

# Data Science Methods for Probabilistic Natural Gas Price Forecasting

## Masterarbeit - Exposé

---

Erdgas gehört bereits heute zu den wichtigsten Primärenergiequellen Deutschlands und der politisch forcierte Atom- und Kohleausstieg bei der Stromerzeugung wird die Rolle von Erdgas mittelfristig zusätzlich erhöhen.

Die Masterarbeit entstand in Zusammenarbeit mit der Gas Portfolio-Management Abteilung der RWE Supply and Trading GmbH (RWEST). In der Arbeit wurden zwei Preisvorhersagemodelle mithilfe von modernen Methoden des Machine Learnings entwickelt. Eines für das Day-Ahead-Produkt, bei dem die Lieferung am folgenden Werktag erfolgt und ein weiteres für das Month-Ahead-Produkt, bei dem die Lieferung im Folgemonat erfolgt. Die Modellentwicklung basierte dabei einerseits auf der Auswertung der wissenschaftlichen Literatur und andererseits auf einer ausführlichen explorativen Datenanalyse. Insgesamt konnte die Prognosegüte der Modelle um  $\approx 10\%$  für beide Produkte verbessert werden.

Die Modelle wurden in Python implementiert. Zur Auswertung und Visualisierung der Ergebnisse wurde außerdem R und Latex genutzt. Die besonderen statistischen Eigenschaften der Daten machten eine manuelle Implementierung der den Modellen zugrundeliegenden Verteilungsfunktion unumgänglich. Diese darauf aufbauend entwickelten probabilistischen Machine-Learning Modelle können die gesamte Verteilung zukünftiger Preise vorhersagen. Sie berücksichtigen dabei eine Vielzahl interner und externer Faktoren. Das macht die Modelle nicht nur für die RWEST interessant, sondern beispielsweise auch für Betreiber von Erdgasspeichern, um die erwartete Ein- bzw. Ausspeisung zu prognostizieren oder für das Risiko- und Portfoliomanagement von Stromproduzenten.

Die Verteilungsfunktion wurde im Rahmen der Arbeit als Python Paket auf [PiPy](#) veröffentlicht. Darüberhinaus wurde eine [Dokumentation](#) für das Paket erstellt, welche anderen Entwicklern den Einstieg erleichtert. Die Entwicklung des Pakets erfolgt öffentlich auf [GitHub](#).

Auf dieser Arbeit aufbauende Forschung wurde bereits am 9. September 2020 auf der [International Ruhr Energy Conference \(INREC\)](#) Konferenz präsentiert. Im Oktober 2020 erfolgt außerdem eine Präsentation auf dem International Symposium of Forecasters. Zudem wird eine Veröffentlichung der weiterführenden Arbeit in einem Journal der Energieinformatik angestrebt. Die Arbeit zeigt, dass moderne Methoden der Informatik und Statistik im Energiebereich von hoher Bedeutung sind. Zudem liefert die Arbeit nicht nur einen akademischen und praktischen Beitrag, sie liefert außerdem einen Beitrag zu quelloffener Software.