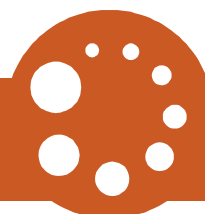




Le projet d'innovation des données «NOGAuto»

Management Summary

EXPERIMENTAL STATISTICS



Neuchâtel, 2024

Éditeur: Office fédéral de la statistique (OFS)
Renseignements: info.urd@bfs.admin.ch
Rédaction: Lorenz Helbling ; Mathias Constantin ;
Daniele Marx ; Cindia Duc Sfez ; OFS
Domaine: 00 Bases statistiques et généralités
Langue du texte original: Français

Concept de mise en page: Section PUB
Téléchargement: www.statistique.ch
Copyright: OFS, Neuchâtel 2024
La reproduction est autorisée, sauf à des fins
commerciales, si la source est mentionnée

Le projet d'innovation des données "NOGAuto"

Management Summary

Le projet NOGAuto s'inscrit dans la stratégie d'innovation sur les données de l'Office fédéral de la statistique (OFS) et fait ainsi partie des principaux projets pilotes retenus dans le but d'améliorer la production statistique publique. Le programme NOGAuto a pour vocation de prédire la classification de l'activité économique des établissements et des entreprises par l'utilisation de méthodes d'apprentissage automatique. Cette classification automatisée servira de soutien aux personnes qui codent la Nomenclature Générale de l'Activité économique (NOGA) à l'OFS. Comme de nombreuses données nécessaires à l'apprentissage automatique sont disponibles au sein de l'OFS (données d'enquêtes, description d'activité au registre du commerce, notes explicatives de la classification, etc.), ce projet s'est naturellement focalisé sur l'utilisation de ces dernières.

Dans un premier temps, les données disponibles dans le Registre suisse des Entreprises et Établissements (REE) ont été sélectionnées puis préparées afin de les rendre utilisables pour la suite du processus. Parmi ces données se trouvent les descriptions d'activités économiques qui sont la principale source pour attribuer un code NOGA aux établissements et aux entreprises. Une technique de « natural language processing » a été appliquée à ces descriptions pour extraire le maximum d'informations utiles à l'algorithme de prédiction des codes NOGA. Enfin, ces données ont été utilisées pour former des modèles (un par langue) à l'aide de techniques d'apprentissage automatique, qui ont ensuite été testés sur un échantillon aléatoire afin d'évaluer la qualité de la prédiction.

Sur la base de ces modèles, différents outils ont été développés pour être transférés en production au sein de l'OFS. Un premier outil de contrôle de listes est déjà mis en place pour assurer une meilleure qualité dans les codes NOGA déjà présents dans les différents registres. Un second produit devrait suivre sous la forme d'une application qui peut être utilisée comme soutien lors du codage de la NOGA sur la base d'une description de l'activité économique.