [ ] Deep Learning
  - [ ] Natural Language Processing (NLP)
  - [ ] Time Series Analysis
  - [ ] Ensemble Methods
  - [ ] Model Evaluation and Validation
  - [ ] Hyperparameter Tuning
  - [ ] Dimensionality Reduction
  - [ ] Outlier detection
  - [ ] Clustering methods

* 5.10 Within the domain "Model Deployment and MLOps ": Please indicate the positions and skills where you believe your NSI has at least sufficient expertise and resources.
  - [ ] Model Deployment Strategies CI/CD
  - [ ] Pipelines for ML Containerization (e.g., Docker)
  - [ ] Orchestration/Scheduling Tools for workflows (e.g., Apache Airflow)
  - [ ] Model Monitoring
  - [ ] Versioning of Models
  - [ ] Infrastructure as Code (IaC)
  - [ ]

Scaling ML Models

- [ ] Model Governance and Compliance
- [ ] ML Workflow Automation (e.g., MLFlow)
- [ ] API Integration
- [ ] Containerization and Orchestration Tools (e.g. Docker and Kubernetes)

* 5.11 Approximately how many employees work in your statistical agency?

- ( ) Less than 50
- ( ) Between 50 and 500
- ( ) Between 500 and 1000
- ( ) Between 1000 and 2000
- ( ) Between 2000 and 3000
- ( ) More than 3000

5.12 Approximately how many Data Scientists work in your statistical agency?

5.13 Do you plan to expand ML-related team(s)?

5.14 Do you mostly consume/use/configure ML-Algorithms coded in available Libraries, or do you implement/program yourself ML-Algorithms from scratch?

* 5.15 Which strategies has your organization found most effective in developing ML skills and expertise among staff? Please choose only the ones that have been particularly impactful for your organization.

- [ ] ML Trainings within your institution
- [ ] ML Trainings outside your institution
- [ ] By recruiting new Data Scientists
- [ ] ML Conferences
- [ ] Partnering with universities or research institutions
- [ ] ML hackathons or competitions
- [ ] Access to online ML learning platforms (e.g., Coursera, edX)
- [ ] Creating an internal ML community of practice

6 Current issues and challenges

6.1 To what extent do the following organizational issues limit your organization's ability to use Machine Learning?

	Does not limit use	Slightly limits use	Moderately limits use	Severely limits use	Prevents use

* Coordination between internal stakeholders (e.g. R&D, methodology, IT, subject-matter, operations)					
* Resistance from stakeholders inside the organization (e.g. business units)					
* Resistance from stakeholders outside of the organization (e.g. data users)					
* Uncertainty over ML applications/project ownership and responsibilities					
* Lack of clear organizational strategy to identify and prioritize machine learning opportunities across different business units					
* Lack of a central unit in charge of coordinating the implementation of ML solutions across the organization					
* Lack of quality frameworks specific to ML solutions					
* Budget constraints					

6.2 To what extent does your organization invest resources to tackle the following organizational issues?

	No need to invest	Insufficiently	Sufficiently	Substantially
* Coordination between internal stakeholders (e.g. R&D, methodology, IT, subject-matter, operations)				
* Resistance from stakeholders inside the organization (e.g. business units)				
* Resistance from stakeholders outside of the organization (e.g. data users)				
* Uncertainty over ML applications/project ownership and responsibilities				
* Lack of clear organizational strategy to identify and prioritize machine learning opportunities across different business units				
* Lack of a central unit in charge of coordinating the implementation of ML solutions across the organization				
* Lack of quality frameworks specific to ML solutions				

* Budget constraints				

6.3 To what extent do the following technical issues limit your organization's ability to effectively use machine learning?

	Does not limit use	Slightly limits use	Moderately limits use	Severely limits use	Prevents use
* Availability of staff with appropriate machine learning algorithm skills					
* Availability of staff with appropriate programming skills to deploy the models					
* Availability of staff to move the project into production and maintain it					
* Fragmented data infrastructure with isolated servers and diverse hardware /software configurations hinders seamless integration of ML solutions across the organization					
* Insufficient MLOps practices and tools					
* Access to computer hardware					
* Access to software					
* Data quality and bias issues in the training data					
* Access to suitable training data					
* Concerns regarding data privacy and protection					

Indication: MLOps (Machine Learning Operations) is the practice of applying DevOps principles to the development, deployment, and maintenance of machine learning models. It aims to streamline the entire ML lifecycle, from data preparation and model development to production deployment and monitoring.

6.4 To what extent does your organization invest resources to tackle the following technical issues?

	No need to invest	Insufficiently	Sufficiently	Substantially
* Availability of staff with appropriate machine learning algorithm skills				
* Availability of staff with appropriate programming skills to deploy the models				

* Availability of staff to move the project into production and maintain it				
* Fragmented data infrastructure with isolated servers and diverse hardware/software configurations hinders seamless integration of ML solutions across the organization				
* Insufficient MLOps practices and tools				
* Access to computer hardware				
* Access to software				
* Data quality and bias issues in the training data				
* Access to suitable training data				
* Concerns regarding data privacy and protection				