



Mittelstand-Digital
**Zentrum
Hannover**

Leitfaden

Anomalie Erkennung



Quelle: sdecoret/stock.adobe.com

Gefördert durch:



Bundesministerium
für Wirtschaft
und Energie

Mittelstand-
Digital

aufgrund eines Beschlusses
des Deutschen Bundestages

Anomalie Erkennung

Häufig sind geeignete Daten und erste Ideen für den Einsatz einer KI-Methode bereits vorhanden. Die konkrete Umsetzung stellt jedoch oftmals eine Herausforderung dar. Der auf GitHub bereitgestellte Leitfaden zeigt daher ein strukturiertes Vorgehen zur Vorbereitung und Umsetzung eines KI-Modells zur Anomalieerkennung und unterstützt die Übertragung auf eigene Anwendungsfälle.

Der auf GitHub bereitgestellte Leitfaden „Anomalie Erkennung“ bietet eine umfassende und an-fängerfreundliche Einführung in die KI-basierte Anomalie Erkennung. Dies geschieht am Beispiel von Energieverbrauchsdaten unter Nutzung des öffentlich verfügbaren LEAD-1.0-Datensatzes. Der Leitfa-den nutzt dabei Jupyter Notebooks, eine interaktive Webanwendung, die es erlaubt ausführbaren Pro-grammcode, mathematische Formeln, Visualisierungen (Diagramme) und erklärenden Text zu vereinen.

Ziel ist es, den gesamten KI bzw. Machine Learning-Lebenszyklus nachvollziehbar abzubilden – von der ersten Datenanalyse über die Datenaufbereitung bis hin zum Training und der Evaluierung eines Vorhersagemodells. Im Mittelpunkt steht die Identifikation ungewöhnlicher Muster in den analysier-ten Werten, hier im Energieverbrauch, beispielsweise atypischer Sensorwerte oder unerwarteter Verbrauchsspitzen. Die Nutzenden lernen dabei, ein leistungsfähiges XGBoost-Klassifizierungs-mo-dell zu entwickeln, das normales von potenziell anomalem Verhalten unterscheiden kann.

Die Anleitung ist in **drei aufeinander aufbauende Phasen** gegliedert, die in einzelnen Bereichen der Anleitung zu finden sind. In der ersten Phase erfolgt eine Einführung in die Grundlagen der Anomalie-erkennung sowie in den LEAD-1.0-Datensatz, der stündliche Energieverbrauchsdaten von 200 Ge-bäuden über einen Zeitraum von einem Jahr umfasst. Im Rahmen der explorativen Datenanalyse werden Datenverteilungen untersucht, fehlende Werte identifiziert und typische saisonale sowie wöchentliche Verbrauchsmuster visualisiert.

In der zweiten Phase werden die Rohdaten in ein für Machine-Learning-Verfahren geeignetes Format überführt. Dazu gehören die Erstellung eines konsistenten Stundenrasters, die Imputation fehlender Werte mithilfe gebäudespezifischer Medianwerte sowie die Entwicklung zusätzlicher Merkmale, dar-unter zyklische Zeitmerkmale sowie zeitlich verzögerte Merkmale für vergangene Stunden und Tage. Zudem wird eine Aufteilung der Daten in Trainings-, Validierungs- und Testdatensätze vorgenommen, um eine realistische Modellbewertung sicherzustellen.

In der abschließenden Phase wird ein XGBoost-Klassifikationsmodell trainiert und optimiert. Dazu wird zunächst eine statistische Referenzmethode etabliert, bevor durch Hyperparameter-Optimierung die Modellleistung verbessert wird. Die Bewertung erfolgt anhand etablierter Metriken wie Recall, Precision und F1-Score, um die Fähigkeit des Modells zur zuverlässigen Erkennung von Anomalien zu beurteilen.

Durch die Kombination von Codebeispielen und Erklärung der einzelnen Schritte wird ein praxisori-entierter Zugang zu den vorgestellten Methoden geschaffen. Die Ansätze können auf eigene Daten-sätze und spezifische Anforderungen im Bereich der Anomalieerkennung übertragen und darüber hinaus auf weitere Anwendungsfälle angepasst werden.

Link zum Leitfaden:

https://git.l3s.uni-hannover.de/mdzh/anomaly_detection

Autorin

Dr. Claudia Niederée
Research Group Leader

L3S Research Center
Leibniz Universität Hannover
Appelstrasse 9a
30167 Hannover



Mittelstand-Digital
**Zentrum
Hannover**