



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Uczenie maszynowe [N1ZiIP2>UcM]

Przedmiot

Kierunek studiów

Zarządzanie i inżynieria produkcji

Rok/Semestr

3/6

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

pierwszego stopnia

Język oferowanego przedmiotu

polski

Forma studiów

niestacjonarne

Wymagalność

obieralny

Liczba godzin

Wykład

8

Laboratorium

8

Inne

0

Ćwiczenia

0

Projekty/seminaria

0

Liczba punktów ECTS

2,00

Koordynatorzy

Wykładowcy

Wymagania wstępne

Wiedza podstawowa dotycząca matematyki i informatyki

Cel przedmiotu

Przekazanie wiedzy pozwalającej na samodzielny analizę danych z użyciem algorytmów uczenia maszynowego.

Przedmiotowe efekty uczenia się

Wiedza:

Student zna pojęcia dotyczące sztucznej inteligencji i uczenia maszynowego. Student zna podstawowe algorytmy dotyczące nienadzorowanych, półnadzorowanych i nadzorowanych metod uczenia maszynowego. Student ma wiedzę z zakresu metod oceny uzyskanych modeli. Student ma wiedzę z zakresu predykcji modelowanych wielkości łącznie z oceną ich niepewności

Umiejętności:

Student potrafi zastosować gotowe narzędzia do grupowania danych i potrafi zinterpretować uzyskane wyniki. Student potrafi zastosować gotowe narzędzia do klasyfikacji danych, potrafi ocenić jakość klasyfikacji oraz zinterpretować uzyskane wyniki. Student potrafi zastosować gotowe narzędzia do budowy modeli regresyjnych, potrafi ocenić jakość predykcji za pomocą takich modeli oraz

zinterpretować uzyskane wyniki.

Kompetencje społeczne:

Student rozumie rolę informatyzacji w tym eksploracji danych i sztucznej inteligencji we współczesnej gospodarce. Potrafi w niej twórczo uczestniczyć. Student widzi potrzebę ciągłego dokształcania się co wynika z bardzo szybkiego rozwoju metod sztucznej inteligencji.

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Wykład: Zaliczenie na podstawie testu w formie pytań zamkniętych, próg zaliczenia wynosi 50% maksimum punktów.

Przyporządkowanie ocen do przedziałów procentowych wyników: <90-100> bardzo dobry; <80-90) dobry plus; <70-80) dobry; <60-70) dostateczny plus; <50-60) dostateczny; <0-50) niedostateczny.

Laboratorium: Zaliczenie na podstawie sprawozdań z zadań dotyczących analizy danych realizowanej na ćwiczeniach

Przyporządkowanie ocen do przedziałów procentowych wyników: <90-100> bardzo dobry; <80-90) dobry plus; <70-80) dobry; <60-70) dostateczny plus; <50-60) dostateczny; <0-50) niedostateczny.

Treści programowe

Podstawowe zagadnienia teoretyczne związane z uczeniem maszynowym. Praktyczne zastosowania metod uczenia maszynowego w analizie danych.

Tematyka zajęć

Wykład.

Ogólne pojęcia dotyczące sztucznej inteligencji i uczenia maszynowego w tym uczenie nadzorowane, półnadzorowane, nienadzorowane, uczenie przez wzmacnianie. Pojęcia klasyfikacji, grupowania, estymacji, predykcji, asocjacji, transformacji danych. Przygotowanie danych do dalszego przetwarzania w tym: środkowanie, normalizacja, standaryzacja, wybielanie, transformacje i redukcje wymiarowości cech (PCA, LLE, MDS, t-SNE, nieujemna faktoryzacja macierzy, analiza czynnikowa, analiza ICA, SSA), kodowanie cech kategoryalnych, dyskretyzacja, uzupełnianie brakujących danych. Ogólny algorytm budowy i testowania modelu w tym test hold-out, repeated hold-out, k - krotna walidacja, metoda leave-one-out. Regresja liniowa, nieliniowa, wieloraka i prosta. Funkcje strat, metoda najmniejszych kwadratów, funkcja Hubera i inne. Metody regularyzacji w tym regresja grzbietowa, metoda LASSO, metoda elastycznej siatki. Regresja k-NN, regresja za pomocą drzewa binarnego i lasu losowego. Regresja z wykorzystaniem sieci neuronowej. Budowa modelu dla wielu zmiennych zależnych. Ocena modelu regresji. Predykcja wartości i ocena błędów predykcji ex-ante i ex-post. Kompromis pomiędzy wariancją a obciążeniem modelu. Klasyfikacja - ogólne pojęcia. Klasyfikatory binarne i wieloklasowe, metody OvA i OvO, klasyfikacja jedno i wieloetykietowa, klasyfikacja wielowyjściowo- wieloklasowa. Funkcje strat w klasyfikacji. Metody: regresja logistyczna, k-NN, LDA, naiwny klasyfikator Bayesa, klasyfikator SVM, nieliniowy klasyfikator SVM, klasyfikacja za pomocą drzew binarnych i niebinarnych, uczenie zespołowe, bagging, pasting, boosting, lasy losowe, metoda extra-trees. Zastosowanie sieci neuronowych do klasyfikacji. Ocena jakości klasyfikatorów binarnych i niebinarnych. Miary odległości między wektorami i klastrami. Metody grupowania danych: algorytm k- means, algorytm hierarchiczny, grupowanie rozmyte. Miary jakości grupowania. Indukcja reguł, indukcja reguł rozmytych, indukcja reguł asocjacyjnych. Głębokie sieci neuronowe i ich zastosowanie w problemach uczenia maszynowego. Laboratorium.

1. Wykorzystanie języka Python i modułów Pandas, Numpy, Matplotlib do wczytywania i wstępnego przetwarzania, oczyszczania, dyskretyzacji i wizualizacji danych. Poszukiwanie korelacji. Ilustracja działania skalowania, podstawowych transformacji danych i redukcji wymiarowości przestrzeni cech.
2. Zastosowanie modułu Scikit-Learn. Wydzielenie zbioru testowego. Uczenie modelu regresji liniowej wielorakiej bez i z wykorzystaniem metod regularyzacji.
3. Dobór zmiennych niezależnych do budowy modelu za pomocą zbioru walidacyjnego i sprawdzianu krzyżowego. Ustalanie hiperparametrów modelu za pomocą sprawdzianu krzyżowego. Ocena dopasowania modelu. Ocena modelu za pomocą zbioru testowego - ocena jakości predykcji ex-post.
4. Budowa klasyfikatorów na przykładzie klasyfikatora k-NN, SVM. Ustalanie hiperparametrów za pomocą sprawdzianu krzyżowego. Dobór sposobu skalowania i transformacji danych na podstawie testów. Klasyfikacja i wizualizacja wyników. Porównanie klasyfikatorów.

5. Zastosowanie drzewa CART oraz lasu losowego do klasyfikacji. Ustalanie hiperparametrów za pomocą sprawdzianu krzyżowego. Klasyfikacja i wizualizacja wyników. Ocena przydatności poszczególnych cech.
6. Klasyfikacja za pomocą sieci perceptronów sigmoidalnych. Dobór struktury sieci. Budowa płytkiej sieci neuronowej do klasyfikacji za pomocą biblioteki Keras dla różnych typów danych.
7. Grupowanie danych z wykorzystaniem biblioteki Scikit-Learn
8. Wykorzystanie biblioteki TensorFlow do budowy sieci konwolucyjnej w celu klasyfikacji danych.

Metody dydaktyczne

Wykład: Zajęcia prowadzone w formie prezentacji multimedialnej ilustrowany przykładami analizy danych z wykorzystaniem języka Python i gotowych modułów do uczenia maszynowego.

Laboratorium: Zajęcia przy komputerach dotyczące zastosowania metod uczenia maszynowego do wybranych zagadnień w oparciu o biblioteki języka Python

Literatura

Podstawowa:

Wes McKinney, Python w analizie danych, Wydanie II, Helion, Gliwice 2018

Aurelien Geron, Uczenie maszynowe z użyciem Scikit-Learn i TensorFkow, Helion, Gliwice 2018

Stanisław Orłowski, Metody i narzędzia eksploracji danych, BTC, Legionowo 2013

Daniel T. Larose, Odkrywanie wiedzy z danych, PWN, Warszawa 2006

Daniel T. Larose, Metody i modele eksploracji danych, PWN, Warszawa 2008

Giuseppe Bonaccorso, Algorytmy uczenia maszynowego, Helion, Gliwice 2019

Uzupełniająca:

Mark Lutz, Python, wprowadzenie, Helion, Gliwice 2013

Michael Dawson, Python dla każdego, podstawy programowania, Wydanie III, Helion, Gliwice 2014

Claus O. Wilke, Podstawy wizualizacji danych, Helion, Gliwice 2020

Mirosław Krzyśko i inni, Systemy uczące się, WNT, Warszawa 2008

Marcin Szeliga, Data Science i uczenie maszynowe, PWN, Warszawa 2017

Jacek Kornacki, Jan Ćwik, Statystyczne systemy uczące się, WNT, Warszawa 2005

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	50	2,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	16	0,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	34	1,50