

# Menschliches Wissen und Künstliche Intelligenz in einem Prognosemodell für Erneuerbare Energien

**Alexander Wallis, Alfons Haber, Sascha Hauke**

Hochschule Landshut, Am Lurzenhof 1, D-84028 Landshut, +49 (871) 506 690,  
alexander.wallis@haw-landshut.de, alfons.haber@haw-landshut.de, sascha.hauke@haw-  
landshut.de, www.haw-landshut.de

**Kurzfassung:** Mit der zunehmenden Integration von volatilen, erneuerbaren Energiequellen (EEQ) in das bestehende Stromnetz, geht auch eine damit ansteigende Komplexität in der Bereitstellung elektrischen Stroms innerhalb eines Smart Grids (SG) einher. Neben den Laständerungen haben Leistungsänderungen der Erzeugungsanlagen der einzelnen EEQ sowie das Lade- und Entladeverhalten von Energiespeichern einen direkten Einfluss auf die Lastflüsse und die Spannungsqualität des Netzes. Daraus resultierend kann es im schlimmsten Fall zum Stromausfall kommen. Um sich diesem Problem vertieft zu widmen, bedarf es einer strengen Koordination und Überwachung der ans Netz angeschlossenen Anlagen, damit leistungstechnischer Austausch zwischen den einzelnen Akteuren in einem SG gesichert ist. In dieser Arbeit wird ein datengetriebener Ansatz mit Hilfe von Methoden des maschinellen Lernens zur Prognose des Stromverbrauchs für den nächsten Tag sowie der Stromerzeugung von EEQ, unter Berücksichtigung von menschlichen Wissens und Künstlicher Intelligenz, vorgestellt. Diese Methoden bilden die wesentliche Grundlage für die weitere strategische Koordinierung innerhalb eines SG und die Stromversorgung der Zukunft.

**Keywords:** Erneuerbare Energien, Künstliche Intelligenz, Smart Grid, Verbrauchsprognose, Erzeugungsprognose, maschinelles Lernen

## 1 Einleitung

Ein aktives Mitwirken an den Klimazielen ist ein wichtiger Beitrag, es geht darum, hier mit dem Tun die erforderlichen Schritte zu setzen. So baut das „*Clean Energy for all Europeans Package*“ der Europäischen Kommission auf einen unverzüglichen Ausbau von erneuerbaren Energiequellen [1]. Die Rahmenbedingungen in Deutschland und auch in Europa führen zu einem Paradigmenwechsel in der Integration von Erzeugungs- und Speicheranlagen in das Stromnetz. Die Aggregation und Steuerung der jeweiligen Teilnehmer führt zu einem Komplexitätsanstieg innerhalb des Netzes und des Systems, den es über Expertenwissen, sprich menschlicher Intelligenz zu lösen gilt. Verschiedene Ansätze befassen sich mit strenger Überwachung und schneller Steuerung, um einen stabilen Betrieb zu gewährleisten [2]–[4]. Insbesondere die Integration von unterschiedlichen erneuerbaren, volatilen Energiequellen baut auf eine solide Produktionsvorhersage, um die einhergehenden Fluktuationen zu bewältigen. Auf Basis einer soliden Vorhersage können im Folgenden z.B. Betriebsstrategien für ein Energiespeichersystem anhand dynamischer Randbedingungen optimiert werden [5].

Im Rahmen dieser Arbeit wird ein datengetriebener Ansatz mit Hilfe von Methoden des *maschinellen Lernens* (ML), einem Teilgebiet der *Künstlichen Intelligenz* (KI), und unter

Berücksichtigung von menschlichen Wissen für die Vorhersage des Stromverbrauchs sowie der Stromerzeugung einer Forschungseinrichtung vorgestellt und mit state-of-the-art Methoden verglichen. Verbrauchsvorhersagen spielen in der Forschung aktuell eine große Rolle und werden anhand ihres Vorhersagehorizonts in die folgenden Kategorien untergliedert: *Long-Term Load Forecast* (ein bis zehn Jahre), *Medium-Term Load Forecast* (ein Monat bis ein Jahr) und *Short-Term Load Forecast* (eine Stunde bis ein Tag) [6]. Aufgrund der Vorhersage für den nächsten Tag steht letzter Kategorie im Fokus dieser Arbeit.

Die restliche Arbeit gliedert sich wie folgt: nachdem in Kapitel 2 das methodische Vorgehen für die Entwicklung des Prognosemodells mit den dazugehörigen Unterschritten beschrieben wird, werden anschließend Vergleichsmodelle definiert, die Ergebnisse präsentiert und evaluiert (Kapitel 3). Abschließend wird die Arbeit in Kapitel 4 zusammengefasst und Ansätze für sowohl mögliche Verbesserungen als auch weiterführende Forschungen aufgezeigt.

## 2 Methodik

Das zunehmende Interesse an Künstlicher Intelligenz wird vorrangig von zwei Faktoren bestimmt: Anzahl der vorhandenen Daten und sowohl steigende als auch gleichzeitig kostengünstigere Rechenkapazitäten. Für die Entwicklung eines Vorhersagemodells ist eine solide Datengrundlage essentiell. Dabei wird die Prognosegüte hauptsächlich – frei nach dem Motto: mehr Daten übertreffen clevere Algorithmen, aber bessere Daten übertreffen mehr Daten – durch die Qualität der zugrundeliegenden Daten beeinflusst [7]. Allerdings liefern gute Daten ohne zusätzliche Analyse in den meisten Fällen keine besseren Ergebnisse. Erst durch das menschliche, domänenspezifische Expertenwissen können Methoden des maschinellen Lernens ihr volles Potential zeigen [8]. Dieses Wissen über die Zusammenhänge innerhalb der Daten – welches extern hinzugefügt wird – ist umso wichtiger für Ansätze, bei denen nicht ausreichend große Datenmengen für einen Deep Learning-Ansatz vorhanden sind [9]. Durch eine detaillierte Betrachtung der rohen Messdaten werden Informationen gewonnen, die wiederum nach einer anschließenden Analyse zu neuen Erkenntnissen führen können (vgl. Abbildung 1).

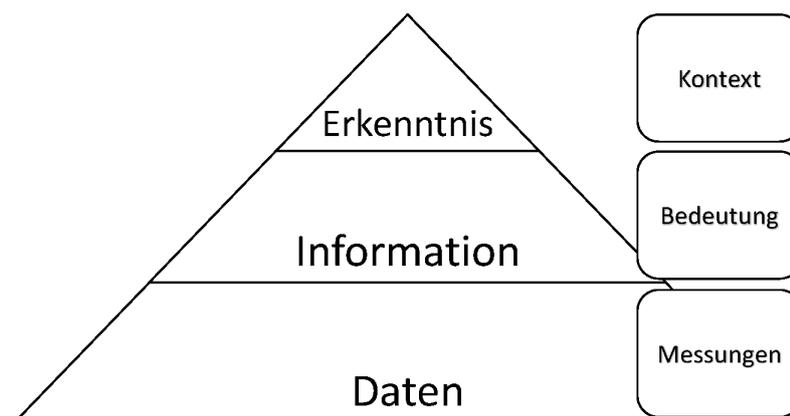


Abbildung 1: Aus Daten werden Informationen gewonnen, die wiederum in einem bestimmten Kontext zu Erkenntnissen führen. Quelle: eigene Darstellung angelehnt an [7].

Dieser Schritt von den Daten hin zu neuen Erkenntnissen wird in diesem Kapitel näher beschrieben [10]. Dabei werden zuerst die zugrunde liegenden Daten genauer untersucht und

mögliche Einflussfaktoren analysiert. Im Anschluss wird der Entwicklungsprozess für das Vorhersagemodell erläutert.

## 2.1 Datenbeschreibung

Im Rahmen dieser Forschungsarbeit diente das *Technologiezentrum Energie* (TZE) in Ruhstorf a. d. Rott, eine Forschungseinrichtung der *Hochschule Landshut*, als Datenquelle für Verbrauchs-, Erzeugungs- und Speicherdaten. Am TZE werden sekundengenaue Leistungswerte aufgezeichnet, in einer Datenbank abgespeichert und für eine anschließende Analyse bereitgestellt. Bei diesen Leistungswerten über die Zeit handelt es sich insbesondere sowohl um Verbrauchsdaten der einzelnen Büros und Labore als auch Erzeugungsdaten der installierten PV-Anlagen. In einem ersten Schritt wurde in Bezug auf die Leistungsänderungen ein möglichst günstiges Zeitintervall gesucht. Die folgende Analyse und Modellentwicklung basiert auf Daten zwischen dem 1. Januar 2017 und dem 31. Dezember 2017. Das Ziel des Modells ist es, den Verbrauch und die Erzeugung des nächsten Tages als stündlich diskrete Werte zu prognostizieren. Daher werden die sekundlichen Daten im nächsten Schritt zu Stundenwerten zusammengefasst, was 8.760 Datenpunkte liefert. Falls fehlenden Messwerte vorhanden sind, werden diese interpoliert. Mit einer vollständigen plausibilitätsgeprüften Datengrundlage, definiert nach dem Verbraucherzählpeilsystem, geht es im nächsten Schritt darum, durch gezielte Analyse Informationen über das zugrundeliegende Verhalten zu erlangen [11].

## 2.2 Analyse und Feature Engineering

Nachdem im vorherigen Schritt eine valide Datengrundlage erzeugt wurde, geht es nun an die Analyse der Verbrauchsdaten und an das Identifizieren von möglichen Indikatoren (*engl.* feature engineering), die das Nutzerverhalten beschreiben und somit als Eingabeparameter (*engl.* features) für das Vorhersagemodell verwendet werden können.

Wenn man – ausgehend von einem Persistenzmodell „gestern wie heute“ – überlegt, welcher Messpunkt in der Vergangenheit am besten den aktuellen Verbrauch zu einer bestimmten Stunde  $h$  beschreibt, gelangt man über eine erste Annäherung an den gestrigen Wert  $h - 24$ . Diese Vermutung belegt Abbildung 2. Die Graphik zeigt den *Pearson Korrelationskoeffizienten* eines Messwerts mit anderen Messwerten innerhalb des angegebenen Zeitfensters.

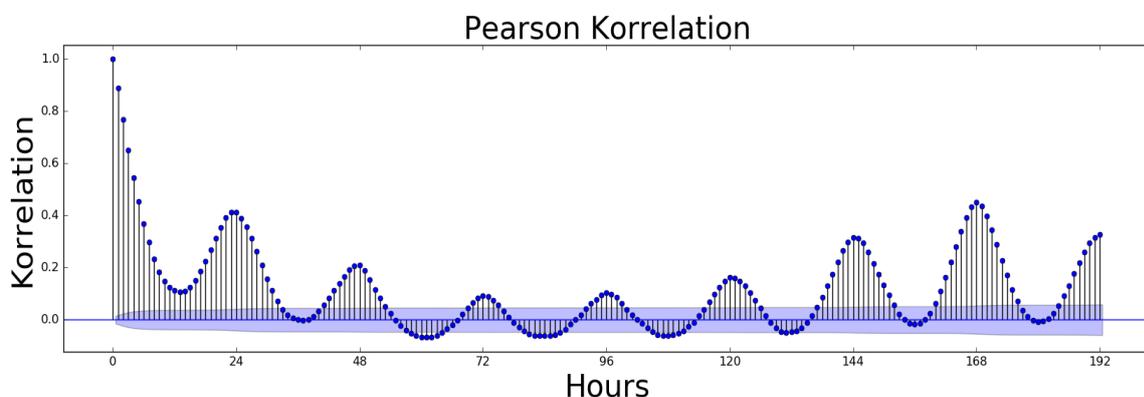


Abbildung 2: Korrelation zwischen einem Messwert zu einem bestimmten Zeitpunkt und einem zurückliegenden Datenpunkt. Quelle: eigene Darstellung.

Die Berechnung von diesem Koeffizienten wird durch Gleichung (1) beschrieben.

$$\rho_{x_t, x_{t-i}} = \frac{\text{cov}(x_t, x_{t-i})}{\sigma_{x_t} \sigma_{x_{t-i}}}, i \in \{0, \dots, 192\} \quad (1)$$

Dabei ist  $\text{cov}$  die Kovarianz zwischen zwei betrachteten Messpunkten und  $\sigma_{x_t}, \sigma_{x_{t-i}}$  beschreibt die Standardabweichungen eines beliebigen Messwerts  $x_t$  und einem davon ausgehend zurückliegenden Wertes  $x_{t-i}$ . Wie erwartet, gibt es eine höhere Korrelation (0,41) mit dem gestrigen Verbrauchswert, allerdings kann der aktuelle Messwert noch genauer mit einem Korrelationsfaktor von 0,5 durch den 168 Stunden (genau eine Woche) zurückliegenden Wert beschrieben werden. Dieses Verhalten ist durch den signifikanten Verbrauchsunterschied zwischen Werktagen und Wochenenden zu erklären, falls der betrachtete Wert ein Montag und ein Werktag ist.

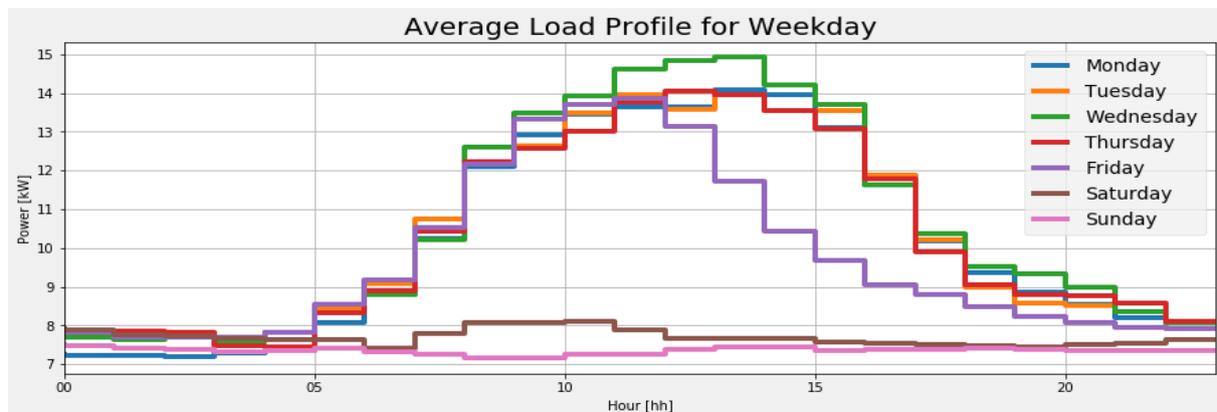


Abbildung 3: Standard Lastprofil für jeden Wochentag auf Basis der aufgezeichneten Daten am TZE. Quelle: eigene Darstellung.

Dieser Unterschied wird in Abbildung 3 dargestellt. Über die gesamte Zeitdauer der aufgezeichneten Daten werden die durchschnittlichen Verbrauchsprofile des TZE für jeden einzelnen Wochentag gezeigt. Hier ist die Differenzierung zwischen Werktagen und Wochenenden, an denen nur die Grundlast verbraucht wird, klar zu erkennen. Des Weiteren können bereits erste grobe Annahmen über das Verbrauchsverhalten, hier am Beispiel TZE getroffen, werden. Der Mittwoch ist in der Regel der Tag mit den durchschnittlich maximalen Verbräuchen und am Freitag flachen die Werte ab Mittag langsam ab. Daraus folgt, dass auf jeden Fall eine Kategorisierung in Werktag und Wochenende sinnvoll ist. Darüber hinaus verhalten sich Feiertage, die auf einen Werktag fallen, in der Regel wie ein durchschnittlicher Wochenendtag.

Nachdem die relevanten Eingabeparameter identifiziert wurden, müssen diese noch kodiert werden, damit das Prognosemodell diese richtig interpretieren kann. Dabei reicht es nicht die Wochentage von 1 bis 7 durchnummerieren, da diese dadurch eine numerische Ordnung bekommen. Eine gängige Methode, um dieses Problem zu umgehen, ist das *one hot encoding*, dabei werden kategorische Variablen als Binärvektoren dargestellt [12]. In diesem Fall wird die kategorische Variable *Wochentag* mit den Werten  $\{\text{Montag}, \dots, \text{Sonntag}\}$  als ein Binärvektor der Länge 7 bestehend aus 0 oder 1 dargestellt, wobei immer nur die Stelle für den jeweiligen Tag mit 1 gesetzt ist z.B.  $\text{Montag} = [1,0,0,0,0,0,0]$  oder  $\text{Mittwoch} = [0,0,1,0,0,0,0]$ .

Zudem werden auch die jeweiligen durchschnittlichen Temperaturen an den beiden Tagen  $d - 1$  und  $d - 7$  berücksichtigt. Insgesamt ergibt das 60 Features. Tabelle 1 liefert eine Übersicht über alle verwendeten Eingabeparameter mit ihren Wertebereichen und Einheiten/Typen.

Tabelle 1: Identifizierte Features für das Prognosemodell. Quelle: eigene Darstellung.

Feature Nummer	Name	Wertebereich	Einheit/Typ
1–24	Leistungswerte für $d - 1$	[0,15; 34,1]	[kW]
25–49	Leistungswerte für $d - 7$	[0,15; 34,1]	[kW]
50 und 51	Temperatur für $d - 1$ und $d - 7$	[-17,9; 32,6]	[°C]
52	Feiertag	{0,1}	[Boolean]
53–60	Wochentag	{0,1}	[Boolean]

### 2.3 Entwicklung des Prognosemodells

Auf den Vorarbeiten der vorherigen Schritte aufbauend, wird in diesem Abschnitt die Entwicklung des Prognosemodells beschrieben. Dafür muss zunächst das zugrundeliegende Problem beschrieben werden, um davon ausgehend das Modell zu trainieren.

**Problemdefinition:** ML-Methoden können als *überwachtes* (engl. supervised), *unüberwachtes* (engl. unsupervised) und *bestärkendes* (engl. reinforcement) Lernen kategorisiert werden [13]. In dieser Arbeit liegt ein Problem des überwachten Lernens vor, da ein Zusammenhang zwischen Eingabe- und Ausgabewerten auf der Basis von Trainingsdaten, die aus einer Menge von Beispielen der Form *Eingabe-Ausgabe* bestehen, vorliegen. Des Weiteren wird zwischen einer *Regression* und einer *Klassifikation* unterschieden, je nachdem, ob es sich bei den Ausgabewerten um kontinuierliche oder diskrete Werte handelt. Hier handelt es sich um ein Regressionsproblem, da ausgehend von einer historischen Datenmenge  $\mathcal{D}$  von  $N$  Messwerten der Form  $\{(x_1, y_1), \dots, (x_N, y_N)\}$ , wobei  $x_i$  der Eingabevektor des  $i$ -ten Beispiels ist und  $y_i$  der dazugehörige Ausgabevektor ist, eine Funktion  $g: X \rightarrow Y$  gesucht wird, wobei  $X$  den Eingaberaum und  $Y$  den Ausgaberaum beschreibt [14].

**Modellparameter:** Nachdem das Problem definiert wurde, geht es an die Entwicklung des Prognosemodells. Dafür wird die komplette Datenmenge  $\mathcal{D}$  in zwei disjunkte Mengen  $\mathcal{D}_T$  und  $\mathcal{D}_V$  für Training respektive Validierung unterteilt. Diese Aufteilung ist notwendig, um eine möglichst optimale Generalisierbarkeit des Modells zu gewährleisten.

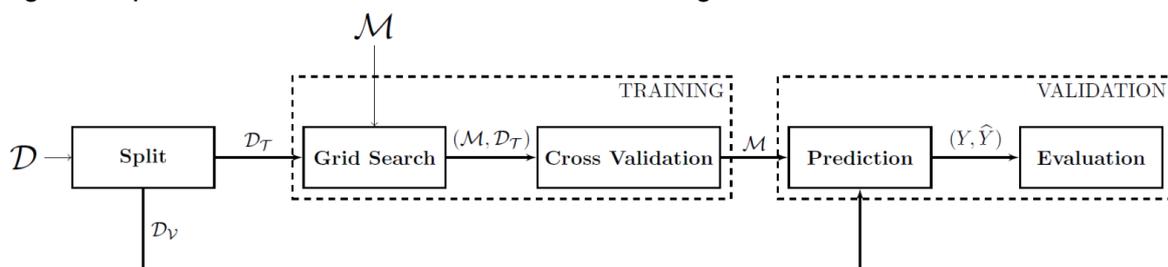


Abbildung 4: Ablaufdiagramm für den Entwicklungsprozess eines Prognosemodells. Quelle: eigene Darstellung

Dabei steht  $\mathcal{M}$  für die Menge der verwendeten Modelle mit ihren jeweiligen Parametern. In dieser Arbeit werden folgende Lernmodelle verwendet: *Random Forest* (RF), *Support Vector Machine* (SVM) und ein *Multi-Layer-Perceptron* (MLP). Alle drei können sowohl für Klassifikationen als auch Regressionen verwendet werden und haben ihre Validität in den unterschiedlichsten Bereichen bewiesen [15]–[20]. Im ersten Schritt werden die Hyperparameter, welche die Architektur des Modells beschreiben, mit Hilfe einer *Grid Search* ermittelt [21]. Die besten Resultate liefern dabei ein RF bestehend aus 1.000 Einzelbäumen und einer maximalen Tiefe von 15, ein NN mit einer verdeckten Schicht und 100 Neuronen sowie eine SVM mit einer *Radial Basis Function* als Kernel und einer C-Rate von 10. Danach werden die variablen Modellparameter (Gewichte) dynamisch in einem Trainingsprozess, der aus einer *10-Fold Cross Validation* besteht, iterativ angepasst [22]. Ist dieser Trainingsprozess abgeschlossen, werden Vorhersagen auf den Validierungsdaten  $\mathcal{D}_v$  erstellt.

### 3 Ergebnisse

Um die Prognosen der jeweiligen ML-Methoden in Relation setzen zu können, werden Vergleichsmethoden definiert. Sei  $L(d, h)$  die bezogene Leistung für ein Stundenintervall  $h$  an einem gegebenen Tag  $d$ , dann beschreibt ein Persistenzmodell die bezogene Leistung  $L(d_t, h)$  anhand desselben Stundenintervalls eines vorherigen Tages  $L(d_{t-i}, h)$ . Ausgehend davon, werden die folgenden Modelle zum Vergleich definiert: (1) „*Heute wie Gestern*“ (HWG)  $L(d_t, h) = L(d_{t-1}, h)$ ; (2) „*Heute wie vor einer Woche*“ (HWVEW)  $L(d_t, h) = L(d_{t-7}, h)$ ; (3) „*Heute wie der heutige Durchschnitt*“ (HWDHD)  $L(d_t, h) = \overline{L(d_t, h)}$ , wobei  $\bar{L}$  den Durchschnitt aller des heutigen Wochentages beschreibt.

Damit die Prognosewerte der jeweiligen Modelle miteinander verglichen werden können, werden gängige statistische Metriken herangezogen. Dabei gilt es, die richtigen Metriken für das vorliegende Problem auszuwählen (siehe *Problemdefinition*). In dieser Arbeit werden die folgenden vier, allgemein angewandten Metriken verwendet: *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE) und *Mean Percentage Error* (MPE). Diese sind in den Gleichungen (2)–(5) beschrieben:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^N (Y_n - \widehat{Y}_n)^2 \quad (2)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (3)$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |Y_n - \widehat{Y}_n| \quad (4)$$

$$\text{MPE} = \frac{100\%}{N} \sum_{n=1}^N \left( \frac{Y_n - \widehat{Y}_n}{Y_n} \right) \quad (5)$$

Dabei beschreibt  $Y_n$  den gemessenen und  $\widehat{Y}_n$  den vorhergesagten Wert.  $N$  ist die Anzahl der zu testenden Daten. Die vorhergesagten Ergebnisse — durch die Testdaten erstellt — sind in Tabelle 2 aufgelistet. Da  $D$  zufällig in  $D$  und  $D$  aufgeteilt wird, besteht die Möglichkeit, dass

eine „ungünstige“ Partitionierung gewählt wird, was die Ergebnisse verfälschen. Um dem entgegen zu wirken, wird der Ablauf aus Abbildung 4 insgesamt zehn Mal wiederholt und der Durchschnitt über die jeweiligen Fehlermaße gebildet. Der Wert in der Klammer steht für die Standardabweichung. Maximale und minimale Werte sind kursiv respektive fett gedruckt.

Tabelle 2: Ergebnisse für die unterschiedlichen Vorhersagemethoden.

		Metriken			
		MSE	RMSE	MAE	MPE
Modelle	HWG	10,98 (2,33)	3,29 (0,34)	1,98 (0,19)	<b>-7,84</b> (8,17)
	HWVEW	12,54 (3,18)	3,51 (0,45)	2,22 (0,25)	-9,14 (8,05)
	HWDHD	6,95 (1,83)	2,61 (0,34)	1,88 (0,23)	-10,85 (6,86)
	RF	<b>5,60</b> (1,49)	<b>2,24</b> (0,24)	<b>1,42</b> (0,11)	-11,12 (10,46)
	SVM	6,14 (1,62)	2,46 (0,32)	1,79 (0,26)	-9,03 (13,42)
	MLP	6,08 (1,05)	2,46 (0,21)	1,76 (0,22)	-8,55 (13,05)

Bei den zuvor beschriebenen Metriken ist es das Ziel, den Wert möglichst gering zu halten und damit den Fehler zu minimieren. Aus diesen Werten in der Tabelle 2 ist zu entnehmen, dass der Random Forest in den meisten Metriken (bis auf den MPE) die besten Resultate liefert. Die anderen ML-Methoden schneiden ähnlich ab, mit einem leichten Vorteil für das MLP. Die erstellten Prognosen werden in Abbildung 5 nochmals dargestellt.

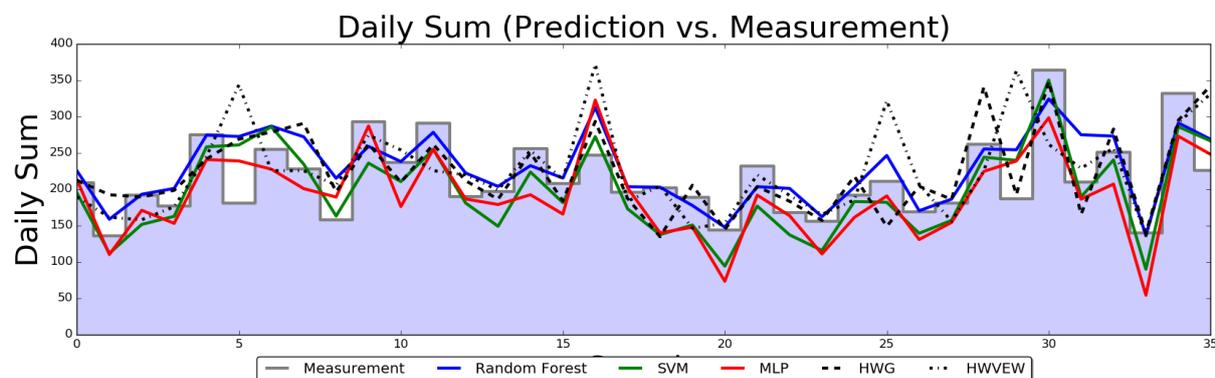


Abbildung 5: Vergleich der Verbrauchsprognosen aufsummiert für die einzelnen Tage zur besseren Betrachtung. Quelle: eigene Darstellung.

Für eine bessere Darstellung der täglichen Verbrauchsprognosen werden die stündlichen Leistungsintervalle über den Tag aufsummiert. Jeder Index auf der x-Achse steht für einen zufällig ausgewählten Tag aus dem Validierungsdaten, wodurch diese Darstellung nicht als zeitlicher Verlauf, sondern als jeweils alleinstehender Datenpunkt bewertet werden muss. Die beiden unterteilten Linien stellen die Vergleichsmodelle (HWG und HWVEW) dar, während die übrigen für die ML-Modelle stehen. Zu erkennen ist, dass sowohl das NN als auch die SVM

eher dazu tendieren den tatsächlichen Verbrauch etwas zu unterprognostizieren, während dem gegenüber der RF und die Vergleichsmodelle teilweise auch mal deutlich überschätzen. Diese gesammelten Erkenntnisse können dazu verwendet werden, zusammen mit weiteren Verbesserungen an den jeweiligen Modellen, die Prognosegüte zu erhöhen.

Aus Platzgründen wurde in dieser Arbeit der Entwicklungsprozess des Prognosemodells nur für die Verbrauchsvorhersage aufgezeigt. Allerdings funktioniert dieses Vorgehen ebenfalls für die Prognose von volatilen Primärenergiequellen (z.B. Solar, Wind). Dazu zeigt Abbildung 6 die beiden Prognosen für die Erzeugung durch eine PV-Anlage und den Verbrauch für den nächsten Tag.

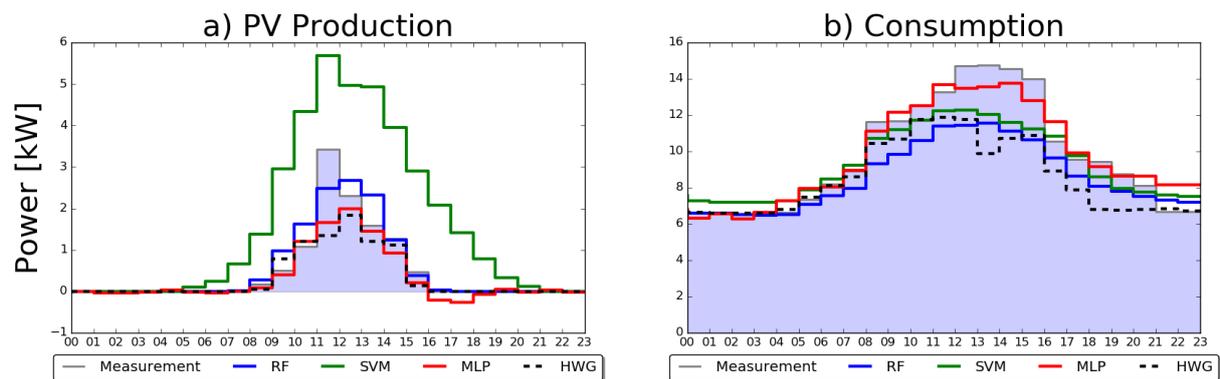


Abbildung 6: Vorhersage für a) PV-Erzeugung und b) Verbrauch anhand der zuvor entwickelten Prognosemodelle beispielhaft für einen zufällig ausgewählten Tag. Quelle: eigene Darstellung.

Aus Gründen der Übersicht wurden für diese Darstellung allerdings nur das Vergleichsmodell HWG herangezogen. Anhand dieser beiden Prognosen für die Erzeugungsmenge und den Verbrauch für den nächsten Tag, können nun die einzelnen Akteure in einem SG oder auch nur ein einzelner Haushalt mit einem Energiespeicher, ihre Strategie für den nächsten Tag anhand vorgegebenen Bedingungen anpassen und im Rahmen einer gewissen Fehlertoleranz optimieren.

## 4 Fazit und Ausblick

Ausgehend von der Tatsache, dass EEQ eine immer wichtigere Rolle in der Stromversorgung spielen werden, um die eingangs erwähnten Klimaziele der EU zu erreichen, ist eine strenge Überwachung und Steuerung mittels Informations- und Kommunikationstechnologien unabdingbar. Um die damit verbundene Komplexität zur stetigen Überwachung und Messung in den Griff zu bekommen, ist eine Möglichkeit auf Basis von historischen Daten und mittels Methoden des maschinellen Lernens Prognosemodelle für die jeweiligen Erzeuger und Verbraucher innerhalb eines SG zu erstellen. Dazu wurden in dieser Arbeit vorhandene Potentiale für die weitere Entwicklung und Forschung an diesen Methoden aufgezeigt. Ausgehend von historischen Verbrauchs-, Erzeugungs- und Speicherdaten über einen längeren Zeitraum, z.B. ein Jahr, wurde in dieser Arbeit ein Modell zur Prognose von Lasten- und Erzeugungsanlagen einer kleinen und bis mittelgroßen Forschungseinrichtung vorgestellt

und an Daten der realen Welt auf Validität geprüft. Die auf einheitlichen Metriken in Tabelle 2 aufgelisteten Werte zeigen, dass die ML-Modelle gegenüber den gängigen Persistenzmodellen (z.B. „heute wie gestern“) eine Minimierung in der Prognoseabweichung aufweisen. Innerhalb den drei betrachteten ML-Methoden ist der Random Forest sowohl in den durchschnittlichen Abweichungen als auch deren jeweiligen Standardabweichungen bis auf eine Ausnahme das akkurateste Modell. Auf Basis dieses Entwicklungsprozesses (vgl. Abbildung 4) wurde zudem, ohne große Abänderungen, ein Prognosemodell für die Stromerzeugung mittels einer PV-Anlage entwickelt, das im Rahmen einer gewissen Fehlertoleranz ebenfalls brauchbare Prognosen liefert. Des Weiteren können diese Modelle auch andere volatile primäre Energiequellen abbilden.

Allerdings sollte diese Arbeit auch den Einfluss von menschlichen Wissens für die Entwicklung eines Prognosemodells aufzeigen. Liegen nicht genügend Daten (Big Data) für die Verwendung von Deep-Learning-Anwendungen vor, müssen die verfügbaren Daten durch Expertenwissen nachträglich angereichert werden, um mit Hilfe von ML-Methoden brauchbare Prognosen zu erlangen.

In weiterführenden Arbeiten wird die Modellqualität, die vorrangig durch die Prognosegüte bestimmt wird, weiter verbessert. Dabei können auch explizit Methoden zur Zeitreihenanalyse und –vorhersage verwendet und mit den Ergebnissen aus dieser Arbeit verglichen werden. Neben der stetigen Verbesserung der Modellgenauigkeit, ist ein möglicher Ansatzpunkt die Verwendung von Energiespeicher, um erwartete Prognosefehler ausgleichen zu können und trotzdem insgesamt zu einer Entlastung des Stromnetzes zu führen. Dieses enge Ineinandergreifen von Vorhersage und Optimierung ist ein sehr interessantes Forschungsfeld mit reichlich Potential für neue Ideen.

## 5 Referenzen

- [1] European Commission, “Clean energy for all Europeans,” 2019.
- [2] S. Kakran and S. Chanana, “Smart operations of smart grids integrated with distributed generation: A review,” *Renew. Sustain. Energy Rev.*, vol. 81, pp. 524–535, Jan. 2018.
- [3] E. Mengelkamp, J. Gärttner, K. Rock, S. Kessler, L. Orsini, and C. Weinhardt, “Designing microgrid energy markets,” *Appl. Energy*, vol. 210, pp. 870–880, Jan. 2018.
- [4] P. D. Curtis and N. Mehravari, “Evaluating and improving cybersecurity capabilities of the energy critical infrastructure,” in *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*, 2015, pp. 1–6.
- [5] O. Greil, J. Kainz, M. Kain, and A. Haber, “Optimierungsmodell für unterschiedliche Lade- und Entladestrategien von PV-Speichersystemen,” 2019, pp. 1–8.
- [6] S. Rahman and O. Hazim, “A generalized knowledge-based short-term load-forecasting technique,” *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 508–514, May 1993.
- [7] A. Halevy, P. Norvig, and F. Pereira, “The Unreasonable Effectiveness of Data,” *IEEE Intell. Syst.*, vol. 24, no. 2, pp. 8–12, Mar. 2009.
- [8] C. Baumgartner, C. Böhm, and D. Baumgartner, “Modelling of classification rules on metabolic patterns including machine learning and expert knowledge,” *J. Biomed. Inform.*, vol. 38, no. 2, pp. 89–98, Apr. 2005.

- [9] G. Piatetski and W. Frawley, *Knowledge Discovery in Databases*. MIT Press, 1991.
- [10] R. L. Ackoff, "From Data to Wisdom," *J. Appl. Syst. Anal.*, vol. 16, no. 1, pp. 3–9, 1989.
- [11] A. Haber, "Batteriespeicher," Hrsg. J. Böttcher and P. Nagl, De Gruyter Oldenbourg, 2018, pp. 233-254.
- [12] K. J. Berry, P. W. Mielke, and H. K. Iyer, "Factorial Designs and Dummy Coding," *Percept. Mot. Skills*, vol. 87, no. 3, pp. 919–927, Dec. 1998.
- [13] O. Simeone, "A Very Brief Introduction to Machine Learning with Applications to Communication Systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, 2018.
- [14] S. J. Russell, P. Norvig, and E. Davis, *Artificial intelligence: A modern approach*. Upper Saddle River, New Jersey, 2010.
- [15] A. Lahouar and J. Ben Hadj Slama, "Day-ahead load forecast using random forest and expert input selection," *Energy Convers. Manag.*, 2015.
- [16] Ying-Ying Cheng, P. P. . Chan, and Zhi-Wei Qiu, "Random forest based ensemble system for short term load forecasting," in *2012 International Conference on Machine Learning and Cybernetics*, 2012, pp. 52–56.
- [17] K. Gajowniczek and T. Zabkowski, "Short term electricity forecasting using individual smart meter data," in *Procedia Computer Science*, 2014.
- [18] R. Garcia-Valle and J. A. P. Lopes, *Electric vehicle integration into modern power networks*. 2013.
- [19] A. Baliyan, K. Gaurav, and S. Kumar Mishra, "A review of short term load forecasting using artificial neural network models," in *Procedia Computer Science*, 2015.
- [20] L. Ekonomou, "Greek long-term energy consumption prediction using artificial neural networks," *Energy*, 2010.
- [21] P. M. Lerman, "Fitting Segmented Regression Models by Grid Search," *Appl. Stat.*, vol. 29, no. 1, p. 77, 1980.
- [22] D. M. Allen, "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction," *Technometrics*, vol. 16, no. 1, pp. 125–127, Feb. 1974.