

Future Lab: Einsatz von KI in der Datenaufbereitung

Julia Schieber-Knöbl

Stv. Bereichsleitung Volkswirtschaftliche Gesamtrechnungen

Koblenz, 03.06.2026

www.statistik.at

Unabhängige Statistiken für faktenbasierte Entscheidungen

Rahmen und Ziel des Projektes

The image features a blue-tinted photograph of a modern building's interior, showing a multi-level atrium with glass railings and potted plants. On the right side, there is a clear view through a window of a modern building facade with a grid of windows and balconies.

Future Lab - Auftrag und Zielsetzung

- Hausweites Projekt bei Statistik Austria
- Fachbereiche reichen Ideen ein, 2 (zusammengefasste) Projekte pro Jahr werden umgesetzt
- innovative Projekte initiieren
- Fokus auf künstliche Intelligenz (KI) und maschinelles Lernen (ML)
- Routineaufgaben automatisieren und Datenqualität verbessern
- Projektorientiertes Arbeiten: Meilensteine, Ziele etc.
- Kollaboratives Arbeiten: Methodik gemeinsam mit Fachbereichen
- Ziel: Mehrwert generieren, langfristige **Integration der erfolgreichen Projekte in den regulären Produktionsprozess**

Einreichung: Datenextraktion Jahresabschlüsse

- Gemeinsames Projekt mit Unternehmensstatistiken (LSE)
- Teil von zusammengefassten Einreichungen für Datenextraktion aus PDFS

Gewinn- und Verlustrechnung 2022/2023

	2022/2023	Vorjahr
	EUR	TEUR
1. Umsatzerlöse	781.279.039,21	590.208
2. Aktivierte Eigenleistungen	0,00	1
3. Sonstige betriebliche Erträge	3.622.334,95	3.352
4. Aufwendungen für Material und sonstige bezogene Herstellungsleistungen	-535.362.205,82	-431.636
5. Personalaufwand	-13.786.897,74	-12.740
6. Abschreibungen	-124.033.535,09	-120.429
<small>(davon aus außerplanmäßigen Abschreibungen EUR 384.129,31; Vorjahr TEUR 521)</small>		
7. Sonstige betriebliche Aufwendungen	-25.676.557,80	-25.698
8. Zwischensumme aus 1 - 7 (Betriebsergebnis)	86.042.177,71	3.058

Ausgangslage

- Jahresabschlüsse (JAB) zunehmend wichtige Quelle für das aktuellste Jahr in der RGR- und VGR-Jahresrechnung
 - Besonders im Bereich Energiewesen, Bau
- Zielgrößen Produktionswert, Vorleistungen und Bruttowertschöpfung lassen sich (meistens zuverlässig) aus Gewinn und Verlustrechnung (GuV) ableiten
- Gesetzliche Regelung in Ö:
 - JAB müssen spätestens 9 Monate nach Bilanzstichtag ans Firmenbuch übermittelt werden
 - Eigentlich maschinell lesbar (xml), aber niedrige compliance

→ Problem: oft nur pdf, daher aufwändige „händische“ Berechnung
→ Beschränkung auf wenige große Einheiten nötig

Projektziele

- Automatisierte Lösung, d.h. Datenextraktion aus den pdf-JAB nach Gesamtkostenverfahren (GKV)
- Übersichtliche Ableitung der BWS inkl. Teilpositionen
- Dadurch: Zeitersparnis, viel mehr Einheiten können korrekt berücksichtigt werden, v.a. hinsichtlich Vorleistungen bzw. Nettoquote
- Einzeldaten sind näher an der LSE
- Weitere Steigerung der Qualität der V/RGR-Daten für das aktuellste Jahr und Reduktion des Revisionsausmaßes

Umsetzung



Training des Modells in Google Vertex AI

- Ca 100 Jahresabschlüsse als PDF zur Verfügung gestellt
- Trainingsdaten: mehrere PDFs im Vertex AI Labeling Interface markiert
- Inhalte (Texte und Zahlenwerte), räumliche Positionierung (x-/y-Koordinaten) und Seitenzuordnung erfasst
- Auf dieser Basis 1. Trainingsmodell für GKV-Abschlüsse (Gesamtkostenverfahren) entwickelt
- Weitere JAB hinzugefügt zum Lernen
- Extraktion der relevanten Datenfelder mittels eines vortrainierten LLM (Large Language Model), welches auf die bereitgestellten PDFs angepasst wurde

Ablauf mittels R-Programm

- Arbeitsablauf in R implementiert:
 - Zugriff auf PDFs (Ordner) und Hochladen in Google Cloud Speicher
 - Verarbeitung durch Extraktionsmodell
 - Daten aus Cloud herunterladen und in strukturierter Form als Excel-Datei aufbereiten
 - Struktur orientiert sich an LSE
 - Auf Basis der extrahierten GuV-Positionen automatisierte Berechnung der Aggregate: P1, P2, BWS
 - Weitere erfasste Metadaten: Firmenbuchnummer, Unternehmensname, Bilanzstichtag, Geschäftsjahr, Währungseinheit (€, Tsd. €...)
- Modell kann durch zusätzliche Trainingsdaten weiter verbessert werden

Probleme und Einschränkungen

- Umsatzkostenverfahren (UKV): nicht genügend Trainingsdaten, Problem bei Vorleistungen
→ Training läuft derzeit, Ergebnisse Ende des Jahres
 - GuV mit „Rohergebnis“ → P1 und P2 können verzerrt sein
 - Bereiche mit speziellen GuV-Positionen: Leasing, Banken, Versicherungen, Realitätenwesen
 - Sonstiger Betriebsaufwand (SBA): enthält oft Bestandteile, die nicht VGR-relevant sind
 - Kritische Anwendung bzw. Plausibilisierung nötig
 - Kosten: derzeit ca. 3 Cent pro Seite
-
- Manueller Aufwand beim Herunterladen der pdfs aus dem Web-UR und Abspeichern in Ordner
→ Lösung durch LSE und IT im Laufe des Jahres

Aktuelle Arbeiten bis Ende 2026

- Ausweitung des Modells:
 - Bilanz
 - GuV nach UKV
 - Anlagenspiegel
 - Evtl. Leasing, Banken, Realitäten,
 - Als nicht sinnvoll herausgestellt: Anhänge
- Aufbau einer Datenbank, die **sämtliche** GuV enthält
 - Xml Schnittstelle im Firmenbuch, gesetzliche Verschärfung ab 1.1.2026
 - Wenn nur pdf verfügbar: Datenextraktion mittels Modell

Konkrete Auswirkungen in den VGR

- RGR: Änderung des Jahresarbeitsablaufes möglich, Einbau und Plausibilisierung der neuen Quelle jeweils im Oktober und November für das aktuellste Jahr (Wertschöpfung und Bruttoregionalprodukt t+12 Rechnung)
- Strukturierte Auswertungen evtl. auch für Jahresrechnung verwendbar (Abdeckung zu unterschiedlichen Zeitpunkten wird derzeit analysiert)
- Für Einheiten, die keine GuV haben: KJE bzw. UVA für P1, Schätzung für P2 anhand Vorjahresstruktur (LSE) – „individuelle Nettoquoten“
- R-Programm für pdf-Datenextraktion kann vom Fachbereich verwendet werden

Rückfragen bitte an

Julia Schieber-Knöbl

julia.schieber-knoebl@statistik.gv.at

Christian Leupold

christian.leupold@statistik.gv.at

STATISTIK AUSTRIA

Guglgasse 13, 1110 Wien

 statistik_at

 statistik_at

 statistics-austria

 statistik_austria

 statistik.at

Unabhängige Statistiken für faktenbasierte Entscheidungen

