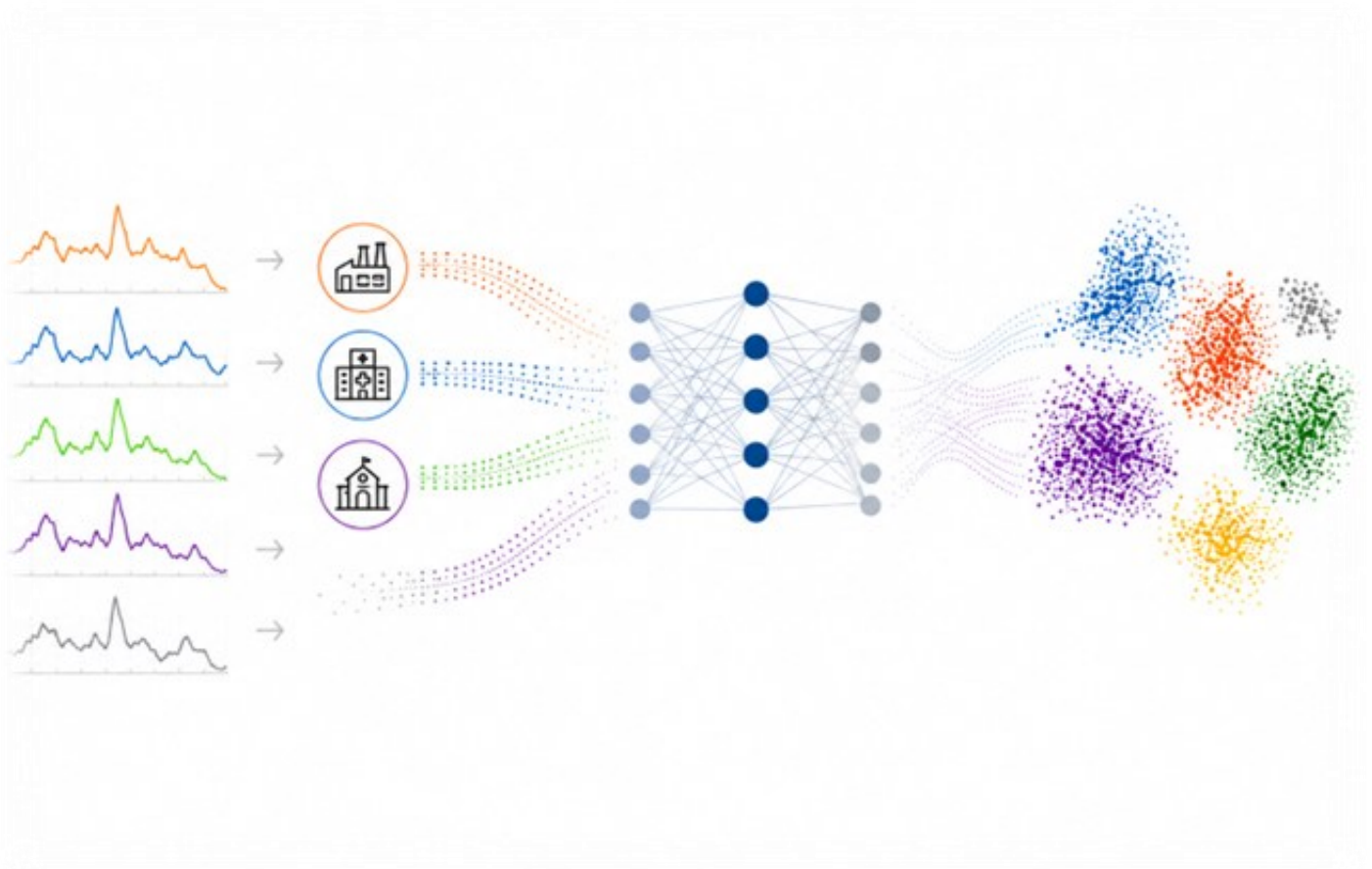




Bericht vom 01. Mai 2026

Verbrauchsmonitoring Schweiz: Clustering und KI-gestützte Branchenzuordnung

Eine branchenbezogene Verbrauchsanalyse





Auftraggeberin:

Bundesamt für Energie BFE
CH-3003 Bern
www.bfe.admin.ch

Auftragnehmer/in:

Supercomputing Systems AG
Technoparkstrasse 1, CH-8005 Zürich
www.scs.ch

Autor/in:

Janine Biner, Supercomputing Systems AG, janine.biner@scs.ch

BFE-Bereichsleitung: Lucas Tochtermann, lucas.tochtermann@bfe.admin.ch

BFE-Programmleitung: Dr. Matthias Galus, matthias.galus@bfe.admin.ch

BFE-Vertragsnummer: SI/600614-01

Für den Inhalt und die Schlussfolgerungen sind ausschliesslich die Autoren dieses Berichts verantwortlich.

Bundesamt für Energie BFE

Pulverstrasse 13, CH-3063 Ittigen; Postadresse: Bundesamt für Energie BFE, CH-3003 Bern
Tel. +41 58 462 56 11 · Fax +41 58 463 25 00 · contact@bfe.admin.ch · www.bfe.admin.ch

Management Summary

Das Bundesamt für Energie (BFE) erhöht stetig die Transparenz in der Schweizer Energieversorgung. Dazu betreibt es unter anderem das nationale Energiedashboard mit aktuellen Informationen zur Schweizer Energieversorgung in den Bereichen Strom, Gas, Preise und vielem mehr. In der Vergangenheit und insbesondere vor dem Hintergrund einer angespannten Stromversorgungslage im Winter 2022 zeigte sich ein dringender Bedarf, Stromverbrauchsdaten nach Branchen und wirtschaftlichen Tätigkeiten zu differenzieren, um allfällige nationale Massnahmen gezielt, wirksam und mit möglichst geringen negativen Auswirkungen umsetzen zu können. Dies liess sich ad hoc nicht rasch und belastbar realisieren, da viele Daten kurzfristig nicht zur Verfügung standen. Entsprechend wurden Arbeiten lanciert, um diesen Herausforderungen künftig zu begegnen.

Vor diesem Hintergrund wurde diese Pilotstudie durchgeführt und es wurde geprüft, inwieweit sich von Verteilnetzbetreibern (VNB) gelieferte Lastprofile mit Metadaten verknüpfen und mittels NOGA-Codes (Nomenclature Générale des Activités économiques¹) wirtschaftlichen Tätigkeiten zuordnen lassen. NOGA-Codes dienen als etablierter Standard zur Brancheneinteilung und können bei erfolgreicher Verknüpfung mit Stromverbrauchsdaten als Grundlage für Clusterbildungen, Machine Learning (ML) gestützte Klassifikationen und statistische Auswertungen dienen. Die Datenverarbeitung erfolgte auf Basis entsprechender Vereinbarungen mit den beteiligten VNB und jederzeit unter Einhaltung datenschutzrechtlicher Vorgaben.

Teilprojekt 1: Metadatenbasierte Zuordnung von Verbrauchsdaten zu NOGA-Codes

Fünf VNB stellten Daten zur Verfügung, aus denen Verbrauchspunkte strukturiert aufbereitet und per Unternehmens-Identifikationsnummer (UID)-Abfrage einem NOGA-Code zugeordnet wurden. Auf Basis der ersten zwei Stellen des NOGA-Codes konnten Messpunkte branchenbezogen gruppiert und erste Auswertungen über die Verteilung der Branchen sowie deren relativen Stromverbrauch erstellt werden. Dieser Ansatz zeigte, dass eine Zuordnung der Messpunkte zu wirtschaftlichen Tätigkeiten und damit eine branchenbezogene Betrachtung grundsätzlich möglich ist. Für schweizweit repräsentative Aussagen sind jedoch zusätzliche Daten von weiteren VNB sowie harmonisierte Datenstrukturen und standardisierte Metadaten notwendig.

Teilprojekt 2: KI-gestützte Klassifikation von Verbrauchsprofilen

Untersucht wurde, ob sich Lastprofile mittels Machine Learning automatisiert NOGA-Codes zuordnen lassen. Die KI-gestützte Klassifikation baute auf den im ersten Teilprojekt aufbereiteten und mit NOGA-Codes angereicherten Lastprofilen auf. Erste Tests zeigten, dass Profile einzelner Branchen (z. B. Handel, Industrie) besser unterscheidbar waren als andere. Die Ergebnisse zeigen, dass die Vielzahl an NOGA-Codes, die begrenzte Datenbasis und unausgewogene Verteilungen zwischen Branchen die Modellqualität begrenzen. Reduzierte Kategoriensets sowie Feature-Engineering konnten die Genauigkeit teilweise verbessern, zeigten aber auch, dass für robuste Modelle erheblich umfangreichere und standardisierte Daten erforderlich sind.

¹ [NOGA 2025 | KUBB - Coding assistance application](#)

Erkenntnisse

Die Arbeiten dieser Pilotstudie zeigen, dass eine Zuordnung von NOGA-Codes zu Messpunkten und damit branchenbezogene Analysen des Stromverbrauchs grundsätzlich machbar sind. Bei regelmässiger Anwendung könnte ein solcher Auswertungsansatz zudem dazu beitragen, strukturelle wirtschaftliche Veränderungen in der Schweiz und in einzelnen Regionen sichtbar zu machen und Entwicklungen im Stromverbrauch nach Wirtschaftszweigen über die Zeit zu verfolgen. Gleichzeitig wurde deutlich, dass insbesondere die Qualität der Messpunktmetadaten sowie Datenverfügbarkeit und Standardisierung bzw. Harmonisierung entscheidend für Genauigkeit und Aussagekraft sind. Machine Learning Ansätze sind grundsätzlich vielversprechend, aktuell jedoch nur für eingeschränkte Anwendungsfälle belastbar. Die Pilotstudie versteht sich als methodischer Konzeptnachweis und nicht als ausgereifte Lösung.

Im Rahmen der Pilotstudie wurden zwei Software-Werkzeuge entwickelt und dem BFE übergeben, welche die NOGA-Zuordnung von Messpunkten sowie die prototypische KI-gestützte Klassifikation unterstützen. Beide Code-Repositories befinden sich im BFE-GitHub-Account², stehen im Sinne des Open-Source-Prinzips zur Verfügung und bilden einen integralen Bestandteil der Projektergebnisse. Die Ergebnisse schaffen damit auch eine Grundlage für mögliche Folgearbeiten bei VNB zur Bereinigung ihrer Datenbasis sowie für die Weiterentwicklung der Applikation des nationalen Energiedashboards des BFE und datenbasierter Transparenz- und Analysegrundlagen im Strombereich.

Empfehlungen für eine Weiterentwicklung

- Verbesserung der Datenqualität bei den Verteilnetzbetreibern sowie Harmonisierung und Standardisierung von Lastprofilen und Metadaten, auch im Hinblick auf den Aufbau einer nationalen Datenplattform mit einer einheitlichen und hochwertigen Stammdatenbasis.
- Erweiterung der Datengrundlage für den Machine Learning Ansatz durch die Einbindung zusätzlicher VNB und eine breitere Abdeckung der Wirtschaftszweige
- Weiterentwicklung der ML-Ansätze bei besserer Datenbasis
- Prüfung einer abgestuften Klassifikationsstrategie (z. B. erst Hauptbranchen, danach Unterkategorien)

² <https://github.com/SFOE/NOGA-Klassifizierung>

Inhaltsverzeichnis

Management Summary.....	3
Inhaltsverzeichnis	5
1. Ausgangslage.....	6
2. Ziele.....	6
3. Teilprojekt 1: Zuordnung von Stromverbrauchsdaten zu NOGA-Codes	7
3.1. Vorgehen.....	7
3.1.1. Vorgehen bei der Datenauswertung.....	9
3.1.2. Technische Umsetzung und Datenverarbeitung.....	14
3.1.3. Datenschutz.....	14
3.2. Resultate.....	15
3.3. Zusammenfassung und Ausblick.....	17
4. Teilprojekt 2: KI-gestützte Zuordnung von Verbrauchskurven zu NOGA-Abschnitten.....	18
4.1. Zielsetzung und Datengrundlage.....	18
4.2. Vorgehen und Resultate	19
4.2.1. Top 5 NOGA-Abschnitte.....	21
4.2.2. Top 4 NOGA-Abschnitte und Sammelkategorie «Andere»	23
4.2.3. Feature Engineering.....	26
4.2.4. LSTM-Autoencoder.....	28
4.2.5. Alternative BFE-NOGA-Kategorien.....	29
4.3. Diskussion und Ausblick	31
5. Software-Artefakte und Repositories.....	32
6. Zusammenfassung und Ausblick.....	33
7. Appendix.....	34
NOGA-Abteilung aus Lastprofilen.....	34
Top 4 NOGA-Abschnitte und Sammelkategorie «Andere».....	38
Feature Engineering.....	39

1. Ausgangslage

Das Bundesamt für Energie (BFE) arbeitet daran, die Transparenz in der Schweizer Energieversorgung weiter zu erhöhen. Dazu stellt es unter anderem mit dem nationalen Energiedashboard aktuelle Informationen zur Energieversorgung in den Bereichen Strom, Gas, Preise und weiteren Themen bereit. Insbesondere vor dem Hintergrund der angespannten Stromversorgungslage im Winter 2022 zeigte sich ein dringender Bedarf, Stromverbrauchsdaten nach Branchen und wirtschaftlichen Tätigkeiten differenzierter auszuwerten, um mögliche Massnahmen fundierter, wirksamer und mit möglichst geringen negativen Auswirkungen gestalten zu können.

Die bisher im BFE verwendete Verbrauchsklassifizierung auf dem Energiedashboard unterscheidet drei Kategorien: Private, KMU und Grossverbraucher. Diese Einteilung bietet einen ersten Überblick, stösst jedoch an ihre Grenzen, wenn es um branchenbezogene Aussagen geht. Für eine differenziertere Beurteilung wirtschaftlicher Tätigkeiten ist daher eine feinere Unterteilung erforderlich. Diese ermöglicht fundiertere Analysen und eine präzisere Planung und Umsetzung von Massnahmen im Bereich der Versorgungssicherheit.

Für diese Verfeinerung mussten die von VNB gelieferten Lastprofile mit Metadaten verknüpft werden. Eine zentrale Rolle spielten dabei NOGA-Codes, die Unternehmen ihrer wirtschaftlichen Tätigkeit zuordnen. Sie ermöglichten es, Verbrauchsdaten nach Wirtschaftszweigen zu aggregieren, systematisch zu vergleichen und datenschutzkonform auszuwerten.

Neben klassischen Datenanalysemethoden sollten auch moderne Verfahren geprüft werden, die eine automatisierte Zuordnung erlauben. Machine Learning Ansätze konnten bei ausreichender Datenqualität Muster in Lastprofilen erkennen und diese wirtschaftlichen Tätigkeiten zuordnen.

Die vorliegende Pilotstudie diente dazu, erste Erfahrungen mit der Bereitstellung, Aufbereitung und Klassifikation von Verbrauchsdaten zu sammeln. Sie schafft Grundlagen für die Weiterentwicklung entsprechender datenbasierter Analyseansätze und zeigt auf, welche Anforderungen an Datenumfang, Standardisierung und Qualität erfüllt sein müssen, damit eine verlässliche und automatisierte branchenbezogene Verbrauchsanalyse möglich wird.

2 Ziele

Ziel der Pilotstudie ist es, die Machbarkeit einer Zuordnung von Messpunkten zu NOGA-Codes und branchenbezogenen Analysen von Stromverbrauchsdaten zu prüfen. Im Zentrum steht die Frage, ob sich die von VNB gelieferten Lastprofile mit Metadaten verknüpfen und mittels NOGA-Codes wirtschaftlichen Tätigkeiten zuordnen lassen. Neben der metadatenbasierten Zuordnung wird untersucht, ob Machine Learning Methoden grundsätzlich in der Lage sind, Muster in Verbrauchskurven zu erkennen und diese automatisiert einer wirtschaftlichen Tätigkeit zuzuweisen. Dabei soll auch evaluiert werden, welche Merkmale für eine erfolgreiche Klassifikation relevant sind und welche Anforderungen an Datenumfang, Datenqualität und Standardisierung erfüllt sein müssen, um belastbare Ergebnisse zu erzielen.

Die Pilotstudie versteht sich als methodischer Konzeptnachweis. Sie soll erste Erfahrungen im Umgang mit heterogenen Datenbeständen liefern, grundlegende Prozessschritte für die Zuordnung von Messpunkten zu wirtschaftlichen Tätigkeiten aufzeigen und eine Einschätzung ermöglichen, ob und unter welchen Rahmenbedingungen eine verbesserte Transparenz in der Stromversorgung durch

branchenbezogene Auswertungen des Stromverbrauchs realistisch umgesetzt werden kann. Ebenso soll sie aufzeigen, welche Datenstrukturen, Formatvorgaben und Metadaten künftig erforderlich wären, um eine einheitliche Stammdatenbasis und entsprechende Analysen technisch und organisatorisch zu stützen.

Nicht Ziel der Studie ist die Entwicklung einer abschliessenden, produktionsreifen Lösung zur automatisierten Klassifizierung oder eine schweizweite, vollständig repräsentative Branchenanalyse. Ebenso umfasst der Projektumfang weder eine Optimierung der NOGA-Systematik noch eine Bewertung von Energieeffizienz, regulatorischen Massnahmen oder potenziellen Einsparwirkungen. Die Untersuchung stützt sich ausschliesslich auf verfügbare Daten, betrachtet daher nur ausgewählte Kategorien und ersetzt keine detaillierte nationale Gesamtstatistik.

Im Ergebnis soll die Studie eine belastbare Grundlage schaffen, um Chancen, Grenzen und Voraussetzungen einer branchenbezogenen Stromverbrauchsanalyse zu beurteilen und Entscheidungsträgern aufzuzeigen, wie datenbasierte Transparenz- und Analysegrundlagen im Bereich Strom künftig weiterentwickelt werden könnten.

3 Teilprojekt 1: Zuordnung von Stromverbrauchsdaten zu NOGA-Codes

Ziel des ersten Teilprojekts war es, den Stromverbrauch von Unternehmen branchenbezogen auszuwerten, indem Messpunkte wirtschaftlichen Tätigkeiten zugeordnet werden. Dazu wurden Verbrauchsdaten mehrerer VNB mit Unternehmensmetadaten verknüpft und anhand von NOGA-Codes klassifiziert. Für die Analyse stellten fünf VNB Daten zur Verfügung.

Der Ansatz war regel- und datenbasiert. Die Zuordnung der Messpunkte zu wirtschaftlichen Tätigkeiten erfolgte nicht anhand der Lastprofile selbst, sondern über vorhandene administrative Informationen. Je nach Datenlage war der NOGA-Code bereits direkt verfügbar oder konnte über die Unternehmens-Identifikationsnummer (UID) ermittelt werden. In Fällen mit unvollständigen Angaben erfolgte die Zuordnung über Unternehmensnamen und Adressinformationen. Dieses Vorgehen ermöglichte es, den Stromverbrauch aggregiert nach Branchen darzustellen und zwischen unterschiedlichen Netzgebieten zu vergleichen.

Die im Teilprojekt entwickelten Prozesse zur Datenaufbereitung, UID-Ermittlung und NOGA-Zuordnung wurden technisch umgesetzt und so weit wie möglich automatisiert. Die entsprechenden Software-Artefakte werden in Kapitel 0 beschrieben.

3.1 Vorgehen

Das Vorgehen zur Zuordnung von Stromverbrauchsdaten zu wirtschaftlichen Tätigkeiten lässt sich in drei grundlegende Szenarien unterteilen, die in Abbildung 1 schematisch dargestellt sind. Welches Szenario zur Anwendung kam, hing davon ab, welche Metadaten die VNB gemeinsam mit den Lastprofilen zur Verfügung stellten.

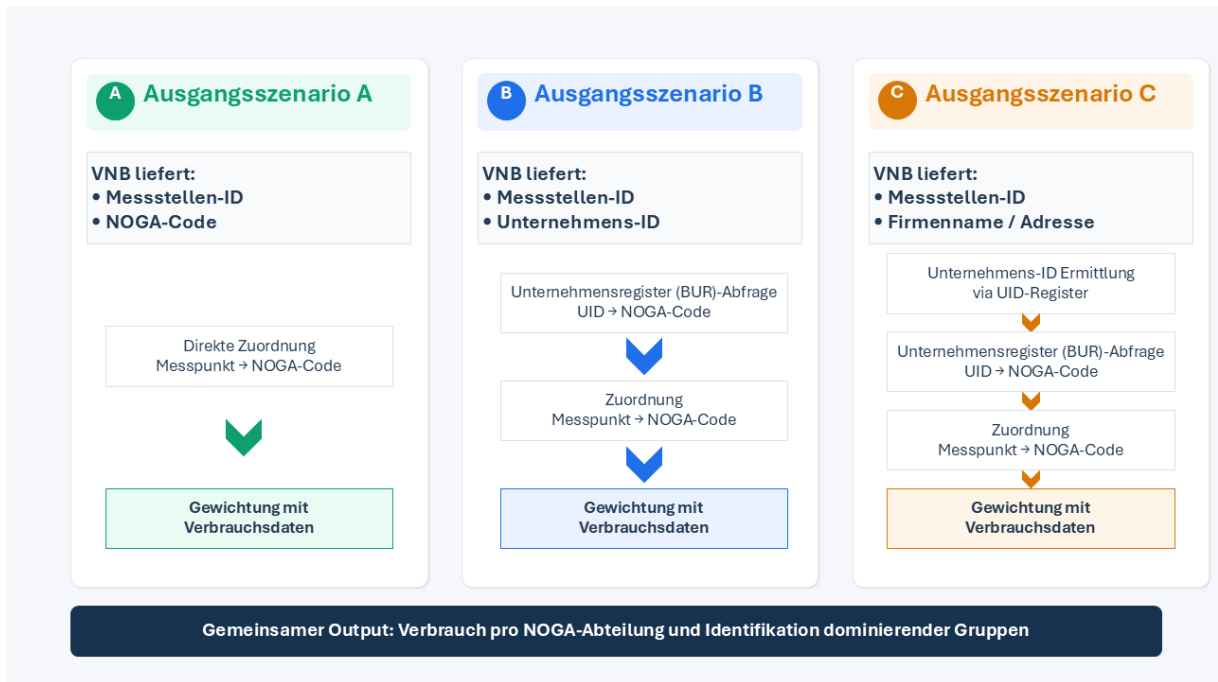


Abbildung 1 Methodik im Detail

Im **Ausgangsszenario A** wurden die Lastprofile bereits direkt mit den zugehörigen NOGA-Codes geliefert. In diesen Fällen war keine zusätzliche Anreicherung der Daten notwendig, sodass die statistische Auswertung der Stromverbrauchsdaten unmittelbar auf Ebene der NOGA-Abteilungen erfolgen konnte.

Im **Ausgangsszenario B** stellten die VNB die Lastprofile zusammen mit der Unternehmens-Identifikationsnummer (UID) der zugehörigen Unternehmen bereit. Mithilfe der UID konnte der entsprechende NOGA-Code über das Betriebs- und Unternehmensregister (BUR) des Bundesamts für Statistik ermittelt werden. Anschliessend erfolgte die Zuordnung zwischen NOGA-Code und Lastprofil, worauf aufbauend die statistischen Auswertungen durchgeführt wurden.

Im **Ausgangsszenario C** war die UID nicht direkt verfügbar. In diesen Fällen wurde die UID zunächst anhand von Unternehmensnamen und Adressinformationen über das Unternehmensregister ermittelt. Nach erfolgreicher Ermittlung der UID entsprach das weitere Vorgehen dem im Szenario B beschriebenen Prozess.

Die drei Szenarien verdeutlichen, wie das methodische Vorgehen flexibel an unterschiedliche Datenlagen der VNB angepasst werden konnte.

3.1.1 Vorgehen bei der Datenauswertung

Datenaufarbeitung

Die Datenauswertung erfolgte in mehreren aufeinander aufbauenden Schritten und basierte auf den von den VNB gelieferten Verbrauchsdaten sowie den zugehörigen Unternehmens- und Adressinformationen. Für jeden VNB wurde zunächst eine konsolidierte Ausgangstabelle erstellt, welche alle relevanten Informationen zu den einzelnen Messpunkten zusammenführt. Eine solche einheitliche Ausgangstabelle lag zu Beginn nicht vor.

Die gelieferten Datensätze variierten stark hinsichtlich Struktur, Vollständigkeit und Detaillierungsgrad. Dies lag insbesondere daran, dass die Metadaten aus unterschiedlichen Systemen der VNB exportiert wurden und je nach VNB unterschiedlich strukturiert, benannt und formatiert waren. Teilweise mussten Verbrauchsdaten und Unternehmensinformationen über Zwischenschlüssel wie Anlagen-IDs oder interne Codestrukturen zusammengeführt werden. Auch Profilkategorien, Datenmodelle und Tabellenstrukturen unterschieden sich zwischen den VNB.

Entsprechend war eine vorgängige Datenaufbereitung erforderlich. Diese umfasste die Vereinheitlichung von Unternehmensnamen, die konsistente Behandlung von Einträgen, die Bereinigung fehlerhafter oder unvollständiger Adressangaben sowie die Harmonisierung unterschiedlicher Datenmodelle in eine einheitliche Struktur. Freitextfelder und zusätzliche Bemerkungen wurden in diesem Schritt nicht weiter berücksichtigt. Grundsätzlich zeigte sich eine heterogene Datenbasis mit uneinheitlichen Strukturen und teilweise erheblichen Qualitätsproblemen. Eine maschinelle Verarbeitung war deshalb erst nach zusätzlichen Aufbereitungs- und Bereinigungsschritten möglich.

Ermittlung der UID

Auf Basis der bereinigten Unternehmens- und Adressdaten wurde anschliessend die UID ermittelt, sofern diese nicht bereits in den gelieferten Daten enthalten war. Die UID dient als eindeutiger Schlüssel zur Verknüpfung der Messpunkte mit dem Betriebs- und Unternehmensregister (BUR) und damit zur Ermittlung des NOGA-Codes. Die Ermittlung der UID erfolgte über den öffentlich verfügbaren UID-Webservice³ des Bundesamtes für Statistik (BFS) anhand von Unternehmensnamen und vollständigen Adressinformationen. Um Fehlzuordnungen zu vermeiden, wurden ausschliesslich Treffer mit hoher Übereinstimmung berücksichtigt.

Die öffentlich verfügbare Abfrage ist auf 20 Einträge pro Minute beschränkt. Daher wurde die Abfragegeschwindigkeit gedrosselt, wodurch die maschinelle Verarbeitung entsprechend länger dauerte.

Zu jedem Suchergebnis wird ein Qualitätsrating vom Unternehmensregister zurückgegeben. Das Qualitätsrating ist nicht dokumentiert. Es scheint vor allem die korrekte Schreibweise des Unternehmensnamens, sowie in zweiter Priorität die Übereinstimmung mit den Adressfeldern zu berücksichtigen. Für die Abfrage wurden nur Qualitätsratings über 90 berücksichtigt (Voreinstellung).

Zur Optimierung des Vorgehens wurden verschiedene Abfragearten und Einschränkungen der Suchergebnisse getestet. Dazu gehörten Suchanfragen nur mit Unternehmensnamen, Suchanfragen mit Unternehmensnamen und Ort sowie zusätzliche Filter auf übereinstimmende Ortsangaben. Dabei zeigte sich, dass eine manuelle Nachbearbeitung rasch zu komplexen Einzelfallbeurteilungen führte.

³ <https://www.uid.admin.ch/>

Der Grund dafür lag vor allem in der Qualität und Aktualität der verfügbaren Metadaten. Wenn Unternehmensnamen, Adressen oder Standortangaben unvollständig, veraltet oder uneinheitlich erfasst sind, kann die automatische Suche im Unternehmensregister keinen eindeutigen Treffer liefern.

Solche Fälle betrafen beispielsweise Firmenübernahmen, Zweigstellen, Firmen in Liquidation, ausländische Hauptsitze oder uneinheitliche bzw. veraltete Registereinträge. So kann etwa ein Messpunkt noch unter einem früheren Firmennamen geführt werden, während im Register bereits der neue Unternehmensname hinterlegt ist. Ebenso kann eine Filiale an einer bestimmten Adresse betrieben werden, während der Registereintrag auf den Hauptsitz verweist. In beiden Fällen sind nicht die Verbrauchsdaten das Problem, sondern die Qualität der zugehörigen Stammdaten.

Auch wenn in einzelnen Fällen durch zusätzliche Recherchen eine Zuordnung möglich gewesen wäre, hätte dies den Prozess erheblich verlangsamt und die Nachvollziehbarkeit der Methodik erschwert. Für das weitere Vorgehen wurde deshalb entschieden, die Suche standardisiert mit Unternehmensnamen und vollständiger Adresse (Strasse mit Hausnummer, Ort mit Postleitzahl) durchzuführen. Damit konnte ein konsistentes, reproduzierbares und für die Zielsetzung der Studie ausreichend robustes Verfahren sichergestellt werden.

Ermittlung des NOGA-Codes

Nach erfolgreicher Identifikation der UID wurde der zugehörige NOGA-Code aus dem Betriebs- und Unternehmensregister (BUR)⁴ des BFS abgerufen. Damit konnte jedem Messpunkt, für den eine UID ermittelt oder bereits mitgeliefert wurde, ein NOGA-Code zugeordnet werden.

Im letzten Schritt wurden die Messpunkte mitsamt ihren Verbrauchsdaten den entsprechenden NOGA-Codes zugeordnet. Auf dieser Basis konnten sowohl die Verteilung der Messpunkte auf die einzelnen Wirtschaftszweige als auch die relativen Verbrauchsanteile ausgewertet werden. Für die eigentlichen Auswertungen wurden die NOGA-Codes anschliessend auf Ebene der NOGA-Abteilungen zusammengefasst. Die Begründung dieser Aggregation wird im folgenden Abschnitt erläutert.

Abteilungen nach NOGA-Code

Zum besseren Verständnis der verwendeten Begriffe ist die hierarchische Struktur der NOGA zu berücksichtigen (Abbildung 2). Die oberste Ebene bildet der NOGA-Abschnitt, der mit einem Grossbuchstaben bezeichnet wird. Darunter folgen die NOGA-Abteilungen mit zwei Ziffern. Weitere Ebenen sind die Gruppen mit drei Ziffern, die Klassen mit vier Ziffern und die Arten mit sechs Ziffern. Im vorliegenden Bericht wird der Begriff NOGA-Code für den vollständigen sechsstelligen Code verwendet, der einem Unternehmen im BUR zugewiesen ist. Die NOGA basiert auf der europäischen Systematik NACE (Nomenclature statistique des activités économiques dans la Communauté européenne) und berücksichtigt zusätzlich schweizerische Besonderheiten.

Für die vorliegende Analyse war jedoch nicht die vollständige sechsstellige Detailtiefe erforderlich. Die sechsstelligen NOGA-Codes wurden deshalb auf die ersten beiden Ziffern reduziert und damit auf Ebene der NOGA-Abteilungen aggregiert (z.B. 104100 → 10). Diese Aggregation ermöglichte eine

⁴ <https://www.burweb2.admin.ch/BurWeb/Login.aspx>

konsistente branchenbezogene Auswertung auch bei begrenzter Datenbasis und stellte eine bessere Vergleichbarkeit zwischen den Netzgebieten der beteiligten VNB sicher.



Abbildung 2: Systematik der Wirtschaftszweige nach NOGA

Die NOGA-Abteilungen lassen sich wiederum übergeordneten NOGA-Abschnitten zuordnen. So gehören beispielsweise die Abteilungen 10 bis 33 zum Abschnitt C „Verarbeitendes Gewerbe/Herstellung von Waren“. Tabelle 1 zeigt diese Zuordnung der verwendeten NOGA-Abteilungen zu übergeordneten Wirtschaftsbereichen bzw. Abschnitten.

Die alternative Brancheneinteilung aus dem BFE-Bericht *Energieverbrauch in der Schweiz 2021 nach NOGA Stufe 2 (Abteilungen)*⁵ wurde nicht übernommen, da sie für diese Studie methodisch weniger geeignet war. Der BFE-Bericht unterscheidet 19 Branchen, die teilweise auf zweistelligen NOGA-Abteilungen und teilweise auf vierstelligen NOGA-Klassen beruhen. Eine Anwendung dieser Einteilung hätte deshalb zusätzliche Zuordnungen auf unterschiedlichen Detailstufen erfordert. Zudem werden einzelne Branchen in dieser Klassifikation bewusst nicht berücksichtigt, was für die vorliegende Analyse nachträgliche Ergänzungen notwendig gemacht hätte.

In den Kapiteln drei und vier desselben Dokuments werden weitere alternative Gruppierungen verwendet, darunter eine Unterteilung in 72 Abteilungen sowie eine Klassifikation in 41 NOGA-Gruppen. Beide Varianten erwiesen sich für die Zielsetzung dieses Projekts als zu granular.

⁵ <https://www.bfe.admin.ch/bfe/de/home/versorgung/statistik-und-geoda-ten/energiestatistiken/teilstatistiken.exturl.html/aHR0cHM6Ly9wdWJkYi5iZmUuYWWRtaW4uY2qyZGUvcHVib-GljYX/Rpb24vZG93bmxvYWQvMTEyMjA=.html>

Aus diesen Gründen wurde für die vorliegende Analyse die einheitliche Aggregation auf Ebene der zweistelligen NOGA-Abteilungen gewählt. Sie bietet einen geeigneten Kompromiss zwischen Detaillierungsgrad, Vergleichbarkeit und Datenverfügbarkeit und ist für die verfügbare Datenbasis robuster als die geprüften alternativen Gruppierungen.

Tabelle 1: Abschnitte nach NOGA-Abteilung

Bezeichnung Abschnitt	Abteilungen
Verarbeitendes Gewerbe/Herstellung von Waren	10-33
Handel; Instandhaltung und Reparatur von Kraftfahrzeugen	45-47
Verkehr und Lagerei	49-53
Öffentliche Verwaltung, Verteidigung; Sozialversicherung	84
Grundstücks- und Wohnungswesen	68
Gesundheits- und Sozialwesen	86-88
Erbringung von Finanz- und Versicherungsdienstleistungen	64-66
Gastgewerbe/Beherbergung und Gastronomie	55-56
Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen	69-75
Wasserversorgung; Abwasser- und Abfallentsorgung und Beseitigung von Umweltverschmutzungen	36-39
Baugewerbe/Bau	41-43
Energieversorgung	35
Erbringung von sonstigen Dienstleistungen	94-96
Land- und Forstwirtschaft, Fischerei	1-3
Information und Kommunikation	58-63
Erziehung und Unterricht	85
Kunst, Unterhaltung und Erholung	90-93
Erbringung von sonstigen wirtschaftlichen Dienstleistungen	77-82
Bergbau und Gewinnung von Steinen und Erden	5-9
Private Haushalte mit Hauspersonal; Herstellung von Waren und Erbringung von Dienstleistungen durch private Haushalte für den Eigenbedarf ohne ausgeprägten Schwerpunkt	97-98
Exterritoriale Organisationen und Körperschaften	99

Verbrauch nach NOGA-Abteilung

Im darauffolgenden Schritt wurden die Verbrauchsdaten mit den bereits gelieferten oder im Projekt ermittelten NOGA-Codes verknüpft. Die Zuordnung erfolgte über die Messpunkt-ID. Damit konnten sowohl Lastprofile, bei denen der NOGA-Code bereits vorhanden war, als auch Lastprofile, bei denen der NOGA-Code über UID, Unternehmensname oder Adresse ermittelt wurde, einheitlich weiterverarbeitet werden. Für die Auswertung wurden die NOGA-Codes gemäss der zuvor beschriebenen Systematik auf Ebene der NOGA-Abteilungen aggregiert. Auf dieser Basis konnte der Stromverbrauch nach NOGA-Abteilungen ausgewertet werden.

Zur Analyse der Verbrauchsstruktur wurden die Verbrauchsdaten pro VNB zunächst über den gesamten verfügbaren Zeitraum von 2018 bis 2023 nach NOGA-Abteilungen aggregiert und visualisiert. Damit konnte aufgezeigt werden, welche Wirtschaftszweige welchen Anteil am Stromverbrauch im jeweiligen Netzgebiet hatten. Ergänzend dazu erfolgte eine jährliche Auswertung der Verbrauchsdaten pro VNB, um zu prüfen, ob sich die Verteilung des Stromverbrauchs zwischen den Wirtschaftszweigen über die Zeit veränderte.

Die Gegenüberstellung der jährlichen Auswertungen pro VNB zeigte, dass sich die relative Verteilung des Stromverbrauchs nach NOGA-Abteilungen über die betrachteten Jahre nur geringfügig veränderte. Da die jährliche Betrachtung keine wesentlichen strukturellen Veränderungen sichtbar machte, wurde für die weiteren Analysen jeweils ein repräsentatives Jahr herangezogen.

3.1.2 Technische Umsetzung und Datenverarbeitung

Die im Teilprojekt angewendeten Verarbeitungsschritte wurden in einer Python-basierten Pipeline umgesetzt, um die Datenaufbereitung, UID-Ermittlung und NOGA-Zuordnung reproduzierbar durchführen zu können. Die Pipeline unterstützt die in Kapitel 3.1 beschriebenen Ausgangsszenarien und ermöglicht es, unterschiedlich strukturierte Eingangsdaten der VNB in eine einheitliche Auswertungsstruktur zu überführen.

Die technische Umsetzung umfasst insbesondere die Bereinigung und Harmonisierung der Metadaten, die Abfrage bzw. Ergänzung von UID-Informationen, die Zuordnung der NOGA-Codes sowie die Aggregation auf Ebene der NOGA-Abteilungen. Damit konnten die einzelnen Verarbeitungsschritte systematisch dokumentiert und bei Bedarf für weitere Datensätze erneut angewendet werden. Die zugehörigen Software-Artefakte werden in Kapitel 5 beschrieben.

3.1.3 Datenschutz

Die Verarbeitung der im Rahmen des Projekts verwendeten Daten erfolgte unter Einhaltung der geltenden datenschutzrechtlichen Bestimmungen sowie auf Basis entsprechender Vereinbarungen mit den beteiligten VNB.

Im Fokus standen primär verbrauchsbezogene Daten mit Unternehmensbezug. Es wurden keine Personendaten von Privatpersonen verarbeitet.

Die Rollen und Verantwortlichkeiten waren klar geregelt. Die Datenverarbeitung erfolgte ausschliesslich im Rahmen der Vereinbarungen mit den Datenlieferanten. Eine Weitergabe der Daten war nur im Rahmen der vertraglichen Vereinbarungen zwischen VNB, BFE und dem beauftragten Dienstleister

sowie unter Einhaltung der gesetzlichen Vorgaben zulässig. Eine Weitergabe an Dritte war ausgeschlossen. Die Datenlieferanten hatten jederzeit die Möglichkeit, die Einhaltung der Vorgaben zu überprüfen.

Es wurden angemessene technische und organisatorische Massnahmen getroffen, um die Sicherheit der Daten zu gewährleisten. Das Sicherheitsniveau orientierte sich am Stand der Technik sowie an den potenziellen Risiken für betroffene Personen.

Der Grundsatz der Datenminimierung wurde berücksichtigt, indem nur jene Daten verarbeitet wurden, die für die Zielsetzung der Studie erforderlich waren. Nach Abschluss des Projekts wurden die Daten entsprechend den vertraglichen Vereinbarungen gelöscht oder zurückgegeben. Rohdaten wurden am Ende des Projekts vollständig gelöscht.

Für den Umgang mit allfälligen Datenschutzverletzungen waren entsprechende Melde- und Abstimmungsprozesse definiert.

3.2 Resultate

Abbildung 4 und Abbildung 4 zeigen exemplarisch sowohl die Verteilung der Messpunkte pro NOGA-Abschnitt als auch die relativen Stromverbrauchsanteile pro NOGA-Abschnitt für ausgewählte VNB.

Die Verteilung der Messpunkte auf die NOGA-Abteilungen liefert Hinweise darauf, wie stark einzelne Branchen in einem Netzgebiet vertreten sind und ob eine weitere Differenzierung sinnvoll erscheint. Dabei ist zu beachten, dass die Anzahl der Messpunkte keine Aussage über das Verbrauchsvolumen einer Branche zulässt. Als pragmatischer Grenzwert für eine weitergehende Unterteilung innerhalb der NOGA-Abteilung wurde ein Anteil von 25 % der Messpunkte innerhalb eines Wirtschaftszweigs definiert.

Ergänzend zur Betrachtung der Messpunktverteilung wurde der Stromverbrauch nach Branchen für die teilnehmenden VNB ausgewertet. Diese Perspektive erlaubt es, die relativen Beiträge der einzelnen Wirtschaftssektoren zum Stromverbrauch eines Netzgebiets zu analysieren. Die exemplarischen Auswertungen verdeutlichen, dass Branchen mit vielen Messpunkten nicht zwingend die höchsten Verbrauchsanteile aufweisen und umgekehrt.

Der Vergleich zwischen verschiedenen VNB verdeutlicht weiter, dass sich die Zusammensetzung der Wirtschaftszweige regional stark unterscheidet. Diese Unterschiede unterstreichen den Mehrwert einer branchenbezogenen Betrachtung pro Netzgebiet.

Ein Vergleich der Verbrauchsverteilungen über mehrere Jahre (2018–2023) pro VNB zeigt, dass die relativen Anteile der Branchen am Gesamtverbrauch eines VNB über die Zeit weitgehend stabil bleiben. Wesentliche strukturelle Veränderungen konnten in den betrachteten Daten nicht beobachtet werden. Gleichzeitig zeigt diese Auswertung, dass ein solcher Ansatz grundsätzlich geeignet ist, strukturelle Veränderungen im Stromverbrauch nach Wirtschaftszweigen über die Zeit sichtbar zu machen.

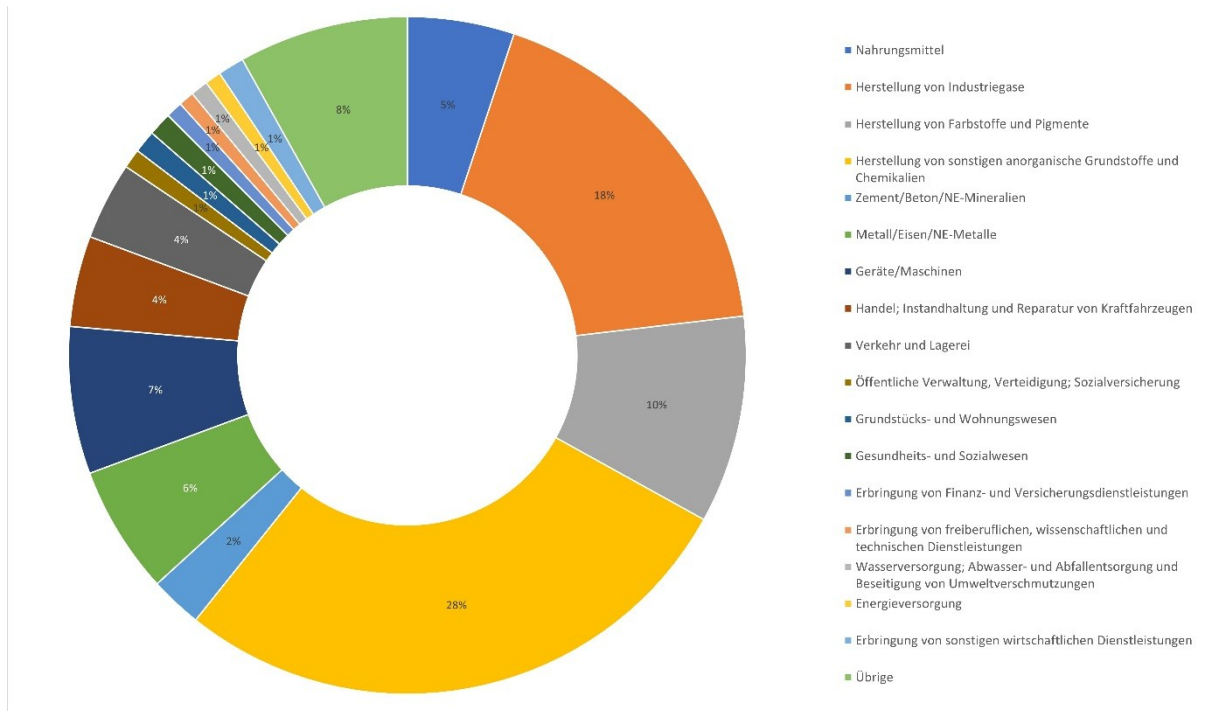


Abbildung 3: Beispielhafter Verbrauch aufgeteilt nach NOGA-Abschnitten

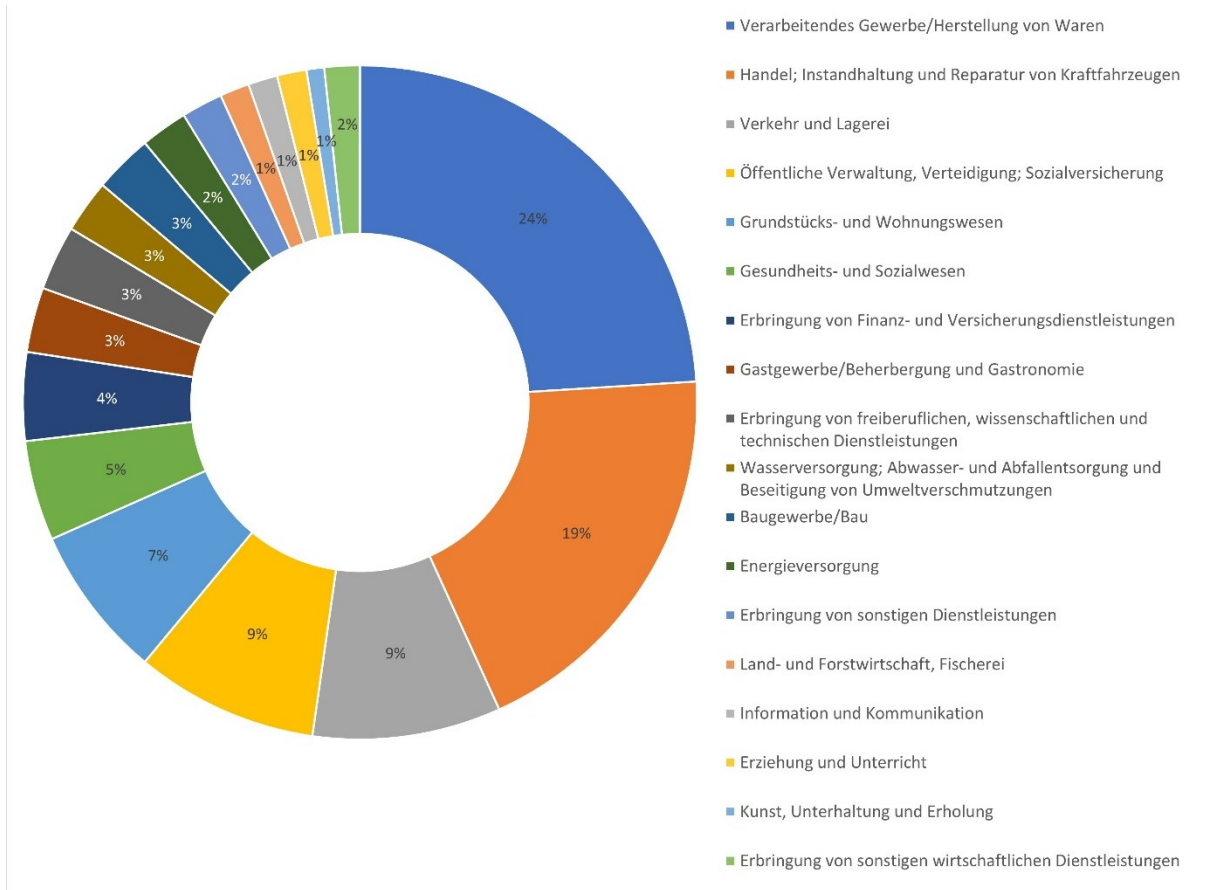


Abbildung 4: Exemplarische Verteilung der Messpunkte nach NOGA-Abschnitten

3.3 Zusammenfassung und Ausblick

Die Aufteilung des Stromverbrauchs eines VNB nach NOGA-Abteilungen und somit nach wirtschaftlichen Tätigkeiten ist ein geeignetes Instrument, um differenzierte regionalspezifische Analysen durchzuführen. Dabei sollte nicht nur der Stromverbrauch nach wirtschaftlicher Tätigkeit innerhalb eines Verteilnetzgebiets betrachtet werden, sondern auch die Verteilung der Messpunkte über die Wirtschaftszweige im jeweiligen Gebiet. Zudem kann die Analyse in ausserordentlichen Versorgungslagen dazu beitragen, Massnahmen zielgerichtet und differenziert auszugestalten.

Das Teilprojekt zeigt, dass auch bei einer zunächst wenig vielversprechenden Metadatengrundlage eine sinnvolle Unterteilung der Stromverbrauchsdaten nach NOGA-Abteilungen möglich ist. Die Ergebnisse liefern damit einen ersten belastbaren Überblick über branchenbezogene Verbrauchsstrukturen und bilden eine Grundlage für weiterführende Auswertungen.

Bei regelmässiger Anwendung könnte ein solcher Auswertungsansatz auch dazu beitragen, strukturelle wirtschaftliche Veränderungen in der Schweiz und in einzelnen Regionen sichtbar zu machen. Damit liessen sich Entwicklungen im Stromverbrauch nach Wirtschaftszweigen über die Zeit beobachten und mit politischen Instrumenten oder wirtschaftlichen Veränderungen in Beziehung

setzen. Dies könnte insbesondere die Analyse der Wirkung von energiepolitischen Massnahmen unterstützen.

Eine solche Auswertung könnte eine belastbare Datengrundlage schaffen, die für die Evaluation von Stromsparzielen sowie zur Überprüfung bereits umgesetzter Massnahmen genutzt werden kann. Zudem würde die Aufschlüsselung des Gesamtstromverbrauchs nach Wirtschaftssektoren in einem Netzgebiet wertvolle Informationen für VNB liefern. In Kombination mit der Information, wie die Messpunkte im Verteilnetzgebiet verteilt sind, ergäben sich zusätzliche Einblicke in die Struktur des Stromverbrauchs und in das Kundenportfolio. Diese Erkenntnisse könnten in strategische Planungen der VNB bzw. der EVU einfließen. Mit einem derartigen Wissen könnten EVU ihre Versorgung optimieren und besser auf ihre Kundenbasis zuschneiden. Um ein stabiles und für die gesamte Schweiz aussagekräftiges Bild des Stromverbrauches der wirtschaftlichen Sektoren zu ermöglichen, bräuchte es eine Harmonisierung der Metadaten der Messpunkte, einheitliche Datenstandards für Lastprofile sowie eine breitere, qualitativ hochwertige Datengrundlage.

Insgesamt zeigt das Teilprojekt, dass die Bereinigung der Messpunktmetadaten der VNB grundsätzlich möglich ist und die Verknüpfung von Lastprofilen mit NOGA-Codes und später mit NOGA-Abteilungen einen praktikablen und aussagekräftigen Ansatz für differenzierte Analysen auf regionaler Ebene darstellt. Damit dieser Ansatz grossflächig bzw. auf kantonaler und bundesweiter Ebene genutzt werden kann, sind jedoch standardisierte Datenstrukturen, eine breitere Datengrundlage und eine institutionell abgestützte Weiterentwicklung erforderlich. VNB sollten für die Bereinigung ihrer Messpunktmetadaten gezielt Massnahmen vorsehen. Einige ressourcenschonende Ansätze werden durch das in diesem Teilprojekt entwickelte Vorgehen bereitgestellt, müssen jedoch jeweils an die Verhältnisse des jeweiligen VNB angepasst werden.

Schliesslich kann die einheitliche Verwendung und Integration von NOGA-Codes in die Messpunktmetadaten als nationales Klassifikationssystem dienen und so wichtige weiterführende Auswertungen und Analysen unterstützen.

4 Teilprojekt 2: KI-gestützte Zuordnung von Verbrauchskurven zu NOGA-Abschnitten

Im zweiten Teilprojekt wurde untersucht, wie Lastprofilaten mithilfe von Machine Learning (ML) klassifiziert und spezifischen NOGA-Abschnitten zugeordnet werden können.

4.1 Zielsetzung und Datengrundlage

Im Gegensatz zum ersten Teilprojekt, bei dem Verbrauchsdaten über administrative Metadaten mit NOGA-Abteilungen verknüpft wurden, steht hier die direkte Klassifikation von Lastprofilen anhand ihrer Verbrauchsmuster im Vordergrund. Ziel war es, das Potenzial datengetriebener Ansätze für eine automatisierte Branchenzuordnung zu evaluieren und deren Grenzen aufzuzeigen.

Als Datengrundlage dienten die im Rahmen des ersten Teilprojekts aufbereiteten und mit NOGA-Codes angereicherten Verbrauchsdaten, die anschliessend auf Ebene der NOGA-Abteilungen aggregiert wurden. Die Lastgänge lagen in einer zeitlichen Auflösung von 15 Minuten und nach der Daten-

aufbereitung in einem einheitlichen Datenformat vor. Die Datenbasis umfasste Messpunkte aus unterschiedlichen NOGA-Abteilungen, wies jedoch eine stark ungleiche Verteilung der Daten zwischen den einzelnen Wirtschaftszweigen auf.

Eine zentrale Herausforderung ergab sich aus der hohen Anzahl möglicher NOGA-Codes und NOGA-Abteilungen sowie aus der begrenzten Datenbasis und der ungleichen Verteilung der Daten zwischen den Wirtschaftszweigen. Für die KI-gestützte Klassifikation wurden die im ersten Teilprojekt ermittelten NOGA-Codes und NOGA-Abteilungen deshalb den übergeordneten NOGA-Abschnitten zugeordnet. Eine direkte Klassifikation aller feingranularen NOGA-Codes bzw. NOGA-Abteilungen war mit der verfügbaren Datenbasis nicht möglich. Der Fokus wurde daher auf ausgewählte NOGA-Abschnitte mit ausreichender Datenbasis gelegt. Die übrigen NOGA-Abschnitte wurden in der Sammelkategorie „Andere“ zusammengefasst.

Vor der eigentlichen Modellierung wurden explorative Analysen durchgeführt, um Unterschiede und Gemeinsamkeiten der Lastprofile zwischen verschiedenen Wirtschaftszweigen zu untersuchen. Dabei zeigte sich, dass einige NOGA-Abschnitte charakteristische Verbrauchsmuster aufweisen, während sich andere nur schwer voneinander unterscheiden lassen. Diese Beobachtungen bildeten die Grundlage für die Auswahl der betrachteten Abschnitte sowie für die nachfolgenden methodischen Entscheidungen.

Die im zweiten Teilprojekt verwendeten Analyse- und Modellierungsansätze wurden in Python und Jupyter-Notebooks umgesetzt. Die zugehörigen Software-Artefakte werden in Kapitel 5 beschrieben.

4.2 Vorgehen und Resultate

In einem ersten Schritt wurde mittels Datenexploration analysiert, wie sich die Muster der Lastgänge je nach NOGA-Abschnitt unterscheiden. Dafür wurden die sieben NOGA-Abschnitte mit den meisten Messpunkten aus allen verfügbaren Abschnitten ausgewählt.

Anschliessend wurde mittels Feature Extraction und eines Sliding-Window-Ansatzes eine Reihe von Merkmalen definiert, die für die weitere Klassifizierung verwendet wurden:

- `time_series_minutely`: 15-Minuten Lastprofilaten in einem Fenster von -18h bis +18h um den zentralen Zeitstempel
- `time_series_daily`: 24-h-Mittelwert der Lastprofilaten in einem Fenster von -3.5d bis +3.5d um den zentralen Zeitstempel
- `additional`: tsfresh-Features⁶
- `weekday`: One-Hot-kodierter Wochentag des zentralen Zeitstempels
Bei der One-Hot-Kodierung werden kategoriale Daten in ein binäres Format umgewandelt.

Wie in Kapitel 4.1 dargestellt, war eine direkte Klassifikation aller feingranularen NOGA-Codes bzw. NOGA-Abteilungen mit der verfügbaren Datenbasis nicht zielführend. Für die weiteren Analysen wurde deshalb ein schrittweises Vorgehen gewählt. Zunächst wurden die sieben NOGA-Abschnitte mit

⁶ <https://tsfresh.readthedocs.io/en/latest/>

den meisten Messpunkten betrachtet. Anschliessend wurden reduzierte Kategoriensets mit drei bis vier NOGA-Abschnitten sowie eine ergänzende Sammelkategorie „Andere“ untersucht.

Die ersten Auswertungen machten deutlich, dass sich einige NOGA-Abschnitte deutlich besser klassifizieren lassen als andere. Während ein Teildatensatz mit nur drei Abschnitten hohe Genauigkeitswerte erreichte, sank die Genauigkeit deutlich, sobald alle sieben zunächst betrachteten Abschnitte miteinbezogen wurden. Die detaillierten Resultate zu diesen sieben Abschnitten sind im Appendix dargestellt.

Auf dieser Grundlage wurden verschiedene Modellierungsansätze unter realistischeren Bedingungen evaluiert. Ziel war es nicht mehr, alle NOGA-Abschnitte gleichzeitig zu klassifizieren, sondern zu prüfen, welche Abschnitte mit der vorhandenen Datenbasis zuverlässig unterscheidbar sind.

Eine Übersicht über die NOGA-Abschnitte, die mehr als zehn Messpunkte enthalten, ist in Tabelle 2 zu finden. Die verwendeten NOGA-Abschnitte sind blau eingefärbt.

Tabelle 2: NOGA-Abschnitte mit mehr als zehn Messpunkten

Bezeichnung	NOGA-Abschnitt	Anzahl Messpunkte
Gastgewerbe/ Beherbergung und Gastronomie	I	114
Handel; Instandhaltung und Reparatur von Kraftfahrzeugen	G	105
Verarbeitendes Gewerbe/ Herstellung von Waren	C	81
Öffentliche Verwaltung, Verteidigung, Sozialversicherung	O	78
Grundstücks- und Wohnungswesen	L	49
Verkehr und Lagerei	H	28
Energieversorgung	D	23
Gesundheits- und Sozialwesen	Q	23
Land- und Forstwirtschaft, Fischerei	A	21
Bergbau und Gewinnung von Steinen und Erden	B	12
Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen	M	12
Baugewerbe/Bau	F	11

4.2.1 Top 5 NOGA-Abschnitte

In einer ersten Analyse wurde untersucht, wie gut sich die fünf grössten NOGA-Abschnitten anhand eines einfachen Feature-Sets klassifizieren lassen.

In einer ersten Analyse wurden die fünf oben ersichtlichen NOGA-Abschnitte (blau eingefärbt) verwendet. Zur Generierung der Trainings- und Testsamples wurde ein Sliding-Window-Ansatz angewandt. Dabei wurde ein vierwöchiges Zeitfenster schrittweise über die Lastprofilaten eines Messpunkts verschoben. Jedes Zeitfenster bildete ein Sample und erhielt den NOGA-Abschnitt des zugehörigen Messpunkts als Label. Ziel war nicht, Veränderungen im Vierwochenrhythmus zu erkennen, sondern aus längeren Lastprofilen mehrere vergleichbare Trainingsbeispiele zu erzeugen, anhand derer das Modell typische Verbrauchsmuster pro NOGA-Abschnitt lernen konnte. Für die Klassifikation wurde folgendes Feature-Set verwendet:

- `additional`: tsfresh Features für eine 4-Wochen Zeitreihe.

Nach der Erzeugung der Samples und der Zuordnung zu den NOGA-Abschnitten standen 8'170 Samples zur Verfügung. Wie in Tabelle 3 ersichtlich, sind in Abschnitt L deutlich weniger Samples vorhanden als in den übrigen Abschnitten. Es ist zu erwarten, dass diese Ungleichverteilung einen direkten Einfluss auf die Klassifizierungsgenauigkeit (Accuracy) haben wird.

Tabelle 3: Anzahl Samples pro NOGA-Abschnitt

NOGA-Abschnitt	Anzahl Messpunkte pro Abschnitt	Anzahl Samples
I	114	2445
G	105	1823
C	81	1900
O	78	1325
L	49	677

Für die Klassifizierung wurde ein XGBoost-Modell⁷ verwendet. Zudem wurde untersucht, wie sich die Cross-Validated Accuracy auf (a) dem gesamten, unausgeglichenen Datensatz und (b) einem ausgeglichenen Datensatz verhält. Für den ausgeglichenen Datensatz diente die Kategorie mit der geringsten Anzahl an Samples als Referenz. Die übrigen Kategorien wurden entsprechend auf diese Anzahl heruntergesampelt. Die erreichte Accuracy ist in Tabelle 4 zusammengefasst. Zusätzlich wurde die Länge der für tsfresh-Features verwendeten Zeitreihen variiert (1, 3, 4 und 5 Wochen). Das 4-Wochen-Fenster lieferte die beste Performance. Grundsätzlich könnte eine längere Zeitreihe die Ergebnisse weiter verbessern. Hierfür wären jedoch mehr Daten erforderlich, um robuste und verlässliche Aussagen zu ermöglichen.

Tabelle 4: Accuracy mit Cross-Validation

Abschnitte	unausgeglichener Datensatz	ausgeglichener Datensatz
	Accuracy mit Cross-validation	Accuracy mit Cross-validation
I, G, C, O, L	70%	68.5%
I, G, C, O	77%	80%
I, G, C	85%	86%

⁷ <https://xgboost.readthedocs.io/en/stable/index.html>

4.2.2 Top 4 NOGA-Abschnitte und Sammelkategorie «Andere»

Darauf aufbauend wurde geprüft, ob sich die Klassifikation verbessern lässt, indem zusätzliche Features verwendet, ungleich verteilte Trainingsdaten ausgeglichen und die übrigen NOGA-Abschnitte in einer Sammelkategorie „Andere“ zusammengefasst werden.

In einem nächsten Schritt wurde die Anzahl der betrachteten NOGA-Abschnitte auf vier beschränkt. Die übrigen NOGA-Abschnitte wurden in der Sammelkategorie „Andere“ zusammengefasst. Für die weiteren Analysen wurden ausgeglichene Trainingsdatensätze verwendet, damit stark vertretene Abschnitte das Modell nicht dominieren. Dazu wurde die Anzahl Samples pro Abschnitt angepasst. Ergänzend wurde mit SMOTE⁸ ein Verfahren eingesetzt, das für schwächer vertretene Abschnitte künstliche Samples erzeugt.

Neben den bereits in Kapitel 4.2.1 verwendeten tsfresh-Feature `additional`, wurden zwei weitere Features integriert:

- `max_value`: Maximalwert der Zeitreihe sowie Features, die den Zeitstempel des Maximalwertes abbilden (Tag, Wochentag, Kalenderwoche, Monat in dem der Maximalwert auftritt)
- `min_value`: Minimalwert der Zeitreihe sowie Features, die den Zeitstempel des Minimalwertes abbilden (Tag, Wochentag, Kalenderwoche, Monat in dem der Minimalwert auftritt).

Untersucht wurde, wie sich die Cross-validated Accuracy verhält, wenn (a) nur das Feature `additional` genutzt wird, im Vergleich zu (b) der Kombination von `additional`, `max_value` und `min_value`. Die Resultate sind in Tabelle 5 zusammengefasst.

Tabelle 5: Vergleich der Cross-validated Accuracy für zwei Feature-Sets

Abschnitt	ausgeglichener Datensatz		SMOTE	
	<code>additional</code>	<code>alle</code>	<code>additional</code>	<code>alle</code>
I, G, C, O, Andere	61%	63%	65.5%	65.8%
I, G, C, Andere	70%	71%	74%	74.5%

In einem weiteren Schritt wurde eine Principal-Components-Analyse (PCA) durchgeführt. PCA ist eine Dimensions-Reduktions-Methode, welche die Varianz innerhalb eines Datensatzes auf ein neues Koordinatensystem abbildet. Dabei erklärt der erste Principal Component (PC) die grösste Varianz des gesamten Datensatzes. Die Feature-Dimension wurde so reduziert, dass insgesamt 95 % der

⁸ <https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/>

Varianz abgedeckt sind. PCA wurde sowohl separat für die Abschnitte I, G, C und O als auch gemeinsam für die vier Abschnitte und die Sammelkategorie „Andere“ angewendet.

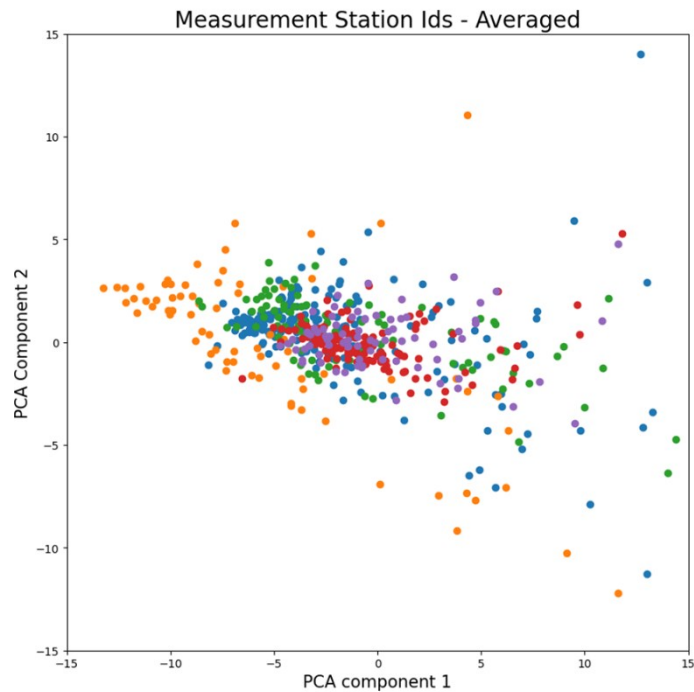


Abbildung 5: PCA-Visualisierung für die Abschnitte I, G, C, O und Sammelkategorie «Andere»

Abbildung 5 zeigt exemplarisch eine Darstellung mit Principal Component 1 auf der x-Achse und Principal Component 2 auf der y-Achse. Eine klare visuelle Trennung der Abschnitte ist nicht erkennbar. Dies unterstreicht die Schwierigkeit, eine entsprechende Klassifizierung ausschliesslich anhand des Lastprofils in Kombination mit den ausgewählten Features `additional`, `max_value` und `min_value` vorzunehmen. Basierend auf den Visualisierungen der einzelnen Abschnitte wurden in einem nächsten Schritt Ausreisser aus dem Trainingsset entfernt. Aus den Testdaten, welche für die Modellevaluation verwendet wurden, wurden die Ausreisser nicht entfernt. Die Anzahl der entfernten Ausreisser pro Abschnitt, ist in Tabelle 6 aufgelistet.

Tabelle 6: Anzahl entfernter Ausreisser pro NOGA-Abschnitt

NOGA-Abschnitt	Anzahl entfernte Ausreisser
I	3
G	5
C	4
O	3

Nach der Ausreisserbereinigung erfolgte eine erneute Auswertung mit einem XGBoost-Modell. Die Performanz wurde wiederum (a) nur mit `additional` und (b) mit allen drei Features (`additional`, `max_value` und `min_value`) ermittelt. Die Datensätze wurden hierbei mittels SMOTE ausbalanciert. Die Cross-Validated Accuracies sind in Tabelle 7 ersichtlich. Die Auswertung zeigt, dass die Ausreisserbereinigung keinen Effekt auf die Performanz der Modelle hat.

Tabelle 7: Vergleich der Cross-validated Accuracy für zwei Feature-Sets nach Ausreisserbereinigung

NOGA-Abschnitt	SMOTE	
	additional	alle
I, G, C, O, Andere	63.47%	63.98%
I, G, C, Andere	73.17%	74.25%

Die Aufteilung in vier NOGA-Abschnitte (I, G, C, O) sowie die Sammelkategorie «Andere» folgte der Idee eines mehrstufigen Vorgehens: Zunächst wurden die vier grössten NOGA-Abschnitte separat klassifiziert. Anschliessend wurde geprüft, ob sich die Sammelkategorie „Andere“, welche die übrigen 17 Abschnitte umfasst, weiter in sinnvolle Untergruppen aufteilen lässt. Ziel dieses rekursiven Vorgehens war es, die Sammelkategorie schrittweise zu verfeinern und dadurch weitere NOGA-Abschnitte separat klassifizieren zu können.

Zur explorativen Untersuchung der Sammelkategorie „Andere“ wurden zunächst verschiedene Vorverarbeitungsschritte durchgeführt. Die Features wurden standardisiert, damit sie vergleichbar skaliert vorlagen. Anschliessend wurde die Dimension der Daten reduziert, um mögliche Gruppenstrukturen besser sichtbar zu machen. Dafür wurden zwei Verfahren eingesetzt:

- PCA, wobei die Anzahl Features so reduziert wurde, dass 95 % der Varianz abgedeckt blieb

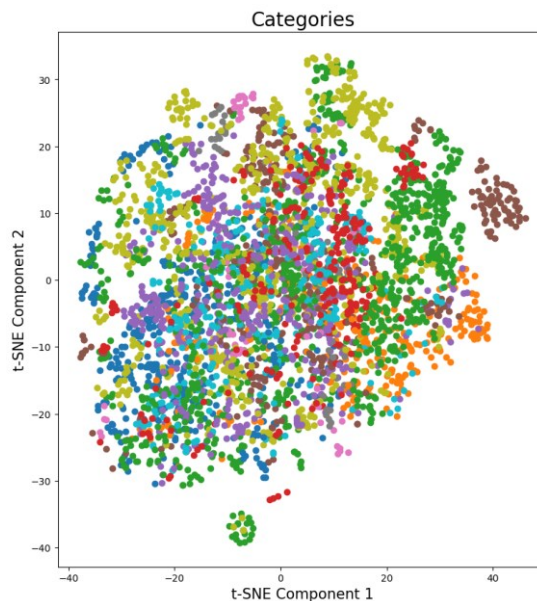
- t-SNE⁹, angewendet auf die drei Features `additional`, `max_value` und `min_value`

t-SNE ist im Gegensatz zu PCA ein nicht-lineares Verfahren zur Dimensionsreduktion und kann auch komplexere Strukturen in den Daten sichtbar machen.

Zur Bestimmung einer geeigneten Clusterzahl für die Kategorie «Andere» wurden die Metriken Sum of Squared Error (SSE) und der Silhouette-Koeffizient verwendet. Die entsprechenden Plots zu SSE und zum Silhouette-Koeffizienten sind im Appendix dargestellt.

Die Visualisierung in

Abbildung 6 zeigt Trennung innerhalb der „Andere“ erkennbar ist. Silhouette-Koeffizienten dass die gebildeten voneinander abgegrenzt weitere Unterteilung der „Andere“ war mit den den verwendeten möglich.



jedoch, dass keine klare Sammelkategorie Auch die niedrigen (< 0.5) deuten darauf hin, Cluster nicht ausreichend sind. Eine belastbare Sammelkategorie verfügbaren Daten und Features daher nicht

⁹ <https://medium.com/data-science/t-sne-clearly-explained-d84c537f53a>

Abbildung 6: t-SNE Visualisierung der Sammelkategorie "Andere"

4.2.3 Feature Engineering

Feature Engineering war ein weiterer Ansatz, der evaluiert wurde. Für jeden Messpunkt wurde aus den verfügbaren Rohdaten eine 4-Wochen-Zeitreihe mit einer Verschiebung von zwei Wochen extrahiert, um ausreichend Stichproben bei gleichzeitigem Erhalt des zeitlichen Kontexts zu generieren. Alle aus demselben Messpunkt abgeleiteten 4-Wochen-Zeitreihen tragen dabei denselben NOGA-Abschnitt als Label. Auf dieser Basis wurden Klassifikatoren trainiert und anschliessend zur Vorhersage des NOGA-Abschnitts auf die Samples des Testdatensatzes angewendet. Die endgültige Entscheidung auf Messpunktebene erfolgte mittels Mehrheitsvotum über alle zu einem Messpunkt gehörenden Samples.

Die zugrunde liegenden Zeitreihen lagen in einer Auflösung von 15 Minuten vor. Zusätzlich wurden aus den Rohdaten zwei weitere Zeitreihen durch Aggregation über 6 Stunden bzw. 24 Stunden erzeugt. Damit stehen pro Sample drei Zeitskalen (15 Minuten, 6 Stunden und 24 Stunden) zur Verfügung, um sowohl kurzfristige als auch tages- und mehrtägige Muster abzubilden.

Für das Feature Engineering wurden auf allen drei Zeitskalen konsistent Merkmale abgeleitet. Hierzu zählen der Zeitstempel der Extremwerte (Maximum und Minimum) einer Zeitreihe sowie kalendarische Merkmale wie Stunde, Tag, Monat, Wochentag, Kalenderwoche, ein Wochenendindikator und ein Indikator für Schweizer Feiertage. Zur Modellierung zyklischer Zusammenhänge wurden diese zeitlichen Merkmale zusätzlich zyklisch kodiert (z.B. über Sinus- und Kosinus-Transformationen). Ergänzend wurden deskriptive Statistiken berechnet, darunter Mittelwert, Median, Standardabweichung, Kurtosis und Schiefe. Die genannten Merkmale wurden für alle drei Zeitskalen erzeugt. Für die Zeitreihe mit 15-Minuten-Auflösung wurden nach Min-Max-Normalisierung zusätzlich Fourier-Koeffizienten bestimmt, um periodische Komponenten explizit zu erfassen. Darüber hinaus wurden für den zentralen Zeitstempel eines Samples weitere Merkmale gebildet, mit Ausnahme der zyklischen Kodierungen. Der Trainingsdatensatz wurde mit SMOTE vergrössert. Dabei wurden zusätzliche synthetische Zeitreihen-Samples durch Interpolation generiert.

Die Cross-validated Accuracy ist in Tabelle 8 zusammengefasst. Zusätzlich wurde die Wichtigkeit der einzelnen Features analysiert. Die verwendeten Features sind in der entsprechenden Reihenfolge mit dem wichtigsten Feature zuoberst in Tabelle 22 im Appendix aufgelistet.

Tabelle 8: Cross-validated Accuracy für Feature-Engineering-Anwendung

	Features	Cross-validation Accuracy	
		Top-4 und «Andere»	Top-3 und «Andere»
3-Wochen Zeitreihe	Center-timestamp, Min, Max, Stats	56.76%	69.22%
	Center-timestamp, Min, Max, Stats, FFT	62.97%	72.52%
4-Wochen Zeitreihe	Center-timestamp, Min, Max, Stats	59.32%	69.46%
	Center-timestamp, Min, Max, Stats, FFT	63.75%	72.35%
5-Wochen Zeitreihe	Center-timestamp, Min, Max, Stats	56.09%	68.25%
	Center-timestamp, Min, Max, Stats, FFT	61.89%	72.13%

4.2.4 LSTM-Autoencoder

Zusätzlich wurde geprüft, ob mithilfe von LSTM-Autoencodern gelernte latente Merkmale einen Mehrwert für die Klassifikation bieten. Dazu wurden Autoencoder trainiert, welche aus den Zeitreihen komprimierte Repräsentationen lernen. Anschliessend wurde untersucht, ob diese latenten Merkmale die Zuordnung zu NOGA-Abschnitten verbessern.

Die Rohdaten lagen mit einer zeitlichen Auflösung von 15 Minuten vor. Zusätzlich wurden drei weitere Zeitskalen durch Aggregation über 1 Stunde, 6 Stunden und 24 Stunden erzeugt. Für jeden Messpunkt wurden 4-Wochen-Zeitreihen mit einer zeitlichen Verschiebung von 2 Wochen extrahiert. Alle Samples eines Messpunktes tragen denselben NOGA-Abschnitt als Label. Als Eingangsmerkmale dienten normierte Verbrauchszeitreihen sowie zyklische Zeitfeatures (Sinus/Cosinus) für Stunde, Tag, Wochentag, Woche und Monat.

Es wurde ein sequenzieller LSTM-Autoencoder eingesetzt. Der Encoder besteht aus zwei LSTM-Schichten mit `HIDDEN_DIM = 16` und `ENCODING_DIM = 10`, wobei die zweite Schicht die komprimierte, 10-dimensionale Repräsentation liefert. Der Decoder spiegelt diese Struktur über einen Repeat-Vector und zwei LSTM-Schichten wider. Die Rekonstruktion erfolgt über eine zeitverteilte Dense-Schicht. Das Modell umfasst rund 5'500 trainierbare Parameter. Grössere Modelle wurden aufgrund des erhöhten Risikos für Overfitting bewusst vermieden.

Für das Training standen 7'379 Samples zur Verfügung, zur Validierung 1'845 und der Testdatensatz beinhaltet 2'239 Samples.

Nach dem Training wurden die im Encoding gelernten latenten Merkmale extrahiert und als Eingaben für Klassifikationsmodelle genutzt. Um die Unausgeglichenheit der NOGA-Abschnitte zu adressieren, wurde das Trainingsset der Klassifikatoren mittels SMOTE (Synthetic Minority Over-sampling Technique) vergrössert. Dabei wurden synthetische Stichproben der unterrepräsentierten NOGA-Abschnitte erzeugt. Die Vorhersagen erfolgen für jede Zeitreihe einzeln. Die endgültige Entscheidung auf Messpunktebene wird mittels Mehrheitsvotum über alle Zeitreihen eines Messpunktes getroffen. Dieses Vorgehen erlaubt es, die Aussagekraft der Autoencoder-Features hinsichtlich der Klassifikationsaufgabe systematisch zu evaluieren.

Die Resultate sind in Tabelle 9 zusammengefasst.

Tabelle 9: Accuracy für die Anwendung der LSTM-Autoencoder

	Features	Accuracy	
		Top-4 und «Andere»	Top-3 und «Andere»
Zeitreihe Stündliche	Center-timestamp, Min, Max, Stats	58.97%	67.52%
	Center-timestamp, Min, Max, Stats, latente Features	57.26%	65.81%
	Latente Features	51.28%	61.54%
Zeitreihe 6-Stunden	Center-timestamp, Min, Max, Stats	58.97%	67.52%
	Center-timestamp, Min, Max, Stats, latente Features	58.12%	63.25%
	Latente Features	59.82%	70.08%
Zeitreihe Täglich	Center-timestamp, Min, Max, Stats	58.97%	67.52%
	Center-timestamp, Min, Max, Stats, latente Features	59.83%	66.67%
	Latente Features	47.86%	63.25%

4.2.5 Alternative BFE-NOGA-Kategorien

Abschliessend wurde geprüft, wie sich das Modell verhält, wenn anstelle der zuvor verwendeten NOGA-Abschnitte die 19 Kategorien aus einem bestehenden BFE-Bericht verwendet werden.

Der BFE-Bericht «Energieverbrauch in der Schweiz 2011 – 2020 nach NOGA Stufe 2»¹⁰ gruppiert die NOGA-Codes in 19 verschiedene Kategorien. Es wurde analysiert, wie sich die Modellgüte des XGBoost-Modells verhält, wenn diese 19 Kategorien als Zielklassen verwendet werden. Das Vorgehen entsprach grundsätzlich dem in Kapitel 4.2.1.

In Tabelle 10 sind jene BFE-NOGA-Kategorien aufgeführt, die Daten von mehr als zehn Messpunkten enthalten.

Tabelle 10: Anzahl Messpunkte pro BFE-NOGA-Kategorie

BFE-NOGA-Kategorien	Anzahl Messpunkte in der Kategorie
19	129
14	114
13	106
16	78
1	25
18	24
6	19
9	17
2	13
11	12
12	11

Der Fokus liegt im Folgenden auf den ersten sechs Kategorien, da diese jeweils mehr als 20 Messpunkte umfassten.

Die Samples wurden mittels Sliding-Window-Ansatz generiert. Ausgehend von einem zentralen Zeitstempel wurde ein vierwöchiges Zeitfenster betrachtet. Darauf basierend wurde folgendes Feature verwendet:

¹⁰ <https://www.bfe.admin.ch/bfe/de/home/news-und-medien/publikationen.ex-turl.html/aHR0cHM6Ly9wdWJkYi5iZmUuYWWRtaW4uY2qvZGUvcHVib-GljYX/Rpb24vZG93bmxvYWQvMTlyODk=.html>

- `additional: tsfresh-Features` für eine 4-Wochen Zeitreihe.

Nach der Erzeugung der Samples, der Zuordnung zu den BFE-NOGA-Kategorien und dem Ausschluss unvollständiger Zeitfenster standen insgesamt 8'995 Samples zur Verfügung. In Tabelle 11 ist ersichtlich, dass die Kategorien 1 und 18 im Vergleich zu den anderen vier Kategorien deutlich weniger Samples enthalten. Es war daher zu erwarten, dass die Klassifikationsgenauigkeit (Accuracy) für diese beiden Kategorien niedriger ausfällt als bei den übrigen Kategorien.

Tabelle 11: Anzahl Samples pro BFE-NOGA-Kategorie

BFE-NOGA-Kategorie	Anzahl Messpunkte pro Kategorie	Anzahl Samples
19	129	2252
14	114	2445
13	106	1847
16	78	1325
1	25	585
18	24	541

Für die Klassifizierung wurde ein XGBoost-Modell verwendet. Zusätzlich wurde evaluiert, wie sich die Cross-Validated Accuracy unterscheidet, wenn (a) ein Datensatz mit allen Samples verwendet wird und (b) ein ausgeglichener Datensatz. Für den ausgeglichenen Datensatz wurden Kategorien mit vielen Samples auf die Anzahl Samples der kleinsten verwendeten Kategorie reduziert. Die erreichte Accuracy ist in Tabelle 12 aufgelistet.

Tabelle 12: Accuracy mit Cross-Validation für zwei verschiedene Datensätze

BFE-NOGA-Kategorien	unausgeglichener Datensatz	ausgeglichener Datensatz
	Accuracy mit Cross-validation	Accuracy mit Cross-validation
19, 14, 13	80%	80%
19, 14, 13, 16	71%	68.5%
19, 14, 13, 16, 1, 18	65.5%	57.5%

4.3 Diskussion und Ausblick

Aus den oben dargestellten Ergebnissen ergibt sich, dass eine Klassifizierung der Lastgänge auf Ebene aller NOGA-Codes bzw. NOGA-Abteilungen mit den in der Studie verfügbaren Daten nicht realistisch ist. Eine Reduktion auf ausgewählte NOGA-Abschnitte mit den meisten Samples, ergänzt durch die Sammelkategorie «Andere» ermöglicht hingegen eine Klassifikation mit einer kreuzvalidierten Genauigkeit (cross-validated Accuracy) von etwa 65–75%. Dabei ist zu

berücksichtigen, dass es sich um reale Daten handelt. Viele Datensätze weisen Lücken, fehlerhafte Inhalte oder sehr ungewöhnliche Muster auf.

Wie bereits in der Datenexploration und in einer ersten Klassifikation ersichtlich, liefert die Einordnung der drei Gruppen «Herstellung von Metallerzeugnissen» (NOGA-Code: 25), «Detailhandel» (NOGA-Code 47) und «Gastronomie» (NOGA-Code 56) mit zusätzlichen Features eine Accuracy von über 95%. Dabei handelt es sich jedoch um ein stark reduziertes Kategorienset auf Ebene einzelner NOGA-Abteilungen. Gleichzeitig ist die Zahl der Messpunkte pro Kategorie sehr klein. Entsprechend stellt sich die Frage, ob eine Klassifikation der drei bis fünf grössten Abschnitte bzw. ausgewählter Abteilungen bereits einen ausreichenden Mehrwert bietet. Die vorliegenden Ergebnisse sprechen dafür, dass eine fokussierte Klassifikation auf ein begrenztes Kategorienset funktionieren kann.

Eine der grössten Herausforderungen waren die begrenzten Datenmengen und die daraus resultierende ungleiche Repräsentation der einzelnen NOGA-Abschnitte. Dieses Problem ist nur schwer zu beheben, da die Branchen in der Schweiz unterschiedlich stark vertreten sind. So sind beispielsweise Messpunkte aus der Forstwirtschaft deutlich seltener vertreten als Messpunkte aus dem Handel oder dem Gastgewerbe. Zudem sind die NOGA-Abschnitte unterschiedlich zusammengesetzt. Manche Abschnitte umfassen mehrere NOGA-Abteilungen (z.B. Baugewerbe umfasst die Abteilungen 41-43), während andere nur aus einer einzigen Abteilung bestehen (z.B. Erziehung und Unterricht mit der Abteilung 85). Dadurch entsteht eine gewisse Unausgewogenheit. Würden die Abschnitte wieder auf einzelne NOGA-Codes oder NOGA-Abteilungen reduziert, ergäbe sich erneut die Herausforderung eine deutlich grössere Anzahl an Klassen zu berücksichtigen. Die Unterschiede zwischen den Klassen würden folglich ebenfalls zunehmen, da es wenige Klassen mit vielen Messpunkten gäbe, aber viele Klassen mit nur wenigen oder gar keinen Messpunkten.

Die Ergebnisse zeigen damit klar, dass die Leistungsfähigkeit der Modelle weniger von der Wahl des Algorithmus als von Umfang, Qualität und Ausgewogenheit der verfügbaren Daten abhängt. Es ist notwendig, mehr und vor allem repräsentativere sowie ausgewogenere Daten zu sammeln. Mit einem umfangreicheren Datensatz könnten grössere Machine Learning Modelle angewendet werden, die vermutlich bessere Ergebnisse liefern. Dennoch bleibt das Problem der unausgeglichenen Daten bestehen, da die Anzahl Messpunkte je NOGA-Abschnitt stark variiert. Dies hängt auch mit der realen Wirtschaftsstruktur zusammen. Messpunkte in Bereichen wie Fischerei oder Forstwirtschaft kommen deutlich seltener vor als Messpunkte im Detailhandel oder im Gastgewerbe.

Insgesamt deutet das Teilprojekt darauf hin, dass KI-gestützte Ansätze für klar abgegrenzte und ausreichend vertretene NOGA-Abschnitte oder ausgewählte NOGA-Abteilungen einen Mehrwert bieten können. Für eine breit abgestützte und robuste Klassifikation über viele wirtschaftliche Tätigkeiten und NOGA-Ebenen hinweg sind jedoch deutlich umfangreichere und standardisiertere Datengrundlagen erforderlich.

5 Software-Artefakte und Repositories

Im Rahmen beider Teilprojekte wurden die entwickelten Methoden und Workflows technisch umgesetzt. Dazu entstanden zwei Code-Repositories, die vollständig im Auftrag des BFE entwickelt wurden und sich im BFE-GitHub-Account¹¹ befinden. Sie enthalten die Implementierung der

¹¹ <https://github.com/SFOE/NOGA-Klassifizierung>

Datenaufbereitung, UID-Ermittlung und NOGA-Zuordnung sowie die Werkzeuge für die Machine-Learning-Analysen. Damit sind die wesentlichen Schritte der Studie technisch dokumentiert, reproduzierbar und für zukünftige Weiterentwicklungen nutzbar. Die Repositories stellen einen integralen Bestandteil der Projektergebnisse dar und bilden eine Grundlage für mögliche Folgeprojekte sowie für die Weiterentwicklung datenbasierter Analyse- und Transparenzgrundlagen im Strombereich.

6 Zusammenfassung und Ausblick

Die vorliegende Analyse bietet einen soliden Ausgangspunkt für eine erste differenzierte Betrachtung des Stromverbrauchs in ausgewählten Gebieten der Schweiz. Um belastbare und repräsentative Aussagen zur Verteilung und zum Stromverbrauch einzelner Branchen in den Regionen bzw. pro VNB machen zu können, werden jedoch zusätzliche Daten weiterer VNB benötigt. Daten aus zusätzlichen Regionen tragen dazu bei, regionsspezifische Effekte zu glätten und die Repräsentativität zu erhöhen.

Die Ergebnisse des ersten Teilprojektes der Studie können als Basis für weitergehende Analysen dienen und bilden eine gute Grundlage für die schweizweite Verbesserung der Transparenz im Stromverbrauch durch regelmässige Auswertungen. Das erste Teilprojekt bestätigt, dass eine Verknüpfung von Verbrauchsdaten und NOGA-Codes grundsätzlich möglich und nutzstiftend ist. Es ermöglicht Einblicke in die Verteilung des Stromverbrauchs nach Branchen sowie in die Struktur der Messpunkte innerhalb der Schweiz. Eine solche Transparenz kann nicht nur im Kontext von Massnahmen in ausserordentlichen Lagen einen Mehrwert bieten, sondern auch bei der Bewertung der Wirkung von politischen Instrumenten auf nationaler oder kantonaler Ebene, beispielsweise Energiesparmassnahmen oder Aufrufen zum Energiesparen. Auch für VNB sind Einblicke aufschlussreich z.B. hinsichtlich der Bewirtschaftung ihres Kundenportfolios. Zudem schafft eine solche Transparenz die Grundlage für weiterführende Auswertungen, etwa zum Verbrauch nach Branchen und Jahreszeiten oder zur Entwicklung über mehrere Jahre oder dem Einfluss des Zubaus von Photovoltaik. Darüber hinaus ermöglicht die Verknüpfung von Messpunkt mit NOGA-Code die vergleichsweise einfache Generierung gelabelter (annotierter) Daten, die für weitere Machine Learning Anwendungen, insbesondere für überwachte (supervised) Verfahren, essenziell sind.

Im zweiten Teilprojekt wurde deutlich, dass eine direkte Klassifizierung von Lastgängen auf Ebene feingranularer NOGA-Codes oder zahlreicher NOGA-Abteilungen sehr ambitioniert und herausfordernd ist. Dabei ist zu unterscheiden, welches Ziel verfolgt wird und welche Anwendungsfälle bereits einen Nutzen mit sich bringen. So zeigen die Ergebnisse, dass eine Klassifizierung in stark reduzierten Kategorien bzw. ausgewählten NOGA-Abschnitten vielversprechende Resultate liefern kann. Dies kann insbesondere für weitere Arbeiten der VNB zur Bereinigung ihrer Messpunktmetadaten im Hinblick auf die nationale Datenplattform im Stromsektor relevant sein. Eine robuste Einordnung in alle 99 NOGA-Abteilungen oder noch feingranularere NOGA-Codes setzt jedoch deutlich mehr und qualitativ bessere Daten voraus, die zusätzlich eine ausgewogenere und repräsentativere Verteilung der Wirtschaftszweige abbilden müssten.

Weitere methodische Ansätze, die künftig geprüft werden könnten, umfassen die Einbindung latenter Merkmale aus Autoencodern oder handgefertigter Merkmale. Auch musterbasierte Ansätze, die spezifische Branchenmuster wie etwa das Verhältnis von Grundlast zu Verbrauchsspitzen erfassen, stellen eine mögliche Erweiterung dar. Ebenso könnte die Einbindung branchenspezifischer Merkmale, etwa Feiertage oder das Nacht- und Wochenendverhalten, die Klassifikation weiter verbessern. Im zweiten Teilprojekt hat sich deutlich gezeigt, dass für die Untersuchung dieser weiteren Ansätze umfangreichere und qualitativ hochwertigere Datengrundlagen erforderlich sind.

Insgesamt zeigt sich, dass die vorhandenen und fortlaufend generierten Daten erhebliches Potenzial aufweisen. Bei angemessener Aufbereitung und Nutzung erscheint eine schweizweite standardisierte Zuordnung von Messpunkten zu NOGA-Codes bzw. NOGA-Abteilungen perspektivisch realistisch. Voraussetzungen sind jedoch eine verbesserte Datenqualität, möglichst einheitliche Datenformate und ein einheitliches Vorgehen, das auf der einen Seite Datenschutz gewährleistet aber auch aussagekräftige Analysen ermöglicht. Aktuell sind insbesondere Datenmenge, Datenqualität und Repräsentativität die zentralen Faktoren, die noch nicht erfüllt sind, um in einem nächsten Schritt eine belastbare Klassifizierung zu realisieren.

Grundsätzlich lässt sich festhalten, dass ein klarer Bedarf an einem grösseren, ausgewogeneren und qualitativ hochwertigen Datensatz besteht, sowohl für weiterführende Machine Learning Anwendungen als auch für die wesentliche Verbesserung der Transparenz im Schweizer Stromverbrauch.

7 Appendix

NOGA-Abteilung aus Lastprofilen

Für die Datenexploration und eine erste Klassifizierung wurden ausgewählte NOGA-Abteilungen aus einem Set von Messpunkten von vier VNB verwendet. Die Anzahl Messpunkte pro NOGA-Abteilung unterscheidet sich daher von den übrigen Auswertungen in Teilprojekt zwei.

Die verwendeten NOGA-Abteilungen und die Anzahl Messpunkte pro Abteilung sind in Tabelle 13 dargestellt.

Tabelle 13: NOGA-Abteilung mit Anzahl Messpunkten pro Kategorie

NOGA-Abteilung	Bezeichnung	Anzahl Messpunkte pro Abteilung
47	Detailhandel	83
84	Öffentliche Verwaltung	78
56	Gastronomie	61
55	Beherbergung	53
68	Grundstücks- und Wohnungswesen	49
25	Herstellung von Metallerzeugnissen	26
35	Energieversorgung	23

Klassifizierung aller sieben NOGA-Abteilungen

Für die Klassifizierung von allen sieben Kategorien wurden zwei verschiedenen Klassifikatoren verwendet. Einmal ein 7-class Random Forest Classifier und einmal ein 7-class MLP Classifier. Die Resultate sind in den folgenden Tabellen zusammengefasst. Eine Klassifizierung der sieben Kategorien nach Zufallsklassifikation ergibt eine Accuracy von 0.143.

Tabella 14: 7-class RandomForestClassifier ohne additional Features

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	0.75	0.86
35	5	0.50	0.20	0.29
47	17	0.85	0.65	0.73
55	11	0.56	0.45	0.50
56	13	0.81	1.00	0.90
68	10	0.00	0.00	0.00
84	16	0.48	0.81	0.60
Accuracy	76			0.61
Macro avg	76	0.60	0.55	0.55
Weighted avg	76	0.60	0.61	0.58

Tabella 15: 7-class RandomForestClassifier mit additional Features

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	0.75	0.86
35	5	0.50	0.20	0.29
47	17	0.79	0.65	0.71
55	11	0.36	0.36	0.36
56	13	0.79	0.85	0.81
68	10	0.00	0.00	0.00
84	16	0.50	0.81	0.62
Accuracy	76			0.57
Macro avg	76	0.56	0.52	0.52
Weighted avg	76	0.55	0.57	0.54

Tabelle 16: 7-class RandomForestClassifier mit additional Features und Wochentagen

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	0.75	0.86
35	5	0.50	0.20	0.29
47	17	0.85	0.65	0.73
55	11	0.56	0.45	0.50
56	13	0.81	1.00	0.90
68	10	0.00	0.00	0.00
84	16	0.48	0.81	0.60
Accuracy	76			0.61
Macro avg	76	0.60	0.55	0.55
Weighted avg	76	0.60	0.61	0.58

Tabelle 17: 7-class MLPClassifier mit additional Features

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	0.75	0.86
35	5	0.33	0.20	0.25
47	17	0.93	0.82	0.88
55	11	0.24	0.45	0.31
56	13	0.67	0.77	0.71
68	10	0.38	0.30	0.33
84	16	0.50	0.31	0.38
Accuracy	76			0.54
Macro avg	76	0.51	0.45	0.47
Weighted avg	76	0.59	0.54	0.55

Klassifizierung von drei Abteilungen (25, 47, 56):

Da sich zeigte, dass sich die drei NOGA-Abteilungen 25, 47 und 56 vergleichsweise gut klassifizieren lassen, wurde die Klassifizierung mit diesen drei Abteilungen erneut durchgeführt. Die Resultate sind in den folgenden Tabellen zusammengefasst. Eine Zufallsklassifikation der drei Abteilungen ergibt eine Accuracy von 0.33.

Tabelle 18: 3-class RandomForestClassifier ohne additional Features

Klasse	Support	Precision	Recall	F1-Score
25	4	0.75	0.75	0.75
47	17	0.88	0.82	0.85
56	13	0.86	0.92	0.89
Accuracy	34			0.85
Macro avg	34	0.83	0.83	0.83
Weighted avg	34	0.85	0.85	0.85

Tabelle 19: 3-class RandomForestClassifier mit additional Features

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	0.75	0.86
47	17	0.94	1.00	0.97
56	13	1.00	1.00	1.00
Accuracy	34			0.97
Macro avg	34	0.98	0.92	0.94
Weighted avg	34	0.97	0.97	0.97

Tabelle 20: 3-class RandomForestClassifier mit additional Features und Wochentagen

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	1.00	1.00
47	17	1.00	0.94	0.97
56	13	0.93	1.00	0.96
Accuracy	34			0.97
Macro avg	34	0.98	0.98	0.98
Weighted avg	34	0.97	0.97	0.97

Tabelle 21: 3-class MLPClassifier mit additional Features

Klasse	Support	Precision	Recall	F1-Score
25	4	1.00	1.00	1.00
47	17	1.00	0.94	0.97
56	13	0.93	1.00	0.96
Accuracy	34			0.97
Macro avg	34	0.98	0.98	0.98
Weighted avg	34	0.97	0.97	0.97

Top 4 NOGA-Abschnitte und Sammelkategorie «Andere»

Visualisierung von SSE und Silhouette-Koeffizient für PCA und t-SNE in Abhängigkeit der Anzahl Cluster. Diese Analyse dient dazu, eine geeignete Clusterzahl für die weitere Untersuchung der Sammelkategorie «Andere» zu evaluieren.

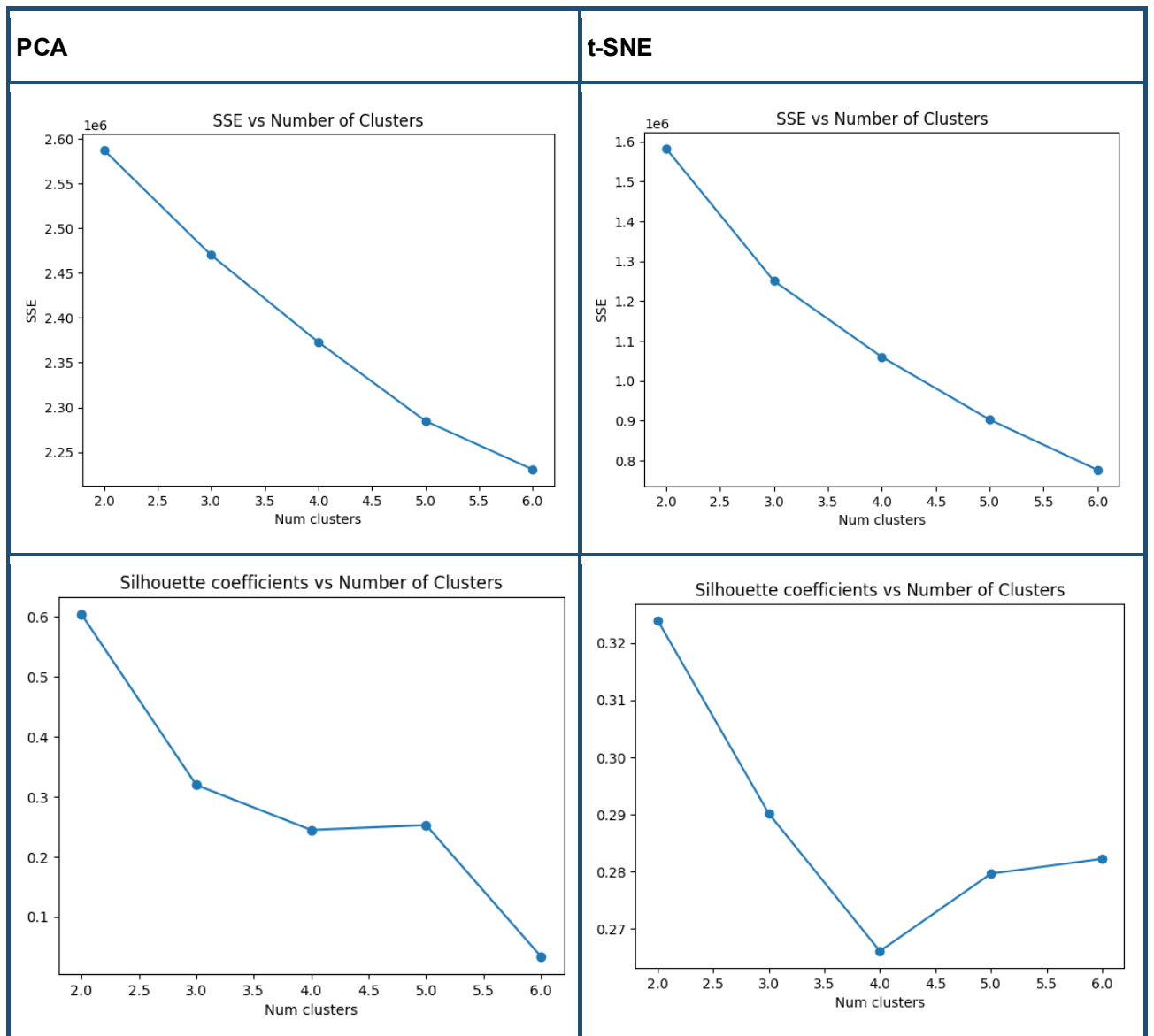


Abbildung 7: SSE und Silhouette-Koeffizient in Abhängigkeit der Anzahl Cluster für die Sammelkategorie "Andere"

Feature Engineering

Tabelle 22 zeigt die verwendeten Features, absteigend nach ihrer Wichtigkeit geordnet.

Tabelle 22: Verwendete Features für die Feature Engineering Anwendung nach Wichtigkeit geordnet

Center-timestamp, Min, Max, Stats	Center-timestamp, Min, Max, Stats, FFT
min_daily_aggregated_day_is_weekend	max_daily_aggregated_day_is_weekend
center_timestamp_day_is_weekend	fft_coeff_img_32
max_6_hours_aggregated_hours_sine	fft_coeff_img_12
max_daily_aggregated_day_is_weekend	min_raw_value_day_of_week_cosine
max_daily_aggregated_day_is_weekend	fft_coeff_real_84
min_daily_aggregated_day_of_week_cosine	min_daily_aggregated_day_is_weekend
min_daily_aggregated_day_of_week_sine	min_daily_aggregated_day_of_week_cosine
center_timestamp_hours	max_daily_aggregated_day_of_week_cosine
min_raw_value_day_of_week_cosine	fft_coeff_real_4
max_6_hours_aggregated_day_is_weekend	max_daily_aggregated_month_sine
max_6_hours_aggregated_day_of_week_sine	min_6_hours_aggregated_day_of_week_cosine
stats_6_hours_aggregated_std_dev	fft_coeff_real_28
stats_6_hours_aggregated_kurtosis	fft_coeff_img_4
max_daily_aggregated_day_of_week_cosine	min_daily_aggregated_day_of_week_sine
max_6_hours_aggregated_hours_cosine	fft_coeff_real_36
min_raw_value	stats_6_hours_aggregated_skewness
min_6_hours_aggregated_hours	center_timestamp_hours
min_6_hours_aggregated_hours_cosine	max_6_hours_aggregated_hours
max_raw_value_hours	stats_raw_mean
stats_6_hours_aggregated_skewness	max_6_hours_aggregated_hours_cosine
max_daily_aggregated	