Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

ML_SoSi – Machine Learning Soziale Sicherheit

Pilotprojekt im Rahmen der Dateninnovationsstrategie des BFS

Management Summary

EXPERIMENTAL STATISTICS

Neuchâtel, 2023

Herausgeber:	Bundesamt für Statistik (BFS)	**Layoutkonzept:**	Sektion PUB
Redaktion:	Luzius von Gunten, Frank Schubert, Brandon Qorri Gonzalez, Athanassia Chalimourda	**Download:**	www.statistik.ch
Themenbereich:	00 Statistische Grundlagen	**Copyright:**	BFS, Neuchâtel 2023 Wiedergabe unter Angabe der Quelle für nichtkommerzielle Nutzung gestattet
Originaltext:	Deutsch		

ESPACE DE L'EUROPE 10 CH-2010 NEUCHÂTEL WWW.STATISTIK.CH

Management Summary

Arbeitslosigkeit kann für die betroffenen Personen sehr unterschiedlich verlaufen. Die so entstehenden Verlaufsbiographien sind unter anderem geprägt durch (wiederholte) Sozialleistungsbezüge aus dem System der Sozialen Sicherheit (Arbeitslosen-, Invalidenversicherung, Sozialhilfe), Wiedereintritt in die Erwerbsarbeit oder auch Rückzug aus dem Erwerbsleben und Migration. Im Pilotprojekt «ML_SoSi» werden Angaben zu den individuellen Verläufen unter Anwendung induktiver statistischer Methoden analysiert und typische Verlaufsmuster identifiziert. Neben den so erzielten Resultaten ist es das Ziel des Projektes, einen datengetriebenen Ansatz zur Analyse individueller Verläufe in Längsschnittdaten für die öffentliche Statistik zu entwickeln.

Als Datenbasis dient ein anonymisierter verknüpfter Datensatz, der monatsgenaue Informationen zu individuellen Sozialleistungsbezügen aus der Sozialhilfe (SH), der Invalidenversicherung (IV) und der Arbeitslosenversicherung (ALV), sowie zur Erwerbstätigkeit (IK) enthält. Im Rahmen dieses Berichts wird für diesen Datensatz die Abkürzung „SHIVALV+IK“ verwendet. Zur Grundgesamtheit zählen Personen zwischen 18 und 65 Jahren, die im Zeitraum 2010–2015 neu Taggelder der Arbeitslosenversicherung (ALV) beziehen. Die Analyse wird auf der Basis von Jahreskohorten umgesetzt. In die Analyse einbezogen werden Informationen zum Bezug von Sozialversicherungs- und Sozialhilfeleistungen sowie zur Erwerbstätigkeit während den folgenden 48 Monaten bzw. 4 Jahren.

In der methodischen Umsetzung werden zunächst mit der Kohorte 2010 in einem zweistufigen Sequenzclusteringverfahren (unsupervised machine learning) typische Verlaufsmuster identifiziert und mittels grafischer Darstellungen («State Distribution Plots») sowie Verlaufsindikatoren analysiert und inhaltlich interpretiert. Anschliessend wird diese initiale Clusterlösung auf die Kohorten der Folgejahre 2011–2015 mittels «supervised machine learning» übertragen (Prädiktion). Bei jeder Übertragung wird dabei die Validität des Modells geprüft, wobei verschiedene Kriterien evaluiert werden.

Mit diesem Vorgehen liegt der Fokus auf der Erkennung und Analyse aggregierten typischen Verlaufsmustern und deren Übertragung auf weitere Kohorten. Die Nutzbarmachung individueller Vorhersagen zu jedweden Zwecken wird ausgeschlossen.

Insgesamt zehn Cluster zur Beschreibung der typischen Verlaufsmuster von neuen Arbeitslosentaggeldbeziehender wurden identifiziert (Tabelle 1), die mehrheitlich auch im Kohortenvergleich 2010-2015 stabil bleiben (8 von 10 der typischen Verlaufsmuster). Inhaltlich zeichnen sich mehrere Cluster ab, in denen sich die Personen nach einer Phase des Bezugs von Arbeitslosentaggeld wieder in den Arbeitsmarkt integrieren. Dabei unterscheiden sich die Cluster nach der Dauer des Taggeldbezugs (Cluster 1 und 2), nach Vorhandensein einer Zwischenverdienstphase (Cluster 3) sowie durch mehrfache ALV-Bezugsperioden mit zwischenzeitlicher Erwerbstätigkeit (Cluster 4). Daneben entstehen Cluster, die klare Tendenzen entweder zum dauerhaften Bezug von IV-Renten oder von Leistungen der Sozialhilfe zeigen (Tabelle 2; Cluster 5, 6, 7, 8 und 9). Hierzu gehören zwei Cluster mit Neubezug dieser Leistungen (Cluster 5, 9) und zwei Cluster, mit jeweils ausgeprägten Phasen mit ergänzendem Erwerbseinkommen (Cluster 6, 7) sowie ein Cluster, mit Personen, die bereits vor dem Bezug von Arbeitslosentaggeld dauerhaft oder wiederholt auf Sozialhilfe angewiesen waren. Ein letztes Cluster schliesslich vereint diejenigen Personen, die während des Beobachtungszeitraumes dauerhaft nicht mehr von den untersuchten Systemen (SH, IV, ALV, IK/Erwerb) erfasst werden (Cluster 10). Das Projekt hat gezeigt, dass das «Sequence Clustering» ein vielversprechendes Verfahren ist, um inhaltlich valide und analytisch relevante Resultate zu erzeugen. Es erlaubt eine deutliche Verringerung der Komplexität der Verlaufsdaten und erweitert damit die Analysemöglichkeiten durch die Erkennung von Muster, die deduktiv nicht antizipiert werden konnten.

Um diese Informationen für das Publikum der öffentlichen Statistik, unter anderem für die politische Steuerung, noch relevanter zu machen, sind Zeitreihendaten von grosser Bedeutung. Die initiale Clusterlösung kann jedoch nicht ohne weiteres in einer neuen Kohorte reproduziert werden. Im vorliegenden Projekt wurde diese Schwierigkeit gelöst indem die initiale Lösung mittels Prädiktion auf neue Kohorten übertragen wurden. Dieses Vorgehen funktioniert gut und die Kriterien, nach welchen entschieden wurde, ab welchem Zeitpunkt die Übertragung nicht mehr valide ist, hat sich in diesem Fall bewährt. Die Erkenntnisse haben einen konkreten Mehrwert für die statistische Standardproduktion, sowohl bezüglich der neu entwickelten Längsschnittindikatoren und deren Visualisierung als auch für die Bildung von beschreibenden, quantitativen Verlaufsprofilen[1].

Erkenntnisse, Möglichkeiten und Beschränkungen bei der Anwendung von datengetriebenen Methoden in der öffentlichen Statistik werden im Bericht vertieft diskutiert. Auf der Basis von key learnings werden Empfehlungen für ähnlich gelagerte Projekte im BFS dargestellt. Die Schlussfolgerungen münden in einem generischen, induktiven Analyseansatz für individuelle Verlaufsdaten in der Statistik-Produktion im BFS.

[1] Siehe Publikation «Verläufe im System der sozialen Sicherheit 2021»: https://www.bfs.admin.ch/bfs/de/home/statistiken/soziale-sicherheit/analysen-verlaeufe-system/analysen.assetdetail.25385873.html