UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting of the 2020/2021 Bureau
Geneva (online), 17-18 February 2021

ECE/CES/BUR/2021/FEB/09/Add.2
25 January 2021

For discussion and
recommendations

Item III (f) of the Provisional
Agenda

REPORT ON THE HLG-MOS MACHINE LEARNING PROJECT

Prepared by the HLG-MOS Machine Learning Project

This document is the main report from the High-level Group for Modernisation of Official Statistics (HLG-MOS) project on machine learning carried out as a priority in 2019 and 2020. The objective of the project was to investigate where machine learning can add value, increase the capability of statistical organizations to use machine learning (ML), and identify common challenges in advancing its use in statistics.

This report presents summary of the work and recommendations on how to advance the use of machine learning in statistical organizations. It is based on lessons learned and experiences from three work packages: pilot studies (WP1), quality (WP2) and integration (WP3).

Separate reports on the three work packages as well as accompanying materials (codes, data) are available on the [UNECE Statistics Wiki](#) and include the following:

WP1 – Pilot studies:
- *[Executive Summary of the pilot studies](#)*
- *[Reports and other documents on 19 pilot studies](#) conducted to assess the added value of machine learning in coding and classification, edit and imputation and the use of imagery data*
- *Theme reports analysing the approaches and results from the pilot studies:*
  - *[WP1 - Theme 1 Coding and classification report](#)*
  - *[WP1 - Theme 2 Edit and imputation report](#)*
  - *[WP1 - Theme 3 Imagery analysis report](#)*

WP2 – Quality:
- *[A Quality Framework for Statistical Algorithms (QF4SA)](#)*

WP3 – Integration:
- *[Report](#) identifying and addressing the common challenges in integrating machine learning into production processes based on a survey conducted by the project team*

Other materials to assist users in the development of machine learning:
- *code used in some of the pilot studies - [available at Studies and Codes](#)*
- *two sets of data on which to learn and experiment with machine learning - [available at Learning and Training](#)*
- *links to learning and training material - [available at Learning and Training](#)*

The work packages investigated by the ML2020 project and potential themes for ML2021 project are presented in the table annexed to this report.

The Bureau is invited to review the report and advise on its further distribution (for example, submitting to CES for information), as well as on potential themes for the project's continuation in 2021.

I. Introduction

1.   The Machine Learning (ML) project was launched by the UNECE High-Level Group for the Modernisation of Official Statistics in March 2019 and concluded its work in December 2020. During this period, over 120 participants from 23 countries, 33 national organisations and 4 international organisations got together to work and collaborate on advancing the use of ML in the production of official statistics. They did so by demonstrating the added value of machine learning in the production of official statistics, developing a quality framework to guide its further development and identify and address challenges in integrating ML solutions in production processes. Reports, documents, code, data and numerous references were released on the public UNECE Statistics Wiki on November 13, 2020 for the benefit of the official statistics community. This release was followed by a webinar held on 16 and 17 November 2020.

2.   Based on the knowledge, experience and insights gained during the project, it is clear that **machine learning** is not just a buzzword. The studies conducted demonstrate that it **can be integrated into coding and classification operations to produce better quality results at the same or lower cost**. It produces some **positive results for edit and imputation in some contexts, but more studies are needed** in this area. **It is essential to exploit big data**, such as the analysis of satellite or aerial images. Its success highly depends on combining the knowledge and efforts of experts in varied disciplines, notably to produce and maintain a sufficient quantity of high quality data to train the algorithms and monitor the performance of ML assisted operations in an efficient manner.

3.   Although the value added of ML is demonstrated by the pilot studies and other recent developments, their **integration into production processes remains a challenge**. The project proposes **a quality framework for statistical algorithms** and addresses **other integration challenges to facilitate its development and acceptance of ML in organisations**.

4.   This report provides background for launching the project and describes how it was conducted. After listing its main outputs, **the report shares the main lessons learned on accepting and facilitating the advancement of machine learning, as well as highlighting related project outputs and suggesting future work**.

II. Modernisation of statistical organisations

5.   National statistical organisations (NSOs) are being challenged to be more responsive to the increasing need for more relevant, timely, detailed and accessible statistical information and data services that can be trusted and used to make clear and effective data-driven policy decisions. NSOs are also being pressured to be highly efficient and meet these expectations within existing budget levels. They are also challenged by the ever-increasing amount of data available in a wide variety of sources, formats and level of quality.

6.   Finally, NSOs must compete with an increasing number of public and private organisations, big and small, who produce and promote statistics in a more timely and accessible manner that attracts the attention of policy makers and many other users, despite relevance or quality deficiencies, at times. These organisations can produce these statistics for several reasons including the development or quick access to alternative data methods and their adoption in the production process, such as machine learning algorithms; greater IT capacity; and by imposing fewer constraints on quality, transparency, ethics and privacy.

7.   At the same time, these are the areas where NSOs hold a competitive advantage. By publishing details on data sources, methods and various indicators, NSOs demonstrate a high-level of transparency. They have considerable collective expertise in efficiently integrating diverse sources of data and a legal obligation to respect privacy and protect against disclosure. Furthermore, they have the opportunity to further augment this capacity through networks of professionals around the world who collaborate on common needs and priorities.

8.   In addition to counting on their individual and collective expertise, statistical organisations must have an adaptive culture to remain relevant by responding in a timely and

continuously responsible manner to data needs of stakeholders, both in terms of statistical information and services. One of the bases allowing NSOs to fulfil their mandate is the development of sound methods and processes that are integrated in the production of official statistics. As mentioned above, this is challenged by the proliferation of data demands, data sources and data producers who are increasingly able to more easily link the various data sources to the demands.

III.   Proposal for a machine learning project

9.   These challenges are the focus of the UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS), a group of committed Chief Statisticians actively steering the modernisation of national statistical organisations. Their mission is to work collaboratively to identify trends, threats, and opportunities in modernising national statistical organisations. HLG-MOS provides a common platform for experts to develop solutions in a flexible and agile way.

10.   HLG-MOS is supported by an Executive Board responsible for the strategic management of on-going activities that include four standing working groups and two time-limited projects. One of the working groups, the Blue-Sky Thinking Network (BSTN), is forward-looking and the "ideas factory" for the statistical modernisation community. BSTN recognized that many new data producers exploit a wide variety of data sources with different approaches and methods than those traditionally used in statistical organisations. Many of the data sources require innovative approaches and methods, such as machine learning and artificial intelligence. At its November 2018 workshop, HLG-MOS further recognised the importance of bringing these technologies into the production of official statistics by supporting a proposal from BSTN to launch a Machine Learning Project. The BSTN position paper justified the project as follows:

> "The interest in the use of Machine Learning for official statistics is rapidly growing. For the processing of some secondary data sources (including administrative sources, big data and the Internet of Things) it seems essential to look into opportunities offered by modern ML techniques, while ML techniques might offer added value also for primary data, as illustrated in the ML position paper. Although ML seems promising, there is only limited experience with concrete applications in the UNECE statistical community, and some issues relating to quality and transparency of results obtained from ML still have to be solved."

IV.   About the Machine Learning Project

11.   UNECE hired a project manager in late February 2019 and launched the project in March with 11 participants from 6 organisations at the time. Virtual meetings were held to scope and plan the project. The plans were finalised at a first face-to-face meeting hosted by the UK Office for National Statistics in May 2019. At that time, the project had already grown to 27 participants from 14 organisations. The participants agreed on the following main goals:

Based on mutual interest and building on existing national developments, the goal of the project is to advance the research, development and application of machine learning techniques to add value to the production of official statistics. To achieve this goal the Machine Learning project team will aim to:

- *Investigate and demonstrate the value added of ML in the production of official statistics, where "value added" is measured as an increase in relevance, better overall quality or reduction in costs.*

- *Advance the capability of ML to add value to the production of official statistics.*

- *Advance the capability of national statistical organisations to use ML in the production of official statistics.*

- *Enhance collaboration between statistical organisations in the development and application of ML.*

12.   Based on the first work of BSTN, the expectations of HLG-MOS and the interests of the team members, the work of the project was organised around three work packages (WP):

- WP1 - Conduct pilot studies on the following themes:

  - coding and classification (C&C)

  - edit and imputation (E&I)

  - use of imagery.

- WP2 - Develop a quality framework to underpin the use of ML

- WP3 - Identify and address integration challenges.

13.   In addition to demonstrating the value added of ML, the purpose of the pilot studies was to facilitate learning, sharing and collaboration within the group. At the beginning of the project, the lack of a quality framework to guide the use of ML was considered as the main hurdle to having proven ML solutions accepted within NSOs. Finally, other challenges related to the operationalisation and further expansion of ML solutions were initially identified and more were raised as the pilot studies progressed.

14.   Combining the project goals and work packages in a single sentence, the work of the project began to: "Integrate in production processes (WP3) demonstrated ML solutions (WP1) in a sound and efficient manner (WP2)". Participants from Canada, Germany, Mexico and the UK volunteered to lead five sub-groups (3 on WP1; WP2; WP3). The sub-groups met virtually to collaborate and share their progress and results. Project update meetings were held monthly.

15.   As the project progressed, more participants joined. In addition, other people working on the pilot studies (collaborators) and people interested in following the progress of the project (followers) also joined. At the time of writing this report, the project had 124 members from 23 countries, 33 national organisations and 4 international organisations. As a result, the monthly meetings were regularly attended by 40 to 60 persons. While the project focused on machine learning, it benefited from the large membership that brought people involved in other international groups working on topics that included issues or interests on machine learning, e.g., the UN Global Group on Big Data and the UNECE HLG-MOS Group on Capabilities and Communication.

16.   To further facilitate sharing and collaboration, four project meetings were held:

- May 2019 hosted by UK ONS to finalize the project goals, organisation and initial deliverables

- September 2019 hosted by the Statistical Office of the Republic of Serbia to share progress, further collaboration and agree on project deliverables and timelines

- April 2020 held virtually due to the pandemic (originally to be hosted by Statistics Poland) to share and discuss pilot study results

- October 2020 held virtually to share and discuss pilot study reports, pilot study theme reports, project outputs and future directions. This meeting also invited project members to introduce other uses of machine learning in their organisations.

17.   All project documents were shared among project members on the UNECE Statistics Wiki. They included working documents, monthly meeting updates and action items and many references shared by project members. The documents also included talks on the project and its results delivered by members at organised events, e.g., BigSurv20, Geo Week 2019, as well as the reports tabled at the annual HLG-MOS Workshop on Modernisation. Finally, the members used different means to share ML code (mostly on GitHub) and some data to facilitate and accelerate learning and experimentation by others.

18. Many of the documents and other elements mentioned above were released on the UNECE public statistics wiki on 13 November 2020. This was followed by a webinar held on 16-17 November 2020. The webinar showcased the project outputs and concluded with an open discussion on future directions. The webinar was attended by 203 people from 33 countries and 60 national and international organisations.

19. **The final output from the ML project is a large group of experts from various disciplines and organisations engaged in advancing the use of ML who are willing to pursue this advancement in a collaborative manner. The UK Office for National Statistics' Data Science Campus has come forward to lead a ML 2021 group with support from the UNECE Secretariat and the HLG-MOS Executive Board.**

IV. Project outputs, lessons learned and other thoughts

20. The three work packages produced 21 pilot studies[1], a quality framework and a report identifying and addressing the common challenges in integrating ML into production processes. The many lessons from this work are shared at the UNECE's wiki (see: HLG-MOS Machine Learning Project). Results and lessons from each pilot study (see: pilot study reports) are analysed and summarised in reports on each of the three areas where they were conducted (see: Coding and Classification Report; Edit and Imputation Report; Imagery Analysis Report). The latter reports are further analysed and summarised to highlight the main results and lessons in the application of ML in statistical processes (see: WP1 executive summary). The developments from the work packages on quality and integration address key aspects needed to facilitate the advancement and acceptance of ML (see: Quality Framework for Statistical Algorithms; Integration report).

21. This report aims at sharing lessons learned and other thoughts from the whole project while avoiding repeating those from the other reports, as much as possible. The group agreed that:

- the project achieved its goals

- most participants gained significant knowledge and experience in the use of ML

- the project did advance the use of ML in the production of official statistics in most participating organisations

- yet there is still limited evidence of ML solutions that "made it" to the production processes

- there is even more interest and need to bring ML from being a good potential solution to an actual production solution.

22. **Based on the knowledge and experience gained during the project, the advancement of ML for the production of official statistics can be summarised in two words: acceptance and facilitation.** Most of the burden of making ML accepted within a statistical organisation lies on those who develop and operationalise the methods. Most of the burden on facilitating the development and operationalisation of ML lies on the organisation. Most importantly, both the acceptance and facilitation require the support of all employees.

23. The next sections elaborate on **key aspects to accept and facilitate ML solutions**. Each aspect is briefly discussed, linked to advances from the project and concluded with suggestions on future work for the ML2021 group. **Table 1 in the appendix lays out these key aspects in a journey from idea generation to production maintenance supported by foundational elements and organisational support. The table highlights what was covered by the ML project and some areas identified as potential themes to be explored by the ML 2021 group.**

---

[1] The 21 studies include the development of a generic pipeline for the use of ML in analyzing imagery data and a "ideas and hints" document on the use of ML for data editing. These outputs are not based on a particular application of ML.

VI. Key aspects to accepting ML solutions

A. Alignment with business needs

24.   Ultimately, ML solutions must be accepted by the people responsible for producing data (usually subject-matter analysts) and, more importantly, those who use the data. Like any approach or technology, ML is one of the means to an end. It should not be considered or adopted simply for what it is, but for what it can do to better address the business needs (increased relevance, detail, timeliness, accuracy, cost efficiency, etc.). The pilot studies generally focussed on improving timeliness and accuracy in three statistical processes (see: pilot study reports). Applications of ML to address other business needs in other processes abound (see: Other applications of Machine Learning for some examples).

25.   One of the characteristics that contributed to the delivery of the project was its hands-on approach. Early on the project pondered the idea of producing a "cookbook" to guide the use of ML. The project worked on developing some of the basic elements of a good recipe (frameworks and good practices) while many participants learned by experimenting with a mix of ingredients (pilot studies) to satisfy the nutritional needs of their respective organisations. Going forward, it is important that the developments in the group continue to be based on the needs of participating organisations (applications) and that they provide input to the more foundational aspects (see the "Supported by" row in Table 1 in appendix).

B. Guidance from a quality framework

26.   The ML solutions must contribute to good or better-quality results towards fulfilling business needs. To do this, one needs to define what "quality" is. Definitions of quality are provided in many widely accepted quality frameworks developed by national and international statistical organisations. The Quality Framework for Statistical Algorithms (QF4SA) provides a supplement to these frameworks and focuses on aspects that are more prominent to the acceptance of ML solutions (see: Quality Framework for Statistical Algorithms). QF4SA provides guidance on the choice of algorithms (including traditional algorithms) for the production process. It purposely uses the terminology statistical algorithm as it covers both traditional and modern methods typically used by official statisticians to strengthen the mutual comprehension between proponents of each types. There is no set formula to ascertain that results from ML solutions are good enough or better than alternatives. As with most quality frameworks, QF4SA proposes five dimensions that must be considered jointly. One may choose to place more emphasis on one or two dimensions, but none should be ignored.

27.   QF4SA was developed while the pilot studies were being conducted. They supplied some key input and inspiration to the framework but did not have enough time to formally experiment some of the practices recommended in the framework. One of the qualities of the project expressed by some participants was its hands-on approach. Going forward, it is recommended that such experimentations take place in pilot studies to provide valuable feedback towards improving and expanding the framework.

C. Demonstration of added value

28.   Most of the pilot studies focused on this aspect. On coding and classification, the studies demonstrated that ML can deliver better quality than a strictly manual operation. The common challenge faced by the pilot studies was a lack of a statistically sound baseline against which to compare ML results. Consequently, many studies start with the goal of replicating an existing operation, e.g., producing the same product classes as a manual classification operation, and focussing the added value on timeliness and, indirectly, cost. There are three severe issues with this goal. First, the accuracy of the existing (or competing) operation is often either not known or not supported by a sound assessment method. Second, ML can never fully replicate another operation. Studies conducted in the project show that

ML solutions can replicate, with a high level of confidence, between 40% to 85% of the results from an existing (manual or other automated) operation. However, third, and more importantly, the goal of ML should not be limited to replicating another operation, unless it can do so much more quickly and at a significantly lower cost. The goal should be on improving the operation by combining the respective strengths of each. In the context of a classification operation, this could mean using ML predictions to automatically assign a class (on the predictions known to be very accurate, e.g., over 98%); using the accurate-but-not-quite-enough predictions to aid coders; and, ignoring the not-good-enough predictions and relying on coders to classify the rest (often less common classes). Variants of this strategy are used in production (see: Workplace injury and illness; Industry and occupation; Standard Industrial Classification). An experiment on this strategy using ML code and data shared was also conducted (see: A user's experiences with the ML code and data shared; Shared code; Product description dataset).

29.   On edit and imputation, the studies showed results ranging from no added value (a simple imputation method did better than other options) to promising. There are no indications that ML methods can't work. They may require less programming and be quicker to implement than current methods. On the downside, creating and maintaining good training data for such algorithms is a challenge and, explaining what they produce and how they produce it, even if it is quicker or more accurate, can be very challenging to explain and, thus, make it accepted by stakeholders. More studies and foundational developments (such as Hints and ideas on data cleaning) are needed to guide the use ML in this area and determine the characteristics of a favourable context in which to apply it.

30.   From the beginning of the project, it was thought that ML is essential to exploit large volumes of data in an efficient manner. This was confirmed in the pilot studies on the analysis of imagery data (satellite and air photography). As access to large volumes of such data is increasing, one of the challenges is to provide users with information on the complex processes needed to correctly and efficiently exploit them, including when machine learning is to be called upon. To provide some of this information, the project proposes a generic pipeline to produce official statistics using satellite data and machine learning (see: Generic pipeline). The pipeline was used to describe two studies on satellite and aerial imagery (see: Aerial Image Address Use Classification and Integrating EO with Official Statistics using Machine Learning).

31.   Going forward, organisations are encouraged to continue advancing their current ML developments towards their operationalisation, and to do so while they continue to collaborate and share with others. The development could be broadened to other areas of interest to organisations (see: Other applications of Machine Learning for some examples), particularly in business needs that are labour intensive, stable over time and offer considerable data to train the algorithms. These developments should consider piloting some of the practices from QF4SA and integration documents and return valuable feedback based on their experiences.

D.   Robust performance over time

32.   The pilot studies focussed on assessing the added value of different ML algorithms and identifying the best model (algorithm and parameters) based on the data they had. As stated before, there are still many challenges in bringing a demonstrated ML solution into production. As importantly, the ML solution must, over time, not only continue to perform as well, but to perform even better as it "learns" more and adapts as the data being entered into the ML solution evolves. As the investigations in ML solutions advanced in the project, participants have been asking when to update or refresh the ML algorithms and/or its parameters, how frequently and how to proceed. Comparable questions are raised in other applications of ML.

33.   Only one application in the project has been in production long enough to have extensive experience in how to update the ML algorithms (see: Workplace injury and illness). The central element in putting in place and maintaining an efficient ML solution is the data used for training, not only at the start, when setting on the initial algorithm and its parameters,

but throughout the use of the ML solution. Another key element is data used for evaluation, sometimes referred to as the "gold standard". This data is needed to assess not only how the ML algorithm performs, but the entire operation, as it usually includes some clerical operations. It must be independent from the training data. These data are essential, but they usually come at a significant cost and must respect certain characteristics, e.g., collection of ground truth data, texts classified by subject-matter experts. The project has gained a good appreciation of the value and the characteristics of good training and evaluation data.

34.   Going forward, it is recommended to document and share this knowledge.

E.   Respect of ethical and legal consideration

35.   "Machine Learning has become more powerful over the past decade, sparking an expansion of new applications. Some of these applications fall within the social domain, in which models based on data profiles can have a significant impact on the life of individuals. To prevent unwanted discrimination in these models, different methods have been proposed within the field of algorithmic fairness." This text is taken from the abstract to a working paper from the Statistics Netherlands Center for Big Data Statistics (see: fair algorithms in context). This excerpt stresses the importance of ethical issues. They were raised in several discussions during the project, but not addressed specifically.

36.   Going forward, these issues could be addressed either by the ML group or in collaboration with other working groups looking at this issue on a broader context. In these developments, it will be important to distinguish the issues about the data sources from the methods to exploit them. It will also be important to focus on issues specific to official statistics that are mostly aggregates rather than direct outcomes on individuals, as often raised in discussions, such as being accepted for a loan or getting a medical diagnosis.

F.   Development on solid scientific grounds from many disciplines

37.   National and international official statistical organisations have always produced relevant and trusted information because they are based on sound methods and processes. When ML methods are developed and implemented on the same basic principles, they go a long way towards dealing with the aspects above and being accepted. The science needed to underpin the processes encompasses knowledge and skills in many disciplines: subject-matter, statistics, informatics, methodology, data science and operations. Compared to traditional methods, these disciplines need to work even more closely together from start (fleshing out an idea and connecting it with a business need) to finish (operationalisation). This is particularly the case for subject-matter knowledge, where ML is not just another solution that has to work for subject-matter business needs, but also a solution that particularly needs subject-matter knowledge to work. While the idea to use ML can come from a single individual (like in some of the pilot studies), the development of this idea needs to quickly involve other disciplines, notably subject matter, to correctly and efficiently advance. Borrowing a sentence from the quality aspect, one may count on one or two particular disciplines, but none should be left out. This is highlighted in an experiment with ML code and data shared by the project. It was conducted by someone with limited knowledge on ML, who, along the way, learned many lessons and made many mistakes (see: A user's experiences with the ML code and data shared).

38.   The ML project benefited from having experts from many disciplines. This allowed learning and sharing different perspectives and aspects to consider in developing, assessing and advancing ML solutions. Going forward, the ML group should seek more participants in data science, subject-matter and IT, depending on the group's plans and goals. Statistical organisations will continue to face the challenge to acquire, develop and organise the varied expertise needed to effectively and efficiently use ML towards their business needs. The acquisition and development (e.g., training) of expertise was the most pressing need expressed in a poll conducted during the webinar (see: ML Webinar; Poll results). This aspect is further discussed below on facilitating ML solutions.

VII. Key aspects to facilitating ML solutions

A. Combination of multi-disciplinary skills

39.   The production of official statistics has always and continues to combine the knowledge and expertise from many disciplines. This is still the case, and even more so, with the proliferation of data sources (big, medium or small), users wanting to exploit them and technologies enabling their use. While many of these skills are present in data science (a relatively new discipline), the breadth and depth of the skills needed in each of the disciplines cannot be found in a single or small number of individuals. Bringing together the required skills is one of the main challenges facing statistical organisations. This can be broken down into four sub-challenges: identification, acquisition, development and organisation. Some of them were addressed by the project (see: Integration). In doing so, many concrete actions by NSOs to facilitate and expand the use of ML were discovered. Many are very recent (see: Initiatives to accelerate the integration of machine learning solutions).

40.   These initiatives include setting up separate organisations dedicated to data science, laboratories and internal or external forums to exchange on data science and machine learning. The leaders who manage these entities likely connect and exchange on an informal basis. Going forward, it is recommended that they create a formal network to share challenges, practices, experiences and results. This network should focus on managerial aspects, e.g., corporate strategies, alignment with needs, culture change, communication. It would also closely interact with the ML 2021 group, e.g., suggest and prioritise projects. The network or the ML2021 project should also connect with other groups working on similar aspects, e.g., the UNECE HLG-MOS Capabilities and Communication Group working on, among other things, change management, organisational frameworks for collaboration and building competencies.

B. Computing infrastructure

41.   From the beginning, the ML project decided to focus on demonstrating added value, quality and integration. The project discussed this issue at times, but just barely scratched the surface. Going forward, this aspect should be considered among the ML 2021 projects, but this should be preceded by a scan to find and connect with any working groups or other developments already addressing this aspect.

C. Research and development

42.   The first key aspect to having ML solutions accepted mentioned above is to align them with business needs. Discussions within the project subgroup on integration did not come to a full consensus on this aspect. Some emphasized the importance of starting with a business need, moving to R&D, producing a prototype and then bringing in the other areas like IT. Others emphasized the importance of building ML experience first, through R&D, which in turn allows one to identify suitable business problems which might be solved by machine learning. Going forward, whichever path taken to advance the use of ML, it should be driven, if not by a specific business need (e.g., from a single statistical program), at least by a clear corporate-wide strategy to continuously increase its relevance by giving access to more information of better quality in a timelier manner and potentially at a lower cost.

D. Sharing and collaboration

43.   Within the project, members shared working documents, methodological and technical references, links to learning resources, and presentations at meetings and sprints. Access to expert resources and meeting facilities were often obtained through the supporting structure of HLG-MOS. Sharing of ML code and data greatly facilitated and accelerated learning and experimentation by others. Many of the documents and other elements

mentioned above were packaged and released on the public UNECE Statistics Wiki on 13 November so that anyone in the official statistics community can benefit from knowledge and materials accumulated within the project team over two years (see: pilot study reports; Shared code; Learning, references and shared data).

44.   Going forward, sharing and collaboration will not only be beneficial to the advancement of ML, but also to quickly avoid its application in areas or contexts where it isn't known to add value. Although virtual sprints and other sessions were successful given the circumstances, it will be important to continue to support the occasional face-to-face meetings, in the form of experts seminar or workshops, where more collaboration can take place and be reinforced at later exchanges at a distance. The supporting structure of HLG-MOS should continue to promote the development of ML by providing access to and sharing information with a vast network of interested participants.

E.   Senior management support

45.   The Machine Learning project would not have existed and be successful without the engagement of many people from numerous organisations and the support from Chief Statisticians through HLG-MOS. Going forward, it will be essential to continue to count on their support to pursue research and development within their respective organisation and with others in collaborative initiatives. In return, these initiatives must be accountable to working on statistical organisations' priorities. In the case of ML2021, it is recommended that it does so by continuing reporting to HLG-MOS through its Executive Board.

F.   Engagement from all employees

46.   New technologies, such as machine learning or artificial intelligence, have a significant impact on an organisation's culture. ML changes what an organisation and each individual employee can do and how to do it. All the studies conducted in the project used supervised learning methods that need essential input, notably from subject matter and clerical staff who are likely the most affected by the change. The studies conducted indicate that ML methods cannot totally replace the work of staff and should not be perceived as such, but rather a means of introducing or strengthening automated processes to achieve better results at the same or lower costs and allowing time for staff to do more added value work. As ML solutions are proven to add value in more operations, employees in all functions and at all levels of the organisation must be encouraged to consider ML as a potential solution to their business needs. They should also have access to experts or a center of expertise on ML to quickly determine if ML should be further considered.

VIII.   Conclusion – is machine learning a buzz, a must or bust?

47.   In its position paper written in December 2018, proposing the Machine Learning project, the Blue-Sky Thinking Network wrote: "Although ML seems promising there is only limited experience with concrete applications in the CES statistical community, and some issues relating to e.g., quality and transparency of results obtained from ML still have to be solved." At that time, one could have shortened this statement to the following question in the context of producing official statistics: "Is machine learning a buzz, must or bust?"

48.   Two years later, the work of the Machine Learning project leads it to conclude that: ML is not just a buzz; that it is a must where it can add value and it should not be used where it does not (i.e., avoid becoming a bust); and that it still has some challenges being accepted and facilitated by some.

49.   The interest and progress in the use of ML is clearly demonstrated by the constant growth in the membership of the project that now has 124 members from 23 countries, 33 national organisations and 4 international organisations. The strong engagement from many participants assisted by many others enabled the project to deliver and share many reports, documents, code, data, learning material and other references for the benefit of the official

statistics community. Beyond the project, the number of talks, organized sessions or whole dedicated events on machine learning and data science are rapidly increasing (see a list of sessions and talks at BigSurv20 - Big Data Meets Survey Science listed on Other applications of Machine Learning; the next Statistics Canada Symposium will be on data science[2]).

50.   Machine learning is a must where it has proven to contribute to producing data that is more relevant, better quality, timelier or cost efficient without any significant reduction to any of these dimensions. It is more likely to render better results in operations that are labour intensive, repetitive and stable, such as in coding and classification. One would say that this is the case when introducing any automation in such operations. In the case of ML, this automation could be carried out more quickly. ML is essential in many uses of large volumes of data. It is more challenging to use processes that have a higher degree of subjectivity, such as in edit and imputation.

51.   With all new and evolving technologies comes a certain degree of resistance from different parties. Some will challenge them with scientific arguments that will only make them stronger. Others will simply resist them like most changes. The former will be convinced as long as the ML solutions are developed on solid scientific grounds from the different disciplines that they need, and the development of those ML solutions is guided by a quality framework and ethical considerations. The latter can be dealt with through clear and strong senior management support. Sharing and collaboration within and between statistical organisations are also essential to advancing the use of ML based on lessons learned on where ML adds value, where it shows promise and where it should not be considered.

52.   In a year and a half, the Machine Learning project generated considerable momentum towards advancing the responsible use of ML in the production of official statistics. The project advanced the use of ML in participating organisations and has shared many lessons learned with the official statistics community. As importantly, it concluded with a strong group of experts from various fields willing to continue, an organisation stepping up to lead it and interest from other individuals in joining and supporting the group on its journey to move ML solution from potential ideas to operationalisation.

---

[2] A link to its announcement is forthcoming

Table 1
Work packages investigated by the ML 2020 project (hyperlinks) and potential themes to be explored by the ML 2021 project (red text)

	Moving from idea to valid solution (demonstration)		Moving from valid solution to production (Operationalisation)		Ensuring production robustness (Maintenance)		
The journey	**All WP1 pilot studies**		**Some WP1 pilot studies**		**Very few WP1 pilot studies**		
	Other applications of Machine Learning		**Some other applications of Machine Learning**		**Very few other applications of Machine Learning**		
	WP3 Integration (Q5 & Q6)		**WP3 Integration (Q5)**				
	Workstream 1: Support current studies towards production; welcome new studies in other processes (e.g. record linkage) and/or data sources (e.g. satellite data)						

	Quality (accuracy, timeliness, efficiency, explainability and reproducibility)	Good Training Data	Skills/Competences	Computing Infrastructure	Interoperability / Business Process	Ethics and Legal	Security
	WP2 Quality		**WP3 Integration (Q3 & Q4)**				
Supported by	*Workstream 2: Experiment with practices and methods on some dimensions of QF4SA (WP2);* *Workstream 3: Review and improve the Framework*	*Workstream 4: How to get good training data, how to keep it up to date, when to relearn a model, what does 'good' mean, how to measure that?*	*Workstream 5: What skills? How to learn? Where to find them?*	*To be defined*	*To be defined*	*Workstream 6: Ethics handbook, regulations , etc.*	*To be defined*

	Organisation		Sharing and Collaboration	
	WP3 Integration (Q1 & Q2)		**HLG-MOS Machine Learning Project**	
Facilitated by	**Initiatives to accelerate the integration of machine learning solutions**		**ML Studies and Codes**	
	Workstream 7: Create/maintain a network of data science unit leaders;		**Learning and Training**	
	Workstream 8: Beyond 2021: How can we better prepare for next 2-5 years? What technology and data sources can we expect? What skills will we need?		**HLG-MOS ML Project webinar**	