# Research and Implementation in ML workshop

Eric Deeben, International Team Lead, ONS Data Science Campus

InKyung Choi, Statistician, UNECE

26 January 2022



Eric.Deeben@ons.gov.uk

# Housekeeping

## No Recording

Today's workshop is a closed session and will not be recorded; however, we will take a few photos of the session for publicity.

Please let us know if you do not want to be photographed.

## Interactive session

We encourage you to share your experiences and questions throughout the workshop and participate actively in the discussion. The session is an opportunity to discuss challenges and develop solutions that are relevant to your specific context.

## Feedback

At the end of the seminar, we will ask you to help us evaluate the session by telling us what you have learnt.

# Content

## Morning: Scoping a Business Case 10:45 - 12:00

- Machine learning - 10 min
- Understand business needs - 10 min
- Research workshop overview - 10 min

Round table - 30 min

*Lunch break 12:00-13:00*

## Afternoon: ML POC workshop 13:00 -14:30

- Feasibility of ML implementation - 10 min
- Built a Proof of Concept "POC" - 10 min
- Round table - 20 min
- Integration - 10 min
- Quality Framework for Statistical Algorithm "QF4SA" - 25 min
- Evaluation - 10 min

# Machine Learning

# Machine Learning

- Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with data, without being explicitly programmed

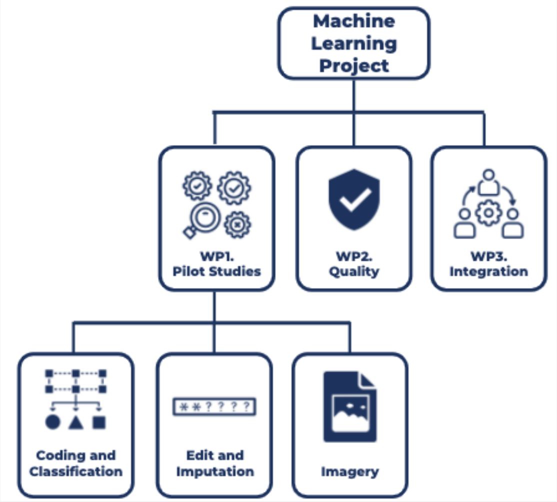
- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Source: IBM

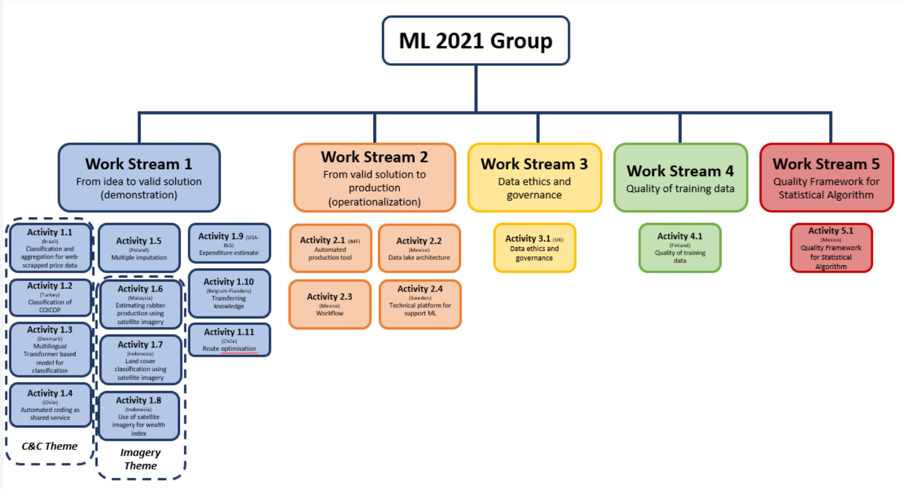
- Machine learning has the advantage to assess Big Data sets

# Is ML for official statistics....?

UNECE HLG-MOS
Machine Learning Project
(2019-20)




UK ONS-UNECE
Machine Learning Group
(2021)




**250+ members around 33 countries**

**38 pilot studies & researches**

**Coming up**
UK ONS-UNECE
Machine Learning Group **2022**

# Join the ONS-UNECE ML Group

- Global Data Squad - opportunity to engage in research international statistical organisations

- Capability building - Coffee and Coding sessions, expert presentations

- A hub for ML news and knowledge sharing

- Connect and network with statisticians and data scientists working on ML in official statistics from around the world

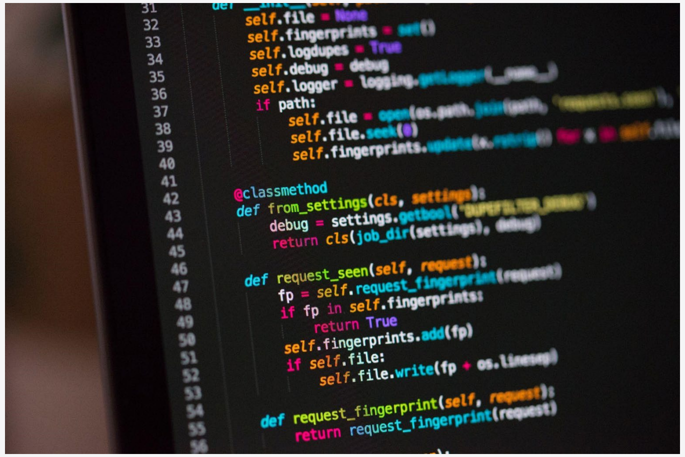



Photo by [Chris Ried](#) on [Unsplash](#)

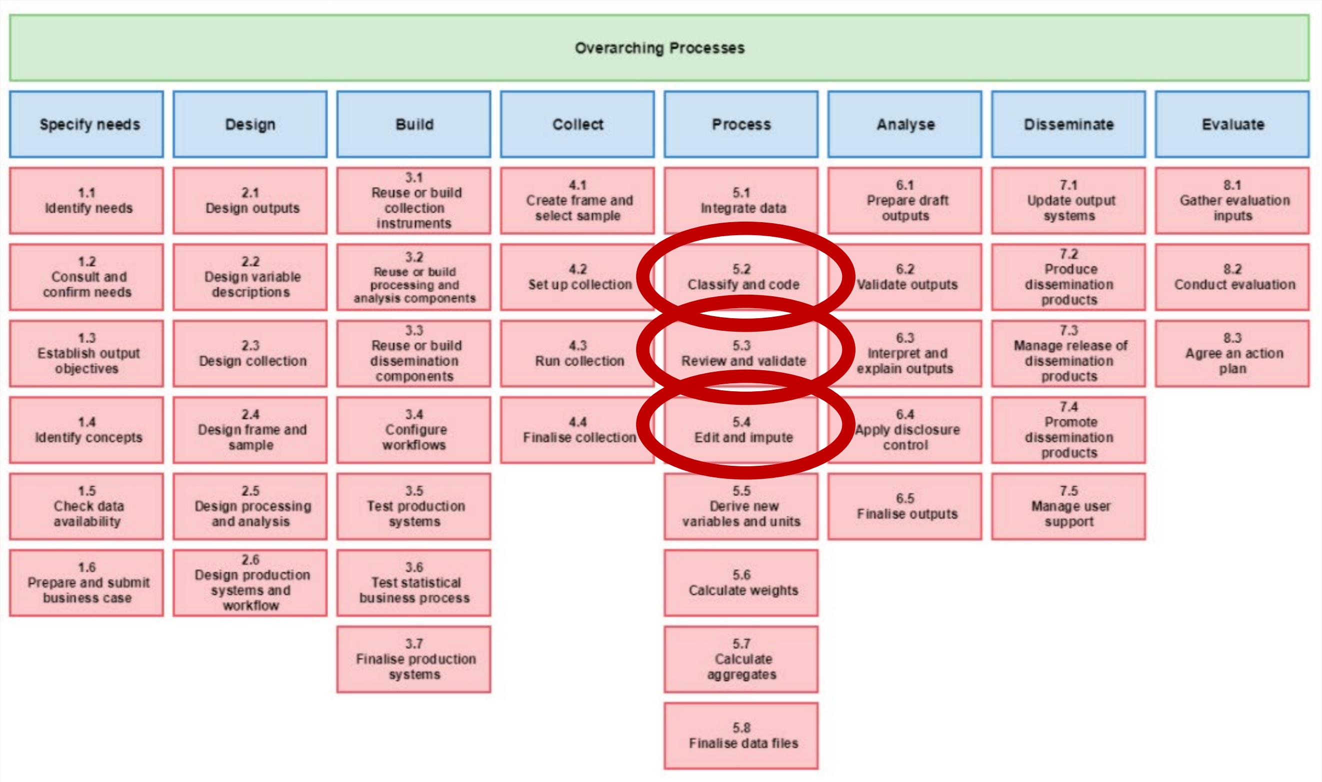
# Understand business needs

# Business needs: What is your value proposition?

- Statistics in a changing world
- Changes in technology and demographics are impacting traditional data sourcing methods
- Machine learning and opportunities for operational efficiency

# Statistical production process

Inputs




Outputs

Areas with manual, repetitive work can be automated with the help of machine learning

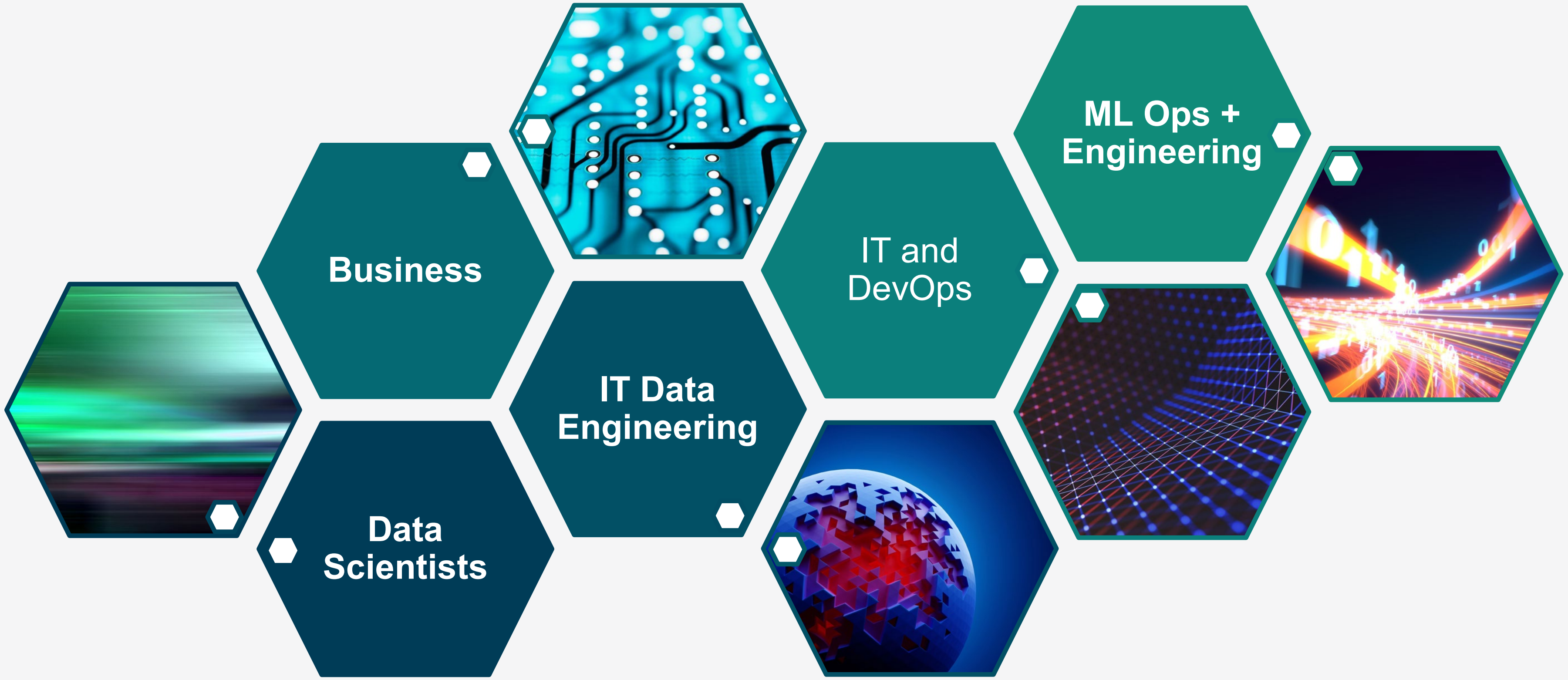
# Research workshop overview

# Workshop objective guidance

Based on mutual interest and building on existing national developments, the objective of the workshop is to **advance the research, development and application of machine learning techniques to add value to the production of official statistics**.

- Investigate and demonstrate the value added of ML in the production of official statistics, where "value added" is 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.

# The ML project team



# Areas to explore at your table

1. What process / task do you want to assist with machine learning?
2. What is the cost of the current way in terms of, for example, time, number of staff involved or budget?
3. What data sets are involved in the process (e.g., survey micro-level data, web-scrapped data)?
4. Who is the business owner?
5. Who are the stakeholders involved?
6. How do you expect machine learning can contribute in terms of quality?
7. Does ML help your organisation, institute, Government to produce better, more accurate, faster statistics?

# Lunch break

We will break for lunch until 13:00

After lunch each group will be invited to
share the main points of their discussion

# Feasibility of ML implementation

# ML Implementation – key building blocks

Organisational experience

Data access

Tools and training

Infrastructure

# How to start ML implementation #1

Organisation experience:

1. Do you know any other team in your organisation or other statistical organisation that has done a similar work? If not, do you know where you can consult?

Data access:

1. Do you have access to the data set ? Is there any security issue that could prevent you from accessing the data? If so, do you have any work-around (e.g., synthetic data, publicly available data that has similar nature)?
2. Is data supply sustainable?

# How to start ML implementation #2

Tools and training:

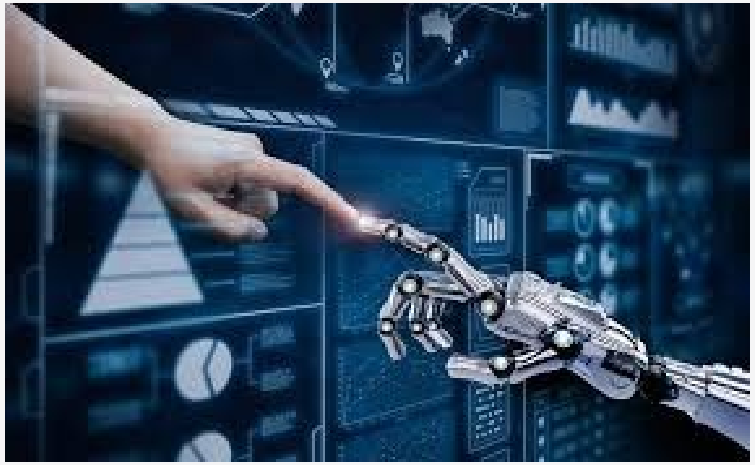
1. Do you (or your team) have the capability to run the machine learning project? If not, how do you plan to acquire it (e.g., training staff, borrowing staff time from other teams, collaboration with university, outsource to private company)?
2. What software (e.g., R, python) do you need? Is it currently supported by corporate IT system? If not, how do you plan to connect the machine learning solution, once it is production-ready, with corporate system (e.g., re-work machine learning codes in the software language supported by corporate, build APIs)?
3. Do have a trained team, and to what level?

Infrastructure:

1. Would you need a large computational resource? If so, how do you plan to obtain it (e.g., use cloud service, purchase GPU)?

# Proof of Concept

# Human in the loop



- Exploring how survey stakeholders will interact with ML
- Where are the humans in the process?
- Applying a human-centred lens
- Business processes don’t exist in a vacuum and often they don’t run themselves.
- Successful machine learning projects serve the needs of the ‘humans in the loop’. Identifying these humans and understanding their needs is critical to success.

# Proof of Concept- guidance areas

- Organisation
- Project Stage
- Context
- Problem
- Value added
- ML Solutions
- Challenges
- Expectations
- Contributions

| Organisation | Project | Context | Problem | Value added | ML Solutions | Challenges | Expectations | Contributions |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| UK | Living cost and Food (LCF) survey, Edit and Imputation process | Input data: Survey data from the LCF survey, these are highly personal data about respondents life situation, income, spending, debt, benefits and any other aspect of their financial situation<br><br>Output: I envisage that that the ML process can carry out the editing | The ML solution has to be able to duplicate the currently manual editing and imputation process to a very high degree of accuracy to speed up the process and save costs. | It is mainly costs. The LCF survey will be merged with another two surveys, no manual editing is carried out for these two surveys. The new survey HFS will need to have an editing and imputation process that can handle the much larger combined data sets. | Don't know yet | Building a training set for this ML solution will be a challenge as the data models keep changing on a monthly basis with considerable changes on a quarterly basis.<br><br>Issues like computing power, cloud based computing with it's legal and ethical issues | To learn which ML approach has to be taken for this POC.<br><br>I hope to find out if ML can substitute or at least assist in the editing process and that it can be established what the | We do not have a software solution yet and we can not share data as they are highly personal.<br><br>But we can share our experience of the currently manual editing & imputation process and any code that comes from this POC |

| Organisation | Project | Context | Problem | Value added | ML Solutions | Challenges | Expectations | Contributions |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| xxxx | Use of deep learning to learn neighborhood characteristics from Google Street View images | Input data:<br><br>Satellite data/aerial images/Google Street View<br>Neighborhood characteristics from survey and admin data<br>Output:<br><br>A (deep) neural network trained to predict neighborhood characteristics from images<br>Stage:<br><br>Collecting the necessary input data | ML solution:<br><br>Deep learning<br>Problems:<br><br>Low response burden<br>High survey costs<br>Long time between data collection and production of official statistics<br>Low frequency<br>Opportunities:<br><br>Estimate neighborhood characteristics without (or with | Value added:<br><br>ML is better equipped to handle image data than model-based techniques. (Design-based techniques not applicable.)<br>Improvement:<br><br>See B. importance<br>Assessment:<br><br>Cross validation: train on subset of annotated data, predict left-out sample and compare with true values<br>Compare model-based | Deep learning is an artifical neural network with several hidden layers. It has been shown to be able to classify images very well, provided a large training set is available. | Obtaining training data: high resolution images annotated with neighborhood characteristics<br>Optimizing network architecture and hyperparameter values<br>IT infrastructure | Learn from other organizations how to solve this problem. | We have register data about neighborhoods and aerial photo's but as far as I know no images (yet) from Google Street View.<br>Some colleagues have experience with training a deep neural net. |

# Machine learning - Integration

Data Science Campus

# Machine learning - Integration

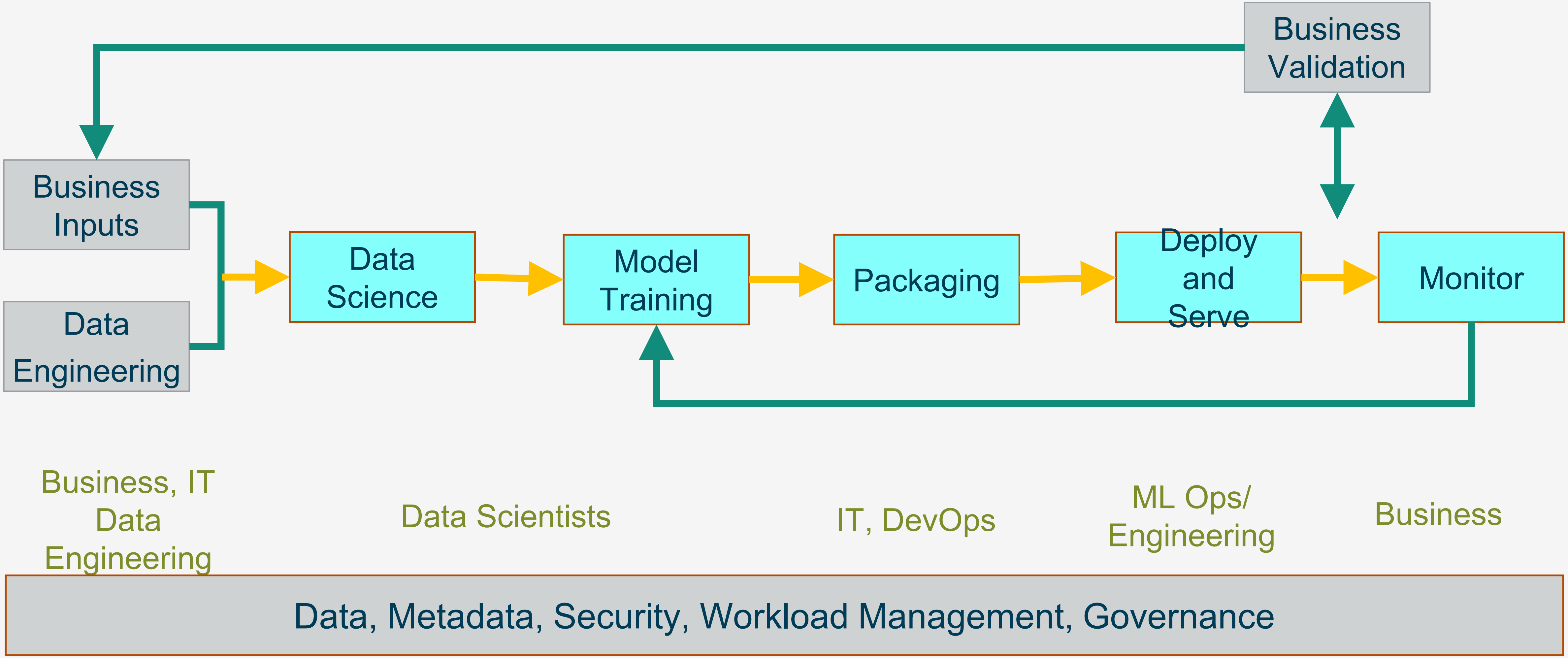
Making ML count : the challenge of integration

*“Successful progression into production only happens for about 20% of PoCs”*

1. Where should ML fit in a statistical organization?
2. What should the ML pipeline look like with regards to the organization structure?
3. What machine learning skills are needed and where are they needed in the organization?
4. How can organizations efficiently acquire the ML skills they need?
5. How should organizations demonstrate and communicate the value-added of ML techniques?
6. How should statistical organizations identify the right problems for ML?

# End-to-end Machine Learning in an Enterprise setting

## Production ML is an iterative team sport






Eric.Deeben@ONS.gov.uk

# What are the Challenges

1. Business Knowledge
   - Stakeholders - Who can make decisions?
   - Data usage - Impact on data users
   - Existing Data Pipeline - Integration
2. Knowing the Data
   - Precision □ □ Recall : to satisfy business needs
3. Value Proposition, what is the ML value?
   - Speed?
   - Cost?
   - Accuracy?

What is the driver for all this?

# Quality Framework for Statistical Algorithm (QF4SA)

InKyung Choi
UN Economic Commission for Europe
Statistics Division

EXPO2020

Mobilizing Big Data and Data Science for the Sustainable Development Goals Event

(January 26, 2022)

UNECE

modernstats
BY HLG - MOS

# Content

1. Background
2. Quality dimensions in QF4SA
    1) Accuracy
    2) Explainability
    3) Reproducibility
    4) Timeliness
    5) Cost-effectiveness
3. Summary

# 1. Background

- National Statistical Offices (NSOs) are the provider of official statistics and have a responsibility to ensure that the highest quality outputs are produced
- Quality frameworks to support quality assurance
- With increasing interest in machine learning methods, existing quality frameworks need to be looked at

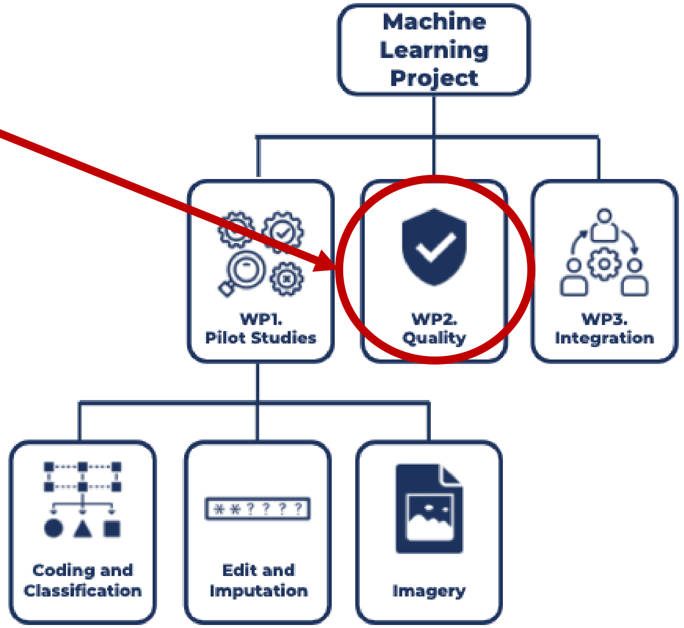
**United Nations National Quality Assurance Framework quality principles and supporting Fundamental Principles of Official Statistics**

| Quality principles | Fundamental Principles of Official Statistics | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **Level A: Managing the statistical system** | | | | | | | | | | |
| 1: Coordinating the national statistical system | | | | | | | | * | | |
| 2: Managing relationships with data users, data providers and other stakeholders | * | | | | * | | | O | | O |
| 3: Managing statistical standards | | | | | | | | | * | |
| **Level B: Managing the institutional environment** | | | | | | | | | | |
| 4: Assuring professional independence | O | * | | | | | O | | | |
| 5: Assuring impartiality and objectivity | * | O | O | O | O | | O | | | |
| 6: Assuring transparency | | | * | | | | O | | | |
| 7: Assuring statistical confidentiality and data security | | | | | | * | | | | |
| 8: Assuring commitment to quality | | * | | | | | | | | |
| 9: Assuring adequacy of resources | O | | | | | | | | | |
| **Level C: Managing statistical processes** | | | | | | | | | | |
| 10: Assuring methodological soundness | | * | | | O | | | | O | O |
| 11: Assuring cost-effectiveness | | | | | * | | | | O | |
| 12: Assuring appropriate statistical procedures | | * | | | O | | | | | |
| 13: Managing the respondent burden | | | | | * | | | | | |
| **Level D: Managing statistical outputs** | | | | | | | | | | |
| 14: Assuring relevance | * | | O | | O | | | | | |
| 15: Assuring accuracy and reliability | * | | | | O | | | | | |
| 16: Assuring timeliness and punctuality | * | | | | O | | | | | |
| 17: Assuring accessibility and clarity | * | | O | | | | | | | |
| 18: Assuring coherence and comparability | * | | O | | | | | | O | |
| 19: Managing metadata | | | * | | | | | | O | |

UN National Quality Assurance Framework

# 1. Background

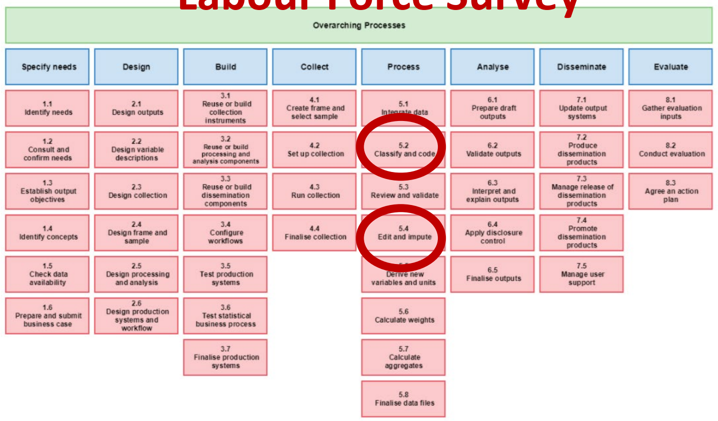
- Developed by UNECE HLG-MOS Machine Learning Project Work Package 2 – Quality Team in 2020
- Contributors: Siu-Ming Tam (Australia), Bart Buelens (Belgium-VITO) Wesley Yung (Canada), Gabriele Ascari and Fabiana Rocci (Italy), Florian Dumpert (Germany), Joep Burger (Netherlands), Hugh Chipman (Acadia University), InKyung Choi (UNECE)
- Will appear in the Statistical Journal of the IAOS (2022)



# 2. Quality dimensions in QF4SA

- Why “Statistical” Algorithm? -> Applicable to both traditional statistical methods as well as ML methods
- Targeted for intermediate outputs, not necessarily for the final statistical output

**Labour Force Survey**




**Final statistical output (e.g., employment rate for 2021 Q4)**

# 2.1. Quality dimensions - Accuracy

- Closeness of computations or estimates to the true values that were intended to measure
- Accuracy metric changes according to the process and to the target, when the focus is on unit wise predictive accuracy (often in ML application)

| ID | Job description | Actual code | Predicted code | Result |
|---|---|---|---|---|
| 1 | I manage crane | 8221 | 8221 | Correct |
| 2 | Fork-lift | 8222 | 8229 | Incorrect |
| 3 | I drive lift trucks | 8222 | 8222 | Correct |
| … | … | … | … | …. |
| 3500 | Plowing machine driver | 8223 | 4133 | Incorrect |

# 2.1. Quality dimensions - Accuracy

- Closeness of computations or estimates to the true values that were intended to measure
- Accuracy metric changes according to the process and to the target, when the focus is on unit wise predictive accuracy (often in ML application)
  - For regression: RMSE (absolute or relative), etc.
  - For classification: accuracy, recall, precision and F1 score

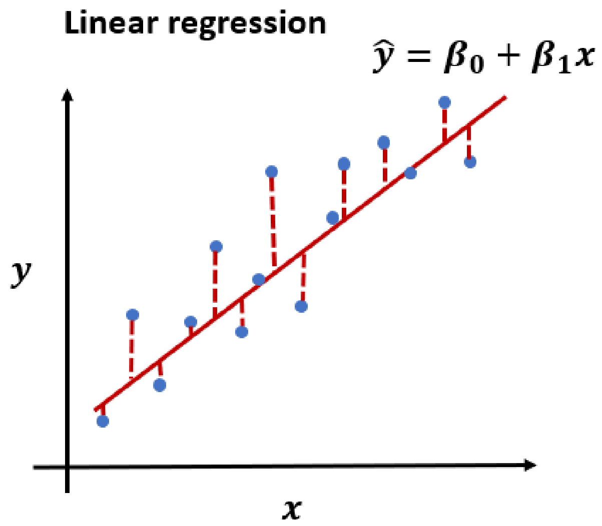
| | | **Predicted category** | |
|---|---|---|---|
| | | Positive | Negative |
| **Actual category** | Positive | True Positive (TP) | False Negative (FN) |
| | Negative | False Positive (FP) | True Negative (TN) |

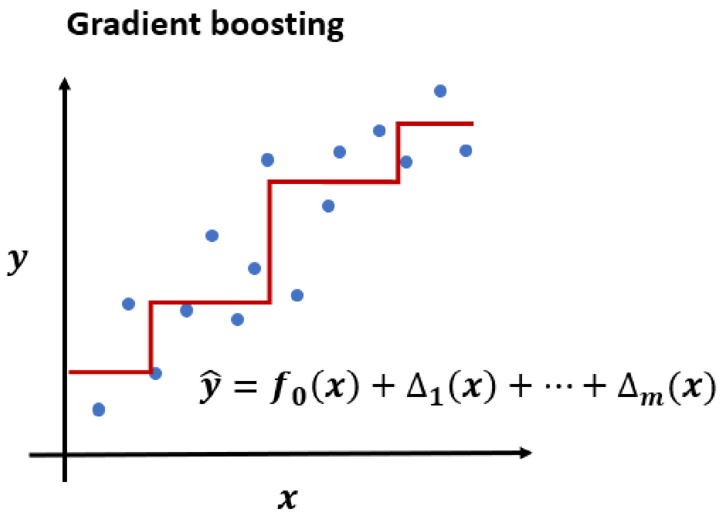
Precision = True positives among all predicted positives

Recall = True positives among all actual positives

# 2.1. Quality dimensions - Accuracy

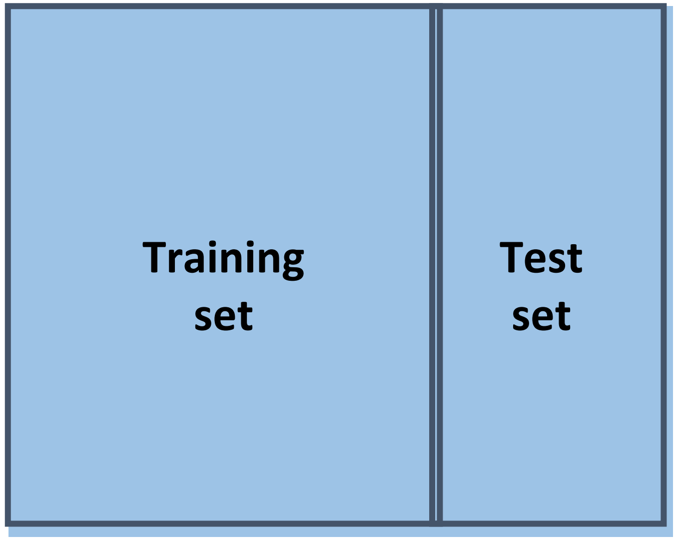
- ML methods often do not have as much restriction as traditional statistical methods
- There is a risk of overestimating accuracy



# 2.1. Quality dimensions - Accuracy

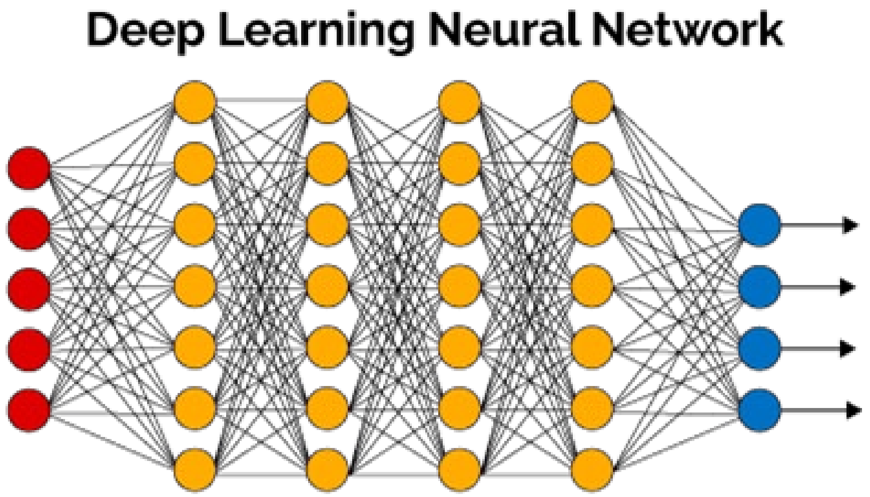
- ML methods often does not have as much restriction as traditional statistical methods
- Risk of overfitting to observed data
- Cross-validation scheme (split data set into training set vs. test set)




For realistic estimation of accuracy

# 2.2. Quality dimensions - Explainability

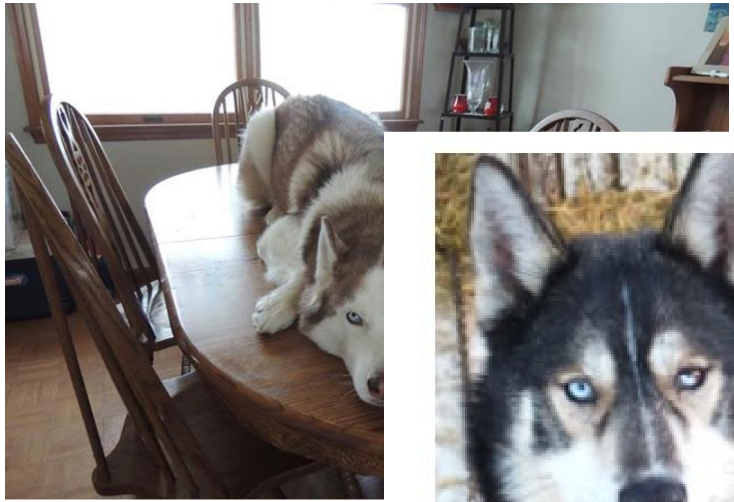
- Degree to which a human can understand how a prediction is made from a statistical or an ML algorithm using its input features
- Increased model complexity might improve accuracy but at the expense of model explainability

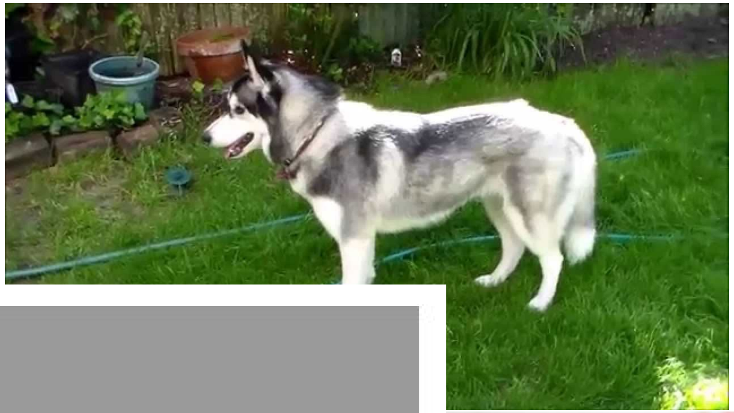


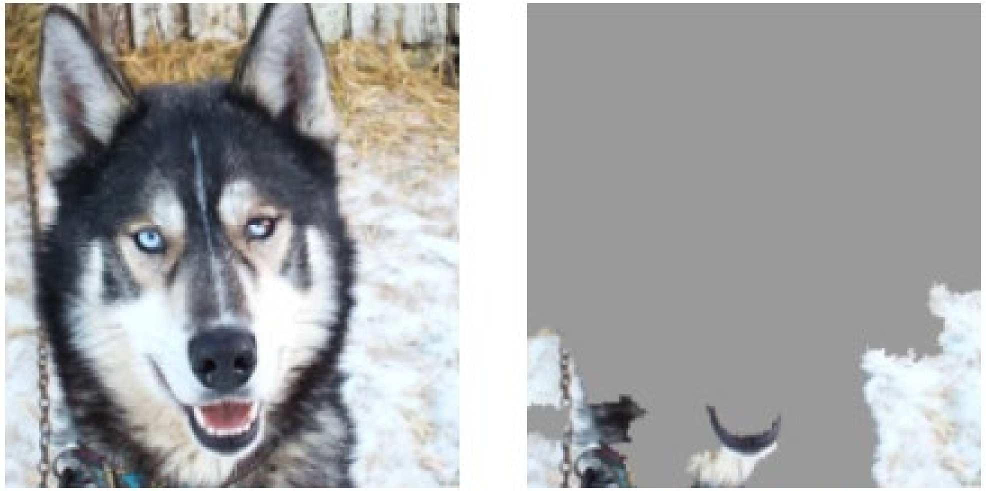
# 2.2. Quality dimensions - Explainability

Husky? Wolf?

# 2.2. Quality dimensions - Explainability







(a) Husky classified as wolf (b) Explanation

Ribeiro et. Al. (2016) "Why Should I Trust You?": Explaining the Predictions of Any Classifier

# 2.2. Quality dimensions - Explainability

E.U. General Data Protection Regulation (GDPR)

"…such processing should be subject to safeguards, which should include… the right to obtain an explanation of the decision…"

L 119/14 EN Official Journal of the European Union 4.5.2016

(71) The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her. However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Such measure should not concern a child.

In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect. Automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions.

UNECE

modernstats

# 2.3. Quality dimensions - Reproducibility

- Three types of reproducibility: methods reproducibility, inferential reproducibility and results reproducibility
  - Methods reproducibility is defined as the **ability to implement, as exactly as possible, the experimental and computational procedures, with the same data and tools, to obtain the same results**
- Machine learning methods are often complex with a lot of parameters, hyperparameters, on top of dependency issues

# 2.3. Quality dimensions - Reproducibility

- How to make “reproducible”?
- Providing enough details about algorithms, assumptions and data so the same procedures could be exactly repeated, in theory or in practice, sharing analytical data sets (original raw or processed data), relevant metadata, analytical code and related software
- Analyses be repeated in-house and by another individual, who should be at arm’s length from the original researcher, to assess reproducibility

# 2.4. Quality dimensions - Timeliness

- The length of time between the reference period and the availability of information
- Also, recommend to consider
  - the length of time it takes to develop or put in place a process
  - the length of time it takes to process data
- The former can take long. But once in production, ML can process vast amounts of data in a short time
- Aspects to consider: Data cleaning, IT infrastructure, Preparation of training data, Evaluation of data quality, Model re-training

Time for development

Time in production

# 2. Quality dimensions – Cost effectiveness

- Degree to which results are effective in relation to the costs of obtaining them (e.g., RMSE reduction per unit cost)
- Fixed cost and on-going cost. Decomposing it into different cost components is useful to better assess potential savings and accuracy improvements against future ongoing costs. This would also help estimate the time needed to recoup the initial investment
- Some ML methods may introduce more cost than others

| Cost component | Type | Purpose |
|---|---|---|
| IT infrastructure | Fixed | Necessary hardware and software |
| Initial staff training | Fixed | Training current staff; hiring new staff |
| Cloud storage | On-going | Cloud storage space |
| Quality assurance | On-going | Conducting quality assurance and control |
| … | | |

# 2. Quality dimensions – Cost effectiveness

- Fixed costs may represent the main challenge, but they also have to be compared with the future savings that ML would grant
- Once the fixed and ongoing costs of training are considered, automation should make it possible to save in terms of staff needed to execute operations. Also potential of using new data sources

# 3. Summary

**Different importance for different stakeholders at different stages**

| **Quality dimension** |
|---|
| Accuracy |
| Explainability |
| Reproducibility |
| Timeliness |
| Cost effectiveness |

# 3. Summary

**Different importance for different stakeholders at different stages**

| Quality dimension | Method 1 | Method 2 | Method 3 (legacy) |
| --- | --- | --- | --- |
| Accuracy | 80% | 85% | 78% |
| Explainability | High (easy) | Low (hard) | High (easy) |
| Reproducibility | High (easy) | Middle | Low (hard) |
| Timeliness | High | Middle | Low |
| Cost effectiveness | High | Middle | Low |

# 3. Summary

- Evaluating quality for statistical algorithms is multi-dimensional
- The proposed QF4SA presents five dimensions to help guide official statisticians when comparing different methods (ML and non-ML)
- The QF4SA is not a replacement for existing quality frameworks but is a supplement to them

# Thank you for your attention

UNECE

modernstats BY HLG - MOS

# Evaluation

- PLACEHOLDER FOR POC ROUNDTABLE FEEDBACK

# Research and Implementation in ML

- ML can be used not only for big data but also for non-big data
- There are advanced methods, but simple methods work well too
- Depending on use case, different emphasis on different quality dimensions
- Sharing and collaboration is key to facilitating ML

# Research and Implementation in ML

- Moving from exploration to production is a key inflection point in enterprises’ ML journeys
- Machine learning solutions are most useful when they align with user needs. A good understanding of existing processes and the ways they meet and frustrate user needs is helpful first step.

# Research and Implementation in ML

Making it to production currently only 35% of organizations indicate that analytical models are fully deployed in production and are often challenged in the "last mile" of the complex and iterative ML workflow*

IDC's Advanced and Predictive Analytics survey and interviews, n = 400, 2017 – 2019

# Research and Implementation in ML

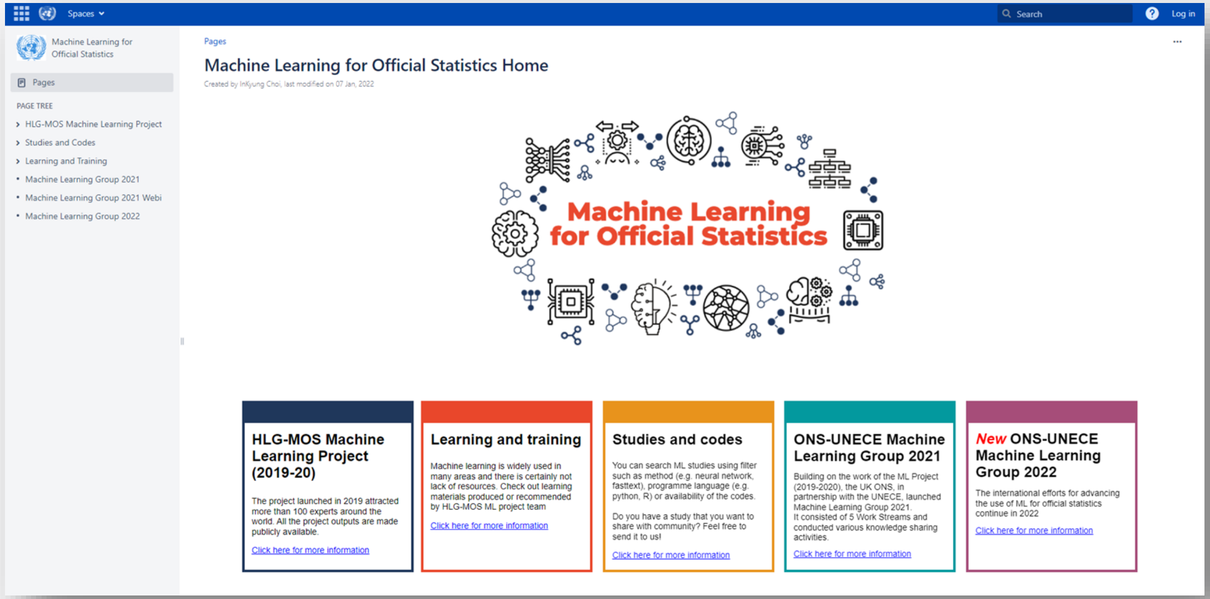
## The challenge of making it to production

Currently only 35% of organizations indicate that analytical models are fully deployed in production and are often challenged in the "last mile" of the complex and iterative ML workflow*

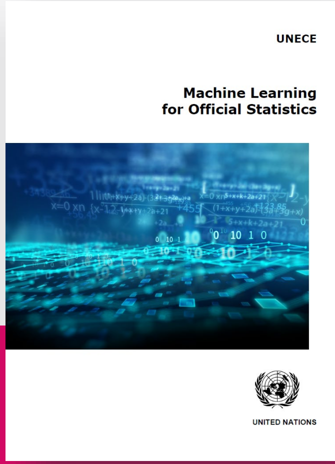
IDC's Advanced and Predictive Analytics survey and interviews, n = 400, 2017 – 2019

# Thank you for your attention

# Resources




**[Machine Learning for Official Statistics Wiki](#)**: all reports, pilot studies, codes, learning resources, etc.




*UNECE publication on Machine Learning for Official Statistics*