Work Package 3: Machine Learning Integration
Author: Alex Measure (U.S. Bureau of Labor Statistics)
Date: December 14th, 2020
Version: 0.4 Final

Introduction

The purpose of the HLG-MOS Machine Learning Project is to advance the use of machine learning in official statistics. To this end, much of the initial work focused on demonstrating value through pilot projects (see Work Package 1). Although the results of many of these projects show great promise, the path from pilot project to production system is far from trivial. This is supported by the fact that despite dozens of participants, only a few members of the ML project report using machine learning in production currently.

One challenge is demonstrating the methodological suitability of these techniques. This is the focus of Work Package 2 (Quality Framework). The purpose of WP3 is to identify and address the remaining challenges to integration and production deployment.

The WP3 team pursued two activities to further this goal, a short online questionnaire designed to get a high level overview of the key challenges and successes, and a deeper investigation into 6 key questions:
- Where should machine learning fit in a statistical organization?
- What should the pipeline of a machine learning project look like?
- What machine learning skills are needed and where are they needed?
- How can organizations efficiently acquire the machine learning skills they need?
- How can organizations demonstrate the value added of machine learning?
- How should statistical organizations identify the right problems for machine learning?

The results are presented in this report.

Online Questionnaire

Our online questionnaire was designed and administered using SurveyMonkey. All members of the HLG-MOS Machine Learning project were encouraged to participate and also to forward the survey to colleagues with relevant expertise. Between September 15th and October 15th of 2020, 28 responses were collected and form the basis of this report. The questionnaire remains available online however at https://www.surveymonkey.com/r/6G5VVFH and additional responses are welcome and may be incorporated into future products.

Our 28 respondents include representatives of national statistical organizations covering 14 countries and regions, all in either North America or Europe. Most report having a role of "Statistician / Data Scientist", followed by "Analyst / Subject Matter Expert" and "Manager / Policy Maker." Only one respondent reported a role of "Software Engineer / Information Technology Specialist". Most respondents also report belonging to large national statistical

organizations (54%) defined as those having more than 2000 employees, followed by 32% of respondents reporting the next largest grouping, between 500 and 2000 employees.

Which option best describes your role in your organization?
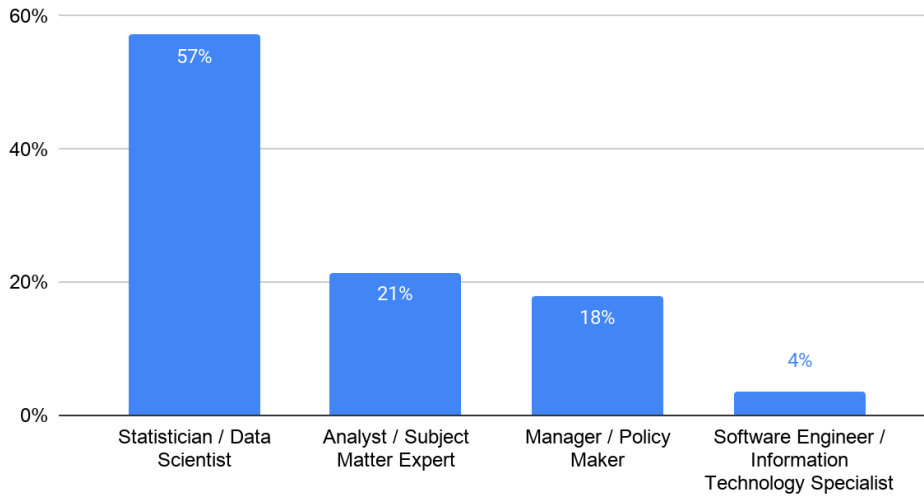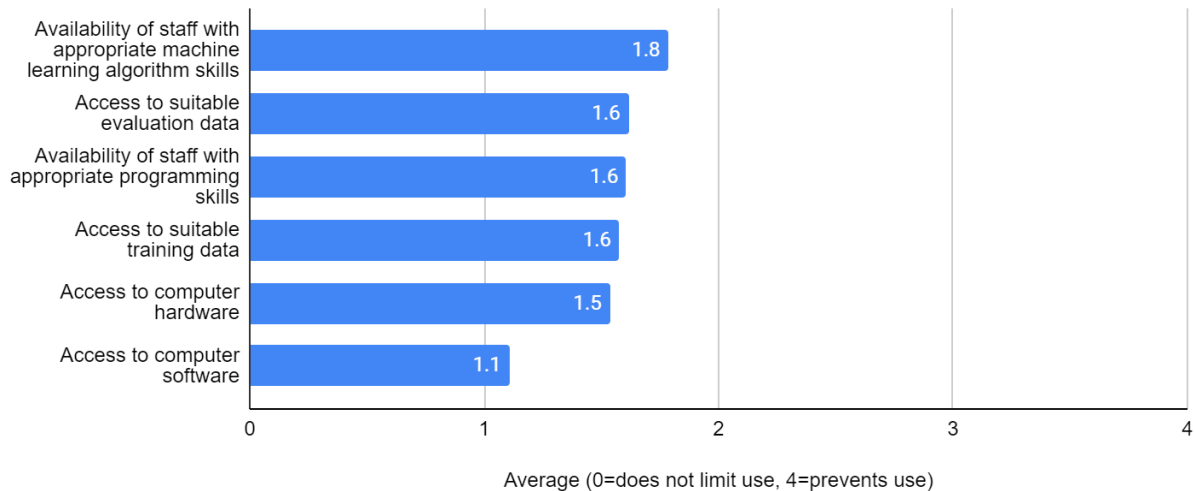

What are the biggest challenges facing statistical agencies in ML? Our questionnaire divides this into two sub questions, one asking about "organizational issues" and the other about "technical issues."

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

Average (0=does not limit use, 4=prevents use)
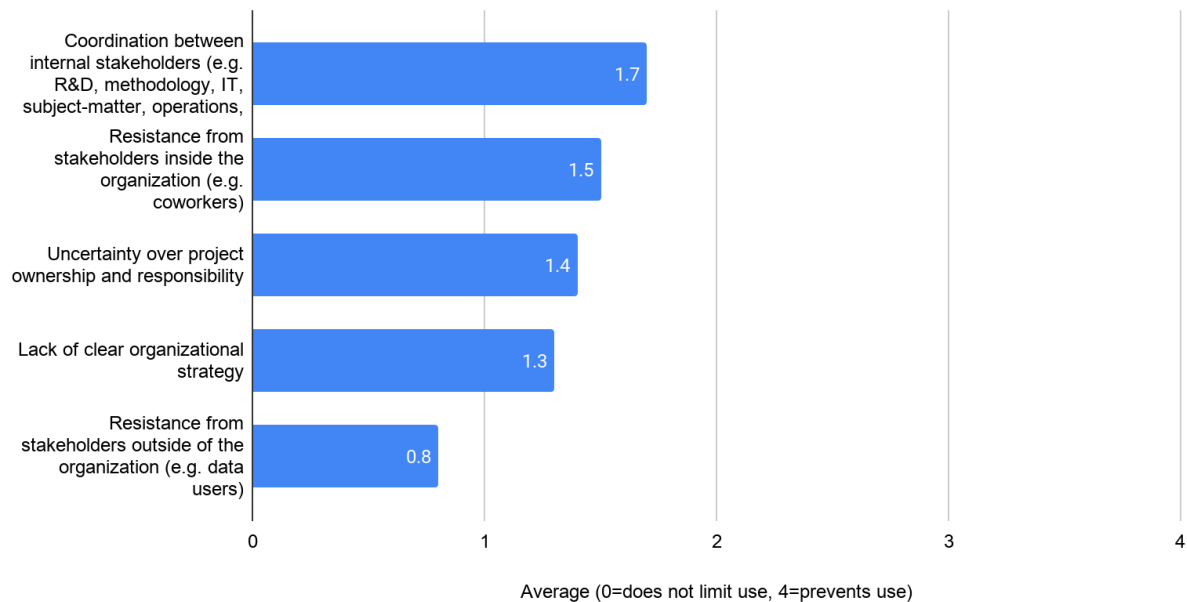
Among organizational issues, "coordination between internal stakeholders" ranked among the largest challenges with 16/27 (59%) reporting this moderately limits, severely limits, or prevents use.
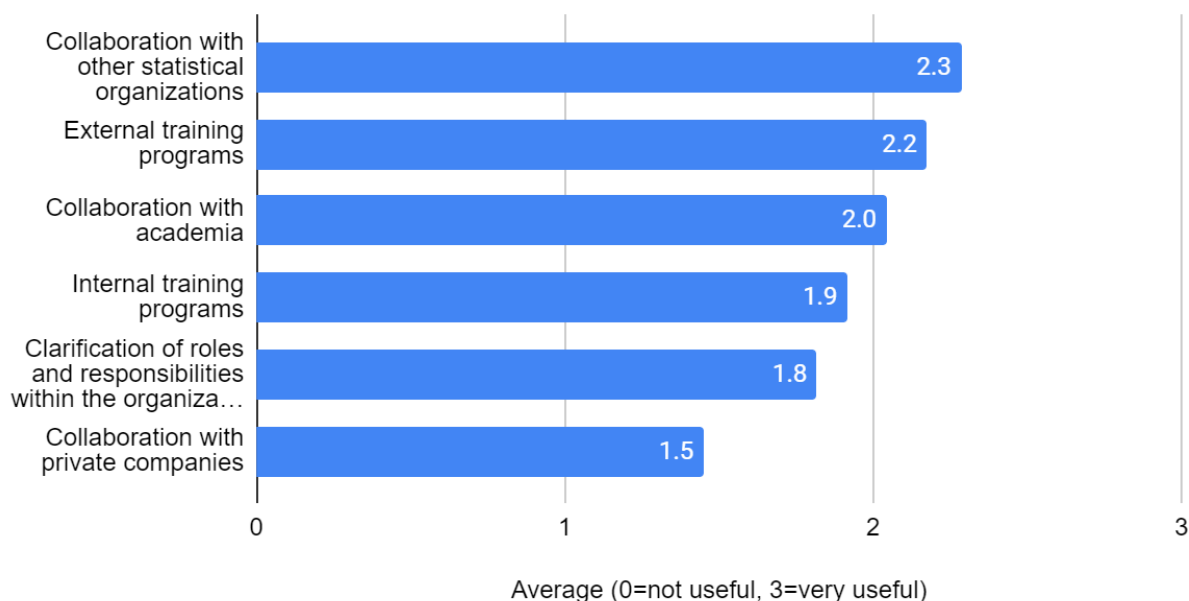
To what extent do the following technical issues limit your organization's ability to effectively use machine learning?

Category	Average
Availability of staff with appropriate machine learning algorithm skills	1.8
Access to suitable evaluation data	1.6
Availability of staff with appropriate programming skills	1.6
Access to suitable training data	1.6
Access to computer hardware	1.5
Access to computer software	1.1

Average (0=does not limit use, 4=prevents use)

Among technical issues, "availability of staff with appropriate machine learning algorithm skills" was the most limiting factor with 10/28 respondents (36%) reporting that it severely limits use. The average score of 1.8 makes this the most problematic issue identified in our survey.

Our survey ends with a question about which activities have been most useful. Among the options presented, "collaboration with other statistical organizations" ranked as the most useful, with 14/28 respondents indicating it is "very useful", followed closely by external training programs with 10/28 indicating "very useful".

How useful have the following activities been in helping your organization more effectively use machine learning?

Activity	Average
Collaboration with other statistical organizations	2.3
External training programs	2.2
Collaboration with academia	2.0
Internal training programs	1.9
Clarification of roles and responsibilities within the organiza…	1.8
Collaboration with private companies	1.5

Average (0=not useful, 3=very useful)

See the appendix for additional details on the survey results.

Long form investigation

While the short questionnaire gives us a high level overview of challenges and potential solutions, it lacks detail. To compliment this information we asked project participants to describe how they addressed six key questions. We received detailed responses from 4 organizations, the UK Office of National Statistics (ONS), the Australian Bureau of Statistics (ABS), Statistics Flanders, and the U.S. Bureau of Labor Statistics (BLS), and related comments from many others. The questions, and a high level overview of the responses are below.

Where should machine learning fit in a statistical organization?

Participants indicated 4 broad approaches:
- **Machine learning as a branch of methodology** - In Statistics Flanders, machine learning is an experimental branch of methodology. Machine learning techniques are clearly related to traditional statistical techniques so methodology is a reasonable starting point, especially for organizations still determining whether they want to use ML. Several other NSO's reported similar models at least early in their investigation. It is of course not a complete solution to production deployment but not all organizations are at that stage yet.
- **Machine learning as a multidisciplinary collaboration** - The Australian Bureau of Statistics' approach emphasizes the importance of multidisciplinary collaboration. In this model different pieces of the organization play lead roles on different aspects of the

project. Methodology or research often develop initial prototypes which are then handed off or co-owned by information technology and subject matter experts. An advantage is that many different pieces of the organization are involved. A frequent challenge is coordination. For example, the tools preferred by researchers and methodologists, such as R and Python, are often quite different from those preferred by software engineers. Another challenge can be in getting alignment with the needs and interests of subject matter experts, who are often the most direct users of the technology and often must also assume key roles in creating training and evaluation data.

- **Machine learning as a decentralized process** - Although the Bureau of Labor Statistics generally follows the multidisciplinary approach, in the case of machine learning it has instead adopted a largely decentralized approach in which the program offices assume primary ownership of machine learning systems and consult with methodologists to verify the integrity of the system, IT to integrate the system with existing infrastructure, and field staff to facilitate data collection and processing activities as needed. This reduces the difficulty of aligning different divisions, but at the cost of the program office assuming a more active role in methodology, systems development and maintenance.
- **Centers of excellence** - For the Office of National Statistics, a key aspect of machine learning strategy is the Data Science Campus, a separate division made up of experts in data science and machine learning which provides advice on machine learning projects not just to ONS, but to many parts of the UK government and even other countries. This allows the sharing of often limited machine learning expertise across many areas. A number of NSO's have recently developed their own versions of this approach, including INEGI (Mexico), Stats Canada, Statistics Finland, and Statistics Sweden. One variation of this approach is the "hub and spoke" model, in which limited machine learning expertise is initially concentrated in the hub (the center of excellence) with the goal of ultimately transferring much of it to the spokes (the specific business areas).

What should the machine learning pipeline look like in regards to organizational structure? Where should projects start, who should control what aspects when?

Interestingly, the responses to this question resulted in two seemingly opposite ideas. One emphasized the importance of starting with a business need, moving to R&D, producing a prototype and then bringing in other areas like IT. The other emphasized the importance of building ML experience first, which in turn allows one to identify suitable business problems which might be solved by machine learning.

In retrospect, it is clear that both are needed. An organization cannot determine whether machine learning is suitable if it knows nothing about machine learning, but it is also clear that the ultimate goal is to serve business needs.

What machine learning skills are needed and where are they needed in the organization?

On this question, there was general agreement among the responses. In organizations that distribute machine learning responsibilities across many divisions, machine learning requires new skills in many areas. Specifically:
- Everyone must understand the basics, such as the key ideas and common terminology. This allows effective communication between parties.
- Research and methodology often must become familiar with new algorithms and new tools, like R and Python, which are popular for machine learning.
- Information technology must learn how to integrate these tools and processes in existing systems. In some cases they must also support new hardware needs, such as powerful Graphical Processing Units for training deep neural networks.
- Subject matter and clerical workers must understand their role in supporting, using, and maintaining these systems as they often play a lead role in creating the training and evaluation data.
- Senior management must understand the needs of ML teams, including the need for careful alignment and coordination across these activities.

Because of the difficulty of coordinating broadly distributed activities, another increasingly popular approach is to rely on positions and operational units that increasingly blur the distinctions between research, methodology, information technology, and subject matter. See, for example, Google's Hybrid Approach to Research, and Data Scientist: The Sexiest Job of the 21st Century. In some organizations, a data scientist spends some of their time researching and evaluating different machine learning solutions to a problem (R&D, methodology), some of it building and running the model in production (IT), and some of it assisting with use and maintenance (subject matter). This blurring of boundaries reduces the extent to which machine learning skills need to be distributed across the organization, but requires individuals and teams with a broad range of skills and the organizational and IT infrastructure necessary to make it work.

How can organizations efficiently acquire the ML skills they need?

Responses identified 4 strategies:
1. **Acquire and train internally** - In this strategy, an outside expert is hired permanently or temporarily and used to train additional experts internally. Statistics Flanders, ONS, and ABS all report using some variant of this approach.
2. **External training** - In the case of machine learning, many high quality trainings are available (often for free), and many NSO's report using these extensively. There are also increasingly suitable trainings available through academia.
3. **Communities of practice** - A community of practice is a group of individuals with a shared interest and willingness to share what they know. The HLG-MOS ML project is partly a community of practice, but many NSO's also have internal communities. The BLS, for example, has a popular data science user's group that frequently features machine learning work.
4. **Research projects** - At some point learning requires doing. Research projects play an important role in supporting skill acquisition.

How should organizations demonstrate and communicate the value-added of ML techniques?

One of the recurring challenges of working on projects involving many parties is the need to convince others to adopt or support new techniques. This is supported both by numerous anecdotes among participants in the ML group, and by questionnaire responses indicating coordination and resistance issues from internal stakeholders. Responses identified 3 potential strategies.

1. **Clearly demonstrate value added** - When replacing or augmenting an existing process, it is often easy to demonstrate speed and cost improvements with machine learning but quality is also an important consideration and frequently much harder to evaluate. In many cases the most readily available evaluation data for a machine learning project is just a subset of the data currently produced by the existing process. In this case, standard quality metrics (accuracy, mean squared error, etc.) only measure how closely the machine learning approach matches the existing process, not the more relevant question of whether one is better or worse.

   One solution is to construct the evaluation data in such a way that it is independent of all processes being evaluated. This can be accomplished, for example, by asking a trusted panel of experts to reprocess the evaluation data without knowledge of how either the machine learning or existing processes would handle it. The resulting "gold standard" can then be used to evaluate and directly compare both the existing process and the machine learning process. In the case of the BLS injury and illness coder, this comparison played a critical role in justifying the use of the machine learning option.

2. **Use ML as a decision-support, at least initially** - Replacing an existing process with something new is also a potentially dangerous task. There is always the potential for some unanticipated issue to occur, and this is especially concerning to stakeholders who might have little familiarity with machine learning. One solution is to instead use machine learning as an assistive tool, at least initially. If we are automating an occupation classification task which was previously done manually, for example, we might start by only using machine learning to provide suggestions to a human coder. This allows stakeholders to get hands-on experience working with the machine learning model in a low-risk setting.

3. **Use ML for things that aren't otherwise possible** - Another way to introduce machine learning is to use it for new projects where no other option is feasible. Analysis of satellite imagery is a good example, it simply is not possible to do this at scale and frequency without prohibitive amounts of labor. Here, machine learning can make an otherwise impossible task possible.

How should statistical organizations identify the right problems for machine learning?

Our investigation uncovered 3 strategies.

1. **Learn from others**. Learning from the successes and failures of others working on machine learning is a relatively cheap and easy way to identify promising areas and

avoid less promising ones. By organizing and promoting the sharing of this information, the HLG-MOS ML project greatly facilitates this.

2. **Look for tasks that meet machine learning friendly criteria**. Machine learning tends to be well suited for tasks that have certain characteristics. These often include the following:
   a. Stable over time, i.e. the task is largely the same task year to year. This is important because machine learning learns from previously processed data and when things change, adjustments are often required. Processes that are more stable over time will thus require less frequent adjustments to continue operating correctly.
   b. Lots of training data showing all relevant inputs to a task and the desired outcomes. Ultimately machine learning requires data to learn. The more that's available and the better the quality, the more effective it tends to be.
3. **Start with lightweight research projects**. Pilot studies provide a relatively low cost and low risk way to explore and test initial ideas.

Appendix: SurveyMonkey results.

In which country or region does your organization operate?	
Country	**Responses**
Canada	6
United States	3
United Kingdom	2
Switzerland	2
Belgium	2
Netherlands	2
Sweden	2
Norway	2
Australia	1
Germany	1
Italy	1
Mexico	1
Europe	1

Which option best describes your role in your organization?

Option	Count	Percentage
Statistician / Data Scientist	16	57%
Analyst / Subject Matter Expert	6	21%
Manager / Policy Maker	5	18%
Software Engineer / Information Technology Specialist	1	4%

Approximately how many employees work in your statistical agency?

Option	Count	Percentage
More than 2000	15	54%
Between 500 and 2000	9	32%
Between 50 and 500	3	11%
Less than 50	1	4%

To what extent do the following **organizational** issues limit your organization's ability to effectively use machine learning?						
Option	Does not limit use (1)	Slightly limits use (2)	Moderately limits use (3)	Severely limits use (4)	Prevents use (5)	Average
Coordination between internal stakeholders (e.g. R&D, methodology, IT, subject-matter, operations, etc.)	6	5	7	8	1	**2.7**
Resistance from stakeholders inside the organization (e.g. coworkers)	4	10	10	2	1	**2.5**
Uncertainty over project ownership and responsibilities	8	5	9	3	1	**2.4**
Lack of clear organizational strategy	9	8	4	5	1	**2.3**
Resistance from stakeholders outside of the organization (e.g. data users)	12	6	2	1	1	**1.8**

To what extent do the following **technical** issues limit your organization's ability to effectively use machine learning?						
	Does not limit use (1)	Slightly limits use (2)	Moderately limits use (3)	Severely limits use (4)	Prevents use (5)	Average
Availability of staff with appropriate machine learning algorithm skills	4	8	6	10	0	**2.8**
Access to suitable evaluation data	5	7	8	5	1	**2.6**
Availability of staff with appropriate programming skills	6	6	9	7	0	**2.6**
Access to suitable training data	5	7	9	4	1	**2.6**
Access to computer hardware	8	5	8	6	1	**2.5**
Access to computer software	10	8	7	3	0	**2.1**

How useful have the following activities been in helping your organization more effectively use machine learning?					
	Not useful (1)	Slightly useful (2)	Moderately useful (3)	Very useful (4)	Average
Collaboration with other statistical organizations	0	6	8	14	**3.3**
External training programs	1	4	8	10	**3.2**
Collaboration with academia	0	9	4	10	**3.0**
Internal training programs	1	9	5	9	**2.9**
Clarification of roles and responsibilities within the organization	1	6	11	4	**2.8**
Collaboration with private companies	4	7	5	4	**2.5**