



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Umwelt, Verkehr, Energie und
Kommunikation UVEK

Bundesamt für Energie BFE
Sektion Energieforschung

Schlussbericht vom 02. Juni 2016

Smart-Meter-Datenanalyse für au- tomatisierte Energieberatungen („Smart Grid Data Analytics“)



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Otto-Friedrich-Universität Bamberg



Datum: 02.06.2016

Ort: Bamberg

Auftraggeberin:

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

Kofinanzierung:

Arbon Energie AG, Salwiesenstrasse 1, CH-9320 Arbon

Auftragnehmer/in:

ETH Zürich
D-MTEC
Lehrstuhl für Informationsmanagement Weinbergstrasse 56/58
CH-8092 Zürich
www.im.ethz.ch

Autoren:

Mariya Sodenkamp, Otto-Friedrich-Universität Bamberg, Lehrstuhl für Energieeffiziente Systeme,
mariya.sodenkamp@uni-bamberg.de

Konstantin Hopf, Otto-Friedrich-Universität Bamberg, Lehrstuhl für Energieeffiziente Systeme,
konstantin.hopf@uni-bamberg.de

Ilya Kozlovskiy, Otto-Friedrich-Universität Bamberg, Lehrstuhl für Energieeffiziente Systeme,
ilya.kozlovskiy@uni-bamberg.de

Thorsten Staake, ETH Zürich, D-MTEC, Lehrstuhl für Informationsmanagement, tstaake@ethz.ch

Mit Beiträgen von:

Christian Beckel, ETH Zürich, Distributed Systems Group, beckel@inf.ethz.ch

BFE-Bereichsleitung: Dr. Michael Moser, michael.moser@bfe.admin.ch

BFE-Programmleitung: Roland Brüniger, roland.brueeniger@r-brueniger-ag.ch

BFE-Vertragsnummern: SI/501053-01, SI/501202-01

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

Bundesamt für Energie BFE

Mühlestrasse 4, CH-3063 Ittigen; Postadresse: CH-3003 Bern
Tel. +41 58 462 56 11 · Fax +41 58 463 25 00 · contact@bfe.admin.ch · www.bfe.admin.ch



Zusammenfassung

Kommunikationsfähige Stromzähler ermöglichen die Erfassung individueller Lastprofile mit hoher zeitlicher Auflösung (typischerweise in 15-Minuten-Intervallen). Projektgegenstand ist die Weiterentwicklung von Methoden des maschinellen Lernens, um aus Lastprofilen und zusätzlichen verbrauchs-relevanten Informationen (Wetter, soziodemographische Daten, Adressinformationen, usw.) automatisiert Merkmale von Haushalten abzuleiten, welche für eine individuelle und spezifische Energieberatung von Nutzen sind.

Mit den im Rahmen des Projektes entwickelten Smart-Meter-Klassifikations-Verfahren konnten 38 Eigenschaften privater Haushalte mit zum Teil hoher Sicherheit (über 70%) aus Lastprofilen und zusätzlichen frei verfügbaren Daten unter Einhaltung von Datenschutzbestimmungen vorhergesagt werden. Neben Umständen der Lebenssituation (z.B. Familien, Rentner, Kinder, sozialer Status) lassen sich auch Energieeffizienz-Charakteristika (z.B. Heizungstyp, Hausalter und -grösse, Geräte im Haushalt) sowie Einstellungen (z.B. gegenüber erneuerbaren Energieträgern, Interesse an Ökostrom oder an Solaranlagen) mit den entwickelten Algorithmen abschätzen. Mit Hilfe der Projektergebnisse können autorisierte Energiedienstleister wirkungsvolle und skalierbare Effizienzkampagnen realisieren. Zugleich unterstützen die Projektergebnisse eine faktenbasierte Diskussion über die Vorteile (z.B. Steigerung der Energieeffizienz) und Kosten (z.B. Wirkung auf die Privatsphäre) solcher Verfahren.

Résumé

Des compteurs électriques intelligents permettent l'enregistrement de profils de charge individuels à fréquence élevée (typiquement à intervalles de 15 minutes). L'objectif du projet est de développer des méthodes d'apprentissage mécanique afin d'identifier automatiquement des caractéristiques propres aux foyers concernés à partir des profils de charges et d'informations complémentaires pertinentes (météo, données sociodémographiques, adresse, etc.). Ces informations pourront être mises à profit dans le cadre d'une consultation énergétique individuelle et spécifique.

Le procédé de classification développé dans le cadre du projet a permis de prédire 38 caractéristiques de foyers, en partie avec une grande fiabilité (plus de 70%) en se basant sur les profils de charge et les données complémentaires disponibles. Les algorithmes développés permettent non seulement d'estimer la situation de vie des habitants (familles, retraités, enfants, statut social...) mais aussi les caractéristiques d'efficacité énergétique (type de chauffage, date de construction et taille du logement, appareils ménagers...) ainsi que les opinions (au sujet des sources d'énergie renouvelable, de l'intérêt pour l'électricité verte ou les panneaux solaires...). Les résultats du projet permettent d'adapter les campagnes sur l'efficacité énergétique et d'améliorer leur impact. Dans le même temps, les résultats invitent à une discussion basée sur les faits au sujet des avantages (par exemple de l'amélioration de l'efficacité énergétique) et des inconvénients (entre autres l'influence sur la vie privée) de tels procédés.



Abstract

Smart electricity meters allow for capturing consumption data of individual households at a high resolution in time (typically at 15-minute intervals). The key objective of this project is to develop further and evaluate feature extraction and machine learning techniques for automatic identification of household properties based on electricity load profiles and additional consumption-related information (weather, socio-demographic data, holidays, etc.). The gained information shall render highly targeted and scalable energy efficiency services possible.

The developed classification methods enable recognition of 38 household characteristics with accuracy of partially above 70%, based on smart meter load profiles and additional freely available data and under adherence to data privacy and security regulations. The characteristics describe inhabitants' life situation (e.g., families, retirees, children, social status), energy efficiency (e.g., heating type, age and size of house, appliances in the household) as well as attitudes (e.g., toward renewable energy sources, interest on green electricity or solar panels). The project results will help authorized energy service providers in realization of effective and scalable energy efficiency campaigns. At the same time, the results support a fact-based discussion of advantages (e.g., enhancement of energy efficiency) and costs (e.g., privacy implications) of such approaches.



Inhaltsverzeichnis

Zusammenfassung	3
Résumé	3
Abstract	4
Inhaltsverzeichnis	5
Abkürzungsverzeichnis	7
1. Ausgangslage	8
2. Ziel der Arbeit	8
3. Abgrenzung	10
4. Vorgehen	11
4.1 Erhebung von Datenquellen für die Haushaltsklassifikation	11
4.1.1 Daten des Projektpraxispartners Arbon Energie AG	12
4.1.2 Wetter- und Klimadaten	16
4.1.3 Kundenstammdaten	16
4.1.4 Sozialdemographische Daten	16
4.1.5 Daten aus Geographischen Informationssystemen (GIS)	17
4.1.6 Feiertags- und Ferienkalender für Datenbereinigung	17
4.2 Anpassung der Feature-Extraktionsalgorithmen.....	17
4.2.1 Feature-Extraktion	17
4.2.2 Korrelationsanalyse zwischen Smart-Meter-Daten und zusätzlichen Daten	21
4.2.3 Anwendung von Feature-Selektion (FS) Algorithmen	22
4.2.4 Evaluation von Features, die oft von Feature-Selektion-Algorithmen ausgewählt werden ...	24
4.2.5 Performanz von Feature-Selektion-Algorithmen.....	33
4.3 Entwicklung der Haushaltsklassifikations-Algorithmen	33
4.3.1 Anwendung von Klassifikationsalgorithmen	34
4.3.2 Umgang mit unbalancierten Klassen	34
4.3.3 Steigerung der Anzahl der geschätzten Charakteristika	35
4.3.4 Anwendung und Weiterentwicklung von Supervised-Learning-Verfahren für mehrdimensionale Daten	37
4.3.5 Repetitive Anwendung der Verfahren, Nutzung von Majority-Voting zur Klassenbestimmung.....	37
4.3.6 Betrachtung der Stabilität der Ergebnisse bei Anwendung eines Standard-Klassifizierers ..	38
4.3.7 Test der Übertragbarkeit der Klassifizierer auf verschiedene geographische Regionen	38
4.4 Tool-Anpassung	38
4.4.1 Beschreibung des Software-Tools	38
4.4.2 Dokumentation für Endbenutzer	40
4.5 Validierung und Feldtests.....	40



5. Ergebnisse	41
5.1 Erweiterung der erkennbaren Haushaltscharakteristika	41
5.2 Deskriptive Ergebnisse aus dem Vergleich der Datensätze aus den beiden Projektphasen	42
5.5 Test der Übertragbarkeit des Tools auf die Schweiz	50
6. Würdigung der Ergebnisse	51
7. Schlussfolgerungen	52
8. Ausblick	52
9. Referenzen	52
9.1 Veröffentlichungen über die Ergebnisse dieses Projekts	52
9.2 Referenzierte Literatur	53
Anhang	55
Anhang 1: Weitere Tabellen	55
Anhang 2: Datenschutzkonzept	63
Anhang 3: Online-Umfragebogen	66
Anhang 4: Vergleich der Anteile an mehrfachverglasten Fenstern in Umfragen in Arbon und Irland	77



Abkürzungsverzeichnis

API	Application Programming Interface
BFE	Bundesamt für Energie
BFS	Bundesamt für Statistik
BRG	Biased Random Guess
FS	Feature-Selektion
GIS	Geographisches Informationssystem
kNN	k-Nearest-Neighbor (Klassifikationsalgorithmus)
LDA	Linear Discriminant Analysis (Klassifikationsalgorithmus)
NIALM	Non Intrusive Appliance Load Monitoring
OSM	OpenStreetMap
SMD	Smart-Meter-Daten
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine (Klassifikationsalgorithmus)
UML	Unified Modelling Language
WGS	World Geodetic System



1. Ausgangslage

Die vielerorts fortschreitende Marktliberalisierung, der intensiver werdende Wettbewerb unter Energiedienstleistern und der Druck der Gesetzgeber sowie der Gesellschaft auf die Branche, Energieeffizienzziele zu erreichen, zwingen diese in zahlreichen Ländern bereits heute zum Aufbau eines besseren Verständnisses des Kundeverhaltens. Dadurch wollen die Energiedienstleister schlussendlich die Kundenloyalität und -Bindung erhöhen, Vertriebskanäle für Zusatzprodukte und -dienstleistungen aufbauen, und die ambitionierten Energieeffizienzziele möglichst umsatzneutral und kostengünstig erreichen. Ein weiterer Trend ist die Digitalisierung, die auch vor den Kundenbeziehungen nicht Halt macht. Energiedienstleister verfügen zunehmend über umfangreichere Verbrauchsdaten ihrer Haushaltskunden, insbesondere bedingt durch den fortschreitenden Ausbau von intelligenten Zählern. Eine zentrale Herausforderung für Energiedienstleister wird in der Zukunft darin bestehen, die steigende Menge an Smart-Meter-Daten intelligent zu nutzen und für Effizienzkampagnen und neue Dienstleistungen gewinnbringend einzusetzen.

In früheren Arbeiten unserer Forschungsgruppe wurden Haushaltsklassifikationsmethoden als rudimentäre Algorithmen entwickelt (Beckel et al., 2012; 2013). Es konnte gezeigt werden, dass Stromverbrauchsdaten als Zeitreihen genutzt werden können, um Eigenschaften privater Haushalte (Single-/Familienhaushalt, Alter des Hauses, Anzahl der Kinder, Anzahl der Bewohner, Art des Herds, etc.) zu erschliessen. Im diesem Forschungsprojekt wurde dieser Ansatz weiterentwickelt und in der Schweiz getestet.

2. Ziel der Arbeit

Übergeordnetes Ziel des Projektes ist die Entwicklung von Methoden zur automatisierten Erkennung von energieeffizienzrelevanten Charakteristika individueller Haushalte (z.B. Anzahl der Bewohner, Grösse und Alter der Wohnung, Nutzung elektrischer Heizungen usw.) anhand von Smart-Meter-Daten, historischen Rechnungen, Wetterindikatoren, Nutzerangaben usw. (siehe Abbildung 1). Durch Kenntnis dieser Charakteristika lässt sich die Wirkung von Energieeffizienz-Kampagnen steigern (z.B. durch die Wahl relevanter Vergleichsgruppen, die Selektion zielgerichteter Empfehlungen, das Setzen realistischer aber ambitionierter Ziele, usw.) und energierelevante Produkte und Dienstleistungen zielgerichteter vermarkten. Weiter kann die konventionelle Energieberatung die gewonnenen Informationen nutzen, um besonders beratungsrelevante Haushalte bevorzugt anzusprechen und so (vor allem durch eine Reduktion der Streuverluste) die Kosten je gesparter Kilowattstunde zu reduzieren. Gesamthaft schätzen wir, dass durch den konsequenten Einsatz der gewonnenen Informationen ein energetisches Potenzial in der Schweiz von jährlich ca. 180 GWh gehoben werden kann, entsprechend etwa einem Prozentpunkt des Stromverbrauchs privater Haushalte in den Jahren 2013 und 2014 (BFE, 2015, S. 28), resultierend aus zusätzlichen Einsparungen gegenüber konventioneller Smart-Meter-Anwendungen in Haushalten.

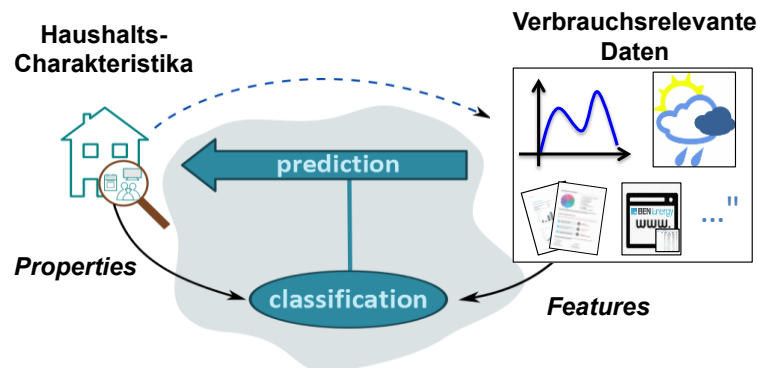


Abbildung 1: Erkennung von Haushaltseigenschaften als Machine-Learning-Ansatz

Abbildung 1 visualisiert das Haushaltsklassifikationssystem, welches im Rahmen des Projektes realisiert wurde: Aus Stromverbrauchsdaten und weiteren Daten werden *Kennzahlen* (sog. *Features*) berechnet, welche unter Verwendung von Machine-Learning-Algorithmen, zur Vorhersage von *Haushaltseigenschaften* (auch *Properties*, *Merkmale* oder *Haushaltscharakteristika* genannt) dienen.

Die **Definition von Features** ist dabei eine entscheidende analytische Aufgabe, bei der es darum geht, eine überschaubare Menge von wichtigen Kennzahlen aus Daten zu definieren (Sodenkamp et al., 2015). Dabei wird domänenspezifisches Fachwissen benötigt.

Auch die **Definition von Klassenlabels für die Haushaltscharakteristika** und die Erhebung solcher Daten mittels Umfragen bei Verbrauchern ist ein Bestandteil der Systementwicklung und benötigt analytisches und domänenspezifisches Fachwissen (Beckel et al., 2012, Sodenkamp et al., 2015).

Das Gesamtprojekt wurde in zwei Projektphasen untergliedert (siehe Tabelle 1). In der ersten Projektphase wurden **rudimentäre Verfahren des maschinellen Lernens weiterentwickelt und** mit bestehenden Trainings- und Validierungsdaten von über 4'000 Haushalten aus Irland **erprobt**. Die Merkmale umfassen die Charakteristika *Single-Haushalt*, *Kinder*, *Berufsstand* (jeweils ja/nein), *Anzahl der Einwohner*, *Trocknernutzung*, *Spülmaschinennutzung*, *Alter des Hauses*, *Wohnfläche* (über- vs. unterhalb des Medians), sowie *Informationen zu typischen Anwesenheitszeiten* (etwa für Lastverschiebungskampagnen). Je nach Merkmal haben wir eine Treffsicherheit (Accuracy) von 70% bis 90% angestrebt. Als Datengrundlage dienen Zeitreihen von Stromverbräuchen, welche bei gängigen Smart-Meter-Infrastrukturen erfasst und übertragen werden. Dadurch kann die konventionelle Energieberatung die gewonnenen Informationen nutzen, um besonders beratungsrelevante Haushalte bevorzugt anzusprechen und so (vor allem durch eine Reduktion der Streuverluste) die Kosten je gesparter Kilowattstunde zu reduzieren.

Ein weiteres Ziel der ersten Projektphase, folgend auf die Verbesserung der Methoden, war die Steigerung der **Effizienz und Skalierbarkeit der Algorithmen** zur Haushaltsklassifikation. Hierbei sollen mit gängiger Hardware (z.B. mit einem Server mit Anschaffungskosten von 15'000 CHF) 10'000 Haushalte pro Stunde klassifiziert werden können. Weiter sollen die Methoden in Form eines nutzerfreundlichen Software-Pakets verfügbar gemacht werden. Dieses soll einen einfachen Import der Daten und eine einfache Parametrisierung der Algorithmen ermöglichen sowie unterschiedliche Filtermethoden für Merkmale und Haushaltstypen enthalten.

Das Ziel der zweiten Projektphase war die Erweiterung der bestehenden Smart-Meter-Klassifikationsalgorithmen auf eine **mehrdimensionale Vorhersage mit zusätzlichen verbrauchsre-**



levanten Daten erweitert. Dazu wurden Wetterinformationen (Temperatur, Niederschlag, Windstärke- und Richtung usw.), Kalenderdaten (Feiertage, Schulferien usw.), Daten aus geografischen Informationssystemen (Landnutzung, Gebäudetyp, Infrastruktur, usw.) sowie sozio-demographische Statistiken (Altersstruktur, Familien, Kinder usw.) hinzugezogen. Schlussendlich musste die Qualität der Merkmalerkennung verbessert werden als auch die **Anzahl der erkannten Haushaltscharakteristika mindestens 30 erreichen**. Somit reichen den Energiedienstleistern alleine die Verbrauchsdaten und Haushaltsadresse aus, um mit (hoher) Sicherheit (zum Teil über 70 Prozent) Kennzahlen zur Lebenssituation (z. B. Familien, Rentner, Kinder, sozialer Status), zur Energieeffizienz (z. B. Heizungstyp, Hausalter und -Grösse, Geräte im Haushalt) und zur Affinität bezüglich erneuerbarer Energieträger (z. B. Interesse an Ökostrom oder Solaranlagen) ihrer Kunden unter Einhaltung von Datenschutzbestimmungen zu erkennen.

Neben der Verbesserung der Qualität der Klassifikation und der Steigerung der Anzahl der zu erkennenden Haushaltscharakteristika sollten die **Algorithmen mit Daten eines Schweizerischen Versorgers validiert werden**. Die neuen Algorithmen wurden mit Daten des schweizerischen Energieversorgers Arbon Energie AG weiterentwickelt und validiert. Ein wichtiges weiteres Ergebnis ist die Bewertung der **Übertragbarkeit von Klassifizierern**. Dabei wurden Algorithmen mit Daten eines (Irishen) Versorgers trainiert, und auf den Daten eines anderen (Schweizerischen) Versorgers getestet. Dadurch lassen sich Aussagen über Machbarkeit und Kosten bei einer Vergrößerung des geographischen Anwendungsbereiches entsprechender Software-Lösungen ableiten.

Schlussendliches Projektziel war es, die entwickelten Algorithmen auf Daten von Arbon Energie AG anzuwenden und in einem Pilottest zu bewerten. Die Optimierung von Rücklaufquoten bei einer Fiber-to-the-Home Cross-Selling Kampagne wurde für die Feldstudie gewählt.

Tabelle 1: Ziele der zwei Projektphasen

	SGDA - I	SGDA - II
	1 Apr 2014	31 März 2015
		31 Jan 2016
Projektlaufzeit	12 Monate	10 Monate
Ziele	<ul style="list-style-type: none"> • Entwicklung eines rudimentären Haushaltsklassifizierers • 12 Haushalts-Charakteristika • Vorhersage-Genauigkeit über 70% 	<ul style="list-style-type: none"> • Mehrdimensionale Klassifikation • Anpassung des Tools für die Schweiz und Test der Übertragbarkeit • Erweiterung der erkennbaren Haushalts-Charakteristika (18 zusätzliche) • Steigerung der Genauigkeit • Pilotstudie

3. Abgrenzung

Angrenzende Arbeiten nutzen zentral erfasste feingranulare Zeitreihendaten von Haushaltsverbräuchen zur Identifikation einzelner Geräte. Diese Ansätze werden unter dem Begriff NIALM (Non Intrusive Appliance Load Monitoring) zusammengefasst. Mit Hilfe dieser Ansätze kann die Effizienz einzel-



ner Geräte bewertet werden. Allerdings erfordern NIALM-Techniken nach bisherigem Forschungsstand sehr hochauflösende Zählerdaten im Hertz- bis Megahertz-Bereich. Diese werden durch die Smart-Meter-Infrastruktur, wie sie für die nächsten Jahre vorgesehen ist, jedoch nicht bereitgestellt. Daher ist ein flächendeckender Einsatz der Technologie bis auf weiteres nicht zu erwarten. Unser Ziel ist es, auf der Basis gängiger Smart-Meter-Daten die bestmöglichen Werkzeuge zu entwickeln, welche zur effektiven Bereitstellung von Energieeffizienzangeboten benötigt werden.

Smart-Meter-Daten (mit Messabständen von 15-min oder grösser) wurden zur Kundensegmentierung bisher mit beschreibenden Segmentierungsverfahren verwendet, welche Kunden anhand ähnlicher Verhaltensmuster in Cluster einteilen (Figueiredo et al., 2005; Kwac et al., 2013; Rá a en et al., 2008; Sa´chez et al., 2009; Verdu et al., 2006). Diese Cluster benötigen jedoch eine manuelle Interpretation durch Experten und sind demnach nicht für eine automatisierte Energieberatung nutzbar.

4. Vorgehen

4.1 Erhebung von Datenquellen für die Haushaltsklassifikation

Das erste Arbeitspaket behandelt die Identifikation und Akquisition von Datenquellen, welche die Grundlage für die Erweiterung des Klassifikationsalgorithmus sind. Die verwendeten Datenquellen und deren Einbettung in die weitere Methodik ist in Abbildung 2 dargestellt.

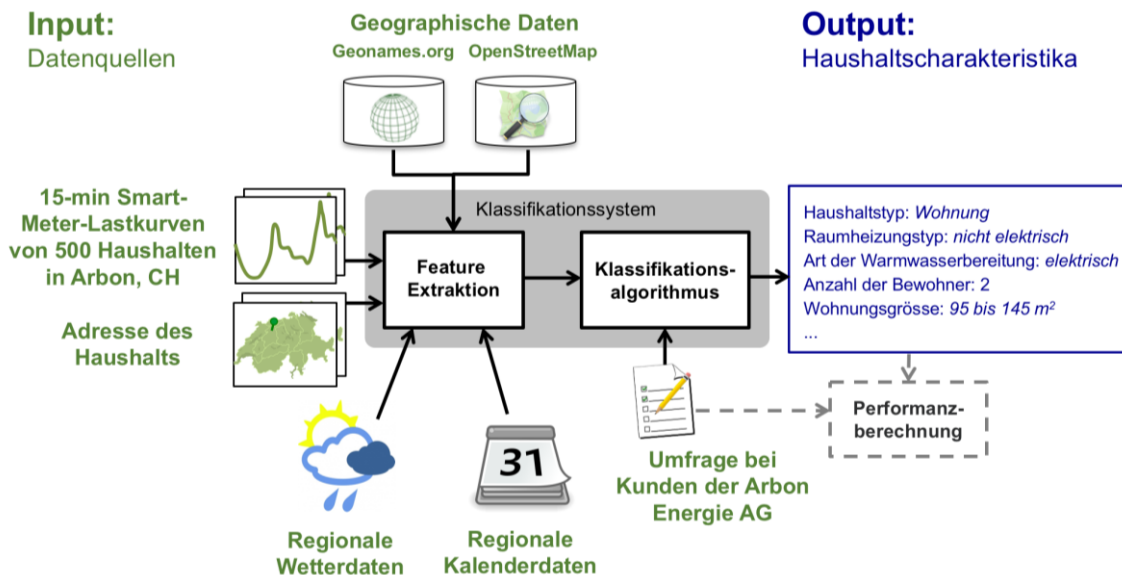


Abbildung 2: Schematischer Überblick über das Klassifikationssystem mit den Daten der Arbon Energie AG in der Schweiz



4.1.1 Daten des Projektpraxispartners Arbon Energie AG

Im Zuge der ersten Projektphase hat die Arbon Energie AG Smart-Meter-Daten zur Verfügung gestellt, welche für die spätere Verwendung vorbereitet wurden. Weiterhin wurde eine Umfrage vorbereitet und bei Kunden der Arbon Energie AG durchgeführt.

Smart-Meter-Daten

Von der Arbon Energie AG wurden 15-minütige Smart-Meter-Daten für den Zeitraum 01.06.2014 bis 31.05.2015 zur Verfügung gestellt. Die genaue Anzahl der Messstellen variiert von Monat zu Monat (zwischen 8'861 und 8'906). Für jeden Haushalt stehen 35'040 Datenpunkte zur Verfügung. Die Messdaten bestehen aus 12 Textdateien (eine CSV-Datei pro Monat) in der Formatierung, die beispielhaft in Tabelle 2 dargestellt ist. Die Variablen der Smart-Meter-Daten sind in der Tabelle 3 beschrieben.

Tabelle 2: Ein Schnappschuss der rohen Smart-Meter-Daten

	V1	V2	V3	V4	V5	V6
1	43179	100000001	2015-04-01 00:00:00.000	2.65900	Niedertarif	36146
2	3422	100000034	2015-04-01 00:00:00.000	2.94300	Niedertarif	5930
3	3422	100000034	2015-04-01 00:15:00.000	2.66300	Niedertarif	5930
4	3422	100000034	2015-04-01 00:30:00.000	0.12500	Niedertarif	5930
5	22955	100000054	2015-04-01 00:00:00.000	0.23300	Niedertarif	1810
6	3422	100000034	2015-04-01 00:45:00.000	0.11900	Niedertarif	5930
7	22955	100000054	2015-04-01 00:15:00.000	0.23500	Niedertarif	1810
8	22955	100000054	2015-04-01 00:30:00.000	0.23500	Niedertarif	1810
9	3422	100000034	2015-04-01 01:00:00.000	0.15700	Niedertarif	5930

Tabelle 3: Variablen im Smart-Meter-Datensatz der Arbon Energie AG

Attribut	Beschreibung	Skala	Verwendung
V1	Kundennummer	Nominal	Nein
V2	Zählernummer	Nominal	Nein
V3	Zeitstempel der Messung	Datum im ISO 8601-Format (Jahr-Monat-Tag), gefolgt von der Uhrzeit (Format: Stunde: Minute: Sekunde).	Ja
V4	Verbrauch in den vergangenen 15 min	Intervall, kWh	Ja
V5	Aktueller Tarif für den Messzeitraum	Kategorisch (Niedertarif, Hochtarif)	Ja
V6	Abo-ID (Vertragsnummer der Messstelle)	Nominal	Ja (anonymisiert)

Für die Zuordnung von Energieverbrauchsdaten und Kundenstammdaten (z.B. Adresse) wurde die Variable V6 verwendet. Die Variablen V1 und V2 wurden nicht benötigt.

In roher Form haben die Daten ein Volumen von 18,34 GB. Im ersten Datenvorbereitungsschritt wurden die Daten eingelesen und pro Tag eine Datei generiert, die alle 96 Messwerte für jeden Zähler enthält. Mit diesem Schritt konnte die Datengrösse auf 1,8 GB verringert werden.



Kundenumfragedaten

Für das Training der Algorithmen wurden Umfragedaten online erhoben. Um die Privatsphäre der Studienteilnehmer zu gewährleisten, haben wir mit der Arbon Energie AG eine **Datenschutzvereinbarung** abgeschlossen. Die von beiden Projektpartnern unterzeichnete Kopie des Datenschutzvereinbarungsdokuments haben wir in die Online-Umfrage beigefügt. Zustimmung der beschriebenen Bedingungen war die Voraussetzung für die Teilnahme an der Umfrage. Die Datenschutzvereinbarung ist als Anhang 2 beigefügt.

Daraufhin wurde eine **Online-Umfrage** mit der Software „surveygizmo“¹ erstellt und durchgeführt. Die Einladung zur Umfrage hat die Arbon Energie AG zusammen mit der Stromrechnung an alle Kunden in zwei Teilen gesendet (in den Kalenderwochen 25 und 29).

Neben den Haushaltseigenschaften wurden in der Kundenumfrage auch persönliche Einstellungen (z.B. Einstellung ggü. Energieeffizienz, Kaufbereitschaft ggü. Solaranlagen) von Kunden abgefragt. Dafür wurden vier validierte Skalen verwendet. So haben wir uns bei Frage 21 an der Skala zur Kaufintention nach Kim et al. (2007) orientiert, welche drei Fragen beinhaltet. Die Skala zur Einstellung gegenüber Energieeffizienz in Frage 25 wurde von Kaiser et al. (2007) übernommen. Die Kaufintention gegenüber Fiber-to-the-Home haben wir von der Skala der Kaufabsicht und Kaufwahrscheinlichkeit nach Juster (1966) übernommen und bezüglich der Zufriedenheit mit der Arbon Energie AG haben wir relevante Fragen aus der Skala von Walsh & Beatty (2007) verwendet.

Als Resultat haben 527 Kunden das Datenschutzkonzept akzeptiert und ein Fragebogen mit 29 Fragen ausgefüllt (im Zeitraum 10.06.2015 - 15.09.2015). Die Rücklaufquote der Umfrage liegt damit bei 5,6%. 451 Teilnehmer haben den Bogen vollständig ausgefüllt.

Tabelle 4 enthält eine Auflistung der Fragen mit möglichen Antwortalternativen. Antworten auf die Fragen 27 bis 29 wurden gemäss Datenschutzkonzept in anonymisierter Form an Arbon Energie AG weitergegeben. Ein Ausdruck der Online-Umfrage ist als Anhang 3 beigefügt.

Tabelle 4: Fragen und Antwortalternativen der durchgeführten Kundenumfrage

	Fragen (entsprechend der Haushaltscharakteristika)	Antwortmöglichkeiten (entsprechen der Klassen)	Antworttyp N: Numerisch K: Kategorial F: Freitext L: Logisch
Seite 1: Allgemeines			
1	Akzeptanz der Datenschutzbedingungen	Ja, Nein (bei „Nein“ wird die Umfrage beendet)	L
2-7	Vorname, Nachname, Strasse, Hausnummer, PLZ, Ort		F
Seite 2: Personen im Haushalt			
8	Anzahl der Personen im Haushalt (ohne Kinder bis einschliesslich 16 Jahre)	1, 2, 3, 4, 5 oder mehr	K
9	Anzahl der Kinder im Haushalt (bis einschliesslich 16 Jahre)	0, 1, 2, 3, 4, 5 oder mehr	K
10	Wie viele Personen sind regelmässig tagsüber zuhause (z.B. 5-6 Stunden pro Tag)	0, 1, 2, 3, 4, 5 oder mehr	K

¹ <https://www.surveygizmo.com/>, zuletzt abgerufen am 09.11.2015



Seite 3: Wohnungseigenschaften			
11	Wohnungstyp	Wohnung in einem Mehrfamilienhaus, Doppelhaushälfte, Einfamilienhaus (freistehend), Reihenhaus	K
12	Wohnfläche (ggf. geschätzt)	Quadratmeter	N
13	Eigentumsverhältnisse	Miete, Eigentum	K
14	Kennen Sie das Baujahr Ihres Hauses?	Ja, Nein	L
14a	Baujahr des Hauses (Kategorien)	neuer als 5, 10, 30, 75, älter als 75 Jahre	K
14b	Baujahr des Hauses (genau bzw. geschätzt)	Baujahr	N
Seite 4: Heizung und Geräte im Haushalt			
15	Art der Heizung (getrennt nach Haupt-, Nebenheizung und Wassererwärmung)	Erdgas, Heizöl, Elektrospeicherheizung, Zentralheizung (Mehrfamilienhaus), Fernwärme, Wärmepumpe, Solarthermie, Pellets, Stückholz, andere / keine	K
16	Wann wurde Ihre Heizung installiert (getrennt nach Haupt- Nebenheizung und Wassererwärmung)	Installationsjahr	N
17a	Wie viele der folgenden Geräte sind in Ihrem Haushalt vorhanden (Elektroherd, Kühlschrank, separater Gefrierschrank, Waschmaschine, Wäschetrockner, Geschirrspüler, Fernseher)	0, 1, 2, 3 oder mehr	K
17b	Schätzen Sie das Alter der Geräte in Jahren (Elektroherd, Kühlschrank, separater Gefrierschrank, Waschmaschine, Wäschetrockner, Geschirrspüler, Fernseher)	2, 5, 7, 10, 15, älter als 15 Jahre	K
Seite 5: Solar- und Geothermie-Potential			
18a	Haben Sie eine der folgenden Anlagen installiert: Photovoltaikanlage, Solarthermie, Luft-Wärmepumpe, Erdsonde-Wärmepumpe	Ja, Nein	K
18b	Wenn ja, Jahr des Einbaus	Installationsjahr	N
19	Wie stark ist Ihr Dach geneigt? (Mit Abbildungen)	0-40°, 40-70°, 70-90°	K
20	Zeigt eine Seite Ihres Daches zur Mittagssonne? (Mit Abbildungen)	Nord/Süd, Süd-West/Süd-Ost, Ost/West, weiss nicht	K
21	Ich könnte mir vorstellen, in den nächsten 1-2 Jahren eine Solaranlage anzuschaffen	7-wertige Likert Skala	K
	Ich beabsichtige, in den nächsten 1-2 Jahren eine Solaranlage anzuschaffen	7-wertige Likert Skala	K
	Ich plane, in den nächsten 1-2 Jahren eine Solaranlage anzuschaffen	7-wertige Likert Skala	K
Seite 6: Energieeffizienz			
22	Anteil der Energiesparlampen (geschätzt)	100%, ~75%, ~50%, ~25%, keine	K
23	Anteil der Doppelt oder dreifachverglasten Fenster (geschätzt)	100%, ~75%, ~50%, ~25%, keine	K
24	Welche der folgenden Energiesparmassnahmen wurde in den letzten 15 Jahren umgesetzt? (Dämmung von Dach oder Obergeschoss, Gebäudeisolierung, Kellerisolierung, Fenstererneuerung, keine, weiss nicht)	ja, nein	K



25	Wenn ich Pullover oder Hosen einen Tag lang getragen habe, kommen sie immer in die Wäsche.	7-wertige Likert Skala	K
	Wenn ich als letzte Person den Raum verlasse, mache ich immer das Licht aus.	7-wertige Likert Skala	K
	Ich lasse strombetriebene Geräte (TV, Stereoanlagen, Drucker, ...) immer auf Standby laufen.	7-wertige Likert Skala	K
	Im Winter drehe ich das Thermostat runter, wenn ich die Wohnung für mehr als vier Stunden verlasse.	7-wertige Likert Skala	K
	Im Winter ist es in meiner Wohnung immer warm genug, um nur ein T-Shirt zu tragen.	7-wertige Likert Skala	K
	In Hotels lasse ich die Handtücher immer täglich austauschen.	7-wertige Likert Skala	K
26	Mein Haushalt spart Energie, um der Umwelt zu helfen.	7-wertige Likert Skala	K
	Mein Haushalt spart Energie, um Geld zu sparen.	7-wertige Likert Skala	K
	Mein Haushalt spart Energie, um den Kindern ein gutes Vorbild zu sein.	7-wertige Likert Skala	K
	Mein Haushalt spart Energie, weil wir generell sparsam leben.	7-wertige Likert Skala	K
	Mein Haushalt spart Energie, weil es in der Familie oder im Freundeskreis so üblich ist.	7-wertige Likert Skala	K
	Mein Haushalt interessiert sich für neue Technologien.	7-wertige Likert Skala	K
Seite 7: Zufriedenheit mit Arbon Energie & Interesse an FTTH			
27	Zufriedenheit mit der Arbon Energie	7-wertige Likert Skala	K
28	Kaufbereitschaft für Fiber-to-the-Home	10-teilige Skala zur Kaufbereitschaft	K
29	Anmerkungen		F

Datenverknüpfung von Smart-Meter-Daten und Umfragedaten

Die Verknüpfung von Smart-Meter-Daten und Umfragedaten fand am 03.11.2015 in den Geschäftsräumen der Arbon Energie AG statt. Gemäss dem Datenschutzkonzept wurden die Kundennummern und alle kundenbezogenen Informationen nach der erfolgreichen Datenverknüpfung gelöscht.

Bei der Verknüpfung der Daten konnte der Grossteil der Umfrage-Antworten mit den Kundendaten und den Smart-Meter-Daten verknüpft werden. Bei einigen Haushalten war jedoch keine Zuordnung möglich. Die Ursachen für eine fehlende Verknüpfbarkeit sind in Tabelle 5 dargestellt.

Tabelle 5: Übersicht über die Datenverknüpfung

Fall	Anzahl betroffener Umfragen	Gründe
Eindeutige Verknüpfung gefunden (Umfrage & Kundendaten)	476 Haushalte	
Nicht eindeutige Verknüpfung (Umfrage & Kundendaten)	53 Haushalte	Detaillierte Untersuchung nötig
Keine Verknüpfung möglich (Umfrage & Kundendaten)	29 Haushalte	Gründe: 28x Name nicht gefunden, 1x Unsinnige Angaben in der Umfrage
Keine Smart-Meter-Daten zu den Umfragedaten	Unterschiedlich pro Woche	



4.1.2 Wetter- und Klimadaten

Historische Wetterdaten sind in stündlicher oder dreistündlicher Auflösung für den betrachteten Zeitraum verfügbar. Dabei greifen wir auf die Datengrundlage des US National Climatic Data Center² zurück, welche die Daten kostenfrei zu nichtkommerziellen Zwecken zum Download anbieten. Für die Schweiz sind die Variablen **Temperatur**, **Wind** (Richtung und Geschwindigkeit), **Niederschlag**, **Luftdruck**, sowie der **Bedeckungsgrad des Himmels** von 175 Wetterstationen in der Datenbank verfügbar.

Aus den 175 verfügbaren Schweizer Stationen wurden die 5 nächsten Stationen ausgewählt (wie in Abbildung 3 dargestellt). Die Wetter-Variablen wurden für die gewählten Stationen gemittelt, um die Wetterdaten für den Ort Arbon zu bestimmen.

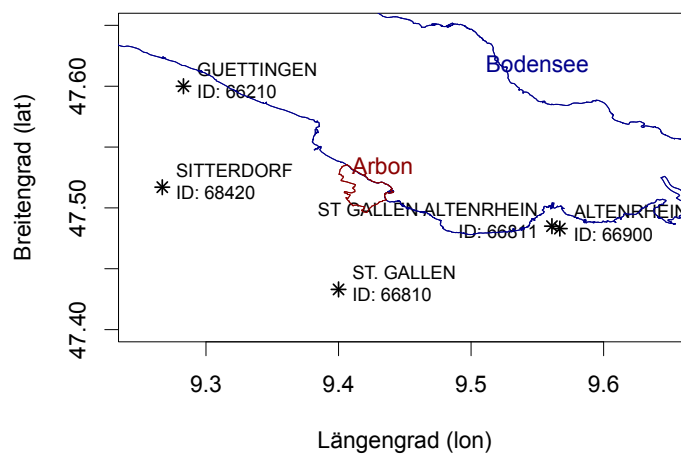


Abbildung 3: Kartendarstellung der fünf nächsten Wetterstationen um Arbon

4.1.3 Kundenstammdaten

Im Rahmen der Umfrage haben die Teilnehmer ihre **Haushaltsadressen** eingetragen. Aus den Smart-Meter-Daten kann der tageszeitabhängige **Tarif** (Hochtarif, Niedertarif) extrahiert werden. Weitere Informationen über den Haushalt sind nicht bekannt.

4.1.4 Sozialdemographische Daten

Das Bundesamt für Statistik bietet 527 Datensätze online über das Web-Portal „STAT TAB“³ an. In diesen Datensätzen sind folgende energieverbrauchs-relevante Statistiken enthalten:

- Gebäude- und Wohnungserhebung (bis 2000),
- Gebäude- und Wohnungsstatistik GWS (seit 2009),
- Leerwohnungszählung,
- Strukturhebung und
- Arealstatistik der Schweiz.

Die statistischen Erhebungen werden auf Ebene einzelner Gemeinden veröffentlicht.

² National Climate Data Center, NCDC DSI 3505. (Online, zuletzt aufgerufen am 09.11.2015). <https://gis.ncdc.noaa.gov/geoportal/catalog/search/resource/details.page?id=gov.noaa.ncdc:C00532>

³ https://www.pxweb.bfs.admin.ch/default.aspx?px_language=de, last access 04.07.2015



4.1.5 Daten aus Geographischen Informationssystemen (GIS)

Es wurden Daten aus verschiedenen frei verfügbaren Geoinformationssystemen herangezogen:

- Google Maps (Geocoding API, Elevation API) zur Umwandlung der Adressen in Geokoordinaten
- OpenStreetMap als Datenquelle für zahlreiche geographische Informationen: Verkehrsnetz, Gebäude, Landnutzungsinformationen, Points of Interests (öffentliche Einrichtungen, Einkaufsmöglichkeiten, Gaststätten usw.)
- GeoNames: Distanzen zu Stadtzentren verschiedener Grösse (1'000, 5'000 und 15'000 Einwohner)

4.1.6 Feiertags- und Ferienkalender für Datenbereinigung

Für die Bestimmung der Feiertage greifen wir auf die Veröffentlichung des Bundesamtes für Justiz „*Gesetzliche Feiertage und Tage, die in der Schweiz wie gesetzliche Feiertage behandelt werden - Verzeichnis gestützt auf Artikel 11 des Europäischen Übereinkommens vom 16. Mai 1972 über die Berechnung von Fristen (SR 0.221.122.3). In Kraft getreten für die Schweiz am 28. April 1983*“ zurück. Daraus ergeben sich folgende Feiertage für Arbon:

01.01.2015 (Neujahrstag), 02.01.2015 (Berchtoldstag), 03.04.2015 (Karfreitag), 06.04.2015 (Ostermontag), 01.05.2015 (Tag der Arbeit), 14.05.2015 (Auffahrt), 25.05.2015 (Pfingstmontag), 01.08.2014 (Bundesfeiertag), 25.12.2014 (Weihnachtstag), 26.12.2014 (Stephanstag)

Aus dem Ferienplan 2014-2018 der Schulgemeinden Steinach und Arbon ergeben sich folgende Ferientermine im Zeitraum der Betrachtung:

05.07.2014 – 11.08.2014 (Sommerferien), 04.10.2014 – 20.10.2014 (Herbstferien), 20.12.2014 – 05.01.2015 (Weihnachtsferien), 24.01.2015 – 02.02.2015 (Sportferien), 03.04.2015 – 20.04.2015 (Frühlingsferien), 14.05.2015 – 26.05.2015 (Vorsommerferien)

4.2 Anpassung der Feature-Extraktionsalgorithmen

Wir haben die bestehende Feature-Extraktionsalgorithmen des rudimentären Klassifikationsalgorithmus angepasst und erweitert. Darüber hinaus haben wir Feature-Selektion (FS) Algorithmen angewandt und die Performanz der Algorithmen für die Anwendung bestimmt.

4.2.1 Feature-Extraktion

Aus den 15-minütigen Stromverbrauchsdaten und den zusätzlichen Datenquellen wurden neue Features abgeleitet, die nachfolgend beschrieben werden.

Features aus Smart-Meter-Daten

Bei der Klassifikation mit Smart-Meter-Daten einer Woche (336 Werte bei 30-minütigen, bzw. 672 Werte bei 15-minütigen Messintervallen) konnte durch die Extraktion von Lastkurvenmerkmalen eine



starke Reduktion der Eingabedaten erreicht werden, welche neben der Steigerung der Ergebnisqualität auch Problemen mit hoher Dimensionalität / Overfitting entgegen.

In der ersten Projektphase wurden 93 Eigenschaften aus 30-minütigen Lastgangdaten für je eine Woche empirisch definiert und ihre automatische Ableitung aus den Lastkurven umgesetzt. Die Features wurden in vier Kategorien unterteilt:

1. **Verbrauchswerte** (z.B. minimaler Verbrauch, maximaler Verbrauch, Tagesverbrauch, Verbrauch am Wochenende)
2. **Statistische Momente** (z.B. Durchschnittsverbrauch am Tag, Korrelation von auf einander folgenden Tagen, Varianz des Verbrauchs am Wochenende)
3. **Zeitliche Kennzahlen** (z.B. Zeiten der Verbrauchsspitzen, der erste Zeitpunkt mit Verbrauch > 1 kWh)
4. **Verhältnisse** (z.B. Verbrauch in der Mittagszeit / Verbrauch am Abend, min. Verbrauch / max. Verbrauch, max. Verbrauch / Durchschnittsverbrauch)

In der zweiten Projektphase wurden die 30-minütigen Smart-Meter-Features auf die 15-minütigen Daten angepasst, wobei 62 neue Features entstanden sind (davon 5 in der Kategorie „Verbrauch“, 29 in der Kategorie „Statistische Momente“, 19 in der Kategorie „Zeitliche Kennzahlen“ und 9 in der Kategorie „Verhältnisse“). Bei der Anpassung von 30-minütigen auf 15-minütige Daten wurde in allen Fällen die Formeln zur Berechnung der 30-minütigen Features verwendet. Das Feature *t_above_0.5kwh* (Durchschnittliche Tageszeit der ersten Überschreitung der 0.5kWh-Grenze) wurde speziell für die 15-min Daten hinzugenommen, und die bisherige 1-kWh Grenze für 30-minütige Daten (*t_above_1kwh*) somit angepasst. Insgesamt haben sich 155 Features ergeben. Eine vollständige Liste dieser Smart-Meter-Features ist in Tabelle A1 im Anhang 1 zu finden.

Features aus Wetter- und Klimadaten

Wir haben Wetter- und Klimadaten auf zwei Weisen in die Klassifikation integriert. Im ersten Ansatz entsprechen die Wetter-Features den Messwerten aus dem Quelldatensatz. Es wurden mit diesem Ansatz erfolgreiche Tests mit 30-minütigen Werten für Temperatur, Windstärke und Niederschlag auf einem irischen Datensatz durchgeführt. Die unverarbeiteten Wettervariablen haben wir jedoch mit Stromverbrauchswerten normalisiert, um die Klassifikationsgüte zu verbessern. Die genaue Methodik dieser Wetterdaten-Normalisierung ist in unserer Publikation (Sodenkamp et al, 2015) beschrieben.

Im zweiten Ansatz haben wir auch Features aus der Korrelation zwischen Wetter- und Stromverbrauchsdaten extrahiert:

1. **cor_overall** (*Korrelation über die gesamte Woche*)
2. **cor_daily** (*Linearer Zusammenhang zwischen Wetter und Stromverbrauch an einem durchschnittlichen Tag*)
- 3-5. **cor_night, cor_daytime, cor_evening** (*Wettereinfluss zu verschiedenen Tageszeiten*)
6. **cor_minima** (*Korrelation der Minima*)
7. **cor_maxmin** (*Korrelation der Wetter-Minima mit den Verbrauchsspitzen*)
8. **cor_weekday_weekend** (*Verhältnis der Korrelation Wochentags zu den Wochenendtagen*)

Alle Features wurden auf die Wetter-Variablen Temperatur, Luftdruck, Windgeschwindigkeit, Niederschlag und der Bedeckungsgrad des Himmels einer Woche berechnet.



Features aus Kundenstammdaten

Aus den Kundenstammdaten konnten zwei Features extrahiert werden:

- **Adresse:** Die Kundenadresse (aus den Umfragedaten) wurde in eine geographische Koordinate im WGS 84 Format umgewandelt. Mit dieser Koordinate wurde die Anbindung geographischer Features ermöglicht.
- **Tarif:** Aus den Smart-Meter-Daten wurde der tageszeitabhängige Tarif (Hochtarif, Niedertarif) abgeleitet und als Feature aufgenommen. Da dieser Tarif jedoch für alle Kunden einheitlich ist, kann das Feature nicht als Prädiktor fungieren.

Die Kundenstammdaten fließen also indirekt in die Berechnung weiterer Features ein.

Features aus sozialdemographischen Daten

Aus den online verfügbaren statistischen Daten des Bundesamtes für Statistik (BFS) konnten 16 statistische Features extrahiert werden. Diese beinhalten:

- Relative Häufigkeit verschiedener Haustypen (Ein- oder Mehrfamilienhäuser, Wohn- und Geschäftshäuser, usw.)
- Mittelwerte des Alters von Gebäuden
- Mittelwerte der Grösse von Häusern
- Kennzahlen über die Bevölkerung (z.B., Alter, Geschlecht, Herkunft, Anzahl der Emigranten / Immigranten)

Das Vorgehen der Datenverknüpfung von Haushalten zu Statistikdaten ist in Abbildung 4 aufgezeigt. Alle Features werden auf Ebene der Gemeinden berechnet, da lediglich für diese Aggregationsstufe Daten vorliegen.

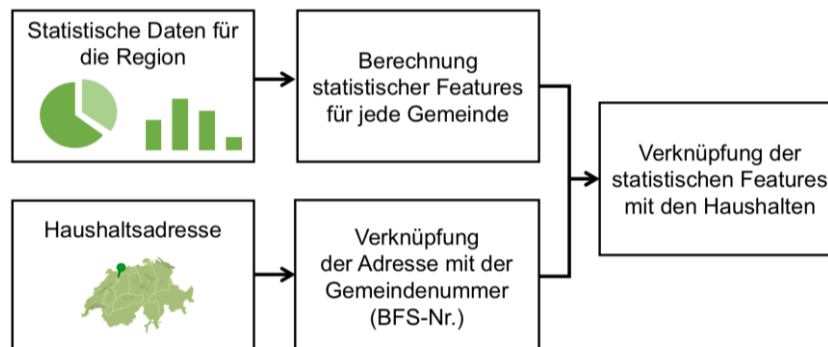


Abbildung 4: Schema für die Verknüpfung von Verbrauchsdaten und Statistiken

Features aus Geo-Informationsdaten

Zur Berechnung von geographischen Features wurden verschiedene GIS-Datenquellen kombiniert (Darstellung in Abbildung 5).



Wir unterscheiden bei den Features aus Geo-Informationen zwischen topologischen und semantischen Features.

Topologische Features beschreiben Struktur und Relation zwischen einem Haushalt und den benachbarten geometrisch-geographischen Objekten. Dabei wird nur die geometrische Form betrachtet. So werden zum Beispiel Seen, Felder, Landnutzungsregionen, usw. auf Polygone reduziert und interessante Orte, wie Briefkästen, Stadtzentrum oder Bibliotheken werden nur als Punkte interpretiert.

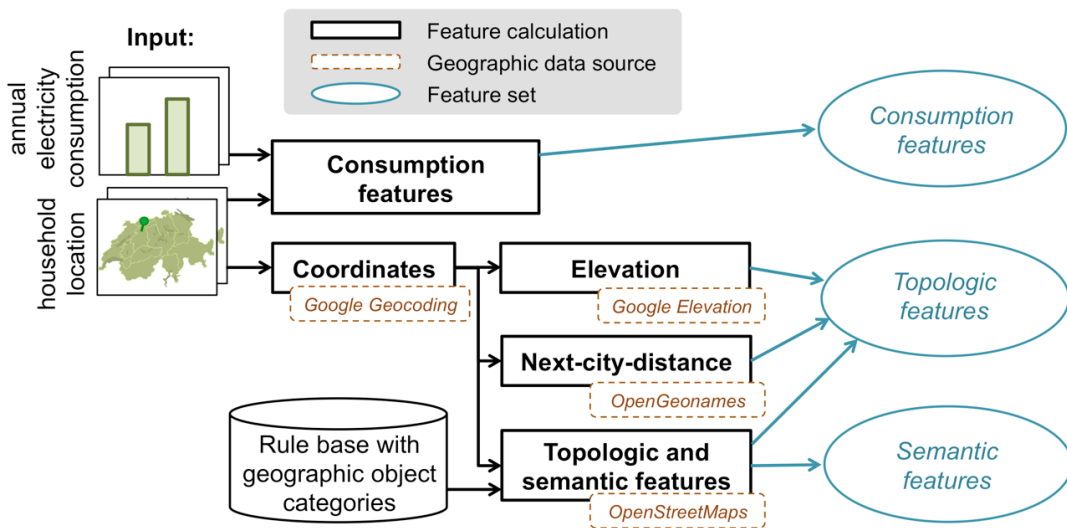


Abbildung 5: Übersicht über die Datenquellen und die Berechnung der geographischen Features

Semantische Features beschreiben die Bedeutung von geographischen Objekten innerhalb ihres geographischen Kontexts. So werden zum Beispiel Seen und Flüsse als Wasserflächen interpretiert und nicht nur als Polygone. Die Bedeutung der Objekte wird aus Beschriftungen, Namen oder Kategorien abgeleitet. Um die Vielzahl der Begriffe zu generalisieren, haben wir die 11 folgenden geographischen **Kategorien von Objekten** definiert:

1. Gebäude
2. Landnutzungsgebiete
3. Öffentliche Einrichtungen
4. Geschäfte und Büros
5. Gaststätten und Cafés
6. Öffentlicher Personenverkehr (Bahn-, Tram- und Buslinien, Flughäfen und Haltestellen)
7. Erholung und Freizeit
8. Kultur
9. Sehenswürdigkeiten
10. Objekte in ländlichen Regionen (Land- und Fortwirtschaft, Natur, usw.)
11. Verkehrssystem (Strassennetz, Schilder, Raststätten, usw.)

Die Kategorien 1 (Gebäude) und 2 (Landnutzungsgebiete) sind **Polygon-basiert**. Für sie können die folgenden Features berechnet werden:



Polygonbasierte Features für **Gebäude**:

- Grösse der Gebäude (Mittelwert und Varianz)
- Distanz zu Gebäuden im Umfeld (Mittelwert und Varianz)
- Typ des nächsten Gebäudes
- Typ des Gebäudes, das die Koordinaten des Haushalts beinhaltet
- Häufigster Gebäudetyp im Betrachteten Kartenausschnitt
- Gesamtgrundfläche der jeweiligen Gebäudetypen (Ein- oder Mehrfamilienhaus, gewerbliches Gebäude, nicht spezifiziert)

Polygon-basierte Features für **Landnutzungsgebiete**:

- Art der Landnutzung, in der sich der betrachtete Haushalt befindet
- Fläche des Gebiets, in der sich der betrachtete Haushalt befindet
- Anteil der jeweiligen Landnutzungstypen im Kartenausschnitt (Wohngebiet, Stadtgebiet, Landgebiet)

Für alle **Punkt-basierten Objektkategorien (Kategorien 3-11)** werden vier Features berechnet:

1. Häufigkeit: Die absolute Häufigkeit der Objekte in einer Objektkategorie
2. Gewichtete Häufigkeit: Die Summe der normalisierten Distanz (im Wertebereich [0;1], wobei 1 die absolute Nähe und 0 die maximale Euklidische Distanz im betrachteten Kartenausschnitt ist)
3. Mittlere Distanz: Der arithmetische Mittelwert aller Euklidischen Distanzen zu Objekten in einer Kategorie
4. Minimale Distanz: Die kleinste Euklidische Distanz zu Objekten in einer Kategorie

Eine vollständige Liste der implementierten geographischen Features mit detaillierter Beschreibung findet sich in Tabelle A2 in Anlage 1.

4.2.2 Korrelationsanalyse zwischen Smart-Meter-Daten und zusätzlichen Daten

Wir haben eine Korrelationsanalyse zwischen Stromverbrauchs- und Wetterdaten durchgeführt. Hierzu wurden 30-minütige Werte für Temperatur (S_1), und Windstärke (S_2), sowie tägliche Werte für Niederschlag (S_3) untersucht. Da in der Region typischerweise nur Heizgeräte und keine Klimaanlage vorhanden sind, wurde im Model für den Stromverbrauch (X) ein linearer Zusammenhang angenommen:

$$X = \alpha + \beta_1 * S_1 + \beta_2 * S_2 + \beta_3 * S_3 + \varepsilon.$$

Die Analyse der Irischen Daten zeigt eine Abhängigkeit der Korrelation von der Tageszeit (siehe Tabelle A3 und Tabelle A4 im Anhang 1). Der Zusammenhang zwischen der Temperatur und Stromverbrauch ist signifikant und auch in Abbildung 6 zu sehen.

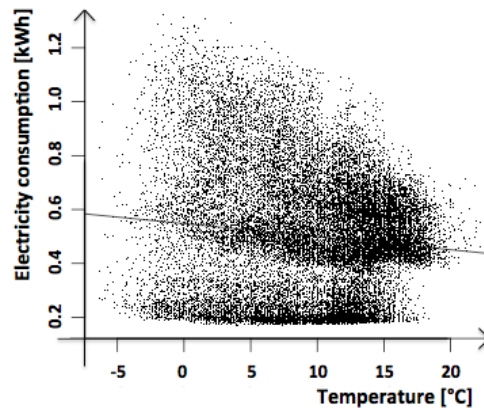


Abbildung 6: Temperatur und Stromverbrauch im Punktdiagramm mit der Regressionsgerade

Der Zusammenhang zwischen der Windstärke und Stromverbrauch ist auch signifikant aber die Effektgrösse ist vernachlässigbar (siehe Tabelle 6). Der Zusammenhang zwischen dem Niederschlag und den Stromverbrauch war nicht signifikant (siehe Tabelle 6). Mit dem gesamten Modell konnte ein Wert von 0.2257 für R^2 erreicht werden. R^2 bezeichnet den Anteil der erklärten Varianz durch das Regressionsmodell (D’Agostino, 1986). Für dieses Modell kann der erreichte Wert als hoch angesehen werden, da allein durch die Wetterdaten knapp 23% der Varianz in den Stromverbrauchsdaten erklärt werden kann.

Tabelle 6: Ergebnisse des kompletten linearen Modells

	Schätzwert	Std. Fehler	Signifikanz Index (**** für <0.0001)
A	0,6069	0,00779	***
β_1	-0,0032	0,0001555	***
β_2	0,0049	0,0003382	***
β_3	0,6884	0,6558081	

4.2.3 Anwendung von Feature-Selektion (FS) Algorithmen

Um für jede Haushaltseigenschaft die aussagekräftigsten Features auszuwählen, wurden Feature-Selektions-Algorithmen erprobt. Für die nicht-numerischen Features aus geografischen Informationssystemen wurden bestehende FS-Methoden erweitert und passende Korrelationsmasse implementiert. Es wurden zwei Kategorien von FS-Methoden verwendet:

Filter-Methoden

Eine Feature-Auswahl wird allein auf Grundlage der Feature-Vektoren getroffen, ohne eine Klassifikation durchzuführen. Folgende Filter-Methoden wurden angewendet:

- Korrelation zwischen Features
- Korrelation zwischen Features und Properties mittels Effektgrössen-Metriken (Cramers V, Kovarianz, Produkt-Moment-Korrelation, Pearsons η^2 , Punkt-biseriale-Korrelation)
- Statistische Tests auf unterschiedliche Verteilungen und somit die Aussagekraft von Features für spezifische Properties (Kolmogorov-Smirnov-Test, t-Test, χ^2 -Test)
- Wichtigkeit der Features (durch Random-Forest-Algorithmus ermittelt)
- Entropie und Entropiebasierte Masse (Information Gain, GainRatio)
- Konsistenz der Feature-Menge



Wrapper-Methoden

Sie führen die Klassifikation mit unterschiedlichen Feature Submengen mehrfach durch und vergleichen direkt die erreichte Klassifikationsgenauigkeit mit anderen Feature-Mengen. Es wurden folgende Methoden getestet:

- Forward-Selection
- Backward-Selection
- Sequential-Forward-Selection
- Parallel-Sequential-Forward-Selection

Es hat sich herausgestellt, dass für die einzelnen Haushaltseigenschaften die FS-Methoden unterschiedlich gut funktionieren. Tabelle 7 zeigt die Methoden, welche in unserer Implementierung die beste durchschnittliche Accuracy über alle Klassifizierer erreicht haben.

Tabelle 7: Beste FS-Methode für einzelne Haushaltseigenschaften

Haushaltseigenschaft	Beste FS Methode
Art der Raumheizung (<i>pSpaceHeatingType</i>)	ohne
Art der Warmwasserbereitung (<i>pWaterHeatingType</i>)	χ^2 -Test
Anzahl der Bewohner im Haushalt (<i>pNumResidentsB</i>)	Konsistenz der Feature-Menge
Alter der Raumheizung (<i>pSpaceHeatingAge</i>)	Forward-Selection
Grösse (Fläche) der Wohnung (<i>pLivingAreaB</i>)	ohne
Eigentumsverhältnisse (<i>pHouseOwnership</i>)	Korrelation zwischen Features
Alter des Hauses (<i>pHouseAge</i>)	Konsistenz der Feature-Menge
Haushaltstyp 4 Klassen (<i>pHouseholdType</i>)	Korrelation zwischen Features
Alter der Geräte (<i>pAppliancesAge</i>)	Konsistenz der Feature-Menge
Art der Kochstelle (<i>pCookingType</i>)	Korrelation zwischen Features
Wärmepumpe vorhanden (<i>pHeatPump</i>)	χ^2 -Test
Solaranlage vorhanden (<i>pSolar</i>)	χ^2 -Test
Anzahl der umgesetzten Energieeffizienzmassnahmen (<i>pEnergyEfficiencyMeasure</i>)	Korrelation zwischen Features
Anzahl der Geräte (<i>pAppliancesNum</i>)	Backward-Selection
Kaufinteresse für Solaranlagen (<i>pInterestSolar</i>)	ohne
Zufriedenheit mit dem Energieversorger (<i>pSatisfactionUtility</i>)	Backward-Selection
Kinder im Haushalt (<i>pChildren</i>)	ohne
Single-Haushalt (<i>pSingle</i>)	Konsistenz der Feature-Menge
Zwei Erwachsene mit Kind/Kindern (<i>pFamily</i>)	ohne
Haushaltstyp 2 Klassen (<i>pHouseholdType2</i>)	Konsistenz der Feature-Menge



4.2.4 Evaluation von Features, die oft von Feature-Selektion-Algorithmen ausgewählt werden

Nach der Klassifikation haben wir berechnet, wie häufig jedes Feature von FS-Methoden ausgewählt und in der Klassifikation verwendet wurde. Diese Auswertung ist somit eine Aggregation der FS-Methoden zur Bestimmung einer optimalen Feature-Menge. Die besten 10 Features für jede Haushaltseigenschaft kann Tabelle 8 entnommen werden.

Tabelle 8: Top-10 Features, welche für die Klassifikation einer Property häufig verwendet wurden

Haushaltscharakteristika	Gewählte Features und ihre Beschreibungen (<i>kursiv</i> als Variablennamen)
Art der Raumheizung (pSpaceHeatingType)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none"> - Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>); Verhältnis des Verbrauchs morgens zum Verbrauch abends ohne des minimalen Verbrauchs (<i>r15_evening_noon_no_min</i>) auf Basis von 15-minütigen Daten - Zeitliche Kennzahlen: Anzahl kleiner Verbrauchsspitzen auf Basis von 15-minütigen Daten (<i>t15_number_small_peaks</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none"> - Fläche des Landnutzungsgebiets „Stadtgebiet“ im Kartenausschnitt von 500x500m um den Haushalt (<i>totalarea.city</i>) - Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“, „Öffentlicher Personenverkehr“, „Öffentliche Einrichtungen“, „Gaststätten und Cafés“, „Geschäfte und Büros“ und „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.culture, mindist.transportation, mindist.publicInstitutions, mindist.food, mindist.business, mindist.recreation</i>)
Art der Warmwasserbereitung (pWaterHeatingType)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none"> - Verhältnisse: Verhältnis des Stromverbrauchs am Morgen zum Verbrauch am Abend an Wochentagen auf Basis von 30-minütigen Daten (<i>r30_wd_morning_noon</i>); Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>) - Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>) und mit einem Verbrauch oberhalb des Grundlast-Verbrauchs (<i>t15_time_above_base2</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none"> - Korrelation des Luftdrucks und des Niederschlags mit dem Stromverbrauch in der Nacht (<i>w_airPr_cor_night, w_prec_cor_night</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none"> - Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“, „Kultur“, und „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food, mindist.culture, mindist.recreation</i>) - Fläche des Landnutzungsgebiets „Stadtgebiet“ im Kartenausschnitt von 500x500m um den Haushalt (<i>totalarea.city</i>)
Anzahl der Bewohner im Haushalt (pNumResidentsB)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none"> - Statistische Momente: Anzahl an Verbrauchsspitzen in 15-minütigen Daten berechnet mit und ohne Glättung (<i>s15_num_big_peaks</i>,



	<p><i>s15_num_peaks</i>),</p> <ul style="list-style-type: none">- Zeitliche Kennzahlen: Anzahl der Verbrauchsspitzen, berechnet mittels des gleitenden gewichteten Mittelwerts (<i>t15_number_small_peaks</i>); Geschätztes Zeitintervall mit Grundlast-Verbrauch bei 15-minütigen Daten (<i>t15_const_time</i>), Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>)- Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none">- Korrelation des Niederschlags mit dem Stromverbrauch am Abend und im Durchschnitt pro Tag (<i>w_prec_cor_evening</i>, <i>w_prec_cor_daily</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentlicher Personenverkehr“ und „Öffentliche Einrichtungen“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.transportation</i>, <i>mindist.publicInstitutions</i>)
Alter der Raumheizung (pSpaceHeatingAge)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verbrauch: Durchschnittlicher Verbrauch am Morgen innerhalb von 30-minütigen Zeitintervallen (<i>c30_morning</i>) und innerhalb 15-minütigen Zeitintervallen abzüglich des Verbrauchs-Minimums (<i>c15_morning_no_min</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none">- Korrelation der Minima in der Windgeschwindigkeit und im Stromverbrauch (<i>w_windSp_cor_minima</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Transport“, „Verkehrssystem“ und „ländlicher Raum“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.transportation</i>, <i>mindist.roadSystem</i>, <i>mindist.countryside</i>)- Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Öffentlicher Personenverkehr“, „Verkehrssystem“ und „ländlicher Raum“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>meandist.transportation</i>, <i>meandist.roadSystem</i>, <i>meandist.countryside</i>)- Gesamte Grundfläche von Gebäuden im Kartenausschnitt von 500x500m um den Haushalt bei denen der Typ bekannt ist und kein Wohngebäude identifiziert, z.B. Industriegebäude, Einkaufszentren, öffentliche Einrichtungen (<i>totalarea.nonresidential</i>)
Grösse (Fläche) der Wohnung (pLivingAreaB)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verbrauch: Durchschnittlicher Verbrauch wochentags am Morgen und Abend innerhalb von 30-minütigen Zeitintervallen (<i>c30_wd_morning</i>, <i>c30_wd_noon</i>), durchschnittlicher Verbrauch am Morgen innerhalb von 15-minütigen Zeitintervallen (<i>c15_morning_no_min</i>)- Zeitliche Kennzahlen: Geschätzte Grundlast in 15-minütigen Zeitintervallen (<i>t15_value_min_guess</i>), Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>) und die Summe der Werte über der Grundlast-Grenze (<i>t15_value_above_base</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“ und „ländlicher Raum“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.recreation</i>, <i>mindist.countryside</i>)- Gesamte Grundfläche von Gebäuden im Kartenausschnitt von 500x500m die als Mehrfamilienhäuser gekennzeichnet sind (<i>totalarea.apartments</i>)



	<ul style="list-style-type: none">- Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>meandist.recreation</i>)
Eigentumsverhältnis (pHouseOwnership)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Statistische Momente: 25%-Quantil aller 15-minütigen Verbrauchsmessungen einer Woche (<i>s15_q1</i>), kleinste Verbrauchsmessung eines 15-minütigen Intervalls innerhalb der gesamten Woche (<i>s15_min</i>) und innerhalb der Wochentage (<i>s15_wd_min</i>), 60%-Quantil der Abweichungen zum vorhergehenden Messwerts (<i>s15_bg_variety</i>)- Zeitliche Kennzahlen: Schätzung der Grundlast auf Grundlage von 15-minütigen Verbrauchsmessungen (<i>t15_value_min_guess</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“, „Verkehrssystem“, „ländlicher Raum“ und „Geschäfte und Büros“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.recreation</i>, <i>mindist.roadSystem</i>, <i>mindist.countryside</i>, <i>mindist.business</i>)- Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>meandist.recreation</i>)
Alter des Hauses (pHouseAge)	<p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“, „Gaststätten und Cafés“, „Geschäfte und Büros“, „ländlicher Raum“, „Erholung und Freizeit“ und „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.culture</i>, <i>mindist.food</i>, <i>mindist.business</i>, <i>mindist.countryside</i>, <i>mindist.recreation</i>, <i>mindist.transportation</i>)- Gesamte Grundfläche von Gebäuden im Kartenausschnitt von 500x500m bei denen der Typ nicht spezifiziert ist (<i>totalarea.notspecified</i>), bei denen der Typ bekannt ist und kein Wohngebäude identifiziert, z.B. Industriegebäude, Einkaufszentren, öffentliche Einrichtungen (<i>totalarea.nonresidential</i>)- Fläche des Landnutzungsgebiets „Wohngebiet“ im Kartenausschnitt von 500x500m um den Haushalt (<i>totalarea.residential</i>)- Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Kultur“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>meandist.culture</i>)
Haushaltstyp 4 Klassen (pHouseholdType)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“, „Erholung und Freizeit“, „Geschäfte und Büros“, „ländlicher Raum“, und „Öffentliche Einrichtungen“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food</i>, <i>mindist.recreation</i>, <i>mindist.business</i>, <i>mindist.countryside</i>, <i>mindist.publicInstitutions</i>)- Gesamte Grundfläche von Gebäuden im Kartenausschnitt von 500x500m bei denen der Typ nicht spezifiziert ist (<i>totalarea.notspecified</i>) und die als Einfamilienhäuser gekennzeichnet sind (<i>totalarea.singlefamily</i>)- Fläche des Landnutzungsgebiets „Wohngebiet“ (<i>totalarea.residential</i>) und „ländlicher Raum“ (<i>totalarea.countryside</i>) im Kartenausschnitt von 500x500m um den Haushalt



Alter der Geräte (pAppliancesAge)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verbrauch: Durchschnittlicher Verbrauch während der ganzen Woche basierend auf 30-minütigen Daten (<i>c30_week</i>)- Verhältnisse: Verhältnis von minimalem zu maximalem Verbrauch in 15-minütigen Messintervallen (<i>r15_min_mean</i>); Verhältnis des Verbrauchs am Morgen zum Verbrauch am Abend während des Wochenendes basierend auf den 30-minütigen Messintervallen (<i>r30_wd_morning_noon</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none">- Korrelation des Luftdrucks mit dem Stromverbrauch in der Nacht (<i>w_airPr_cor_night</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Landnutzungstyp (<i>thislanduseType</i>) und Gebäudetyp (<i>thisbuildingType</i>) der die Koordinaten des Haushalts einschliesst und der Gebäude- / Landnutzungstyp eines Polygons das der Koordinate am nächsten liegt (<i>nextbuildingType</i>, <i>nextlanduseType</i>)- Minimale Distanz zu geographischen Objekten in der Kategorie „ländlicher Raum“, und „Geschäfte und Büros“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.countryside</i>, <i>mindist.business</i>)
Art der Kochstelle (pCookingType)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Statistische Momente: Anzahl der Verbrauchsspitzen basierend auf 15-minütigen Messintervallen (<i>s15_num_peaks</i>)- Verhältnisse: Verhältnis von minimalem Verbrauch Wochentags und am Wochenende (<i>r15_min_mean</i>), Verhältnis des Verbrauchs abends und mittags (<i>r15_evening_noon_no_min</i>), Tagsüber und nachts (<i>r15_day_night_no_min</i>) je abzüglich der Minima und basierend auf 15-minütigen Daten- Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>) und die Summe der Werte über der Grundlast-Grenze (<i>t15_value_above_base2</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“, „Kultur“, und „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food</i>, <i>mindist.culture</i>, <i>mindist.recreation</i>)- Fläche des Landnutzungsgebiets „Stadtgebiet“ (<i>totalarea.city</i>) im Kartenausschnitt von 500x500m um den Haushalt
Wärmepumpe vorhanden (pHeatPump)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verbrauch: Durchschnittlicher 30-minütiger Verbrauch abends am Wochenende (<i>c30_we_evening</i>)- Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>); Verhältnis von Verbrauch nachts zu tagsüber basierend auf 30-minütigen Daten (<i>r30_night_day</i>)- Zeitliche Kennzahlen: Schätzung der Grundlast auf Grundlage von 15-minütigen Verbrauchsmessungen (<i>t15_value_min_guess</i>), Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“, „Verkehrssystem“, „Geschäfte und Büros“ und „Erholung und Freizeit“ in-



	<p>nerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.culture, mindist.roadSystem, mindist.business, mindist.recreation</i>)</p> <ul style="list-style-type: none">- Fläche des Landnutzungsgebiets „Stadtgebiet“ (<i>totalarea.city</i>) im Kartenausschnitt von 500x500m um den Haushalt
Solaranlage vorhanden (pSolar)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>); Verhältnis des Verbrauchs tagsüber zu nachts (<i>r15_day_night_no_min</i>); Verhältnis vom durchschnittlichen zum maximalen Verbrauch (<i>r15_mean_max_no_min</i>) je abzüglich dem Wochen-Minimum; das Verhältnis des Verbrauchs am Abend und am Mittag basierend auf den 30-minütigen Daten (<i>r30_evening_noon</i>)- Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch über dem Mittelwert (<i>t15_above_mean</i>); Zeitpunkt des ersten Erreichens des Tages-Minimum gemittelt über alle Wochentage (<i>t15_daily_min</i>)- Statistische Momente: Anzahl der Verbrauchsspitzen nach einer Glättung auf Basis von 15-minütigen Daten (<i>s15_num_big_peaks</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“, „Erholung und Freizeit“ und „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.culture, mindist.recreation, mindist.transportation</i>)
Anzahl der umgesetzten Energieeffizienzmassnahmen (pEnergyEfficiencyMeasure)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Zeitliche Kennzahlen: Summe der 15-minütigen Messwerte über der Grundlast-Grenze (<i>t15_value_above_base2</i>); Anzahl der Zeiträume über dem Mittelwert (<i>t15_above_mean</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none">- Korrelation des Niederschlags mit dem Stromverbrauch in der Nacht (<i>w_prec_cor_night</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentliche Einrichtungen“, „ländlicher Raum“, „Geschäfte und Büros“, „Gaststätten und Cafés“, „Sehenswürdigkeiten“ und „Kultur“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.publicInstitutions, mindist.countryside, mindist.business, mindist.food, mindist.sights, mindist.culture</i>)- Landnutzungstyp (<i>thislanduseType</i>) der die Koordinaten des Haushalts einschliesst



Anzahl der Geräte (pAppliancesNum)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>); Summe der Werte über dem Mittelwert (<i>t15_value_above_mean</i>); Anteil der 15-minütigen Messzeiträume mit einem Verbrauch oberhalb der Grundlast (<i>t15_percent_above_base</i>)- Verhältnisse: Verhältnis von Durchschnitt zum maximalen Verbrauch (<i>r15_mean_max_no_min</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none">- Korrelation von Luftdrucks und Windgeschwindigkeit mit dem Stromverbrauch abends (<i>w_airPr_cor_evening</i>, <i>w_windSp_cor_evening</i>), Korrelation des Luftdrucks mit dem Stromverbrauch in der Nacht (<i>w_airPr_cor_night</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „ländlicher Raum“, „Öffentliche Einrichtungen“ und „Gaststätten und Cafés“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.countryside</i>, <i>mindist.publicInstitutions</i>, <i>mindist.food</i>)
Kaufinteresse für Solaranlagen (plnterestSolar)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verhältnisse: Verhältnis des Verbrauchs tagsüber zu nachts (<i>r15_day_night_no_min</i>) je abzüglich der Minima und basierend auf 15-minütigen Daten; Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>); Verhältnis von Durchschnitt zum maximalen Verbrauch (<i>r15_mean_max_no_min</i>); Verhältnis des Verbrauchs am Morgen zum Verbrauch am Abend während des Wochenendes basierend auf den 30-minütigen Daten (<i>r30_wd_morning_noon</i>)- Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>); Summe der Messwerte über der Grundlast-Grenze (<i>t15_value_above_base2</i>)- Statistische Momente: Anzahl der Verbrauchsspitzen basierend auf 15-minütigen Daten (<i>s15_num_peaks</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“, „Gaststätten und Cafés“ und „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.culture</i>, <i>mindist.food</i>, <i>mindist.recreation</i>)
Zufriedenheit mit dem Energieversorger (pSatisfactionUtility)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Zeitliche Kennzahlen: Mittlere Autokorrelation der 15-minütigen Messungen über eine Zeitspanne von 3h über alle tage (<i>ts15_acf_mean3h</i>) und nur an Wochentagen (<i>ts15_acf_mean3h_weekday</i>)- Statistische Momente: Durchschnittliche Korrelation der Verbrauchskurven zwischen allen Tagen in einer Woche (<i>s15_cor</i>), zwischen den Tagen Mo.-Fr. (<i>s15_cor_wd</i>), Geschätztes Zeitintervall mit Grundlast-Verbrauch bei 15-minütigen Daten (<i>t15_const_time</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none">- Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“, „Kultur“, „Geschäfte und Büros“ und „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food</i>, <i>mindist.culture</i>, <i>mindist.business</i>, <i>mindist.transportation</i>)
Kinder im Haushalt (pChildren)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none">- Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen



	<p>Daten (<i>r15_min_wd_we</i>); Verhältnis von Durchschnitt zum maximalen Verbrauch (<i>r15_mean_max_no_min</i>),</p> <ul style="list-style-type: none"> - Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervallen mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>); Summe der Messwerte über der Grundlast-Grenze (<i>t15_value_above_base2</i>), Mittlere Autokorrelation der 15-minütigen Messungen über eine Zeitspanne von 3 Stunden (<i>ts15_acf_mean3h</i>), - Statistische Momente: Anzahl der Verbrauchsspitzen nach einer Glättung aus Basis von 15-minütigen Daten (<i>s15_num_big_peaks</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none"> - Korrelation der Temperatur mit dem Stromverbrauch tagsüber (<i>w_temp_cor_daytime</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none"> - Minimale Distanz zu geographischen Objekten in der Kategorie „Geschäfte und Büros“ und „Kultur“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food</i>, <i>mindist.culture</i>)
<p>Single-Haushalt (pSingle)</p>	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none"> - Verhältnisse: Verhältnis von Verbrauch am Abend zu Verbrauch mittags am Wochentag auf Basis von 30-minütigen Daten (<i>r30_wd_evening_noon</i>), - Verbrauch: Durchschnittlicher Verbrauch am Abend während der ganzen Woche basierend auf 30-minütigen Daten, mit und ohne Subtraktion des Minimums (<i>c30_noon</i>, <i>c15_noon_no_min</i>) - Statistische Momente: Anzahl der Verbrauchsspitzen nach einer groben Glättung auf Basis von 15-minütigen Daten (<i>s15_num_big_peaks</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none"> - Korrelation des Niederschlags mit dem Stromverbrauch gemittelt über alle Tage (<i>w_prec_cor_daily</i>) und zur Tageszeit (<i>w_prec_cor_daytime</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none"> - Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“, „Kultur“ und „Öffentliche Einrichtungen“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food</i>, <i>mindist.culture</i>, <i>mindist.publicInstitutions</i>) - Fläche des Landnutzungsgebiets „Wohngebiet“ (<i>totalarea.residential</i>) im Kartenausschnitt von 500x500m um den Haushalt
<p>Zwei Erwachsene mit Kind/Kindern (pFamily)</p>	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none"> - Verhältnisse: Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten (<i>r15_min_wd_we</i>), Verhältnis von Durchschnitt zum maximalen Verbrauch ohne Berücksichtigung des Wochen-Minimums (<i>r15_mean_max_no_min</i>) - Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>) und über der Grundlast (<i>t15_above_base</i>); Durchschnittlicher Zeitpunkt des ersten Erreichens des täglichen Minimums (<i>t15_daily_min</i>); Mittlere Autokorrelation der 15-minütigen Messungen über eine Zeitspanne von 3 Stunden (<i>ts15_acf_mean3h</i>); Anteil der 15-minütigen Messzeiträume mit einem Verbrauch oberhalb der Grundlast (<i>t15_percent_above_base</i>), Summe der Messwerte über der Grundlast-Grenze (<i>t15_value_above_base2</i>) <p>Features aus Wetter- und Klimadaten:</p> <ul style="list-style-type: none"> - Korrelation der Temperatur mit dem Stromverbrauch tagsüber (<i>w_temp_cor_daytime</i>)



	<p>Features aus geografischen Daten:</p> <ul style="list-style-type: none"> - Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.food</i>)
Haushaltstyp 2 Klassen (pHousehold-Type2)	<p>Features aus Smart-Meter-Daten:</p> <ul style="list-style-type: none"> - Verbrauch: Durchschnittlicher Verbrauchs nachts in 30-minütigen Daten (<i>c30_night</i>), - Zeitliche Kennzahlen: Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh (<i>t15_above_2kwh</i>); Durchschnittliches Erreichen des täglichen Maximums (<i>t15_daily_max</i>) <p>Features aus geografischen Daten:</p> <ul style="list-style-type: none"> - Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentliche Einrichtungen“, „Erholung und Freizeit“, „Gaststätten und Cafés“, „ländlicher Raum“ und „Verkehrssystem“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>mindist.publicInstitutions</i>, <i>mindist.recreation</i>, <i>mindist.food</i>, <i>mindist.countryside</i>, <i>mindist.roadSystem</i>) - Durchschnittliche Distanz zu geographischen Objekten in der Kategorien „Erholung und Freizeit“ und „ländlicher Raum“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt (<i>meandist.recreation</i>, <i>meandist.countryside</i>)

Die folgenden Features in Tabelle 9 wurden über alle Properties hinweg am häufigsten gewählt. Die Auflistung zeigt die Features in absteigender Sortierung nach der Häufigkeit.

Tabelle 9: Über alle Properties hinweg am häufigsten gewählte Features

Rang	Feature	Kategorie	Beschreibung
1	<i>mindist.food</i>	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
2	<i>mindist.recreation</i>	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
3	<i>mindist.publicInstitutions</i>	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentliche Einrichtungen“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
4	<i>t15_above_2kwh</i>	Verbrauch/ Zeitliche Kennzahlen	Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch von mehr als 2 kWh
5	<i>mindist.culture</i>	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
6	<i>mindist.business</i>	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Geschäfte und Büros“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
7	<i>mindist.countryside</i>	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „ländlicher Raum“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt



8	meandist.recreation	Geographisch/ Punkt-basiert	Durchschnittliche Distanz zu allen geographischen Objekten in der Kategorie „Erholung und Freizeit“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
9	meandist.food	Geographisch/ Punkt-basiert	Durchschnittliche Distanz zu allen geographischen Objekten in der Kategorie „Gaststätten und Cafés“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
10	t15_value_min_guess	Verbrauch/ Zeitliche Kennzahlen	Schätzung der Grundlast auf Grundlage von 15-minütigen Verbrauchsmessungen
11	r30_wd_evening_noon	Verbrauch/ Verhältnisse	Verhältnis von Verbrauch am Abend zu Verbrauch mittags am Wochentag auf Basis von 30-minütigen Daten
12	r15_mean_max_no_min	Verbrauch/ Verhältnisse	Verhältnis von Durchschnitt zum maximalen Verbrauch ohne Berücksichtigung des Wochen-Minimums
13	meandist.countryside	Geographisch/ Punkt-basiert	Durchschnittliche Distanz zu allen geographischen Objekten in der Kategorie „ländlicher Raum“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
14	t15_time_above_base2	Verbrauch/ Zeitliche Kennzahlen	Anzahl der 15-minütigen Zeitintervalle mit einem Verbrauch oberhalb des Grundlast-Verbrauchs
15	r30_evening_noon	Verbrauch/ Verhältnisse	Verhältnis von Verbrauch am Abend zu Verbrauch mittags auf Basis von 30-minütigen Daten
16	mindist.transportation	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
17	t15_const_time	Verbrauch/ Zeitliche Kennzahlen	Geschätztes Zeitintervall mit Grundlast-Verbrauch bei 15-min Daten
18	t15_daily_min	Verbrauch/ Zeitliche Kennzahlen	Zeitpunkt des ersten Erreichens des Tages-Minimum, gemittelt über alle Wochentage
19	r15_min_wd_we	Verbrauch/ Zeitliche Kennzahlen	Verhältnis des minimalen Verbrauchs an Wochentagen zu dem minimalen Verbrauch am Wochenende auf Basis von 15-minütigen Daten
20	totalarea.apartments	Geographisch/ Polygon-basiert	Gesamte Grundfläche von Gebäuden im Kartenausschnitt von 500x500m die als Mehrfamilienhäuser gekennzeichnet sind
21	r15_var_wd_we	Verbrauch/ Verhältnisse	Verhältnis der Varianz Wochentags – Wochenendtags, berechnet auf Basis der 15-minütigen Daten
22	w_prec_cor_night	Wetter	Korrelation des Niederschlags mit dem Stromverbrauch in der Nacht
23	meandist.publicInstitutions	Geographisch/ Punkt-basiert	Durchschnittliche Distanz zu allen geographischen Objekten in der Kategorie „Öffentliche Einrichtungen“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt
24	t15_percent_above_base	Verbrauch/ Zeitliche Kennzahlen	Anteil der 15-minütigen Messzeiträume mit einem Verbrauch oberhalb der Grundlast
25	w_airPr_cor_night	Wetter	Korrelation des Luftdrucks mit dem Stromverbrauch in der Nacht
26	mindist.roadSystem	Geographisch/ Punkt-basiert	Minimale Distanz zu geographischen Objekten in der Kategorie „Verkehrssystem“ innerhalb eines Kartenausschnitts von 500x500m um den Haushalt



4.2.5 Performanz von Feature-Selektion-Algorithmen

Ein Projektziel der ersten Projektphase war es, die Effizienz und Skalierbarkeit der Algorithmen für die praktische Anwendung zu steigern, damit gängige Hardware (z.B. mit einem Server mit Anschaffungskosten von 15'000 CHF) in der Lage ist, 10'000 Haushalte in der Stunde zu klassifizieren.

Als grösstes Potential für Effizienz-Verbesserungen haben sich die FS-Algorithmen herausgestellt. Die Laufzeit dieser haben wir bei 2'911 verschiedenen Konfigurationen unseres Tools (verschiedene Klassifizierer, Feature-Mengen, Parameter, usw.) gemessen und mit einer einfachen linearen Regression ausgewertet. In Tabelle 10 kann der Einfluss von einzelnen FS-Methoden auf die Laufzeit abgelesen werden. Die erklärte Varianz des Modells beträgt 56% mit einem Standardfehler von 84,29 bei 7702 Freiheitsgraden. Man erkennt, dass die Wrapper-Methoden (Forward- und Backward-Selektion) sowie die Konsistenz-Methode sehr lange brauchen, wohingegen die übrigen Filter-Methoden sehr schnell sind. Für eine skalierbare Implementierung sind demnach Filter-Methoden zu bevorzugen.

Tabelle 10: Geschätzte Koeffizienten für die Erklärung der Feature-Selection-Laufzeit mit einem linearen Regressionsmodell

	Geschätzter Parameter	Std. Fehler	Signifikanz (. für <0.01, *** für <0.0001)
(Intercept)	0,00	2,09	
cfs	50,15	3,39	***
consistency	177,40	3,97	***
chi.squared	1,01	3,13	
gain.ratio	1,37	3,72	
random.forest.importance	7,66	3,98	.
symmetrical.uncertainty	1,56	3,98	
backward.search	445,20	5,36	***
forward.search	130,80	5,36	***
KS-test	0,39	5,37	
eta-squared	1,77	7,31	
combined	0,52	7,31	

4.3 Entwicklung der Haushaltsklassifikations-Algorithmen

In der ersten Projektphase wurde ein Klassifikationstool für 12 Haushaltseigenschaften auf Basis von 30-minütigen SMD aus Irland weiterentwickelt. Details einer frühen Version dieses Systems wurden in den Publikationen von Beckel et al. (2013 & 2014) beschrieben. In der zweiten Phase haben wir dieses Klassifikations-Tool auf 15-minütige Smart-Meter-Daten angepasst und um die mehrdimensionale Klassifikation mit externen Daten erweitert. Es wurden zudem 18 neue Haushaltseigenschaften definiert und vier existierende -Definitionen der Haushaltseigenschaften aus Projektphase 1 auf die Schweiz übertragen.



4.3.1 Anwendung von Klassifikationsalgorithmen

In der ersten Projektphase wurden massgeblich Klassifikationsalgorithmen verwendet, welche kontinuierliche bzw. numerische Daten verarbeiten. Es wurden dort k-Nearest-Neighbor (kNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Mahalanibos Distances, Neuronale Netze, AdaBoost und Random Forest verwendet.

Um den vollen Informationsgehalt von Geodaten abzubilden wurde es notwendig, auch kategoriale Features zu definieren (z.B. Haustyp oder Landnutzungstyp). Viele Klassifikatoren können jedoch nicht gut mit solchen Kategorialen Daten umgehen. Aus diesem Grund wurde Naïve Bayes als weiterer Klassifikator verwendet und lediglich drei der Klassifikatoren aus dem Projektphase 1 mit den besten Klassifikationsergebnissen verwendet (kNN, SVM, Random Forest).

An zahlreichen Stellen haben wir die Parameter der Klassifizierer optimiert. Dabei verfügte der SVM-Klassifizierer über das grösste Verbesserungspotential. Wir haben deshalb eine Parameteroptimierung durchgeführt bei der 230 verschiedene Parameterkonstellationen verwendet wurden. Abbildung 7 zeigt die Bandbreite (Minimumale, maximale und mittlere Genauigkeit) über alle Konfigurationen.

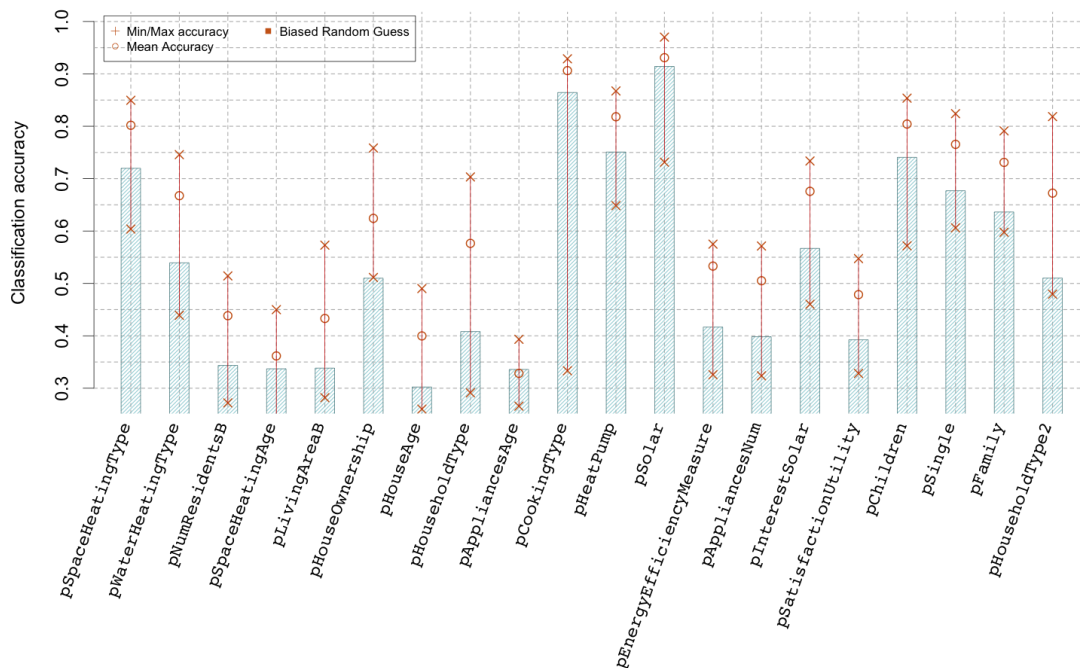


Abbildung 7: Klassifikationsergebnisse einer Woche mit verschiedenen SVM Parameter-Konstellationen

4.3.2 Umgang mit unbalancierten Klassen

Kleine Klassen (z.B. Single 20% vs. Nicht Single 80%) werden mit herkömmlichen Algorithmen schlecht erkannt, wenn die Genauigkeit als Auswertungsmetrik herangezogen wird. Als Lösung wurde SMOTE (Synthetic Minority Oversampling Technique) implementiert. Hierbei werden zum Training zusätzliche Repräsentanten der kleineren Klasse generiert und als (zufällig)-gewichtetes Mittel von mehreren ähnlichen tatsächlichen Repräsentanten generiert. Wir definieren die Klassen als unbalanciert, falls die dreifache Grösse der kleinsten Klasse einer Property kleiner ist als die Grösse deren grössten Klasse. So lässt sich die Genauigkeit als Precision für die Properties „Family“, „#bedrooms“,



„Children“, „Single“ und „Floor_area“ um durchschnittlich 27% verbessern. Die Verbesserungen durch SMOTE sind in Abbildung 8 dargestellt.

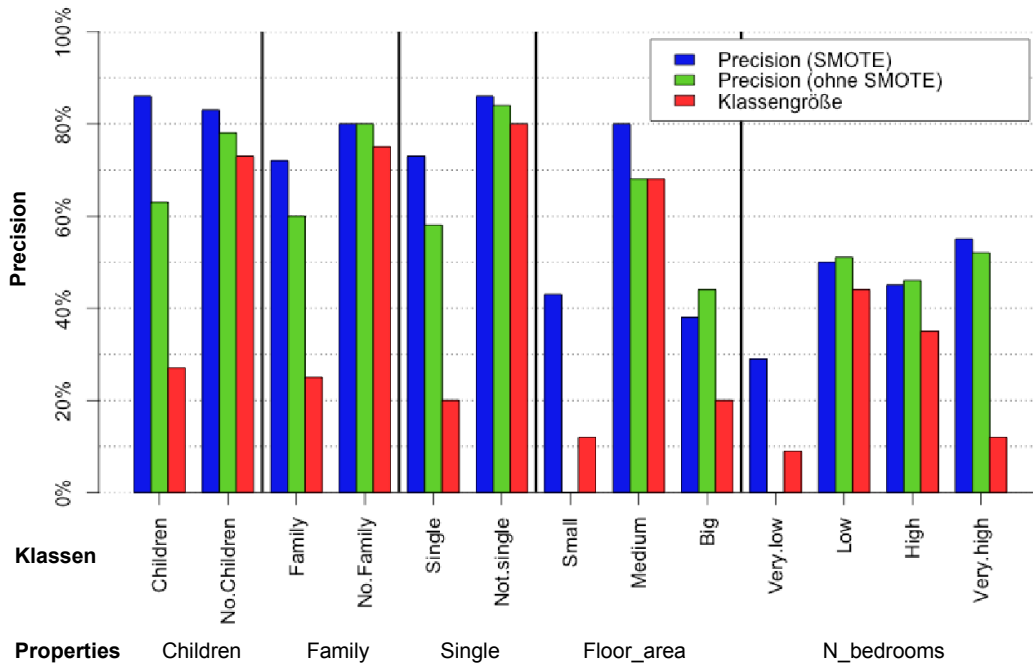


Abbildung 8: Verbesserung der Klassifikationsgenauigkeit mit und ohne SMOTE beim irischen Datensatz

4.3.3 Steigerung der Anzahl der geschätzten Charakteristika

Wir haben neue Properties und Klassen entsprechend den Fragen und Antwortmöglichkeiten aus der Kundenumfrage verwendet. Die Property-Definitionen können der Tabelle 11 entnommen werden.

Tabelle 11: Übersicht über alle Haushaltseigenschaften mit ihrer Definition und der Referenz auf die Frage-Nummer (FN) der Kundenumfrage (s. Tabelle 4 in Abschnitt 4.1.1)

Haushaltscharakteristika	Klasse	Definition	Klassengröße	
			Abs.	Rel.
Hohes Kaufinteresse an Fiber-to-the-Home (ftth.pi.classesC)	Hoch	$FN-28 \geq 8$	137	34,51%
	Nicht hoch	$FN-28 < 8$	260	65,49%
Kein Wissen über Fiber-to-the-Home (ftth.unknown)	Nein	$FN-28 \neq \text{„Ich weiss nicht“}$	488	92,60%
	Ja	$FN-28 = \text{„Ich weiss nicht“}$	39	7,40%
Alter der Geräte (pAppliancesAge)	Neu	Mittl. Gerätealter (FN-17a) $< q_{0,25}$	153	33,41%
	Mittel	Mittl. Gerätealter zwischen $q_{0,25}$ und $q_{0,75}$	152	33,19%
	Alt	Mittl. Gerätealter $> q_{0,75}$	153	33,41%



Anzahl der Geräte (<i>pAppliancesNum</i>)	Wenige	Geräteanzahl $\emptyset < q_{0,25}$ (FN-17a)	149	28,27%
	Mittel	Geräteanzahl \emptyset zwischen $q_{0,25}$ und $q_{0,75}$	280	53,13%
	Viele	Geräteanzahl $\emptyset > q_{0,75}$	98	18,60%
Kinder im Haushalt (<i>pChildren</i>)	Kinder	Anz. Kinder (FN-9) > 0	69	13,09%
	Keine Kinder	Anz. Kinder = 0	458	86,91%
Art der Kochstelle (<i>pCookingType</i>)	Elektrisch	Anzahl el. Herd (FN-17a) > 0	484	91,84%
	Nicht elektrisch	Anzahl el. Herd = 0	43	8,16%
Anzahl der umgesetzten Energieeffizienzmassnahmen (<i>pEnergyEfficiencyMeasure</i>)	Keine	Anzahl umgesetzte Energiesparmassnahmen FN-24	304	57,69%
	Eine		109	20,68%
	Mehrere		114	21,63%
Zwei Erwachsene mit Kind/Kindern (<i>pFamily</i>)	Familie	Anz. Erwachsene (FN-8) > 1 und Anz. Kinder (FN-9) > 0	111	21,06%
	Keine Familie	sonst	416	78,94%
Wärmepumpe vorhanden (<i>pHeatPump</i>)	Nein	Vorhandene Wärmepumpe (FN-18a)	453	85,96%
	Ja		74	14,04%
Alter des Hauses (<i>pHouseAge</i>)	10-29	Alter des Hauses (FN-14, FN-14a, FN-14b)	147	28,27%
	30-74		219	42,12%
	<10		71	13,65%
	>=75		83	15,96%
Haushaltstyp 4 Klassen (<i>pHouseholdType</i>)	Apartment	FN-11 = „Wohnung im Mehrfamilienhaus“	323	61,88%
	Einfamilienhaus	FN-11 = „Einfamilienhaus (freistehend)“	130	24,90%
	Doppelhaushälfte	FN-11 = „Doppelhaushälfte“	46	8,81%
	Reihenhaus	FN-11 = „Reihenhaus“	23	4,41%
Haushaltstyp 2 Klassen (<i>pHouseholdType2</i>)	Apartment	FN-11 = „Wohnung im Mehrfamilienhaus“	323	61,88%
	Haus	FN-11 \neq „Wohnung im Mehrfamilienhaus“	199	38,12%
Eigentumsverhältnisse (<i>pHouseOwnership</i>)	Eigentum	Besitzverhältnisse (FN-13)	280	53,95%
	Miete		239	46,05%
Kaufinteresse für Solaranlagen (<i>pInterestSolar</i>)	Gering	Summe aus FN-21 < $q_{0,50}$	387	73,43%
	Mittel	Summe aus FN-21 zwischen $q_{0,50}$ und $q_{0,75}$	49	9,30%
	Hoch	Summe aus FN-21 > $q_{0,75}$	91	17,27%
Grösse (Fläche) der Wohnung (<i>pLivingAreaB</i>)	≤ 95	Wohnungsgrösse (FN-12) $\leq 95 \text{ m}^2$	166	32,49%
	≤ 145	Wohnungsgrösse zwischen 95 und 145 m^2	180	35,23%
	> 145	Wohnungsgrösse (FN-12) > 145 m^2	165	32,29%
Anzahl der Bewohner im Haushalt (<i>pNumResidentsB</i>)	1 Person	Anzahl der Bewohner im Haushalt (FN-8 + FN-9)	121	23,05%
	2 Personen		238	45,33%
	3-5 Personen		150	28,57%
	>5 Personen		16	3,05%
Zufriedenheit mit dem Energieversorger (<i>pSatisfaction-Utility</i>)	Gering	Summe aus FN-27 < $q_{0,25}$	127	25,45%
	Mittel	Summe aus FN-27 zwischen $q_{0,25}$ und $q_{0,75}$	261	52,30%
	Hoch	Summe aus FN-27 > $q_{0,75}$	111	22,24%
Single-Haushalt (<i>pSingle</i>)	Single	Anz. Erwachsene (FN-8) = 1 und Anz. Kinder (FN-9) = 0	121	23,05%
	Nicht Single	sonst	404	76,95%



Solaranlage vorhanden (p_{Solar})	Ja	FN-18a ist „Photovoltaik Anlage“ oder „Solarthermie-Anlage“	29	5,50%
	Nein	FN-18a ist weder „Photovoltaik Anlage“ noch „Solarthermie-Anlage“	498	94,50%
Alter der Raumheizung ($p_{Space-HeatingAge}$)	Neu	Alter der Raumheizung (FN-16) $< q_{1/3}$	128	33,51%
	Mittel	Heizungsalter zwischen $q_{1/3}$ und $q_{2/3}$	135	35,34%
	Alt	Heizungsalter $> q_{2/3}$	119	31,15%
Art der Raumheizung ($p_{Space-HeatingType}$)	Elektrisch	Typ der Raumheizung (FN-15) = „Elektro-Speicherheizung“	21	3,98%
	Wärmepumpe	Raumheizung = „Wärmepumpe“	66	12,52%
	Andere	Raumheizung \notin {„Elektro-Speicherheizung“; „Wärmepumpe“}	440	83,49%
Art der Warmwassersbereitung ($p_{WaterHeatingType}$)	Elektrisch	Typ der Warmwasseraufbereitung (FN-15) = „Elektro-Speicherheizung“	81	15,37%
	Wärmepumpe	Warmwasseraufbereitung = „Wärmepumpe“	63	11,95%
	Andere	Warmwasseraufbereitung \notin {„Elektro-Speicherheizung“; „Wärmepumpe“}	383	72,68%

4.3.4 Anwendung und Weiterentwicklung von Supervised-Learning-Verfahren für mehrdimensionale Daten

Die Berechnungen der Korrelationsanalyse, wie in Abschnitt 4.2.2 beschrieben haben wir verallgemeinert und in unserem Tool als generalisierte Algorithmen implementiert. Die dabei entwickelte Methode wurde veröffentlicht (Sodenkamp et al., 2015).

4.3.5 Repetitive Anwendung der Verfahren, Nutzung von Majority-Voting zur Klassenbestimmung

Um die Klassifikationsergebnisse zu verbessern und das gesamte Datenvolumen auszunutzen haben wir die Klassifikation auf mehrere Wochen erweitert. Dabei wurde jeweils die beste Konfiguration der Klassifikationsalgorithmen für eine Woche gewählt und auf alle anderen Wochen angewandt. Bei dieser Klassifikation wurde neben der Majority-Voting-Methode, die probabilistische Version der Klassifizierer verwendet, welche für jede Klasse eine Zugehörigkeitswahrscheinlichkeit liefern. Die sich ergebenden Wahrscheinlichkeiten für jede Woche werden dann mit Hilfe vom Mittelwertfunktion zu einem Wert aggregiert. Der Haushalt wird dann der entsprechend der wahrscheinlichsten Klasse klassifiziert. Durch diese Verfahren konnte die Klassifikations-Genauigkeit gesteigert werden.



4.3.6 Betrachtung der Stabilität der Ergebnisse bei Anwendung eines Standard-Klassifizierers

Wir haben die Stabilität der Klassifikation in zweierlei Hinsicht untersucht:

1. Verlässlichkeit der Vorhersagen bei mehreren Durchläufen – Wie gross sind die Abweichungen in der Klassifikations-Genauigkeit bei mehrfach wiederholter Durchführung von Trainings- und Testschritten?
2. Saisonale Stabilität der Klassifikation – Wie verändert sich die Klassifikations-Genauigkeit bei der Verwendung von unterschiedlichen Wochen im Jahr?

Die Ergebnisse dieser Aufwertungen beschreiben wir im Abschnitt 5. Ergebnisse.

4.3.7 Test der Übertragbarkeit der Klassifizierer auf verschiedene geographische Regionen

Mit dieser Analyse haben wir getestet, ob ein Haushaltsklassifikationsalgorithmus der auf einem Kundendatensatz trainiert wurde, auch gute Klassifikationsergebnisse für einen anderen Datensatz (z.B. mit Kunden aus einem anderen Land) erzielt. Für diesen Übertragbarkeitstest verwendeten wir neben dem Smart-Meter- und Kundendatensatz der Arbon Energie AG (basierend auf 15-min Smart-Meter-Daten) auch den Datensatz irischer Verbraucher (30-min Smart-Meter-Daten) aus Projektphase 1. Die Datengranularität beider Datensätze haben wir dabei auf 30-min vereinheitlicht.

Die Übertragbarkeit wurde auf vier Haushaltseigenschaften getestet, die in beiden Datensätzen erhoben wurden: Kinder im Haushalt (pChildren), Art der Kochstelle (pCookingType), Zwei Erwachsene mit Kind/Kindern (pFamily), Single-Haushalt (pSingle). Es wurde ausserdem eine für beide Datensätze repräsentative Woche im Juni verwendet. Ergebnisse dieser Auswertung präsentieren wir im Abschnitt 5. Ergebnisse.

4.4. Tool-Anpassung

Aus den rudimentären Algorithmen haben wir in der Projektphase 1 ein Tool entwickelt, das in der zweiten Projektphase erweitert wurde. Alle Algorithmen wurden in Statistik-Software R entwickelt (GNU General Public License Version 3) und entsprechend einschlägiger Coding-Guidelines kommentiert.

4.4.1 Beschreibung des Software-Tools

In der ersten Projektphase wurde ein Tool mit fünf Modulen entwickelt (1. Laden und Transformieren von Daten, 2. Datenvorbereitung, 3. Klassifikation mit verschiedenen Algorithmen, 4. Evaluation mit verschiedenen Metriken, 5. Export und Visualisierung). Diese wurden in der zweiten Projektphase erweitert und neu strukturiert. Die finale Softwarestruktur ist im Komponentendiagramm in Abbildung 9 dargestellt.

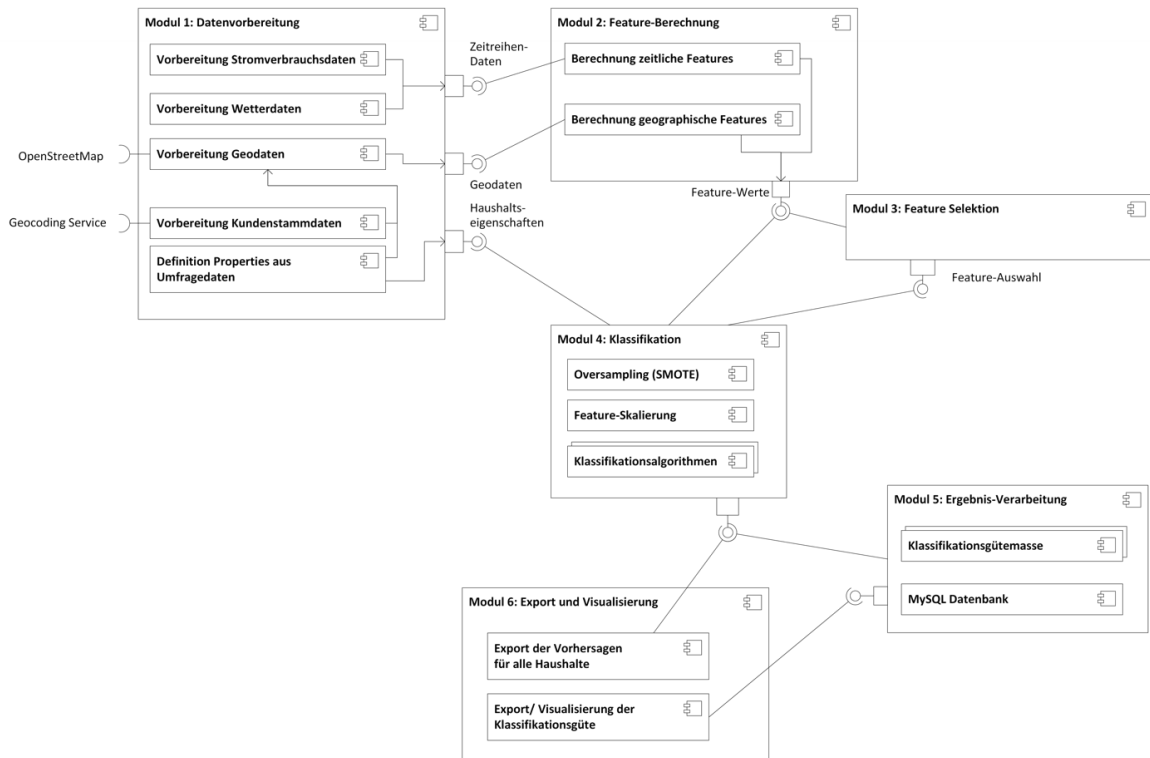


Abbildung 9: UML-Komponentendiagramm des entwickelten Tools

Den Programmablauf für die Haushaltsklassifikation haben wir in Abbildung 10 vereinfacht als UML-Aktivitätsdiagramm illustriert.

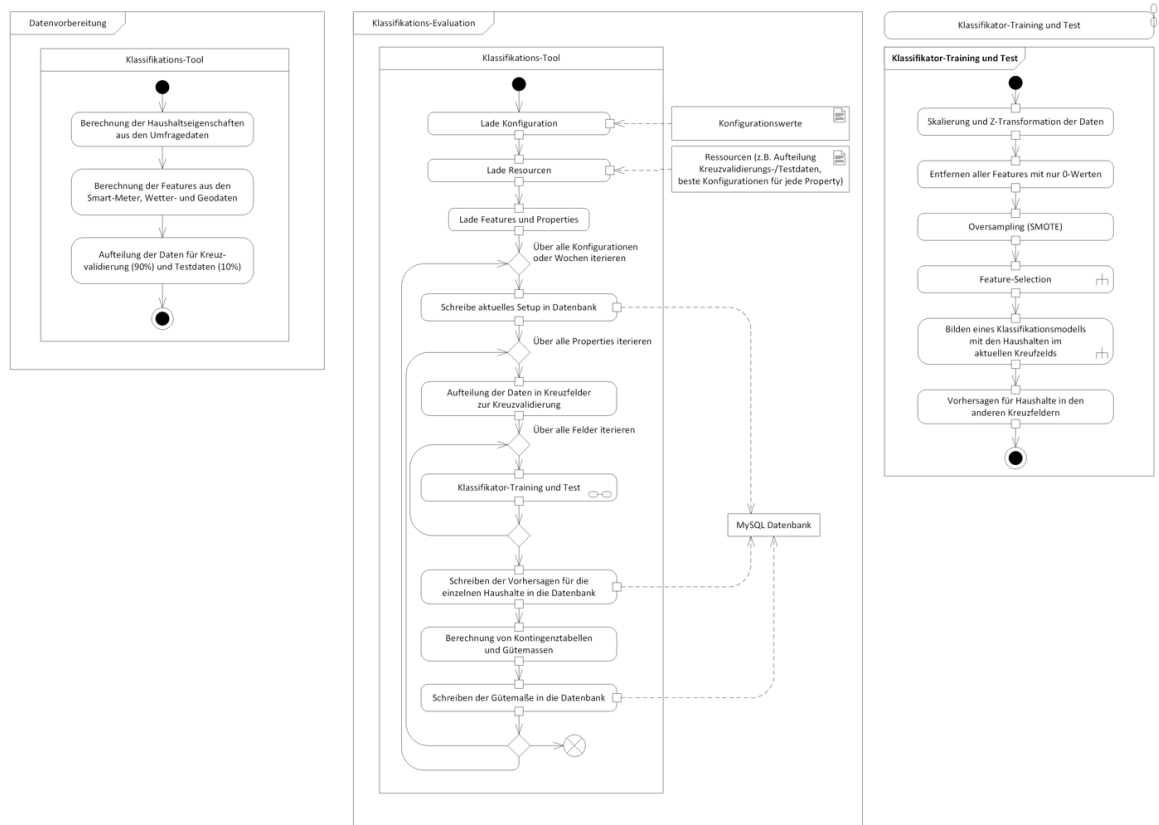


Abbildung 10: UML Aktivitätsdiagramm der Datenvorbereitung und der Klassifikations-Evaluation.

4.4.2 Dokumentation für Endbenutzer

Wir haben unser Tool für einen Anwender dokumentiert, der Grundkenntnisse in der Statistikumgebung GNU-R hat. Dabei erfolgte die Dokumentation des Tools Quellcodenah, d.h. alle Programmschritte sind für den Anwender in derartiger Weise dokumentiert, dass er in der Lage ist, alle Ergebnisse des Projekts zu reproduzieren und die Klassifikation für neue Daten anzuwenden.

Um auch Anwendern ohne R-Kenntnisse die Hauptfunktionen und den Aufbau zu vermitteln haben wir Programm-Modelle (wie in den Abbildungen 9 und 10) erstellt.

4.5 Validierung und Feldtests

Zusammen mit der Arbon Energie AG wurde für das Jahr 2016 ein Feldtest der Haushaltsklassifikationsalgorithmen geplant. Es handelt sich um die Optimierung von Rücklaufquoten bei einer Fiber-to-the-Home Cross-Selling Kampagne. Die Vorhersagealgorithmen wurden bereits entwickelt, der Zeitpunkt dieses Berichts liegt jedoch vor dem Beginn der Feldtests, weshalb an dieser Stelle noch nicht über die Ergebnisse berichtet werden kann.



5. Ergebnisse

5.1 Erweiterung der erkennbaren Haushaltscharakteristika

Im Rahmen der zweiten Projektphase wurde die Anzahl der erkennbaren Haushaltscharakteristika von 12 auf 22 erhöht. 18 neue Haushaltscharakteristika konnten durch eine Kundenumfrage bei der Arbon Energie AG erhoben werden. Ein besonderer Fokus der Umfrage lag dabei auf Properties der Energieeffizienz und den Erneuerbaren Energieträgern. Die erhobenen Daten werden in Tabelle 12 mit den Daten aus Projektphase 1 gegenübergestellt.

Tabelle 12: Vergleich der Datensätze die im Projekt verwendet wurden

Art der Daten		Schweiz	Irland
Smart-Meter-Daten	Granularität	15 Minuten	30 Minuten
	Zeitraum	Jun 2014 – Mai 2015	Jul 2009 – Dez 2010
	# Haushalte	451	4200
Umfrage-Daten	# Haushalts-Charakteristika	22	12
	Lebenssituation	Anzahl der Bewohner im Haushalt (<i>pNumResidents</i>), Grösse (Fläche) der Wohnung (<i>pLivingArea</i>), Haushaltstyp (<i>pHouseholdType</i>), Kinder im Haushalt (<i>pChildren</i>), Single-Haushalt (<i>pSingle</i>), Zwei Erwachsene mit Kind/Kindern (<i>pFamily</i>)	Anzahl der Zimmer (<i>N_bedrooms</i>), Anstellungsverhältnis (<i>Employment</i>), Zwei Erwachsene mit Kind/Kindern (<i>Family</i>), Grösse (Fläche) der Wohnung (<i>floor area</i>), Kinder im Haushalt (<i>Children</i>), Anzahl der Bewohner im Haushalt (<i>N_residents</i>), Single-Haushalt (<i>Single</i>), Ruhestand (<i>Retirement</i>), Soziale Klasse nach der britischen NRS Skala (<i>social_class</i>)
	Energieeffizienz	Art der Raumheizung (<i>pSpaceHeatingType</i>), Art der Warmwasserbereitung (<i>pWaterHeatingType</i>), Alter der Raumheizung (<i>pSpaceHeatingAge</i>), Alter des Hauses (<i>pHouseAge</i>), Alter der Geräte (<i>pApplianceAge</i>), Art der Kochstelle (<i>pCookingType</i>), Anzahl der umgesetzten Energieeffizienzmassnahmen (<i>pEnergyEfficiencyMeasure</i>), Anzahl der Geräte (<i>pAppliancesNum</i>)	Anzahl der Geräte (<i>N_devices</i>), Art der Kochstelle (<i>Cooking</i>), Alter des Hauses (<i>age_house</i>),
	Erneuerbare Energieträger	Wärmepumpe vorhanden (<i>pHeatPump</i>), Solaranlage vorhanden (<i>pSolar</i>), Kaufinteresse für Solaranlagen (<i>pInterestSolar</i>)	
	Weitere	Zufriedenheit mit dem Energieversorger (<i>pSatisfactionUtility</i>), Kein Wissen über Fiber-to-the-Home (<i>ftth.unknown</i>), Hohes Kaufinteresse an Fiber-to-the-Home (<i>ftth.pi.classesC</i>)	
Adresse		Ja	Nein



5.2 Deskriptive Ergebnisse aus dem Vergleich der Datensätze aus den beiden Projektphasen

Wir haben die Kundenumfrage bei der Arbon Energie AG so gestaltet, dass ein Teil der Fragen welche in der irischen Studie gefragt wurden, auch bei den Kunden der Arbon Energie AG erhoben wurden. Die gemeinsamen Variablen haben wir untersucht und präsentieren die Ergebnisse in den Abbildungen 11 bis 19.

Die Haushalte in Arbon weisen im Vergleich zu den Haushalten in Irland keinen Unterschied im Anteil der mehrfach verglasten Fenstern auf (vgl. Abbildung 17), der Mittelwert in der Umfrage bei der Arbon Energie AG ist 86,5%, in der irischen Umfrage beträgt er 88,3%. Die Werte sind nicht signifikant unterschiedlich (t-Test mit $t=-1,3383$, $df=534,05$, $p\text{-Wert}=0,1814$). Darüber hinaus kann nicht darauf geschlossen werden kann, dass in der Schweiz weniger doppelt- oder mehrfachverglaste Fenster verbaut sind. Eine detaillierte Auseinandersetzung mit dem angegebenen Anteil an doppelt- oder mehrfachverglasten Fenstern in beiden Umfragen ist in Anhang 4 zu finden.

Allerdings geben die Haushalte in der Schweiz einen signifikant höheren Anteil an Energiesparlampen im Mittel an (CH: 56,4%, IR: 45,8%, t-Test mit $t=7.1791$, $df=557,07$ und $p\text{-Wert} < 0,0001$ positiv, vgl. Abbildung 16). Dies ist erwartungsgemäss, da „seit 2012 die früher gewohnte Glühbirne verboten“ ist (EnergieSchweiz, 2015).

Die Unterschiede zwischen den Umfragen sind zunächst rein deskriptive Ergebnisse und können aufgrund der fehlenden bzw. nicht überprüften Repräsentanz der Umfrage innerhalb der Arbon Energie AG Kunden nicht auf die schweizerische Gesamtbevölkerung übertragen werden. Daneben können verschiedene Umfragemethoden Ursache für diese Unterschiede sein (Telefonumfrage in Irland, Online-Umfrage in der Schweiz). Auch das konkrete Verständnis der Umfrageteilnehmer für Begrifflichkeiten kann Ursache der Unterschieden sein: es kann zum Beispiel unklar sein, ob LED-Leuchten zur Antwort auf die Frage „Wie hoch schätzen Sie den Anteil (in %) der Energiesparlampen in Ihrem Haushalt“ hineinzählen, oder nur Kompaktleuchtstofflampen. Schliesslich unterscheiden sich auch die konkreten Fragen zur Motivation für Energieeffizienz in den Umfragen: In Arbon haben wir fünf Arten der Motivation gefragt (Monetär, Umwelt, Vorbild für Kinder, Generell sparsamer Lebensstil und ob Energieeffizientes Handeln in Familie oder Freundeskreis üblich ist). In der irischen Umfrage wurden nur die ersten beiden Punkte im Kontext von Sparmöglichkeiten an der Stromrechnung abgefragt.

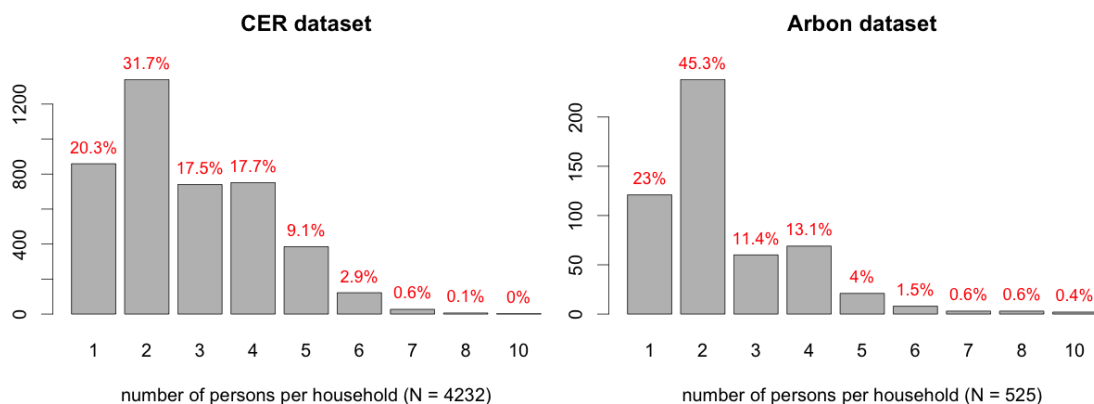


Abbildung 11: Häufigkeit der Antworten zur Frage nach der Anzahl an Personen im Haushalt in der irischen Umfrage (links) und der Arbon Energie AG Kundenumfrage (rechts), Lesebeispiel: In beiden Umfragen ist der Anteil der Einpersonenhaushalte nahezu gleich (20.3% in Irland vs. 23% in Arbon)

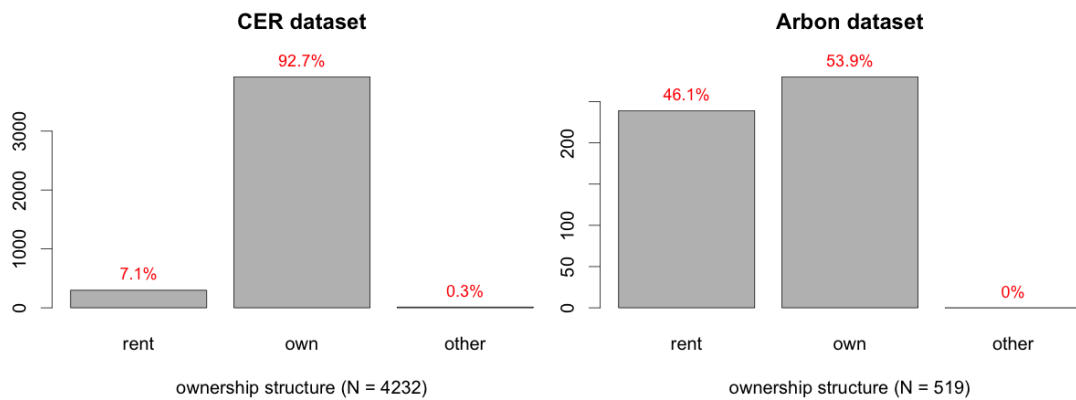


Abbildung 12: Verteilung der Eigentumsverhältnisse an den Wohnungen in beiden Umfragen im Vergleich; in der irischen Umfrage ist der Anteil an Haushalten die zur Miete wohnen viel geringer (7.1% in Irland vs. 46.1% in Arbon)

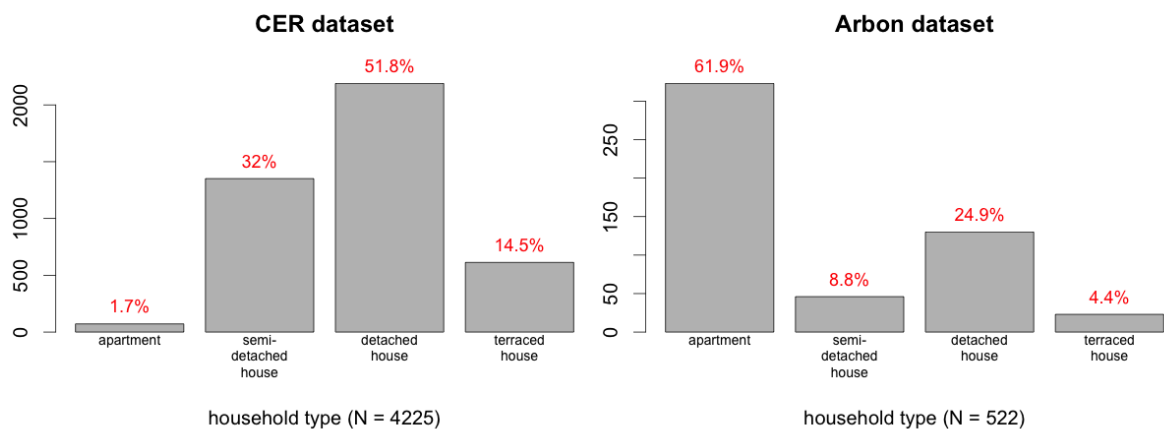


Abbildung 13: Verteilung der Wohnungstypen in beiden Umfragen im Vergleich; der Anteil an Wohnungen in Mehrfamilienhäusern (*apartment*) ist in der Umfrage in Arbon mit 61,9% wesentlich höher als in der irischen Umfrage mit 1,7%, auch die anderen Haustypen Doppelhaushälfte (*semi-detached house*), Reihenhaus (*terraced house*) und freistehendes Einfamilienhaus (*detached house*) unterscheiden sich stark

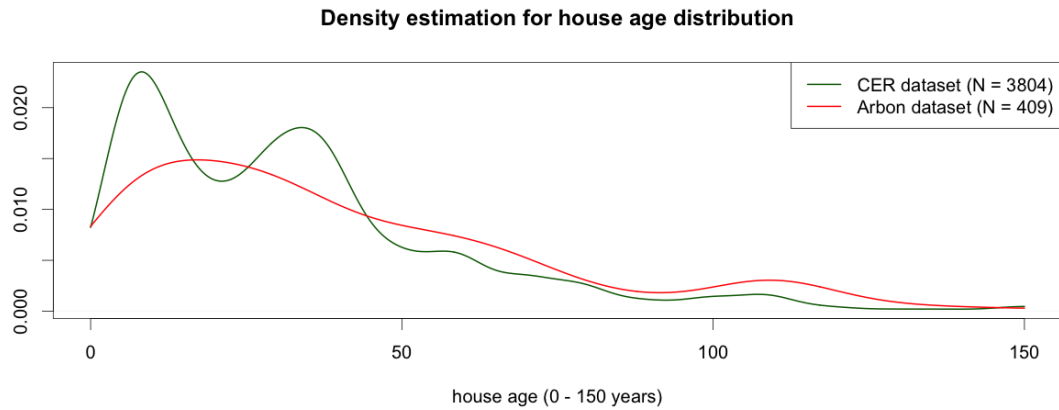


Abbildung 14: Verteilung der Variable Hausalter in beiden Umfragen im Vergleich

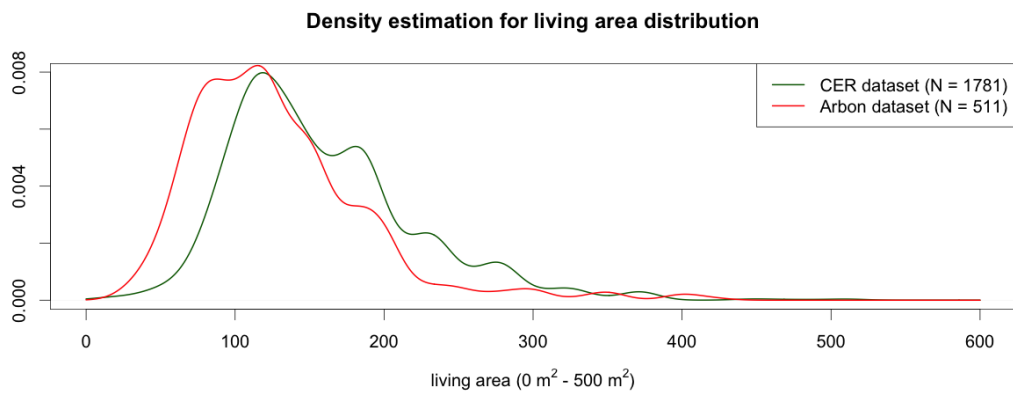


Abbildung 15: Verteilung der Variable Wohnungsgröße in beiden Umfragen im Vergleich

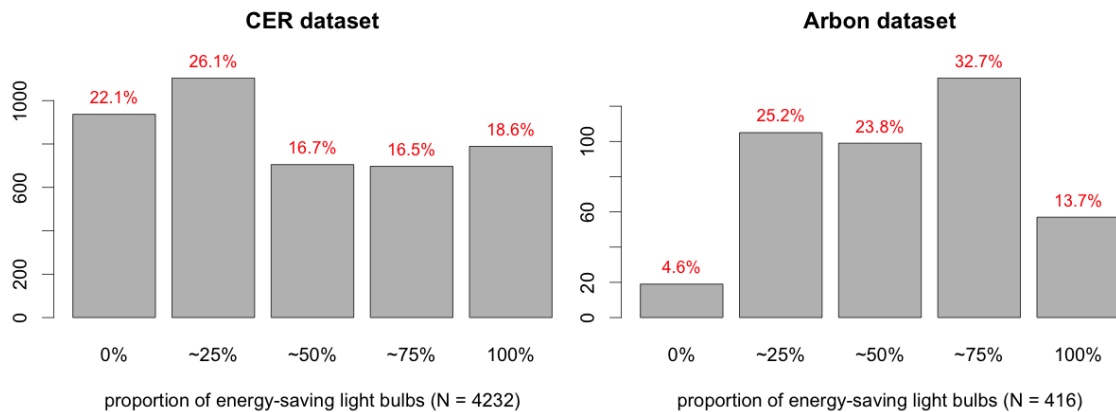


Abbildung 16: Verteilung des Anteils an Energiesparlampen in beiden Umfragen im Vergleich

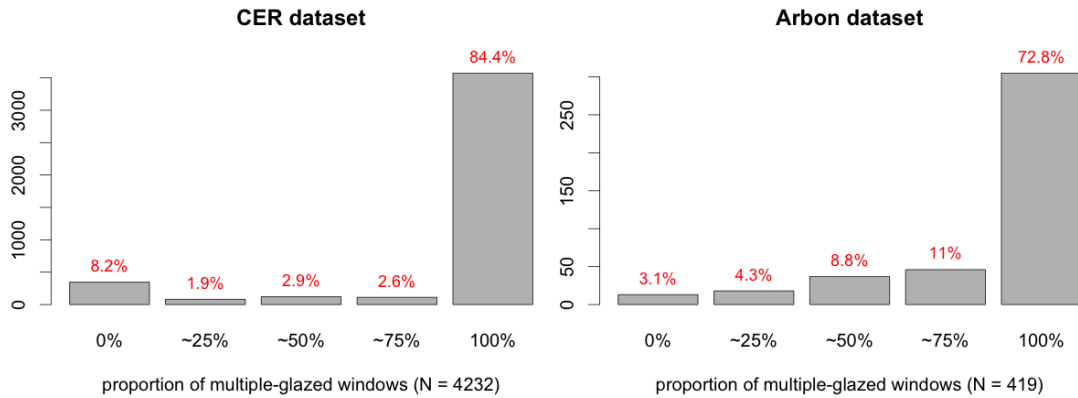


Abbildung 17: Verteilung des Anteils an mehrfachverglasten Fenstern in beiden Umfragen im Vergleich

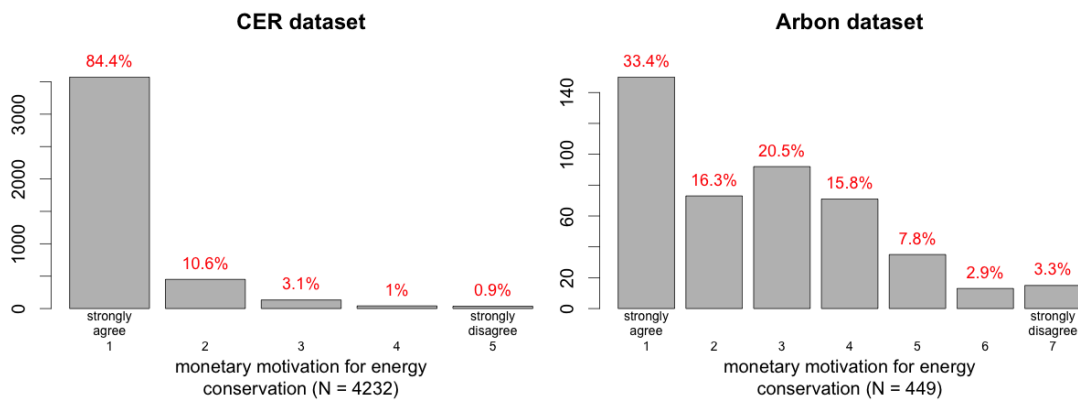


Abbildung 18: Verteilung der Zustimmung zu der Aussage, aus monetären Gründen energieeffizient zu handeln in beiden Umfragen im Vergleich

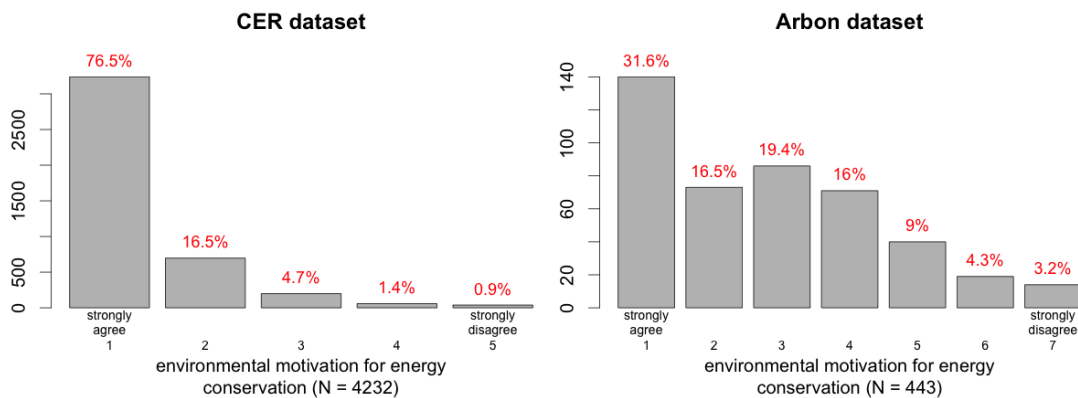


Abbildung 19: Verteilung der Zustimmung zu der Aussage, aus Umweltschutzgründen energieeffizient zu handeln in beiden Umfragen im Vergleich



5.3 Ergebnisse der mehrdimensionalen Klassifikation

Die Klassifikationsergebnisse aller Properties mit allen Klassifizierern sind in Abbildung 20 dargestellt. Als Vergleich für die Klassifikationsgüte betrachten wir die Ergebnisse welche durch den **Biased Random Guess (BRG)** Klassifikator erreicht werden. Dieser Klassifikator ordnet jedem Haushalt zufällig eine Klasse zu, wobei jede Klasse genauso häufig gewählt wird, wie sie in den Daten auftritt. Die Genauigkeit von BRG lässt sich durch $\sum_{k=1}^K h_k^2$ berechnen, wobei h_k die relative Häufigkeit der Klasse k ist.

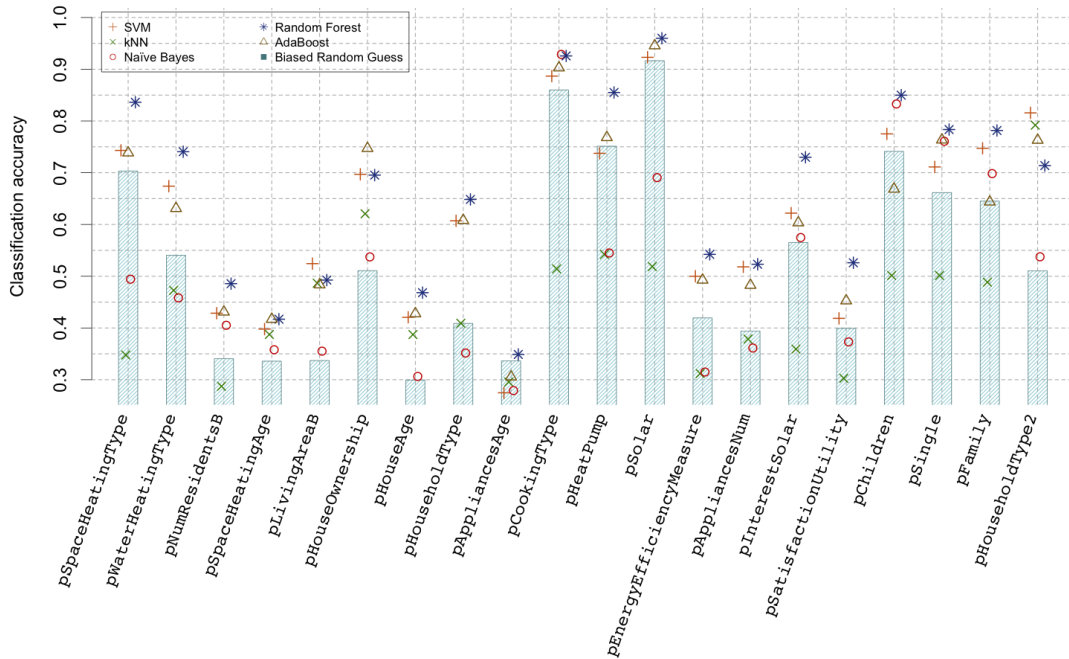


Abbildung 20: Klassifikationsergebnisse einer Woche mit allen Klassifizierern ohne Feature-Selection

5.3.1 Verbesserung der Klassifikation durch externe Daten

In den Abschnitten 4.2.1 und 4.2.2 haben wir beschrieben, wie Wetterinformationen in der Klassifikation durch Korrelationsanalysen berücksichtigt wurden. Durch die Normalisierung des Stromverbrauchs mit der Temperatur konnte die Fehlerrate in der Klassifikation um durchschnittlich 2% verringert werden, die Klassifikationsgenauigkeit konnte somit um 1% verbessert werden (siehe Tabelle 13).



Tabelle 13: Klassifikationsgenauigkeit mit (AccT) und ohne Betrachtung (Acc) der Temperatur am Beispiel des irischen Smart-Meter-Datensatzes, in % Accuracy

Haushaltseigenschaft	Woche		Woche		Woche		Woche	
	46 / 2009		02 / 2010		16 / 2010		37 / 2010	
	Acc	AccT	Acc	AccT	Acc	AccT	Acc	AccT
Single-Haushalt (Single)	83,3	83,6	84,0	84,2	82,4	82,6	83,4	83,6
Anzahl der Geräte (N_devices)	53,1	53,9	53,5	54,4	53,8	54,4	51,2	51,4
Art der Kochstelle (Cooking)	73,0	73,3	72,2	72,4	73,3	73,6	72,2	72,6
Zwei Erwachsene mit Kind/Kindern (Family)	76,4	77,0	78,8	79,1	77,3	78,0	76,0	76,5
Kinder im Haushalt (Children)	74,7	75,2	75,5	75,6	74,1	74,1	73,4	74,1
Alter des Hauses (age_house)	61,1	61,5	61,3	61,2	62,0	62,2	61,4	61,4
Soziale Klasse nach der britischen NRS Skala (social_class)	51,5	51,9	51,1	51,3	49,7	50,4	51,3	51,5
Grösse (Fläche) der Wohnung (floor area)	63,5	65,2	63,1	64,7	63,9	65,7	63,6	65,0
Anzahl der Bewohner im Haushalt (N_residents)	72,8	73,7	71,0	72,2	72,1	72,9	75,2	76,4
Anzahl der Zimmer (N_bedrooms)	49,7	50,0	49,4	49,5	49,5	49,9	48,3	48,5
Anstellungsverhältnis (Employment)	68,4	69,2	67,4	68,4	69,2	70	68,4	69,4
Ruhestand (Retirement)	71,4	72,3	71,5	72,3	72,3	73,1	72,3	73,0

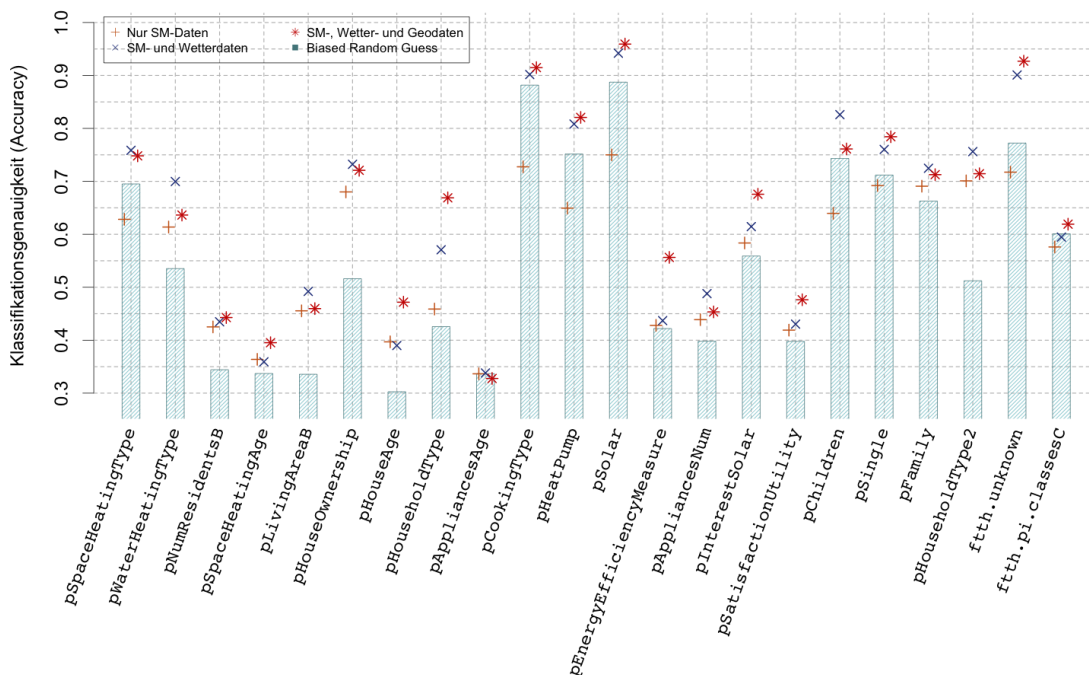


Abbildung 21: Klassifikationsergebnisse (Mittlere Accuracy in Woche 6 mit verschiedenen Feature-Selection-Methoden) bei der Anwendung unterschiedlicher Feature-Mengen (mit / ohne Wetterdaten, mit / ohne Geodaten)

Darüber hinaus haben wir Wetterdaten und Geodaten in Form von Features in die Klassifikation aufgenommen. Abbildung 21 zeigt die Klassifikationsergebnisse mit verschiedenen Feature-Mengen (Stromverbrauchsdaten mit / ohne Wetterdaten und mit / ohne Geodaten).



Die Hinzunahme von Wetter-Features konnte die Klassifikation bei 19 von 22 Properties verbessern. Die zusätzliche Erweiterung der Klassifikation mit geographischen Features verbesserte die Klassifikation bei 13 von 22 Fällen. Eine genaue Auflistung der Features, welche für jede Property häufig gewählt wurden, findet sich in Tabelle 8 in Abschnitt 4.2.4.

5.3.2 Stabilität der Ergebnisse

Wie in Abschnitt 4.3.6 beschrieben, haben wir die Stabilität der Klassifikation in zweierlei Hinsicht untersucht. Die Verlässlichkeit der Vorhersagen bei mehreren Durchläufen (Stabilität 1) ist dargestellt in Abbildung 22. Die Standardabweichungen in der Klassifikationsgenauigkeit je Property sind als Fehlerbalken eingezeichnet. Man erkennt, dass die Vorhersagen bei mehrmaligem Durchlauf sehr stabil sind. Die Saisonale Stabilität (Stabilität 2) ist dargestellt in Abbildung 23. Man erkennt auch hier, dass die Klassifikation stabil ist, jedoch in den Sommermonaten (Woche 1 – 17) leicht schlechter ist, als im übrigen Jahr.

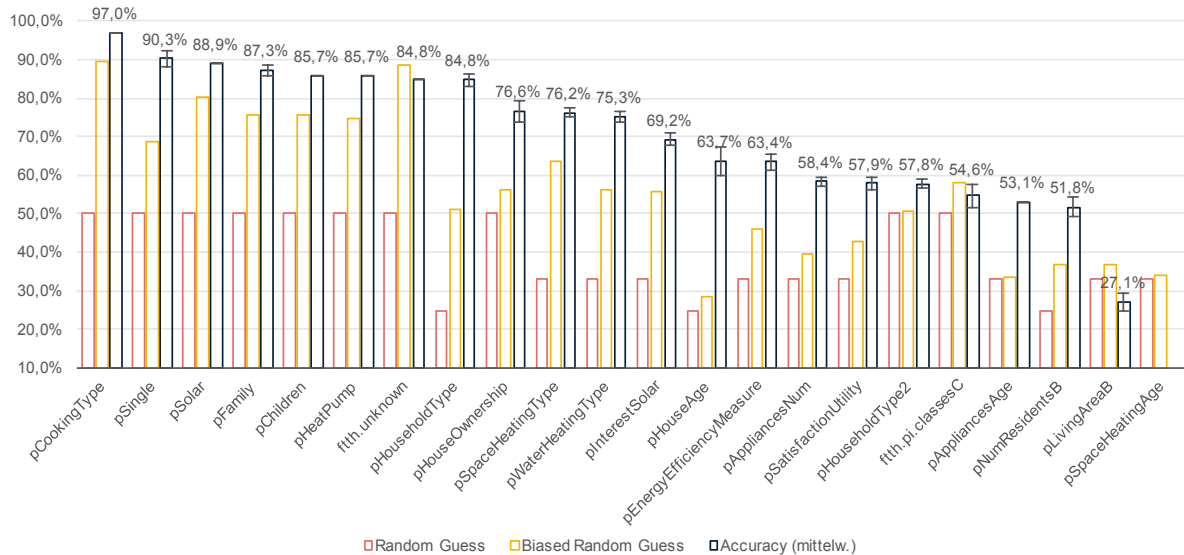


Abbildung 22: Die Klassifikationsergebnisse sind bei der mehrmaligen Ausführung stabil und weisen nur eine geringe Standardabweichung bei der Accuracy auf (vgl. Fehlerbalken im Diagramm)

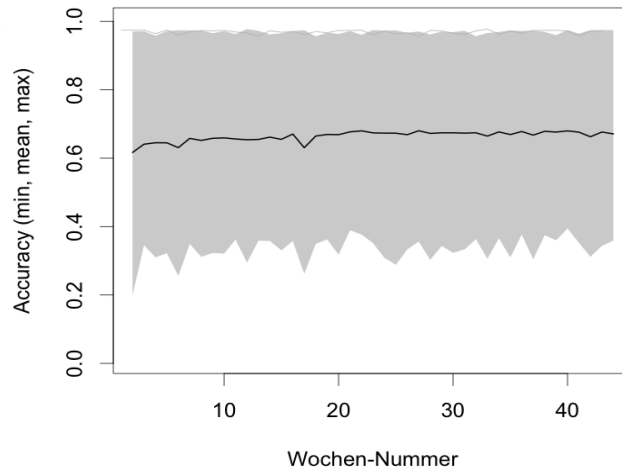


Abbildung 23: Durchschnittliche Klassifikationsgenauigkeit über alle Properties in den 52 Wochen des Arbon Energie AG Datensatzes (Juni 2014 bis Mai 2015).

5.3.3 Klassifikation mit mehreren Wochen

Durch die Verwendung mehrerer Wochen in der Klassifikation konnte die Genauigkeit verbessert werden. Die Ergebnisse von dieser mehrwöchigen Klassifikation sind in der Abbildung 24 dargestellt.

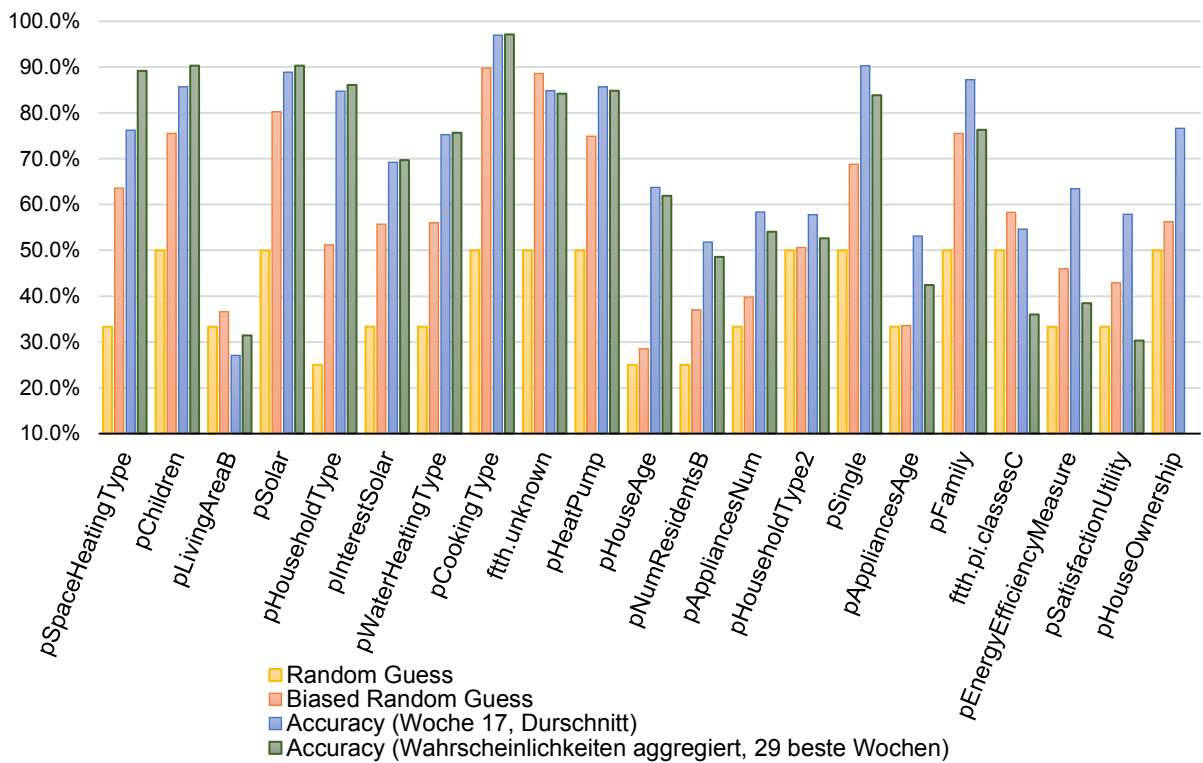


Abbildung 24: Ergebnisse Ein- / Mehrwöchige Klassifikation verglichen mit BRG und Random Guess



5.5 Test der Übertragbarkeit des Tools auf die Schweiz

Zum Testen der Übertragbarkeit haben wir den irischen Datensatz aus der ersten Projektphase verwendet und den in dieser Projektphase erhobenen schweizerischen Datensatz verwendet. Auf Basis der Daten eines Landes wurde jeweils ein Klassifikator trainiert, der dann auf jeweils beide Datensätze angewandt wurde. Wir vergleichen das Ergebnis welches der Klassifikator, der auf den original Daten trainiert worden ist, mit denen, die Klassifikator mit Daten eines anderen Landes erreichen kann. Das Vorgehen zum Testen der Übertragbarkeit ist Abbildung 25 Beispielhaft dargestellt. Die Ergebnisse für alle Properties werden in Abbildung 26 dargestellt.

Als Ergebnis können wir feststellen, dass die Übertragbarkeit für verschiedene Properties unterschiedlich gut möglich ist:

- Haushalte mit Familien in der Schweiz können mit dem in Irland trainierten Algorithmus ähnlich gut erkannt werden.
- Single-Haushalte und Kinder im Haushalt können mit leichten Einbussen in der Accuracy erkannt werden.
- Der mit den irischen Daten trainierte Klassifizierer kann elektrische Kochstellen in der Schweiz nur schlecht erkennen.

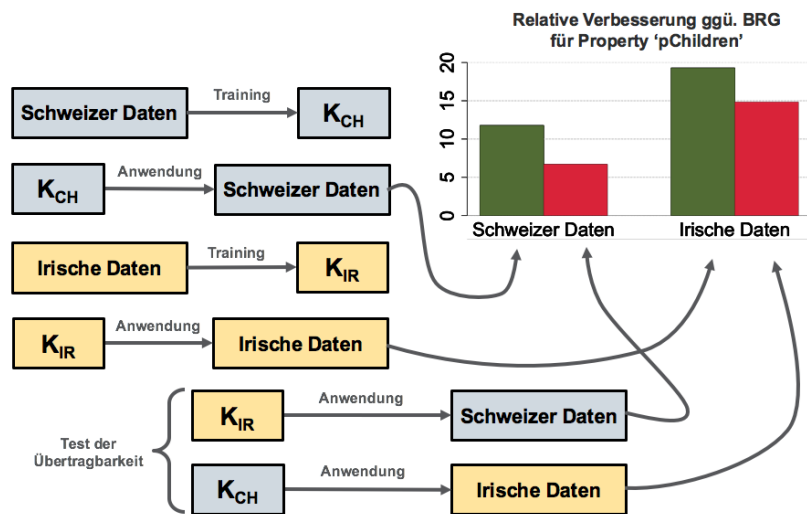


Abbildung 25: Schaubild zum Test der Übertragbarkeit und zur Lesart der Ergebnisdarstellung für die Haushaltseigenschaft Kinder im Haushalt (pChildren)

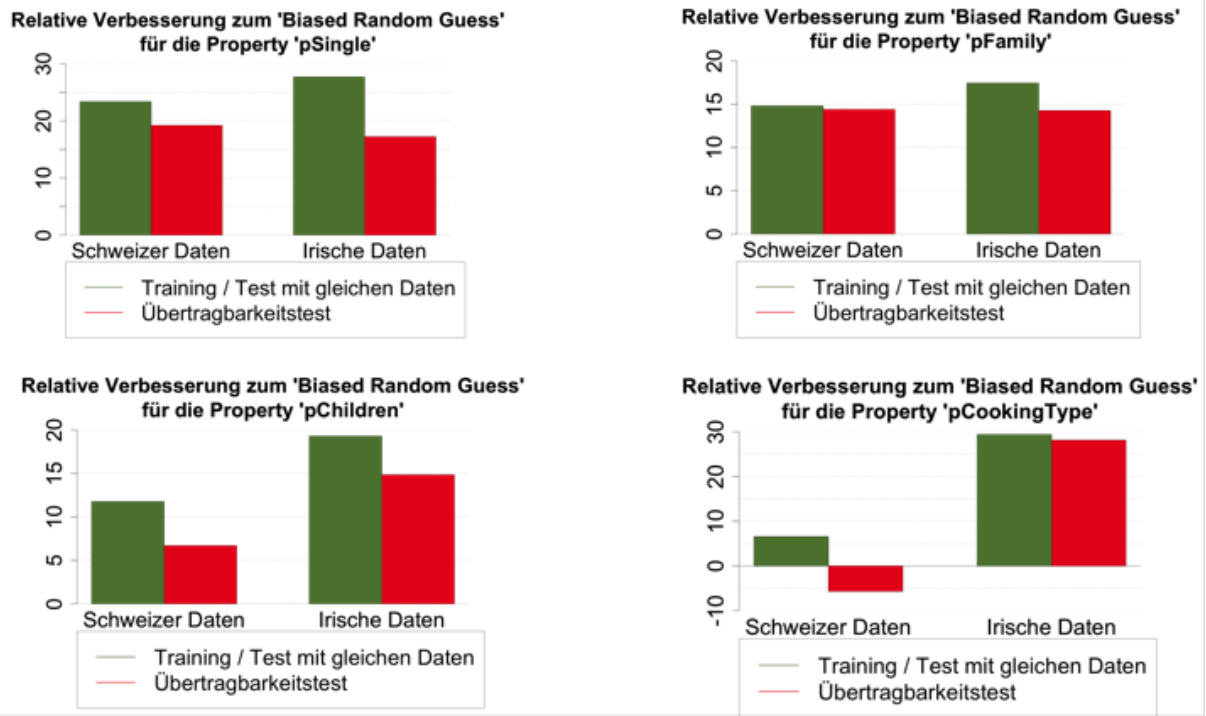


Abbildung 26: Ergebnisse des Übertragbarkeits-Testes für die gemeinsame Properties

6. Würdigung der Ergebnisse

Im Rahmen des Projekts wurde ein Haushalts-Klassifikations-Tool in der Statistiksoftware GNU R entwickelt, mit dem 38 Haushaltscharakteristika zu Lebenssituation, Energieeffizienz und erneuerbaren Energien aus Verbrauchsinformationen intelligenter Stromzähler unter Einhaltung von Datenschutzbestimmungen erkannt werden können. Dabei wird eine durchschnittliche Genauigkeit von 70% erreicht.

Durch die Anbindung von zusätzlichen Datenquellen (aus Geoinformationssystemen, Wetterdaten, usw.) konnte die Fehlerrate um durchschnittlich 5% reduziert werden. Bei 30% der Charakteristika sogar um 15%. Die Relevanz der geographischen Informationen zeigt auch, dass 8 von 10 der am häufigsten ausgewählten Features aus geografischen Daten stammen. Allerdings können die präsentierten Ergebnisse nicht ohne Smart-Meter-Daten erreicht werden, was ein typisches Phänomen im Bereich Machine-Learning bestätigt: Features die in einer Feature-Menge relevant sind, können alleine genommen irrelevant sein (vgl. Guyon & Elisseeff, 2006, S. 11).

Die regionale Übertragbarkeit des Klassifizierers konnte gezeigt werden. Sie funktioniert in drei von vier Fällen zufriedenstellend.

Die gesteckten Projektziele für die beiden Projektphasen wurden vollständig erreicht. Im Nachgang des Projekts soll das entwickelte Tool in einer Feldstudie zusammen mit der Arbon Energie AG in einer Fiber-to-the-Home Cross-Selling Kampagne getestet werden.



7. Schlussfolgerungen

Mit den Smart-Meter-Klassifikations-Verfahren konnten im Rahmen des Projektes 38 Eigenschaften privater Haushalte mit hoher Sicherheit (zum Teil über 70%) aus Lastprofilen und zusätzlichen frei verfügbaren Daten vorhergesagt werden. Neben Umständen der Lebenssituation (z.B. Familien, Rentner, Kinder, sozialer Status), können auch Energieeffizienz-Charakteristika (z.B. Heizungstyp, Hausalter und -grösse, Geräte im Haushalt) und die Einstellungen gegenüber erneuerbarer Energieträgern (z.B. Interesse an Ökostrom oder an Solaranlagen) mit den entwickelten Algorithmen vorhergesagt werden. Die Nutzung der Daten für die Klassifikationsentwicklung setzt eine Einwilligung der Kunden (Zustimmung zu den Datenschutzbestimmungen) voraus.

Gesamthaft schätzen wir, dass durch den konsequenten Einsatz der mit Smart-Meter-Datenanalyse gewonnenen Informationen ein energetisches Potenzial in der Schweiz von ca. 180 GWh gehoben werden kann, entsprechend etwa einem Prozentpunkt an zusätzlichen Einsparungen gegenüber konventioneller Smart-Meter-Anwendungen in Haushalten.

8. Ausblick

Die im Projekt erzielten Ergebnisse sollen über verschiedene Wege der Öffentlichkeit und den Energiedienstleistern zur Verfügung gestellt werden. Die folgenden Massnahmen zur Förderung des Wissens- und Technologietransfers der Projektergebnisse sind vorgesehen:

- Durchführung einer Feldstudie zusammen mit dem Praxispartner Arbon Energie AG, in der das Cross-Selling-Potential der Haushaltsklassifikationsmethoden im Rahmen einer Fiber-to-the-Home Marketingkampagne getestet werden soll
- Austausch mit dem ETH-Spinoff-Unternehmen BEN Energy AG, um eine Markteinführung des Tools zu unterstützen und die Technologie weiteren Energiedienstleistungsunternehmen zugänglich zu machen
- Einbeziehung der Ergebnisse und der Methoden welche im Projekt entwickelt wurden in der Lehre an der ETH Zürich und der Universität Bamberg in der Vorlesung „Energieinformatik“ an der ETH Zürich in den Vorlesungen und „Energieeffiziente Systeme“, „Data Analytics in der Energieinformatik“ und „Business Intelligence & Analytics“ an der Universität Bamberg
- Vorstellung der Ergebnisse und Methoden in nationalen und internationalen Fachzeitschriften bzw. auf Konferenzen

9. Referenzen

9.1 Veröffentlichungen über die Ergebnisse dieses Projekts

Im Rahmen des Projekts wurden folgende Veröffentlichungen herausgegeben, weitere befinden sich in Vorbereitung:

Sodenkamp, M., Kozlovskiy, I., & Staake, T. (2016). Supervised classification with interdependent variables to support targeted energy efficiency measures in the residential sector. *Decision Analytics*, 3(1).



- Sodenkamp, M., Kozlovskiy, I., & Staake, T. (2015). Gaining IS Business Value through Big Data Analytics: A Case Study of the Energy Sector. *Proceedings of the Thirty Sixth International Conference on Information Systems (ICIS), Fort Worth, USA, 13-16 December 2015*.
- Sodenkamp, M., Hopf, K., & Staake, T. (2014). Using Supervised Machine Learning to Explore Energy Consumption Data in Private Sector Housing. *Handbook of Research on Organizational Transformations Through Big Data Analytics*, 320.
- Hopf, K., Sodenkamp, M., Kozlovskiy, I., & Staake, T. (2014). Feature extraction and filtering for household classification based on smart electricity meter data. *Computer Science-Research and Development*, 1-8.
- Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, 78, 397-410.

9.2 Referenzierte Literatur

- Beckel, C., Sadamori, L., & Santini, S. (2012). Towards automatic classification of private households using electricity consumption data. In G. J. Pappas (Hrsg.), *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* (S. 169–176). Toronto and Ontario and Canada: ACM.
- Beckel, C., Sadamori, L., & Santini, S. (2013). Automatic socio-economic classification of households using electricity consumption data. In D. Culler & C. Rosenberg (Hrsg.), *Proceedings of the Fourth International Conference on Future Energy Systems* (S. 75–86). Berkeley and California and USA: ACM.
- BFE (2015). Analyse des schweizerischen Energieverbrauchs 2000 - 2014 nach Verwendungszwecken, Autoren: Infras AG, Prognos AG, TEP, Im Auftrag des Bundesamtes für Energie (BFE), Bern, zuletzt abgerufen unter http://www.bfe.admin.ch/themen/00526/00541/00542/02167/index.html?lang=de&dossier_id=02169 am 19.05.2016
- D'Agostino, R. B. (1986). *Goodness-of-fit-techniques* (Vol. 68). CRC press.
- EnergieSchweiz (2015). Energieetikette für Lampen und Leuchten: Faktenblatt, 02.07.2015, zuletzt abgerufen unter http://www.bfe.admin.ch/energieetikette/00887/02072/index.html?lang=de&dossier_id=05370 am 19.05.2016
- Figueiredo, V., Rodrigues, F., Vale, Z., & Gouveia, J. B. (2005). An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, 20(2), 596–602.
- Guyon, I., & Elisseeff, A. (2006). An Introduction to Feature Extraction. In I. Guyon, M. Nikravesh, S. Gunn, & L. Zadeh (Eds.), *Feature Extraction: Foundations and Application* (pp. 1–25). Berlin, Heidelberg: Springer.
- Juster, F. T. (1966). Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. *Journal of the American Statistical Association*, 61(315), 658–696.
- Kaiser, F. G., Oerke, B., & Bogner, F. X. (2007). Behavior-based environmental attitude: Development of an instrument for adolescents. *Journal of Environmental Psychology*, 27(3), 242–251.
- Kim, H.-W., Chan, H. C., & Gupta, S. (2007). Value-based Adoption of Mobile Internet: An empirical investigation. *Decision Support Systems*, 43(1), 111–126.



- Kwac, J., Tan, C.-W., Sintov, N., Flora, J., & Rajagopal, R. (2013). Utility customer segmentation based on smart meter data: Empirical study. In *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on* (S. 720–725). IEEE.
- Räsänen, T., Ruuskanen, J., & Kolehmainen, M. (2008). Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Applied Energy*, 85(9), 830–840.
- Sánchez, I. B., Espinos, I. D., Moreno Sarrion, L., Quijano López, A., & Burgos, I. N. (2009). Clients segmentation according to their domestic energy consumption by the use of self-organizing maps. In *Energy Market, 2009. EEM 2009. 6th International Conference on the European* (S. 1–6). Abgerufen von <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5207172>
- Verdu, S. V., Garcia, M. O., Senabre, C., Marin, A. G., & Franco, F. J. G. (2006). Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps. *IEEE Transactions on Power Systems*, 21(4), 1672–1682.
- Walsh, G., & Beatty, S. E. (2007). Customer-based corporate reputation of a service firm: scale development and validation. *Journal of the Academy of Marketing Science*, 35(1), 127–143.



Anhang

Anhang 1: Weitere Tabellen

Tabelle A1: Liste aller implementierten Zeitreihen-Features (Stromverbrauch und Wetterdaten); ein Kreuz in der Spalte „30-min“ bedeutet, dass das Feature auf Basis eines 30-min Zeitintervalls berechnet wurde, ein zusätzliches Kreuz in der Spalte „15-min zusätzlich“ bedeutet, dass dieses Feature zusätzlich auf Basis eines 15-min Zeitintervalls berechnet wurde; Features mit schwarzem Kreuz wurden in der ersten Projektphase, Features mit rotem Kreuz in der zweiten Projektphase definiert; Features mit grauem Hintergrund in diesen Spalten wurden schlussendlich im Klassifikationstool verwendet.

Kategorie	Name	Beschreibung	30-min	15-min zusätzlich	Alternativwert (für NA-Werte)
Statistisch	b_day_diff	Abweichung der Messwerte an Wochentagen	X	X	
Statistisch	b_day_weak	Schwache Version von b_day_diff	X	X	
Statistisch	bg_variety	60%-Quartil der Verteilung der Abweichung zum vorhergehenden Messwert	X	X	
Statistisch	c_max_avg	Mittleres Tagesmaximum	X	X	
Statistisch	c_min_avg	Mittleres Tagesminimum	X	X	
Statistisch	c_sm_max	Maximum bei einfacher Glättung	X	X	
Statistisch	number_big_peaks	Anzahl der Verbrauchsspitzen nach einer groben Glättung	X	X	
Statistisch	number_small_peaks	Anzahl kleiner Spitzen an Wochentagen (Berechnung mittels gleitendem gewichtetem Mittelwert)	X	X	
Statistisch	number_zeros	Anzahl von Null-Werten	X	X	
Statistisch	s_cor	Korrelation zwischen Tag 1 und Tag 2 der Woche	X	X	
Statistisch	s_cor_wd	Mittlere Korrelation zwischen den Wochentagen	X	X	
Statistisch	s_cor_wd_we	Korrelation zwischen Wochentagen und Wochenendtagen	X	X	
Statistisch	s_cor_we	Korrelation zwischen Sa und So	X	X	
Statistisch	s_diff	Summe der Differenzen zum Vorgänger (Betrag)	X	X	
Statistisch	s_max	Wochen-Maximum	X	X	
Statistisch	s_min	Wochen-Minimum	X	X	
Statistisch	s_num_peaks	Anzahl der Spitze (lokales Maximum bei der Betrachtung von drei Messwerten)	X	X	
Statistisch	s_q1	Unteres Quartil	X	X	
Statistisch	s_q2	Median	X	X	
Statistisch	s_q3	Oberes Quartil	X	X	
Statistisch	s_sm_max	Gleitendes Maximum	X	X	
Statistisch	s_var_wd	Varianz an Wochentagen	X	X	
Statistisch	s_var_we	Varianz an Wochenendtagen	X	X	
Statistisch	s_variance	Varianz	X	X	
Statistisch	s_wd_max	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X	X	
Statistisch	s_wd_min	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X	X	
Statistisch	s_we_max	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X	X	
Statistisch	s_we_min	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X	X	
Statistisch	sm_variety	20%-Quartil der Verteilung der Abweichung zum vorhergehenden Messwert	X	X	
Verbrauch	c_afternoon	Mittlerer Verbrauch am Nachmittag (14:00 - 17:59)	X		
Verbrauch	c_afternoon_no_min	c_afternoon abzgl. Minimum	X	X	
Verbrauch	c_evening	Mittlerer Verbrauch am Mittag (18:00 - 21:59)	X		
Verbrauch	c_evening_no_min	c_evening abzgl. Minimum	X	X	
Verbrauch	c_morning	Mittlerer Verbrauch am Morgen (6:00 - 9:59 Uhr)	X		
Verbrauch	c_morning_no_min	c_morning abzgl. Minimum	X	X	
Verbrauch	c_night	Mittlerer Verbrauch in der Nacht (1:00 - 5:59 Uhr)	X		



Verbrauch	c_night_no_min	c_night abzgl. Minimum	X	X	
Verbrauch	c_noon	Mittlerer Verbrauch am Mittag (10:00 - 13:59)	X		
Verbrauch	c_noon_no_min	c_noon abzgl. Minimum	X	X	
Verbrauch	c_wd_afternoon	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X		
Verbrauch	c_wd_evening	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X		
Verbrauch	c_wd_morning	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X		
Verbrauch	c_wd_night	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X		
Verbrauch	c_wd_noon	Obige Definition, beschränkt auf Wochentage (Mo-Fr)	X		
Verbrauch	c_we_afternoon	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X		
Verbrauch	c_we_evening	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X		
Verbrauch	c_we_morning	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X		
Verbrauch	c_we_night	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X		
Verbrauch	c_we_noon	Obige Definition, beschränkt auf Wochenendtage (Sa,So)	X		
Verbrauch	c_week	Mittlerer Verbrauch über die gesamte Woche	X		
Verbrauch	c_weekday	Mittlerer Verbrauch an Wochentagen (Mo-Fr)	X		
Verbrauch	c_weekend	Mittlerer Verbrauch am Wochenende (Sa,So)	X		
Verhältnis	r_afternoon_wd_we	Verhältnis des Verbrauchs Nachmittags - Wochentags – Wochenendtags	X		
Verhältnis	r_day_night_no_min	r_night_day (Minimum ist jeweils abgezogen)	X	X	
Verhältnis	r_evening_noon	Verhältnis c_evening / c_noon	X		
Verhältnis	r_evening_noon_no_min	r_evening_noon (Minimum ist jeweils abgezogen)	X	X	
Verhältnis	r_evening_wd_we	Verhältnis des Verbrauchs Abends - Wochentags - Wochenendtags	X		
Verhältnis	r_max_wd_we	Verhältnis des Maximums Wochentags - Wochenendtags	X	X	
Verhältnis	r_mean_max	Verhältnis c_week / c_max	X	X	
Verhältnis	r_mean_max_no_min	r_mean_max (Minimum ist jeweils abgezogen)	X	X	
Verhältnis	r_min_mean	Verhältnis c_min / c_week	X	X	
Verhältnis	r_min_wd_we	Verhältnis des Minimums Wochentags - Wochenendtags	X	X	
Verhältnis	r_morning_noon	Verhältnis c_morning / c_noon	X		
Verhältnis	r_morning_noon_no_min	r_morning_noon (Minimum ist jeweils abgezogen)	X	X	
Verhältnis	r_morning_wd_we	Verhältnis des Verbrauchs Morgens - Wochentags – Wochenendtags	X		
Verhältnis	r_night_day	Verhältnis c_night / c_week	X		
Verhältnis	r_night_wd_we	Verhältnis des Verbrauchs Nachts - Wochentags – Wochenendtags	X		
Verhältnis	r_noon_wd_we	Verhältnis des Verbrauchs Mittags - Wochentags – Wochenendtags	X		
Verhältnis	r_var_wd_we	Verhältnis der Varianz Wochentags - Wochenendtags	X	X	
Verhältnis	r_wd_evening_noon	Verhältnis c_wd_morning / c_wd_noon	X		
Verhältnis	r_wd_morning_noon	Verhältnis c_wd_morning / c_wd_noon	X		
Verhältnis	r_wd_night_day	Verhältnis c_wd_night / c_wd_weekend	X		
Verhältnis	r_we_evening_noon	Verhältnis c_we_morning / c_we_noon	X		
Verhältnis	r_we_morning_noon	Verhältnis c_we_morning / c_we_noon	X		
Verhältnis	r_we_night_day	Verhältnis c_we_night / c_we_weekend	X		
Wetter	w_airPr_cor_daytime	Linearer Zusammenhang zwischen Luftdruck und Stromverbrauch jeweils tagsüber 6:00 - 17:59 Uhr Mo-Fr	X		0
Wetter	w_airPr_cor_evening	Linearer Zusammenhang zwischen Luftdruck und Stromverbrauch jeweils abends 6:00 - 17:59 Uhr Mo-Fr	X		0
Wetter	w_airPr_cor_maxmin	Linearer Zusammenhang zwischen den Maxima des Stromverbrauch und Minima des Luftdruck	X		0
Wetter	w_airPr_cor_weekday_weekend	Korrelation von Luftdruck und Stromverbrauch Wochentags / Wochenende	X		0
Wetter	w_airPr_cor_daily	Linearer Zusammenhang zwischen Luftdruck und Stromverbrauch innerhalb eines Tages Woche	X		0
Wetter	w_airPr_cor_minima	Linearer Zusammenhang zwischen den Minima des Luftdrucks und Stromverbrauch	X		0



Wetter	w_airPr_cor_night	Linearer Zusammenhang zwischen Luftdruck und Stromverbrauch jeweils in der Nacht 0:00 - 5:59 Uhr	x		0
Wetter	w_airPr_cor_overall	Linearer Zusammenhang zwischen Luftdruck und Stromverbrauch in einer Woche	x		0
Wetter	w_prec_cor_weekday_weekend	Korrelation von Niederschlag und Stromverbrauch Wochentags / Wochenende	x		0
Wetter	w_prec_cor_daily	Linearer Zusammenhang zwischen Niederschlag und Stromverbrauch innerhalb eines Tages Woche	x		0
Wetter	w_prec_cor_daytime	Linearer Zusammenhang zwischen Niederschlag und Stromverbrauch jeweils tagsüber 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_prec_cor_evening	Linearer Zusammenhang zwischen Niederschlag und Stromverbrauch jeweils abends 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_prec_cor_maxmin	Linearer Zusammenhang zwischen den Maxima des Stromverbrauch und Minima des Niederschlags	x		0
Wetter	w_prec_cor_minima	Linearer Zusammenhang zwischen den Minima des Niederschlags und Stromverbrauch	x		0
Wetter	w_prec_cor_night	Linearer Zusammenhang zwischen Niederschlag und Stromverbrauch jeweils in der Nacht 0:00 - 5:59 Uhr	x		0
Wetter	w_prec_cor_overall	Linearer Zusammenhang zwischen Niederschlag und Stromverbrauch in einer Woche	x		0
Wetter	w_skyc_cor_weekday_weekend	Korrelation von Bedeckungsgrad des Himmels und Stromverbrauch Wochentags / Wochenende	x		0
Wetter	w_skyc_cor_daily	Linearer Zusammenhang zwischen Bedeckungsgrad des Himmels und Stromverbrauch innerhalb eines Tages Woche	x		0
Wetter	w_skyc_cor_daytime	Linearer Zusammenhang zwischen Bedeckungsgrad des Himmels und Stromverbrauch jeweils tagsüber 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_skyc_cor_evening	Linearer Zusammenhang zwischen Bedeckungsgrad des Himmels und Stromverbrauch jeweils abends 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_skyc_cor_maxmin	Linearer Zusammenhang zwischen den Maxima des Stromverbrauch und Minima des Bedeckungsgrad des Himmels	x		0
Wetter	w_skyc_cor_minima	Linearer Zusammenhang zwischen den Minima des Bedeckungsgrad des Himmels und Stromverbrauch	x		0
Wetter	w_skyc_cor_night	Linearer Zusammenhang zwischen Bedeckungsgrad des Himmels und Stromverbrauch jeweils in der Nacht 0:00 - 5:59 Uhr	x		0
Wetter	w_skyc_cor_overall	Linearer Zusammenhang zwischen Bedeckungsgrad des Himmels und Stromverbrauch in einer Woche	x		0
Wetter	w_temp_cor_daytime	Linearer Zusammenhang zwischen Temperatur und Stromverbrauch jeweils tagsüber 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_temp_cor_evening	Linearer Zusammenhang zwischen Temperatur und Stromverbrauch jeweils abends 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_temp_cor_maxmin	Linearer Zusammenhang zwischen den Maxima des Stromverbrauch und Minima der Temperatur	x		0
Wetter	w_temp_cor_minima	Linearer Zusammenhang zwischen den Minima von Temperatur und Stromverbrauch	x		0
Wetter	w_temp_cor_weekday_weekend	Korrelation von Temperatur und Stromverbrauch Wochentags / Wochenende	x		0
Wetter	w_temp_cor_daily	Linearer Zusammenhang zwischen Temperatur und Stromverbrauch innerhalb eines Tages Woche	x		0
Wetter	w_temp_cor_night	Linearer Zusammenhang zwischen Temperatur und Stromverbrauch jeweils in der Nacht 0:00 - 5:59 Uhr	x		0
Wetter	w_temp_cor_overall	Linearer Zusammenhang zwischen Temperatur und Stromverbrauch in einer Woche	x		0
Wetter	w_windSp_cor_daytime	Linearer Zusammenhang zwischen Windgeschwindigkeit und Stromverbrauch jeweils tagsüber 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_windSp_cor_evening	Linearer Zusammenhang zwischen Windgeschwindigkeit und Stromverbrauch jeweils abends 6:00 - 17:59 Uhr Mo-Fr	x		0
Wetter	w_windSp_cor_maxmin	Linearer Zusammenhang zwischen den Maxima des Stromverbrauch und den Minima in der Windgeschwindigkeit	x		0
Wetter	w_windSp_cor_minima	Linearer Zusammenhang zwischen den Minima in der Windgeschwindigkeit und Stromverbrauch	x		0
Wetter	w_windSp_cor_	Linearer Zusammenhang zwischen Windgeschwindigkeit und	x		0



	overall	Stromverbrauch in einer Woche			
Wetter	w_windSp_cor_weekday_weekend	Korrelation von Windgeschwindigkeit und Stromverbrauch Wochentags / Wochenende	X		0
Wetter	w_windSp_cor_daily	Linearer Zusammenhang zwischen Windgeschwindigkeit und Stromverbrauch innerhalb eines Tages Woche	X		0
Wetter	w_windSp_cor_night	Linearer Zusammenhang zwischen Windgeschwindigkeit und Stromverbrauch jeweils in der Nacht 0:00 - 5:59 Uhr	X		0
Zeitlich	above_base	Anzahl der Messpunkte über der Grundlast-Grenze	X	X	
Zeitlich	const_time	Geschätzte Zeit der Grundlast	X	X	
Zeitlich	dist_big_v	Abstand zwischen hohen Werten	X	X	
Zeitlich	first_above_base	Erstes Überschreiten einer als Grundlast angenommenen Grenze	X	X	
Zeitlich	percent_above_base	Anteil der Messpunkte über der Grundlast-Grenze	X	X	
Zeitlich	t_above_0.5kwh	Zeitpunkt der ersten Überschreitung der 0.5kWh-Grenze, (Gemittelt über alle Wochentage)		X	-1
Zeitlich	t_above_1kwh	Zeitpunkt der ersten Überschreitung der 1kWh-Grenze, (Gemittelt über alle Wochentage)	X	X	-1
Zeitlich	t_above_2kwh	Zeitpunkt der ersten Überschreitung der 2kWh-Grenze, (Gemittelt über alle Wochentage)	X	X	-1
Zeitlich	t_above_mean	Anzahl der Datenpunkte über dem Wochenmittel, (In der gesamten Woche)	X	X	-1
Zeitlich	t_daily_max	Zeitpunkt der ersten Erreichens des Tages-Maximum, (Gemittelt über alle Wochentage)	X	X	-1
Zeitlich	t_daily_min	Zeitpunkt der ersten Erreichens des Tages-Minimum, (Gemittelt über alle Wochentage)	X	X	-1
Zeitlich	time_above_base2	Anzahl der Messpunkte über der Grundlast-Grenze (Alternative Berechnung)	X	X	
Zeitlich	ts_acf_mean3h	Mittlere Autokorrelation (über eine Zeitspanne von 3h)	X	X	
Zeitlich	ts_acf_mean3h_weekday	Mittlere Autokorrelation (über eine Zeitspanne von 3h) an Wochentagen	X	X	
Zeitlich	ts_stl_varRem	Mittlerer Rest bei Saison- / Trend- Dekomposition	X	X	
Zeitlich	value_above_base	Summe der Messpunkte über der Grundlast-Grenze	X	X	
Zeitlich	value_min_guess	Geschätzte Grundlast	X	X	
Zeitlich	wide_peaks	Anzahl der Spitzen (alternative Berechnung)	X	X	
Zeitlich	width_peaks	Durchschnittliche Ausprägung des Peaks	X	X	

Tabelle A2: Liste aller implementierten geographischen Features.

Kategorie	Feature Name	Beschreibung	Werte
Topologie	num_Nodes	„Node“ ist ein Datentyp in OpenStreetMap der einen einzelnen Punkt auf der Karte darstellt. Dieses Feature ist die Anzahl aller Nodes im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt.	Ganzzahl
Topologie	num_NodeTags	„Tags“ beschreiben Attribute (z.B. Verkehrsschilder, Kontaktinformationen) von Objekten in OpenStreetMap, hier von Nodes. Dieses Feature zählt durchschnittliche Anzahl an Tags pro Nodes im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt.	Ganzzahl
Topologie	num_Ways	„Way“ ist ein Datentyp in OpenStreetMap der einen Polygonzug auf Basis mehrerer „Nodes“ beschreibt. Dieses Feature ist die Anzahl aller Ways im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt.	Ganzzahl
Topologie	num_WayTags	„Tags“ beschreiben Attribute (z.B. Gebäudetyp, Anzahl an Stockwerken) von Objekten in OpenStreetMap, hier von Ways. Dieses Feature zählt durchschnittliche Anzahl an Tags pro Way im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt.	Ganzzahl
Topologie	num_Relations	„Relation“ ist ein Datentyp in OpenStreetMap der geographischen Beziehungen (z.B. Gemeindegliederungen oder Plätze/Innenhöfe, die durch angrenzende Gebäude definiert sind) auf Basis mehrerer „Nodes“, „Ways“ und anderer „Relations“ beschreibt. Dieses Feature ist die Anzahl aller Relations im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt.	Ganzzahl



Topologie	num_Relation Tags	„Tags“ beschreiben Attribute (z.B. Herkunft der Daten, Kontaktinformationen) von Objekten in OpenStreetMap, hier von Relations. Dieses Feature zählt durchschnittliche Anzahl an Tags pro Relation im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt.	Ganzzahl
Topologie	elevation	Beschreibt die Höhe des Haushalts über dem Meeresspiegel	Fliesskommazahl
Topologie	longitude	Beschreibt den Längengrad des Haushalts im WGS 84 Format	Fliesskommazahl
Topologie	latitude	Beschreibt den Breitengrad des Haushalts im WGS 84 Format	Fliesskommazahl
Semantisch / Polygon-basiert	nextlanduseType	Beschreibt den Landnutzungstyp eines Polygons das der Koordinate des Haushalts am nächsten liegt und eine Flächennutzung beschreibt (z.B. Wohngebiet, Stadtgebiet, ländlicher Raum)	Kategorial: "city", "residential", "countryside"
Semantisch / Polygon-basiert	thislanduseType	Beschreibt den Landnutzungstyp eines Polygons das die Koordinate des Haushalts umschliesst und eine Flächennutzung beschreibt (z.B. Wohngebiet, Stadtgebiet, ländlicher Raum)	Kategorial: "city", "residential", "countryside"
Semantisch / Polygon-basiert	buildingType Mode	Der flächenmässig am meisten auftretende Gebäudetyp im rechteckigen Kartenausschnitt mit 500x500m um den Haushalt	Kategorial: "apartments", "singlefamily", "nonresidential", "notspecified"
Semantisch / Polygon-basiert	nextBuildingType	Gebäudetyp eines Hauses das der Koordinate des Haushalts am nächsten liegt (z.B. Mehrfamilienhaus).	Kategorial: "apartments", "singlefamily", "nonresidential", "notspecified"
Semantisch / Polygon-basiert	thisBuildingType	Gebäudetyp eines Hauses das die Koordinate des Haushalts einschliesst (z.B. Mehrfamilienhaus).	Kategorial: "apartments", "singlefamily", "nonresidential", "notspecified"
Semantisch / Punkt-basiert	num_public Institutions	Anzahl der Objekte aus der Kategorie „Öffentliche Einrichtungen“ (z.B. Stadthalle, Rathaus) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_business	Anzahl der Objekte aus der Kategorie „Geschäfte und Büros“ im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_food	Anzahl der Objekte aus der Kategorie „Gaststätten und Cafés“ im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_transportation	Anzahl der Objekte aus der Kategorie „Öffentlicher Personenverkehr“ (z.B. Bushaltestellen, Bahnhöfe) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_recreation	Anzahl der Objekte aus der Kategorie „Erfolg und Freizeit“ (z.B. Schwimmbad, Fitnessstudio) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_culture	Anzahl der Objekte aus der Kategorie „Kultur“ (z.B. Theater, Kino) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_sights	Anzahl der Objekte aus der Kategorie „Sehenswürdigkeiten“ (z.B. Museum) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_countryside	Anzahl der Objekte aus der Kategorie „ländlicher Raum“ (z.B. Scheune im Bauernhof, Felder) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	num_roadSystem	Anzahl der Objekte aus der Kategorie „Verkehrssystem“ (z.B. Verkehrsschilder) im Kartenausschnitt mit 500x500m um den Haushalt	Ganzzahl
Semantisch / Punkt-basiert	mindist_public Institutions	Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_business	Minimale Distanz zu geographischen Objekten in der Kategorie „Geschäfte und Büros“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_food	Minimale Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_transportation	Minimale Distanz zu geographischen Objekten in der Kategorie „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_recreation	Minimale Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_culture	Minimale Distanz zu geographischen Objekten in der Kategorie „Kultur“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_sights	Minimale Distanz zu geographischen Objekten in der Kategorie „Sehenswürdigkeiten“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl



Semantisch / Punkt-basiert	mindist_country-side	Minimale Distanz zu geographischen Objekten in der Kategorie „ländlicher Raum“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	mindist_road System	Minimale Distanz zu geographischen Objekten in der Kategorie „Verkehrssystem“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_public Institutions	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Öffentliche Einrichtungen“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_business	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Geschäfte und Büros“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_food	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Gaststätten und Cafés“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_transportation	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Öffentlicher Personenverkehr“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_recreation	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Erholung und Freizeit“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_culture	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Kultur“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_sights	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Sehenswürdigkeiten“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_countryside	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „ländlicher Raum“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	meandist_road System	Durchschnittliche Distanz zu geographischen Objekten in der Kategorie „Verkehrssystem“ innerhalb eines Kartenausschnitts mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.city	Fläche des Landnutzungsgebiets „Stadtgebiet“ im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.residential	Fläche des Landnutzungsgebiets „Wohngebiet“ im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.countryside	Fläche des ländlichen Raums (Felder, Wald, Wiesen, Bauernhöfe, usw.) im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.not-specified	Gesamte Grundfläche von Gebäuden im Kartenausschnitt mit 500x500m die nicht näher spezifiziert sind	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.single-family	Gesamte Grundfläche von Gebäuden im Kartenausschnitt mit 500x500m die als Einfamilienhäuser gekennzeichnet sind	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.apartments	Gesamte Grundfläche von Gebäuden im Kartenausschnitt mit 500x500m die als Mehrfamilienhäuser gekennzeichnet sind	Fliesskommazahl
Semantisch / Polygon-basiert	totalarea.non-residential	Gesamte Grundfläche von Gebäuden im Kartenausschnitt mit 500x500m um den Haushalt bei denen der Typ bekannt ist und kein Wohngebäude identifiziert, z.B. Industriegebäude, Einkaufszentren, öffentliche Einrichtungen	Fliesskommazahl
Semantisch / Polygon-basiert	building-area.mean	Durchschnittliche Grundfläche aller Gebäude im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	building-area.median	Median der Grundfläche aller Gebäude im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	buildingarea.var	Varianz der Grundfläche aller Gebäude im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Polygon-basiert	nextbuilding.area	Grundfläche des Gebäudes das die Koordinate des Haushalts einschliesst (z.B. Mehrfamilienhaus).	Fliesskommazahl
Semantisch / Punkt-basiert	buidling.distMean	Durchschnittliche Distanz zu allen Gebäude im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl
Semantisch / Punkt-basiert	buidling.distVar	Varianz der Distanz zu allen Gebäude im Kartenausschnitt mit 500x500m um den Haushalt	Fliesskommazahl



Tabelle A3: Koeffizienten für Temperatur im Modell (1) für verschiedene Tageszeiten

Tageszeit	R ²	β ₂	Signifikanz Index ⁴
05:00	0,2142	-0,0016	***
09:00	0,3691	-0,0069	***
19:00	0,7135	-0,0391	***

Tabelle A4: Koeffizienten für Temperatur im Modell (1) für verschiedene Tageszeiten

Tageszeit	R ²	β ₂	Signifikanz Index
00:00	0.3676	-0.0086	***
00:30	0.2966	-0.006	***
01:00	0.2583	-0.0051	***
01:30	0.2274	-0.0037	***
02:00	0.2229	-0.0033	***
02:30	0.2093	-0.0025	***
03:00	0.2217	-0.0024	***
03:30	0.2054	-0.0019	***
04:00	0.216	-0.0019	***
04:30	0.2251	-0.0016	***
05:00	0.2142	-0.0016	***
05:30	0.1393	-0.0012	***
06:00	0.0914	-0.0015	***
06:30	0.0816	-0.0025	***
07:00	0.0448	-0.0036	***
07:30	0.068	-0.0057	***
08:00	0.123	-0.007	***
08:30	0.303	-0.0064	***
09:00	0.3691	-0.0069	***
09:30	0.2864	-0.0081	***
10:00	0.2198	-0.0098	***
10:30	0.1923	-0.0102	***
11:00	0.1821	-0.0118	***
11:30	0.1825	-0.0115	***
12:00	0.1922	-0.0132	***
12:30	0.1976	-0.0124	***
13:00	0.2403	-0.0147	***
13:30	0.2489	-0.0138	***
14:00	0.2917	-0.0153	***
14:30	0.3552	-0.0146	***
15:00	0.4628	-0.0169	***
15:30	0.5806	-0.0176	***
16:00	0,6843	-0.0234	***
16:30	0.7121	-0.0281	***
17:00	0.6653	-0.0362	***
17:30	0.6805	-0.0366	***
18:00	0.6932	-0.0399	***
18:30	0.7313	-0.0371	***
19:00	0.7135	-0.0391	***
19:30	0.6931	-0.0343	***

⁴ Der Signifikanz-Index zeigt das Signifikanzlevel der Koeffizienten: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < < 1



20:00	0.6489	-0.0336	***
20:30	0.6199	-0.0281	***
21:00	0.6077	-0.0262	***
21:30	0.6274	-0.0205	***
22:00	0.6364	-0.0178	***
22:30	0.6243	-0.0135	***
23:00	0.5838	-0.0126	***
23:30	0.457	-0.0099	***



Anhang 2: Datenschutzkonzept

Datenschutzkonzept | 2015-06-05 | Bits to Energy Lab



Datenschutzkonzept für das BFE Projekt Smart-Meter-Datenanalyse für automatisierte Energieberatungen

Bamberg, 06. Juni 2015

1. Einleitung

Mit Unterstützung der Arbon Energie AG führt das Bits to Energy Lab am Lehrstuhl für Energieeffiziente Systeme an der Universität Bamberg und am Lehrstuhl für Informationsmanagement an der ETH Zürich ein wissenschaftliches Projekt durch, in dessen Fokus die Ableitung von Haushaltsscharakteristika aus Lastgängen steht. Die im Rahmen des Projekts erhobenen und zur Verfügung gestellten Daten werden vom Forscherteam ausschliesslich zu nicht-kommerziellem Zweck (Forschung) genutzt.

2. Ziel des Projekts

Ziel des Projektes ist die Entwicklung von Methoden zur automatisierten Erkennung von energieeffizienzrelevanten Charakteristika individueller Haushalte (z.B. Anzahl der Bewohner, Größe und Alter der Wohnung, Nutzung elektrischer Heizungen usw.) anhand von Smart-Meter-Daten, historischen Rechnungen, Wetterindikatoren, Nutzerangaben usw. Durch Kenntnis der Merkmale lässt sich die Wirkung von Energieeffizienz-Kampagnen steigern (z.B. durch die Wahl relevanter Vergleichsgruppen und die Auswahl zielgerichteter Empfehlungen).

Schlussendliches Projektziel ist es, die zu entwickelnden Algorithmen auf Daten der Arbon Energie AG anzuwenden und in einem noch zu definierenden Pilottest zu bewerten. Eine solche Pilotstudie könnte z.B. die Optimierung von Mailing-Rücklaufquoten, die Vorhersage von Vertragsabschlüssen von Ökostrom-Kunden oder die Steigerung der Quote einer Inanspruchnahme von Energieberatungen bei Haushalten mit hohem Einsparpotential sein.

3. Methode

Es werden Supervised-Learning-Methoden in Kombination mit einem Trainingsdatensatz von Arbon Energie verwendet, bei dem Lastprofile und über Umfragen erhobene Haushaltsmerkmale vorhanden sind. Für die Steigerung der Erkennungsgenauigkeit und die Anbindung von geografischen Daten werden anonymisierte Kundenstammdaten benutzt.

4. Datenerhebung

Die Daten werden unterschiedlich erhoben und in verschiedenen IT-Systemen gespeichert. Potentiell teilnehmende Haushalte sind Kunden der Arbon Energie AG.



Smart-Metering-Daten (Arbon Energie AG)

Die Lastgangdaten über den Energieverbrauch von den Haushalten werden von Arbon Energie im Rahmen der Forschungsarbeit zur Verfügung gestellt.

Kundenstammdaten (Arbon Energie AG)

Die Kundenstammdaten (Anrede, Titel, Ortsinformationen usw.) werden zwecks Anbindung von Geoinformationen und Steigerung der Klassifikationsgenauigkeit verwendet. Diese Daten werden in einem IT-System von Arbon Energie AG gespeichert und vor Ort bearbeitet.

Umfragedaten (Arbon Energie AG und Forscherteam)

Im Rahmen der Untersuchung werden potenziellen Teilnehmerinnen und Teilnehmern zu einer Online-Umfrage eingeladen. Die Umfragedaten werden in folgende Kategorien aufgeteilt: Lebenssituation, Sozio-demografische Daten, Energieverbrauch, und Kaufbereitschaft. Erhebung von Umfragedaten erfolgt in zwei Schritten:

a. Anschreiben der Haushalte (Arbon Energie AG)

Das Anschreiben mit der Bitte um Teilnahme an der Pilotstudie erfolgt mit einer Anmeldefrist von ca. 14 Tagen.

b. Registrierung und Befragungen (Forscherteam)

Die angeschriebenen Haushalte können sich innert des im Anschreiben angegebenen Zeitraums auf der Projektwebseite für die Studie registrieren. Anschliessend erhalten sie eine Einladung zum Eingangsfragebogen. Die Identifikation erfolgt über den Namen und die Adresse. Im Eingangsfragebogen erhalten die potentiellen Teilnehmer die Gelegenheit, sich über die Verwendung ihrer Daten zu informieren und ihr Einverständnis dafür abzugeben. Für die Teilnahme an der Studie ist zudem das Einverständnis des Teilnehmers vonnöten, dass das Forscherteam die vorliegenden Daten auswerten kann. Ohne Einverständniserklärung zum Datenschutzkonzept ist eine Teilnahme nicht möglich. Bei der Registrierung erfolgt zudem die Online-Befragung. Die Haushaltsdaten sind für das geplante Training von Algorithmen und die Auswertung der Pilotstudie notwendig.

5. Datenpseudonymisierung

Die Verbrauchsdaten, die Kundenstammdaten, und die Haushaltsdaten aus dem Online-Befragungstool werden in die Studienmanagementdatenbank überführt. Diese wird von dem Bits to Energy Lab verwaltet. Den Umfragedaten werden die Verbrauchsdaten und die Kundenstammdaten zugeordnet. Sobald die Zuordnung erfolgt ist, werden die Datensätze vom Forscherteam pseudonymisiert. Den registrierten Studienhaushalten wird hierzu eine pseudonymisierte Nummer (VID) zugewiesen. Die Liste mit der Zuordnung von Adresse und VID wird an Ar-



bon Energie AG übergeben und aus der Studiendatenbank gelöscht. Diese Liste wird erforderlich, um die Effizienzinterventionen Haushaltsspezifisch zu testen.

6. Datenbearbeitung

Die Erkennung von Haushaltsmerkmalen erfolgt durch Forscherteam nur mit den pseudonymisierten Daten. Somit können persönliche Haushaltsdaten (Angaben aus der Umfrage, Verbrauchsdaten und Kundenstammdaten) von dem Forscherteam nicht zurückverfolgt werden. Die erhobenen Daten werden nur für wissenschaftliche Zwecke (z.B. Publikation in Fachzeitschriften und auf Konferenzen) verwendet. Die Anonymität der Haushalte bleibt ausserhalb des Forscherteams somit vollumfänglich gewahrt. Das Forscherteam holt das Einverständnis der Arbon Energie ein, wenn die Daten an Dritte weitergegeben werden sollen.

7. Datenlöschung

Mit dem Befragungstool erhobene Daten werden nach dem Auslesen innerhalb von Monatsfrist nach Projektende aus der Datenbank des Online-Befragungstools gelöscht, nachdem sie in die Studienmanagement-Datenbank überführt und pseudonymisiert wurden. Eine Kopie der pseudonymisierten Daten wird bei der Bits to Energy Lab aufbewahrt und kann allenfalls zu einem späteren Zeitpunkt für weitere Studienzwecke verwendet werden, um beispielsweise den Energieverbrauch und das Verhalten in der Bevölkerung langfristig zu verfolgen, sofern die Arbon Energie ihr Einverständnis abgibt. Diese Kopie bleibt an dem Bits to Energy während einer Frist von zehn Jahren nach Abschluss des Projektes bestehen. Zusätzlich wird eine Zuordnungstabelle (Studienhaushalt zu pseudonymisierter Nummer (VID)) für allfällige Rückfragen gespeichert. Sollte ein Teilnehmer nach Studienbeginn den Wunsch äussern, aus der Studie auszutreten, hat er das Anrecht auf die Löschung der von ihm erhobenen Daten.

8. Datenzugriff sowie organisatorische und technische Massnahmen

Organisation und technische Umsetzung

Zugang zu den Personen- und Studiendaten hat nur die operative Studienkoordination des Forscherteams (max. sechs Personen). Alle Interaktionen mit den Haushalten werden über die Studienkoordination abgewickelt. Alle weiteren Personen erhalten nur die zur Ausführung ihrer Arbeit notwendigen Informationen und Daten in pseudonymisierter Form.

Der Zugang zu den Daten wird über eine Zugangsverwaltung mit einer Liste der Rollen und Rechte für jedes IT-System geregelt. Diese Liste wird vom Forscherteam verwaltet.

Studienmanagementdatenbank: Zugriff auf die Studienmanagementdatenbank mit den umfassenden Kontakt- und Studiendaten hat nur die operative Studienkoordination.

Befragungstool: Das Befragungstool wird von den Forschern des Bits to Energy Labs betrieben. Zugriff auf das Befragungstool und die dahinterliegende Datenbank haben nur die autorisierten Forscher des Forscherteams.



Anhang 3: Online-Umfragebogen

Energieeffizienz-Umfrage Arbon Energie

Herzlich willkommen!

Der Fragebogen informiert Sie zunächst über die Teilnahme- und Datenschutzbedingungen. Anschliessend stellen wir Ihnen Fragen zu Ihrem Wohnungs-Umfeld, Strom-Verbrauchsverhaltens sowie Fragen zu Ihrer Einstellung in Bezug auf Energieeffizienz. Der Fragebogen besteht aus 7 Seiten und dauert 5 bis 10 Minuten.

Sämtliche Daten werden vor der Auswertung **anonymisiert**, **ausschliesslich zu wissenschaftlichen Zwecken genutzt und nicht an Dritte weitergegeben**. Bitte antworten Sie bei allen Fragen so **spontan und ehrlich** wie möglich. Es gibt **keine richtigen oder falschen Antworten**.

Herzlichen Dank für Ihre Unterstützung!
Ihr Forscherteam an der ETH Zürich und der Otto-Friedrich Universität Bamberg

Die [Datenschutzbedingungen](#) zu dieser Studie besagen im Wesentlichen, dass sämtliche Daten

- nur in anonymisierter Form ausgewertet werden
- nur zu Forschungszwecken verwendet werden und
- von den Forschern nicht an Dritte weitergegeben werden.

1. Erklären Sie sich mit den Datenschutzbedingungen einverstanden? *

- Ja.
- Nein, ich bin mit den Datenschutzbedingungen nicht einverstanden. Ich möchte nicht an der Umfrage teilnehmen.

* = Pflichtfelder

Weiter

0% |



Energieeffizienz-Umfrage Arbon Energie

Herzlich willkommen!

Der Fragebogen informiert Sie zunächst über die Teilnahme- und Datenschutzbedingungen. Anschliessend stellen wir Ihnen Fragen zu Ihrem Wohnungs-Umfeld, Strom-Verbrauchsverhaltens sowie Fragen zu Ihrer Einstellung in Bezug auf Energieeffizienz. Der Fragebogen besteht aus 7 Seiten und dauert 5 bis 10 Minuten.

Sämtliche Daten werden vor der Auswertung anonymisiert, ausschliesslich zu wissenschaftlichen Zwecken genutzt und nicht an Dritte weitergegeben. Bitte antworten Sie bei allen Fragen so spontan und ehrlich wie möglich. Es gibt keine richtigen oder falschen Antworten.

Herzlichen Dank für Ihre Unterstützung!
Ihr Forscherteam an der ETH Zürich und der Otto-Friedrich Universität Bamberg

Die [Datenschutzbedingungen](#) zu dieser Studie besagen im Wesentlichen, dass sämtliche Daten

- nur in anonymisierter Form ausgewertet werden
- nur zu Forschungszwecken verwendet werden und
- von den Forschern nicht an Dritte weitergegeben werden.

1. Erklären Sie sich mit den Datenschutzbedingungen einverstanden? *

- Ja.
- Nein, ich bin mit den Datenschutzbedingungen nicht einverstanden. Ich möchte nicht an der Umfrage teilnehmen.

Vorname *

Nachname *

Strasse *

Hausnummer *



Postleitzahl *

Ort *

* = Pflichtfelder

Weiter



Energieeffizienz-Umfrage Arbon Energie

Personen im Haushalt

8. Anzahl der **Personen** im Haushalt (ohne Kinder bis einschliesslich 16 Jahre): *

1 2 3 4 5 oder mehr

9. Anzahl der **Kinder** im Haushalt (bis einschliesslich 16 Jahre): *

0 1 2 3 4 5 oder mehr

10. Wie viele **Personen** sind regelmässig tagsüber zuhause (z.B. 5-6 Stunden pro Tag)? *

0 1 2 3 4 5 oder mehr

[Zurück](#) [Weiter](#)



Energieeffizienz-Umfrage Arbon Energie

Wohnungseigenschaften

11. Wie wohnen Sie? *

- Wohnung in einem Mehrfamilienhaus Doppelhaushälfte
- Einfamilienhaus (freistehend) Reihenhaus

12. Wie gross ist Ihre **Wohnfläche** (ggf. bitte schätzen)? *

 m²

13. Wie sind Ihre **Eigentumsverhältnisse**? *

- Miete Eigentum

14. Kennen Sie das **Baujahr** Ihres Hauses? *

- Ja Nein Ungefähr

Schätzen Sie, wie alt ist das Haus, in dem Sie wohnen: *

- neuer als 5 Jahre
- neuer als 10 Jahre
- neuer als 30 Jahre
- neuer als 75 Jahre
- älter als 75 Jahre

* = Pflichtfelder

Zurück Weiter



Energieeffizienz-Umfrage Arbon Energie

Heizung und Geräte im Haushalt

15. Wie wird Ihr Haushalt beheizt und Ihr Warmwasser gewärmt?
Falls Sie nicht alle Optionen sehen, können Sie den Scrollbalken
(Bildlaufleiste unter der Tabelle) nach rechts schieben. *

	Erdgas	Heizöl	Elektro- speicherheizung	Zentralheizung (Mehrfamilienhaus)
Haupt- Heizung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neben- Heizung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Heizung für Warmwasser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16. Wann wurden Ihre Heizungsanlagen installiert?

Haupt-Heizung	<input type="text"/>
Neben-Heizung	<input type="text"/>
Heizung für Warmwasser	<input type="text"/>

17. Bitte geben Sie an, wieviele der folgenden Geräte in Ihrem Haushalt
vorhanden sind und schätzen Sie deren Alter in Jahren.

	Anzahl regelmässig benutzter Geräte *				Alter des ältesten regelmässig benutzten Geräts (in Jahren)
	0	1	2	3 oder mehr	
Elektroherd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --
Kühlschrank	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --
Separater Gefrierschrank	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --
Waschmaschine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --
Wäschetrockner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --
Geschirrspüler	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --
Fernseher	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-- Bitte auswählen --

* = Pflichtfelder

Zurück Weiter



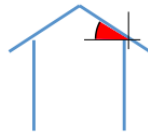
Energieeffizienz-Umfrage Arbon Energie

Solar- und Geothermie-Potential

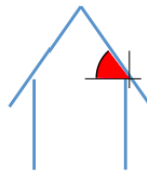
18. Haben Sie einen oder mehrere der folgenden Anlagen installiert?

	Vorhanden? *		Jahr des Einbaus
	Ja	Nein	
Photovoltaik-Anlage (Solarstrom)	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Solarthermie (Heizung oder Warmwasseraufbereitung)	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Luftwärmepumpe	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Erdsonde-Wärmepumpe	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

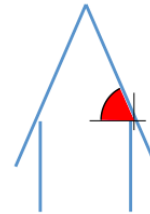
Die Dachneigung beträgt...



keine oder leichte Neigung
0 – 40°



mittlere Neigung
40 – 70°



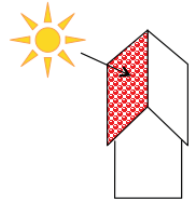
Starke Neigung
70 – 90°

19. Wie stark ist Ihr Dach geneigt? *

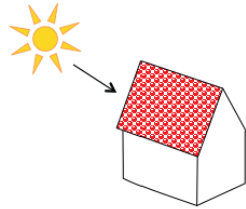
- keine (Flachdach) bis leichte Neigung (0-40°)
- mittlere Neigung (40-70°)
- starke Neigung (40-90°)
- weiss nicht



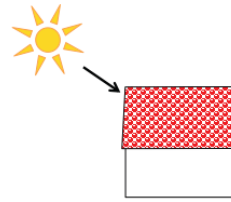
Stand der Mittagssonne (von Süden)



Dachausrichtung:
Nord / Süd



Dachausrichtung:
Süd-West / Süd-Ost



Dachausrichtung:
Ost / West

20. Zeigt eine Seite Ihres Daches zur Mittagssonne? *

- Nord / Süd
- Süd-West / Süd-Ost
- Ost / West
- weiss nicht

21. Wie stehen Sie zu folgenden Aussagen über Ihre Bereitschaft zur Anschaffung einer Solaranlage (Photovoltaik oder Solarthermie)?

1 = "Trifft überhaupt nicht zu", 7 = "Trifft vollkommen zu". *

	trifft überhaupt nicht zu 1	2	3	4	5	6	trifft vollkommen zu 7	keine Angabe
Ich könnte mir vorstellen, in den nächsten 1-2 Jahren eine Solaranlage anzuschaffen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich beabsichtige die Anschaffung einer Solaranlage in den nächsten 1-2 Jahren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich plane, in den nächsten 1-2 Jahren eine Solaranlage anzuschaffen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück Weiter



Energieeffizienz-Umfrage Arbon Energie

Fragen zur Energieeinsparung

22. Wie hoch schätzen Sie den Anteil (in %) der Energiesparlampen in Ihrem Haushalt? *

- 0 %
 25 %
 50 %
 75 %
 100 %
 weiss nicht

23. Wie hoch schätzen Sie den Anteil (in %) der doppelt oder dreifach verglasten Fenster in Ihrem Haushalt? *

- 0 %
 25 %
 50 %
 75 %
 100 %
 weiss nicht

24. Welche der folgenden Energiesparmassnahmen wurden an Ihrem Haus in den letzten 15 Jahren umgesetzt? *

- Dämmung des Dachs bzw. Obergeschosses
 Gebäudeisolierung
 Kellerisolierung

 Fenstererneuerung
 keine
 weiss nicht

Mit den folgenden Fragen möchten wir ein Gefühl dafür entwickeln, wie wichtig ihnen Energiesparen ist:

25. Wie stehen Sie zu folgenden Aussagen?

1 = "Trifft überhaupt nicht zu", 7 = "Trifft vollkommen zu". *

	trifft überhaupt nicht zu	1	2	3	4	5	6	trifft vollkommen zu	7	keine Angabe
Wenn ich Pullover oder Hosen einen Tag lang getragen habe, kommen sie immer in die Wäsche.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Wenn ich als letzte Person den Raum verlasse, mache ich immer das Licht aus.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Ich lasse strombetriebene Geräte (TV, Stereoanlagen, Drucker, ...) immer auf Standby laufen.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Im Winter drehe ich das Thermostat runter, wenn ich die Wohnung für mehr als vier Stunden verlasse.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Im Winter ist es in meiner Wohnung immer warm genug, um nur ein T-Shirt zu tragen.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
In Hotels lasse ich die Handtücher immer täglich austauschen.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

26. Wie stehen Sie bzw. Ihr Haushalt zu folgenden Aussagen?

1 = "Trifft überhaupt nicht zu", 7 = "Trifft vollkommen zu". *

	trifft überhaupt nicht zu	1	2	3	4	5	6	trifft vollkommen zu	7	keine Angabe
Mein Haushalt spart Energie, um der Umwelt zu helfen.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Mein Haushalt spart Energie, um Geld zu sparen.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Mein Haushalt spart Energie, um den Kindern ein gutes Vorbild zu sein.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Mein Haushalt spart Energie, weil wir generell sparsam leben.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Mein Haushalt spart Energie, weil es in der Familie oder im Freundeskreis so üblich ist.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>
Mein Haushalt interessiert sich für neue Technologien.	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

* = Pflichtfelder



Energieeffizienz-Umfrage Arbon Energie

Letze Seite

Die Arbon Energie möchte die Gelegenheit nutzen, um etwas über Ihre Einstellung zum Unternehmen und zu einem neuen Service-Angebot zu erfahren. Natürlich sind alle Angaben freiwillig und werden nur in anonymisierter Form an Arbon Energie weitergegeben.

27. Wie stehen Sie zu folgenden Aussagen?

1 = "Trifft überhaupt nicht zu", 7 = "Trifft vollkommen zu".

Die Arbon Energie...

	trifft überhaupt nicht zu						trifft vollkommen zu		keine Angabe
	1	2	3	4	5	6	7		
... liefert qualitativ hochwertige Dienstleistungen und Produkte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
... ist ein starkes, verlässliches Unternehmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
... entwickelt innovative Dienstleistungen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
... kümmert sich um Ihre Kunden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
Ich bin mit der Arbon Energie zufrieden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

28.

Die Arbon Energie plant das Angebot von **Fiber-to-the-home** (schnelles Internet über ein Glasfasernetz). Arbon-Energie würde sich über Ihre Meinung dazu freuen.

Wie hoch schätzen Sie die Chance ein, dass Sie innerhalb der nächsten 12 Monate Fiber-to-the-Home beziehen werden?

- 10 - Sicher, auf jeden Fall!
- 9 - Fast sicher.
- 8 - Sehr wahrscheinlich.
- 7 - Wahrscheinlich.
- 6 - Gut möglich.
- 5 - Eventuell möglich.
- 4 - Mit geringer Wahrscheinlichkeit.
- 3 - Mit sehr geringer Wahrscheinlichkeit.
- 2 - Unwahrscheinlich.
- 1 - Sehr unwahrscheinlich.
- 0 - Nein, überhaupt nicht!
- Ich weiss nicht, was Fiber-to-the-home ist

29. Gibt es noch Infos oder Anmerkungen, die Sie uns mitteilen möchten?

[Zurück](#) [Weiter](#)



Energieeffizienz-Umfrage Arbon Energie

Danke!

Der Fragebogen ist nun beendet.

Herzlichen Dank für das Beantworten unseres Fragebogens und die Unterstützung unserer Studie!



Anhang 4: Vergleich der Anteile an mehrfachverglasten Fenstern in Umfragen in Arbon und Irland

Ausgangslage

Das Bits to Energy Lab hat gemeinsam mit der Arbon Energie AG das Projekt „Smart-Meter-Datenanalyse für automatisierte Energieberatungen“ des Bundesministeriums für Energie durchgeführt. Das Ziel des Projekts war die Erkennung von energieeffizienzrelevanten Charakteristika individueller Haushalte anhand von Smart-Meter-Daten und zusätzlichen Kundeninformationen.

Im Rahmen des Projekts wurde zwischen Juni und September 2015 eine Umfrage zur Energieeffizienz, Technologieaffinität, Lebenssituation sowie Kaufbereitschaft eines Glasfaser-Internetanschlusses bei den Kunden der Arbon Energie AG durchgeführt, an der 527 Haushalte erfolgreich teilgenommen haben.

Daneben standen zur vergleichenden Auswertung die Ergebnisse einer Umfrage aus dem Jahr 2009 in Vorbereitung einer nationalen Smart-Meter-Studie in Irland, durchgeführt von der Commission for Energy Regulation (CER, 2011) zur Verfügung. Diese Umfrage umfasst 4'232 Antworten, darunter auch demographische Merkmale wie Alter, Geschlecht, Berufsstand und soziale Klasse.

Die Ergebnisse der Fragen zum Anteil an mehrfachverglasten Fenstern (s. Tabelle A5, Ergebnisse in Abbildung A1) ergeben, dass der Anteil an mehrfachverglasten Fenstern in der Schweiz geringer ist als in Irland. **Da dieses Ergebnis kontraintuitiv ist, soll in diesem Bericht erläutert werden, ob dieses Ergebnis korrekt ist und was gegebenenfalls Gründe für die Abweichung sein können.**

Aus der Tabelle A5 ist es zu sehen, dass die beiden Fragen verschiedene Formulierungen auf zwei Sprachen haben. Diese Tatsache kann das Verständnis und Interpretation der Fragen bei den Kunden beeinflussen und schliesslich die Umfrageresultate in Arbon und in Irland unvergleichbar machen.

Tabelle A5: Vergleich der Fragetexte in den Arbon und Irland Umfragen

	Frage 23 in Arbon (gestellt in 2015)	Frage 4906 in Irland (gestellt in 2009)
Fragetext	Wie hoch schätzen Sie den Anteil (in %) der <u>doppelt oder dreifach verglasten</u> Fenster in Ihrem Haushalt?	Please indicate the approximate proportion of windows in your home which are <u>double glazed</u> ?
Antwortmöglichkeiten	~ 0 %	None
	~ 25 %	About a quarter
	~ 50 %	About half
	~ 75 %	About three quarters
	~ 100 %	All
	weiss nicht	



Deskriptive Resultate

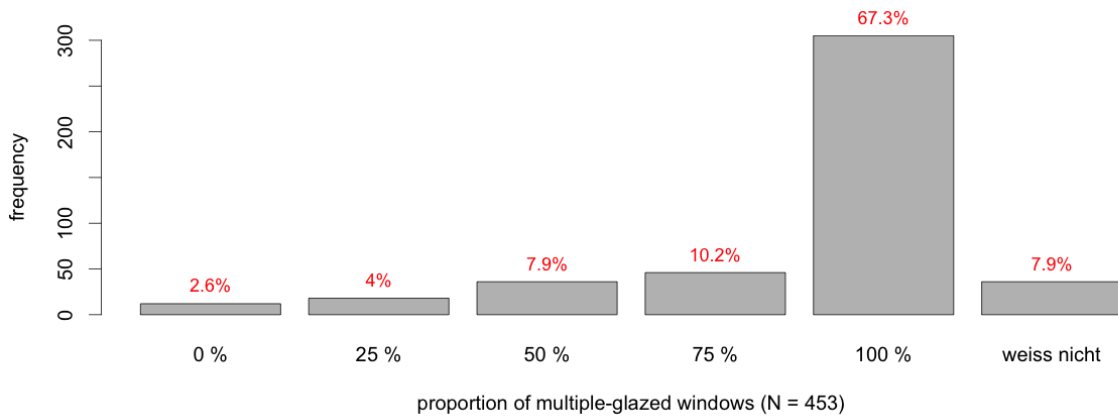


Abbildung A1: Verteilung der Antworten auf die Frage 23 in Arbon

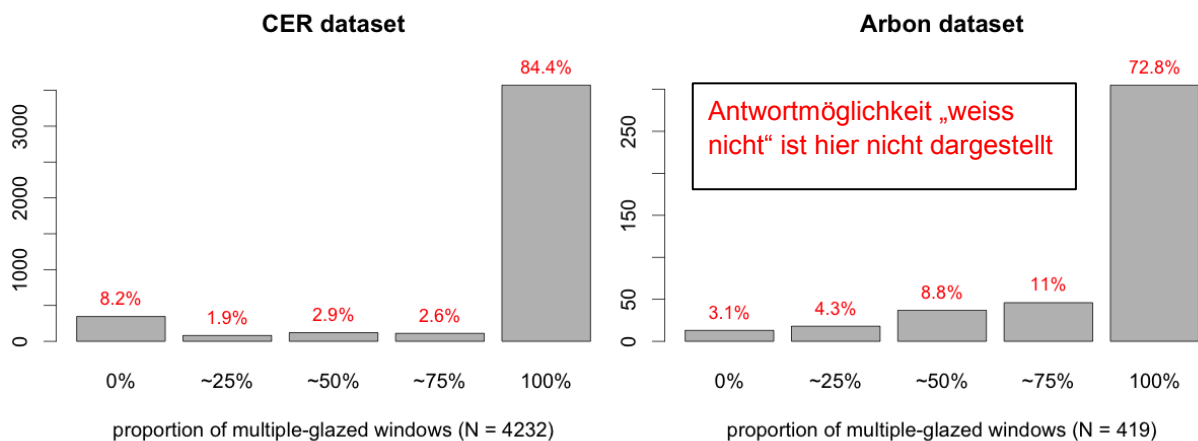


Abbildung A2 (Abb. 17 im Schlussbericht): Verteilung des Anteils an mehrfachverglasten Fenstern in beiden Umfragen im Vergleich

Durchschnittlicher Anteil der mehrfachverglasten Fenster in allen Haushalten:

- Arbon: 86.8%
- Irland: 88.3%

Detailliertere deskriptive Resultate

Tabelle A6 vergleicht den Anteil der mehrfachverglasten Fenster für unterschiedliche Haushaltsklassen. Der Anteil der Haushalte mit einer kompletten Ausstattung an mehrfach verglasten Fenstern ist in Irland höher (84 % vs. 73 % in Arbon), aber der Anteil der Häuser komplett ohne mehrfach verglaste Fenstern ist geringer (2.9 % in Arbon, 8.1 % in Irland). Im Vergleich der beiden Umfragen lässt sich feststellen, dass der durchschnittliche Anteil der mehrfachverglasten Fenster im irischen Datensatz mit 88.3% leicht höher ist, als im schweizerischen Datensatz mit 86.8%. Ein statistisch signifikanter Unterschied kann jedoch nicht nachgewiesen werden. Der leicht höhere Anteil der mehrfachverglasten



Fenster im Durchschnitt ist auch vorhanden, wenn man die Haushalte nach den Gebäudeeigenschaften Hausart und Baujahr separiert. Bei Aufteilung nach Wohnungen im Eigentum und Wohnungen zur Miete ist aber der Anteil der mehrfachverglasten Fenster im schweizerischen Datensatz grösser.

Tabelle A6: Gegenüberstellung des relativen Anteils an mehrfach verglasten Fenstern in den beiden Umfragen in unterschiedlichen Kategorien

	Arbon	CER
Durchschnittlicher Anteil der mehrfachverglasten Fenster in allen Haushalten	86.8%	88.3%
→ Bei Wohnungen/Häuser im Eigentum	89.9 %	88.9 %
→ Bei Wohnungen/Häuser zur Miete	82.4 %	81.0 %
→ Bei Einfamilienhäusern	87.6 %	88.3 %
→ Bei Mehrfamilienhäusern	86.2 %	86.6 %
→ Bei Häusern vor 1980 gebaut	81.9 %	82.9 %
→ Bei Häusern nach 1980 gebaut	92.1 %	95.7 %
Anteil der Haushalte mit 100% mehrfachverglasten Fenstern	73.1 %	84.4 %
Anteil der Haushalte mit 0% mehrfachverglasten Fenstern	2.9 %	8.1 %

Vergleich der Umfrageergebnisse mit Statistiken

Wir vergleichen nun die Ergebnisse der Umfragen mit veröffentlichten Statistiken und Studien:

Für Irland wurde von Brophy et al. (1999, S.13) ein Anteil der mehrfachverglasten Fenster von 36 % dokumentiert. Laut der Autoren war dies zum Zeitpunkt der Studie einer der schlechtesten Werte in Nordeuropa. Allerdings weisen die Autoren darauf hin, dass 50% der noch nicht mehrfachverglasten Fenster in einem Zehnjahreszeitraum durch neue Fenster ersetzt werden können, sodass im Jahr 2009 (zum Zeitpunkt der irischen CER Studie) ein Anteil von 66% denkbar wäre. Watson & Williams (2003, S. IX) dokumentieren allerdings bereits einen Anteil von 92% für Häuser nach 1990 gebaut wurden und mehr als 51% über alle Häuser in Irland. Für Gewerbeimmobilien wurde von der Sustainable Energy Authority Ireland (2015, S. 7, 50) ein Anteil von ca. 50-60% Mehrfachverglasung festgestellt. Die Ergebnisse der CER Umfrage liegen somit weit über den sonst dokumentierten Anteilen an mehrfachverglasten Fenstern.

Zum tatsächlichen Einsatz an mehrfachverglasten Fenstern **in der Schweiz** konnten wir keine Statistiken oder Studien finden. In Ermangelung solcher Erhebungen für die Schweiz, können wir die in Arbon erhobenen Zahlen nicht vergleichen, müssen aber annehmen, dass die Zahlen einen Hinweis auf die tatsächlichen Verhältnisse in der Schweiz geben.

Vergleich der Umfrageergebnisse zu übrigen Gebäude-Eigenschaften mit Statistiken

Zunächst betrachten wir die Frage, ob die Antworten bzgl. der Gebäudeeigenschaften in der schweizerischen Umfrage die statistischen Erhebungen des Kantons Thurgau bzw. die Statistiken der Gesamtschweiz widerspiegeln. In Tabelle A7 haben wir die Statistiken für die Schweiz (aus dem Jahr



2014) den relativen Häufigkeiten der Gebäudeeigenschaften des in Arbon erhobenen Datensatz gegenübergestellt. Laut dieser Umfrage ist der Anteil der Wohnungen und Wohnhäuser im Eigentum, der Anteil der neuen Häuser (nach 1980 gebaut) und der Anteil der Einfamilienhäuser leicht höher als im übrigen Gebiet der Schweiz. Aus diesem Grund ist wahrscheinlich auch die durchschnittliche Wohnungsgrösse höher als in der übrigen Schweiz. Im grossen und ganzen spiegeln die Antworten der Umfrage aber die Verteilungen in der Schweiz wider.

Ausserdem zeigen wir mit diesem Vergleich, dass die Umfrage in Irland nicht repräsentativ ist, insb. in Bezug auf die durchschnittliche Haushaltsgrösse, Eigentumsverhältnisse (Miete/Eigentum), das Hausalter sowie den Anteil der Häuser/Wohnungen. Dieser Umstand ist in Tabelle A7 dargestellt (zwei Spalten rechts). Die relativen Häufigkeiten der Gebäudeeigenschaften im irischen Datensatz weichen im Anteil des Wohneigentums stark ab. Von daher können die Ergebnisse basierend auf der irischen Umfrage nicht einfach übertragen werden.

Tabelle A7: Vergleich der Umfragen mit den statistischen Erhebungen in der Schweiz und in Irland

	Arbon 2015	Thurgau 2014	Schweiz 2014	CER 2009	Irland
Wohnungseigentumsquote (Anteil der Wohnungen im Eigentum)	59.24%	Ca. 47 % ¹	Ca. 38% ¹ 44 % ⁴	93 %	Ca. 70 % ⁴
Gebaut nach 1980	48.44%	42 % ³	36% ²	42.23%	52.5% ⁶
Anteil der Einfamilienhäuser	41.25%	Ca. 38 % ¹	Ca. 27 % ¹	98%	64% ⁵
Durchschnittliche Wohnungs- grösse in m ²	133.6	111 ⁷	98 ⁷	206.5	104 ⁶

1) BFS (2015). Statistik „Bau- und Wohnungswesen 2014“ (908-1400)

2) BFS (2014). Wohngebäude nach Bauperiode und Anzahl Wohnungen im Gebäude

3) BFS (2014). Gebäude nach Kategorie, Bauperiode und Kantonen

4) Eurostat (2016). „Distribution of population by tenure status, type of household and income group“ (source: SILC) [filc_lvho02], Zahl für das Jahr 2009

5) Central Statistical Office (2009). „Survey on Income and Living Conditions (SILC), Housing Module 2007“, heruntergeladen von: <http://www.cso.ie/en/silc/releasesandpublications/> am 24.05.2016

6) Dol, K. & Haffner, M. (2010) „Housing Statistics in the European Union 2010“, OTB Research Institute for the Built Environment, Delft University of Technology für The Hague: Ministry of the Interior and Kingdom Relations, S. 51, 54

7) BFS (2016). „Gebäude- und Wohnungserhebungen 1970, 1980, 1990 und 2000“, Datenwürfelbeschreibung «Gebäude- und Wohnungszählung 1970-2000»:

Schlussfolgerung

Wir haben die Ergebnisse der Befragung zum Anteil an mehrfachverglasten Fenstern in Arbon und Irland analysiert. Wir kommen **auf Grundlage der Umfragedaten zu dem Schluss, dass in der schweizerischen Stichprobe aus dem Jahr 2015 ein geringerer Anteil an mehrfachverglasten**



Fenstern von den Befragten angegeben wurde, als in der irischen Stichprobe aus dem Jahr 2009. Allerdings ist dieser Unterschied klein und nicht statistisch signifikant.

Mögliche Erklärungen

Unterschiedliche Auswahlmethoden der Umfrageteilnehmer

Bei der Irischen Studie wurde darauf geachtet, eine möglichst repräsentative Stichprobe in Bezug auf verfügbare demographische Merkmale, Verhalten und Stromverbrauchsprofilen auszuwählen. Dies geschah vor dem Hintergrund, dass in vorhergehenden Smart-Meter-Studien besonders gebildete Bevölkerungsschichten teilgenommen haben. Dieser Bias sollte vermieden werden. Da bei der Auswahl der Teilnehmer die Eigenschaften von Häusern nicht berücksichtigt wurden, können die Ergebnisse dieser Umfrage nur bedingt mit den Ergebnissen aus Arbon verglichen werden.

Unterschiedliche Durchführung der Umfrage

Die Umfrage in Irland wurde per Telefon durchgeführt, in Arbon online. Es kann sein, dass z.B. ältere Personen, die mehr Finanzkraft haben um Fenstersanierungen durchzuführen, nicht an der Umfrage in Arbon teilgenommen haben.

Unterschiedliche Sprachen

Die Umfrage in Irland wurde auf Englisch durchgeführt, und in Arbon auf Deutsch. Es kann sein, dass es sich dabei um das unterschiedliche Verständnis / Interpretation der Fragen und Antwortmöglichkeiten bei den Kunden handelt.

Fehlende „weiss nicht“ Option in der irischen Umfrage

In den Antworten der irischen Umfrage gibt es keine fehlenden Werte und auch keine Antwortoption „weiss nicht“. Deshalb ist es möglich, dass Umfrageteilnehmer falsche Angaben gemacht haben. In der Arbon-Umfrage ist der Anteil der „weiss nicht“ Antworten 7%.

Verständnisprobleme beim Fragetext

Weiterhin kann es sein, dass die Teilnehmer in beiden Umfragen den Begriff „doppelt oder mehrfachverglaste Fenster“ bzw. „double glazed windows“ nicht korrekt verstehen, da z.B. ein Fenster wie in Abbildung A3 dargestellt ist als „double hung window“ bezeichnet wird. Solche Fenster sind in Irland verbreitet.

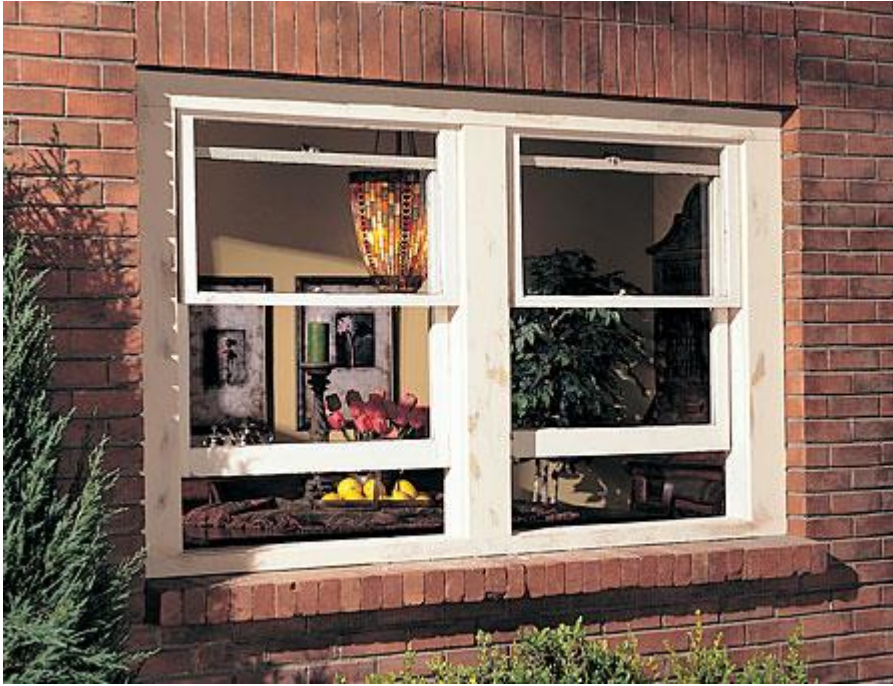


Abbildung A3: Beispiel für ein *double hung window*, wie es in Nordamerika und Grossbritannien eingebaut wird
(Quelle: <http://www.vizionwindows.com/images/nuevas/double-hung-window-02.jpg>)



Literaturverweise zu Anhang 4

Brophy, V. Clinch, J.P., Convery, F.J., Healy, J.D., King, C. & Lewis, J.O. (1999). „Homes for the 21st Century – the Costs and Benefits of Comfortable Housing for Ireland“, report prepared for Energy Action Ltd, by the Energy Research Group and Environmental Institute, University College Dublin, Dublin: Energy Action Ltd.

Commission for Energy Regulation. (2011). Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report (Information Paper No. CER11080a) (S. 146). Abgerufen von [http://www.ucd.ie/issda/data/commissionforenergyregulation/cer/](http://www.ucd.ie/issda/data/commissionforenergyregulation/cer/am) am 29.03.2016

Watson, D. & Williams, J. (2003) Irish National Survey of Housing Quality 2001-2002, The Economic and Social Research Institute

Sustainable Energy Authority Ireland (2015) „Extensive survey of the commercial buildings stock in the Republic of Ireland“, Insights Paper November 2015