



Pilotprojekte im Rahmen der Dateninnovationsstrategie

# Kurze methodische Evaluation im Hinblick auf eine mögliche Integration in die Produktion

**EXPERIMENTAL STATISTICS**



Neuchâtel, 2020

<b>Herausgeber:</b>	Bundesamt für Statistik (BFS)	<b>Layoutkonzept:</b>	Sektion DIAM
<b>Auskunft:</b>	Jean-Pierre.Renfer@bfs.admin.ch, Tel. 058 463 66 62	<b>Download:</b>	<a href="http://www.statistik.ch">www.statistik.ch</a>
<b>Redaktion:</b>	Jean-Pierre Renfer	<b>Copyright:</b>	BFS, Neuchâtel 2020 Wiedergabe unter Angabe der Quelle für nichtkommerzielle Nutzung gestattet
<b>Themenbereich:</b>	00 Statistische Grundlagen		
<b>Originaltext:</b>	Französisch		
<b>Übersetzung:</b>	Sprachdienste BFS		

# 1 Evaluierte Pilotprojekte

Die ersten fünf Pilotprojekte im Zusammenhang mit der Dateninnovationsstrategie 1.0 des BFS wurden auf Basis der Projektberichte von Ende Juni 2019 dahingehend evaluiert, ob sie in die Produktion überführt werden können. Konkret geht es um folgende Pilotprojekte:

- a) **ADELE**: Luftbildinterpretation im Rahmen der Statistik der Bodennutzung;

Verwendete komplementäre Analysemethoden: für das Training auf den Luftbildern wurden mittels Keras verschiedene «Convolutional Neural Networks» (CNN) angewandt;

- b) **ML\_SoSi**: Verläufe im System der sozialen Sicherheit;

Verwendete komplementäre Analysemethoden: k-Means und Optimal Matching (agnes mit Optimal-Matching-Datenmatrix aus seqdist) für «Sequence Clustering» sowie «State Distribution Plots» (package TraMineR) und «Sankey Plots» für die Visualisierung;

- c) **NOGAuto**: Automatisierung der Kodierung der wirtschaftlichen Tätigkeit von Unternehmen im Betriebs- und Unternehmensregister;

Verwendete komplementäre Analysemethoden: in der Vorbereitungsphase «Text Mining» («Natural Language Processing», NLP); in der Modellierungsphase künstliche neuronale Netze und «Random Forests»;

- d) **Plausi++**: Automatische Plausibilitätsprüfung zur Überprüfung der Qualität und Zuverlässigkeit von Administrativ- und Umfragedaten;

Verwendete komplementäre Analysemethoden: im prädiktiven Teil «Gradient Boosting Machines» und «Random Forests»; im explikativen Teil «Ceteris Paribus» und «Local Interpretable Model-Agnostic Explanations» (LIME)

- e) **SAE\_BESTA**: Evaluation des Potenzials zur Verbesserung der Präzision im Rahmen der Beschäftigungsstatistik (BESTA)

Verwendete komplementäre Analysemethoden: «Small Area Estimation» (lmer) und robustes Modell (rlm mit Huber-M) unter Verwendung von Extrapolationsgewichten.

## 2 Evaluation

### 2.1 Evaluationskriterien

Die Pilotprojekte werden im Hinblick auf eine mögliche Überführung in die Produktion einzeln evaluiert. Dazu werden fünf Kriterien herangezogen: **Dokumentation, Ziele, Mehrwert, Qualität der Validierung** und **Übereinstimmung mit den Qualitätskriterien des BFS**.

Die Kriterien in Bezug auf die Dokumentation, die Qualität der Validierung sowie die Übereinstimmung mit den Qualitätskriterien des BFS haben einen massgeblichen Einfluss auf die Gesamtbeurteilung. Diese drei Kriterien müssen deshalb vollständig erfüllt sein.

### 2.2 Status der Pilotprojekte

Die Evaluation dient als Grundlage für eine Empfehlung bezüglich des Status, der den Pilotprojekten zugewiesen werden soll, d.h.:

- **Im Gange**: Fortsetzung der Arbeiten gemäss den Empfehlungen aus der Evaluation
- **Abgeschlossen**: Abschluss des Projekts ohne Überführung in die Produktion
- **Angehalten**: Unterbrechung des Projekts aufgrund der aktuellen Umstände (Prioritäten, Ressourcen, Kenntnisse, Infrastruktur usw.) mit Möglichkeit auf Reaktivierung, sofern es die Umstände erlauben
- **In die Produktion überführt**: Lancierung eines Projekts zur Integration in die Produktion

## 2.3 Evaluation und Empfehlungen für die Pilotprojekte

- a) ADELE: Die Dokumentation fasst die Vorgehensweise und die verwendeten Methoden gut zusammen. Einige informatische Aspekte könnten noch ergänzt werden. Die Projektziele wurden erreicht und eine Integration in das Revisionsprojekt der Arealstatistik ist geplant. Der Mehrwert für die Produktion wird im Rahmen dieses Revisionsprojekts ermittelt. Dasselbe gilt für die Kriterien hinsichtlich der Qualität der Validierung sowie der Übereinstimmung mit den Qualitätskriterien des BFS.

Empfohlener Status

**In die Produktion überführt:** Lancierung eines Integrationsprojekts für die Produktion.

- b) ML\_SoSi: Die Dokumentation erläutert mehrere Aspekte gut, liesse sich jedoch noch optimieren, indem bestimmte angewandte Methoden detaillierter beschrieben werden. Das Projekt ist weit fortgeschritten, es wurden jedoch noch nicht alle festgelegten Ziele erreicht. Insbesondere mit den ersten Ergebnissen und deren Darstellung wird bereits ein gewisser Mehrwert geschaffen. Die Qualität der Validierung und die Übereinstimmung mit den Qualitätskriterien des BFS können erst beurteilt werden, wenn das Projekt abgeschlossen ist.

Empfohlener Status

**Im Gange:** Fortsetzung der Arbeiten gemäss den Empfehlungen.

- c) NOGAuto: Die Dokumentation gibt einen guten Überblick über die Problematik und die geplante Vorgehensweise, könnte aber noch detaillierter sein. Die Arbeiten müssen fortgesetzt werden, wobei insbesondere die Prozeduren verallgemeinert und optimiert werden sollten. Die ersten Ergebnisse sind in Bezug auf den Mehrwert vielversprechend, müssen jedoch noch validiert und mit den Qualitätskriterien des BFS abgeglichen werden.

Empfohlener Status

**Im Gange:** Fortsetzung der Arbeiten gemäss den Empfehlungen.

- d) Plausi++: Die Dokumentation beschreibt die Vorgehensweise detailliert und verständlich und enthält alle nötigen Elemente für eine Reproduktion. Die Ziele wurden vollständig erreicht und der Mehrwert für die Produktion konnte anhand einer umfassenden Validierung, insbesondere unter Berücksichtigung mehrerer Methoden, aufgezeigt werden. Bei der Fortsetzung der Arbeiten muss der Fokus auf das Qualitätssicherungsverfahren innerhalb der Produktion gelegt werden, damit die Qualitätskriterien des BFS erfüllt werden.

Empfohlener Status

**Im Gange:** Abschluss der Arbeiten gemäss den Empfehlungen.

- e) SAE\_BESTA: Die Dokumentation ist relativ detailliert und die am Ende erwähnten offenen Punkte sind noch anzugehen. Die Arbeiten sind gut fortgeschritten, decken jedoch noch nicht alle Analysebereiche ab. Der Mehrwert, die Qualität der Validierung und die Übereinstimmung mit den Qualitätskriterien des BFS können erst beurteilt werden, wenn das Projekt abgeschlossen ist.

Empfohlener Status

**Im Gange:** Fortsetzung der Arbeiten gemäss den Empfehlungen.