Modulbeschrieb

Explainable AI

Allgemeine Informationen

Modulbezeichnung

Explainable AI

Modulkategorie

Fachliche Vertiefung

Modulverantwortliche/r

Klaus Frick

Anzahl der Credits

3

Ziele, Inhalt und Methoden

Lernziele, zu erwerbende Kompetenzen

Machinelles Lernen und speziell Deep Learning sind inhärente Bestandteile nahezu aller Bereiche unseres Alltags. Smartphones, (autonome) Fahrzeuge, medizinische Diagnostik, Börsenprognosen, Epidemiologie, Produktionsanlagen, Roboter-Mensch Interaktion u.v.m. sind abhängig von Entscheidungen, die durch Machine Learning Algorithmen getroffen werden. Typischerweise fokussiert sich die Forschung bzw. die universitäre Ausbildung auf das Training und hier insbesondere auf die Genauigkeit *(accuracy)* der Vorhersagen von solchen Modellen. Explainable AI (XAI) ist ein junges Forschungsgebiet, das der Frage nach geht: *«Warum wurde diese Entscheidung vom Modell getroffen?»* XAI ist die Grundlage für Erklärbarkeit (*explainability*) und Interpretierbarkeit (*interpretability*) von Machine Learning Modellen und ermöglicht somit den Schritt von blossen Vorhersagen (predictions) zur Wissenserzeugung aus Daten (knowledge discovery).

Modulinhalt

In diesem Seminar führen wir eine genaue Begriffsdefinition von XAI durch und lernen die grundlegendenden Ansätze:

1) Einführung, Begriffsklärung und Beispiele zu XAI
2) Intrinsisch erklärbare Methoden: Allgemeine lineare Modelle, Entscheidungsbäume & Co
3) Model-agnostische Ansätze für XAI: Partial Dependence Plots, LIME, SHAP
4) XAI und Deep Learning: Activation Atlas, CAM,

Lehr- und Lernmethoden

Dies ist eine interaktive Buchdiskussionsrunde, welche sich jede Woche im Semester für zwei Stunden trifft. Während dieser Treffen wird über vorgegebene Abschnitte im Buch diskutiert. Die Aufgabe der Teilnehmer ist es, diese Abschnitte vorgängig gelesen und verstanden zu haben, so dass sie an der Diskussion aktiv teilnehmen können. Neben diesem theoretischen Teil gibt es auch einen praktischen, wo die Teilnehmer die gelesene und diskutierte Theorie mit Python Übungen vertiefen. Überdies gibt es theoretische Übungen, welche die Studierenden selbständig zu lösen haben, um zu zeigen, dass der Stoff verstanden wurde.

Die wöchentlichen Diskussionen finden online/hybrid via MS Teams statt und können von Studierenden aller drei Standorte besucht werden.

Voraussetzungen, Vorkenntnisse, Eingangskompetenzen

Grundlagen der Linearen Algebra und der Analysis.
Grundlagen der Statistik bzw. des Maschinellen Lernens: Lineare Modelle, Logistische Regression und Entscheidungsbäume.
Kenntnisse über neuronale Netze/Deep Learning von Vorteil aber nicht notwendig.

Bibliografie

L. Gianfagna and A. Di Cecco: *Explainable AI with Python.* Springer 2021

Leistungsbewertung

Prüfungsart

schriftliche Prüfung

Zulassungsbedingungen

Besuch von 75% der Diskussionen, Bearbeiten von 75% der Übungsserien/Python Übungen

Prüfungsdauer

2h

Hilfsmittel

Das Buch *"Explainable AI with Python"*