

SYLABUS
DOTYCZY CYKLU KSZTAŁCENIA 2025/2026-2028/2029
ROK AKADEMICKI 2028/2029

1. PODSTAWOWE INFORMACJE O PRZEDMIOCIE

Nazwa przedmiotu	<i>Big Data</i>
Kod przedmiotu*	
nazwa jednostki prowadzącej kierunek	<i>Instytut Informatyki, Wydział Nauk Ścisłych i Technicznych</i>
Nazwa jednostki realizującej przedmiot	<i>Instytut Informatyki, Wydział Nauk Ścisłych i Technicznych</i>
Kierunek studiów	<i>Informatyka i ekonometria</i>
Poziom studiów	<i>studia pierwszego stopnia</i>
Profil	<i>praktyczny</i>
Forma studiów	<i>studia stacjonarne</i>
Rok i semestr/y studiów	<i>rok IV, sem. VII</i>
Rodzaj przedmiotu	<i>przedmiot kierunkowy</i>
Język wykładowy	<i>polski</i>
Koordinator	<i>dr hab. prof. UR Jan Bazan</i>
Imię i nazwisko osoby prowadzącej / osób prowadzących	

1.1. Formy zajęć dydaktycznych, wymiar godzin i punktów ECTS

Semestr (nr)	Wykł.	Ćw.	Konw.	Lab.	Sem.	ZP	Prakt.	Inne (jakie?)	Liczba pkt. ECTS
7	10			20					2

1.2. Sposób realizacji zajęć

☒ zajęcia w formie tradycyjnej

☐ zajęcia realizowane z wykorzystaniem metod i technik kształcenia na odległość

1.3. Forma zaliczenia przedmiotu (z toku) (egzamin, zaliczenie z oceną, zaliczenie bez oceny)

Zaliczenie z oceną

2. WYMAGANIA WSTĘPNE

Znajomość zagadnień realizowanych na przedmiotach: Algorytmy i struktury danych, Bazy danych, Programowanie obiektowe.

3. CELE, EFEKTY UCZENIA SIĘ, TREŚCI PROGRAMOWE I STOSOWANE METODY DYDAKTYCZNE

3.1 Cele przedmiotu

C1	Celem przedmiotu jest zaznajomienie studentów z podstawowymi metodami eksploracji danych zarówno w stosunku do małych danych, jak i dużych danych.
----	--

3.2 Efekty uczenia się dla przedmiotu

EK (efekt uczenia się)	Treść efektu uczenia się zdefiniowanego dla przedmiotu	Odniesienie do efektów kierunkowych
EK_01	Ma wiedzę dotyczącą metod eksploracji danych zarówno dla małych, jak i dużych danych, w tym metod konstrukcji klasyfikatorów, metod grupowania danych, metod wykrywania reguł asocjacyjnych i sekwencyjnych.	K_Wo2, K_Ko5

EK_02	Potrafi stosować wybrane narzędzia i biblioteki informatyczne do eksploracji zarówno małych, jak i dużych danych, w tym do realizacji w nich podstawowych zadań eksploracji (klasyfikatory, metody grupowania danych, metody wykrywania reguł asocjacyjnych i sekwencyjnych).	K_U07, K_U11
EK_03	Potrafi realizować projekty informatyczne związane z budową systemów eksploracji danych.	K_U07, K_U11, K_K05

3.3. Treści programowe

A. Problematyka wykładów

Treści merytoryczne
1. Wprowadzenie do komputerowych metod eksploracji danych.
2. Przegląd najbardziej znanych systemów i bibliotek programistycznych do eksploracji danych.
3. Środowisko programowania w języku Python jako nowoczesne środowisko do eksploracji danych.
4. Przygotowanie zbiorów danych do eksploracji (obliczanie statystyk dotyczących zbiorów danych, radzenie sobie z pustymi miejscami w danych, skalowanie i standaryzacja cech, kodowanie cech symbolicznych).
5. Klasyfikatory.
6. Grupowanie danych.
7. Reguły asocjacyjne i sekwencyjne.
8. Filtracja cech.
9. Odkrywanie nietypowych przypadków.
10. Eksploracja danych z wykorzystaniem języka SQL.
11. Duże dane i problemy z ich eksploracją.
12. Zestawienie współczesnych podejść do eksploracji dużych zbiorów danych.
13. Strumieniowa eksploracja danych.
14. Eksploracja danych z wykorzystaniem równoległości obliczeń.
15. Podejście do obliczeń oparte na paradygmacie Map-Reduce z wykorzystaniem Apache Spark.
16. Biblioteka ML-lib dla Apache Spark
17. Koncepcja jeziora danych i jego realizacja za pomocą Apache Spark.
18. Kolokwium z teorii dotyczącej eksploracji danych.

B. Problematyka laboratoriów

Treści merytoryczne
1. Zadania na przygotowanie danych do eksploracji.
2. Zadania na konstruowanie klasyfikatorów.
3. Zadania na grupowanie danych.
4. Zadania na wyliczanie reguł asocjacyjnych i sekwencyjnych.
5. Zadania na filtrację cech i wykrywanie nietypowych przypadków.
6. Zadania na eksplorację danych z wykorzystaniem języka SQL.
7. Zadania na strumieniową eksplorację danych.
8. Zadania na realizację eksploracji danych z wykorzystaniem równoległości obliczeń.
9. Zadania na zastosowanie paradygmatu Map-Reduce do obliczeń.
10. Zadania na wykorzystanie biblioteki ML-lib dla Apache Spark.
11. Kolokwium praktyczne przy komputerze z metod eksploracji danych.
12. Zaliczenie projektu z eksploracji danych.

3.4. Metody dydaktyczne

Wykład: wykład z prezentacją multimedialną

Laboratorium: rozwiązywanie zadań z użyciem komputera.

4. METODY I KRYTERIA OCENY

4.1 Sposoby weryfikacji efektów uczenia się

Symbol efektu	Metody oceny efektów uczenia się (np.: kolokwium, egzamin ustny, egzamin pisemny, projekt, sprawozdanie, obserwacja w trakcie zajęć)	Forma zajęć dydaktycznych (w, ćw, ...)
EK_01	Kolokwium na papierze na końcu wykładów.	Wykład
EK_02	Kolokwium praktyczne przy komputerze na laboratorium	Lab.
EK_03	Zaliczenie projektu	Lab.

4.2 Warunki zaliczenia przedmiotu (kryteria oceniania)

Zasady uzyskania oceny końcowej:

Zaliczenie ćwiczeń i laboratorium następuje na podstawie zaliczenia wszystkich efektów weryfikowanych przez planowane w danym okresie metody weryfikacji. Przy tym zakłada się, że każda metoda weryfikacji dostarcza osobne oceny dla każdego z weryfikowanych przez nią efektów uczenia się. Jeśli dany efekt jest weryfikowany przez więcej niż jedną metodę, to ocena weryfikująca osiągnięcie tego efektu jest obliczana jako średnia arytmetyczna ocen uzyskanych w poszczególnych metodach weryfikowania tego efektu.

Student otrzymuje z zaliczenia ocenę **niedostateczny**, gdy metody weryfikacji wykażą, iż co najmniej jeden z efektów nie został osiągnięty (średnia ocena dla tego efektu jest niższa niż 3.0);

Student otrzymuje ocenę **dostateczny**, gdy przeciętnie każdy z efektów zostanie osiągnięty na poziomie co najmniej 3.0, ale chociaż jeden z efektów został osiągnięty na poziomie mniejszym

5. CAŁKOWITY NAKŁAD PRACY STUDENTA POTRZEBNY DO OSIĄGNIĘCIA ZAŁOŻONYCH EFEKTÓW W GODZINACH ORAZ PUNKTACH ECTS

Forma aktywności	Średnia liczba godzin na zrealizowanie aktywności
Godziny kontaktowe wynikające z harmonogramu studiów	30
Inne z udziałem nauczyciela akademickiego (udział w konsultacjach, egzaminie)	2
Godziny niekontaktowe – praca własna studenta (przygotowanie do zajęć, egzaminu, napisanie referatu itp.)	20
SUMA GODZIN	52
SUMARYCZNA LICZBA PUNKTÓW ECTS	2

6. PRAKTYKI ZAWODOWE W RAMACH PRZEDMIOTU

wymiar godzinowy	-
zasady i formy odbywania praktyk	-

7. LITERATURA

Literatura podstawowa:

1. Morzy T., Eksploracja danych; Metody i algorytmy, PWN, 2013.
2. Harrison, M.: Uczenie maszynowe w Pythonie, Helion, 2020.
3. Albon, C: Uczenie maszynowe w Pythonie. Receptury, Helion, 2019.
4. Jules S. Damji, J.,S., Wenig, B., Das, T., Lee, D.: Spark. Błyskawiczna analiza danych, Helion, 2023.
5. Zestaw notebooków dotyczących eksploracji danych w języku Python w środowisku Jupyter autorstwa Jana Bazana - dostępne na Teams.
6. Dokumentacja pakietu Scikit-learn: <https://scikit-learn.org>

Literatura uzupełniająca:

1. Beazley, D.: Python, zwężte kompendium dla programisty, Helion, 2022.
2. Moroney, L: Sztuczna inteligencja i uczenie maszynowe dla programistów, Helion, 2021.
3. Raschka, S: Python, Uczenie maszynowe, Helion, 2019.