

Sylabus przedmiotu / modułu kształcenia		
Nazwa przedmiotu/modułu kształcenia:		Uczenie maszynowe
Nazwa w języku angielskim:		Machine learning
Język wykładowy:	polski	
Kierunek studiów, dla którego przedmiot jest oferowany:		informatyka
Jednostka realizująca:	Wydział Nauk Ścisłych i Przyrodniczych	
Rodzaj przedmiotu/modułu kształcenia (obowiązkowy/fakultatywny):		fakultatywny
Poziom modułu kształcenia (np. pierwszego lub drugiego stopnia, jednolitych magisterskich):		pierwszego stopnia
Rok studiów:	trzeci	
Semestr:	szósty	
Liczba punktów ECTS:	3	
Imię i nazwisko koordynatora przedmiotu:		dr Piotr Świtalski
Imię i nazwisko prowadzących zajęcia:		dr Piotr Świtalski, mgr Mateusz Przychodzki
Założenia i cele przedmiotu:		Założeniem tego przedmiotu jest znajomość przez słuchaczy zagadnień związanych z metodami sztucznej inteligencji oraz analizy danych. Celem przedmiotu jest zaznajomienie studentów z problematyką metod uczenia maszynowego, m.in.: • metodami klasyfikacji danych, • metodami regresji danych, • typowymi algorytmami uczenia maszynowego, • wprowadzeniem do generatywnych sieci przeciwnych.
Symbol efektu	Efekty uczenia się	Symbol efektu kierunkowego
	WIEDZA Student zna i rozumie:	
W_01	najważniejsze paradygmaty i metody uczenia maszynowego	K_W01, K_W10
W_02	algorytmów nadzorowanego i nienadzorowanego uczenia maszynowego	K_W10
W_03	jak ocenić i poprawić jakość modeli uczenia maszynowego	K_W01, K_W10
	UMIEJĘTNOŚCI Student potrafi:	
U_01	wykorzystać biblioteki implementujące algorytmy uczenia maszynowego do rozwiązania postawionego problemu	K_U02, K_U22
U_02	zaprojektować i przygotować implementację danego algorytmu uczenia maszynowego, odpowiednią dla problemu i dopasowaną do posiadanych danych	K_U07, K_U09
Forma i typy zajęć:		• studia stacjonarne: wykłady (21 godz.), ćwiczenia laboratoryjne (24 godz.) • studia niestacjonarne: wykłady (15 godz.), ćwiczenia laboratoryjne (15 godz.)
Wymagania wstępne i dodatkowe:		

Warunkiem uczestnictwa w zajęciach jest znajomość zagadnień z przedmiotów „Analiza danych” oraz „Sztuczna inteligencja” lub znajomość literatury obowiązującej w tych przedmiotach.

Treści modułu kształcenia:

Treści dotyczące wykładu:

1. **Wprowadzenie do uczenia maszynowego.** Rodzaje systemów uczenia maszynowego. Uczenie nadzorowane i uczenie nienadzorowane. Uczenie półnadzorowane. Uczenie przez wzmacnianie. Uczenie wsadowe i uczenie przyrostowe. Główne problemy uczenia maszynowego.
2. **Ocena przydatności danych.** Źródła danych. Przykład oceny przydatności danych źródłowych. Rozkład częstości zmiennych. Zmienne numeryczne i kategoriowe. Brakujące dane. Reprezentatywność danych. Duplikaty. Szeregi czasowe.
3. **Klasyfikacja.** Zbiór danych MNIST. Podział zbioru danych. Uczenie klasyfikatora binarnego. Miary wydajności. Macierz pomyłek. Precyzja klasyfikatora. Pełność. Wynik F1. Wykres krzywej ROC. Klasyfikacja wieloklasowa. Analiza błędów. Klasyfikacja wieloetykietowa. Klasyfikacja wielowyjściowa.
4. **Regresja.** Model regresji liniowej. Gradient prosty. Stochastyczny spadek wzdłuż gradientu. Regresja wielomianowa. Krzywe uczenia. Regresja logistyczna. Regresja softmax.
5. **Maszyny wektorów nośnych.** Liniowa klasyfikacja SVM. Klasyfikacja miękkiego marginesu. Nieliniowa klasyfikacja SVM. Jądro wielomianowe. Cechy podobieństwa. Gaussowskie jądro RBF. Regresja SVM. Funkcja decyzyjna i prognozy. Programowanie kwadratowe. Problem dualny. Kernelizowane maszyny SVM. Przyrostowe maszyny SVM.
6. **Drzewa decyzyjne i lasy losowe.** Uczenie i wizualizowanie drzewa decyzyjnego. Wylączenie prognoz. Szacowanie prawdopodobieństw przynależności do klas. Algorytm uczący CART. Entropia jako miara zanieczyszczenia. Hiperparametry regularyzacyjne. Drzewa decyzyjne w regresji. Niestabilność. Lasy losowe. Istotność cech w lasach losowych.
7. **Techniki uczenia nienadzorowanego.** Klasyfikacja technik uczenia nienadzorowanego. Algorytm centroidów (algorytm k-średnich). Wyszukiwanie optymalnej liczby skupień. Analiza skupień w segmentacji obrazu. Analiza skupień w przetwarzaniu wstępnym. Analiza skupień w uczeniu półnadzorowanym. Algorytm DBSCAN. Inne algorytmy analizy skupień.
8. **Uczenie przez wzmacnianie.** Uczenie się optymalizowania nagród. Wyszukiwanie strategii. Wprowadzenie do narzędzia OpenAI Gym. Sieci neuronowe jako strategie. Ocenianie czynności: problem przypisania zasługi. Gradienty strategii. Procesy decyzyjne Markowa. Uczenie metodą różnic czasowych. Q-uczenie. Przegląd popularnych algorytmów RN. AutoML i AutoGluon.
9. **Przetwarzanie sekwencji za pomocą sieci rekurencyjnych i splotowych.** Neurony i warstwy rekurencyjne. Komórki pamięci. Sekwencje wejść i wyjść. Uczenie sieci rekurencyjnych. Prognozowanie szeregów czasowych. Obsługa długich sekwencji. Duże modele językowe LLM.
10. **Uczenie reprezentacji za pomocą autokoderów.** Efektywne reprezentacje danych. Analiza PCA za pomocą niedopełnionego autokodera liniowego. Autokodery stosowe. Autokodery splotowe. Autokodery rekurencyjne. Autokodery odsumiające. Autokodery rzadkie. Autokodery wariacyjne.
11. **Generatywne sieci przeciwstawne.** Generator i dyskryminator. Problemy związane z uczeniem sieci GAN. Głębokie splotowe sieci GAN. Rozrost progresywny sieci GAN. Sieci StyleGAN.

Treści dotyczące laboratorium:

1. **Zastosowania Pythona w uczeniu maszynowym.** Pliki. Kolekcje. Typy danych. Menedżer pakietów pip. Struktury danych. Funkcje wyższego rzędu.
2. **Jupyter Notebook – NumPy, Pandas.** Instalacja Jupyter Notebook. Interfejs użytkownika. Komórki w Jupyter Notebook. Wykonywanie kodu w Jupyter Notebook. Zapisywanie i eksportowanie notatników. Biblioteka NumPy. Pandas – operacje na DataFrame.

3. **Wizualizacja danych z Matplotlib.** Wstęp do Matplotlib. Podstawowe funkcje i metody Matplotlib. Dostosowywanie wyglądu wykresów. Obiektowe podejście w Matplotlib. Rodzaje wykresów w uczeniu maszynowym.
4. **Analiza zbioru Titanic w kontekście uczenia maszynowego.** Pobieranie i eksploracja zbioru Titanic. Sprawdzanie struktury danych i wykrywanie brakujących wartości. Eksploracja danych. Kodowanie zmiennych kategorycznych. Normalizacja i standaryzacja cech numerycznych. Porównanie metod skalowania. Analiza korelacji.
5. **Klasyfikacja.** Zbiór danych MNIST. Klasyfikacja binarna. Ocena modelu klasyfikacyjnego. Miary wydajności algorytmu. Klasyfikacja wieloetykietowa.
6. **Regresja liniowa.** Wprowadzenie do Scikit-learn. Przygotowanie danych. Budowa i trenowanie modelu regresji liniowej. Ocena jakości modelu.
7. **Maszyny wektorów nośnych.** Klasyfikacja liniowa SVM. Zbiór Iris. Wyjaśnienie wizualizacji granicy decyzyjnej w LinearSVC. Klasyfikacja nieliniowa SVM (jądro RBF). Zbiór Moons. Regresja SVM. GridSearchCV – wyszukiwanie optymalnych parametrów.
8. **Drzewa decyzyjne i lasy losowe.** Drzewa decyzyjne (CART) – klasyfikacja i regresja. Lasy losowe – zespoły drzew decyzyjnych. Hiperparametry. Wizualizacja drzewa i granic decyzyjnych. Drzewa w scikit-learn. Klasyfikacja w drzewach decyzyjnych. Regresja w drzewach decyzyjnych.
9. **Uczenie nienadzorowane.** Algorytm centroidów. Granice algorytmu centroidów. Algorytm DBSCAN.
10. **Uczenie przez wzmacnianie.** Wyszukiwanie strategii. OpenAI Gym. Gradienty strategii. Q-uczenie.
11. **Automatyzacja procesu uczenia się.** AutoML i AutoGluon. Presety w AutoGluon. Klasyfikacja danych tabelarycznych z AutoGluon. AutoML w scikit-learn: Auto-sklearn.
12. **Autokodery i generatywne sieci przeciwstawne.** Efektywne reprezentacje danych. Analiza PCA za pomocą niedopełnionego autokodera liniowego. Autokodery rekurencyjne. Autokodery wariacyjne. Generatywne sieci przeciwstawne: generator i dyskryminator.

Literatura podstawowa:

1. Géron A.: Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Wydanie III, Helion, 2023
2. Szeliga M.: Data science i uczenie maszynowe, PWN, 2017
3. Morzy T.: Eksploracja danych – metody i algorytmy, PWN, 2024

Literatura dodatkowa:

1. Raschka S. i Mirjalili V.: Python. Uczenie maszynowe. Wydanie 2, Helion, 2019
2. Sztuczna inteligencja i uczenie maszynowe dla programistów. Praktyczny przewodnik po sztucznej inteligencji, Helion, 2021

Planowane formy/działania/metody dydaktyczne:

Wykład tradycyjny wspomagany jest technikami multimedialnymi. Ćwiczenia laboratoryjne – zajęcia praktyczne z wykorzystaniem wybranych narzędzi programowych. Na stronie internetowej prowadzącego zamieszczane są materiały z problemami i zadaniami laboratoryjnymi.

Sposoby weryfikacji efektów uczenia się osiągniętych przez studenta:

Symbol efektu	Metody weryfikacji efektów uczenia się
W_01	Efekt realizowany na egzaminie w formie pisemnej. Przykładowe pytanie: „Przedstaw metodę klasyfikacji wieloetykietowej”.

W_02	Efekt realizowany na egzaminie w formie pisemnej. Przykładowe pytanie: „Opisz dwie techniki doboru właściwej liczby skupień podczas stosowania algorytmu centroidów”.
W_03	Efekt realizowany na egzaminie w formie pisemnej. Przykładowe pytanie: „Którego modelu uczącego regresji liniowej należy użyć, jeżeli zbiór danych uczących zawiera miliony cech? Odpowiedź uzasadnij”.
U_01	Efekt realizowany w trakcie zajęć laboratoryjnych. Przykładowe zadanie: „Zainstaluj bibliotekę scikit-learn. Na podstawie zbioru Titanic zbuduj model regresji liniowej, który przewiduje cenę biletu (`fare`) pasażera na podstawie samodzielnie dobranego zestawu cech.”.
U_02	Efekt realizowany w trakcie zajęć laboratoryjnych. Przykładowe zadanie: „Dobierz klasyfikator rozróżniający dwa gatunki irysów: Iris-versicolor oraz Iris-virginica. W tym celu wybierz dwie cechy opisujące płatki kwiatu. Przygotuj zbiór danych zawierający tylko powyższe dwie klasy.”.

Forma i warunki zaliczenia:

Ocena z przedmiotu składa się z dwóch ocen częściowych:

- oceny z zajęć laboratoryjnych,
- oceny z egzaminu końcowego.

Na ocenę z zajęć laboratoryjnych składają się oceny częściowe uzyskane na regularnych zajęciach z nauczycielem akademickim, za które można uzyskać sumarycznie 50 pkt. Zaliczenie zajęć laboratoryjnych możliwe po uzyskaniu co najmniej 51% liczby punktów z tej formy zaliczenia

Egzamin jest egzaminem pisemnym. Do egzaminu mogą przystąpić osoby, które uzyskały zaliczenie laboratorium. Na egzaminie można uzyskać maksymalnie 50 pkt. Egzamin będzie zaliczony w przypadku uzyskania co najmniej 51% liczby punktów z tej formy zaliczenia.

Ocena końcowa, w zależności od sumy uzyskanych punktów (maksymalnie 100 pkt) jest następująca (w nawiasach ocena wg skali ECTS):

- 0 – 50 pkt: niedostateczna (F),
- 51 – 60 pkt: dostateczna (E),
- 61 – 70 pkt: dostateczna plus (D),
- 71 – 80 pkt: dobra (C),
- 81 – 90 pkt: dobra plus (B),
- 91 – 100 pkt: bardzo dobra (A).

Bilans punktów ECTS:

Studia stacjonarne

Aktywność	Obciążenie studenta
Udział w wykładach	21 godz.
Udział w ćwiczeniach laboratoryjnych	24 godz.
Przygotowanie się i udział w egzaminie	10 godz.
Przygotowanie się do ćwiczeń laboratoryjnych	18 godz.
Udział w konsultacjach z przedmiotu	2 godz.

Sumaryczne obciążenie pracą studenta	75 godz.
Punkty ECTS za przedmiot	3 ECTS
Studia niestacjonarne	
Aktywność	Obciążenie studenta
Udział w wykładach	15 godz.
Udział w ćwiczeniach laboratoryjnych	15 godz.
Przygotowanie się i udział w egzaminie	21 godz.
Przygotowanie się do ćwiczeń laboratoryjnych	22 godz.
Udział w konsultacjach z przedmiotu	2 godz.
Sumaryczne obciążenie pracą studenta	75 godz.
Punkty ECTS za przedmiot	3 ECTS