**Nazwa przedmiotu:**

Zaawansowane metody uczenia maszynowego

**Koordynator przedmiotu:**

prof. dr hab. Jan Mielniczuk

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Inżynieria i Analiza Danych

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

1120-DS000-MSP-0115

**Semestr nominalny:**

2 / rok ak. 2019/2020

**Liczba punktów ECTS:**

6

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

1. godziny kontaktowe – 83 h; w tym
a) obecność na wykładach – 30 h
b) obecność na laboratoriach – 30 h
c) obecność na zajęciach projektowych – 15 h
d) konsultacje – 5 h
e) obecność na egzaminie – 3 h
2. praca własna studenta – 60 h; w tym
a) zapoznanie się z literaturą – 5 h
b) przygotowanie projektu – 25 h
c) rozwiązanie zadań domowych – 5 h
d) przygotowanie do zajęć projektowych – 10 h
e) przygotowanie raportu/prezentacji – 5 h
f) przygotowanie do egzaminu – 10 h
Razem 158 h, co odpowiada 6 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. obecność na zajęciach projektowych – 15 h
4. konsultacje – 5h
5. obecność na egzaminie – 3 h
Razem 83 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. obecność na zajęciach projektowych – 15 h
3. przygotowanie projektu – 25 h
4. rozwiązanie zadań domowych – 5 h
5. przygotowanie do zajęć laboratoryjnych – 5 h
6. przygotowanie do zajęć projektowych – 10 h
7. przygotowanie raportu/prezentacji – 5 h
Razem 95 h, co odpowiada 4 pkt. ECTS

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

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

**Wymagania wstępne:**

Wstęp do Uczenia Maszynowego

**Limit liczby studentów:**

Liczba grup: 2, Laboratoria – 15 osób / grupa

**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 (program przedmiotu):
1. Regresja liniowa: metoda MNK, diagnostyka dopasowania i podstawowe testy.
2. Regresja liniowa: podstawowe odstępstwa od modelu i metody jego adaptacji.
3. Podejście bayesowskie w regresji liniowej, bayesowskie porównanie modeli
4. Regularyzacja w regresji: regresja grzbietowa i metoda Lasso
5. Regresja wysokowymiarowa, selekcja zmiennych: filtry, wrappery, kryteria informacyjne, metody zachłanne.
6. Regresja nieliniowa: nieliniowe metody parametryczne, metody nieparametryczne, drzewa regresyjne CART, metoda MARS.
7. Nieliniowe metody klasyfikacyjne: nieliniowe empiryczne reguły bayesowskie, estymatory gęstości.
8. Nieliniowe metody klasyfikacyjne: QDA, drzewa klasyfikacyjne CART, metoda jądrowa SVM. Klasyfikacja wieloetykietowa
9. Komitety klasyfikatorów (bagging, boosting, algorytmy gradientowe, lasy losowe).
10. Modele graficzne: sieci bayesowskie.
11. Modele graficzne: losowe pola Markowa, wnioskowanie w modelach graficznych.
12. Praktyczna budowa systemu uczącego się dla danych wysokowymiarowych, metody Multisplit , Random Subspace Methods i Nearest Shrunken Centroids.
13. Szeregi czasowe: charakterystyki procesów stacjonarnych, problem prognozy, metoda Yule’a-Walkera, algorytm innowacyjny.
14. Procesy liniowe: modelowanie procesami ARMA(p,q).
15. Analiza danych funkcjonalnych FDA.
Laboratorium: Analiza i modelowanie zbiorów danych wykorzystująca metody regresyjne, klasyfikacyjne, i prognozy szeregów czasowych – w siatce ćwiczenia Projekt: Praktyczna wieloaspektowa analiza zbiorów danych i konstrukcja systemu automatycznego modelowania spełniającego określone kryteria (zadana precyzja, kryterium LIFT 5% na zbiorze testowym, zadany błąd średniokwadratowy prognozy)

**Metody oceny:**

Wykład: wykład problemowy i omawianie studiów przypadku, Laboratoria: samodzielne rozwiązywanie zadań w laboratoriach Projekt: Samodzielna konstrukcja modelu predykcyjnego

**Egzamin:**

tak

**Literatura:**

1.Bishop, Pattern Recognition and Machine Learning, Springer 2006
2. Hastie, Tibshirani, Friedman, Elements of Statistical Learning, wydanie drugie, Springer 2009
3. Izenman, Modern Multivariate Statistical Techniques, Springer 2008

**Witryna www przedmiotu:**

.

**Uwagi:**

program 3 semestralny - 1 semestr
program 4 semestralny - 2 semestr

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

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:** DS2\_W03, DS2\_W05

**Powiązane charakterystyki obszarowe:** I.P7S\_WG

**Charakterystyka W02:**

Zna podstawowe metody nieliniowe klasyfikacji i regresji i wie, kiedy mogą być użyteczne. Wie jak oceniać jakość klasyfikatora i oszacowania funkcji regresji.

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_W04, DS2\_W05

**Powiązane charakterystyki obszarowe:** I.P7S\_WG

**Charakterystyka W03:**

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

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_W04, DS2\_W05

**Powiązane charakterystyki obszarowe:** I.P7S\_WG

**Charakterystyka W04:**

Zna podstawowe charakterystyki procesów stacjonarnych, zna metody identyfikacji komponent periodycznej i trendu procesu, zna podstawowe procesy liniowe (ARMA) , zna analizę danych funkcjonalnych FDA

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_W03, DS2\_W05

**Powiązane charakterystyki obszarowe:** I.P7S\_WG

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

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

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_U03, DS2\_U14

**Powiązane charakterystyki obszarowe:** I.P7S\_UW

**Charakterystyka U02:**

Umie skonstruować klasyfikator nieliniowy i ocenić jego jakość

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_U04

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U03:**

Umie skonstruować nieliniowy estymator regresji i ocenić jego jakość

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_U03, DS2\_U14

**Powiązane charakterystyki obszarowe:** I.P7S\_UW

**Charakterystyka U04:**

Egzamin, ocena projektu

Weryfikacja:

Egzamin, ocena projektu

**Powiązane charakterystyki kierunkowe:** DS2\_U14

**Powiązane charakterystyki obszarowe:** I.P7S\_UW

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K01:**

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:** DS2\_K05

**Powiązane charakterystyki obszarowe:** I.P7S\_KO, I.P7S\_KR