Uniwersytet Gdański

Karta przedmiotu

Nazwa i kod przedmiotu	Techniki eksploracji danych wielowymiarowych, PG_00193526		
Kierunek studiów	Bioinformatyka (O)		
Data rozpoczęcia studiów	październik 2026 r.	Rok akademicki realizacji przedmiotu	2027/2028
Poziom kształcenia	I stopnia - licencjackie	Grupa zajęć	Grupa zajęć obowiązkowych z zakresu kierunku studiów Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki
Forma studiów	stacjonarne	Sposób realizacji	na uczelni
Rok studiów	2	Język wykładowy	polski
Semestr studiów	4	Liczba punktów ECTS	4.0
Profil kształcenia	ogólnoakademicki	Forma zaliczenia	egzamin
Jednostka prowadząca	Rektor -> Wydział Chemii		
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot	prof. dr hab. Tomasz Puzyn	
	Prowadzący zajęcia z przedmiotu		

Formy zajęć							
	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	45.0	0.0	0.0	60
	W tym liczba godzin zajęć na odległość: 0.0						

Aktywność studenta i liczba godzin pracy					
	Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów	Udział w konsultacjach	Praca własna studenta	RAZEM
	Liczba godzin pracy studenta	60	0.0	40.0	100

Cel przedmiotu
1. Zaprezentowanie studentom zakresu możliwości zastosowania technik eksploracji danych wielowymiarowych w bioinformatyce, w szczególności w analizie danych genomicznych, trasnskryptomicznych, proteomicznych i metabolomicznych oraz chemomicznych. 2. Zdobycie przez studentów umiejętności posługiwania się najważniejszymi technik eksploracji danych wielowymiarowych (poprawny dobór metody, wykonywanie analiz i interpretacja uzyskanych wyników). 3. Zapoznanie się przez studentów z możliwościami bibliotek Python w zakresie technik eksploracji danych wielowymiarowych.

Efekty uczenia się przedmiotu

Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
[BIOINL3_U03] Stosuje metody matematyczne i statystyczne do opisu zjawisk i analizy danych; posiada umiejętność analizy danych w profesjonalnych bazach danych wykorzystywanych w bioinformatyce	Student potrafi poprawnie sformułować problem badawczy (pytanie badawcze) i dobrać do niego odpowiednią technikę eksploracji danych wielowymiarowych. Student potrafi poprawnie przeprowadzić analizę struktury wewnętrznej zbioru danych wielowymiarowych w oparciu o techniki eksploracji danych przy wykorzystaniu narzędzi dostępnych dla języka Python i samodzielnie zaprogramowane skrypty oraz poprawnie zinterpretować uzyskane wyniki. Student potrafi poprawnie zaprezentować (w formie pisemnego sprawozdania) przeprowadzone analizy oraz przedyskutować uzyskane wyniki.	[SU2] prezentacja/projekt/referat/ raport [SU3] opracowanie tekstowe/ praca pisemna [SU4] test/egzamin - ustny lub pisemny
[BIOINL3_W03] Ma wiedzę z zakresu metod matematycznych i statystycznych pozwalającą na opis i modelowanie procesów i zjawisk biologicznych	Student zna podstawy teoretyczne (algorytm działania) najważniejszych technik eksploracji danych wielowymiarowych. Student wskaże przykłady zastosowania technik eksploracji danych wielowymiarowych w bioinformatyce. Student zna biblioteki Python wykorzystywane w eksploracji danych wielowymiarowych.	[SW4] test/egzamin - ustny lub pisemny [SW1] wypowiedź ustna/rozmowa/ dyskusja [SW2] prezentacja/projekt/referat/ raport

Treści przedmiotu

Specyfika dużych zbiorów danych (ang. big data); Archiwizacja danych wielowymiarowych; Repozytoria danych; Import/Eksport danych; Metody wstępnej kontroli i preprocessingu danych; Wizualizacja danych wielowymiarowych; Podobieństwo obiektów w wielowymiarowej przestrzeni cech; Metody wyrażania podobieństwa pomiędzy obiektami oraz pomiędzy obiektami i skupieniami obiektów (odległość euklidesowa, Czebyszewa, miejska, dopełniająca, metoda pojedynczego i pełnego wiązania); Problem redukcji wymiarowości zbioru danych; Redundancja informacji; Obiekty odbiegające w przestrzeni wielkowymiarowej; Algorytmy uczenia maszynowego bez nadzoru: jedno- i dwukierunkowa hierarchiczna analiza skupień (HCA), analiza głównych składowych (PCA), Metody wyboru zmiennych i obiektów reprezentatywnych.

Na zajęciach szczególny nacisk położony zostanie na praktyczne zastosowanie technik eksploracji danych wielowymiarowych w analizie danych genomicznych, transkryptomiczych, metabolomicznych oraz chemomicznych (porównanie cech strukturalnych i właściwości związków chemicznych w dużych zbiorach). Modelowanie ścieżek transkryptomicznych i metabolicznych. Identyfikacja markerów chorobowych na poziomie molekularnym w oparciu o wielowymiarowe dane proteomiczne i metabolomiczne. Zastosowanie technik wielowymiarowej eksploracji danych w medycynie spersonalizowanej. Tworzenie skryptów w Python. Łączenie dostępnych narzędzi (R, Python, KNIME).

Wymagania wstępne i dodatkowe

wymagania wstepne:rachunek macierzowy, zagadnienie własne, znajomość środowiska Linux, podstawy programowania w języku Python;

wymagania formalne:

- Metody matematyczne dla bioinformatyki
- Informatyka - wstęp
- Metody numeryczne dla bioinformatyków
- Python z podstawami algorytmiki
- Analiza statystyczna i rachunek prawdopodobieństwa dla bioinformatyków

Sposoby i kryteria oceniania osiąganych efektów uczenia się

Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
egzamin	51.0%	100.0%

<table>
<tr><td rowspan="3">Zalecana lista lektur</td><td>Podstawowa lista lektur</td><td>. Literatura wymagana do ostatecznego zaliczenia zajęć (zdania egzaminu):

A.1. wykorzystywana podczas zajęć

• Skrypt do ćwiczeń laboratoryjnych przygotowywany przez pracowników Zespołu Chemometrii Środowiska

A.2. studiowana samodzielnie przez studenta

• J. Mazerski: Podstawy chemometrii. Gdańsk: Wydawnictwo Politechniki Gdańskiej, 2000
• M. Gągolewski: Programowanie w języku R. PWN, 2016
• M. Lutz: Python. Wprowadzenie. Helion, 2002
• S. Raschka: Python. Uczenie maszynowe. Helion, 2016</td></tr>
<tr><td>Uzupełniająca lista lektur</td><td>publikacje naukowe</td></tr>
<tr><td>Adresy eZasobów</td><td></td></tr>
<tr><td>Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania</td><td colspan="2"></td></tr>
<tr><td>Praktyki zawodowe w ramach przedmiotu</td><td colspan="2">Nie dotyczy</td></tr>
</table>

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.