



LEIBNIZ UNIVERSITÄT HANNOVER
INSTITUT FÜR WIRTSCHAFTSINFORMATIK

Probabilistische Zeitreihenprognose von Verbraucherstrompreisen mit Methoden des Machine Learnings

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

Felix Georg Schmidt



Prüfer:

Prof. Dr. rer. nat. Michael H. Breitner

Hannover, den 01.07.2022

Inhaltsverzeichnis

Kurzfassung	I
Abbildungsverzeichnis	V
Tabellenverzeichnis	VIII
Abkürzungsverzeichnis	IX
Symbolverzeichnis	XI
1 Einleitung	1
2 Theoretische Grundlagen	3
2.1 Verbraucherstrompreis	3
2.1.1 Definition und Abgrenzung	3
2.1.2 Internationale Disparitäten und ihre Ursachen	3
2.1.3 Bestandteile	6
2.2 Zeitreihenprognosen	8
2.2.1 Definition und Überblick	8
2.2.2 Optionen der Zeitreihenprognose	9
2.2.3 Statistische Methoden der Zeitreihenprognose	11
2.3 Machine Learning	13
2.3.1 Überblick und Einordnung	13
2.3.2 Supervised, Unsupervised und Reinforcement Learning	15
2.4 Zeitreihenprognose mit Machine Learning	17
3 Parameter und Prognosemethoden	21
3.1 Parameter	21
3.1.1 Parameterauswahl	21
3.1.2 Datengrundlage	24
3.1.3 Datenaufbereitung	29
3.1.4 Parametervalidierung	42
3.2 Feedforward Neural Network	47
3.3 Long Short-Term Memory Network	50

3.4	Implementierung in Python	52
3.4.1	Entwicklungsumgebung und Programmaufbau	52
3.4.2	Implementierung der Datenaufbereitung	55
3.4.3	Implementierung der Korrelationsanalyse	59
3.4.4	Implementierung des FNN	60
3.4.5	Implementierung des LSTM	65
3.4.6	Anpassung der Hyperparameter	68
4	Ergebnisse	77
4.1	Ergebnisse des FNN	77
4.2	Ergebnisse des LSTM	83
4.3	Ergebnisvergleich	89
5	Implikationen und Limitationen	92
5.1	Methoden	92
5.2	Vergangenheitsdaten	94
5.3	Prognosen	95
6	Diskussion	96
7	Fazit und Ausblick	100
	Literatur	102
A	Anhang	111
A.1	Programmcode	111
A.1.1	main.py	111
A.1.2	data_preparation.py	112
A.1.3	correlation_analysis.py	126
A.1.4	FNN.py	128
A.1.5	LSTM.py	136
B	Research Summary	144
B.1	Einleitung	144
B.2	Parameter	144
B.3	Methoden	146
B.4	Ergebnisse	149

B.5 Diskussion	152
B.6 Fazit und Ausblick	155
Ehrenwörtliche Erklärung	157

1 Einleitung

Der deutsche Verbraucherstrompreis zählt zu den höchsten weltweit. Die Entwicklung der letzten 22 Jahre zeigt eine stetige Steigerung um insgesamt 119 % (Breitkopf (2022) und Statistisches Bundesamt (2022a)). In der jüngeren Vergangenheit wurde seine Spitzenposition jedoch von anderen Staaten übernommen. Die volatile Wirtschaftslage, primär hervorgerufen durch die Corona-Pandemie und den Ukraine-Konflikt, führt zu weltweit steigenden Rohstoffpreisen. Durch die enormen Schwankungen am Energiemarkt stellen mittel- bis langfristige Prognosen eine Herausforderung dar. War beispielsweise der US-amerikanische Rohöl-Preis während des Pandemiebeginns kurzzeitig negativ, liegt er seit Ausbruch des Ukraine-Konflikts mittelfristig über der Marke von 100 US-Dollar je Barrel (Vgl. Abb. 8). Dies führt für Verbraucher und Unternehmen zu Unsicherheiten bezüglich der langfristigen Entwicklung der Rohstoffpreise. Durch die Ausweitung des Einsatzes alternativer Technologien auf Basis der Energiewende, beispielsweise in Form von batterieelektrischen Fahrzeugen oder Wärmepumpen, steigt die Bedeutung des Strompreises für die Verbraucher. Aus der gewerblichen Perspektive beeinflusst er aber auch die Entwicklung neuer Produkte und kann zu einem Innovationstreiber in der Elektroindustrie werden. Ein langfristiger Ausblick auf die Entwicklung des Verbraucherstrompreises kann hier Klarheit schaffen.

Ziel dieser Arbeit ist es, die existierende Forschungslücke einer Langfristprognose des deutschen Verbraucherstrompreises zu schließen. Vergangene Arbeiten wie beispielsweise jene von Pao (2007) zeigen, dass die Künstlichen Neuronalen Netze, als ein Bestandteil des Machine Learnings, den klassischen Methoden mit Blick auf langfristige Prognosen überlegen sind. Unter dem Einsatz zweier solcher Methoden, werden in dieser Arbeit probabilistische Zeitreihenprognosen erarbeitet. Dabei werden auch externe Faktoren berücksichtigt, um ein möglichst präzises und belastbares Ergebnis zu erreichen. Zu ihnen zählen quantifizierbare Kennzahlen, beispielsweise aus der Wirtschaft. Einen Einfluss auf den Verbraucherstrompreis haben auch kurzfristige innen- und außenpolitische Entscheidungen. Sie sind nicht direkt quantifizierbar und ein unvorhersagbarer Einflussfaktor auf den Strompreis (Yousefi et al. (2019):S. 2911). Da sie weder abbildbar sind noch mit Sicherheit umgesetzt werden, finden sie in dieser Arbeit keinen Einzug in die Prognose. Mit

einem abschließenden Vergleich wird die präzisere Methode identifiziert.

Zu Beginn der Arbeit werden im zweiten Abschnitt die theoretischen Grundlagen des Verbraucherstrompreises, der Zeitreihenprognosen sowie des Machine Learnings erläutert. Der Abschnitt schließt mit dem Überblick über vergangene Arbeiten zu Zeitreihenprognosen mit Machine-Learning-Methoden. In Abschnitt 3 folgen die verwendeten Parameter und Prognosemethoden. Dabei werden zuerst potenzielle Parameter identifiziert sowie die Daten akquiriert und aufbereitet. Eine statistische Untersuchung wird zeigen, welche Parameter für ein Prognosemodell zu verwenden sind. Nach der Vorstellung der beiden Prognosemethoden wird die vollständige Implementierung und Optimierung in Python erläutert. Es folgt Abschnitt 4 mit den Ergebnissen der Arbeit sowie deren Vergleich. Auf Implikationen und Limitationen der Methoden und Daten wird im fünften Abschnitt eingegangen. Nach einer anschließenden Diskussion und kritischen Betrachtung der Ergebnisse, auch im Kontext der jüngsten Entwicklungen, schließt diese Arbeit mit einem Fazit und einem Ausblick auf weitere Möglichkeiten in diesem Forschungsfeld.

7 Fazit und Ausblick

Das Ziel dieser Arbeit war die Erarbeitung einer probabilistischen Langfristprognose des deutschen Verbraucherstrompreises. Dafür wurde zuerst auf die theoretischen Grundlagen des Verbraucherstrompreises, der Zeitreihenprognose und des Machine Learnings eingegangen. Anhand vergangener Arbeiten diesen Kontexts war eine Identifikation von zwei Methoden zur Durchführung der Prognosen möglich. Mit dem Feedforward Neural Network und dem Long Short-Term Memory Network wurden Methoden gewählt, die in der Zeitreihenprognose bereits ihr Potenzial gezeigt haben. Während Erstere als eine universelle Methode gilt und auch Anwendungserfolge außerhalb der Zeitreihenprognose aufweisen kann, handelt es sich bei Zweiterer um eine Methode mit Eigenschaften, die explizit auf solche Anwendungsfälle ausgerichtet ist. Nach den Prognosemethoden wurden begleitende Feature-Parameter identifiziert, ihre Daten gesammelt und auf Validität geprüft. Es hat sich gezeigt, dass sechs von elf Zeitreihen die Kriterien erfüllen und mit ihren Prognosen eingesetzt werden können. Anschließend wurde die vollständige Implementierung in Python sowie die Konfiguration und Optimierung der Modelle transparent dargestellt.

Durch die Ergebnisse wird deutlich, dass das FNN in diesem Anwendungsfall dem LSTM überlegen ist. Mit einem Mean Squared Error auf den Validierungsdaten von $0,000948 \text{ €}^2/\text{kWh}^2$ gegenüber $0,001273 \text{ €}^2/\text{kWh}^2$ ist der Fehler des FNN um 25,5 % geringer. Des Weiteren ist die Durchlaufzeit des LSTM durchschnittlich um den Faktor 2,3 höher. Mit Blick auf diese beiden Punkte, lässt sich damit auch ein Kosten-Nutzen-Vorteil für das FNN bestätigen.

Die LSTM-Prognose für den Zeitraum der Jahre 2022 bis 2025 ähnelt dem historischen Verlauf des Verbraucherstrompreises. Hier fällt der Anstieg bis Ende 2025 mit 36,34 ct/kWh moderat aus. Dagegen führt die Prognose des FNN zu dem Ergebnis, dass der Verbraucherstrompreis in diesem Zeitraum deutlich stärker als bisher ansteigen wird. Im Dezember des Jahres 2025 soll nach dieser Prognose der Preis bei 49,42 ct/kWh liegen. Beide Ergebnisse weisen außerdem neben charakteristischen Mustern einen Einbruch der Prognosen von Juni auf Juli 2022 auf, der vermutlich auf den Wegfall der EEG-Umlage zurückzuführen ist.

Die Prognoseergebnisse beider Modelle sind nicht inflationsbereinigt, was die ho-

he Steigerung des Verbraucherstrompreises nach der FNN-Prognose, gerade mit Blick auf die aktuelle Inflationsrate, relativiert. Außerdem zeigt ein Vergleich mit bereits verfügbaren realen Daten, dass der reale Verbraucherstrompreis im April 2022 höher liegt als die Prognose des FNN und LSTM für diesen Monat. Das führt zu dem Schluss, dass die verwendeten aktuellen Prognosen der Feature-Parameter diese Entwicklung noch nicht abdecken und folglich auch die Prognosen der Modelle unterhalb der realen Daten liegen.

In dieser Arbeit, als erste ihrer Art, ist es gelungen, die Forschungslücke im Bereich der Langfristprognosen des deutschen Verbraucherstrompreises zu schließen. Es wurden geeignete Parameter sowie optimale Modellkonfigurationen identifiziert und damit stabile Prognoseergebnisse für den deutschen Verbraucherstrompreis erarbeitet. Dies zeigen auch der geringe Validierungsfehler und seine klare Konvergenz während des Trainings. Auftretende Limitationen wurden erkannt und konnten im überwiegenden Teil beseitigt werden.

Unter anderem wurde die mangelnde Robustheit gegenüber den initialisierten Gewichten der Modelle mittels einer ausführlichen Analyse kompensiert. Dadurch war es nach vorheriger Filterung möglich, ein 95%-Prognoseintervall der beiden Modelle zu erstellen. Eine weitere potenzielle Limitation stellen die verwendeten Daten dar. Ohne eine ausreichende Datenqualität und -frequenz, ist eine verlässliche Prognose nicht möglich. Die ausschließliche Verwendung von Daten bekannter Institute, Ministerien und Fachverbänden führen zu einer höchstmöglichen Qualität der Daten. Zeitreihen, die keine monatliche Auflösung abbildeten, konnten mit Inter- und Extrapolationen zu stetigen Entwicklungen transformiert werden.

Im Bereich der verwendeten Daten ist jedoch noch weiteres Potenzial denkbar. In der Theorie bietet eine Vergrößerung der Datenbasis die Möglichkeit präzisere und validere Prognosen zu erstellen. Dies gilt sowohl für den zeitlichen Horizont, als auch für das Heranziehen weiterer externer Faktoren in Form von Feature-Parametern. Gerade mit Blick auf die volatilen Entwicklungen der aktuellen Zeit, ist auch ein stetiges Aktualisieren der Datenbasis unumgänglich. Inwiefern weitere Ansätze aus dem Bereich des Machine Learnings bei der Bewältigung dieser fundamentalen Aufgabe Erfolg haben, wird sich in weiterführenden Arbeiten zeigen.