



Schlussbericht vom 19. März 2020

SmartLoad – Smart Meter Data Analytics for Enhanced Energy Efficiency in the Residen- tial Sector



Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Datum: 19. März 2020

Ort: Rathausen, Zürich und Bamberg

Subventionsgeberin:

Bundesamt für Energie BFE
Sektion Energieforschung und Cleantech
CH-3003 Bern
www.bfe.admin.ch

Ko-Finanzierung:

Bundesministerium für Wirtschaft und Energie
Projektträger Jülich, Wilhelm-Johnen-Straße, D-52428 Jülich

Subventionsempfänger/innen:

Centralschweizerische Kraftwerke AG
Rathausen 1, CH-6032 Emmen
www.ckw.ch



BEN Energy AG
Badenerstrasse 60, CH-8004 Zürich
www.ben-energy.com



Otto-Friedrich-Universität Bamberg
Lehrstuhl für Wirtschaftsinformatik, insb. Energieeffiziente Systeme
Kapuzinerstrasse 16, D-96047 Bamberg
www.uni-bamberg.de/eesys



Autor/in:

Andreas Weigert, Otto-Friedrich-Universität Bamberg, andreas.weigert@uni-bamberg.de
Konstantin Hopf, Otto-Friedrich-Universität Bamberg, konstantin.hopf@uni-bamberg.de
Thorsten Staake, Otto-Friedrich-Universität Bamberg, thorsten.staake@uni-bamberg.de
André Rast, Centralschweizerische Kraftwerke AG, andre.rast@ckw.ch
Jan Marckhoff, BEN Energy AG, jan.marckhoff@ben-energy.com

BFE-Projektbegleitung:

Roland Brüniger, roland.brueiniger@r-brueniger-ag.ch

BFE-Vertragsnummer: SI/501521-01

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



Zusammenfassung

Im Zuge der Digitalisierung erhoffen sich Unternehmen Wettbewerbsvorteile durch die Entwicklung datenbasierter Dienstleistungen und die Steigerung der Vertriebseffizienz. Dies gilt auch für Energieversorgungsunternehmen (EVUs), die, wenn auch in einem stark regulierten Umfeld, eine Transformation vom Versorger zum Dienstleister anstreben und zugleich eine Schlüsselrolle beim Austausch fossiler Energieträger durch Elektrizität aus erneuerbaren Quellen einnehmen. Ausgehend von den in absehbarer Zeit flächendeckend zur Verfügung stehenden Verbrauch- und Transaktionsdaten planen EVUs, auf ihre Kunden massgeschneiderte Produkte und Dienstleistungen zu entwickeln. Um den Wert dieser Daten zu heben, ist jedoch der Einsatz moderner Analysemethoden – zum Beispiel Verfahren des maschinellen Lernens – erforderlich. In den letzten Jahrzehnten wurden zwar eine Reihe von generellen Datenanalyseverfahren entwickelt, aber es fehlt nach wie vor (insbesondere in Unternehmen) das Verständnis dafür, wie die Analytik auf konkrete Anwendungen der Energiedatenanalyse anzuwenden ist. Das Projekt «SmartLoad» trug dazu bei, diese Lücke zu schliessen und hat untersucht, inwiefern verfügbare Daten bei Energieanbietern genutzt werden können, um Haushalte zu erkennen, die a) bereit sind, nachhaltige Stromprodukte zu nutzen und die sich b) für die Investition in Selbstversorgungs- und Speicherlösungen interessieren. Darüber hinaus hat das Projekt untersucht, wie c) Smart-Meter-Daten das Stromverbrauchs-Feedback unterstützen können, um die Energieeffizienz im Privatsektor zu erhöhen.

Vor diesem Hintergrund hat die Projektgruppe detaillierte Daten von insgesamt 1'373 privaten Haushalten erhoben, mit diesen Daten maschinellen Lernverfahren trainiert und Vorhersagemodelle für das Kaufinteresse für lokalen Ökostrom, die Investition in Photovoltaikanlagen und die Anschaffung von Wärmepumpen entwickelt. Das Modell für den Kauf von Ökostrom wurde in einem Experiment mit 4'712 Privatkunden evaluiert. Die Modelle zur Prädiktion der Investitionsbereitschaft in Photovoltaikanlagen und Wärmepumpen ermöglichten die Entwicklung eines Softwareprototypen, der den Vertrieb der Technologien unterstützt.

Die Erkenntnisse des Projekts liefern Energieanbietern wichtige Hinweise, unter anderen für welche Aufgaben des Vertriebs von nachhaltigen Produkten maschinelle Lernverfahren eingesetzt werden können, welche externen Datenquellen die Kaufvorhersagen verbessern, wie geeignete Trainingsdaten für die Modellentwicklung gesammelt werden sollten und wie die entwickelten Vorhersagemodelle in eine Anwendung für den Vertriebsmitarbeiter überführt werden können.

Résumé

Dans le cadre de la transformation digitale, les entreprises espèrent obtenir des avantages concurrentiels en développant des services basés sur les données et en augmentant l'efficacité des ventes. Cela s'applique également aux entreprises d'approvisionnement en électricité (EAE) qui, bien que dans un environnement très réglementé, cherchent à se transformer de fournisseur électrique en fournisseur de services. En même temps elles jouent un rôle clé dans le remplacement de l'énergie fossile par de l'énergie électrique de sources renouvelables. Sur la base des données de consommation et de transaction prochainement disponibles à l'échelle nationale, les EAE prévoient développer des produits et services sur mesure pour leurs clients. Pour augmenter la valeur de ces données, l'utilisation de méthodes analytiques modernes est nécessaire, par exemple des techniques d'apprentissage automatique. Bien qu'un certain nombre de méthodes générales d'analyse des données aient été développées au cours des dernières décennies, il y a encore un manque de compréhension concernant les applications spécifiques possibles avec l'analyse des données énergétiques (en particulier dans les entreprises). Le projet "SmartLoad" a contribué à combler cette lacune en examinant comment les données disponibles peuvent être utilisées par les fournisseurs d'énergie pour identifier les ménages qui a) sont prêts à utiliser des produits d'électricité renouvelables et b) s'intéressent à investir dans



des solutions d'autosuffisance et de stockage. En outre, le projet a examiné c) comment les données des smart meters peuvent soutenir la visualisation de la consommation d'énergie afin d'accroître l'efficacité énergétique dans le secteur privé.

Dans ce contexte, l'équipe de projet a recueilli des données détaillées auprès d'un total de 1'373 ménages privés. Avec ces données elle a créé des méthodes d'apprentissage automatique et a développé des modèles prédictifs concernant l'intérêt pour l'achat d'électricité écologique locale, l'investissement dans les systèmes photovoltaïques et l'achat de pompes à chaleur. Le modèle pour l'achat d'électricité écologique a été évalué lors d'une expérience menée auprès de 4'712 clients privés. Les modèles de prévision de la volonté d'investissement dans un système photovoltaïque et dans une pompe à chaleur ont permis de développer un logiciel prototype, qui soutient la distribution de ces technologies.

Les résultats du projet livrent d'importants indices aux fournisseurs d'énergie pour identifier les tâches de vente de produit renouvelable, qui peuvent être améliorés avec l'apprentissage automatique, pour identifier les sources de données apte à augmenter les prévisions d'achat, pour définir les données d'apprentissage nécessaire au développement de modèles et pour créer des applications de vente basées sur les modèles de prédiction développés.

Summary

In the course of digitization, companies hope to gain competitive advantage through the development of data-based services and an increase in sales efficiency. This also applies to energy retailers which strive—in a highly regulated environment—for a transformation from utilities to service providers. At the same time, these firms play a key role in the energy transition while replacing fossil energy sources with renewable ones. Energy utilities are planning to develop products and services tailored to the needs of their customers, based on comprehensive data (e.g., from electricity consumption and digital business processes) that will be available nationwide in the near future. In order to increase the value of this data, the use of advanced analytical methods—for example machine learning—is necessary. Although a number of generic data analytic techniques have been developed in recent decades, there is still a lack of understanding (especially in firms), how to apply these methods to specific applications of energy data. The "SmartLoad" project helped to close this gap and as it investigated to what extent available data can be used by retailers in the energy industry to identify households that a) are willing to use sustainable electricity products and b) are interested in investing in self-sufficiency and storage solutions. In addition, the project investigated how c) smart meter data can support electricity consumption feedback to increase energy efficiency in the private sector.

The project consortium has collected detailed data for 1,373 private households, trained machine learning methods with these data and developed prediction models for the purchase interest in a locally produced green electricity tariff, the investment in photovoltaic systems and the purchase of heat pumps. The model for the purchase of green electricity was evaluated in an experiment with 4,712 private customers. The models for predicting the willingness to invest in photovoltaic systems and heat pumps enabled the development of a software prototype that supports the sales of the respective technologies.

Findings of the project provide energy suppliers with important lessons on, amongst other aspects, which tasks in the sales of sustainable products can be supported by machine learning methods, which external data sources improve purchase predictions, how training data should be collected effectively for model development, and how the developed prediction models can be integrated in an application for different stakeholders in the sales process.



Take-home messages

- Unternehmen der Energiebranche besitzen umfangreiche Daten, die für fortgeschrittene Datenanalysen nutzbar sind. Hierbei können Unternehmen im regulierten Bereich bereits solche Daten nutzen, die ihnen ohne rechtliche Hürden zur Verfügung stehen (z. B. frühere Tarifwahl oder Kaufentscheidung von Kunden). Die Einwilligung von Kunden für die Nutzung von Daten aus dem wettbewerblich regulierten Bereich (z. B. Smart-Meter-Daten) kann bei Darstellung des Mehrwerts für den Kunden eingeholt werden.
- Die Verbreitung von nachhaltigen Produkten, wie Ökostromtarifen, Photovoltaik- oder Energiespeicheranlagen kann durch den Einsatz fortgeschrittener Datenanalysen und verfügbaren Daten bei Unternehmen effektiv unterstützt werden.
- Für die Vorhersage von zukünftigen Kundenverhalten sind Daten aus Experimenten oder verganginem Kundenverhalten hilfreich; Kundenumfragen (insb. angegebene Intentionen für Handlungen) sind hingegen für solche Vorhersagen nur bedingt aussagekräftig.
- Für den Erfolg von Vorhersagemodellen auf Basis von maschinellen Lernverfahren ist die Definition des Vorhersageziels, die Identifikation und Vorbereitung der Eingabedaten sowie die Einbettung in Geschäftsprozesse bzw. die Kundenkommunikation einen weitaus grösseren Hebel als die Feineinstellung der Algorithmen mit der Wahl der besten Parameter. Letzteres kann den Einsatz dann verbessern, wenn die vorgenannten Arbeitsschritte erfolgreich durchlaufen sind.



Inhaltsverzeichnis

Zusammenfassung	3
Résumé	3
Summary	4
Take-home messages	5
Inhaltsverzeichnis	6
Abkürzungsverzeichnis	8
1 Einleitung	9
2 Teilprojekt 1: Vertrieb von lokal produziertem Ökostrom	11
2.1 Ausgangslage und Teilziele.....	12
2.2 Kundensegmentierung für den Vertrieb lokaler Ökostromprodukte.....	13
2.2.1 Datenerhebung.....	13
2.2.2 Ergebnisse Kundensegmentierung	14
2.3 Prädiktive Vorhersagemodelle für den Vertrieb von lokalen Ökostromprodukten	16
2.3.1 Konzept zur Entwicklung von Vorhersagemodellen	16
2.3.2 Feldexperiment zur Evaluation von Vorhersagemodellen.....	17
2.3.3 Simulationsexperiment zur Evaluation von geeigneten Ground-Truth-Daten.....	19
2.4 Diskussion der Ergebnisse und Implikationen.....	23
2.4.1 Herausforderungen bei der Entwicklung von Vorhersagemodellen.....	23
2.4.2 Theoretische Implikationen.....	25
2.4.3 Praktische Implikationen.....	25
3 Teilprojekt 2: Vertrieb von Photovoltaik- und Energiespeicheranlagen	25
3.1 Ausgangslage und Teilziele.....	26
3.2 Erhebung von Daten zum Training von Vorhersagemodellen	26
3.3 Konzeptionierung des Vorhersagesystems.....	27
3.4 Modellbildung und Evaluation der Vorhersagen.....	28
3.5 Prototypische Umsetzung des Entscheidungsunterstützungssystems	30
3.6 Diskussion der Ergebnisse	33
4 Teilprojekt 3: Datenanalysen für automatisiertes Energieeffizienz-Feedback	34
4.1 Schätzung von Stand-by-Verbrauch und Grundlast auf Basis von Smart-Meter-Daten	34
4.2 Auswertung von Energieeffizienz-Eigenschaften der Haushalte	35
4.3 Experiment zum Langzeiteffekt von Energieeffizienz-Mailings	38
4.3.1 Status quo Energieeffizienz-Mailings	39
4.3.2 Ergebnisse der Langzeitstudie	40
5 Rechtliche Rahmenbedingungen	41
5.1 Datenschutz.....	41
5.2 Unbundling.....	41



6	Schlussfolgerungen und Fazit	41
7	Ausblick und zukünftige Umsetzung	42
8	Nationale und internationale Zusammenarbeit	43
8.1	Nationale Zusammenarbeit	43
8.2	Internationale Zusammenarbeit.....	43
9	Publikationen	43
10	Literaturverzeichnis	44



Abkürzungsverzeichnis

AUC	Area Under The Curve
BFE	Bundesamt für Energie
CSV	Comma-separated values
CKW	Centralschweizerische Kraftwerke AG
ERP	Enterprise-Resource-Planning
EVU	Energieversorgungsunternehmen
FPR	False Positive Rate
PV	Photovoltaik
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
WP	Work Package



1 Einleitung

Die Digitalisierung stellt Energieunternehmen vor grosse Herausforderungen, schafft aber zugleich vielfältige Möglichkeiten, neue Dienstleistungen zu entwickeln und die Energieeffizienz privater Haushalte zu steigern. Ein Treiber dieser Entwicklung ist der Ausbau intelligenter Stromzähler, sogenannter Smart-Meter, sowie der vermehrte Einsatz vernetzter dezentraler Erzeugungstechnologien, die zunehmend Verbrauchs- und Nutzungsdaten verfügbar machen. Mithilfe moderner Analysemethoden, wie beispielsweise aus dem Bereich des maschinellen Lernens, können Energieanbieter auf dieser Datenbasis aufbauend massgeschneiderte Produkte und Dienstleistungen entwickeln. So lassen sich z. B. Haushalte, die in besonderem Masse von Selbstversorgungs- und Energiespeicherlösungen profitieren können, identifizieren, und Kunden mit einem hohen Potenzial für Energieeffizienzprogramme können gezielt angesprochen werden. Schlussendlich führen die bereits verfügbaren Daten in Kombination mit entsprechenden Analysemethoden zu mehr Energieeffizienz, vertieften Kundenbeziehungen, besserer Servicequalität. In liberalisierten Märkten unterstützt der Einsatz innovativer Analysemethoden darüber hinaus die Kundengewinnung und/oder -bindung.

Private Haushalte konsumieren 27% des Primärenergieverbrauchs in Europa (Eurostat 2017). Sie sind deshalb eine relevante Zielgruppe, um die Energieeffizienz zu steigern. Dabei variiert der Verbrauch der Haushalte allerdings stark. Dies ist vor allem auf Unterschiede in der Effizienz von Geräten (z. B. Wirkungsgrad der Heizung) und im Nutzungsverhalten (z. B. Wahl der Raumtemperatur) der Bewohner zurückzuführen. Diese beiden Bereiche bergen deshalb Potential für erheblichen Energieeinsparungen (Dietz u. a. 2009). Um die Energieeffizienz von Wohngebäuden zu steigern, ist es daher hilfreich, deren individuelle Eigenschaften (z. B. Grösse der Wohnung, Anzahl der Bewohner, Standby-Verbrauch, Grundlast, Beleuchtung) zu kennen. Mit diesen Informationen können Energieversorger oder Energieberater durch personalisierte Verbrauchsberichte, Spartipps und Lastverlagerungshinweise gezielt Anreize zur Effizienzsteigerung geben und massgeschneiderte Dienstleistungen und Produkte entwickeln.

Neben der Energieeffizienz gilt es, erneuerbare Energieträger in das Stromnetz zu integrieren und Ausgleichspotentiale für Last (z. B. durch elektrische oder thermische Speicher) zu schaffen. Dies kann durch die Entwicklung von neuen Energieprodukten durch Energieanbieter unterstützt werden. Um die Steigerung der Energieeffizienz im Privatsektor und die Integration von erneuerbaren Energieträgern sowie Speicherlösungen durch gezielte Kampagnen zu unterstützen, führten die Central-schweizerische Kraftwerke AG (Schweiz), die BEN Energy AG (Schweiz) sowie die Otto-Friedrich-Universität Bamberg (Deutschland) das Forschungsprojekt «SmartLoad» durch. Das Projekt war Teil der ERA-Net Smart Energy Systems Initiative der Europäischen Union und assoziierten Staaten und damit ein internationales Forschungsvorhaben.

Ziel war es, maschinelle Lernverfahren weiterzuentwickeln, die wir in einem früheren, durch das Bundesamt für Energie (BFE) geförderten, Projekt «Smart-Meter-Datenanalyse für automatisierte Energieberatungen» entwickelt haben. Die entwickelten Verfahren erlauben die Extraktion von 38 Haushaltsmerkmalen aus 15- bzw. 30-minütigen Smart-Meter-Lastprofilen sowie frei verfügbaren Daten (Sodenkamp u. a. 2016).

Aufbauend auf den früheren Ergebnissen wurden in diesem Forschungsprojekt folgende übergreifende Projektziele realisiert:

- **Technologieentwicklung:** Im Projekt nutzten wir maschinelle Lernverfahren für die Datenanalyse von Smart-Meter- und weiteren Daten, die beim Energieanbieter vorhanden sind, um Eigenschaften von Haushalten zu erlernen. Beispielsweise sollten Eigenschaften wie das Vorhandensein von hohem Grundlast-Stromverbrauch oder «Energiefressern» sowie Haushalte mit einem hohen Potenzial zum Kauf von Eigenversorgungs- und Speicherlösungen erkannt werden.



- **Validierung der Verfahren und Demonstration im Feld:** Wir erhoben Trainings- und Testdaten von Haushalten im Versorgungsgebiet des Industriepartners Centralschweizerische Kraftwerke AG (CKW AG) und testeten die Algorithmen in Demonstrationskampagnen.
- **Skalierbarkeit und Übertragbarkeit:** Die entwickelten Algorithmen ermöglichen die Verarbeitung der Daten in kurzer Zeit. Die Einbeziehung von Kontextdaten (z. B. öffentliche Statistikdaten und geographische Informationen) erlaubt eine hohe interregionale Transferfähigkeit der Modelle.
- **Replizierbarkeit:** Der Implementierungspartner BEN Energy AG betreut derzeit rund 30 Energiedienstleister in Europa. Die entwickelten algorithmischen Lösungen werden innerhalb der BEN-Tool-Suite implementiert und sind somit auch für andere Energieversorger verfügbar. Darüber hinaus ermöglicht die Verbreitung der Ergebnisse durch die ERA-Net Community und in Form von wissenschaftlichen Publikationen die Nutzung der Ergebnisse durch Analyseabteilungen anderer Versorgungsunternehmen.
- **Datenschutzkonformität:** Sowohl die entwickelten Verfahren als auch deren Anwendung stehen im Einklang mit der EU-Datenschutzgrundverordnung und den schweizerischen Datenschutzvorschriften.
- **Entflechtung und Kartellrecht:** Wir prüften, ob die im gegebenen Kontext gefundenen Ergebnisse unter Berücksichtigung der gesetzlichen Bestimmungen verwendbar sind. Insbesondere betrifft dies Daten aus dem Netzbetrieb (z. B. Smart-Meter-Daten) die nicht für Marktzwecke verwendet werden dürfen.

Das Projekt war Teil der ERA-Net Smart Energy Systems Initiative der Europäischen Union. In diesem Rahmen wurden Forschungsergebnisse mit anderen Forschergruppen ausgetauscht und Synergien genutzt, um den Projekterfolg zu unterstützen.

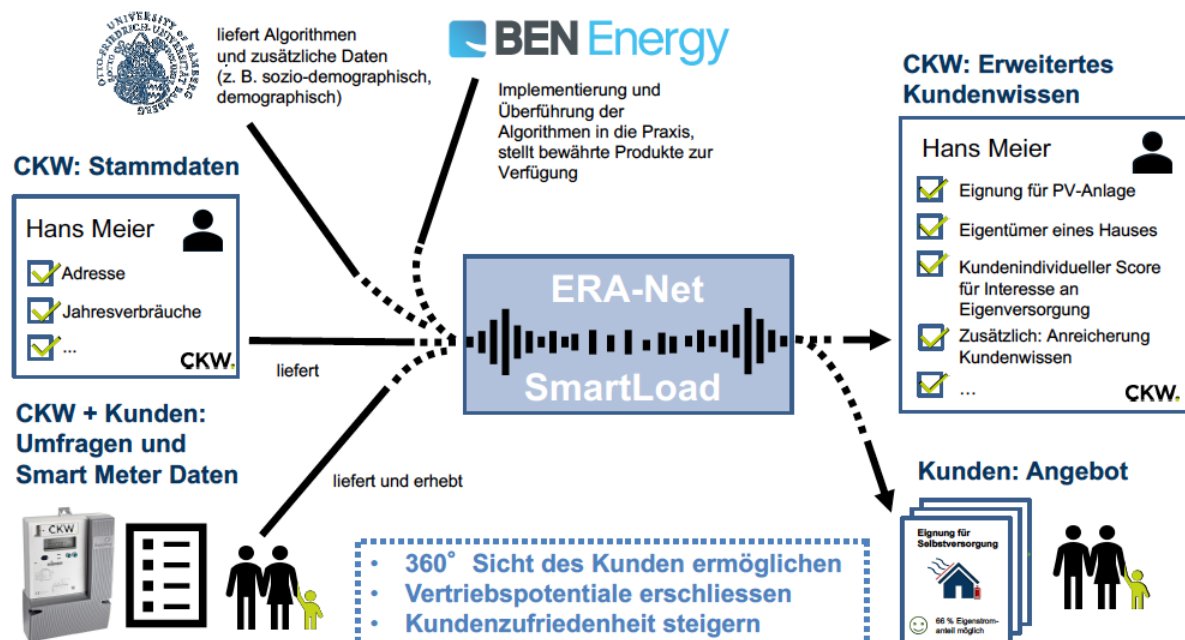


Abbildung 1: Projektvision ERA-Net SES «SmartLoad»

Die entwickelten Methoden können auf eine breite Palette von Anwendungen im Bereich des Vertriebs von nachhaltigen Energieprodukten angewendet werden und so im Privatbereich die Verbreitung von



erneuerbaren Energieträgern, die Energieeffizienz und nachhaltige Lösungen für Mobilität und Wärmeerzeugung verbessern. Aufgrund der Breite der Anwendungspalette wurden das Projekt in drei Teilprojekte aufgeteilt, um sowohl forschungsnahe als auch praxisnahe Lösungen zu entwickeln und diese im Feld zu testen:

1. **Vertrieb von lokal produziertem Ökostrom** – Unterstützung des Energievertriebs im Bereich lokal produzierten Ökostroms durch prädiktive Algorithmen
2. **Vertrieb von Photovoltaik- und Energiespeicheranlagen** – Identifizierung von Haushalten mit Interesse am Kauf von Photovoltaik- und Speicherlösungen
3. **Datenanalysen für automatisiertes Energieeffizienz-Feedback** – Haushaltsspezifisches Elektrizitätsfeedback

Diese zeitliche Einordnung der Anwendungsfelder im Projektplan ist zusammen mit Datenerhebungskampagnen und Arbeitspaketen (WP) in Abbildung 2 dargestellt.

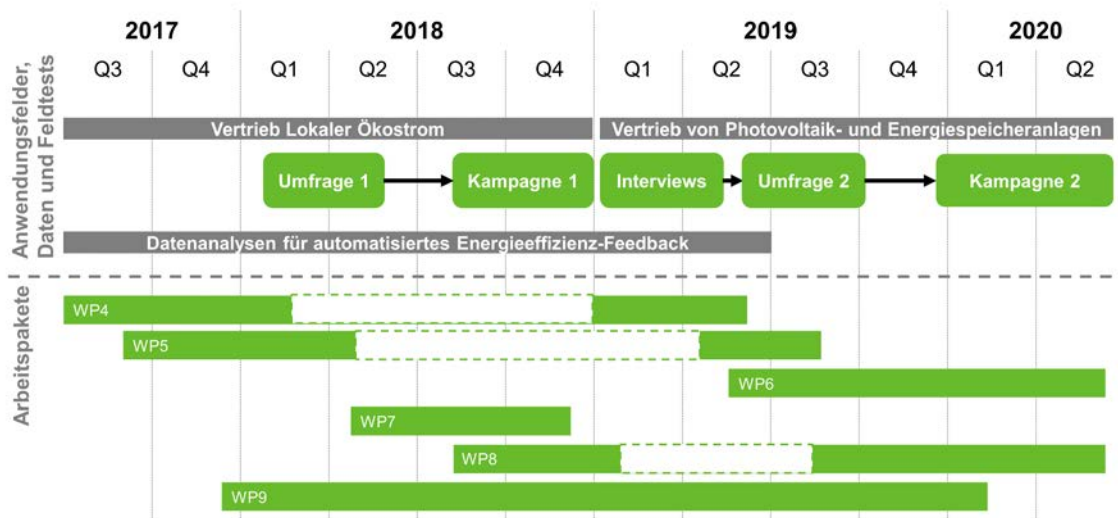


Abbildung 2: Projektplan mit Anwendungsfeldern und Datenerhebungszeitpunkten

2 Teilprojekt 1: Vertrieb von lokal produziertem Ökostrom

In diesem Teilprojekt untersuchten wir im Zeitraum 06/2017 bis 01/2019, inwiefern der Vertrieb von nachhaltigen Stromprodukten durch Datenanalysen und maschinelles Lernen unterstützt werden kann. Im Rahmen einer Kundenumfrage und -segmentierung erhoben wir im Vorfeld der technologischen Entwicklung von prädiktiven Modellen die Einstellung von Endkunden gegenüber lokalen Stromtarifangeboten. Ausserdem betrachteten wir die Einbettung von fortgeschrittenen Datenanalyseverfahren im organisationalen Kontext, da wir dies als kritischen Erfolgsfaktor für die Umsetzung erachteten. Nachfolgend beschreiben wir die beiden Teilziele des Teilprojekts sowie die jeweilige Forschungsmethodik. Am Schluss des Abschnitts diskutieren wir die Ergebnisse und leiten Implikationen für die Forschung und Praxis ab.



2.1 Ausgangslage und Teilziele

Das Energiegesetz vom 30. September 2016 (*Energiegesetz* 2018) hat u. a. zum Ziel, den «Übergang hin zu einer Energieversorgung, die stärker auf der Nutzung erneuerbarer Energien, insbesondere einheimischer erneuerbarer Energien» zu vollziehen. Eine Möglichkeit, dies von Energieversorgern umzusetzen stellt die Erhöhung des Anteils im Energiemix dar, aber auch die Steigerung der Nachfrage nach nachhaltig erzeugtem Strom bietet Potenzial. Der Privatkundenvertrieb von Stromtarifen mit einem hohen Anteil an neuen erneuerbaren Energieträgern ist allerdings herausfordernd, da solche Stromtarife als Nischenprodukte anzusehen sind. In 2018 lag der Lieferantenmix für gelieferten Strom aus Sonnenenergie, Windenergie, Biomasse und Geothermie in der Schweiz bei 2,25% («Stromkennzeichnung» 2018). Ausserdem bleibt der Markt zum Vertrieb solcher Produkte auf das jeweilige Versorgungsgebiet eines Energieversorgers beschränkt¹. Nur ein kleiner Anteil der Endverbraucher wählte in der Vergangenheit aktiv einen anderen Tarif als den vom Energieversorger angebotenen Standardtarif. Dem Marketing und dem Vertrieb von Energieversorgern kommt daher die besondere Aufgabe zu, Stromtarife mit erneuerbaren Energieträgern gesondert zu platzieren. Die aktive Bewerbung solcher Tarife birgt allerdings Risiken: Auf die Tarifauswahl aufmerksam gemacht, informieren sich Privatkonsumenten über die angebotenen Tarife. Statt für den nachhaltigen Tarif, entscheiden sich Kunden möglicherweise für einen günstigeren Tarif, der in der Regel ökologisch weniger wertvoll und einen höheren Anteil an Graustrom, oder Strom aus fossilen Quellen aufweist. Der Vertrieb von nachhaltigen Stromprodukten ist daher eine komplexe Aufgabe, bei welcher Methoden der Datenanalyse und maschinelles Lernen helfen können, gezielt Kunden mit Interesse an entsprechenden Stromtarifen anzusprechen.

In diesem Teilprojekt wurde deshalb untersucht, wie der Vertrieb von nachhaltigen Stromprodukten durch Datenanalysen und maschinelles Lernen unterstützt werden kann. Im Projekt wurde dabei eine Produktumstellung bei der CKW begleitet, bei der alle Stromtarife für Privatkunden auf neue Stromtarife umgestellt wurden, siehe Abbildung 1. Neben der strategischen Zielsetzung der Tarifumstellung, Kunden in ähnliche Tarife zu migrieren (schwarze Pfeile in Abbildung 1), war die Zielsetzung des Forschungsprojekts, Kunden des früheren Standardtarifs für den neu eingeführten lokalen Ökostromtarif zu identifizieren und in den neuen Tarif zu überführen (blauer Pfeil).

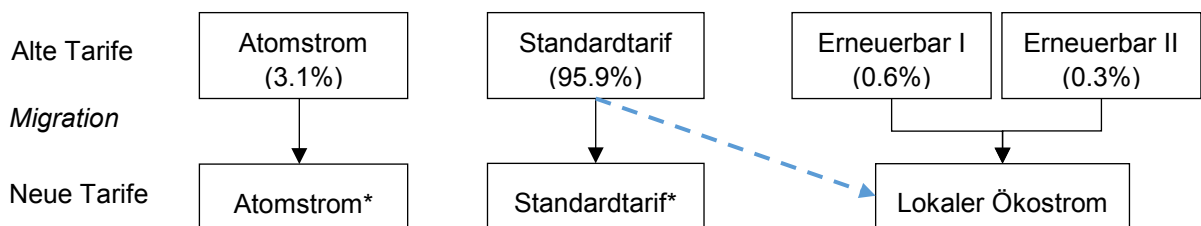


Abbildung 3: Alte und neue (*) Tarife und Standardmigrationswege für die Kunden während der Tarifänderung; Prozentsätze in Klammern: Verteilung der Kunden auf die alten Tarife

Konkret wurde dabei die Einführung eines neuen lokalen Ökostromtarifs «MeinRegioStrom» in zwei wesentlichen Schritten begleitet: Erstens wurde die Kaufbereitschaft für lokalen Ökostrom im Rahmen einer Kundenumfrage und einer anschliessenden Segmentierung mithilfe von unüberwachten maschinellen Lernverfahren untersucht, um Optionen zur Kundenansprache zu entwickeln. Zweitens wurden Kunden mit hohem Interesse am regionalen Ökostromprodukt mithilfe von überwachten maschinellen

¹ Dies gilt nicht für Verbraucher mit einem Stromverbrauch von mehr als 100'000 kWh.



Lernverfahren identifiziert und damit das Vertriebs-Targeting unterstützt. Wir haben das Vorhersagemodel in einem Feldexperiment als Upselling-Kampagne getestet. Die beiden Schritte des Teilprojekts werden nachfolgend im Detail ausgeführt.

2.2 Kundensegmentierung für den Vertrieb lokaler Ökostromprodukte

Der lokale Ökostromtarif «MeinRegioStrom» wurde mit der Tarifumstellung der CKW AG im Jahr 2019 neu eingeführt. Die bisherige Forschung zur Kaufbereitschaft für Ökostrom betrachtet vornehmlich nachhaltig erzeugten Strom, nicht aber den Aspekt der regionalen Erzeugung (Kaenzig und Wüstenhagen 2008; Tabi, Hille, und Wüstenhagen 2014a). Aus diesem Grund untersuchten wir im Projekt Einflussfaktoren der Kaufbereitschaft (wie persönliche Einstellungen oder sozioökonomische Faktoren). Hierzu führten wir eine umfangreiche Kundenumfrage sowie eine Kundensegmentierung durch, bei der neben den Umfragedaten auch verschiedene Datenquellen des Energieanbieters einbezogen wurden. Die Datenerhebung und Analyse mithilfe von Clustering-Verfahren wird nachfolgend beschrieben.

2.2.1 Datenerhebung

Verfügbare Daten stammen einerseits aus Informationssystemen der CKW AG, frei verfügbaren geographischen Informationssystemen und einer Kundenumfrage, die für diesen Projektabschnitt durchgeführt wurde. **Tabelle 1** zeigt einen Überblick über die genutzten Daten.

Tabelle 1: Datenquellen für das Teilprojekt Vertrieb von lokal produziertem Ökostrom

Datenquelle	Anzahl der Datenpunkte
Jahresstromverbräuche (intern)	1'100'000 Messungen des Jahresstromverbrauchs (Jahre 2012-2017)
Kundenkontaktdaten (intern)	1'208'582 Einträge aus der Kundenkontaktdatenbank (Jahre: 2010-2018)
Mahndaten (intern)	65'533 Einträge zwischen 2013-2018
Stamm- und Vertragsdaten (intern)	circa 200'000 Einträge
Geographische Daten aus OpenStreetMap (extern)	Daten aus einer Gesamtfläche von 2'750'000 Hektar
Umfrage (extern)	707 vollständige Antworten von Kunden in Standardtarif (bei einem Rücklauf von insgesamt 12,8%)

Daten aus internen Informationssystemen wurden von der CKW AG vor der Weitergabe pseudonymisiert, sodass kein Bezug zu natürlichen Personen möglich ist. Aus den Daten wurden Merkmale, sog. Features, abgeleitet.

Zur Vorbereitung der Umfrage haben wir den Stand der Forschung in Bezug auf die Kaufbereitschaft von Ökostromprodukten ermittelt, da die Kaufbereitschaft von lokal produziertem Strom noch in keiner Studie veröffentlicht wurde. Das tatsächliche Kaufverhalten von Ökostromprodukten wurde in Grossbritannien (MacPherson und Lange 2013) und in den Vereinigten Staaten (Kotchen und Moore 2007) untersucht. Die Kaufbereitschaft – erhoben durch Umfragen – wurde in Schweden (Hansla u. a. 2008) und in Deutschland (Tabi, Hille, und Wüstenhagen 2014b) untersucht. Aus den genannten Studien konnten wir sieben Faktoren identifizieren, die einen Einfluss auf die Kaufbereitschaft von Ökostromprodukten gezeigt haben: 1) Umwelteinstellung, 2) Altruismus, 3) politische Einstellung (insb. Unterstützung grüner und konservativer Partei), 4) Bildungsniveau, 5) Einkommen, 6) Alter und 7) Geschlecht. Allerdings wurden nicht alle Faktoren in allen Studien untersucht, und die Effekte variieren je



nach Erhebung. Für Einstellungen und Verhalten von Verbrauchern haben wir auf wissenschaftlich fundierte Messinstrumente zurückgegriffen. Das Einverständnis der Kunden zur Datennutzung und -verarbeitung wurde im Rahmen der Kundenumfrage eingeholt.

Im Februar 2018 wurden insgesamt 10'000 Kunden zur Umfrage mit personalisierten Mailings eingeladen. Für 3'690 Kunden (37%) war die E-Mail-Adresse bekannt und sie wurden per E-Mail eingeladen. Die übrigen 6'310 Kunden (63%) erhielten die Einladung per Brief. Von 10'000 Einladungen haben 1'281 Kunden die Umfrage ausgefüllt. Die Rücklaufquote der Umfrage lag insgesamt bei 12,8% und damit deutlich über den Erwartungen.

2.2.2 Ergebnisse Kundensegmentierung

Die Kundensegmentierung führten wir mit dem k-Means Algorithmus durch, einem partitionierenden Algorithmus, der ähnliche Kunden schrittweise gruppiert. Ziel ist es dabei, die Unterschiede der Kunden innerhalb eines Kundensegments zu minimieren und die Unterschiede der Kunden zwischen den Segmenten zu vergrössern. Für die Clusteranalyse wählten wir 18 Merkmale aus den oben beschriebenen Datenquellen, um die Produktmerkmale des neu eingeführten lokalen Ökostromtarifs (Regionalität, Erneuerbare Energieträger, Preis, Service) zu beschreiben.

Zur Abbildung von regionalen Aspekten erhoben wir in der Umfrage Einstellungen von Kunden in Bezug auf die Regionalität, darunter deren Affinität zu in der Region produziertem Strom, Regionalverbundenheit, Teilnahme an regionalen Aktivitäten, Anzahl Punkte in Geodaten (kann als Bebauungs- / Besiedlungsdichte interpretiert werden) sowie die Anzahl an Kunden in der Postleitzahl-Region. Das Produktmerkmal erneuerbare Energieträger wurde abgebildet mit der Wichtigkeit von erneuerbaren Energieträgern, Zustimmung zu altruistischen Verhalten (anderen helfen, Gemeinwohl), die Wahrnehmung zum eigenen Energiesparverhalten, sowie dem tatsächlichen Energiesparverhalten in Form des Stromverbrauchs pro Quadratmeter Wohnfläche. Das Merkmal Preis erhoben wir mittels soziodemographischer Angaben, wie dem monatlichen Nettohaushaltseinkommen, der Wohnart, Eigentumsverhältnisse, und der Bildung. Das Produktmerkmal Service wurde mit der Anzahl an Kontakten zum Energieversorger im Jahr 2017 sowie der Nutzung des Onlineportals erhoben. Die Zustimmung zum Kauf und die Zahlungsbereitschaft für konkrete Produkt erhoben wir, nachdem wir dem Umfrageteilnehmer zunächst die wesentlichen Produkteigenschaften präsentierten.

Für das gewählte Clustering-Verfahren muss die Anzahl an Segmenten im vornherein festgelegt werden. Nach mehreren Tests haben wir eine Anzahl von drei Segmenten (W1, W2, W3) als passend identifiziert. **Tabelle 2** zeigt das Ergebnis der Segmentierung. Für jedes einzelne Merkmal wird die Abweichung des jeweiligen Segments gegenüber dem Durchschnitt aller Kunden im Standardtarif gezeigt. Die Abweichung kann wie folgt interpretiert werden:

- ++ / -- starke positive / negative Abweichung, signifikant mit $\alpha < .01$
- +/- mittlerer oder schwache positive / negative Abweichung, signifikant mit $\alpha < 0,05$
- (+) / (-) geringfügige Abweichung, nicht statistisch signifikant
- Prozentzahlen in eckigen Klammern zeigen den Durchschnitt aller Kunden im Standardtarif

Im Folgenden wird das Segment W2 diskutiert, das hinsichtlich des lokalen Ökostromtarifs eine höhere Kauf- und Zahlungsbereitschaft im Vergleich zu den beiden Segmenten W1 und W3 aufweist. Hinsichtlich des Produktmerkmals Regionalität haben die Umfrageteilnehmer eine starke Ausprägung hin zu höherer Affinität zu in der Region produziertem Strom, gepaart mit einer höheren Regionalverbundenheit und Teilnahme an regionalen Aktivitäten. Umfrageteilnehmer in diesem Segment wohnen eher in gering bebauten Gegenden.



Produktmerkmale aus dem Bereich Erneuerbare Energieträger sind Umfrageteilnehmern im Segment W2 wichtiger und sie stimmen den Aussagen zum Energiesparverhalten und Altruismus (anderen helfen, Gemeinwohl) eher zu. Trotz positiver Angabe zum Energiesparverhalten weichen diese Haushalte nicht im Stromverbrauch pro m² vom Durchschnitt ab.

Hinsichtlich des Produktmerkmals Preis und Service zeigen sich keine statistisch signifikanten Abweichungen zum Durchschnitt (Variablen: Aspekten Einkommen, Wohnart und Bildung). Tendenziell leben die Umfrageteilnehmer dieses Segments aber eher im Haus statt in einer Wohnung, eher im Eigentum als zur Miete und in kleineren Ortschaften. Die Anzahl der Kundenkontakte zum Energieversorger ist im Wesentlichen gleich, wobei die Nutzung des Onlineportals etwas häufiger stattfindet.

Tabelle 2: Ergebnisse der Kundensegmentierung zur Personalisierung von Kundenansprachen

	Segment W1 n=192 (26,6%)	Segment W2 n=326 (45,2%)	Segment W3 n=204 (28,2%)
Regionalität			
F5: Affinität Regionalstrom	0	++	--
F3: Regio. Verbundenheit	--	++	0
F4: Regio. Aktivität	--	++	0
Anzahl Punkte in Geodaten	+	(-)	0
Kunden in der PLZ Region	(+)	0	(-)
Erneuerbare Energieträger			
F1: Erneuerbare Energieträger	+	+	--
F2: Anderen Helfen (Altruismus 1)	+	+	--
F7: Gemeinwohl (Altruismus 2)	0	+	--
F6: Energiesparverhalten	0	+	--
Stromverbrauch pro m ²	-	0	(-)
Preis			
Einkommen	(-)	0	0
Wohnung (statt Haus) [54,3%]	(+) [59,4%]	(-) [50,6%]	(0) [55,4%]
Wohnt im Eigentum [61%]	(-) [51,6%]	(+) [65,6%]	(0) [62,6%]
Bildung Hochschule / Universität [35,8%]	(-) [32,8%]	(-) [31,0%]	(0) [34,3%]
Service			
Anzahl Kundenkontakte Energieversorger (2017)	0	0	(-)
Nutzung Onlineportal [18%]	(-) [16,1%]	(+) [20,6%]	(-) [15,7%]
Lokaler Ökostrom			
Zustimmung Kauf lokaler Ökostromtarif	0	+	-
Zahlungsbereitschaft für lokalen Ökostromtarif [40 Rp./kWh]	(-) [32 Rp./kWh]	(+) [54 Rp./kWh]	(-) [23 Rp./kWh]

Zusammenfassend ermöglicht die Nutzung des k-Means Algorithmus die Ableitung von interpretierbaren Kundensegmenten. Die Eigenschaften des Kundensegments W2 erlaubten der CKW AG die Formulierung einer geeigneten Ansprache für eine konkrete Zielgruppe (siehe nächster Abschnitt). Die Erhebung von Produkteigenschaften in der Umfrage hat ausserdem dabei geholfen, einen potenziell erfolgswirksamen Namen «MeinRegioStrom» für das lokale Ökostromprodukt zu finden. Maschinelle



Lernverfahren, insbes. Clustering-Verfahren erlauben die Nutzbarmachung von Kundeneigenschaften und unterstützen damit die strategische Planung und Ableitung geeigneter Marketingmassnahmen.

2.3 Prädiktive Vorhersagemodelle für den Vertrieb von lokalen Ökostromprodukten

In diesem Abschnitt beschreiben wir, wie wir maschinelle Lernverfahren für ein Tarif-Upselling im Bereich lokaler Ökostrom genutzt und im Rahmen eines Feldexperiments validiert haben. Die Ergebnisse der Feldstudie sowie einer nachgelagerten Simulationsstudie erlauben generelle Empfehlungen für die Nutzung verschiedener Datenquellen für das Training prädiktiver Modelle und stellen einen nennenswerten Beitrag zur Forschung dar.

2.3.1 Konzept zur Entwicklung von Vorhersagemodellen

Für die Entwicklung von Vorhersagemodellen benötigt man zwei Arten von Daten: Daten mithilfe derer Vorhersagen gemacht werden (Features / Prädiktoren) und Ground-Truth-Daten zum Training der Modelle. Letztere repräsentieren die Zielklasse, also im Anwendungsfall die potenziellen Interessenten für einen lokalen Ökostromtarif. Zur Erhebung von Ground-Truth-Daten bei der Einführung von neuen Tarifen existieren mehrere Möglichkeiten, die wir im Gang der Untersuchung evaluiert haben:

1. Die Verwendung von historischen Verhaltensdaten, also Daten von Kunden, die bereits in der Vergangenheit einen zum lokalen Ökostromtarif ähnlichen Tarif gewechselt sind,
2. Umfragedaten von Kunden zur Kaufbereitschaft für einen lokalen Ökostromtarif,
3. Verhaltensdaten, die mit einem separaten Experiment (z. B. mit Testkunden) erhoben werden, also konkrete Wechseldaten von Kunden in einen lokalen Ökostromtarif.

Für das Feldexperiment haben wir uns – in Abstimmung mit allen Projektpartnern – für die zweite Variante entschieden. Gegen die Verwendung historischer Verhaltensdaten sprach, dass wesentliche Eigenschaften des neuen Stromtarifs, wie z. B. die Regionalität, die Serviceausgestaltung und der Preis neuartig waren und sich erheblich von den alten Stromtarifen unterschieden haben. Ein separates Experiment konnten wir aus zeitlichen Gründen nicht durchführen.

Features extrahierten wir aus den Datenquellen, die auch im Abschnitt Segmentierung verwendet wurden. **Abbildung 4** zeigt das Grobkonzept zum Vorgehen bei der Entwicklung von Vorhersagealgorithmen.

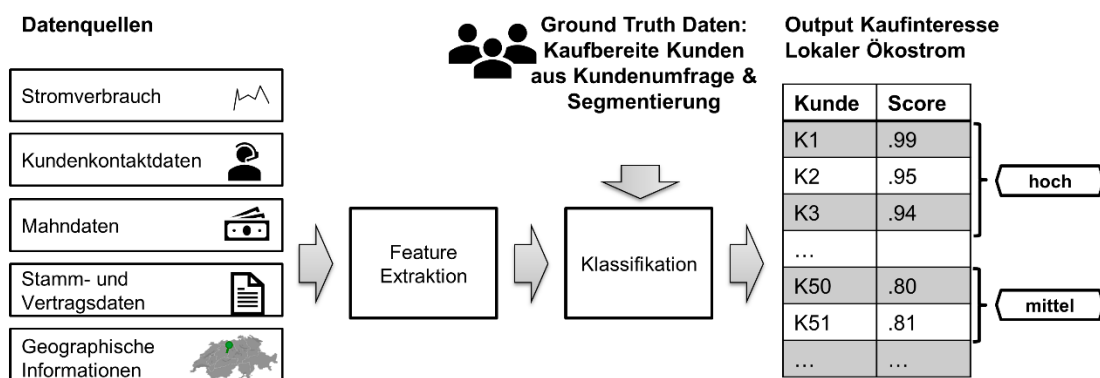


Abbildung 4: Grobkonzept zum Vorgehen zur Klassifikation der Haushalte in das Zielsegment



2.3.2 Feldexperiment zur Evaluation von Vorhersagemodellen

Ziel des Feldexperiments war es, Kunden mit einer hohen Kaufbereitschaft für das lokale Ökostromprodukt mithilfe der Vorhersagemodelle zu identifizieren und diese gezielt anzusprechen.

In Vorbereitung auf das Feldexperiment haben wir verschiedene maschinelle Lernverfahren mit den erhobenen Daten getestet, um die Kaufbereitschaft für lokal produzierten Ökostrom von Kunden im Standardtarif möglichst optimal vorherzusagen. Dazu berechneten wir aus den vom Energieversorger bereitgestellten Daten insgesamt 108 Features, testeten mehrere Feature-Teilmengen und Klassifikationsalgorithmen. Für das finale Vorhersagemodell wählten wir den Klassifikationsalgorithmus logistische Regression und wählten 17 Features: Kundenstammdaten (2), Stromverbräuche (2), Kundenkontaktdaten (3), Zahlungsdaten (3), Geoinformationsdaten (7). Das Vorhersagemodell wurde mittels eines sog. Bagging-Ansatzes (Ziehen ohne Zurücklegen) validiert. Auf Basis von 30 Bagging-Modellen wurden gemittelte Vorhersage-Scores berechnet. Unter Laborbedingungen sind die entwickelten Modelle deutlich besser als eine zufällige Vorhersage (Area Under The Curve (AUC) = 0,74).

Bei der Entwicklung eines zielgruppenorientierten Anschreibens hat die Marketingabteilung der CKW AG in Zusammenarbeit mit einer externen Agentur die genauen Produktbeschreibungen und die Ansprachen für die Kunden zu den jeweiligen Produkten entwickelt. Nach Aussage der CKW AG hat die im Projekt erstellte Segmentierung sehr zum Verständnis der Kundenbasis und zur Schärfung der Produkteigenschaften beigetragen. Beispielhaft zeigt **Abbildung 5** das E-Mail-Anschreiben, das für das Zielsegment «potenzieller Interessent lokaler Ökostrom» entwickelt und in der nachfolgend beschriebenen Kampagne verwendet wurde.

CKW. Kundennummer: 21.08.2019

„Wir machen Energie einfach“ – das ist unser Markenversprechen an Sie. Deshalb vereinfachen wir unsere Stromprodukte.

Ab dem 01. Januar 2019 gibt es nur noch drei unterschiedliche Stromprodukte. Aus Ihrem aktuellen Produkt „CKW WasserKraft“ wird nun „CKW ClassicStrom“. **Damit profitieren Sie von einer Preisreduktion in Höhe von circa 10 Prozent.**

Sie möchten selbstständig bestimmen woher Ihr Strom kommt und die Energiezukunft der Region nachhaltig gestalten? Dann haben wir genau das richtige Angebot für Sie

Unsere Empfehlung: Nutzen Sie die Gelegenheit mit „CKW MeinRegioStrom“ **ausschliesslich ökologischen Strom aus Zentralschweizer Sonnenenergie zu beziehen. So gestalten Sie die Energiezukunft aktiv mit. Zudem werden Sie Teil der „CKW MeinRegioStrom“ Community und haben die Chance, sich eine Reihe attraktiver Vorteile wie Tickets für Sport- und Kulturevents zu sichern.**

Haben wir Ihr Interesse geweckt? Dann bestellen Sie am besten gleich Ihr neues Produkt „CKW MeinRegioStrom“ auf unserem **neuen Kundenportal „MyCKW“**. Melden Sie sich mit Ihrem gewohnten Login-Daten an oder registrieren Sie sich mit Ihrem persönlichen Zugangscode XXXXX auf www.ckw.ch/myckw. Gerne dürfen Sie uns auch **unter 0800 88 77 66 anrufen.**

Wir danken für Ihr Vertrauen.

Was sich ändert.

Alternative.

Produkteigenschaften.

Call to Action: Registrieren auf neuem Portal oder telefonisch anrufen.

Abbildung 5: E-Mail-Anschreiben für das Zielsegment potenzieller Interessent lokaler Ökostrom

Da das Experiment in die Tarifumstellung der CKW AG eingebettet war, begann die Planung des Experiments bereits kurz nach Start des Forschungsprojekts in einem gemeinsamen Workshop im November 2017 und involvierte zahlreiche Abteilungen der CKW AG (z. B. Produktentwicklung, Netz, Recht, Marketing, Kundenkontaktcenter). Das Experiment stand unter der Bedingung, dass die Tarifumstellung insgesamt nicht negativ beeinflusst werden soll. So konnte eine Anzahl von 5'000 Kunden in das Experiment aufgenommen werden. Darüber hinaus war das Ziel der CKW AG Kundenkommunikation im Zeitraum der Tarifumstellung (dies betraf nicht nur die Ansprache der Kunden in Briefen



o. ä., sondern auch die Öffentlichkeitsarbeit, die Homepage, usw.) einen Wechsel der Kunden in günstigere Tarife zu vermeiden. Folglich sollte offensive Werbung für neue Tarife (insb. teurere) vermieden werden.

Von 1'281 Kunden, die an der ersten Kundenumfrage teilgenommen haben, waren 950 Kunden im Standardtarif, wovon 707 die Umfrage vollständig ausgefüllt haben. In dieser Umfrage wurde die Kaufintention für ein zu «MeinRegioStrom» ähnliches Produkt erhoben.

Zum Training zur Verfügung standen

- 69 Kunden mit einer hohen Kaufintention für Produkt A, also allen Kunden die mindestens eine leichte Zustimmung bei der Kaufintention angegeben haben.
- 638 Kunden mit einer niedrigen Kaufintention für Produkt A, also allen Kunden die unentschlossen waren oder keine Zustimmung bei der Kaufintention angegeben haben.

Das Vorhersagemodell wurde in einem Experiment mit 4'712 Kunden getestet (2'840 vom Algorithmus und 1'872 zufällig selektiert). Unser Feldexperiment fand gleichzeitig mit einer Umstellung aller Kunden auf die neue CKW AG Produktlandschaft statt, bei der circa 20 verschiedene Kundensegmente individualisierte Anschreiben erhielten. Aus diesen Kundensegmenten haben wir für unser Feldexperiment eine spezifische Zielgruppe ausgewählt, die homogen hinsichtlich des ursprünglichen Tariftyps und sonstiger Tarifspezifika war. Die Zielgruppe bestand ausschliesslich aus Kunden, die Privatkunden im Standardtarif waren, für die eine E-Mail-Adresse vorhanden und für das Experiment nutzbar war (Kunde hat Einverständnis gegeben bzgl. Produktinformationen kontaktiert zu werden) und die nicht bereits Teil der Umfrage waren. Aus dieser Zielgruppe haben wir zwei Gruppen für das Feldexperiment gezogen:

- **Kontrollgruppe:** 1'872 Kunden aus der Zielgruppe bei Sicherstellung der Randomisierung hinsichtlich der Variablen Postleitzahl-Region, Geburtsjahr des Tarifinhabers, Portalnutzung, Tariftyp, Stromverbrauch und einiger Features des Vorhersagemodells.
- **Selektion der Algorithmusgruppe:** 2'840 Kunden mit den höchsten Vorhersage-Scores (vgl. **Abbildung 4**)

Die Kunden wurden im August 2018 per E-Mail kontaktiert und über die Produktumstellung sowie über das neue lokale Ökostromprodukt «MeinRegioStrom» informiert. Anschliessend hatten die angesprochenen Kunden bis Mitte Oktober Zeit, auf das Anschreiben zu reagieren, bis die Daten ausgewertet wurden.

Insgesamt haben 25 von 2'840 Kunden (0,88%) aus der Algorithmus-Auswahl den Tarif gewechselt. Davon haben 8 ein Upgrade zu «MeinRegioStrom» durchgeführt. In der Kontrollgruppe haben 11 von 1'872 Kunden einen Tarifwechsel vollzogen und 5 ein Upgrade zu «MeinRegioStrom» durchgeführt. Aufgrund der sehr niedrigen Wechselraten lassen sich mit statistischen Tests keine statistischen Unterschiede zwischen Algorithmus-Auswahl und Kontrollgruppe erkennen. Es fällt jedoch auf, dass sich die Wechselraten (0,5-0,9%) der beiden Feldexperimentgruppen Algorithmus-Auswahl und Kontrollgruppe und den restlichen Standardtarifkunden (0,5%) ausserhalb der Kampagne kaum unterscheiden.



Tabelle 3: Ergebnisse der Vorhersagemodelle zum Tarifwechsel für das Feldexperiment

Gruppen	Hat gewechselt	Upgrade zu MeinRegioStrom
	Basis: Alle Kunden in Gruppe	Basis: Aktive Wechsler
Algorithmus-Auswahl	25/2'840 = 0,88 %	8/25 = 32,00 %
Kontrollgruppe	11/1'872 = 0,58 %	5/11 = 45,45 %
Standardtarifkunden ausserhalb der Kampagne	503/95'000 = 0,53 %	61/503 = 12,13 %

2.3.3 Simulationsexperiment zur Evaluation von geeigneten Ground-Truth-Daten

Eines der Kernergebnisse des im vorherigen Abschnitt beschriebenen Feldexperiments ist die Schwierigkeit, die mit Erhebung von geeigneten Ground-Truth-Daten für die Entwicklung eines Vorhersagemodells einhergeht. Für das beschriebene Feldexperiment wurde die Variable Kaufintention mittels Umfrage erhoben, bei der dem Kunden die wesentlichen Produkteigenschaften präsentiert wurden. Allerdings geht mit einer getätigten Kaufintention nicht zwangsläufig ein Kauf (Tarifabschluss bzw. Tarifwechsel) einher (Morwitz, Steckel, und Gupta 2007), weshalb diese Art von Ground-Truth-Daten ggf. nicht ideal sind.

Wir haben deshalb in dem nachfolgend beschriebenen Simulationsexperiment untersucht, welche Daten sich besonders für das Training von Vorhersagemodellen eignen, um Kundenverhalten vorherzusagen. Wir nutzten dabei die Situation bei der CKW AG, dass alle Kunden von einem Tarifsystem mit vier Basistarifen auf ein System mit drei Basistarifen umgestellt wurden (vgl. **Abbildung 3**). Im Zuge dieser Umstellung wurden alle Kunden durch die CKW AG im August 2018 angeschrieben und über die Tarifänderung informiert. Die Migration aller Kunden von der alten auf die neue Tarifregelung trat am 31. Dezember 2018 in Kraft. Nach der Ankündigung der Tarifänderung (per Brief oder E-Mail) konnten die Kunden entweder selbst aktiv einen Tarif wählen oder wurden durch die CKW AG automatisch migriert. Diese Wechselinformationen dienen uns für unser Simulationsexperiment zur Evaluation von geeigneten Ground-Truth-Daten als Datengrundlage. Dieses Experiment soll die Antwort auf die Frage liefern, welche Typen von Trainingsdaten am besten für das Training eines Modells geeignet sind.

Experimentelles Setup: Wir berücksichtigten im Experiment drei verschiedenen Trainingsdatentypen, die wir in dieser Analyse für je zwei Tarifwechselarten testen. Als Trainingsdatentypen wählten wir Historische Verhaltensdaten, Umfragedaten zur Kaufbereitschaft und Verhaltensdaten, die mit einem (hypothetischen) Experiment erhoben wurden. Als Tarifwechselarten wählten wir a) ob der Kunde den teureren Tarif ("Upgrade") auswählt, und b) ob der Kunde den günstigeren Tarif ("Downgrade") auswählt, nachdem er die Information zu den neuen Tarifen erhalten hat. Da die Mehrheit der Kunden (110'313 \approx 95,9%) vor der Tarifumstellung im Standardtarif war, berücksichtigen wir ausschliesslich deren Wechselverhalten für das Simulationsexperiment. In **Tabelle 4** beschreiben wir die Ableitung von Trainingsdaten für die drei Trainingsdatentypen und damit zur Erfassung der abhängigen Variablen. Aus den drei Trainingsdatentypen mit jeweils zwei abhängigen Variablen haben wir sechs Vorhersagemodelle entwickelt. Die wichtigsten Schritte des Modellierungsprozesses sind in **Abbildung 6** dargestellt. Zunächst haben wir Features aus den verschiedenen Datenquellen zusammengeführt. Als Lernalgorithmus haben wir Random Forest (Breiman 2001) gewählt, da es eines der leistungsfähigsten maschinellen Lernverfahren ist (Fernández-Delgado u. a. 2014), eine begrenzte Rechenkomplexität aufweist (insbesondere im Vergleich zu Deep-Learning-Ansätzen) und die Fähigkeit besitzt, Vorhersagen erklärbar zu machen. Da die Klassenverteilung in den historischen Verhaltensdaten und in den Experimentdaten stark unausgewogen ist (es gibt kaum Upgrader und Downgrader und sehr viele



die immer im Standardtarif waren), haben wir mittels Down-Sampling eine annähernd ähnliche Verteilung von positiven und negativen Beispielen simuliert.

Tabelle 4: Ableitung der positiven Trainingsklassen für das Simulationsexperiment

Trainingsdatentypen	Downgrade	Upgrade
Trainingsdatentyp 1 (Historische Verhaltensdaten auf Basis von historischen Tarifwechseln vor der Tarifumstellung)	Kunden die vom Standard- in den Atomstromtarif gewechselt sind	Kunden die vom Standard in den Tarif Erneuerbar I oder Erneuerbar II gewechselt sind
Trainingsdatentyp 2 (Umfragedaten zur Kaufbereitschaft für die neuen Tarife)	Umfrageteilnehmer, die eine hohe Kaufabsicht gegenüber dem neuen Atomstrom*-tarif geäussert haben.	Umfrageteilnehmer, die eine hohe Kaufabsicht gegenüber Lokaler Ökostrom geäussert haben.
Trainingsdaten Typ 3 (Verhaltensdaten aus einem Experiment, bei dem die Tarife aller Kunden umgestellt wurden und deren Wechsel aufgezeichnet wurden)	Kunden, die während der Tarifumstellung aktiv vom alten Standard- in den neuen Atomstrom*-Tarif gewechselt sind oder vom alten Atomstrom- in den neuen Atomstrom*-Tarif migriert wurden.	Kunden, die während der Tarifumstellung aktiv vom alten Standardtarif in den lokalen Ökostrom gewechselt sind oder von einen der alten Erneuerbaren Tarife den lokalen Ökostrom gewechselt sind.

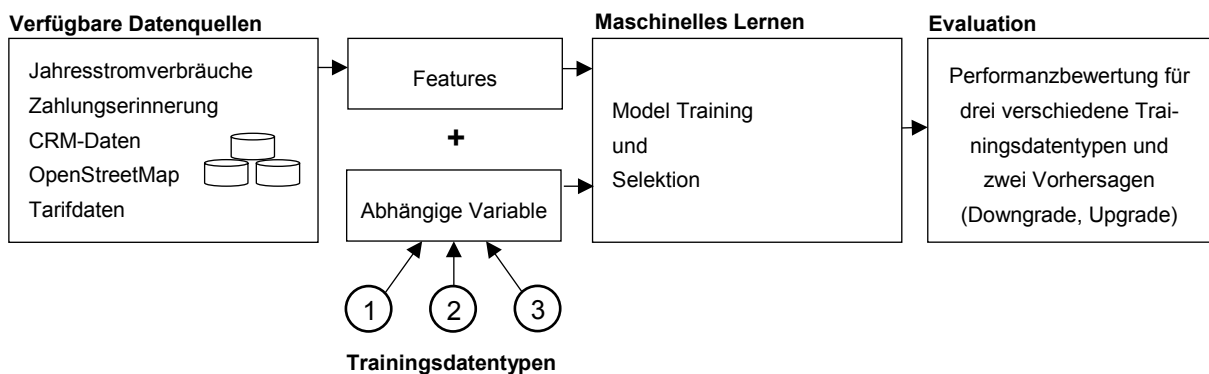


Abbildung 6: Feinkonzept zum Vorgehen zur Klassifikation der Haushalte in das Zielsegment

Evaluation des Simulationsexperiments: Wir bewerten die Qualität der Vorhersagemodelle anhand der AUC-Metrik. Dieses Klassifizierungsleistungsmass für binäre Klassifizierungsprobleme basiert auf der ROC-Kurve (Receiver Operating Characteristic), die den Zusammenhang zwischen der True Positive Rate (TPR) und der False Positive Rate (FPR) zur unterschiedlichen Unterscheidung von Schwellenwerten in einem zweidimensionalen Diagramm veranschaulicht. Die Fläche unter der ROC-Kurve (AUC) ist ein anerkanntes Gütemass und liegt zwischen 0 und 1, während Werte kleiner oder gleich 0,5 Zufallsvorhersagen anzeigen und 1 eine perfekte Klassifizierung darstellt (Fawcett 2006). Für die methodisch korrekte Beurteilung der Güte der Vorhersage verwenden wir einen dreistufigen Ansatz: Zuerst wählen wir zufällig 80% der jeweiligen Trainingsdaten für das Modelltraining (Trainingspartition) aus und behalten 20% für die Modellbewertung (Testpartition). Zweitens wird das Modelltraining und die Parameterauswahl über die Trainingspartition durchgeführt. Um die beste Parameter-



kombination für das jeweilige Vorhersageproblem auszuwählen, wird jede der sechs Parameterkombinationen mit 10-facher Kreuzvalidierung bewertet. Da auch bei diesem Verfahren noch Zufallseffekte auftreten können, wird die Kreuzvalidierung fünfmal wiederholt, was zu 50 Gütemessungen für jede Parameterkombination führt. Wir wählten die Parameterkombination mit dem höchsten Trainings-AUC für die Vorhersage. Drittens verwenden wir für die endgültige Validierung des Modells Validierungsdaten mit den tatsächlich gewählten neuen Tarifen. Diese stammen aus einem Datenexport vom April 2019, also vier Monate nachdem die automatische Migration aller Kunden durchgeführt wurde. So erhalten wir Training-AUC, Test-AUC und Validierung-AUC für jedes betrachtete Vorhersageproblem, mit denen wir unsere Hypothese im nächsten Abschnitt bewerten.

Ergebnisse des Simulationsexperiments: Wir betrachten sechs abhängige Variablen, die sich aus den drei verschiedenen Trainingsdatentypen (historischer Verhaltensdaten, Experiment, Umfrage) und zwei verschiedenen Vorhersageproblemen (Upgrade vom Standardtarif auf den Tarif Lokaler Ökostrom und Herabstufung vom Standardtarif auf den Atomstrom*) ergeben, siehe Abbildung 3. Tabelle 5 zeigt die Trainings-AUC der besten Parameterkombination, Test-AUC und Validierungs-AUC wie oben beschrieben, sowie die Anzahl der verfügbaren positiven und negativen Beispiele, die zum Training des Modells verwendet wurden.

Aus dem Simulationsexperiment lassen sich folgende Kernergebnisse ableiten:

- Sekundäre Daten (historisches Kaufverhalten) erlauben eine Vorhersage des Down- und Upgrade-Verhaltens besser als der Zufall.
- Umfragedaten können nur bedingt für die Vorhersage des Kundenverhaltens verwendet werden, was vermutlich an der unkonkreten Formulierung der Kaufentscheidung liegt. Allerdings kann das Downgrade besser als das Upgrade vorhergesagt werden.
- Daten, die im Rahmen eines Experiments erhoben werden, können besser als Umfragedaten dazu verwendet werden, um das Kaufverhalten für lokale Ökostromprodukte vorherzusagen. Allerdings funktionieren die historischen Verhaltensdaten besser hinsichtlich der Vorhersagegüte.



Tabelle 5: Ergebnisse der Vorhersagemodelle zum Tarifwechsel für die drei verschiedenen Trainingsdatentypen

Abh. Var.	Daten	Positiv	Negativ	AUC	Teststatistiken	ROC-Kurve (Validierung)
Trainingsdatentyp 1: Hist. Verhalten	Downgrade					
	Train	3'257	3'257	0,747	t=84,339; p<0,001; SD=0,021	
	Test	814	23'892	0,755	-	
	Validierung	1'119	109'194	0,551	$\chi^2=47,762$; p<0,001	
	Positiv*	440	32'509	-	-	
	Negativ*	679	76'685	-	-	
Trainingsdatentyp 1: Hist. Verhalten	Upgrade					
	Train	993	993	0,779	t = 66,871; p<0,001; SD=0,030	
	Test	248	23'892	0,793	-	
	Validierung	139	110'174	0,592	$\chi^2=10,185$; p<0,002	
	Positiv*	60	33'438	-	-	
	Negativ*	79	76'736	-	-	
Trainingsdatentyp 2: Kaufintention	Downgrade					
	Train+	307	408	0,488	t=-1,5066; p=0,931; SD=0,057	
	Test	76	101	0,519	-	
	Validierung	1'119	109'194	0,501	$\chi^2=1,168$; p=0,280	
	Positiv*	300	30'921	-	-	
	Negativ*	819	78'273	-	-	
Trainingsdatentyp 2: Kaufintention	Upgrade					
	Train+	227	496	0,546	t=4,146; p<0,001; SD=0,079	
	Test	56	124	0,583	-	
	Validierung	139	110'174	0,509	$\chi^2=0,208$; p=0,648	
	Positiv*	1	1'718	-	-	
	Negativ*	138	108'456	-	-	
Trainingsdatentyp 3: Experiment	Downgrade					
	Train+	289	289	0,654	t=25,087; p<0,001; SD= 0,0434	
	Test	72	1'643	0,603	-	
	Validierung	131	102'527	0,576	$\chi^2=57,153$; p<0,001	
	Positiv*	473	34'615	-	-	
	Negativ*	572	66'998	-	-	
Trainingsdatentyp 3: Experiment	Upgrade					
	Train+	45	45	0,671	t=8,700; p<0,001; SD=0,139	
	Test	11	1'643	0,542	-	
	Validierung	131	102'527	0,579	$\chi^2=1,402$; p=0,236	
	Positiv*	51	34'508	-	-	
	Negativ*	80	68'019	-	-	



2.4 Diskussion der Ergebnisse und Implikationen

Ziele in diesem Teilprojekt waren erstens der Test von Kundensegmentierungsalgorithmen zur Entwicklung von personalisierter Ansprache und zweitens die Entwicklung von Vorhersagemodellen zur Identifikation von Kunden mit hohem Upselling-Potenzial für einen lokalen Ökostromtarif.

Maschinelle Lernverfahren, sog. Clustering-Algorithmen, erlauben die Segmentierung von Kunden in mehrere in sich homogene Kundengruppen mit ähnlichen Einstellungen, Interessen und Zielen. Hierfür wurden Kunden im Standardtarif zu einer Umfrage eingeladen und spezifische Daten für neue Stromtarife und persönlichen Einstellungen erhoben. Ausserdem wurden verschiedene interne Datenquellen (Messwesen, Kundenkontaktcenter, Abrechnung, Onlineportal) und externe Datenquellen (freie öffentliche geographische Informationssysteme) miteinander kombiniert und zusammen analysiert. Die Segmentierung der Kunden im Standardtarif hat im Projekt geholfen, eine zielgerichtete persönliche Kundenansprache zu entwickeln.

Maschinelle Lernverfahren, sog. Vorhersagealgorithmen, erlauben unter bestimmten Voraussetzungen die Identifikation von Kunden, die einen Tarifwechsel zu einem regionales Stromprodukt durchführen. Im Projekt wurden hierfür Vorhersagemodelle auf Basis verschiedener Trainingsdatentypen (historischer Verhaltensdaten, Experiment, Umfrage) getestet, um den Tarifwechsel auf einen lokalen Ökostromtarif von Kunden im Standardtarif vorherzusagen. Das Vorhersagemodell auf Basis von Umfragetrainingsdaten wurde in einem Feldexperiment mit 4'712 Kunden getestet, wobei lediglich 0,88% der selektierten Kunden Ihren Tarif gewechselt haben. Das Ergebnis der Kampagne wurde im Rahmen eines Workshops diskutiert und acht Herausforderungen identifiziert, von denen die wichtigste, die Erhebung von guten Trainingsdaten im Detail mit einem Simulationsexperiment auf Basis von 110'313 Tarifänderungen untersucht wurde. Es zeigt sich, dass historische Verhaltensdaten (alte Tarifwechseldaten von ähnlichen Tarifen) gute Vorhersagen ermöglichen. Zudem zeigt sich, dass bei einer notwendigen Erhebung (z. B., wenn keine historischen Verhaltensdaten zur Verfügung stehen) Experimentdaten Umfragedaten vorzuziehen sind. Nachfolgend werden diese Herausforderungen sowie theoretische und praktische Implikationen im Detail beschrieben.

2.4.1 Herausforderungen bei der Entwicklung von Vorhersagemodellen

Aus den Ergebnissen der Forschungsarbeit in diesem Teilprojekt können wir folgende Herausforderungen bei der Entwicklung von Vorhersagemodellen zur Identifikation von potenziellen Interessenten für Stromtarife ableiten:

Kleine-Klassen-Problem: Die Zielklasse «Upgrader» ist im Vorhersagemodell mit einem Anteil von 9,75% im Vergleich zur Klasse Nicht-Upgrader (90,25%) stark unterrepräsentiert. Das ist für Kunden, die bisher einen Standardtarif hatten, nicht ungewöhnlich, da solche Kunden bisher noch nie einen Stromtarif gewechselt haben. Unterrepräsentierte Klassen führen bei vielen Maschinellen Lernverfahren zu ungenauen Vorhersagen. Sampling-Methoden können ungleiche Verteilungen beim Trainieren von Modellen ausgleichen.

Wenige Trainingsdaten: Eine geringe Anzahl an Trainingsdaten kann dazu führen, dass maschinelle Lernverfahren zu spezifische Regeln ableiten, die nicht ausreichen, um die Grundgesamtheit der Kunden zu generalisieren. Die Bewertung von Vorhersagemodellen, die mit einer geringen Anzahl an Trainingsdaten trainiert worden sind, ist komplex, da die Performanz schon bei geringen Änderungen an Daten, Methoden oder Parametern starken Schwankungen unterliegt. Dadurch ist eine kontinuierliche Verbesserung von Methoden, z. B. eine optimierte Feature-Auswahl oder Parameteroptimierung schwierig, da unklar ist, von welcher Veränderung die Performanz tatsächlich stammt.

Erhebung von Ground-Truth-Daten: Die Erhebung der Zielklasse „Upgrade“ mithilfe einer Kaufintentionsskala ist ggf. ungenau, da mit einer Kaufintention nicht zwangsläufig ein Kauf (Tarifabschluss



bzw. -wechsel) einhergeht (Ajzen 1985). In der Regel ist eine getätigte Kaufintention eher mit einer Wahrscheinlichkeit gleichzusetzen, die in Abhängigkeit des konkreten Produkts steht und vorab kaum erhoben werden kann.

Skalenumwandlung / -transformation: Die Kaufintentionsskala wird üblicherweise mit mehreren Items einer 7er Likert-Skala erhoben, wobei die Zustimmung zum Kauf verschieden stark ausgeprägt ist (schwache, mittlere oder starke Zustimmung). Die Umwandlung der Kaufintentionsskala in Klassen kann also mit unterschiedlich starken Zustimmungen erfolgen. Zum Beispiel können nur Umfrageteilnehmer mit einer sehr starken Zustimmung zur Klasse Upgrader zugeordnet werden oder alle Kunden, die mindestens eine leichte Zustimmung haben. Auch hier kann die Anzahl einer Mindestmenge an Trainingsbeispielen die Zuordnung einschränken, z. B. wenn viele Umfrageteilnehmer eher eine sehr leichte Zustimmung zum Kauf angegeben haben und nur wenige eine starke Zustimmung.

Komplexität bei der Produktentwicklung und beim Einsatz von Vorhersagetools: Der Entwicklungsvorgang von neuen Tarifen ist rechtlichen Vorgaben unterworfen, bei denen langwierige Abstimmung zwischen verschiedenen Abteilungen z. B. Netz, Recht, Marketing und Produktentwicklung notwendig ist. Indirekt hat auch dies Einfluss auf die Erhebung der Zielvariable im Vorhersagemodell «Kaufintention». Die Kaufintention muss möglichst in Bezug auf ein konkretes Produkt erhoben werden, bei denen Hauptmerkmale wie Preis, Aussehen und Eigenschaften feststehen. Wenn dieses Produkt nun mittels einer Umfrage getestet und die Kaufintention von Umfrageteilnehmern erhoben wird, sollten sich die Produkteigenschaften nicht mehr verändern. Im Projekt gab es im Vergleich zum final angebotenen Produkt geringfügige Veränderungen ggü. dem vorgestellten Produkt in der Umfrage. Konkret war der Name und der Preis des Stromtarifs und der alternativen Stromtarife noch unbekannt und konnte deshalb nicht genannt werden. Ausserdem wurde die Produktbeschreibung geringfügig von einem lokalen Stromprodukt zu einem lokalen Ökostromprodukt hin verändert.

Ungewisse Produktnachfrage für neuartiges Produkt: Das neuartige Produkt «MeinRegioStrom» war möglicherweise nur für eine sehr kleine Zielgruppe interessant. Zuerst war die bisherige Nachfrage bei der CKW AG von Kunden nach Stromtarifen mit neuen erneuerbaren Energieträgern gering. Der Anteil an Kunden, die sich explizit für einen der alten Tarife mit neuen erneuerbaren Energieträgern entschieden haben, lag lediglich bei 1% (siehe **Abbildung 3**). Des Weiteren war zur Produkteinführung der Aspekt Regionalisierung (lokal produzierter Strom) begrenzt auf die Ortschaft Kriens und Umgebung. Es ist davon auszugehen, dass der Aspekt Regionalität im CKW-Vertriebsgebiet, welches 73 Gemeinden umfasst, nicht für jeden Kunden als solcher wahrgenommen wurde.

Risiken bei der Kundenansprache: Die Chancen der Produktpromotion eines lokalen und höherpreisigen Ökostromtarifs stehen dem Risiko eines Downgrades zu einem ökologisch weniger wertvollen Produkt gegenüber. Der mittelpreisige Stromtarif ist auch gleichzeitig das Standardprodukt und wurde von den meisten Kunden nicht aktiv gewählt. Es war also davon auszugehen, dass sich Kunden erstmalig mit den angebotenen Alternativen auseinandersetzen. In der Kundenkommunikation wurden daher zwei Kernbotschaften hinterlegt, die ein Downgrade verhindern und gleichzeitig ein Upgrade befördern sollen. Erstere war so formuliert, dass der Kunde bei einer automatischen Migration, also wenn er / sie nicht selbst aktiv den Tarif wechselt, von 10% geringeren Energiekosten profitiert. Es wurde also trotz Promotion eines lokalen Ökostromtarifs auch indirekt ein Anreiz gesetzt, kein Upgrade zum höherpreisigen Stromtarif durchzuführen.

Wahl des optimalen Vertriebskanals: Gleichzeitig mit dem neuen Stromprodukt wurde ein neues Online-Portal vorgestellt, über das der Tarif gewechselt werden konnte. Diesen Vertriebskanal gab es zwar bereits in der Vergangenheit, er wurde aber nur von einem sehr kleinen Anteil an Kunden genutzt. Es ist davon auszugehen, dass der Abschluss eines Stromtarifs via Online-Portal für viele Kunden ungewohnt war.



2.4.2 Theoretische Implikationen

Die Ergebnisse dieser Studie unterstützen den wissenschaftlichen Diskurs über die Datenerhebung, der durch epistemologische Arbeiten (d. h. über die Erhebung von Forschungsdaten) sowie durch Arbeiten aus der Marktforschung geprägt war. Beide Bereiche unterscheiden zwischen Primärdaten (die für ein spezielles Forschungsziel oder Marktforschungsprojekt erhoben werden) und Sekundärdaten (die für andere Zwecke oder zur Lösung eines anderen Problems als des vorliegenden Problems erhoben werden). Mehrere Studien aus der Psychologie und dem Marketing haben gezeigt, dass das zukünftige Verhalten von Individuen durch vergangenes Verhalten erklärt werden kann. In Übereinstimmung mit mehreren früheren Studien, die prädiktive Analysen zur Vorhersage des Kundenverhaltens auswerten, finden wir Unterstützung für die Hypothese, dass vergangene Verhaltensdaten, die als Transaktionsdaten in Enterprise-Resource-Planning-Systemen (ERP) von Unternehmen gespeichert sind, zusammen mit maschinellen Lernverfahren zur Vorhersage des Kaufverhaltens verwendet werden können. Unsere Studie zeigte, dass selbst Daten von leicht unterschiedlichen Kaufentscheidungen verwendet werden können, um das zukünftige Verhalten bei ähnlichen Entscheidungen vorherzusagen.

Wir bewerten zwei Methoden der Datenerhebung: Vorsätzliche Daten, die bei Umfragen erhoben wurden, und Verhaltensdaten, die bei Experimenten (z. B. mit Testkunden) erhoben wurden. Die zwei Erhebungsmethoden sind notwendig, da Daten über vergangenes Verhalten manchmal nicht verfügbar sind. Ausserdem können die Daten nicht im Sinne der Datenschutz- und Datenschutzgesetzgebung geschützt werden, da die Einwilligung der betroffenen Person in die Datenverarbeitung zu einem anderen Zweck erteilt worden sein könnte. Unsere Ergebnisse zeigen, dass die experimentellen Daten für Vorhersagen über das Kaufverhalten der Kunden deutlich besser sind. Wenn Unternehmen Trainingsdaten für die prädiktive Modellierung sammeln wollen, sollten sie sich für die Durchführung von Experimenten und nicht für Umfragen entscheiden. Dies behebt auch einige andere Probleme, die die Datenerhebung mit sich bringt (z. B. Verzerrungen, Abhängigkeit von Messinstrumenten).

2.4.3 Praktische Implikationen

Unsere Studie erlaubt drei Empfehlungen für Praktiker. Erstens, wenn historische Daten aus dem Kundenverhalten (z. B. aus ERP- oder CRM-Systemen) verfügbar sind, sollten diese verwendet werden. Zweitens, falls die Sammlung experimenteller Daten nötig ist, muss das Unternehmen lange vor dem offiziellen Rollout eines Produkts entsprechend Zeit in die endgültigen Produktbeschreibungen investieren und z. B. eine Online-Plattform einrichten, auf der diese Daten erhoben werden können (z. B. Webshop oder Smartphone-App, Self-Service-Portal). Dies kann eine Menge persönlicher und technischer Ressourcen binden. Schliesslich sind experimentelle Daten Umfragedaten vorzuziehen, wenn der Kauf neuer Produkte vorhergesagt werden soll und die Eigenschaften der Produkte nicht vollständig bekannt sind.

3 Teilprojekt 2: Vertrieb von Photovoltaik- und Energiespeichieranlagen

In diesem Teilprojekt untersuchten wir im Zeitraum 01/2019 bis 03/2020, inwiefern der Vertrieb von erneuerbaren Energieanlagen für den Privatsektor durch Datenanalysen und maschinelles Lernen unterstützt werden kann. Hierzu führten wir zwei Umfragen unter Interessenten solcher Anlagen durch und entwickelten Vorhersagemodelle, die von der BEN Energy AG in ein prototypisches Entscheidungsunterstützungssystem für die Vertriebsunterstützung bei der CKW AG im Bereich «SmartEnergy» überführt wurden.



Nachfolgend beschreiben wir dir beiden Teilziele des Teilprojekts sowie die Forschungsmethodik. Am Schluss des Abschnitts diskutieren wir die Ergebnisse und leiten Implikationen für die Praxis und Forschung ab.

3.1 Ausgangslage und Teilziele

Für Haushalte bestehen zahlreiche Möglichkeiten, in erneuerbare Energieträger zu investieren, zum Beispiel in Photovoltaikanlagen, Batteriespeicher und Wärmepumpen. Die Kombination der Energieerzeugung mit einem Batterie- oder Wärmespeichersystem ermöglicht es den Haushalten nicht nur, ihre Eigenversorgung mit selbst erzeugter Energie zu erhöhen, sondern trägt auch zur Entlastung der Stromnetze bei und senkt so die Notwendigkeit eines teuren Netzausbaus. Allerdings bleibt der Ausbau solcher Systeme hinter den Erwartungen zurück. So ist beispielsweise der Absatz von Photovoltaik ausserhalb Asiens in den letzten Jahren vor allem in Europa zurückgegangen (Jäger-Waldau 2018). Eine neue Generation von Entscheidungsunterstützungssystemen, die auf maschinellem Lernen basiert, kann den bisher teuren und expertenzentrierten Verkaufsprozess in kritischen Phasen unterstützen (Syam und Sharma 2018) und so dazu beitragen, die Verbreitung von erneuerbaren Energieanlagen in Privathaushalten zu erhöhen. Studien im Bereich der Wirtschaftsinformatik veranschaulichen den Beitrag von maschinellen Lernverfahren in verschiedenen Phasen des Verkaufsprozesses: Bei der Zielausrichtung und Positionierung (Martens u. a. 2016; Olson und Chae 2012), der Bedarfschätzung (Loureiro, Miguéis, und da Silva 2018; Prinzie und Van den Poel 2007; Shrivastava und Jank 2015) und der Lead-Generierung sowie der Qualifizierung (Cui, Wong, und Wan 2012).

Nachdem die technologische Unterstützung der Vertriebsprozesse im Bereich des Vertriebs erneuerbarer Energieanlagen noch nicht Bestandteil der Forschung war, hatte dieser Projektabschnitt zum Ziel, das Potenzial von Photovoltaik- (PV) und Speicheranlageninstallationen von Privathaushalten mit Hilfe von maschinellen Lernverfahren zu identifizieren. Dies beinhaltet einerseits die Abschätzung des technischen Potenzials, also z. B. die Eignung der baulichen Gegebenheiten für die Installation einer Anlage, andererseits das Adaptionspotenzial in Form der Kaufbereitschaft der Immobilienbesitzer, also wie sehr ein Kunde bzw. Interessent bereit ist, in eine solche Anlage zu investieren. Die Modellentwicklung gliedert sich in vier Phasen, die nachfolgend in Unterabschnitten ausgeführt sind:

1. Erhebung von Daten zum Training von Vorhersagemodellen
2. Konzeptionierung des Vorhersagesystems
3. Modellbildung und Evaluation der Vorhersagen
4. Prototypische Umsetzung

Am Ende wurde das statische Modell durch die BEN Energy AG in einen Prototyp als Webanwendung umgesetzt, welche als Vertriebsunterstützung durch die CKW AG genutzt werden kann.

3.2 Erhebung von Daten zum Training von Vorhersagemodellen

Für die Entwicklung von Vorhersagemodellen werden zwei Arten von Daten, sog. Ground-Truth-Daten und Features benötigt. Erstere repräsentieren die Zielklasse, also Interessenten von PV- oder Speicheranlageninstallationen mit geeigneten Immobilien (technisches Potenzial vorhanden) sowie einer hohen Kaufbereitschaft (Bereitschaft zur Adaption vorhanden). Features werden benötigt, um Regeln mithilfe von Maschinellen Lernverfahren abzuleiten, die es erlauben, Interessenten mit hohem technischem Potenzial und hoher Kaufbereitschaft von Interessenten mit niedrigem technischem Potenzial und niedriger Kaufbereitschaft abzugrenzen. Ground-Truth-Daten und Features wurden aus internen Systemen der CKW AG und aus externen Datenquellen erhoben, bzw. mithilfe von Umfragen erfasst. Die einzelnen verfügbaren Datenquellen werden nachfolgend beschrieben.



Interne Daten wurden von der CKW AG in Form von Anfragen aus zwei Onlinerechnern und zwei CRM-Systemen zur Verfügung gestellt. Bei den Anfragen aus den Onlinerechnern handelt es sich um Exporte des Solarspeicherrechners² und des Heizungsrechners³. Diese Onlinerechner sind in die Webseite der CKW AG integriert und bieten Interessenten die Möglichkeit, unverbindliche Kostenschätzungen für Photovoltaikanlagen, Batteriespeicher und Wärmepumpen auf Basis von Nutzereingaben zu berechnen. Daneben erlauben die Onlinerechner dem Nutzer, eine weitergehende Beratung durch die CKW AG anzufordern. Onlinerechnerdaten enthalten grundlegende Informationen zum Gebäude, zur geplanten Anlage, zu bisherigen Anlagen und errechnete Kennzahlen zum Projekt. Darüber hinaus lassen sich aus der Interaktion des Interessenten mit dem Rechner bzw. der Häufigkeit des Aufrufs weitere Informationen ableiten. Die Daten aus den Onlinerechnern enthalten primär Merkmale zum technischen Potenzial einer Immobilie. Bei den Exporten aus den CRM-Systemen handelt es sich um Kundenauftragsinformationen von den Standorten Luzern und Zürich.

Externe Daten wurden aus dem Solarkataster von Sonnendach.ch⁴ und Bevölkerungsdaten von opendata.swiss⁵ angebunden. Solarkatasterdaten wurden uns freundlicherweise von dem Bundesamt für Energie zur Verfügung gestellt und enthalten Informationen zum Photovoltaikpotenzial von Dachflächen, sowie Schätzwerte zum Strom- und Wärmebedarf von Immobilien in der Schweiz (Klauser und Schlegel 2016). Der Datensatz besteht aus Vektordaten zu Dachflächen, deren Photovoltaikpotenzial und Informationen zum Strom- und Wärmebedarf der Immobilien.

Für die Entwicklung der beiden Umfragen wurden Faktoren herausgearbeitet, die den Kauf von PV- und Heizsystemen beeinflussen. Hierzu wurden fünf Interviews mit Vertriebsmitarbeitern der CKW AG geführt sowie existierende Literatur ausgewertet. Auf dieser Basis wurden zwei Umfragen erstellt zu der alle Nutzer des Onlinerechners mit vorhandener E-Mail-Adresse eingeladen wurden. In den Umfragen wurden die folgenden Themen erhoben: Nutzung des Onlinerechners, Informationen zur Anschaffung von PV-Anlagen, Heizsystemen, bzw. Wärmepumpen, Fortschritt des Kunden im Kauf- / Entscheidungsprozess, Fragen für Käufer eines Systems, Fragen für Interessenten eines Systems, Peergruppeneffekte, Motivatoren für den Kauf, gewünschte Systemeigenschaften, Ortsbezug, Gebäudeinformationen, sozioökonomische Faktoren, sowie persönliche Einstellungen wie Autarkiebedürfnis, Statussymbol, Umwelt und Early-Adopter. In der Umfrage wurden ausserdem Ground-Truth-Daten erhoben, die für die Entwicklung von Vorhersagemodellen essenziell sind.

3.3 Konzeptionierung des Vorhersagesystems

Für die Entwicklung von Vorhersagemodellen ist die Definition von abhängigen Variablen und Features ein wichtiger Schritt. In einem gemeinsamen Workshop haben wir sechs verschiedene abhängige Variablen auf Basis eines üblichen Kaufprozesses von nachhaltigen Investitionsgütern identifiziert, die den Kaufprozess an verschiedenen Stellen unterstützen können, siehe Abbildung 7.

1. Interesse an Beratung: Hat der Lead Interesse an einer Kontaktaufnahme?
2. Produktart: Was für ein Produkt möchte der Lead kaufen?
3. Zeitraum bis zur Beauftragung: Wann vergibt der Kunde voraussichtlich einen Auftrag?
4. Kostenrahmen: Um wie viel unterscheidet sich der finale Anlagenpreis vom geschätzten Anlagenpreis im Onlinerechner?

² <https://solarspeicherrechner.ckw.ch/>

³ <https://heizungsrechner.ckw.ch/>

⁴ <https://www.uvek-gis.admin.ch/BFE/sonnendach/>

⁵ <https://opendata.swiss/de/dataset/bevoelkerung-pro-plz>



5. Kaufintention: Wie sicher möchte der Kunde eine Anlage kaufen?
6. Kaufentscheidung: Hat der Lead bereits eine Solaranlage gekauft?

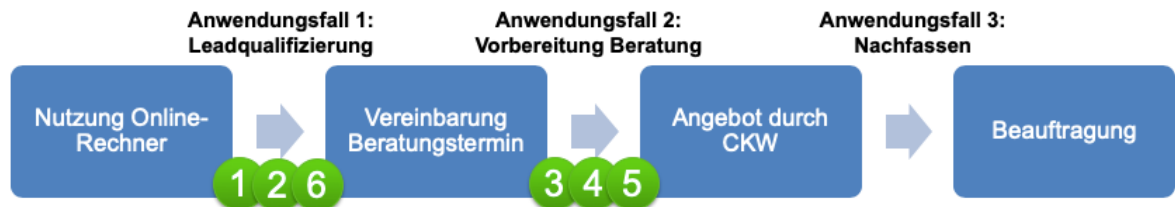


Abbildung 7: Vorhersagen im Kontext des Kaufprozesses

Beispielsweise erlaubt es die abhängige Variable «Kaufintention», Interessenten mit hohem technischem Potenzial und hoher Kaufbereitschaft von Interessenten mit niedrigem technischem Potenzial und niedriger Kaufbereitschaft abzugrenzen. Wir entschieden uns für die Entwicklung von Vorhersagemodellen für drei der sechs abhängigen Variablen. Interesse an einer Beratung, der Kaufintention und der Kaufentscheidung für die beiden Investitionsgüter Photovoltaikanlage und Wärmepumpe.

Für die Entwicklung von Features können prinzipiell alle intern und extern erhobenen Daten verwendet werden. Bei der Auswahl an Daten gilt es allerdings zwischen den beiden Aspekten der Vorhersagegenauigkeit und der Datenverfügbarkeit abzuwägen. Die Verwendung von Umfragedaten als Features erlaubt es beispielsweise detaillierte persönliche Einstellungen wie die Einstellung zur Umwelt oder dem Autarkiebedürfnis aufzunehmen. Solche detaillierten Daten bieten die Chance, die Genauigkeit eines Vorhersagemodells zu erhöhen, allerdings ist die Verfügbarkeit von solchen Daten eingeschränkt auf die Teilnehmer der Umfrage. Ein Modell, das zum Beispiel auf Basis solcher Daten trainiert worden ist, benötigt für die Vorhersage von neuen Interessenten ebenfalls solche Daten, die derzeit nicht erhoben werden. Im Projekt galt es herauszufinden, welche der Datenquellen einen Mehrwert liefern können. Für die Ableitung von Features wurden die Daten miteinander verknüpft, bereinigt, aggregiert und fehlende Werte ergänzt, soweit möglich. **Tabelle 6** zeigt die Anzahl abgeleiteten Features pro Datenquelle.

Tabelle 6: Anzahl der abgeleiteten Features zur Entwicklung von Vorhersagemodellen

Datenquelle	Photovoltaik	Wärmepumpe
Onlinerechner	29	25
Interaktionsdaten aus dem Onlinerechner	9	8
Solarkatasterdatensatz	8	11
Geoinformationen	0	1
Umfrage	115	118

3.4 Modellbildung und Evaluation der Vorhersagen

Wir wählten die Klassifikationsalgorithmen Random Forest und logistische Regression zum Trainieren der Vorhersagemodelle. Die ungleiche Klassenverteilung im hier vorliegenden Fall (siehe **Tabelle 7**) stellt beim Trainieren eine Herausforderung dar, da die Erkennung von kleinen Klassen erschwert wird. Dies ist z. B. der Fall für die positive Klasse «hat Interesse an Beratung für Photovoltaik», bei der lediglich 6% aller Interessenten in diese Kategorie fallen. Eine Lösung stellt die Verwendung von



Sampling-Verfahren dar, bei der die ungleiche Klassenverteilung im Trainingsprozess gezielt harmonisiert wird. Wir haben insgesamt fünf verschiedene Sampling-Verfahren (Downsampling, Upsampling, ROSE, SMOTE, kein Sampling) getestet.

Tabelle 7: Verteilung der positiven und negativen Trainingsbeispiele pro abhängige Variable

Abhängige Variable	Photovoltaik		Wärmepumpe	
	Positiv	Negativ	Positiv	Negativ
Interesse an Beratung	6%	94%	10%	90%
Kaufintention	25%	75%	25%	75%
Kaufentscheidung	36%	64%	28%	72%

Die verschiedenen möglichen Modellparameter (fünf verschiedene Datenquellen, Teilfragen aus der Umfrage, fünf Sampling-Verfahren, zwei Lernalgorithmen, drei Vorhersagevariablen, 10-fache Kreuzvalidierung und die beiden Anwendungsfälle Photovoltaik und Wärmepumpe) führten zu einer Vielzahl an Modellen (circa 20'400). Um belastbare Aussagen über die optimale Parameterkombination aus den Vorhersageergebnissen abzuleiten haben wir eine Varianzanalyse durchgeführt.

Wie sich gezeigt hat, erlauben Modelle mit den Sampling-Verfahren Downsampling und ROSE deutlich verbesserte Vorhersagen im Vergleich zu keinem Sampling in nahezu allen Fällen. Wir wählten Downsampling. Der Algorithmus Random Forest führt im Vergleich zur logistischen Regression zu leicht besseren Vorhersagen, weshalb wir diesen wählten.

Die Vorhersagegüte variiert je nach verwendeter abhängiger Variable. Trotzdem können wir generelle Empfehlungen hinsichtlich der technischen Nutzbarkeit der getesteten Datenquellen und der hierdurch erzielten Verbesserung der Vorhersage ableiten, die in **Tabelle 8** zusammengefasst sind.

Tabelle 8: Empfehlungen hinsichtlich der technischen Nutzbarkeit und der potenziellen Verbesserung der Vorhersage

Datenquelle	Photovoltaik		Wärmepumpe	
	Nutzbar	Verbesserung	Nutzbar	Verbesserung
Rechner	✓	✓	✓	✓
Interaktion	✓	✓	✓	✗
Solar	✓	✗	✗	✓
Geo	✓	-	✓	✓
Umfrage	✗	✓	✗	✓

Die Güte der Vorhersagemodelle für die beiden Anwendungsfälle Photovoltaik und Wärmepumpe sind in **Abbildung 8** dargestellt. Wir haben die Ergebnisse anhand der vier Gütemasse Accuracy, Sensitivität, Spezifität und F2-Metrik (einem gewichteten harmonischen Mittel aus Sensitivität und Spezifität) analysiert. Das Konfidenzintervall erlaubt eine Abschätzung der Stabilität der Vorhersagen für neue Anfragen im Onlinerechner und wurde auf Basis der 10-fachen Kreuzvalidierung berechnet. Die finale Modellauswahl erfolgte mittels der F2-Metrik.

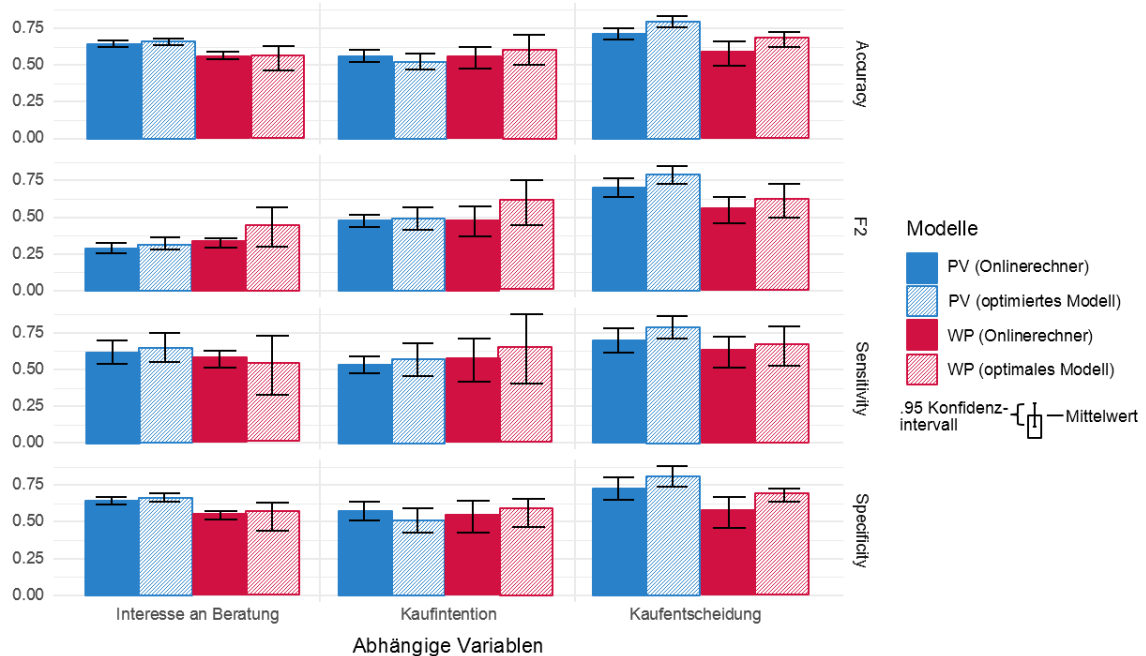


Abbildung 8: Qualität der Vorhersagemodelle

In fast allen Fällen erzielten Vorhersagemodelle mit zusätzlichen Datenquellen (optimiertes Modell) bessere Ergebnisse als Modelle die nur auf Onlinerechnerdaten basierten. Insgesamt fällt auf, dass die Vorhersagequalität für den Anwendungsfall Photovoltaik höher ist und die wichtigen abhängigen Variablen *Kaufentscheidung* (bis zu 79% Accuracy) und *Interesse an Beratung* (bis zu 66% Accuracy) gut vorhergesagt werden können. Die Vorhersage der Variable *Kaufintention* erzielt eine geringere Genauigkeit (bis zu 62% Accuracy) und führte bei einigen Tests zu stark schwankenden Ergebnissen.

3.5 Prototypische Umsetzung des Entscheidungsunterstützungssystems

Die Vorhersagemodelle wurden prototypisch implementiert und in Form einer Weboberfläche als Dashboard visualisiert. Das Dashboard zeigt die vorhergesagten Scores der drei Variablen für die zwei Anwendungsfälle Photovoltaik und Heizung. Insgesamt gibt es drei Ansichten, die im Folgenden beschrieben werden. In der Übersichtsansicht (Ansicht 1) werden Anfragen von Interessenten im aktuellen Monat dargestellt, siehe **Abbildung 9**. Für jede Vorhersagevariable existiert eine kombinierte Karten- und Detailansicht (Ansicht 2) die in **Abbildung 10** und **Abbildung 11** dargestellt ist. Die Kartenansicht erlaubt die geographische Zuordnung und das Erkennen von Clustern von Anfragen in bestimmten Regionen. Je weiter in die Karte gezoomt wird, desto detaillierter werden die Anfragen angezeigt. Auf maximaler Zoomebene wird der Vorhersagescore für die jeweilige Variable für jeden einzelnen Interessenten mittels Farbkodierung angezeigt (dunkelblau = niedriger Score, dunkelrot = hoher Score). Daneben werden weitere Detailinformationen, wie z. B. die Dachform angezeigt. Diese Ansicht erlaubt es dem Vertriebsmitarbeiter Anfragen überblicksartig zu erfassen, die Qualität von Einzelanfragen zu bewerten und diese zu Gruppen für kombinierte Vor-Ort-Besuche zusammenzufassen. Zudem gibt es die Möglichkeit, einen CSV-Export von potenziellen Interessenten in der Nähe einer Kundenanfrage zu erhalten. Die Report- / Downloadansicht (Ansicht 3) erlaubt den Download von Kundengruppen mit ähnlichen Merkmalen, siehe **Abbildung 12**. Es können verschiedene Reports realisiert



werden, von denen beispielhaft die Selektion von «Kunden mit hohem Cross-Selling Potential PV-Anlage» und «PV Pioneer» (Kunden mit hohem Score und geographischer Nähe zueinander) im Bildschirmausschnitt zu sehen sind.

Die Entwicklung und Evaluation des Prototyps erfolgte iterativ in mehreren Feedback- und Optimierungsschleifen und wurde zuletzt mithilfe von Fokusgruppeninterviews in einem Workshop evaluiert.

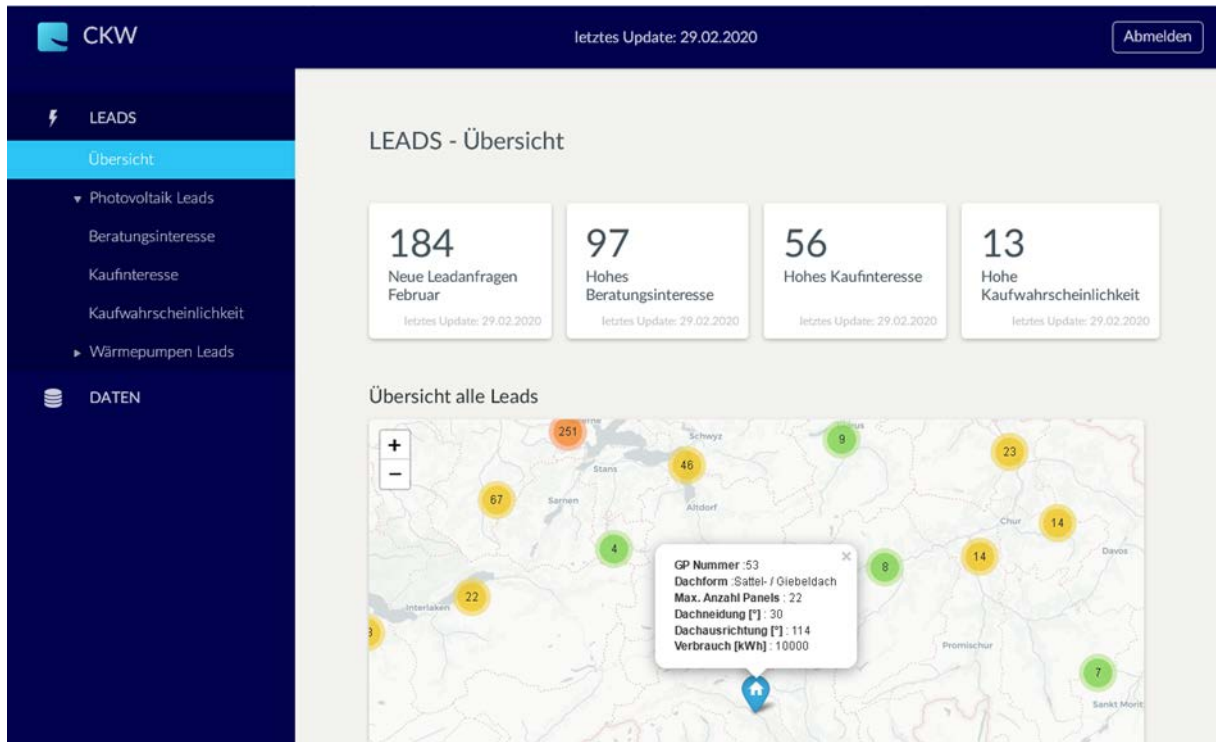


Abbildung 9: Prototyp – Übersicht zu aktuellen Kundenanfragen

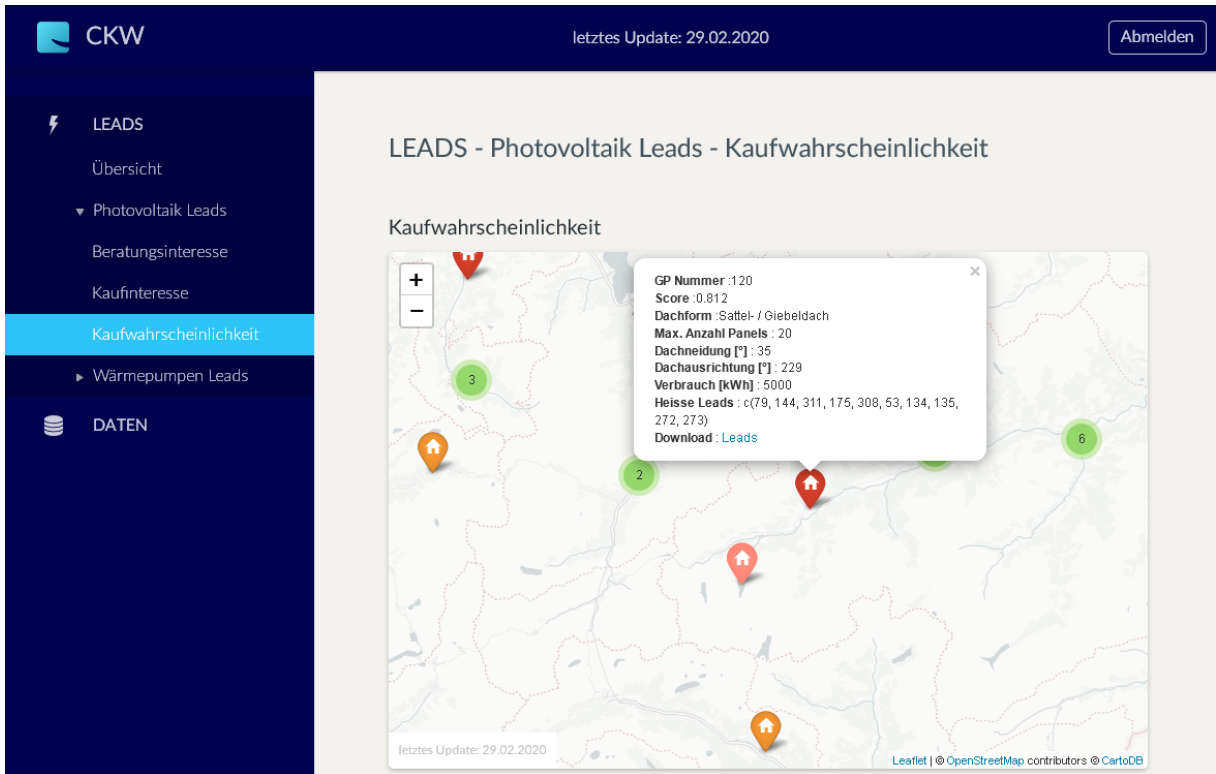


Abbildung 10: Prototyp – Vorhersagescores für die Kaufwahrscheinlichkeit von Photovoltaikanlagen

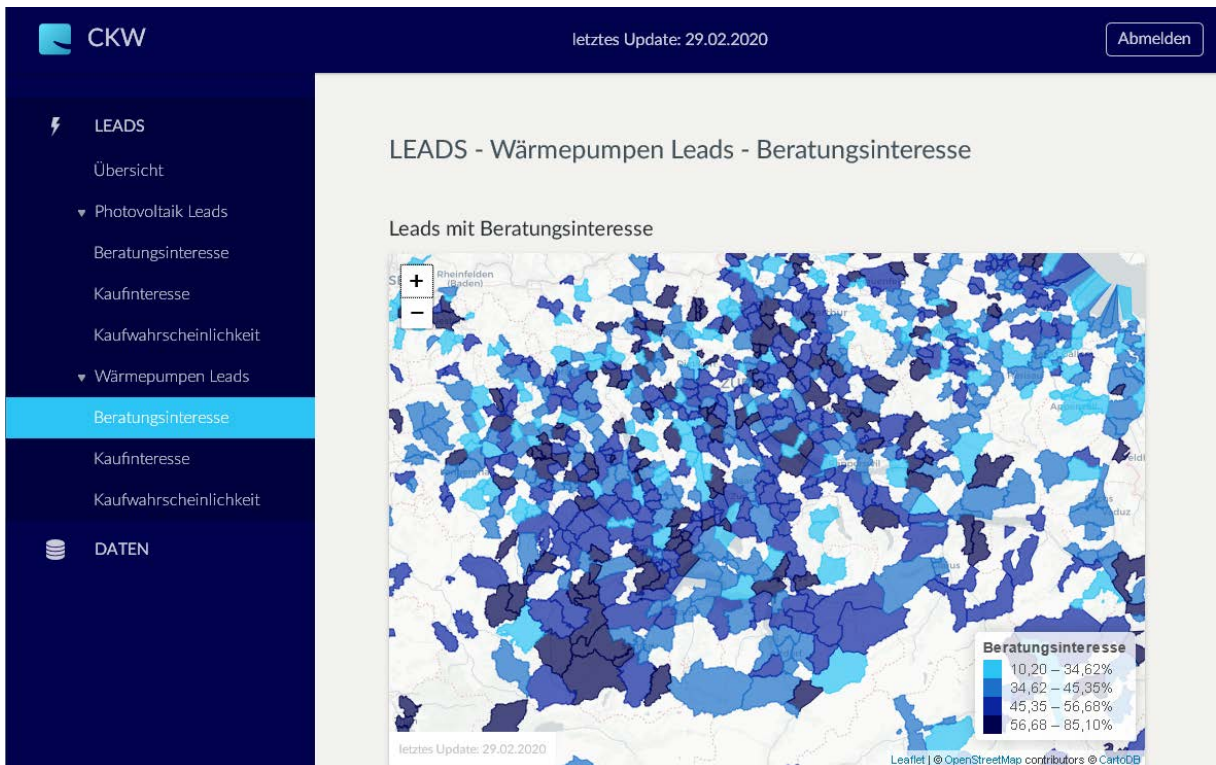


Abbildung 11: Prototyp – Vorhersagescores für die Kaufwahrscheinlichkeit von Photovoltaikanlagen

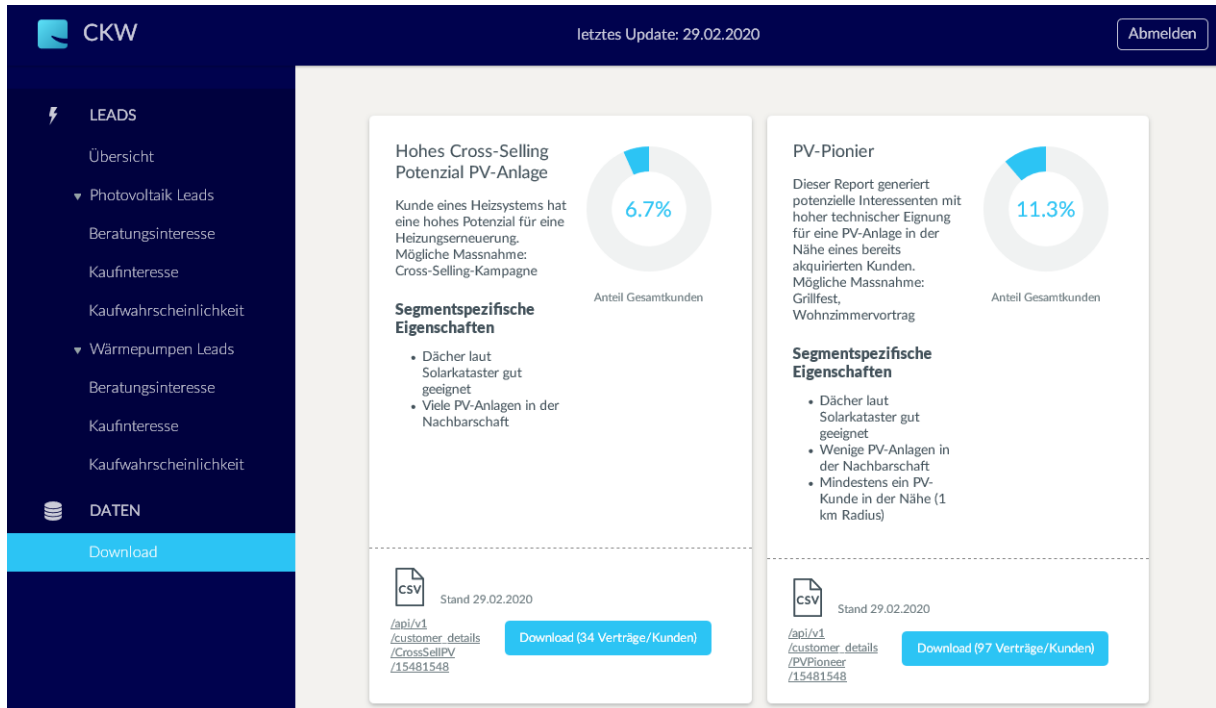


Abbildung 12: Prototyp – Reports zum Download von spezifischen Interessentengruppen

3.6 Diskussion der Ergebnisse

In diesem Teilprojekt wurde untersucht, inwiefern der Vertrieb von erneuerbaren Energieanlagen für den Privatsektor durch Datenanalysen und maschinelles Lernen unterstützt werden kann. Im Detail nutzten wir hierfür Daten aus zwei Onlinerechnern der CKW AG, um die Anwendungsfälle Photovoltaik und Wärmepumpen abzudecken. Mit zwei Umfragen haben wir Daten zum Training erhoben und entwickelten Vorhersagemodelle für drei für den Vertrieb wichtige Variablen *Interesse an Beratung*, *Kaufintention* und *Kaufentscheidung*. Ausserdem testeten wir den Beitrag weiterer Datenquellen für die Vorhersage. Schliesslich implementierte die BEN Energy AG die statischen Modelle in einen Prototyp, welcher als Vertriebsunterstützung durch die CKW AG genutzt werden kann.

Mit den beiden Umfragen konnten wir eine ausreichende Menge an Trainingsdaten sammeln und einen sehr guten Rücklauf von 469 Antworten im Anwendungsfall Photovoltaik und 197 Antworten im Anwendungsfall Wärmepumpe erzielen. Auf Basis der Onlinerechnerdaten können die drei untersuchten Variablen bereits mit einer Güte vorhergesagt werden, die den Vertriebsprozess unterstützen kann. Die Ergebnisse können jedoch durch die Hinzunahme weiterer Datenquellen verbessert werden. Besonders interessant ist, dass ausgewählte Elemente in der Umfrage die Vorhersagemodelle deutlich verbessern können (z. B. Konkurrenzinvestitionen). Diese Daten kann die CKW AG zukünftig einfach in den Onlinerechnern erheben. Auch Daten aus dem Solarkatasterdatensatz (sonnendach.ch) konnten für einzelne Variablen (z. B. Interesse an einer Beratung für eine Wärmepumpe) einen Beitrag leisten. Der umgesetzte Prototyp zeigt exemplarisch, wie die Scores als Ergebnis der Vorhersagemodelle im täglichen Vertrieb eingesetzt werden können. Der entwickelte Prototyp befindet sich nun in der Demonstrationsphase (Technology Readiness Level 5-7) und ist Teil der BEN Energy AG Produktsuite «Kundenmanager». Damit können die Ergebnisse des Forschungsprojekts nun als Dienstleistung in Form einer sog. Platform-as-a-Service auch anderen Installateuren oder Energieversorgern angeboten werden.



Zusammenfassend sehen wir im Geschäftsfeld Vertrieb von erneuerbaren Energieanlagen für EVUs erhebliches Potenzial für erweiterte Datenanalysen, insbesondere da die Daten im Projekt aus wettbewerbsrechtlicher Sicht unkritisch und erzielten Ergebnisse vielversprechend sind.

4 Teilprojekt 3: Datenanalysen für automatisiertes Energieeffizienz-Feedback

Das dritte Teilprojekt betrachtete den Einsatz von maschinellen Lernverfahren für ausgewählte Bereiche des Energieeffizienz-Feedbacks in Haushalten. Hierbei stand die Nutzung von Smart-Meter-Daten für verbessertes Energiefeedback im Zentrum der Untersuchung. Dabei haben wir drei Auswertungen auf Basis verfügbarer Daten durchgeführt: Erstens haben wir untersucht, ob mithilfe von Smart-Meter-Daten auf die Grundlast eines Haushalts geschlossen werden kann. Diese Information kann Endverbrauchern helfen, Stromfresser im Haushalt zu identifizieren und ihren Gesamtstromverbrauch zu reduzieren. Zweitens haben wir evaluiert, inwiefern der Stromverbrauch von grösseren Geräten in Haushalten auf Basis von Smart-Meter-Daten abgeschätzt werden kann, um spezifischeres Feedback zum Stromverbrauch zu geben. Drittens haben wir die Wirkung von Energieeffizienz-Feedback auf Basis von Smart-Meter-Daten auf die Langzeitwirkung hin untersucht.

4.1 Schätzung von Stand-by-Verbrauch und Grundlast auf Basis von Smart-Meter-Daten

Gemäss den Projektzielen haben wir untersucht, inwiefern der Stand-by-Verbrauch von Einzelgeräten und der Grundlast im Haushalt mit typischen Smart-Meter-Daten (1 Messung pro 15 Minuten) geschätzt werden kann.

Für die Schätzung des Stand-by-Verbrauchs mit Smart-Meter-Daten haben wir Literatur zum aktiven Forschungsbereich des Non-Intrusive-Load-Monitoring (Hart 1992; Zeifman und Roth 2011) gesichtet. Wir mussten aber zum Schluss kommen, dass die Schätzung von Stand-by-Verbrauch (also der Betriebszustand des Geräts und dessen Verbrauch in der Stand-by-Zeit) mit den gegebenen Daten nicht möglich ist. Der erste Grund hierfür ist die zu geringe Datengranularität der Smart-Meter-Daten, die zentral an einer Stelle im Haus gemessen werden. Nach Carrie Armel u. a. (2013) sind für die Schätzung der Betriebszustände von Grossverbrauchern (z. B. Kühlschrank oder Klimaanlage) bei dieser Art von Daten hochfrequente Messreihen von mindestens 1 Hz nötig; für Kleinverbraucher ist eine Frequenz im kHz-Bereich notwendig. Zwar können im Allgemeinen Algorithmen auch bei niedriger Messfrequenz den Betriebszustand von Einzelgeräten erkennen, aber dann muss der Verbrauch jedes Geräts einzeln gemessen werden, was bei Smart-Meter-Daten nicht der Fall ist. Der zweite Grund, weshalb wir keine Schätzung des Stand-by-Verbrauchs durchführen konnten, war der Mangel an Daten, denn kein Datensatz mit entsprechenden Betriebszuständen war in geeignetem Umfang verfügbar.

Für die Grundlast-Schätzung betrachteten wir die Summe aller Verbräuche von Geräten, die durchgängig im Betriebszustand «an» sind, also sogenannte Always-on-Geräte (z. B. Kühlschrank, Umwälzpumpe oder WiFi), sowie diejenigen, die sich im Zustand «Stand-by» befinden. Im Rahmen des Projekts haben wir einen Ansatz entwickelt, der die Schätzung von Grundlast für Haushalte mithilfe von Smart-Meter-Daten ermöglicht (siehe Abbildung 13). Das Verfahren wird zeitnah in einer Publikation veröffentlicht.

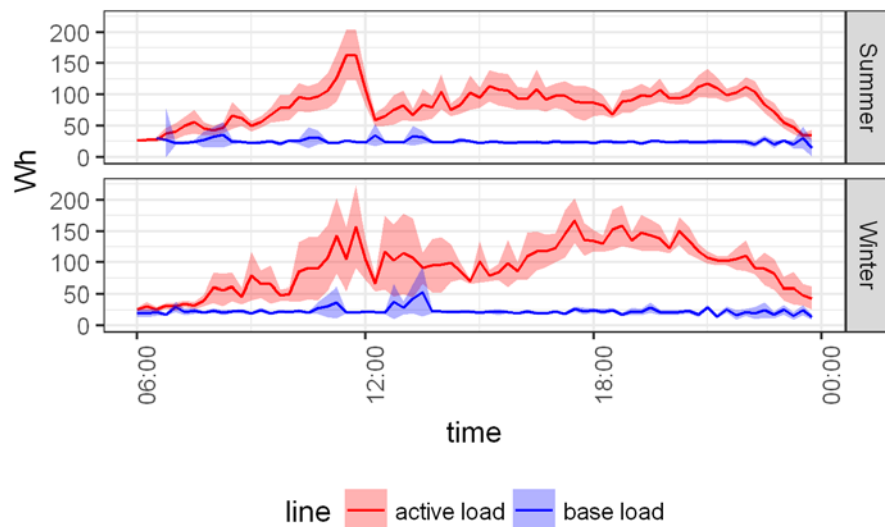


Abbildung 13: Diese Abbildung zeigt die durchschnittliche Grundlastkurve (base load) eines Haushalts in blau, sowie die Aktivlastkurve (active load) in rot. Der schattierte Bereich um die Kurven zeigt das 95 % Konfidenzintervall der zu erwartenden Abweichung auf Basis einer Student-t-Verteilung.

4.2 Auswertung von Energieeffizienz-Eigenschaften der Haushalte

Neben dem Stromverbrauchs-Feedback mit aggregierten Verbrauchswerten, wie wir es in der vorhergehenden Analyse untersucht haben, zielt diese Auswertung darauf ab, den Stromverbrauch auf einzelne grosse Verbraucher herunterzubrechen. Im Projektgesuch war deshalb angedacht, Verbrauchszeiten grosser Geräte, die Schätzung der Effizienz von Heizungs- und Raumklimaanlagen, sowie die Schätzung der Effizienz der Beleuchtung im Haushalt sowie die Schätzung der Effizienz von Haushalten durchzuführen. Im Projektverlauf, insbesondere bei der Entwicklung der ersten Umfrage mit allen Projektpartnern und Pre-Tests der Fragebögen haben wir festgestellt, dass die Erhebung der Daten, welche für diese Analysen notwendig sind, kaum mit Daten erfolgen kann, die in Umfragen erhoben werden. Der Grund hierfür ist, dass die Fragen entweder sehr speziell sein müssten und der Fragebogen dadurch sehr umfangreich sein würde, andererseits ist die Angabe solcher Effizienzeigenschaften durch Verbraucher sehr subjektiv und ohne spezifische Einzelgerätemessungen kaum überprüfbar. Wir haben uns deshalb entschieden, einen vorhandenen Datensatz aus Nordamerika zu verwenden, der Einzelmessungen und Umfragedaten für eine grosse Zahl an Haushalten beinhaltet.

Um den Einfluss einzelner Geräte und Verbrauchsgruppen trotzdem zu analysieren, haben wir auf einen existierenden und öffentlich verfügbaren Datensatz mit Smart-Meter-Daten aus Nordamerika zurückgegriffen. Dieser Datensatz, veröffentlicht von *Pecan Street Inc.* (Glasgo, Hendrickson, und Azevedo 2017) enthält Daten von 323 Haushalten im Jahr 2014. Speziell an diesem Datensatz ist, dass sowohl Verbrauchsdaten von Häusern und Wohnungen auf der Gesamtverbrauchs- als auch auf Einzelverbraucherebene erfasst sind. Der Datensatz enthält zudem Informationen aus Umfragen der Bewohner und Energie-Audits.

Wir haben die Verbrauchsmessungen auf den Anteil der Verbrauchswerte der Einzelmessungen am Gesamtstromverbrauch mit einer Varianzanalyse untersucht, um den Einfluss einzelner Geräte und Verbrauchsgruppen auf den Gesamtstromverbrauch zu schätzen. Dabei wird der Einfluss von nominalskalierten Variablen (hier die Existenz eines Zählers je Verbraucher oder Verbrauchskategorie, binär, siehe Tabelle 9) auf den Stromverbrauch des Haushalts berechnet. Abbildung 14 zeigt die Ergebnisse der Berechnung und gibt den Wert mit einem 90% Konfidenzintervall an. Der Übersichtlichkeit



halber sind die Einflüsse in Kategorien eingeteilt, wurden jedoch als Gesamtmodell berechnet. Der *intercept* beschreibt den y-Achsenabschnitt, der in der vorliegenden Anwendung dem Anteil des nicht gemessenen Stromverbrauchs durch Einzelmessgeräte darstellt, also der Kategorie *other*.

Tabelle 9: Verbraucher bzw. Verbrauchergruppen im Dataport Datensatz, in Kategorien für die Auswertung zusammengefasst

Variablenname (Bezeichnung)	Beschreibung	Häufigkeit im Jahr 2014
Kategorie: Heizung und Raumklimatisierung		
air1	Raumklimaanlage	251
air2	Raumklimaanlage	31
air3	Raumklimaanlage	3
airwindowunit1	Split-Klimaanlage (Fenstereinheit)	2
furnace1	Zentralheizung	203
furnace2	Zentralheizung	20
heater1	Stand-Alone-Heizung	1
housefan1	Lüftungsschacht	1
waterheater1	Boiler	46
waterheater2	Boiler	2
Kategorie: Küchengeräte		
dishwasher1	Spülmaschine	170
disposal1	Spülbeckenhäcksler f. Abfallreste	98
freezer1	Gefrierschrank	6
icemaker1	Eiswürfelbereiter	2
kitchen1	Stromkreis Küche	40
kitchen2	Stromkreis Küche	13
kitchenapp1	Stromkreis für Küchenkleingeräte	122
kitchenapp2	Stromkreis für Küchenkleingeräte	77
microwave1	Mikrowelle	104
oven1	Ofen	69
oven2	Ofen	2
range1	Herd (mit oder ohne Ofen)	74
refrigerator1	Kühlschrank	175
refrigerator2	Kühlschrank	9
venthood1	Abzugshaube	10
winecooler1	Weinkühler	2
Kategorie: Beleuchtung und Kleingeräte		
bathroom1	Stromkreis Bad	45
bathroom2	Stromkreis Bad	4
bedroom1	Stromkreis Schlafzimmer	57
bedroom2	Stromkreis Schlafzimmer	24
bedroom3	Stromkreis Schlafzimmer	4
diningroom1	Stromkreis Esszimmer	13
lights_plugs1	Steckdosen und Lampen allgemein	65
lights_plugs2	Steckdosen und Lampen allgemein	30
lights_plugs3	Steckdosen und Lampen allgemein	9
lights_plugs4	Steckdosen und Lampen allgemein	4
lights_plugs5	Steckdosen und Lampen allgemein	2
lights_plugs6	Steckdosen und Lampen allgemein	0
livingroom1	Stromkreis Wohnzimmer	74
livingroom2	Stromkreis Wohnzimmer	5
office1	Stromkreis Büro	20
outsidelights_plugs1	Aussenbeleuchtung	13
outsidelights_plugs2	Aussenbeleuchtung	1
pump1	Jeglicher Pumpentyp, ausser Poolpumpen	1
security1	Sicherheitssystem	4
utilityroom1	Stromkreis Abstellraum	3
Kategorie: Kategorie: Aussenbereich und Pool		
garage1	Stromkreis Garage	22

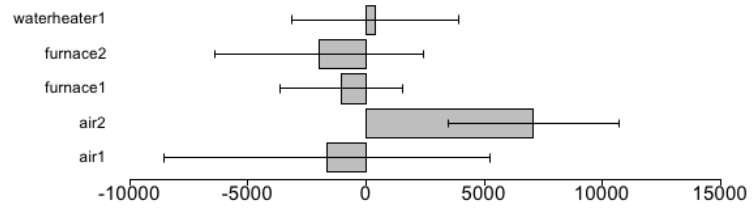


garage2	Stromkreis Garage	2
shed1	Stromkreis Gartenhütte	3
sprinkler1	Stromkreis Gartensprenger	6
Kategorie: Pool		
jacuzzi1	Whirlpool	12
pool1	Kombination aus Poolpumpe und/oder Poolequipment	3
poollight1	Poolbeleuchtung	3
poolpump1	Poolpumpe	19
Kategorie: Kategorie: Waschen, Trocknen, usw.		
clotheswasher1	Waschmaschine	106
clotheswasher_dryg1	Waschmaschine	21
drye1	Trockner	103
dryg1	Trockner	26

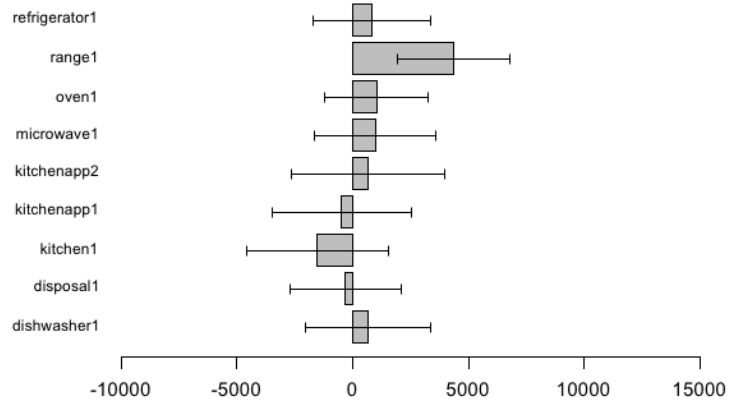
Es ist zu erkennen, dass Grossverbraucher (z. B. Poolpumpe, besondere Typen von Raumklimaanlagen) einen hohen Einfluss auf den Gesamtstromverbrauch haben. Die genaue Höhe dieses Einflusses ist jedoch nicht exakt abzuschätzen (dies erkennt man am breiten Konfidenzintervall), da die Verbrauchswerte stark schwanken und für die einzelnen Gerätekategorien zum Teil nur wenige Datenpunkte verfügbar sind. Zudem ist bei diesen Ergebnissen zu berücksichtigen, dass die Daten aus Nordamerika kaum den Gegebenheiten von Haushalten in Mitteleuropa entsprechen und insbesondere die Schätzung grosser Verbraucher stark unterschiedlich sein wird. Letztendlich müssten für solch eine Auswertung spezifischere Daten in entsprechender Menge erhoben werden, um belastbare Aussagen abzuleiten. Dies war aufgrund der Schwerpunktsetzung des Projekts nicht möglich.



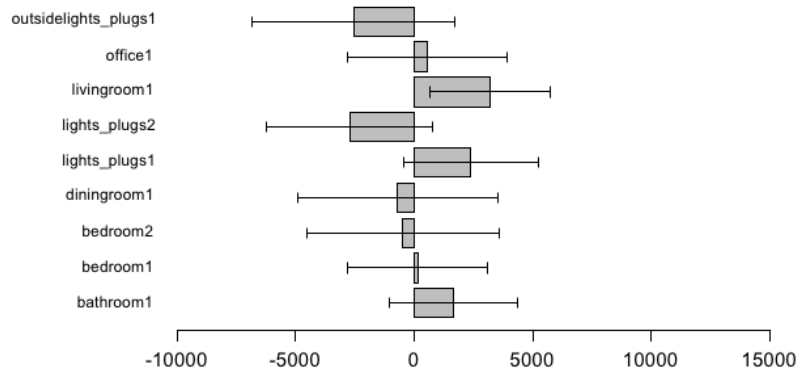
a) Kategorie: Heizung und Raumklimatisierung



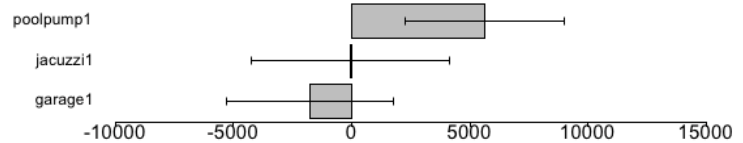
b) Kategorie: Küchengeräte



c) Kategorie: Beleuchtung und Kleingeräte



d) Kategorie: Aussenbereich und Pool



e) Kategorie: Waschen, Trocknen, usw.

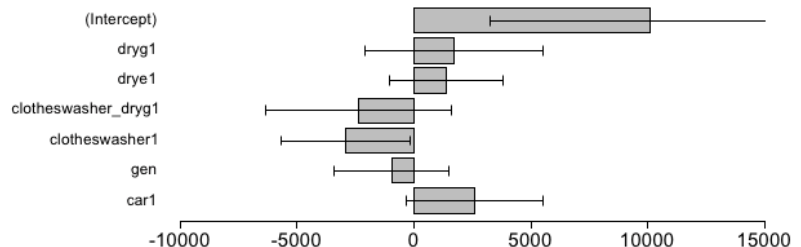


Abbildung 14: Einfluss verschiedener Verbraucher oder Verbrauchskategorien auf den gesamten Stromverbrauch, mit einem 90% Konfidenzintervall, ermittelt mit einer Varianzanalyse

4.3 Experiment zum Langzeiteffekt von Energieeffizienz-Mailings

Für das Anwendungsfeld Energieeffizienz haben wir die Wirkung einer langlaufenden Energieeffizienzkampagne untersucht, bei dem Verbraucher Feedback auf ihren Smart-Meter-Stromverbrauch



per E-Mail erhalten. Energieeffizienz-Mailings für Privatkunden können automatisiert mit Daten, die beim Energieversorger vorliegen, realisiert werden. Solche Mailings sind ein wirkungsvoller Baustein im Bereich der Energieeffizienz (Tiefenbeck 2014; 2017). In früheren Arbeiten der ETH Zürich, der Universität Bamberg und der BEN Energy AG wurden Technologien entwickelt, die automatisierte Energieeffizienz-Mailings für Privatkunden ermöglichen (Sodenkamp u. a. 2016; Hopf und Staake 2019). Der Effekt solcher personalisierter Energieeffizienz-Mailings wurde bereits festgestellt (Hopf 2019), allerdings nur über einen kurzen Studienzeitraum.

Die lange Laufzeit dieser Kampagne erlaubt es uns nun in diesem Projekt zu untersuchen, ob der positive Effekt von Energiefeedbackintervention auch über einen längeren Zeitraum anhält. Im Folgenden wird zunächst der Status quo der bereits existierenden Lösung vorgestellt. Anschliessend werden einerseits die Ergebnisse der Untersuchung der Langzeiteffekt des Mailings als solches (Öffnungs- und Klickraten) untersucht, andererseits der Energieeinspareffekt der Intervention.

4.3.1 Status quo Energieeffizienz-Mailings

BEN Energy AG hat ein Energieeffizienzportal für einen Schweizer Energieanbieter entwickelt. Im Rahmen dessen erhalten alle Nutzer quartalsweise Energiereports per E-Mail. Der Report enthält Informationen zum Stromverbrauch, Verhaltensfeedback und Energiespartipps. Bei dem Design des Energieberichts wurde Wert daraufgelegt, dass die E-Mail dem Kunden einen anschaulichen Überblick über den Verbrauch der vergangenen Periode gibt, der Verbrauchstrend angezeigt wird und der Verbrauch mit der Nachbarschaft verglichen wird. Der Bericht enthält auch eine Bewertung der Energieeffizienz des Haushalts auf Basis von Haushalten mit ähnlichen Merkmalen, sofern der Kunde zuvor Angaben zu seinem Haushalt im Online-Portal hinterlegt hat.

Die Wirkung der Energiereports wurde bei der Einführung im April 2017 von einem Begleitforschungsprojekt untersucht, dessen Ergebnisse in Hopf (2019) dokumentiert sind. BEN Energy AG hat seitdem weitere Daten über die Nutzung dieser Energiereports von 969 teilnehmenden Schweizer Haushalten erhoben, die nachfolgend im Hinblick auf den Langzeiteffekt der Mailings ausgewertet werden. Ein Überblick über den Studienzeitraum, die Versanddaten der Energieberichte und die Kundenbefragung ist in Abbildung 15 dargestellt.

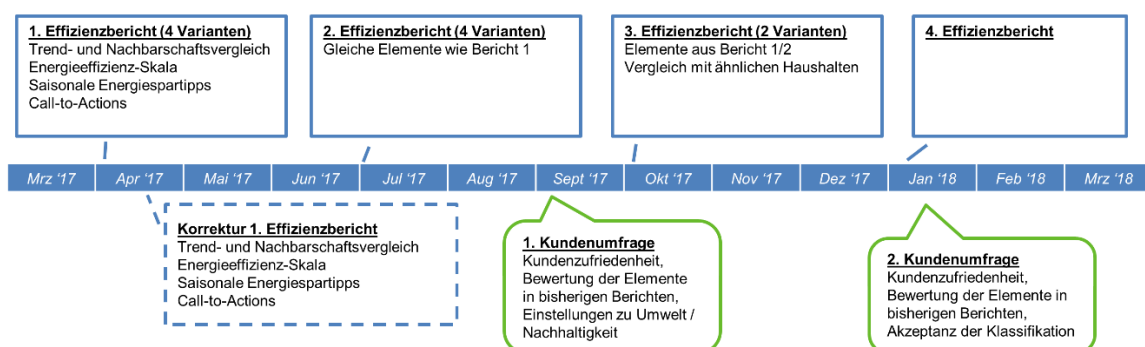


Abbildung 15: Zeitlicher Ablauf des Experiments im Zeitraum 2017-2018

In Vorbereitung auf das Experiment wurden alle Haushalte, die sich in dem Energieeffizienzportal registriert hatten, in Gruppen eingeteilt. Da allerdings im Laufe des Langzeitexperiments neue Nutzer hinzugekommen sind, hat sich die Gruppeneinteilung später verändert. Tabelle 10 zeigt die Gruppenzuweisungen. Davon wurden die Experimentgruppen A und B zu Beginn des Experiments eingeteilt, die Gruppen C und X konnten durch das Forscherteam nicht beeinflusst werden.



Den Energiereport erhielten nur Haushalte mit einem täglichen Stromzähler (n=414 zu Beginn, n=463 am Ende), da der quartalsweise Bericht bei jährlichen Verbrauchsdaten keine regelmässig neuen Informationen enthalten würde. Haushalte mit jährlichen Zählerständen (n=543) stehen als Vergleichsgruppe für mehrere Analysen zur Verfügung. Eine zufällig ausgewählte Gruppe von Kunden erhielt das Feedback zum Energieverbrauch zu Beginn unseres Experiments im April (n=263), nachfolgend Gruppe A genannt. Die übrigen Haushalte mit installierten Tageszählern erhielten im Juli und Oktober ihren ersten Energiebericht (n=151), nachfolgend Gruppe B.

Tabelle 10: Experimentgruppen und deren Stichprobengrössen im Zeitraum 2017-2018

Nutzergruppe		Experiment Gruppe	Datum des ersten Berichts	Haushaltseigenschaften		
				Angegeben	Nicht angegeben	Gesamt
Portal-nutzer (n = 1'006)	Tägliche Zähler (n = 463)	A	April 2017	157	106	263
		B	Oktober 2017	76	75	151
		C	Neue Nutzer (nach Oktober 2017)	26	23	49
	Jährliche Zähler (n = 543)	X	(kein Bericht)	254	289	543

4.3.2 Ergebnisse der Langzeitstudie

Auf Basis ursprünglichen Ergebnisse zu den Energieeffizienz-Mailings werden im Folgenden zunächst die Ergebnisse der Untersuchung der Langzeiteffekt des Mailings als solches (Öffnungs- und Klickraten), anschliessend die Energieeinspareffekte der Intervention vorgestellt.

In den Jahren 2017 und 2018 wurden insgesamt acht Energiereports an die Haushalte versendet. Die Zeitpunkte, die Anzahl an Berichten und Statistiken über die Resonanz von den Nutzern zeigt Tabelle 11. Im Vergleich zu 2017 hat die Rate, wie oft Nutzer den E-Mail-Bericht öffneten oder auch den Link in der E-Mail anklickten, leicht abgenommen. Allerdings erfolgte die Versendung im Jahr 2018 an eine deutlich grössere Nutzerzahl. Aus diesen beiden Tatsachen lässt sich folgern, dass das Mailing nachhaltig von hohem Interesse für die teilnehmenden Haushalte ist. Auch die Stunden bis zur Öffnung, bzw. des ersten Klicks im Jahr 2018 bestätigen diese Annahme. Insgesamt weisen die E-Mails selbst nach zwei Jahren Laufzeit eine Öffnungsrate von rund 66% und eine Klickrate von rund 22% auf. Im Vergleich weisen typische Marketingkampagnen lediglich Öffnungsraten von ca. 20-40%, und Klickraten von ca. 2% auf (Campaign Monitor 2019; mailchimp 2018; EPSILON 2018). Das hier vorgestellte Ergebnis zeigt, dass das Energieeffizienzmailing hinsichtlich der Metriken Öffnungs- und Klickrate auch nach nunmehr zwei Jahren weit bessere Ergebnisse als typische Marketingkampagnen aufzeigen.

Tabelle 11: Kundenreaktionen auf die Energieberichte

Datum	Energiereport	Anzahl E-Mails	Öffnungsrate	Klickrate	Median Stunden Öffnung	Median Stunden erster Klick
2017-04-04	Newsletter Tuesday	66	65,2%	31,8%	3,410	4,060
2017-04-05	Newsletter Wednesday	66	75,8%	37,9%	4,930	5,910
2017-04-06	Newsletter Thursday	65	76,9%	30,8%	2,550	8,320
2017-04-07	Newsletter Friday	68	61,8%	32,4%	3,100	5,530
2017-04-26	Trend Correction	195	70,8%	19,5%	2,130	2,610
2017-07-06	Newsletter	289	69,6%	23,2%	3,300	4,500
2017-10-04	Newsletter	417	63,1%	18,2%	3,920	7,850
2018-01-02	Newsletter	400	65,5%	14,5%	3,630	13,230
2018-04-03	Newsletter	391	48,6%	11,0%	4,590	8,090
2018-07-03	Newsletter	400	64,5%	13,7%	2,300	5,940
2018-10-02	Newsletter	383	59,3%	10,2%	4,290	7,130



Die Auswirkungen der Energieeffizienzberichte auf den Stromverbrauch im Zeitraum 2017-2018 wurden bereits in Hopf (2019) dargestellt. Nachdem alle Portalnutzer mit täglichem Zähler ab Oktober 2018 einen Energiereport erhalten, steht keine Vergleichsgruppe für die Zeitreihenanalyse zur Verfügung. Betrachtet man aber die übrigen Analysen, kann das Energieeffizienz-Mailing auch in der Langzeitbetrachtung als erfolgreich bewertet werden.

5 Rechtliche Rahmenbedingungen

5.1 Datenschutz

Für das Projekt wurde zwischen allen Partnern ein Datenschutzkonzept erarbeitet. Dieses Datenschutzkonzept beschreibt die Schritte der Datenerhebung, -verarbeitung und -löschung. Alle im Projekt verwendeten Daten werden lediglich in pseudonymisierter Form verarbeitet und können nicht zu natürlichen Personen zugeordnet werden. Die CKW AG hat alle Daten vor der Übermittlung pseudonymisiert (Task 12.2). In Einzelfällen wurden in Umfragen zweckgebunden personenbezogene Daten erhoben (E-Mail-Adresse für Gewinnspiel oder Anfrage an die CKW AG). In diesen Fällen wurde explizit das Einverständnis des Teilnehmers eingeholt und die Daten wurden lediglich für den angegebenen Zweck verwendet.

5.2 Unbundling

Für das Projekt sind insbesondere die Bestimmungen der Entflechtung nach Art. 10 StromVG sowie Art. 7 des Kartellgesetzes in der Schweiz zu berücksichtigen. Die involvierten Projekt-Mitarbeitenden sind sich dieser Thematik bewusst und haben entsprechende Geheimhaltungserklärungen unterzeichnet. Für das Forschungsprojekt wurden z. B. Smart-Meter-Daten verwendet, die nicht ohne weiteres für Marktzwecke verwendet werden dürfen. Sollen sie dereinst für Marktzwecke verwendet werden, so muss der Kunde das EVU explizit für diese Zwecke bevollmächtigen. Eine alternative Lösung kann sein, dass diese Daten direkt über die Kundenschnittstelle des Smart-Meters bezogen werden, was ebenfalls die Zustimmung des Kunden benötigt.

Für das Forschungsprojekt wurde eine Lösung der Datennutzung erarbeitet, die sowohl die Ziele des Forschungsprojekts als auch die Ziele der beteiligten Industriepartner berücksichtigt. Die Lösung wurde mit dem Bundesamt für Energie (Schweiz) abgestimmt und schriftlich geregelt.

6 Schlussfolgerungen und Fazit

Ziel des Projekts war es, Datenanalyseverfahren weiterzuentwickeln, um die in absehbarer Zeit flächendeckend zur Verfügung stehenden Daten bei EVUs (z. B. Smart-Meter-Stromverbrauchsdaten und Transaktionsdaten digitaler Geschäftsprozesse) für die Steigerung der Servicequalität und der Energieeffizienz im Privatsektor nutzbar zu machen. Um den Wert solcher Daten zu heben, werden moderne Analysemethoden – zum Beispiel Verfahren des maschinellen Lernens – benötigt. Zwar hat die Forschung in den letzten Jahrzehnten eine Reihe von effektiven Datenanalyseverfahren für den generellen Einsatz entwickelt, aber es fehlt nach wie vor (insbesondere in Unternehmen) das Verständnis dafür, wie die Analytik auf konkrete Anwendungen der Energiedatenanalyse anzuwenden ist.

Konkret hat das Projekt «SmartLoad» untersucht, inwiefern verfügbare Daten bei Energieanbietern genutzt werden können, um Haushalte zu erkennen, die a) bereit sind, nachhaltige Stromprodukte zu nutzen und die sich b) für die Investition in Selbstversorgungs- und Speicherlösungen interessieren.



Darüber hinaus hat das Projekt untersucht, wie c) Smart-Meter-Daten das Stromverbrauchs-Feedback unterstützen können, um die Energieeffizienz im Privatsektor zu erhöhen. Das Projektkonsortium hat hierzu in drei Online-Umfragen detaillierte Daten von insgesamt 1'373 privaten Haushalten erhoben, mit diesen Daten maschinelle Lernverfahren trainiert und Vorhersagemodelle für das Kaufinteresse für lokalen Ökostrom, die Investition in Photovoltaikanlagen und die Anschaffung von Wärmepumpen entwickelt. Das Modell für den Kauf von lokal produziertem Ökostrom wurde in einem Experiment mit 4'712 Privatkunden evaluiert. Die Modelle zur Prädiktion der Investitionsbereitschaft in Photovoltaikanlagen und Wärmepumpen ermöglichten die Entwicklung eines Softwareprototypen, der den Vertrieb von nachhaltigen Technologien unterstützt.

Die Erkenntnisse des Projekts liefern Energieanbietern wichtige Hinweise, unter anderen für welche Aufgaben des Vertriebs von nachhaltigen Produkten maschinelle Lernverfahren eingesetzt werden können, welche externen Datenquellen die Kaufvorhersagen verbessern, wie geeignete Trainingsdaten für die Modellentwicklung gesammelt werden sollten und wie die entwickelten Vorhersagemodelle in eine Anwendung für den Vertriebsmitarbeiter überführt werden können. Die entwickelten Verfahren als auch deren Anwendung stehen im Einklang mit der EU-Datenschutzgrundverordnung, den schweizerischen Datenschutzvorschriften und den Vorschriften der Entflechtung und des Kartellrechts in der Schweiz.

7 Ausblick und zukünftige Umsetzung

Die Ergebnisse aus dem Forschungsprojekt dienen der CKW AG zur Verbesserung der Produkte und der Servicequalität, werden nach Projektabschluss durch BEN Energy AG und wissenschaftlich durch den Lehrstuhl für Wirtschaftsinformatik, insb. Energieeffiziente Systeme an der Otto-Friedrich-Universität Bamberg verwertet.

Wir erwarten, dass die entwickelten Algorithmen in beiden Unternehmen des Verbundprojekts wirtschaftlichen Nutzen erreichen: Bei der CKW AG können die Marketingkosten eines neuen lokalen Ökostromprodukts reduziert werden und der Vertriebsprozess bei Photovoltaik und Stromspeichern signifikant verbessert werden. BEN Energy wird die Algorithmen in bestehenden Analytics- und Smart-Metering-Produkten implementieren und als Software-as-a-Service Dienstleistungen verschiedenen Energieversorgern zur Verfügung stellen. Bei der wirtschaftlichen Nutzung werden jeweils die Bestimmungen aus Datenschutz- und Wettbewerbsrecht gewahrt.

Der entwickelte Prototyp im Teilprojekt 2 befindet sich nun in der Demonstrationsphase (Technology Readiness Level 5-7) und ist Teil der BEN Energy AG Produktsuite «Kundenmanager». Damit können die Ergebnisse des Forschungsprojekts nun als Dienstleistung in Form einer sogenannten Plattform-as-a-Service auch anderen Energieversorgern und Organisationen angeboten werden, insofern sie einen entsprechenden Onlinerechner betreiben und diese Daten von Interessenten strukturiert erheben.

Wissenschaftliche Ergebnisse werden zeitnah durch die Universität Bamberg in wissenschaftlichen Zeitschriften und Konferenzen veröffentlicht. Das aus dem Projekt resultierende Know-how in den Bereichen Dimensionsreduktion, maschinelle Lernverfahren und deren Anwendung wird auch Basis für weitere Forschungsprojekte sein. Die erreichten Erkenntnisse werden auch in der Lehre an Studierende weitergegeben und werden in folgenden Lehrveranstaltungen integriert:

- *Data Analytics in der Energieinformatik*: 6 ECTS Kurs auf Master-Niveau (Inhalte: Analysemethoden der Energieinformatik und deren Anwendung auf praktisch relevante Aufgabenstellungen, um wirkungsvolle Energiedienstleistungen zu entwickeln.)



- *Business Intelligence & Analytics*: 6 ECTS Kurs auf Master-Niveau (Inhalte: Fundamentale Konzepte und Methoden der Datenanalyse und der modernen Entscheidungstheorie und -praxis. Die Modulschwerpunkte liegen auf prädiktiven und präskriptiven Analysen, welche Unternehmen zu einer besseren Einsicht in Prozesse und Entscheidungen verhelfen.)
- *Smart Grid Data Analytics*: 6 ECTS Projekt für Bachelor- und Master-Studierende mit wechselnden Themen der Datenanalyse in der Energieinformatik.
- *Business Intelligence for Renewable Energy Systems*: 6 ECTS Projekt für Master-Studierende mit wechselnden Themen der Datenanalyse in der Energieinformatik.
- *Abschluss- und Seminararbeiten*

8 Nationale und internationale Zusammenarbeit

8.1 Nationale Zusammenarbeit

Das Projekt wurde in Zusammenarbeit mit der CKW AG und der BEN Energy AG durchgeführt. Die CKW AG hat alle erforderlichen anonymisierten Daten bereitgestellt. Alle Industriepartner brachten sich umfassend mit allen erforderlichen personellen Ressourcen ein.

8.2 Internationale Zusammenarbeit

Das Projekt erfolgte in Zusammenarbeit mit der Otto-Friedrich-Universität Bamberg, Lehrstuhl für Energieeffiziente Systeme. Im Projektkonsortium fanden wöchentliche Telefonkonferenzen statt, in denen die Projektplanung, der aktuelle Stand der Arbeiten, sowie Projektergebnisse vorgestellt und besprochen wurden. Zudem haben im Projekt mehrere Workshops stattgefunden.

Das Projekt wurde ausserdem auf folgenden Veranstaltungen des Förderrahmenprogramms ERA-Net SES (früher: Smart Grids+) vertreten:

- 07. – 08.06.2017 in Bukarest
- 23. – 25.05.2018 in Malmö
- 17. – 19.09.2018 in Magdeburg
- 07. – 09.10.2019 in Namur

9 Publikationen

Es befinden sich derzeit mehrere Beiträge für wissenschaftliche Konferenzen oder Fachzeitschriften in Vorbereitung bzw. Begutachtung. Die Veröffentlichung dieser Beiträge wird nach Projektabschluss erfolgen.



10 Literaturverzeichnis

- Ajzen, Icek. 1985. «From Intentions to Actions: A Theory of Planned Behavior». In *Action Control*, herausgegeben von Julius Kuhl und Jürgen Beckmann, 11–39. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69746-3_2.
- Armel, K. Carrie, Abhay Gupta, Gireesh Shrimali, und Adrian Albert. 2013. «Is disaggregation the holy grail of energy efficiency? The case of electricity». *Energy Policy*, Special Section: Transition Pathways to a Low Carbon Economy, 52 (Supplement C): 213–34. <https://doi.org/10.1016/j.enpol.2012.08.062>.
- Breiman, Leo. 2001. «Random forests». *Machine Learning* 45 (1): 5–32.
- Campaign Monitor. 2019. «CTOR – the Email Marketing Metric You May Not Know». *BLOG - EMAIL MARKETING* (blog). 15. Januar 2019. <https://www.campaignmonitor.com/blog/email-marketing/2019/01/ctor-email-marketing-metrics-you-may-not-know>.
- Cui, Geng, Man Leung Wong, und Xiang Wan. 2012. «Cost-Sensitive Learning via Priority Sampling to Improve the Return on Marketing and CRM Investment». *Journal of Management Information Systems* 29 (1): 341–74.
- Dietz, Thomas, Gerald T. Gardner, Jonathan Gilligan, Paul C. Stern, und Michael P. Vandenbergh. 2009. «Household Actions Can Provide a Behavioral Wedge to Rapidly Reduce US Carbon Emissions». *Proceedings of the National Academy of Sciences of the United States of America* 106 (44): 18452–56.
- Energiegesetz. 2018. <https://www.admin.ch/opc/de/classified-compilation/20121295/index.html>.
- EPSILON. 2018. «Email trends and benchmarks remain steady in Q1 2018». *BLOG* (blog). 2. August 2018. <https://us.epsilon.com/a-brand-new-view/email-trends-and-benchmarks-remain-steady-in-q1-2018>.
- Eurostat. 2017. «Final consumption expenditure of households, by consumption purpose - Eurostat (Code: tsdpc520, Last update: 25/01/17)». 25. Januar 2017. <http://ec.europa.eu/eurostat/web/products-datasets/-/tsdpc520>.
- Fawcett, T. 2006. «An introduction to ROC analysis». *Pattern Recognition Letters* 27 (8): 861–74.
- Fernández-Delgado, Manuel, Eva Cernadas, Senén Barro, und Dinani Amorim. 2014. «Do we need hundreds of classifiers to solve real world classification problems?» *The Journal of Machine Learning Research* 15 (1): 3133–81.
- Glasgo, Brock, Chris Hendrickson, und Inês M. L. Azevedo. 2017. «Using advanced metering infrastructure to characterize residential energy use». *The Electricity Journal* 30 (3): 64–70. <https://doi.org/10.1016/j.tej.2017.03.004>.
- Hansla, André, Amelie Gamble, Asgeir Juliusson, und Tommy Gärling. 2008. «Psychological Determinants of Attitude towards and Willingness to Pay for Green Electricity». *Energy Policy* 36 (2): 768–74. <https://doi.org/10.1016/j.enpol.2007.10.027>.
- Hart, George Wiliam. 1992. «Nonintrusive appliance load monitoring». *Proceedings of the IEEE* 80 (12): 1870–91. <https://doi.org/10.1109/5.192069>.
- Hopf, Konstantin. 2019. *Predictive Analytics for Energy Efficiency and Energy Retailing*. 1. Aufl. Bd. 36. Contributions of the Faculty Information Systems and Applied Computer Sciences of the Otto-Friedrich-University Bamberg. Bamberg: University of Bamberg. <https://doi.org/10.20378/irbo-54833>.
- Hopf, Konstantin, und Thorsten Staake. 2019. «Methoden der Energiedatenanalyse». Schlussbericht zum Vorhaben im Rahmen des Eurostars Projekts E! 9859 BEnginell in Veröffentlichung. Bamberg: Universität Bamberg.
- Jäger-Waldau, Arnulf. 2018. «PV Status Report 2018». Text JRC 113626. Joint Research Centre (European Commission). <https://ec.europa.eu/jrc/en/publication/pv-status-report-2018>.
- Kaenzig, Josef, und Rolf Wüstenhagen. 2008. «Understanding the Green Energy Consumer». *Marketing Review St. Gallen* 25 (4): 12–16. <https://doi.org/10.1007/s11621-008-0057-3>.
- Klauser, Daniel, und Thomas Schlegel. 2016. «Sonnendach.ch – Datenmodell». Dokumentation v 1.4. Bern: Bundesamt für Energie.



- Kotchen, Matthew J., und Michael R. Moore. 2007. «Private Provision of Environmental Public Goods: Household Participation in Green-Electricity Programs». *Journal of Environmental Economics and Management* 53 (1): 1–16. <https://doi.org/10.1016/j.jeem.2006.06.003>.
- Loureiro, A. L. D., V. L. Miguéis, und Lucas F. M. da Silva. 2018. «Exploring the use of deep neural networks for sales forecasting in fashion retail». *Decision Support Systems* 114 (Oktober): 81–93. <https://doi.org/10.1016/j.dss.2018.08.010>.
- MacPherson, Ronnie, und Ian Lange. 2013. «Determinants of green electricity tariff uptake in the UK». *Energy Policy* 62 (Supplement C): 920–33. <https://doi.org/10.1016/j.enpol.2013.07.089>.
- mailchimp. 2018. «Email Marketing Benchmarks». 1. März 2018. <https://mailchimp.com/resources/email-marketing-benchmarks/>.
- Martens, David, Foster Provost, Jessica Clark, und Enric Junqué de Fortuny. 2016. «Mining Massive Fine-Grained Behavior Data to Improve Predictive Analytics». *MIS Quarterly* 40 (4): 869–88.
- Morwitz, Vicki G., Joel H. Steckel, und Alok Gupta. 2007. «When Do Purchase Intentions Predict Sales?» *International Journal of Forecasting* 23 (3): 347–64. <https://doi.org/10.1016/j.ijforecast.2007.05.015>.
- Olson, David L., und Bongsug(Kevin) Chae. 2012. «Direct marketing decision support through predictive customer response modeling». *Decision Support Systems* 54 (1): 443–51. <https://doi.org/10.1016/j.dss.2012.06.005>.
- Prinzie, Anita, und Dirk Van den Poel. 2007. «Predicting home-appliance acquisition sequences: Markov/Markov for Discrimination and survival analysis for modeling sequential information in NPTB models». *Decision Support Systems* 44 (1): 28–45. <https://doi.org/10.1016/j.dss.2007.02.008>.
- Shrivastava, Utkarsh, und Wolfgang Jank. 2015. «A data driven framework for early prediction of customer response to promotions». In *AMCIS 2015 Proceedings*. Puerto Rico: AIS electronic library. <http://aisel.aisnet.org/amcis2015/BizAnalytics/GeneralPresentations/18>.
- Sodenkamp, Mariya, Konstantin Hopf, Ilya Kozlovskiy, und Thorsten Staake. 2016. «Smart-Meter-Datenanalyse für automatisierte Energieberatungen («Smart Grid Data Analytics»)». Final Report 291131. Bern, Switzerland: Bundesamt für Energie. <http://www.bfe.admin.ch/dokumentation/energieforschung/index.html?lang=de&publication=11372>.
- «Stromkennzeichnung». 2018. PRONOVO, VSE. <https://www.strom.ch/de/service/stromkennzeichnung>.
- Syam, Niladri, und Arun Sharma. 2018. «Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice». *Industrial Marketing Management* 69 (Februar): 135–46. <https://doi.org/10.1016/j.indmarman.2017.12.019>.
- Tabi, Andrea, Stefanie Lena Hille, und Rolf Wüstenhagen. 2014a. «What makes people seal the green power deal?—Customer segmentation based on choice experiment in Germany». *Ecological Economics* 107: 206–215.
- . 2014b. «What makes people seal the green power deal?—Customer segmentation based on choice experiment in Germany». *Ecological Economics* 107: 206–215.
- Tiefenbeck, Verena. 2014. «Behavioral Interventions to Reduce Residential Energy and Water Consumption». Doctoral Thesis, ETH Zurich. <https://doi.org/10.3929/ethz-a-010263933>.
- . 2017. «Bring behaviour into the digital transformation». *Nature Energy* 2 (6): 17085. <https://doi.org/10.1038/nenergy.2017.85>.
- Zeifman, Michael, und Kurt Roth. 2011. «Nonintrusive appliance load monitoring: Review and outlook». *IEEE Transactions on Consumer Electronics*, 76–84. <https://doi.org/10.1109/TCE.2011.5735484>.