**Nazwa przedmiotu:**

Data Mining

**Koordynator przedmiotu:**

prof. dr hab. Jan Mielniczuk

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Matematyka

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

1120-MASMA-NSP-0019

**Semestr nominalny:**

2 / rok ak. 2021/2022

**Liczba punktów ECTS:**

6

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

1. godziny kontaktowe – 70 h; w tym
a) obecność na wykładach – 30 h
b) obecność na laboratoriach – 30 h
c) obecność na egzaminie – 5 h
d) konsultacje – 5 h
2. praca własna studenta – 60 h; w tym
a) przygotowanie do laboratoriów – 25 h
b) zapoznanie się z literaturą – 10 h
c) przygotowanie do egzaminu – 25 h
Razem 130 h, co odpowiada 6 pkt. ECTS

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

a) obecność na wykładach – 30 h
b) obecność na laboratoriach – 30 h
c) obecność na egzaminie – 5 h
d) konsultacje – 5 h
Razem 70 h, co odpowiada 3 pkt. ECTS

**Język prowadzenia zajęć:**

polski

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

a) obecność na laboratoriach – 30 h
b) przygotowanie do laboratoriów – 25 h
Razem 55 h, co odpowiada 2 pkt. ECTS

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

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

**Wymagania wstępne:**

Stosowana analiza regresji

**Limit liczby studentów:**

30, 2 grupy laboratoryjne, 15 osób/grupa

**Cel przedmiotu:**

Zapoznanie słuchaczy z podstawowymi technikami data miningu i uczenia maszynowego

**Treści kształcenia:**

Wykład:
1. analiza dyskryminacyjna LDA . Kwadratowa analiza dyskryminacyjna QDA
2. Klasyfikator bayesowski i jego wersje empiryczne.
3. Metoda najbliższego sąsiada, naiwna metoda bayesowska, klasyfikacja logistyczna.
4. Metody oceny klasyfikatorów.
5. Drzewa klasyfikacyjne CART.
6.Drzewa regresyjne.
7. Maszyny wektorów podpierających SVM Vapnika,
8.Wprowadzenie do metody przekształcenia jądrowego.
9. Sieci neuronowe, modele addytywne,
10. Metody łączenia klasyfikatorów i estymatorów regresji
11. Problemy wysokiego wymiaru wektora atrybutów: selekcja i ekstrakcja cech.
12. Analiza skupień: metody kombinatoryczne/hierarchiczne
13. Modelowanie mieszankowe, zastosowania
14. Analiza czynnikowa i analiza składowych niezależnych
Laboratorium:
Praktyczna realizacja tematów 1-14 omawianych na wykładzie w oparciu o system R w oparciu o rzeczywiste i symulowane zbiory danych

**Metody oceny:**

Laboratoria: 30%, egzamin 70 %

**Egzamin:**

tak

**Literatura:**

1. J Friedman, T Hastie, R. Tibshirani, Elements of statistical learning, Springer, 2001
2. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT, 2005

**Witryna www przedmiotu:**

brak

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka DAM\_W01:**

Zna ogólne sformułowanie problemu klasyfikacji pod nadzorem i bez nadzoru oraz podstawowe metody klasyfikacji liniowej LDA i klasyfikacji logistycznej. Wie, na czym polega kwadratowa analiza dyskryminacyjna QDA.

Weryfikacja:

Egzamin, zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_W07

**Powiązane charakterystyki obszarowe:**

**Charakterystyka DAM\_W02:**

Zna pojęcie i postać estymatora bayesowskiego w problemie klasyfikacji oraz konstrukcję jego odpowiedników empirycznych. Zna podstawowe metody oceny błędu klasyfikacji.

Weryfikacja:

Egzamin, zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_W07

**Powiązane charakterystyki obszarowe:**

**Charakterystyka DAM\_W03:**

Zna metodologię konstrukcji drzew klasyfikacyjnych oraz maszyn wektorów podpierających. Wie, na czym polegają metody łączenia klasyfikatorów

Weryfikacja:

Egzamin, zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_W08

**Powiązane charakterystyki obszarowe:**

**Charakterystyka DAM\_W04:**

Zna podstawowe metody analizy skupień: metodę k-średnich, dendrogramy, metodę mieszanek, sieci samoorganizujące się Kohonena.

Weryfikacja:

Egzamin, zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_W09

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka DAM\_U01:**

Umie skonstruować klasyfikatory liniowe w oparciu o metodologię LDA i model logistyczny, porównać z klasyfikatorem QDA i ocenić ich błędy klasyfikacji.

Weryfikacja:

Zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_U07

**Powiązane charakterystyki obszarowe:**

**Charakterystyka DAM\_U02:**

Umie stosować metodę CART i SVM do problemu klasyfikacji i estymacji regresji.

Weryfikacja:

Zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_U07

**Powiązane charakterystyki obszarowe:**

**Charakterystyka DAM\_U03:**

Umie stosować metody analizy składowych głównych w konkretnych zagadnieniach, wybierać liczbę kierunków w tej metodzie oraz oceniać jej skuteczność.

Weryfikacja:

Zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_U08

**Powiązane charakterystyki obszarowe:**

**Charakterystyka DAM\_U04:**

Umie przeprowadzać analizę skupień w oparciu o podstawowe stosowane metodologie.

Weryfikacja:

Zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_U09

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka DAM\_K01:**

Potrafi współdziałać i pracować w zespole przyjmując w nim różne role.

Weryfikacja:

Zaliczenie laboratoriów

**Powiązane charakterystyki kierunkowe:** M2SMAD\_K01

**Powiązane charakterystyki obszarowe:**