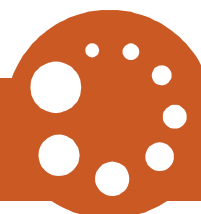




The Data Innovation Project “NOGAuto”

Management Summary

EXPERIMENTAL STATISTICS



Neuchâtel, 2024

Published by: Federal Statistical Office (FSO)

Information: info.urd@bfs.admin.ch

Editors: Lorenz Helbling ; Mathias Constantin ;
Daniele Marx ; Cindia Duc Sfez; UST

Topic: 00 Statistical Basis and Overviews

Original text: French

Translation: FSO language services

Layout concept: Section PUB

Downloads: www.statistics.ch

Copyright: FSO, Neuchâtel 2024
Reproduction with mention of source authorised
(except for commercial purposes)

The Data Innovation Project “NOGAuto”

Management Summary

NOGAuto is one of the flagship pilot projects selected to improve the production of public statistics under the Federal Statistical Office's (FSO) data innovation strategy. The NOGAuto application aims to leverage machine learning methods to predict the classification of economic activities of businesses and enterprises. These automated classifications are intended to assist FSO staff assigning General Classification of Economic Activities (NOGA, Swiss NACE) codes. The project focused on reusing the large range of data already available at the FSO (survey data, enterprise descriptions in the commercial register, classification explanatory notes, etc.) for machine learning.

The first step was to select and prepare existing data from the Swiss Business and Enterprise Register (BER) for the next step in the process. The descriptions of economic activities, which are the main basis for assigning a NOGA code to businesses and enterprises, are among these data. A 'natural language processing' technique was applied to these descriptions to extract the maximum amount of useful information for the NOGA code prediction algorithm. Finally, these data were used to train models (one per language) using machine learning techniques, which were then tested on a random sample to assess the quality of the prediction.

On the basis of these models, various tools have been developed aimed for being transferred into FSO's production. A first application to be rolled out is a list-checking tool to improve the accuracy of existing codes in the various registers. A second application should assist FSO staff with NOGA code assignment based on descriptions of economic activity.

.