**Nazwa przedmiotu:**

Inżynieria uczenia maszynowego

**Koordynator przedmiotu:**

Rajmund Kożuszek

**Status przedmiotu:**

Fakultatywny dowolnego wyboru

**Poziom kształcenia:**

Studia I stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Sztuczna inteligencja

**Kod przedmiotu:**

IUM

**Semestr nominalny:**

7 / rok ak. 2021/2022

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. liczba godzin kontaktowych – 47 godz., w tym
obecność na wykładach 30 godz.,
spotkania projektowe 15 godz.,
konsultacje 2 godz.,
2. praca własna studenta – 70 godz., w tym
przygotowanie do kolokwium 25 godz.,
przygotowanie do zadania projektowego 15 godz.,
realizacja projektu 30 godz.
Łączny nakład pracy studenta wynosi 117 godz., co odpowiada 4 pkt. ECTS.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1,61 pkt. ECTS, co odpowiada 47 godz. kontaktowym

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

2,05 pkt. ECTS, co odpowiada 15 godz. spotkań projektowych plus 45 godz. przygotowania i realizacji projektu

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład: | 30h |
| Ćwiczenia: | 0h |
| Laboratorium: | 0h |
| Projekt: | 15h |
| Lekcje komputerowe: | 0h |

**Wymagania wstępne:**

Podstawy programowania, Wprowadzenie do sztucznej inteligencji

**Limit liczby studentów:**

60

**Cel przedmiotu:**

Celem przedmiotu jest zapoznanie studentów z całością procesu tworzenia modeli uczenia maszynowego: od zbierania wymagań biznesowych i definiowania zadań modelu, po analizę i diagnostykę ich działania.
Przedmiot przedstawia modele uczenia maszynowego jako część działających produkcyjnie systemów informatycznych. W tym kontekście przedstawiany jest całościowy proces tworzenia modeli: od momentu zbierania wymagań biznesowych i definiowania zadań modelowania, poprzez etap tworzenia i wdrażania modeli predykcyjnych, aż po analizę i diagnostykę sposobu ich działania. Poruszany jest również szerszy kontekst związany z rozwojem technik sztucznej inteligencji oraz jego konsekwencjami.

**Treści kształcenia:**

Wykład:
1. Wprowadzenie. (2h)
i) Sprawy organizacyjne, cele i program zajęć, zasady zaliczania.
ii) Specyfika systemów stosujących modele predykcyjne.
iii) Cykl opracowywania modeli predykcyjnych.
iv) Dostępne standardy: CRISP-DM, KDD, SEMMA.
Część I: Przygotowanie modeli predykcyjnych.
2. Definiowanie zadań modelowania. (2h)
i) Zadania biznesowe, a zadania modelowania: zbieranie wymagań, precyzowanie i definiowanie zadań, ustalanie kryteriów sukcesu.
ii) Standardowe zadania: regresja, klasyfikacja, grupowanie, modelowanie sekwencji, analiza szeregów czasowych, modele generatywne.
iii) Wstępny dobór technik modelowania.
3. Inżynieria cech i tworzenie zbiorów do modelowania (4h)
i) Atrybuty wejściowe/wyjściowe:
(1) typy zmiennych, normalizacja, odstające wartości atrybutów,
(2) dane brakujące, imputacja jednokrotna/wielorotna,
(3) dane niezbalansowane/asymetryczne, równoważenie liczby przykładów,
(4) duże zbiory danych, próbkowanie i jego konsekwencje,
(5) błędy w danych,
(6) tworzenie nowych atrybutów, interakcje,
(7) zaawansowane techniki inżynierii cech:
1. kodowanie przy pomocy drzew,
2. “feature hashing”,
3. atrybuty tekstowe, embedding,
(8) selekcja cech i redukcja wymiarowości.
ii) Tworzenie zbiorów do modelowania.
(1) Problemy z pozyskaniem odpowiedniej ilości danych, “gęstość danych”.
(2) Manualne etykietowanie danych, stosowanie usług zewnętrznych (np. Amazon turk).
(3) Zastosowanie technik automatycznego etykietowania danych – SNORKEL, SNUBA.
(4) Reprezentatywność zbioru danych.
4. Dobór funkcji celu. (2h)
i) Omówienie funkcji celu dla: klasyfikacji, regresji oraz rankingu.
ii) Optymalizacja funkcji celu, a zadanie predykcyjne.
iii) Stosowane funkcje celu, a warunki zadania biznesowego.
5. Trenowanie i ewaluacja modeli. (2h)
i) Strojenie hiper-parametrów.
ii) Znaczenie strojenia.
iii) Popularne metody: przeszukiwanie po hipersiatce, przeszukiwanie losowe, podejście bayesowskie, heurystyki.
iv) Dobór metody uczącej.
v) Porównywanie modeli (offline).
(1) Tworzenie rankingów modeli.
(2) Różnice w jakości, a stopień złożoności modeli.
vi) Modele zagregowane.
Część II: Modele na środowisku produkcyjnym.
6. Wdrażanie modelu w systemie informatycznym. (4h)
i) Elementy architektury systemów stosujących modele predykcyjne.
ii) Trenowanie i stosowanie modeli do predykcji.
iii) Zbieranie informacji diagnostycznych.
iv) Dane wejściowe/wyjściowe jako zależności wytrenowanego modelu.
v) Wersjonowanie modeli.
vi) Testy automatyczne, powtarzalność wyników.
vii) Przegląd dostępnych platform i narzędzi.
7. Monitorowanie serwowanych modeli. (2h)
i) Monitorowanie modelu.
(1) Miary techniczne.
(2) Miary biznesowe.
(3) Monitorowanie danych wejściowych.
ii) Znaczenie modeli bazowych/punktów odniesienia.
iii) Interpretacja wyników, adekwatne typy wykresów, paradoks Simpsona.
8. Ewaluacja nowych modeli. (4h)
i) Eksperymenty A/B:
(1) przydział grup eksperymentalnych, sprzężenia zwrotne,
(2) planowanie eksperymentów: dobór wielkości grupy, czas trwania, warunki zatrzymania,
(3) podejście częstościowe, a podejście bayesowskie.
ii) Wdrażanie modeli jako interwencja – przyczynowo-skutkowa ocena przydatności modelu.
(1) Metoda Causal Impact.
(2) Projektowanie interwencji.
iii) Dynamiczny wybór modelu za pomocą podejścia multi-arm bandits.
(1) Definicja zagadnienia, metoda próbkowania Thompsona.
9. Wyjaśnianie predykcji modelu. (2h)
i) Modele poddające się łatwej interpretacji, a „czarne skrzynki”.
ii) Wyjaśnianie predykcji czarnych skrzynek: LIME, DALEX.
Część III: Szerszy kontekst zastosowań modeli predykcyjnych.
10. Kwestie bezpieczeństwa i prywatności. (2h)
i) Modele zagrożeń, wycieki danych.
ii) Zagadnienie „złośliwych danych”.
iii) Ochrona danych wrażliwych, anonimizacja, prywatność różnicowa.
11. Kierunki rozwoju metod uczenia maszynowego. (2h)
i) Wpływ technologii na rozwój metod uczenia maszynowego – np.
(1) Karty GPU, jednostki TPU.
(2) Chmury obliczeniowe.
(3) Rozwiązania z obszaru „edge computing”.
(4) Repozytoria modeli.
ii) Koszty/Zyski z trenowania złożonych modeli predykcyjnych.
iii) Wpływ konkursów i benchmarków na rozwój metod UM.
iv) Aktualne kierunki rozwoju metod UM.
12. Społeczne kwestie technik sztucznej inteligencji. (2h)
i) Zaawansowane techniki modelowania, a ochrona prywatności.
ii) Koncepcja „bańki filtrującej”.
iii) Konsekwencje nadużywania predykcji.
iv) Kwestie obciążenia zbiorów danych, zagadnienie “fairness”.
Projekt:
Celem projektu jest opracowanie kompletnego systemu bazującego na UM rozwiązującego konkretne zadanie biznesowe. System powinien mieć elementy pozwalające na:
● serwowanie przygotowanych modeli,
● cykliczne ich trenowanie i podmianę/aktualizację,
● wprowadzanie i ewaluację nowych modeli,
● monitorowanie sposobu działania modeli.
Realizacja projektu będzie rozbita na kilka etapów (wyniki uzyskane po zakończeniu każdego z nich będą dyskutowane z prowadzącym):
1. analizę przedstawionego problemu biznesowego i zdefiniowanie zadania modelowania,
2. przygotowanie danych uczących i prostego modelu bazowego,
3. implementację systemu realizującego wymagania techniczne projektu,
4. przygotowanie bardziej zaawansowanego modelu rozwiązującego zadanie modelowania,
5. zestawienie eksperymentu umożliwiającego porównanie przygotowanych modeli online.
Przykładowe typy realizowanych systemów:
1. system rekomendujący filmy,
2. przewidywanie ruchów rynków finansowych,
3. kategoryzacja witryn internetowych,
4. system OCR,
5. wykrywanie obiektów na zdjęciach.

**Metody oceny:**

Wykład prowadzony w tradycyjnej formie.
Realizacja projektu będzie rozbita na kilka etapów (wyniki uzyskane po zakończeniu każdego z nich będą prezentowane i dyskutowane z prowadzącym w trakcie 2-3 godzinnego spotkania):
1. analizę przedstawionego problemu biznesowego i zdefiniowanie zadania modelowania,
2. przygotowanie danych uczących i prostego modelu bazowego,
3. implementację systemu realizującego wymagania techniczne projektu,
4. przygotowanie bardziej zaawansowanego modelu rozwiązującego zadanie modelowania,
5. zestawienie eksperymentu umożliwiającego porównanie przygotowanych modeli online.

**Egzamin:**

nie

**Literatura:**

Część I: Przygotowanie modeli predykcyjnych.
1. Sarkar, Dipanjan, Raghav Bali, and Tushar Sharma. Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems. Apress, 2017.
2. Bi, Jinbo, et al. "Dimensionality reduction via sparse support vector machines." Journal of Machine Learning Research 3.Mar (2003): 1229-1243.
3. Weinberger, Kilian, et al. "Feature hashing for large scale multitask learning." Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009.
4. Wickham, Hadley. "Tidy data." Journal of Statistical Software 59.10 (2014): 1-23. 2.
5. Bergstra, James, Dan Yamins, and David D. Cox. "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms." Proceedings of the 12th Python in science conference. 2013.
Część II: Modele na środowisku produkcyjnym.
6. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., … Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. Nips, 2494–2502. Retrieved from http://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf
7. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., … Young, M. (2014). Machine Learning : The High-Interest Credit Card of Technical Debt. NIPS 2014 Workshop on Software Engineering for Machine Learning (SE4ML), 1–9. http://doi.org/10.1007/s13398-014-0173-7.2
8. Dmitriev, Pavel, et al. "A Dirty Dozen: Twelve Common Metric Interpretation Pitfalls in Online Controlled Experiments." Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017.
9. Alipourfard, Nazanin, Peter G. Fennell, and Kristina Lerman. "Can you Trust the Trend: Discovering Simpson's Paradoxes in Social Data." arXiv preprint arXiv:1801.04385 (2018).
10. Goldstein, Alex, et al. "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation." Journal of Computational and Graphical Statistics 24.1 (2015): 44-65.
11. Brodersen, Kay H., et al. "Inferring causal impact using Bayesian structural time-series models." The Annals of Applied Statistics 9.1 (2015): 247-274.
12. Alipourfard, N., Fennell, P. G., & Lerman, K. (2018). Can you Trust the Trend: Discovering Simpson’s Paradoxes in Social Data. Retrieved from http://arxiv.org/abs/1801.04385
Część III: Szerszy kontekst zastosowań modeli predykcyjnych.
13. Herley, Cormac, and Paul C. Van Oorschot. "Sok: Science, security and the elusive goal of security as a scientific pursuit." 2017 IEEE Symposium on Security and Privacy (SP). IEEE, 2017.
14. Barreno, Marco, et al. "The security of machine learning." Machine Learning 81.2 (2010): 121-148.
15. Athalye, Anish, Nicholas Carlini, and David Wagner. "Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples." arXiv preprint arXiv:1802.00420 (2018).
16. O'neil, Cathy. Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books, 2016.

**Witryna www przedmiotu:**

https://usosweb.usos.pw.edu.pl/kontroler.php?\_action=katalog2/przedmioty/pokazPrzedmiot&prz\_kod=103A-INSZI-ISP-IUM

**Uwagi:**

(-)

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

zna podstawowe typy zadań do których stosowane są modele UM

Weryfikacja:

kolokwium

**Powiązane charakterystyki kierunkowe:** W06

**Powiązane charakterystyki obszarowe:** I.P6S\_WG.o, P6U\_W

**Charakterystyka W02:**

zna metody służące do diagnostyki działających produkcyjnie modeli UM

Weryfikacja:

kolokwium

**Powiązane charakterystyki kierunkowe:** W06

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o

**Charakterystyka W03:**

zna metody porównywania działających produkcyjnie modeli UM

Weryfikacja:

kolokwium

**Powiązane charakterystyki kierunkowe:** W06

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o

**Charakterystyka W04:**

zna metody pozwalające na wyjaśnianie predykcji dokonywanych przez modele UM

Weryfikacja:

kolokwium

**Powiązane charakterystyki kierunkowe:** W06

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o

**Charakterystyka W05:**

rozumie kwestie bezpieczeństwa związane z zastosowaniem UM

Weryfikacja:

kolokwium

**Powiązane charakterystyki kierunkowe:** W10

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o, III.P6S\_WG

**Charakterystyka W06:**

rozumie społeczne konsekwencje stosowania UM

Weryfikacja:

kolokwium

**Powiązane charakterystyki kierunkowe:** W13, W15

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WK, III.P6S\_WK

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

potrafi tworzyć systemy informatyczne stosujące komponenty bazujące na UM

Weryfikacja:

ocena projektu

**Powiązane charakterystyki kierunkowe:** U07

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o

**Charakterystyka U02:**

umie definiować zadania UM na podstawie zadań biznesowych

Weryfikacja:

ocena projektu

**Powiązane charakterystyki kierunkowe:** U04

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o

**Charakterystyka U03:**

potrafi przygotowywać dane służące do tworzenia modeli UM

Weryfikacja:

ocena projektu

**Powiązane charakterystyki kierunkowe:** U04

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o

**Charakterystyka U04:**

umie dostroić parametry i wytrenować model UM na przygotowanym zbiorze danych

Weryfikacja:

ocena projektu

**Powiązane charakterystyki kierunkowe:** U02, U03

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o

**Charakterystyka U05:**

potrafi zaprojektować eksperyment A/B i rozumie jego specyfikę

Weryfikacja:

ocena projektu

**Powiązane charakterystyki kierunkowe:** U03

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o