HLG-MOS MACHINE LEARNING PROJECT

Claude Julien

ModernStats World Virtual Workshop 2020

Introduction

- Approved by HLG-MOS in December 2018
- Launched in March 2019
- Introduced by Eric Deeben at ModernStats 2019
- Will end at the end of 2020

Objective, participation and engagement

- Advance the sound and efficient use of machine learning methods in the production of official statistics
- Over 40 active participants and nearly 80 other people collaborating or following our work
- 22 countries; 31 national organisations; 4 international organisations

Outputs

- 21 reports, mostly demonstrations of added value of ML (pilot studies)
- 3 summary analysis reports on the use of ML in Classification & Coding, Edit & Imputation and use of Imagery
- An initial quality framework for statistical algorithms
- Integration challenges and practices
- Shared code from numerous studies and shared product description data to practice some ML
- Links to learning and training material; references
- Most outputs will be made publicly available by November 12
- **Webinar on November 16 and 17**

Findings on the use of ML

- ML should be used for Coding & Classification
  - Numerous positive demonstrations on a variety of data sources and contexts
  - Just a few applications in production or near-production
- ML shows some promise for Edit & Imputation
  - Demonstrations with a varying degree of positivity
  - Paper on hints and ideas for editing
  - No applications in production yet
- ML is essential in the use of imagery, especially in the context of increasing access to large amounts of imagery data
  - More advanced developments
  - Generic Pipeline for Production of Official Statistics Using Satellite Data and Machine Learning

Findings on positive interdependencies between various humans and ML

- ML when combined with the work of humans (sometimes different work) can achieve better quality at a lower (or same) cost
- ML **REALLY NEEDS** the combination of expertise in Statistics, Subject-matter, IT, Data Science, Sampling & QA methodology

Findings on ongoing challenges

- ML still needs more work to be integrated into the production processes (i.e. be accepted)
- One can demonstrate "what" it can do (added value), but also "how" it does it (explainability)
  - The ML's project quality work team provides an initial framework to address this
  - Opportunity to start bridging/translating traditional statistics and ML terminology
- No single individual can make ML "happen"; NSOs need to adapt their organization to facilitate the acquisition, development and combination of the expertise required to benefit from ML
  - The ML project's integration work team has gathered several NSO practices to facilitate the development and use of ML and should share their experiences along their journey

Going forward

- The ML project will end, but the ML group will very most likely continue on its momentum
- Directions on what it could continue to work, share and collaborate on abound

Further demonstrate the added value of ML in other **statistical business processes** or in the use of other sources of data

Facilitate the use and integration of ML by connecting it with other key processes through **standard models**

"Hidden Technical Debt in Machine Learning Systems," Google NIPS 2015

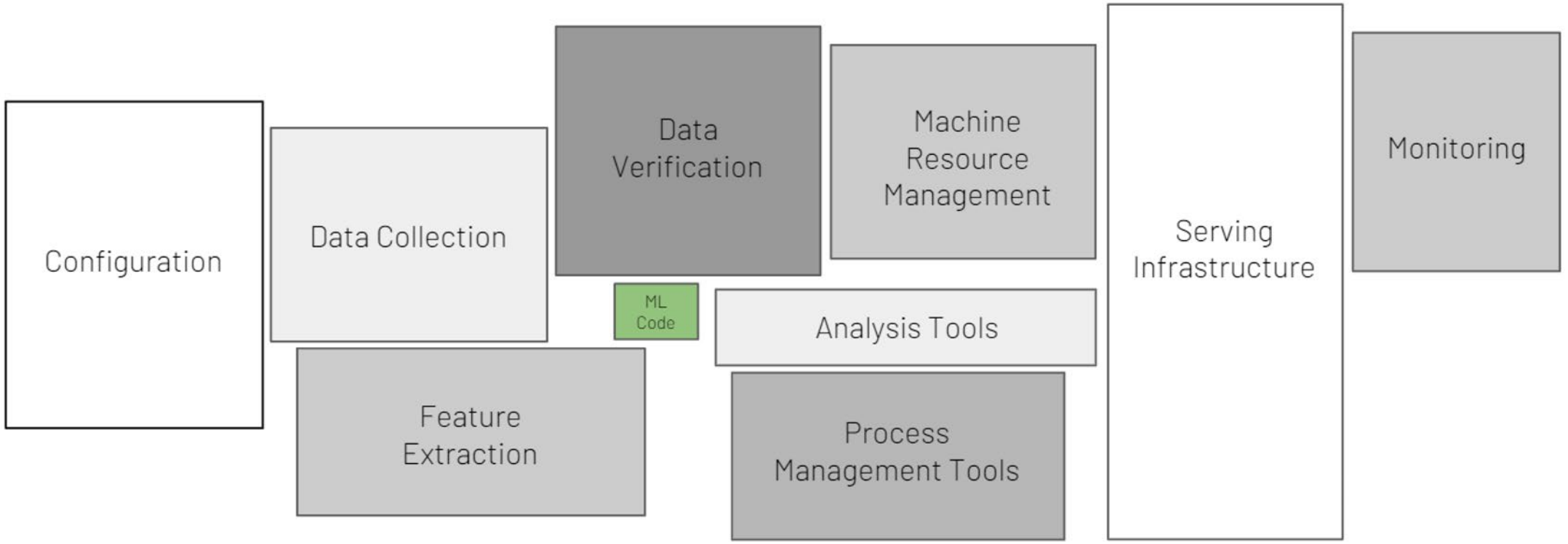


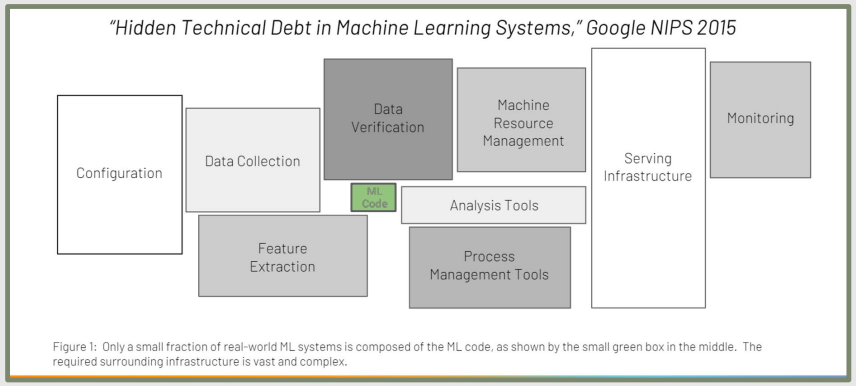
Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small green box in the middle. The required surrounding infrastructure is vast and complex.

Facilitate the use and integration of ML by addressing other mandatory and cultural dimensions

Ethics

Law

Competence



Business processes

Management

Security

Organisation and acceptance

Conclusion (just ideas)

- Say a few words on the ML project
- Emphasize need to continue to communicate and coordinate between groups, projects and other initiatives (e.g. input from Standards/ModernStats group in setting the future directions and activities of the ML group)
- Invitation to register to ML webinar