



Blue Sky Thinking Network – Looking Back & Ahead

By UNECE Modernstats, November 2018



What were the BSTN goals?

- Identify and evaluate new opportunities
- Generate new ideas
- Develop concrete proposals
- Coordinate internationally
- Create synergy with other platforms



Looking Back at 2018

Success

- BSTN 2018 has resulted in
 - Machine Learning Position Paper
 - ML Project Proposal
 - Plan to work on similar IT position paper(s) in 2019

Learning

- “Think Tank” role of BSTN not (yet) fully realised
 - ***Do we need a change to improve this?***

ML position paper (1/4)

- Written by team from Canada, Finland, Italy, Mexico, Netherlands
- Reviewed by Executive Board
- Finalised November 2018, will become available on the HLG-MOS wiki

The use of machine learning in official statistics

UNECE Machine Learning Team, November 2018

Wesley Yung (Canada), Jukka Karkimaa (Finland), Monica Scannapieco (Italy), Giulio Barcarolli (Italy), Diego Zardetto (Italy), José Alejandro Ruiz Sanchez (Mexico), Barteld Braaksma (Netherlands), Bart Buelens (Netherlands), Joep Burger (Netherlands)

Abstract

This paper is written for managers and policy makers to inform them about the possibilities to use machine learning (ML) in the production of official statistics and to demystify ML for those official statisticians unfamiliar with it. After providing some background on official statistics and ML, the paper explains why ML is becoming relevant for official statistics. The possible uses of ML in official statistics are presented for primary data (probability samples) using the GSBPM process model, and for secondary data (registers, big data/ non-probability samples and mixed sources). Finally, issues about quality are discussed, which we believe should be addressed with high priority.

Introduction

Computers have learned to paint in the style of Rembrandt (www.nextrembrandt.com) and to compose music in the style of Bach (Hadjeres et al. 2017). Chess computer Deep Blue already defeated the world champion over twenty years ago, question-answering computer Watson won the

This paper is written for managers and policy makers to inform them about the possibilities to use machine learning (ML) in the production of official statistics and to demystify ML for those official statisticians unfamiliar with it.

ML position paper (2/4)

- Contents (14 pages, we aimed for 10...)
 - 1) Introduction
 - 2) Conclusions and recommendations
(Next slides)
 - 3) Background
(What is it, why relevant?)
 - 4) ML in primary data
(GSBPM based)
 - 5) ML in secondary data
(Big data, register data, multisource)
 - 6) Quality considerations
(Traditional indicators vs. ML-specific issues)
 - 7) References



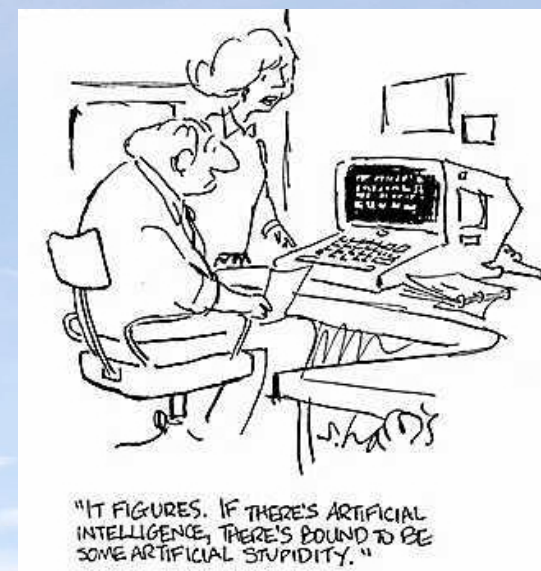
ML position paper (3/4)

- Conclusions
 - ✓ Processing of secondary data is steering NSOs towards use of ML
 - ✓ But use can also extend to more traditional domains

	Family of ML techniques	GSBPM phase
Record linkage	Clustering	2.4, 5.1
Coding	Classification	2.4, 4.3, 5.2
Outlier detection	Clustering	2.4, 4.3, 5.1, 6.2
Stratification	Classification	4.1, 4.3, 5.4, 5.6
Estimation	Regression/classification	4.3
Imputation	Regression/classification	5.4
Calibration	Regression/classification	5.6
Disclosure control	Regression/classification	6.4

ML position paper (4/4)

- Recommendations
 - Develop quality framework
 - ❖ New, or integrated in existing one?
 - ❖ Possibly als guidelines for ML use
 - Create inventory of ML projects
 - ❖ To exchange both methods and applications
 - ❖ Not one-off but living document
 - ❖ Including previous work (Chu&Poirier, Destatis)
 - Set up interdisciplinary teams for ML work
 - ❖ Modelers
 - ❖ Programmers
 - ❖ subject matter specialists



Next year- new IT trends

- Group on new IT trends like
 - Secure multiparty computing
 - Edge computing
 - Blockchain techniques
 - Cloud services
- For each topic short position paper (aim 3 pages, ready in 3 months)
 - What is it
 - What does it mean for OS
 - What is already being done
 - What future work is needed
- Start with physical meeting?
- Core team: Canada, Italy, Netherlands, UK
 - Others welcome!



What more to promote BSTN?

- ✓ Invest in building a real network and core group
- ✓ Combine top-down and bottom-up approaches
 - Monitor relevant ideas from high-level meetings
 - See what's happening elsewhere (i
 - Collect suggestions from HLG-MOS workshops
 - Liaise with HLG-MOS Modernisation Committees
 - Get the people within NSO's innovation labs together
- Arrange meeting(s) where hands-on blue sky thinkers:
 - Share BST happening within their organizations
 - Share, discuss and prioritize any BST ideas that might be promising
 - Agree on (limited!) next steps to realize ideas
 - Translate ideas into proposals
- ✓ Liaise with other communities
 - Geo, academic world, governments, NGOs, private sector, ...
- ✓ Allocate resources to develop and work on BST ideas
- ✓ Secretarial support from UNECE



What you can do:



Blue Sky Thinking Network

The "ideas factory" for statistical modernisation. This network provides a platform where members can share ideas and look for partners to explore how statistical organisations can benefit from innovations to increase efficiency or to develop new products and services. For modernising Official Statistics, the Network aims to:

- Identify and evaluate new opportunities
- Generate new ideas
- Develop concrete proposals
- Collaborate internationally
- Create synergy with other communities

What you can do:

- Join the Network
- Bring your ideas
- Contribute to other ideas
- See what is happening



Join and bring your ideas for innovating statistics to the Network

The Network consists of a flexible pool of statisticians and other experts with a wide range of capabilities and experiences. Anybody can join and submit ideas to transform how we produce statistics or how we work. The Network aims to follow-up on selected ideas through concrete pilots or project proposals. Ideas can be submitted through the [Network's wiki space](#)

Contact us to join:

Network Coordinator: Barteld Braaksma (bbka@cbs.nl)

Network Support: Gerald Muriel (gerald.muriel@unece.org)

Submit your ideas or get more information:

Wiki platform: <http://www1.unece.org/stat/platform/display/BST>

modernstats
by HLG - MOS



– Innovation Manager: Barteld Braaksma (bbka@cbs.nl)

• Wiki platform:

<https://statswiki.unece.org/display/BST>



Blue Sky Thinking Network – Looking Back & Next Steps

By UNECE Modernstats

