

Sylabus pro předmět

POKROČILÉ STROJOVÉ UČENÍ A BIG DATA

Kód předmětu:	4IT533
Název v jazyce výuky:	Pokročilé strojové učení a big data
Název česky:	Pokročilé strojové učení a big data
Název anglicky:	Advanced machine learning and big data
Počet přidělených ECTS kreditů:	5
Forma výuky kurzu:	prezenční; 0/4 (počet hodin přednášek týdně / počet hodin cvičení týdně) při semestrální výuce mimosemestr; 0/52 (počet hodin přednášek za období / počet hodin cvičení za období) při mimosemestrální výuce
Forma ukončení kurzu:	zkouška
Jazyk výuky:	čeština
Doporučený typ a ročník studia:	magisterský navazující (druhý cyklus): 1
Semestr:	— obsah této položky nebyl definován —
Vyučující:	Ing. Pavel Zimmermann, Ph.D. (garant)
Omezení pro zápis:	žádné
Doporučené doplňky kurzu:	žádné
Vyžadovaná praxe:	žádná

Zaměření předmětu:

Tento předmět se zaměřuje na rozšíření znalostí a dovedností v oblasti strojového učení nad rámec bakalářského studia a dále na zpracování velkých dat. Studenti budou pracovat s moderními technikami strojového učení a algoritmy pro analýzu velkých datových souborů. Dále se budou zabývat prací s distribuovanými výpočty a cloudovými službami.

Výsledky učení:

Po absolvování předmětu budou studenti schopni:

- Pochopit a aplikovat techniky pokročilého strojového učení.
- Propojit praktické zadání s informacemi obsaženými v datech a pomocí postupů strojového učení ho vyřešit.
- Řešit aplikované úlohy nad rámec základních postupů strojového učení probraných v rámci bakalářského studia.
- Analyzovat a zpracovat velké datové soubory a použít k tomu moderní algoritmy strojového učení.

Obsah předmětu:

- Opakování a rozšíření úloh strojového učení probíraných v základních bakalářských kurzech
- Detekce anomálií, snižování dimenze dat
- Vybrané speciální úlohy, např. propenzitní úlohy nebo řadící modely.
- Aplikace strojového učení na časové řady, sekvenční data a data s temporální složkou.
- Konvoluční sítě s časovým prvkem (Temporal Convolutional Networks (TCN))
- Strojové učení se v analýze přežívání
- Frameworky pro zpracování velkých dat
- Procesování dat: batch vs stream, MapReduce
- Analýza dat: vizualizace, modelování pomocí PySpark
- Nástroje pro Big data, ekosystémy and platformy

Způsob studia, metody výuky a studijní zátěž (počet hodin):

Druh	Počet hodin studijní zátěže
	Prezenční studium
Účast na cvičeních/seminářích/tutoriálech	48
Příprava semestrální práce	25
Příprava na průběžný test (testy)	33
Příprava na závěrečnou ústní zkoušku	24
Celkem	130

Způsoby a kritéria hodnocení:

Druh	Váha
	Prezenční studium
Vypracování semestrální práce	30 %
Absolvování průběžného testu (testů)	40 %
Absolvování závěrečné ústní zkoušky	30 %
Celkem	100 %

Hodnocení:

Zkoušky

1 Výborně (90 – 100 %)

2 Velmi dobře (75 – 89 %)

3 Dobře (60 – 74 %)

4 Nedostatečně (0 – 59 %)

Zápočty

Z Započteno

NZ Nezapočteno

Zvláštní podmínky a podrobnosti:

Tento kurz byl spolufinancován Evropskou unií, Národním plánem obnovy z projektu NPO_VŠE_MSMT-16603/2022

Literatura:

Typ*	Autor	Název	Místo vydání	Nakladatel	Rok	ISBN
Z	GÉRON, Aurélien	Hands-on machine learning with Scikit-Learn, Keras and TensorFlow : concepts, tools, and techniques to build intelligent systems	Beijing	O'Reilly	2019	978-1-4920-3264-9

* Z – základní literatura D – doporučená literatura