

# Kurzfassung

Kurzfrist-Lastprognosen sind für Netzbetreiber ein wichtiges Verfahren, um potenzielle kritische Energienetz Zustände vorherzusagen und damit den Einsatz ihrer finanziellen und materiellen Vermögenswerte zu optimieren. Da Machine-Learning (ML) Methoden herkömmliche Prognosetechniken übertreffen können, werden sie immer häufiger eingesetzt. Jedoch reicht es bei Entscheidungen mit hohem Risiko, wie sie für den Netzbetrieb getroffen werden, häufig nicht aus, nur die Vorhersage eines Modells zu kennen, und es bedarf ein tieferes Verständnis, wie die Vorhersage gebildet wird. Da neuronale Netze zu komplex sind, um deren interne Prozesse zu interpretieren, entsteht der Bedarf an Erklärungsmethoden, die darauf abzielen, das Verhalten des Modells zu erklären und Vertrauen in dieses aufzubauen. In dieser Arbeit wird eine modellagnostische Erklärungsmethode in Form einer Signifikanzkarte (Engl.: saliency map) implementiert, die Erklärungen durch gezielte Perturbation der Eingabedaten generiert. Das Verhalten wird an einem Encoder-Decoder Lastprognosemodell getestet und bewertet. Die Methode ist, im Gegensatz zu anderen Methoden, dazu in der Lage, die Signifikanz der Eingabedaten nicht nur untereinander, sondern auch über die für Zeitreihen wichtige Zeitdimension hinweg darzustellen. Die Ergebnisse zeigen, dass die Methode in der Lage ist, verständliche Erklärungen für kurzfristige Vorhersagen des neuronalen Netzes zu liefern.

**Stichwörter:** Kurzfrist-Lastprognosen, Interpretierbares Machinelles Lernen, Erklärbare Künstliche Intelligenz, Saliency Map

# Abstract

Short-term load forecasting is an important process for power system operators to be able to predict potential critical power grid states and thereby to optimize the use of their financial and physical assets. As deep neural networks can outperform traditional forecasting techniques, they have become increasingly popular for prediction tasks. However, for high stake decisions, like those made for power system operations, it is often not sufficient to know only the prediction of a model, and a deeper understanding of how the prediction is formed is required. Because deep neural networks are too complex to be interpretable per se, the need for explanation methods arises, that aim to explain and build trust in the behavior of the model. In this thesis, a model-agnostic explanation method in the form of saliency maps is implemented, which generates explanations by perturbations of the input data. The method is tested and evaluated on an encoder-decoder load forecasting model. In contrast to other state-of-the-art explanation methods, the method is able to highlight feature importance, not only in the feature dimension, but also in the time dimension, which is critical for time series prediction tasks. The results show, that the method is able to generate comprehensible explanations for short-term predictions of the neural network.

**Keywords:** Short-term Load Forecasting, Interpretable Machine Learning, Explainable Artificial Intelligence, Saliency Map

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Short Term Load Forecasting . . . . .	1
1.2	Interpretability of Neural Networks . . . . .	1
1.3	Goal . . . . .	2
<b>2</b>	<b>Theoretical Background</b>	<b>3</b>
2.1	Encoder-Decoder Recurrent Neural Network . . . . .	3
2.2	Relevant terms and aspects . . . . .	4
2.2.1	Interpretability vs Explainability . . . . .	4
2.2.2	Classification of Explanation Methods . . . . .	5
2.2.3	Properties of Explanations . . . . .	6
2.3	State of the Art Explanation Methods . . . . .	8
2.3.1	Post-hoc Model-agnostic Methods . . . . .	8
2.3.2	Meaningful Perturbations . . . . .	9
<b>3</b>	<b>Implementation</b>	<b>11</b>
3.1	Data Preparation . . . . .	11
3.2	Creation of References . . . . .	12
3.3	Training the Neural Network for Time Series Prediction . . . . .	12
3.4	Optimizing the Mask Parameters . . . . .	13
<b>4</b>	<b>Evaluation</b>	<b>17</b>
4.1	Dataset . . . . .	17
4.2	Neural Net Parameters . . . . .	18
4.3	Results . . . . .	19
4.3.1	Scenario 1: Monday Maps . . . . .	19
4.3.2	Scenario 2: Sunday Maps . . . . .	19
4.3.3	Scenario 3: High Load Maps . . . . .	20
4.3.4	Scenario 4: Low Load Maps . . . . .	20
4.4	Properties of the Proposed Method . . . . .	20
<b>5</b>	<b>Conclusion</b>	<b>33</b>
5.1	Discussion . . . . .	33
5.2	Outlook . . . . .	33
	<b>List of Figures</b>	<b>35</b>
	<b>List of Tables</b>	<b>37</b>

*Contents*

**Bibliography**

**40**