**Nazwa przedmiotu:**

Zaawansowane metody uczenia maszynowego

**Koordynator przedmiotu:**

Prof. dr hab. Jan Mielniczuk

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka i Systemy Informacyjne

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

1120-INPAD-MSP-XXXX

**Semestr nominalny:**

2 / rok ak. 2018/2019

**Liczba punktów ECTS:**

5

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

.

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

.

**Język prowadzenia zajęć:**

polski

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

.

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

|  |  |
| --- | --- |
| Wykład:  | 0h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 50h |
| Projekt:  | 25h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

Wstęp do uczenia maszynowego, Statystyka obliczeniowa

**Limit liczby studentów:**

.

**Cel przedmiotu:**

Celem przedmiotu jest zapoznanie słuchaczy z zaawansowanymi metodami uczenia maszynowego oraz nauczenie ich aktywnej umiejętności modelowania i analizy danych regresyjnych przy użyciu szeregu nowoczesnych metod statystycznych i data miningowych, w szczególności metod modelowania krzywych regresji, metod klasyfikacji i analizy szeregów czasowych.

**Treści kształcenia:**

Wykład:
Regresja liniowa: metoda MNK, diagnostyka dopasowania i podstawowe testy.
Regresja liniowa: podstawowe odstępstwa od modelu i metody jego adaptacji.
Metoda Lasso, regresja grzbietowa, metody odporne.
Metody selekcji predyktorów w modelu, filtry, wrappery, kryteria informacyjne, metody zachłanne.
Regresja nieliniowa: nieliniowe metody parametryczne, metody nieparametryczne: lokalnie liniowa, spline’y, metoda dopasowania wstecznego w modelu addytywnym.
Nieliniowe metody klasyfikacyjne: nieliniowe empiryczne reguły bayesowskie, estymatory gęstości.
Nieliniowe metody klasyfikacyjne: QDA, drzewa klasyfikacyjne CART, metoda jądrowa SVM.
Drzewa regresyjne CART, metoda MARS.
Komitety klasyfikatorów (bagging, boosting, lasy losowe).
Metody rangowania (rank classification): algorytm Pagerank, HTS.
Modelowanie z użyciem dużych zbiorów danych: metoda stochastycznego spadku gradientu.
Praktyczna budowa systemu uczącego się (case study), metody Multisplit i RSM.
Szeregi czasowe: charakterystyki procesów stacjonarnych, problem prognozy, metoda Yule’a-Walkera, algorytm innowacyjny.
Procesy liniowe: modelowanie procesami ARMA(p,q).
Modelowanie i prognoza dla procesów niestacjonarnych.
Laboratorium:
Analiza i modelowanie zbiorów danych wykorzystująca metody regresyjne, klasyfikacyjne, analizy skupień i prognozy szeregów czasowych.
Projekt:
Praktyczna wieloaspektowa analiza zbioru danych i konstrukcja systemu automatycznego modelowania spełniającego określone kryteria (zadana precyzja, kryterium LIFT na zbiorze testowym, zadany błąd średniokwadratowy prognozy).

**Metody oceny:**

Projekt maks. 50 p., egzamin maks. 50 p., łącznie maks. 100 p. Na zaliczenie konieczne jest uzyskanie łącznie ponad 50 p. na 100 p. możliwych. Ostateczna ocena z przedmiotu wynika z sumy zdobytych punktów: [0, 50] – 2,0; (50, 60] – 3,0; (60, 70] – 3,5; (70, 80] – 4,0; (80, 90] – 4,5; (90, 100] – 5,0.

**Egzamin:**

tak

**Literatura:**

1. T. Hastie, R. Tibshirani, J.H. Friedman, The elements of statistical learning, wydanie drugie, Springer 2009
2. Ch.M. Bishop, Pattern recognition and machine learning, Springer 2006
3. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT 2005

**Witryna www przedmiotu:**

e.mini.pw.edu.pl

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W2\_01:**

Zna metody dopasowania modelu liniowego przy użyciu metody najmniejszych kwadratów, estymatora Lasso i estymatora ridge, zna podstawowe przyczyny odstępstw od modelu i środki zaradcze; zna metody selekcji predyktorów, również dla sytuacji wysokowymiarowej

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_W01, PD\_W08, PD\_W10, PD\_W11, PD\_W13

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W2\_02:**

Zna podstawowe metody nieliniowe klasyfikacji i wie, kiedy mogą być użyteczne; wie jak oceniać jakość klasyfikatora

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_W01, PD\_W08, PD\_W10, PD\_W11, PD\_W13

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W2\_03:**

Wie jak stosować komitety klasyfikatorów i dobierać ich parametry, i jak je wykorzystać do uporządkowania predyktorów

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_W01, PD\_W08, PD\_W10, PD\_W11, PD\_W13

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W2\_04:**

Zna podstawowe charakterystyki procesów stacjonarnych, zna metody identyfikacji komponent periodycznej i trendu procesu, zna podstawowe procesy liniowe (ARMA) i nieliniowe (GARCH)

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_W01, PD\_W08, PD\_W10, PD\_W11, PD\_W13

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U2\_01:**

Umie ocenić dopasowanie modelu liniowego, zidentyfikować ewentualne odstępstwa od modelu i zaadaptować model

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_U05, PD\_U06, PD\_U13, PD\_U14, PD\_U15, PD\_U16, PD\_U23

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_02:**

Umie skonstruować klasyfikator nieliniowy i ocenić jego jakość

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_U05, PD\_U06, PD\_U13, PD\_U14, PD\_U15, PD\_U16, PD\_U23

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_03:**

Umie skonstruować nieliniowy estymator regresji i ocenić jego jakość

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_U05, PD\_U06, PD\_U13, PD\_U14, PD\_U15, PD\_U16, PD\_U23

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_04:**

Umie skonstruować prognozę na podstawie podstawowych modeli liniowych i nieliniowych szeregów czasowych

Weryfikacja:

egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_U13, PD\_U14, PD\_U15, PD\_U16, PD\_U23, PD\_U05, PD\_U06

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K2\_01:**

Rozumie społeczne aspekty podejmowanych działań data miningowych, w szczególności związane z popełnionymi błędami fałszywego sygnału i braku jego detekcji

Weryfikacja:

ocena projektu

**Powiązane charakterystyki kierunkowe:** PD\_K01, PD\_K02

**Powiązane charakterystyki obszarowe:**