



Projets pilotes dans le cadre de la stratégie d'innovation sur les données

Evaluations méthodologiques succinctes en vue d'un transfert possible dans la production

EXPERIMENTAL STATISTICS



Neuchâtel, 2020

Éditeur:	Office fédéral de la statistique (OFS)	Concept de mise en page:	Section DIAM
Renseignements:	Jean-Pierre.Renfer@bfs.admin.ch, tél. 058 463 66 62	Téléchargement:	www.statistique.ch
Rédaction:	Jean-Pierre Renfer	Copyright:	OFS, Neuchâtel 2020
Domaine:	00 Bases statistiques des généralités		La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée
Langue du texte original:	Français		
Traduction:	Services linguistiques de l'OFS		

1 Projets pilotes évalués

Les cinq premiers projets pilotes de la stratégie d'innovation 1.0 de l'OFS suivants ont été évalués en vue d'un transfert possible dans la production sur la base des rapports de projet de fin juin 2019. En bref, il s'agit des projets pilotes suivants :

- a) **ADELE** : Reconnaissance d'images aériennes pour la statistique nationale de la superficie ;

Méthodes d'analyse complémentaires utilisées : « Différents réseaux neuronaux de convolution (convolutional neural network – CNN) ont été appliqués pour l'entraînement sur des images aériennes se basant sur keras » ;

- b) **ML_SoSi** : Parcours individuels dans le système de sécurité sociale ;

Méthodes d'analyse complémentaires utilisées : kmeans et optimal matching (agnes avec matrice de distance d'appariement optimal tirée de seqdist) pour «Sequence Clustering», «State Distribution Plots» (package TraMineR) et «Sankey Plots» pour la visualisation ;

- c) **NOGAuto** : Automatisation de la codification des nomenclatures économiques dans le Registre des entreprises et établissements ;

Méthodes d'analyse complémentaires utilisées : pour la phase préparation « text mining » (« natural language processing », NLP); pour la phase modélisation, « artificial neural networks » et « random forests » ;

- d) **Plausi++** : Contrôles automatiques de plausibilité pour vérifier la qualité et la fiabilité des données administratives et des données collectées par voie d'enquête ;

Méthodes d'analyse complémentaires utilisées : partie prédictive «Gradient Boosting Machines» et «Random Forests». Pour la partie explicative «Ceteris Paribus» et «Local Interpretable Model-Agnostic Explanations» (LIME) ;

- e) **SAE_BESTA** : Evaluation du potentiel d'amélioration de la précision dans le cadre de la Statistique de l'emploi (BESTA)
Méthode d'analyse complémentaire utilisées : « Small area estimation » (lmer) et modèle robuste (rlm avec Huber-M) en utilisant les poids d'extrapolation.

2 Evaluation

2.1 Critères d'évaluation

L'évaluation d'un transfert possible dans la production se base sur les évaluations individuelles des projets pilotes. Pour cela cinq critères ont été retenus, portant sur la **documentation**, les **buts**, la **valeur ajoutée**, la **qualité de la validation** et **l'adéquation avec les critères de qualité de l'OFS**.

L'évaluation des critères concernant la documentation, la qualité de la validation ainsi que l'adéquation avec les critères de qualité de l'OFS joue un rôle prépondérant dans l'évaluation globale et ces derniers doivent de ce fait être entièrement satisfaits.

2.2 Statut des projets pilotes

L'évaluation sert de base à une recommandation concernant le statut du projet pilote lequel désigne un des états suivants :

- **En cours** : poursuite des travaux selon les recommandations de l'évaluation ;
- **Terminé** : clôture du projet sans transfert dans la production ;
- **Arrêté** : interruption du projet sous les conditions actuelles (priorités, données, ressources, savoir-faire, infrastructure, etc.) avec possibilité de réactivation si les conditions le permettent ;
- **Transféré dans la production** : initialisation d'un projet de transfert dans la production.

2.3 Evaluations et recommandations des projets pilotes

- a) ADELE : La documentation synthétise bien le procédé et les méthodes utilisées mais pourrait encore être développée notamment sur certains aspects informatiques qui devraient y figurer. Les buts de ce projet ont été atteints et une intégration dans le projet de révision de la statistique de la superficie est prévu. La valeur ajoutée pour la production sera évaluée dans ce projet de révision. A noter que les critères concernant la qualité de la validation de même que l'adéquation avec les critères de l'OFS seront également adressés dans ce projet de révision.

Statut recommandé

Transféré dans la production : initialisation d'un projet de transfert dans la production.

- b) ML_SoSi : La documentation décrit bien plusieurs aspects mais pourrait encore être optimisée avec la description détaillée de certaines méthodes appliquées. Le projet est bien avancé mais demande encore un effort pour atteindre les buts fixés. Le projet a déjà permis d'apporter une certaine valeur ajoutée pour la production en particulier avec les premiers résultats et leur visualisation. La qualité de la validation ainsi que l'adéquation avec les critères de l'OFS pourront être évaluées une fois le projet finalisé.

Statut recommandé

En cours. Poursuivre les travaux selon les recommandations.

- c) NOGAuto : La documentation permet d'avoir un bon aperçu de la problématique et des procédés envisagés mais devrait être encore plus détaillée. Les travaux doivent être poursuivis en généralisant le procédé et en essayant d'optimiser ce dernier. Les premiers résultats sont prometteurs en termes de plus-value potentielle mais doivent encore être validés et confrontés aux critères de qualité de l'OFS.

Statut recommandé

En cours. Poursuivre les travaux selon les recommandations.

- d) Plausi++ : La documentation détaille bien le procédé de manière compréhensible et fournit tous les éléments pour la reproduction de ce dernier. Les buts ont été entièrement atteints et la valeur-ajoutée pour la production a pu être démontrée en se basant sur une validation approfondie incluant notamment plusieurs méthodes. La suite des travaux doit se concentrer sur le procédé pour garantir la qualité au cours de la production pour être en adéquation avec les critères de qualité de l'OFS.

Statut recommandé

En cours. Finaliser les travaux selon les recommandations.

- e) SAE_BESTA : La documentation est relativement bien détaillée et les points ouverts mentionnés à la fin de celle-ci seront être abordés. Les travaux sont bien avancés mais ne couvrent pas encore tous les domaines d'analyse. La valeur ajoutée, la qualité de la validation ainsi que l'adéquation avec les critères de l'OFS pourront être évaluées une fois le projet finalisé.

Statut recommandé

En cours. Poursuivre les travaux selon les recommandations.