**Nazwa przedmiotu:**

Zaawansowane uczenie maszynowe

**Koordynator przedmiotu:**

Rajmund Kożuszek

**Status przedmiotu:**

Fakultatywny dowolnego wyboru

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Przedmioty techniczne - zaawansowane

**Kod przedmiotu:**

ZUM

**Semestr nominalny:**

2 / rok ak. 2021/2022

**Liczba punktów ECTS:**

4

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

1. liczba godzin kontaktowych – 35 godz.,w tym:
a. obecność na wykładach: 30 godz.,
b. udział w konsultacjach związanych z treścią wykładu: 1 godz.,
c. udział w konsultacjach związanych z realizacją projektu: 4 godz.,
2. praca własna studenta – 75 godz., w tym:
a. przygotowanie do wykładów (przejrzenie materiałów z wykładu i dodatkowej literatury, próba rozwiązania ćwiczeń domowych sformułowanych na wykładzie): 15 godz.,
b. realizacja projektu: 10 godz. (zapoznanie się z literaturą i oprogramowaniem) + 40 godz. (wykonanie zadań projektowych) + 10 godz. (sporządzenie dokumentacji) = 60 godz.,
c. przygotowanie do kolokwiów: 10 godz.
Łączny nakład pracy studenta wynosi: 120 godz., co odpowiada 4 pkt. ECTS.

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

1,25 pkt. ECTS, co odpowiada 35 godz. kontaktowym

**Język prowadzenia zajęć:**

polski

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

2 pkt. ECTS, co odpowiada 60 godz. realizacji projektu

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

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

**Wymagania wstępne:**

 Biegła umiejętność programowania w przynajmniej jednym z języków: C, C++, Java, Python, R.
 Znajomość podstaw statystyki opisowej i testowania istotności statystycznej.
 Znajomość podstaw teoretycznych uczenia maszynowego: model PAC, wymiar VC.
 Znajomość podstawowych algorytmów uczenia maszynowego: drzewa decyzyjne, naiwny klasyfikator bayesowski, model liniowy.

**Limit liczby studentów:**

30

**Cel przedmiotu:**

Celem przedmiotu jest zapoznanie studentów z zaawansowanymi zagadnieniami dotyczącymi uczenia maszynowego i jego zastosowań do tworzenia modeli predykcyjnych. Jest to wiodący obszar badań naukowych w zakresie sztucznej inteligencji, a opracowane w wyniku tych badań algorytmy znajdują liczne zastosowania praktyczne. Przegląd rozszerzonych wariantów wybranych podstawowych algorytmów uczenia się oraz wybranych zaawansowanych algorytmów uczenia się i technik stosowanych do poprawienia jakości tworzonych za ich pomocą modeli predykcyjnych przygotowuje studentów do ich stosowania i implementowania, a także do podjęcia prac badawczych w dziedzinie uczenia maszynowego.

**Treści kształcenia:**

Wykład:
1. Wprowadzenie i rekapitulacja podstaw (4 godz.)
Definicja uczenia się; zadania indukcyjnego uczenia się; nadmierne dopasowanie; obciążenie indukcyjne; wybrane wyniki obliczeniowej teorii uczenia się; przegląd podstawowych algorytmów uczenia się, metod oceny jakości modeli oraz narzędzi ze statystyki i teorii informacji.
2. Rozszerzenia drzew decyzyjnych (2 godz.)
Metody przycinania drzew decyzyjnych; metody obsługi brakujących wartości atrybutów przy tworzeniu i stosowaniu drzew decyzyjnych; drzewa regresji, drzewa modeli.
3. Rozszerzenia naiwnego klasyfikatora bayesowskiego (2 godz.)
Sieć bayesowska; reprezentacja łącznego rozkładu prawdopodobieństwa; nie całkiem naiwny klasyfikator bayesowski; klasyfikator bayesowski jako model liniowy.
4. Regresja liniowa i logistyczna (3 godz.)
Model liniowy; estymacja parametrów modeli liniowego; diagnostyka modelu liniowego; techniki przełamywania ograniczenia liniowości; model logistyczny; logarytmiczna funkcja straty; estymacja parametrów modelu logistycznego; zastosowanie do kalibracji predykcji probabilistycznych innych modeli; regularyzacja.
5. Maszyny wektorów nośnych i metody jądrowe (3 godz.)
Odległość od granicy decyzyjnej; maksymalizacja marginesu klasyfikacji liniowo-progowej; postać prymalna i dualna zadania maksymalizacji marginesu; wektory nośne dla zadania regresji; sztuczka jądrowa; typy funkcji jądrowych.
6. Modele zespołowe (2 godz.)
Zasada modelowania zespołowego; metody tworzenia i łączenia modeli bazowych; algorytmy typu bagging, las losowy, boosting i ich specjalizowane warianty.
7. Modele wrażliwe na koszty pomyłek (2 godz.)
Macierz kosztów pomyłek; uwzględnianie kosztów pomyłek przy predykcji; uwzględnianie kosztów pomyłek przy tworzeniu modeli.
8. Klasyfikacja nienadzorowana i jednoklasowa (3 godz.)
Zadanie grupowania; miary niepodobieństwa; grupowanie typu k-środków; grupowanie gęstościowe; miary jakości grupowania; jednoklasowe maszyny wektorów nośnych; las izolacyjny; wskaźniki nietypowości na podstawie niepodobieństwa do sąsiadów i grup; detekcja anomalii i nowości.
9. Transformacja reprezentacji (2 godz.)
Selekcja atrybutów przez filtrowanie; selekcja atrybutów przez opakowanie; proste transformacje atrybutów: standaryzacja, normalizacja, kodowanie binarne; transformacje redukujące wymiarowość.
10. Pogłębiona ocena jakości modeli (2 godz.)
Informacyjne kryteria jakości modeli; uzupełniające miary jakości klasyfikacji i regresji; obciążenie i wariancja oceny; ocena pośrednia i końcowa; nadmierne dopasowanie oceny; selekcja modeli.
11. Aktualne zagadnienia badawcze (2 godz.)
Przegląd aktualnych kierunków prac badawczych.

Projekt:
Projekt realizowany w zespołach 2-osobowych umożliwi zdobycie praktycznych doświadczeń w zakresie stosowania algorytmów maszynowego uczenia się oraz prowadzenia związanych z nimi prac badawczych. Będzie on swoim zakresem obejmować:
 eksperymenty z użyciem dostępnych bibliotek dostarczających implementacji algorytmów uczenia maszynowego,
 samodzielną implementację lub modyfikację dostępnej implementacji algorytmu uczenia maszynowego i badanie jego właściwości,
 zadania wymagające doboru odpowiednich metod uczenia maszynowego, ich zastosowania oraz analizy jakości uzyskanych wyników.

**Metody oceny:**

Realizacja przedmiotu obejmuje następujące formy zajęć:
 wykład prowadzony w wymiarze 2 godz. tygodniowo,
 projekt realizowany samodzielnie w zespołach,
 konsultacje.
Aktywizacji studentów służą:
 interaktywna formuła wykładu,
 dostarczane po każdym wykładzie notatki podsumowujące jego treść z zachętą do zadawania pytań oraz notatki Q&A zawierające odpowiedzi na pytania studentów,
 dostarczane po każdym wykładzie ćwiczenia sprawdzające przyswojenie omawianych zagadnień,
 wymóg konsultacji interpretacji tematu i zakresu projektu,
 wymóg przedstawienia do oceny wstępnej dokumentacji projektu,
 wymóg konsultacji zmian interpretacji tematu i zakresu projektu wprowadzanych po ocenie dokumentacji wstępnej.
Sprawdzanie założonych efektów kształcenia realizowane jest przez:
 ocenę wiedzy i umiejętności wykazanych na sprawdzianie pisemnym,
 ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych – ocena wykonanych prac implementacyjnych, eksperymentalnych i jakości dokumentacji,
 formatywną ocenę związaną z rozwiązywaniem ćwiczeń domowych formułowanych na wykładzie, udziałem w konsultacjach i interaktywną formą prowadzenia wykładu.

**Egzamin:**

nie

**Literatura:**

1. I. H. Witten, E. Frank, M. A. Hall, C. J. Pal: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2016.
2. Bishop, C: Pattern Recognition and Machine Learning. Springer, 2011.
3. T. Mitchell: Machine Learning. McGraw Hill, 1997.
4. Y. S. Abu-Mostafa, M. Magdon-Ismail, H.-T. Lin: Learning from Data: A Short Course. AMLBook, 2012.
5. P. Cichosz: Data Mining: Explained Using R. Wiley, 2015.
6. Biblioteka Weka.
7. Pakiety języka R.
8. Pakiety języka Python.

**Witryna www przedmiotu:**

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

**Uwagi:**

(-)

## Charakterystyki przedmiotowe