



EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL





Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Federal Statistical Office FSO

Improving efficiency in assignment and quality control of NACE codes combining innovative methodologies with human expertise

Athanassia Chalimourda, Lorenz Helbling, Mathias Constantin

Data Science, AI and statistical methods

Interoperability and Registers

Quality 2024, 4-7 June 2024

Contents

- Introduction
- Automatic Classification of Economic Activities
- Overall Quality: Global Performance Measures
- Quality Measures in Decision Making: Performance by Class
- Interaction with Experts and Expert-Systems
- Quality Control of Existing Codes and Activity Descriptions
- Conclusions

NOGAuto, Shiny App, Methods (Business Registers Data):

Lorenz Helbling, Mathias Constantin,
Cindia Duc Sfez, Daniele Marx

Methodological Support (Statistical Methods):

Athanassia Chalimourda, Daniel Assoulin

Introduction

The *Nomenclature générale des activités économiques*, **NOGA** is the Swiss **NACE** version.

- NOGA has hierarchical categories
- The classification of the economic activities according to NOGA is performed manually by coding experts.

A classification example:

*The operation of a drugstore and the marketing of all drugstore, herbal, dietetic and cosmetic products, medicines and health products (NOGA Code: **477501**)*

- **Sector** (3 classes): **3** – Services
- **Section** (21): **G** – Trade; Maintenance and repair in motor vehicles
- **Two digits** (88): **47** – Retail trade (excluding trade in motor vehicles)
- **Four digits** (615): **4775** – Retail trade in cosmetic and body care products
- **Six Digits** (794): **477501** – Drugstore

Automatic Classification of Economic Activities

NOGAuto is an assistance system for **automatic classification** and **interaction with the coding experts**.

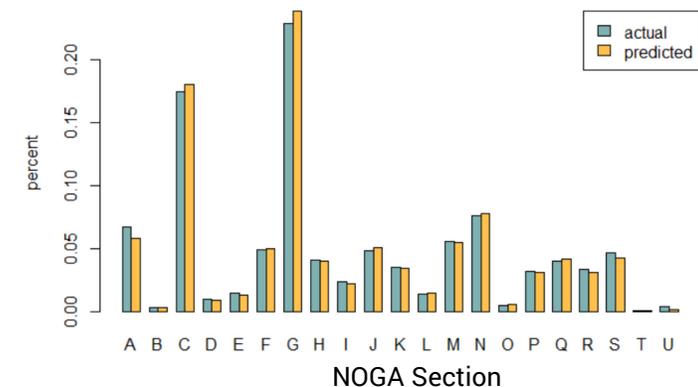
- It involves Natural Language Processing turning an activity description into a numeric vector (text2vec).
- A supervised machine learning with a **gradient boosting machine (GBM)** assigns a NOGA Code to an activity description.
- The **predicted code** is assigned to a description with a **prediction probability** approximated by the number of GBM-trees that voted for that code.
- NOGAuto follows the **hierarchical structure** of NOGA. The GBM proceeds to the next NOGA level for the most probable predictions.

Overall Quality: Global Performance Measures

- **Accuracy:** Overall percentage of elements for which the predicted and the actual class are the same
- **Balanced Accuracy:** Mean value of the agreement percentages per class (with respect to the actual classes)
- **Cohen's Kappa:** The accuracy is corrected for the class agreement expected by chance
- **Comparison of the distributions** of the classes' percentages

	NOGA-Section	NOGA-Division
Accuracy	0.90	0.88
Balanced Accuracy	0.87	0.86
Cohen's Kappa	0.89	0.87

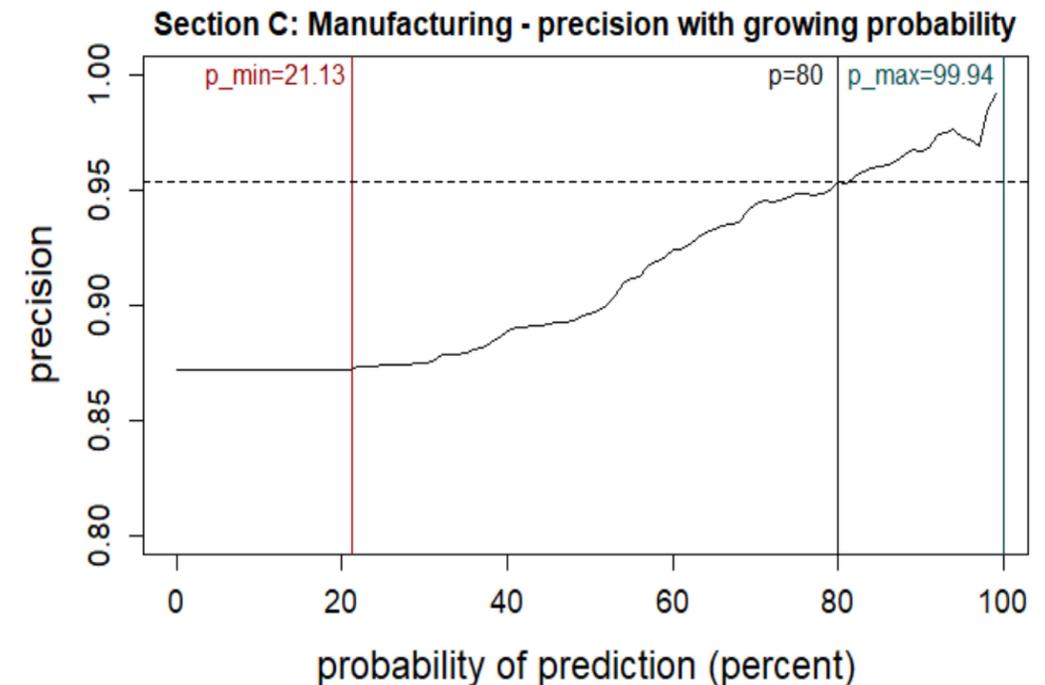
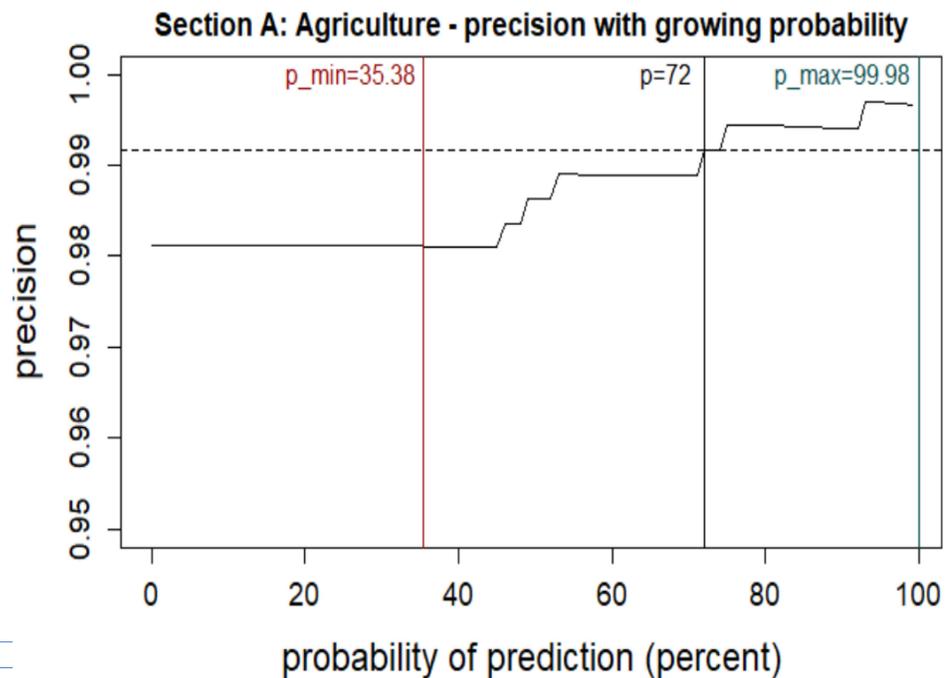
Distributions of the percentages of the classes in a small sample ($n_{\text{test}} = 6408$) of activity descriptions in French

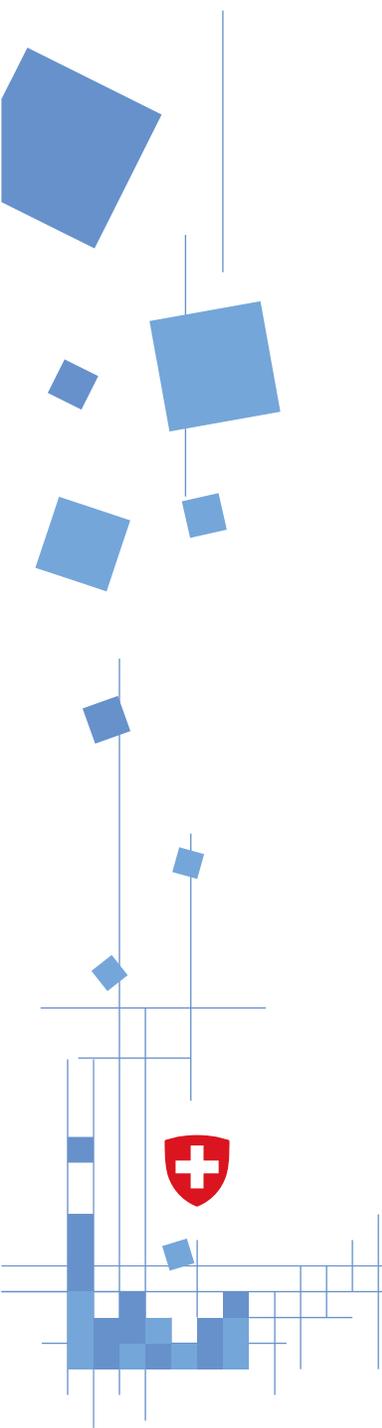


Quality Measures in Decision Making: Performance by Class

We use **precision** by class and an element's **prediction probability** to find subsets of higher precision in the predicted classes.

Precision measures the proportion of correctly predicted elements (true positives) out of all elements predicted in a given class (true positives *and* false positives).





Interaction with Experts and Expert-Systems

Maximize **automation** while controlling **quality** (error propagation, predictive and distributional accuracy)

Assist coding experts, leaving challenging classifications to them.

Interaction with Expert-Systems

Example (originally in German): «*The company provides a full range of graphic design and illustration services. It also supports companies, institutions and individuals in communication matters.*»

Choose language: en

NOGA Code predictions

Insert the needed variables

Write the activity description

Die Gesellschaft erbringt sämtliche Dienstleistungen im Bereich Grafik und Illustration. Ausserdem unterstützt sie Unternehmen, Institutionen und Einzelpersonen in Kommunikationsfragen.

Detected language: German

If the language is not correct, please select:
 Français Allemand Italien

Search

First Prediction

741
Activités spécialisées de design
63.7 %
Train model
Final code

Second Prediction

749
Autres activités spécialisées, scientifiques et techniques n.c.a.
25.9 %
Train model
Final code

Third Prediction

742
Activités photographiques
6.3 %
Train model
Final code

Text translation

Texte original	Die Gesellschaft erbringt sämtliche Dienstleistungen im Bereich Grafik und Illustration. Ausserdem unterstützt sie Unternehmen, Institutionen und Einzelpersonen in Kommunikationsfragen.
Texte traduit	La société fournit tous les services dans le domaine du graphisme et de l'illustration. En outre, elle soutient les entreprises, les institutions et les particuliers en matière de communication.

Text cleaning

Texte original	La société fournit tous les services dans le domaine du graphisme et de l'illustration. En outre, elle soutient les entreprises, les institutions et les particuliers en matière de communication.
Texte nettoyé	fourn serviv domain graphism Illustr soutient entrepris institu particul mat commun

+

Interaction with Expert-Systems

Example (originally in German): «*The company provides a full range of graphic design and illustration services. It also supports companies, institutions and individuals in communication matters.*»

The screenshot shows the KUBB Coding assistance application interface. At the top, there is a navigation menu with 'Federal Administration', 'FDHA', and 'FSO'. The page title is 'KUBB Coding assistance application'. Below the title, there is a search bar and a navigation menu with 'NOGA codes', 'Search', 'Codification rules', and 'Instructions'. The main content area displays the breadcrumb 'NOGA > M > 74 > 741 > 7410 Specialised design activities'. The left sidebar shows a list of NOGA codes: 741001 Industrial and product design, 741002 Graphic design and visual communication, and 741003 Interior and spatial design. The main content area shows a tree view of NOGA codes, with '7410 Specialised design activities' selected. Below the tree view, there is a section titled 'Corresponding CPA codes' with a list of subcategories and their descriptions.

Federal Administration > FDHA > FSO

Contact DE FR IT EN

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

KUBB
Coding assistance application

Search

NOGA codes Search Codification rules Instructions

NOGA > M > 74 > 741 > 7410 Specialised design activities

< 741 Specialised design activities

7410 Specialised design activities

741001 Industrial and product design

741002 Graphic design and visual communication

741003 Interior and spatial design

7410 Specialised design activities

- NOGA
 - M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES
 - 74 Other professional, scientific and technical activities
 - 741 Specialised design activities
 - 7410 Specialised design activities
 - 741001 Industrial and product design
 - 741002 Graphic design and visual communication
 - 741003 Interior and spatial design

Corresponding CPA codes

- **74.10.1** Interior, industrial and other specialised design services
- **74.10.11** Interior design services
This subcategory includes:
 - interior design services such as the planning and designing of interior spaces to meet the physical, aesthetic and functional needs of people
 - drawing up of designs for interior decorating
 - interior decoratingThis subcategory excludes:
 - architectural design services, see 71.11
- **74.10.12** Industrial design services
This subcategory includes:
 - design services for industrial products, i.e. creating and developing designs and specifications that optimise the use, value and appearance of products, including the determination of the materials, construction, mechanism,

Interaction with Expert-Systems

Example (originally in German): «*The company provides a full range of graphic design and illustration services. It also supports companies, institutions and individuals in communication matters.*»

The screenshot shows a web application interface for NOGA Code predictions. At the top right, there is a language selection dropdown set to 'en'. The main content area is titled 'NOGA Code predictions' and is divided into several sections:

- Insert the needed variables:** A text input field containing a German description of a company's services. Below the input, it shows 'Detected language: German' and radio buttons for 'Français', 'Allemand', and 'Italien'. A 'Search' button is at the bottom right of this section.
- First Prediction:** Shows a predicted NACE code '741002' with the description 'Design graphique et communication visuelle' and a confidence gauge at 97.9%. It includes 'Train model' and 'Final code' buttons.
- Second Prediction:** Shows a predicted NACE code '741001' with the description 'Design industriel et de produits' and a confidence gauge at 1.1%. It includes 'Train model' and 'Final code' buttons.
- Third Prediction:** Shows a predicted NACE code '741003' with the description 'Design d'intérieur et design spatial' and a confidence gauge at 1%. It includes 'Train model' and 'Final code' buttons.
- Text translation:** A table showing the original German text and its French translation.
- Text cleaning:** A table showing the original French text and its cleaned version.

Texte original	Texte traduit
Die Gesellschaft erbringt sämtliche Dienstleistungen im Bereich Grafik und Illustration. Ausserdem unterstützt sie Unternehmen, Institutionen und Einzelpersonen in Kommunikationsfragen.	La société fournit tous les services dans le domaine du graphisme et de l'illustration. En outre, elle soutient les entreprises, les institutions et les particuliers en matière de communication.

Texte original	Texte nettoye
La société fournit tous les services dans le domaine du graphisme et de l'illustration. En outre, elle soutient les entreprises, les institutions et les particuliers en matière de communication.	fourn servic domain graphism Illustr soutient entrepris institu particul mati commun

Quality Control of Existing Codes and Activity Descriptions

How can we use an innovative system under development for quality control?

For 50'000 elements of an external administrative database, we focus on cases, where the code and the NOGAuto prediction differ.

The *available description* was the only input for NOGAuto.

UID	code	code description	available description	prediction	code description
xxxxx	711101	Architectural firms	Construction planning	411000	Development of construction projects
xxxxx	711101	Architectural firms	Real estate expertise	683100	Procurement of properties for third parties

Examples of obvious mismatches between code and available description: *code* and *prediction* belong to different economy sectors and the prediction probability is >90%.

UID	code	code description	available description	prediction	code description
xxxxx	014100	Dairy cow farming	Removals and transports	494200	Removal transport
xxxxx	309201	Manufacturing of bicycles	Bicycle repairs and trade	952900	Repair of other consumer goods

Conclusions

We performed **quality assessment** in various ways:

1. We assessed the **overall quality** of the NOGAuto assistance system using global quality measures.
2. We used **quality measures by class** for decision making.
3. We used an innovative system under development for **quality control** of existing codes and activity descriptions.

It is planned that the expert can use NOGAuto to assign NOGA codes to economic activity descriptions :

- in fully or semi-automated mode
- alone or in combination with other expert systems

This process would allow the expert to code in a **timely, convenient** and **efficient** manner.



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Thank you for your attention!



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat 

The conference is partly
financed by the European
Union