



Leibniz Universität Hannover
Wirtschaftswissenschaftliche Fakultät
Institut für Wirtschaftsinformatik

Entwicklung eines Prognoseverfahrens für Energieverbräuche basierend auf Machine Learning

Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M. Sc.)“ im
Studiengang Wirtschaftsingenieur der Fakultät für Elektrotechnik und Informatik,
Fakultät für Maschinenbau und der Wirtschaftswissenschaftlichen Fakultät der
Leibniz Universität Hannover

vorgelegt von

Name: Bieber



Vorname: Clemens



Prüfer: Prof. Dr. M. H. Breitner

Ort, den: Hannover 19.07.2020

Inhalt

Abkürzungsverzeichnis	III
Formelverzeichnis.....	IV
Abbildungsverzeichnis	V
Tabellenverzeichnis	VII
1. Einleitung.....	1
2. Theoretische Grundlagen.....	4
2.1 Energieversorgungssystem in Deutschland.....	4
2.1.1 Netzaufbau / Infrastruktur / Stromanbieter.....	6
2.1.2 Ideales und reales Netz, Diskrepanz und resultierende Probleme in der Praxis	11
2.1.3 Strompreiszusammensetzung in Deutschland	13
2.2 Einblick in die Prognostizierung des Wetters	20
2.3 Machine Learning in Bezug auf Energieprognosemodelle	23
2.3.1 Übersicht bisheriger Erkenntnisse und Bezugnahme auf Machine Learning	24
2.3.2 Machine Learning Einordnung	26
2.3.3 Arten von Machine Learning Algorithmen.....	29
2.3.4 Regression.....	30
2.3.4.1 Regression Allgemein.....	30
2.3.4.2 Random Forest Regression.....	32
2.3.5 Artificial Neural Network.....	35
2.3.5.1 Architektur ANN.....	36
2.3.6.2 Lernverfahren	40
2.3.6.3 Backpropagation	42
2.5 Bewertungskriterien der Modelle	44
2.6 Rohdatenvorverarbeitung	46
2.6.1 Umgang mit Lücken im Datensatz	47
2.6.2 Transformation von Datenquellen	48
2.6.3 Outlier Detection	49
2.6.4 Auswahl der Eingabeparameter	51
2.7 Forschungsstand.....	52
3. Vorstellung der Methoden.....	55
3.1 Datenaufbereitung.....	55
3.2 Struktur Random Forrest.....	62
3.3 Struktur ANN.....	64
4. Analyse und Diskussion der Ergebnisse	66
5. Einordnung des Modells.....	77

5.1 Ausblick.....	77
5.2 Limitationen.....	78
6. Fazit.....	82
Anhang.....	85
Literaturverzeichnis.....	88
Ehrenwörtliche Erklärung.....	94

1. Einleitung

In den letzten Jahren ist ein drastischer Anstieg der Nachfrage an die weltweite Energieversorgung, bei gleichzeitiger Verknappung der fossilen Energieträger, zu verzeichnen. In Kombination mit der globalen Klimaerwärmung und den Folgen des weltweiten Bevölkerungswachstums ist davon auszugehen, dass dieser Trend sich fortsetzen wird. Gemessen an der Sommerdurchschnittstemperatur 2019 berichtete der Deutsche Wetterdienst (DWD) in einer Pressemitteilung, dass dies der zweitheißeste Sommer seit 1960 gewesen ist und dass grundsätzlich ein deutlicher Anstieg der Durchschnittstemperaturen zu erkennen ist (vgl. DWD 2019, S.2). Dieser Anstieg und die damit verbundenen Schwierigkeiten im Sommer, wie beispielsweise die Bereitstellung klimatisierter Büros für Mitarbeiter, betreffen so gut wie alle Menschen in Deutschland. Daher geht das Umweltbundesamt 2020 bereits davon aus, dass alleine im Bereich der Klimatisierung ein Anstieg der verursachten Co₂ Emissionen von 100% zu erwarten sei (vgl. Umweltbundesamt 2020, S.2). Vorangetrieben, durch diese Entwicklung drängen in den letzten Jahren immer mehr alternative Energieerzeugungskonzepte in den Energieerzeugungsmarkt, um der sauberen Energiebereitstellung zukünftig näher zu kommen. Zu diesen Konzepten zählen beispielsweise Photovoltaik- und Windkraftanlagen, oder im Bereich der Wärmeerzeugung solarthermische Anlagen. In der Zeit von 1990 bis 2019 konnten erneuerbare Energien ihren Anteil von ursprünglich 3,4% auf 42,1% an der Bereitstellung des benötigten Strombedarfs in Deutschland erhöhen (vgl. Umweltbundesamt 2019, S.2). Diese Entwicklung wird besonders in Deutschland mit dem Ziel, die gesteckten Klimaziele bis 2020 beziehungsweise 2030 zu erreichen, unter dem Deckmantel der Energiewende vorangetrieben (vgl. BMU 2020, S.1). Aufgrund der Tatsache, dass das Klimaziel bis 2020, welches bezogen auf das Jahr 1990 unter anderem eine Reduktion der Treibhausgasemissionen um 40% vorsah, wahrscheinlich verfehlt werden wird (vgl. IKEM 2018, S.3) und auch die Erreichung des Ziels bis 2030 auf der Kippe steht (vgl. Kersting 2020, S.1), bedarf es einem weiteren Ausbau erneuerbarer Energieerzeuger, um eine Einhaltung der Klimaziele zu gewährleisten.

2012 ergab eine Studie des Fraunhofer Instituts, dass Deutschland ausreichend Ressourcen für eine 100 prozentige Energieversorgung im Bereich Wärme und

Strom zur Verfügung stehen, um durch erneuerbare Energien den Bedarf zu decken (Fraunhofer Institut 2012, S.5). Erneuerbare Energien basieren in der Regel auf volatilen, nicht immer zugänglichen Rohstoffen, wie beispielweise Wind oder Sonne. Diese Verbundenheit mit nicht beeinflussbaren Faktoren erschwert die Einbindung dieser Energielieferanten in das Energieversorgungsnetz. Der Ursprung dafür liegt darin, dass der Bedarf und die Nachfrage der Energie, welche zur Verfügung steht zeitlich nicht immer übereinstimmen und somit häufig die Netzstabilität gefährdet ist. Aber nicht nur die Einbindung erneuerbarer Energien erschwert den weiteren Ausbau dieser. Auch abflachende Subventionen in diesem Bereich und die somit steigenden Kosten der Umstellung für den Endverbraucher hemmen den weiteren Ausbau (vgl. Witsch 2018, S.3).

An dieser Stelle soll die praktische Anwendung der theoretischen Fragestellung, welche im Rahmen dieser Arbeit bearbeitet wird ansetzen. Ziel dieser Arbeit ist es, ein Prognoseverfahren für Energieverbräuche, basierend auf Machine Learning zu konzeptionieren. Ein solches Verfahren könnte, nach ausreichender Prüfung, dazu dienen den Energiebedarf eines Hauses über den Tag bereits vorher zu kennen und somit die Netzbetreiber entsprechend darauf vorzubereiten. Sodass diese volatile Energieverbräuche besser einteilen können. Ein anderer Ansatzpunkt wird durch die Nutzung des Gebäudebetreibers selbst dargestellt, da dieser in der Lage sein könnte das Energiemanagement so anzupassen, dass bei der Benutzung von erneuerbaren Energien auf teure Speicher oder Netzzrückgriffe verzichtet werden kann. Der Fokus dieser Arbeit liegt zunächst in der Entwicklung des Prognoseverfahrens als solches. Daher sollen zwei Prognosemodelle entwickelt werden, welche unterschiedliche Machine Learning Ansätze verfolgen. Zum einen wird ein Random Forest Ansatz und zum anderen ein Artificial Neuronal Network Ansatz gewählt. Beide Prognosemodelle verwenden unter anderem Wetterdaten, um eine Prognose der Energieverbräuche zu erstellen. Anschließend wird im Analyseteil untersucht, welches Modell eine präzisere Vorhersage ausgibt und inwieweit die Forschungsfrage beantwortet werden konnte.

Um die Forschungsfrage zu bearbeiten, wird in Kapitel 2 zunächst auf die theoretischen Grundlagen eingegangen, die für die Auseinandersetzung mit der Thematik obligatorisch sind. Das Kapitel teilt sich in sieben Bereiche auf. Zunächst werden die Energieerzeugung, der Netzaufbau, die damit verbundenen

Problematiken und die Zusammensetzung des Strompreises in Deutschland thematisiert. Anschließend folgt eine Einordnung der Erstellung und der Zuverlässigkeit von Wetterprognosen. Folgend werden die theoretischen Grundsätze von Machine Learning, unterteilt in den Stand der Forschung und den theoretischen Aufbau der verwendeten Modelle erläutert. Weiterhin werden in Kapitel 2 die Bewertungskriterien, welche im Analysekapitel aufgegriffen werden, sowie das grundsätzliche Prinzip der Rohdatenverarbeitung im Bereich des Machine Learnings dargelegt. Nachdem Kapitel 2 grundlegend in die Thematik eingeführt hat, wird in Kapitel 3 auf die Umsetzung der theoretischen Modelle, sowie deren Anpassung an die zu bearbeitende Fragestellung eingegangen. Kapitel 4 folgt mit der Vorstellung der Ergebnisse der erläuterten Modelle und deren spezifischer Auswertung. Nach der Feststellung der Ergebnisse wird in Kapitel 5 knapp aufgegriffen, in welchen Bereichen die thematisierten Prognoseverfahren Anwendung finden können. Anschließend werden in diesem Kapitel Limitationen, welche sich im Rahmen der Arbeit, während der Durchführung oder der Ergebnisanalyse ergaben, thematisiert. Dieses Kapitel stellt den fließenden Übergang zum 6. und letzten Kapitel dar. Dieses ist das Fazit und der Ausblick, welches sich aus dem Verlauf der Arbeit, den Ergebnissen und den Limitationen zusammensetzt, die für weiterführende Arbeiten in diesem Bereich wegweisend sein können.

6. Fazit

Im Rahmen dieser Arbeit lag der Fokus auf der Entwicklung eines Algorithmus, welcher zukünftige Energieverbrauchswerte vorhersagen soll.

Zunächst wurde ein theoretisches Verständnis für die Energieerzeugung, den Energiemarkt und die Strompreisentwicklung in Deutschland geschaffen, um die Notwendigkeit eines Algorithmus zur Vorhersage von Energieverbräuchen zu verdeutlichen. Anschließend wurde sich der Erstellung und Genauigkeit der Wetterprognosen im Allgemeinen gewidmet, da diese einen obligatorischen Teil der Informationen für die späteren Algorithmen liefern und unentbehrlich für die praxisnahe Anwendung sind. Der Theorieteil schließt mit dem Kapitel, welches die Grundlagen des Maschinellen Lernens umfasst und besonders die in dieser Arbeit fokussierten Algorithmen Random Forest und Artificial Neuronal Network dabei herausarbeitete.

Im nächsten Schritt wurden beide Algorithmen in Python implementiert. Im Laufe mehrerer Durchläufe gelang es, optimale Parameter für den jeweiligen Algorithmus zu definieren. Diese wurden letztlich verwendet, um zu den vorgestellten Ergebnissen zu gelangen. So war es möglich im besten Fall der untersuchten Datensätze mit dem ANN-Algorithmus eine Genauigkeit von 98,9% zu erlangen. Bis auf eine Ausnahme lagen alle weiteren Ergebnisse in einem ähnlichen Spektrum. Die einzige Ausnahme stellte der Gebäudetyp des profil16 dar. Die wesentlich ungenaueren Ergebnisse konnten in der Analyse jedoch durch den stark abweichenden Verlauf in Bezug auf die anderen Gebäude begründet werden.

Herausgestochen in der Ergebnisanalyse ist, dass beide Algorithmen nicht die Daten bezogen auf ihre zeitliche Komponente auswendig lernen, sondern, dass Features, wie beispielweise Temperatur, Niederschlag, Sonnenstunden oder Watt_t-1h, einen höheren Beitrag bezüglich der Prognose innehaben. In diesem Zuge ist ebenfalls aufgefallen, dass die zusätzliche Information, welche die Energieverbrauchswerte beinhaltet, und dabei insbesondere diese mit den Werten zu der vorherigen Stunde eine enorme Verbesserung der Ergebnisse zur Folge hatten. Dies spiegelte sich ebenfalls in der benötigten Rechenzeit des ANN-Algorithmus wider. Zusammenfassend kann nach der Ergebnisanalyse festgehalten

werden, dass es möglich ist, mittels Random Forest oder Artificial Neuronal Network einen Vorhersagealgorithmus aufzubauen, welcher unter den richtigen Voraussetzungen sehr exakte Ergebnisse erzielen kann.

Abschließend wurden Limitationen des Modells und mögliche Verbesserungen thematisiert. Dabei stellten sich zwei mögliche Schwachstellen für die praktische Anwendung heraus. Zum einen können beide Modelle erst genutzt werden, nachdem das jeweilige Gebäude einige Jahre in Betrieb genommen wurde und ausreichend Daten zur Verfügung stehen und zum anderen setzen die Algorithmen eine konstante Nutzung der Gebäude voraus.

Um diese Probleme zu umgehen, könnte es in zukünftigen Erweiterungen sinnvoll sein, die Anzahl der im Gebäude befindlichen Personen ebenfalls mit in das Modell einfließen zu lassen. Dies kann zur Folge haben, dass die zu erwartende Korrelation zwischen der Anzahl der Personen und dem Energieverbrauch abgedeckt wird. Eine weitere Möglichkeit stellt der Einsatz eines LSTM-Algorithmus dar, welcher ausschließlich die Daten der vorherigen Stunden benötigt. Diesen könnte man mit einem der hier beschriebenen Algorithmen kombinieren, um eine optimale Vorhersagegenauigkeit zu erlangen.

Um das zuerst beschriebene Problem zu umgehen, dass das Gebäude einige Jahre im Betrieb genommen sein muss, damit ein Algorithmus lernen und für Energieverbrauchsprognosen verwendet werden kann, könnten Parameter wie Raumgrößen, Wärmeisolierung und exakte geografische Daten mit in das Modell aufgenommen werden. Diese könnten eine Übertragung von Algorithmen ähnlicher Gebäude aufeinander mit der Einstellung der entsprechenden Parameter beinhalten.

Abschließen ist zu sagen, dass das Ziel der Algorithmusentwicklung mit der Absicht der Voraussage von Energieverbrauchswerten beziehungsweise der Entwicklung eines Prognosemodells auf Basis von Machine Learning erreicht wurde und dass beide Algorithmen präzise Ergebnisse mit bis zu 98 % liefern. Dabei zeigen die Ergebnisse, dass die Wahl des Algorithmus selbst weniger relevant ist, als eine entsprechend vollständige Datenlage und die korrekte Aufbereitung dieser für den jeweiligen Algorithmus. Durch die gute Eignung zur Vorhersage der Energieverbrauchswerte kann in weiteren Arbeiten an einer Integrationen in

verschiede Lastmanagementsysteme gearbeitet werden, sodass die Probleme der Energiewende minimiert werden könnten.