

## SYLABUS

DOTYCZY CYKLU KSZTAŁCENIA 2022 - 2026

Rok akademicki 2024/2025

### 1. PODSTAWOWE INFORMACJE O PRZEDMIOCIE

Nazwa przedmiotu	<i>komputerowe narzędzia eksploracji danych</i>
Kod przedmiotu*	
Nazwa jednostki prowadzącej kierunek	<i>Kolegium Nauk Przyrodniczych</i>
Nazwa jednostki realizującej przedmiot	<i>Kolegium Nauk Przyrodniczych</i>
Kierunek studiów	<i>informatyka</i>
Poziom studiów	<i>studia inżynierskie I-go stopnia</i>
Profil	<i>ogólnoakademicki</i>
Forma studiów	<i>stacjonarne</i>
Rok i semestr/y studiów	<i>rok III, semestr 6</i>
Rodzaj przedmiotu	<i>przedmiot kształcenia specjalnościowego</i>
Język wykładowy	<i>polski</i>
Koordinator	<i>dr hab. Jan Bazan, prof. UR</i>
Imię i nazwisko osoby prowadzącej / osób prowadzących	<i>dr hab. Jan Bazan, prof. UR</i>

\* -opcjonalnie, zgodnie z ustaleniami w Jednostce

#### 1.1. Formy zajęć dydaktycznych, wymiar godzin i punktów ECTS

Semestr (nr)	Wykł.	Ćw.	Konw.	Lab.	Sem.	ZP	Prakt.	Inne (jakie?)	Liczba pkt. ECTS
6	15			30					3

#### 1.2. Sposób realizacji zajęć

zajęcia w formie tradycyjnej

#### 1.3 Forma zaliczenia przedmiotu (z toku)

egzamin

### 2. WYMAGANIA WSTĘPNE

Znajomość zagadnień realizowanych na przedmiotach: algorytmy i złożoność, programowanie obiektowe, język skryptowy, sztuczna inteligencja, metody eksploracji danych, bazy danych.

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

#### 3.1 Cele przedmiotu

C <sub>1</sub>	Zaznajomienie studentów z przykładowymi konkretnymi systemami eksploracji danych, w których można wykonywać podstawowe zadania eksploracji jak przygotowanie danych, selekcja cech, konstruowanie klasyfikatorów, grupowanie obiektów i odkrywanie wzorców (np. reguła) itd.
C <sub>2</sub>	Kształcenie umiejętności stosowania bibliotek języka Python języka SQL w problemach eksploracji danych o średnim stopniu trudności.
C <sub>3</sub>	Zapoznanie studentów z paradygmatem Map-Reduce.

#### 3.2 Efekty uczenia się dla przedmiotu

EK (efekt uczenia się)	Treść efektu uczenia się zdefiniowanego dla przedmiotu	Odniesienie do efektów kierunkowych <sup>1</sup>
EK_01	Ma wiedzę dotyczącą komercyjnych i niekomercyjnych narzędzi do eksploracji danych.	K_Wo7, K_Wo9
EK_02	Ma podstawową wiedzę dotyczącą metod eksploracji danych przeznaczonych do eksploracji dużych danych.	K_Wo7, K_Wo9
EK_03	Potrafi stosować wybrane narzędzia i biblioteki informatyczne do eksploracji małych, lecz „niełatwych” danych, w tym do realizacji w nich podstawowych zadań eksploracji.	K_U20
EK_04	Potrafi stosować wybrane narzędzia i biblioteki informatyczne do eksploracji „łatwych” dużych danych, w tym do realizacji w nich podstawowych zadań eksploracji	K_U20
EK_05	Potrafi realizować projekty informatyczne związane z budową systemów eksploracji dużych danych.	K_U20

#### 3.3 Treści programowe

##### A. Problematyka wykładu

1. Przegląd komputerowych metod eksploracji danych - przypomnienie.
2. Przegląd najbardziej znanych systemów i bibliotek programistycznych do eksploracji danych.
3. Środowisko programowania w języku Python jako nowoczesne środowisko do eksploracji danych.
4. Eksploracja danych z wykorzystaniem języka SQL.
1. Duże dane i problemy z ich eksploracją.
2. Zestawienie współczesnych podejść do eksploracji dużych zbiorów danych.
3. Strumieniowa eksploracja danych.

<sup>1</sup> W przypadku ścieżki kształcenia prowadzącej do uzyskania kwalifikacji nauczycielskich uwzględnić również efekty uczenia się ze standardów kształcenia przygotowującego do wykonywania zawodu nauczyciela.

4. Eksploracja danych z wykorzystaniem równoległości obliczeń.
5. Podejście do obliczeń oparte na paradygmacie Map-Reduce z wykorzystaniem Apache Spark.
6. Biblioteka ML-lib dla Apache Spark

#### B. Problematyka ćwiczeń, konwersatoriów, laboratoriów, zajęć praktycznych

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 eksploracja 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.

### 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, EK_02	Egzamin	wykład
EK_03	Kolokwium przy komputerze	lab
EK_04, EK_05	Projekt	lab

### 4.2 Warunki zaliczenia przedmiotu (kryteria oceniania)

Efekt	Ocena	Kryteria oceny
EK_01 EK_02	dst	Zna ogólnie funkcjonalności 2-3 wybranych narzędzi informatycznych do eksploracji danych oraz zna techniki realizacji w nich podstawowych zadań eksploracji. W szczególności, zna w jednym narzędziu i jednym języku programowania (SQL, Python) techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji i wizualizacji podstawowych informacji o ich strukturze. Jednak nie ma wystarczającej wiedzy jak poprawnie użyć wyznaczone modele eksploracji danych zarówno w typowej jak i nietypowej sytuacji. Nie zna także technik wyliczania bardziej złożonych statystyk dotyczących danych lub wyznaczonych modeli.
	db	Posiada bardziej rozległą i niezawodną wiedzę o funkcjonalnościach wybranych narzędzi informatycznych do eksploracji danych w porównaniu z wymogami na ocenę dostateczny. Zna techniki wczytywania danych i ich

		przygotowania do eksploracji w co najmniej dwóch językach, aczkolwiek umiejętności w jednym języku są wiodące. Wie jak poprawnie użyć wyznaczone modele eksploracji danych w typowej sytuacji oraz jak wyliczać bardziej złożone statystyki dotyczące danych lub wyznaczonych modeli. Jednak nie wie, jak można poprawnie użyć wyznaczone modele eksploracji danych w nietypowej sytuacji lub w nietypowy sposób, tj. np. wtedy, gdy dane testowe trzeba je zaimportować w nietrywialny sposób.	
	bdb	Posiada rozległą wiedzę o funkcjonalnościach wybranych narzędzi informatycznych do eksploracji danych. Zna na dobrym poziomie w dwóch, a na bardzo dobrym poziomie w jednym języku programowania podstawowe techniki mierzenia jakości wyznaczonych modeli i formułowania opinii na temat jakości tych metod. Wie także jak poprawnie użyć wyznaczone modele eksploracji danych w typowej sytuacji oraz jak wyliczać bardziej złożone statystyki dotyczące danych lub wyznaczonych modeli. Wreszcie wie jak poprawnie użyć wyznaczone modele eksploracji danych w nietypowej sytuacji lub w nietypowy sposób, tj. np. wtedy, gdy dane testowe nie są bezpośrednio dostępne w narzędziu i trzeba je zaimportować w nietrywialny sposób.	
EK_03	dst	Potrafi stosować wybrane narzędzia informatyczne do eksploracji danych, w tym techniki do realizacji w nich podstawowych zadań eksploracji. W szczególności, potrafi stosować techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji i wizualizacji podstawowych informacji o ich strukturze oraz potrafi stosować podstawowe techniki mierzenia jakości wyznaczonych modeli. Jednak nie zawsze potrafi poprawnie użyć wyznaczone modele eksploracji danych zarówno w typowej jak i nietypowej sytuacji. Ma także problemy z wyliczaniem bardziej złożonych statystyk dotyczące danych lub wyznaczonych modeli.	
	db	Potrafi stosować wybrane narzędzia informatyczne do eksploracji danych, w tym techniki do realizacji w nich podstawowych zadań eksploracji. W szczególności, potrafi stosować techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji i wizualizacji podstawowych informacji o ich strukturze oraz potrafi stosować podstawowe techniki mierzenia jakości wyznaczonych modeli. Ponadto, potrafi poprawnie użyć wyznaczone modele eksploracji danych w typowej sytuacji oraz wylicza bardziej złożone statystyki dotyczące danych lub wyznaczonych modeli. Jednak nie potrafi poprawnie użyć wyznaczone modele eksploracji danych w nietypowej sytuacji lub w nietypowy sposób.	
	bdb	Potrafi stosować wybrane narzędzia informatyczne do eksploracji danych, w tym techniki do realizacji w nich podstawowych zadań eksploracji. W szczególności, potrafi stosować techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji i wizualizacji podstawowych informacji o ich strukturze oraz potrafi stosować podstawowe techniki mierzenia jakości wyznaczonych modeli. Ponadto, potrafi poprawnie użyć wyznaczone modele eksploracji danych w typowej sytuacji oraz wylicza bardziej złożone statystyki dotyczące danych lub wyznaczonych modeli. Wreszcie potrafi poprawnie użyć wyznaczone modele eksploracji danych w nietypowej sytuacji lub w nietypowy sposób, tj. np. wtedy, gdy dane testowe	

		nie są bezpośrednio dostępne w narzędziu i trzeba je zaimportować w nietrywialny sposób.	
EK_04	dst	Potrafi realizować projekty informatyczne związane z eksploracją dużych danych przez zrównoleglanie obliczeń, przy czym projekty te mają własność uniwersalności na niskim poziomie (silnie dopasowane do danego zbioru danych).	
EK_05	db	Posiada umiejętności jak na ocenę dostateczny. Ponadto potrafi stosować paradygmat Map-Reduce w prostych sytuacjach.	
	bdb	Posiada umiejętności jak na ocenę dobry. Ponadto potrafi korzystać w realizowanym projekcie informatycznym związanym z eksploracją danych z biblioteki ML-lib dla Apache Spark.	

#### 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	45
Inne z udziałem nauczyciela akademickiego (udział w konsultacjach, egzaminie)	5
Godziny niekontaktowe – praca własna studenta (przygotowanie do zajęć, egzaminu, napisanie referatu itp.)	35
SUMA GODZIN	85
<b>SUMARYCZNA LICZBA PUNKTÓW ECTS</b>	<b>3</b>

*\* Należy uwzględnić, że 1 pkt ECTS odpowiada 25-30 godzin całkowitego nakładu pracy studenta.*

#### 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. Zestaw noteboków dotyczących eksploracji danych w języku Python w środowisku Jupyter autorstwa Jana Bazana - dostępne na Teams.

5. Dokumentacja pakietu Scikit-learn: <https://scikit-learn.org>

Literatura uzupełniająca:

1. Larose D.T., Odkrywanie wiedzy z danych; wprowadzenie do eksploracji danych, PWN, 2006.
2. Beazley, D.: Python, zwięzłe kompendium dla programisty, Helion, 2022.
3. Moroney, L: Sztuczna inteligencja i uczenie maszynowe dla programistów, Helion, 2021.
4. Raschka, S: Python, Uczenie maszynowe, Helion, 2019.