

Zespół Autorski:
dr inż. Paweł Cichosz

ZAAWANSOWANE UCZENIE MASZYNOWE (jęz. polski)
Advanced Machine Learning (jęz. angielski)

Kod przedmiotu (USOS)¹:
Grupa/Grupy przedmiotów (USOS)²:

Poziom kształcenia: *drugiego stopnia*
Forma studiów i tryb prowadzenia przedmiotu: *studia stacjonarne*
Kierunek studiów: *informatyka*
Profil studiów: *ogólnoakademicki*
Specjalność: *pozostawić puste w przypadku braku specjalności*
Jednostka prowadząca: *Wydział Elektroniki i Technik Informatycznych*
Jednostka realizująca: *Instytut ...*
Koordynator przedmiotu: *...*
Poziom przedmiotu: *Zaawansowany*
Status przedmiotu: *Obieralny*
Język prowadzenia zajęć: *polski*
Semestr nominalny:
Minimalny numer semestru: *zdefiniować, jeśli nie określono semestru nominalnego*
Wymagania wstępne/zalecane przedmioty poprzedzające: PIP
WSI
UM
Dyskonta *jeśli występują, wpisać kod USOS przedmiotu/przedmiotów oraz oszacowanie liczby punktów ECTS*
Limit liczby studentów: 60

Powód zgłoszenia przedmiotu: modyfikacja programu studiów II stopnia na kierunku Informatyka

Cel przedmiotu: *(max 256 znaków)*

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.

¹ Kod przedmiotu uzupełnia Dziekanat WEiTI

² W przypadku nowego programu studiów grupy przedmiotów wprowadza Dziekanat WEiTI, w innym przypadku grupy przedmiotów, do których ma należeć zgłoszony przedmiot podaje koordynator przedmiotu

Skrócony opis przedmiotu (max 1000 znaków):

Przedmiot przedstawia wybrane zaawansowane zagadnienia z zakresu uczenia maszynowego i jego zastosowań do tworzenia modeli predykcyjnych. Zakłada się, że studenci posiadają już wiedzę dotyczącą elementów teorii maszynowego uczenia się, takich jak model PAC i wymiar VC, oraz podstawowych algorytmów, takich jak drzewa decyzyjne, naiwny klasyfikator bayesowski i klasyfikator liniowo-progowy. Na wykładzie omawiane są rozszerzone warianty tych algorytmów oraz bardziej złożone algorytmy, takie jak regresja liniowa i logistyczna, maszyny wektorów nośnych i metody jądrowe, algorytmy modelowania zespołowego oraz wybrane metody uczenia nienadzorowanego i klasyfikacji jednoklasowej. Dyskutowane są także zagadnienia związane z uwrażliwianiem modeli na koszty pomyłek, transformacji danych i oceny jakości modeli. Projekt daje możliwość poszerzenia i pogłębienia doświadczeń w implementowaniu i stosowaniu algorytmów uczenia maszynowego, a także prowadzeniu badań dotyczących ich właściwości.

Skrócony opis przedmiotu w języku angielskim (max 1000 znaków):

The course presents selected advanced topics from the area of machine learning and its applications to predictive modeling. It is assumed that the students have knowledge about elements of machine learning theory, such as the PAC model and the VC dimension, as well as basic algorithms, such as decision trees, the naive Bayes classifier, and the linear-threshold classifier. The lectures will cover extended variants of these algorithms and more refined algorithms, such as linear and logistic regression, support vector machines and kernel methods, ensemble modeling algorithms, and selected methods for unsupervised and one-class classification. Issues of achieving misclassification cost sensitivity, data transformation, and model quality evaluation are also discussed. The project assignment provides an opportunity to gain broader and deeper experience in implementing and applying machine learning algorithms as well as conducting research of their properties.

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 przełamania 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.

Inne formy (patrz Wymiar godzinowy zajęć)

Egzamin: *nie*

Literatura i oprogramowanie:

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-Ismael, 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.

Wymiar godzinowy zajęć: *(Należy podać liczbę godzin w semestrze)*

Formy prowadzonych zajęć	Wymiar godzinowy zajęć
Wykład	- 30
Ćwiczenia audytoryjne	-
Zajęcia Projektowe	- 15
Laboratoria	-
Zajęcia komputerowe	-
Seminaria	-
Lektoraty	-
Warsztaty – zajęcia zintegrowane	-
Zajęcia z wykorzystaniem technik kształcenia na odległość	-

Organizacja zajęć:

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.

Wymiar w jednostkach ECTS: 4

Liczba godzin pracy studenta związanych z osiągnięciem efektów kształcenia (opis):

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 kolokwium: **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.

Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym: 2 pkt. ECTS, co odpowiada 60 godz. realizacji projektu.

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.

Efekty uczenia się:

symbol efektu uczenia się	opis efektu uczenia się	forma zajęć/ technika kształcenia	sposób weryfikacji (oceny) ³	odniesienie do efektów uczenia się dla programu
WIEDZA				
w01	zna techniki przycinania i obsługi brakujących wartości atrybutów dla drzew decyzyjnych oraz metody tworzenia drzew regresji i drzew modeli	wykład	kolokwium	W05, W_SI_06, W_SI_07
w02	zna metodę reprezentacji zależności między atrybutami za pomocą sieci bayesowskiej oraz sposób ich wykorzystania do tworzenia niecałkiem naiwnych klasyfikatorów bayesowskich	wykład	kolokwium	W05, W_SI_06, W_SI_07
w03	zna koncepcję, algorytmy tworzenia i zasady stosowania modeli regresji liniowej i logistycznej	wykład	kolokwium	W05, W_SI_06, W_SI_07
w04	zna koncepcję, algorytmy tworzenia i zasady stosowania maszyn wektorów nośnych z funkcjami jądrowymi	wykład	kolokwium	W05, W_SI_06, W_SI_07
w05	zna koncepcję, algorytmy tworzenia i zasady stosowania modeli zespołowych	wykład	kolokwium	W05, W_SI_06, W_SI_07
w06	zna metody uwrażliwiania modeli predykcyjnych na koszty pomyłek	wykład	kolokwium	W05, W_SI_06, W_SI_07
w07	zna metody transformacji danych umożliwiające poprawę jakości modeli predykcyjnych	wykład	kolokwium	W05, W_SI_06, W_SI_07
w08	zna metody pogłębionej oceny jakości i selekcji modeli predykcyjnych	wykład	kolokwium	W05, W_SI_06, W_SI_07
w09	zna aktualnie kierunki badawcze w obszarze uczenia maszynowego	wykład	kolokwium	W_01, W_03, W05, W_SI_06, W_SI_07
UMIĘJĘTNOŚCI				
u01	umie dokonać analizy przebiegu wykonania algorytmów uczenia	projekt	dokumentacja projektu	U_01, U_07

³ Sposoby weryfikacji uzyskania efektów uczenia się: egzamin pisemny, egzamin ustny, kolokwium pisemne, kolokwium ustne, test, sprawozdanie/raport pisemny, projekt, prezentacja, praca domowa, esej, wzajemna ocena przez uczestników zajęć, ocena aktywności podczas zajęć samoocena.

	maszynowego oraz zweryfikować ich wyniki			
u02	umie dokonać analizy oceny jakości modeli predykcyjnych, ich selekcji i strojenia ich parametrów	projekt	dokumentacja projektu	U_01, U_07
u03	umie stosować metody uwrażliwiania na koszty pomyłek i transformacji danych do poprawiania jakości modeli predykcyjnych	projekt	dokumentacja projektu	U_01, U_07
u04	umie implementować algorytmy uczenia maszynowego lub modyfikować ich istniejące implementacje	projekt	dokumentacja i kod źródłowy projektu	U_10
u05	umie prowadzić prace badawcze w obszarze uczenia maszynowego oraz krytycznie analizować i prezentować ich wyniki	projekt	dokumentacja projektu	U_01, U_06, U_07
KOMPETENCJE SPOŁECZNE				
k01	stosuje właściwe metody komunikacji ustnej i pisemnej w zakresie formułowania zadań uczenia się i przedstawiania wyników algorytmów uczenia maszynowego	projekt	obserwacja aktywności w czasie konsultacji, dokumentacja projektowa	K_01
k02	efektywnie współpracuje w zespole przy pracach implementacyjnych i badawczych dotyczących uczenia maszynowego	projekt	obserwacja aktywności w czasie konsultacji, dokumentacja projektowa	K_01

Uwagi:

Data i podpis koordynatora przedmiotu:

Data i podpis kierownika zakładu:

Data i podpis kierownika kierunku / kierownika specjalności:

Data i podpis z-cy dyrektora instytutu d/s dydaktycznych / nauczania / kształcenia:

Data i podpis Dziekana: