

KARTA PRZEDMIOTU

Nazwa przedmiotu: Wyjaśnialne uczenie maszynowe (1000-319bEML)

Nazwa w języku polskim: Wyjaśnialne uczenie maszynowe

Nazwa w jęz. angielskim: Explainable machine learning

Dane dotyczące przedmiotu:

Jednostka oferująca przedmiot: Wydział Matematyki, Informatyki i Mechaniki

Przedmiot dla jednostki: Wydział Matematyki, Informatyki i Mechaniki

Domyślny typ protokołu dla przedmiotu:

Egzamin

Język wykładowy:

angielski

Skrócony opis:

Celem przedmiotu jest Poznanie zagadnień, metod oraz technik wyjaśniania złożonych modeli uczenia maszynowego. Modele predykcyjne są coraz bardziej złożone, komitety drzew, głębokie sieci neuronowe to modele o tysiącach parametrów. Dla modeli o takiej wymiarowości łatwo stracić kontrolę nad tym czego model się wyuczył.

Podczas tego przedmiotu omówimy narzędzia do analizy struktury modelu traktowanego jako czarna skrzynka, oraz do analizy predykcji z tego modelu. Pozwoli to na zwiększenie zaufania do modelu, poprawę skuteczności modelu, oraz możliwość wyciągnięcia użytecznej wiedzy z modelu. Poznamy najpopularniejsze metody wyjaśnień, przedyskutujemy ich silne i słabe strony tak by uczestnik zajęć miał niezbędne kompetencje do dalszego zgłębiania literatury z tego obszaru.

Opis:

- Wprowadzenie do wyjaśnialnej sztucznej inteligencji, interpretowalnego uczenia maszynowego i zagadnienia dyskryminacji
- Metody analizy warunkowej modeli predykcyjnych: metoda Break-Down, Break-Down z interakcjami, SHAP, ASV
- Metody analizy modeli przez perturbacje: metoda LIME, LORE
- Metody analizy modeli kontekstowych i testowania wrażliwości modeli: Ceteris Paribus, Partial Dependence, Accumulated Local Effects
- Metody oceny ważności zmiennych: Variable Importance by Pertmutations, Model Class Reliance
- Sprawiedliwość i stronniczość
- Objaśnienia specyficzne dla sieci neuronowych

Rodzaj przedmiotu

monograficzne

Wymagania (lista przedmiotów)

Głębokie sieci neuronowe (1000-317bDNN)

Uczenie statystyczne (1000-317bSML)

Przynależność do grup przedmiotów w cyklach:

Opis grupy przedmiotów	Cykl pocz.	Cykl kon.
Przedmioty do wyliczania średniej dla programu ERASMUS (1000-ERASMUS)	2022	
Przedmioty obowiązkowe dla II roku Machine Learning (1000-ML2-OBW)	2022	
Przedmioty obieralne dla informatyki (1000-OBIER)	2023	

Punkty przedmiotu w cyklach:

<bez przypisanego programu>			
Typ punktów	Liczba	Cykl pocz.	Cykl kon.
Europejski System Transferu i Akumulacji Punktów (ECTS)	6	2022	