



HLG-MOS Machine Learning Project

Presented to 2020 Workshop on the Modernisation of
Official Statistics

Claude Julien – UNECE Project Manager



The work from great^(*) group of people, their numerous collaborators and supporters

Very happily presented to 2020 Workshop on the Modernisation of Official Statistics

Claude Julien – UNECE Project Manager

(*)Great = professional, engaged, talented, perseverant and very pleasant

Is ML just a buzz, a bust or a must?

- ML is definitely NOT just a BUZZ
- ML is a MUST... where it adds value
- ML is still up against some BUSTers (challengers and resistors)

Note: the ML group icon  in this talk points at ML project outputs.

Background

- Project objective: Advance the sound and efficient use of ML in the production of official statistics
- WP1: Pilot Studies (Eric Deeben – ONS DSC)
 - Coding & Classification (Claus Sthamer - ONS DSC)
 - Edit & Imputation (Florian Dumpert – Destatis)
 - Imagery (Abel Coronado & Jimena Carrillo – INEGI)
- WP2: Quality (Wesley Yung – Statistics Canada)
- WP3: Integration challenges (Alex Measure – BLS)

Background

- Project objective: Advance the sound and efficient use of ML in the production of official statistics
- WP1: Pilot Studies (Eric Deeben – ONS DSC)
 - Coding & Classification (Claus Sthamer – ONS DSC)
 - Edit & Imputation (Florian Dummer – ISTAT)
 - Imagery (Abel Corrales & Jolena Carrillo – INEGI)
- WP2: Quality (Wesley Yung – Statistics Canada)
- WP3: Integration challenges (Alex Measure – BLS)

InKyung Choi

ML is definitely NOT a just a buzz!

- The ML group has continued to grow (40 participants and 80 other collaborators and followers)
- Organisational changes within several NSOs
- Some NSOs have set-up internal and external fora for data science and ML
- Many talks, organized sessions and entire on ML
- 325+ registrants from 46 countries and 95 organisations at ML webinar



ML is a MUST... where it adds value

- 21 study reports and 3 summary theme reports shared on public wiki last week
- On C&C, ML does deliver better quality at same/lower cost
- On E&I, ML can add value in some situations
- On Imagery, ML is essential and must be as accessible as possible
 - Generic Pipeline for Production of Official Statistics Using Satellite Data and Machine Learning



ML is still up against some BUSTers (challengers and resistors)

- More ML methods have made it to production processes, but this “acceptance” is still a challenge
- Quality Framework for Statistical Algorithms (QF4SA)
- Integration challenges and practices



Quality Framework for Statistical Algorithms (QF4SA)

- Guide official statisticians on the choice of algorithms to be used in the production process
- Statistical algorithms cover both *traditional* and *modern* statistical methods
- Accuracy, Timeliness, Efficiency, Explainability and Reproducibility
- The QF4SA is not a replacement for existing quality frameworks but is a supplement to them

Integration of ML

- Skills needed? => pretty clear; everyone needs an appropriate level
- Skills acquisition? => biggest challenge; many sources; what is the best mix?
- Skills organisation? => several good models that should to be shared
- Needs-driven or R&D-driven? => NSO-driven (my opinion)
- How to gain acceptance? => Sound and fair assessment
- How to best align with needs? => Share and learn from experiences

How to advance the use of ML?

Acceptance

- Alignment with business needs
- Add value towards bus. needs
- Quality robustness
- Production robustness
- Scientific grounds
- Ethics
- Roles of staff

Facilitation

- Bring together multiple expertise
- Computing infrastructure
- Senior management support
- Research & Development
- Sharing & Collaboration
- Roles of staff

Machine Learning Project Outputs

- Demonstrate added value (quality, efficiency)
 - **21 pilot studies** + analysis = **3 Theme reports** + analysis = **WP1 report**
- Facilitate further experimentation
 - **ML code**
 - **Learning, training and references**
- Address common challenges in journey towards its use
 - **WP2 - Quality Framework for Statistical Algorithms (QF4SA)**
 - **WP3 - Integration report**
- Set the foundation for future developments
 - **WP1 + WP2 + WP3 = Project Summary Report and Recommendations**

Contributions from the ML project

Acceptance

- WP3** • Alignment with business needs
- WP1** • Add value towards bus. needs
- WP2** • Quality robustness
 - Production robustness
 - Scientific grounds **ML proj.**
 - Roles of staff

Facilitation

- Bring together multiple expertise **ML proj. + WP3**
- Computing infrastructure
- Senior management support **HLG-MOS + WP3**
- Research & Development **ML proj.**
- Sharing & Collaboration **ML proj. + NSOs + Community**
- Roles of staff

Recommendations

- Let the ML group continue to share and collaborate
 - ONS DSC has come forward to “lead” the group
 - Need continued support from the Heads of NSOs (HLG-MOS)
 - Stay connected to the HLG-MOS Exec
- Heads of NSO data science units should create a formal network to share management practices
- The ML group should continue to communicate and share with other groups looking into other aspects or other applications of ML within their respective broader agenda

I thank the ML group for all their great work

- Thank you
- Merci
- Tak
- Dank u
- Kiitos
- Danke
- Takk
- Grazie
- 감사합니다
- Takk
- Dziękuję
- Obrigado
- Хвала / Hvala
- Gracias
- Tack
- Mauruuru

I thank the HLG-MOS and the Exec Board for giving me this
very pleasant and enriching opportunity