**Nazwa przedmiotu:**

Wstęp do uczenia maszynowego

**Koordynator przedmiotu:**

Prof. dr hab. inż. Jacek Mańdziuk

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia I stopnia

**Program:**

Inżynieria i Analiza Danych

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

1120-DS000-ISP-0242

**Semestr nominalny:**

4 / rok ak. 2019/2020

**Liczba punktów ECTS:**

5

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

1. godziny kontaktowe – 67 h; w tym
a) obecność na wykładach – 30 h
b) obecność na laboratoriach – 30 h
c) konsultacje – 5 h
d) obecność na egzaminie – 2 h
2. praca własna studenta – 60 h; w tym
a) zapoznanie się z literaturą – 5 h
b) przygotowanie projektu – 30 h
c) przygotowanie raportu/prezentacji – 10 h
d) przygotowanie do egzaminu – 10 h
Razem 127 h, co odpowiada 5 pkt. ECTS

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

1. obecność na wykładach – 30 h
2. obecność na laboratoriach – 30 h
3. konsultacje – 5h
4. obecność na egzaminie – 2 h
Razem 67 h, co odpowiada 3 pkt. ECTS

**Język prowadzenia zajęć:**

polski

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

1. obecność na laboratoriach – 30 h
2. przygotowanie projektu – 30 h
3. przygotowanie raportu/prezentacji – 10 h
Razem 70 h, co odpowiada 3 pkt. ECTS

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

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

**Wymagania wstępne:**

brak

**Limit liczby studentów:**

.

**Cel przedmiotu:**

Celem przedmiotu jest zapoznanie słuchaczy z podstawowymi metodami uczenia maszynowego nadzorowanego i nienadzorowanego oraz nauczenie ich praktycznej umiejętności ich stosowania i oceny ich przydatności

**Treści kształcenia:**

Wykład (program przedmiotu):
1. Wstępne przetwarzanie danych: redukcja wymiaru danych, ekstrakcja cech: analiza składowych głównych, SVD, skalowanie wielowymiarowe, korelacja cech
2. Generalizacja. Dobre dopasowanie a przeuczenie, walidacja krzyżowa. Miary podobieństwa obiektów, segmentacja danych w oparciu o podobieństwo.
3. Uczenie nienadzorowane: analiza skupień. Miary odległości, algorytmy optymalizacji funkcji kryterialnej (k-means, k-medoids), metody hierarchiczne.
4. Problem klasyfikacji binarnej/wieloklasowej: podejście bayesowskie, empiryczne klasyfikatory bayesowskie (naiwna metod bayesowska, klasyfikator knn)
5. Liniowe metody klasyfikacyjne: LDA, klasyfikator logistyczny, SVM (wersja liniowa)
6. Drzewa jako klasyfikatory. Ekstrakcja reguł z drzew.
7. Reguły asocjacyjne, systemy regułowe.
8. Ewaluacja skuteczności klasyfikatora. Macierz błędów. Miary błędów oraz ich interpretacja.
9. Kompleksowa analiza wszystkich etapów budowy oraz zastosowania narzędzi analizy danych na wybranych przykładach (case studies) z wykorzystaniem metodyki CRISP-DM.
Laboratorium:
Projekt: praktyczna, wieloaspektowa analiza rzeczywistych zbiorów danych obejmująca zagadnienia omawiane na wykładzie.

**Metody oceny:**

Zaliczenie przedmiotu oparte jest o wyniki dwóch zadań realizowanych w ramach laboratorium (w sumie maksymalnie 60 pkt) oraz wyniki egzaminu (maksymalnie 40 pkt). W przypadku zadań realizowanych w laboratorium oceniana jest zarówno jakość rozwiązań jak i terminowość ich realizacji. Szczegółowe zasady przyznawania punktów przedstawiane są studentom podczas pierwszych zajęć laboratoryjnych (dot. laboratorium) oraz wykładowych (dot. egzaminu).
Ocena końcowa zależy od łącznej liczby punktów uzyskanych z zadań punktowanych oraz egzaminu i jest wyznaczana zgodnie z poniższymi regułami: 0-50 pkt – 2.0, 51-60 pkt – 3.0, 61-70 pkt – 3.5, 71-80 pkt – 4.0, 81-90 pkt – 4.5, 91-100 pkt – 5.0. Warunkiem koniecznym uzyskania oceny pozytywnej jest zdobycie co najmniej 30 pkt z laboratorium oraz 20 pkt z egzaminu.

**Egzamin:**

tak

**Literatura:**

1. Cichosz, Systemy Uczące się, WNT 2007
2. Bishop, Pattern Recognition and Machine Learning, Springer 2006
3. Flasiński , Wstęp do Sztucznej Inteligencji, PWN 2011

**Witryna www przedmiotu:**

.

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

Zna podstawowe metody wstępnej obróbki danych, w tym metod redukcji wymiaru danych i ekstrakcji cech.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_W04, DS\_W05

**Powiązane charakterystyki obszarowe:** I.P6S\_WG

**Charakterystyka W02:**

Posiada wiedzę teoretyczną i praktyczną w zakresie podstawowych metod liniowej klasyfikacji oraz kombinatorycznych i hierarchicznych metod analizy skupień.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_W04

**Powiązane charakterystyki obszarowe:** I.P6S\_WG

**Charakterystyka W03:**

Posiada wiedzę dotyczącą podstawowych metod indukcji drzew decyzyjnych oraz ekstrakcji reguł z drzew.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_W04

**Powiązane charakterystyki obszarowe:** I.P6S\_WG

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

Potrafi wykonać i zinterpretować analizę składowych głównych oraz metodę skalowania wielowymiarowego.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_U06

**Powiązane charakterystyki obszarowe:** I.P6S\_UW

**Charakterystyka U02:**

Potrafi skonstruować klasyfikator liniowy i ocenić jego jakość.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_U13, DS\_U16

**Powiązane charakterystyki obszarowe:** I.P6S\_UW

**Charakterystyka U03:**

Potrafi zbudować klasyfikator w postaci drzewa decyzyjnego oraz ocenić jego praktyczną skuteczność.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_U13

**Powiązane charakterystyki obszarowe:** I.P6S\_UW

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K01:**

Umie współpracować w grupie projektowej przyjmując w niej różne role

Weryfikacja:

Ocena projektu

**Powiązane charakterystyki kierunkowe:** DS\_K04

**Powiązane charakterystyki obszarowe:** I.P6S\_KR