


**KAPITAŁ LUDZKI**  
 NARODOWA STRATEGIA SPÓJNOŚCI

 Projekt współfinansowany przez  
 Unię Europejską w ramach  
 Europejskiego Funduszu  
 Społecznego

**UNIA EUROPEJSKA**  
 EUROPEJSKI  
 FUNDUSZ SPOŁECZNY


<b>Nazwa przedmiotu</b>		<b>Kod ECTS</b>	
Techniki eksploracji danych wielowymiarowych		11.0.0215	
<b>Nazwa jednostki prowadzącej przedmiot</b>			
Katedra Chemii i Radiochemii Środowiska			
<b>Studia</b>			
<b>wydział</b>	<b>kierunek</b>	<b>poziom</b>	<b>pierwszego stopnia</b>
Wydział Matematyki, Fizyki i Informatyki	Bioinformatyka	forma	stacjonarne
		moduł	Podstawowa
		specjalnościowy	Podstawowa
		specjalizacja	Podstawowa
<b>Nazwisko osoby prowadzącej (osób prowadzących)</b>			
prof. dr hab. Tomasz Puzyn; mgr Annemarie Danielsson; Michał Kałapus; dr Alicja Mikołajczyk; dr Waldemar Kłobus; dr inż. Karolina Jagiełło; dr Agnieszka Gajewicz-Skrętna			
<b>Formy zajęć, sposób ich realizacji i przypisana im liczba godzin</b>		<b>Liczba punktów ECTS</b>	
<b>Formy zajęć</b>		4	
Wykład, Ćw. laboratoryjne			
<b>Sposób realizacji zajęć</b>			
zajęcia w sali dydaktycznej			
<b>Liczba godzin</b>			
Ćw. laboratoryjne: 45 godz., Wykład: 15 godz.			
<b>Termin realizacji przedmiotu</b>			
2022/2023 letni			
<b>Status przedmiotu</b>		<b>Język wykładowy</b>	
obowiązkowy		polski	
<b>Metody dydaktyczne</b>		<b>Forma i sposób zaliczenia oraz podstawowe kryteria oceny lub wymagania egzaminacyjne</b>	
<ul style="list-style-type: none"> <li>- Metoda projektów (projekt badawczy, wdrożeniowy, praktyczny)</li> <li>- ćwiczenia laboratoryjne: laboratorium komputerowe, metoda projektów, praca własna (przygotowanie się do kolokwium, przygotowanie sprawozdań z realizowanego projektu)</li> <li>- wykład: wykład konwersatoryjny, praca własna (przygotowanie się do egzaminu)</li> </ul>		<b>Sposób zaliczenia</b>	
		<ul style="list-style-type: none"> <li>- Zaliczenie na ocenę</li> <li>- Egzamin</li> </ul>	
		<b>Formy zaliczenia</b>	
		<ul style="list-style-type: none"> <li>- egzamin ustny</li> <li>- egzamin pisemny z pytaniami (zadaniami) otwartymi</li> <li>- egzamin pisemny testowy</li> <li>- ustalenie oceny zaliczeniowej na podstawie ocen cząstkowych otrzymywanych w trakcie trwania semestru</li> </ul>	
		<b>Podstawowe kryteria oceny</b>	

- Wykład:
- Egzamin pisemny składający się z kilkunastu pytań testowych oraz kilku pytań otwartych (zadań) obejmujących zagadnienia wymienione w treściach programowych wykładu i ćwiczeń laboratoryjnych.
  - Warunkiem uzyskania pozytywnej oceny z zaliczenia pisemnego jest zdobycie minimum 51% punktów możliwych do uzyskania. Skala ocen jest zgodna z obowiązującym na Uniwersytecie Gdańskim regulaminem studiów.
  - Studenci, którzy uzyskali w pierwszym terminie zaliczenia pisemnego wynik 51% i więcej, a chcą podwyższyć ocenę, mogą zgłosić się na egzamin ustny. Ocena końcowa jest w tym przypadku średnią arytmetyczną z ocen uzyskanych na zaliczeniu pisemnym i ustnym.
  - Zaliczenie ustne jest obowiązkowe dla studentów, którzy uzyskali z egzaminu pisemnego wynik pomiędzy 41% a 50%. W tym przypadku student otrzymuje szansę uzupełnienia punktów brakujących do uzyskania oceny dostatecznej (omawia sposób poprawnego rozwiązania zadań z zaliczenia pisemnego). W tym przypadku nie ma możliwości poprawienia oceny z pierwszego terminu zaliczenia na wyższą.
  - Negatywna ocena z egzaminu (pisemnego i ustnego) musi być poprawiona podczas egzaminu poprawkowego odbywającego się w oparciu o te same zasady co egzamin w pierwszym terminie.
  - Ocena może być podwyższona o połowę studentom szczególnie aktywnie uczestniczącym w dyskusji naukowej podczas zajęć.
- Ćwiczenia laboratoryjne:
- Samodzielne wykonanie wszystkich zadanych ćwiczeń w pracowni komputerowej. Nieobecność można odrobić podczas zajęć z inną grupą ćwiczeniową lub w trakcie konsultacji u prowadzącego.
  - Potwierdzenie umiejętności prezentacji uzyskanych wyników oraz ich naukowej dyskusji poprzez uzyskanie pozytywnej oceny ze sprawozdań obejmujących wykonane ćwiczenia.
  - Zaliczenie wszystkich kolokwium wejściowych obejmujących podstawowe zagadnienia teoretyczne niezbędne do poprawnego wykonania ćwiczenia. Niezaliczone kolokwia należy poprawić w dodatkowym terminie wyznaczonym przez prowadzącego na zakończenie semestru (poza zajęciami).
  - Ocena końcowa z ćwiczeń jest średnią ważoną ze średnich arytmetycznych ocen otrzymanych z (i) kolokwium pisemnych (waga 40%), oraz (ii) sprawozdań obejmujących wykonane ćwiczenia (waga 60%).
  - Ocena może być podwyższona o połowę studentom szczególnie aktywnie uczestniczącym w dyskusji naukowej podczas zajęć.
  - Niezaliczenie ćwiczeń laboratoryjnych skutkuje niedopuszczeniem do zaliczenia wykładu do chwili uzyskania zaliczenia.

**Sposób weryfikacji założonych efektów uczenia się**

zakładany efekt kształcenia	konwersatorium	kolokwium	sprawozdanie	egzamin pisemny	egzamin ustny
	Wiedza				
KW_03	x		x	x	x
	Umiejętności				
KU_03		x	x	x	x
	Kompetencje				

**Określenie przedmiotów wprowadzających wraz z wymogami wstępnymi**

**A. Wymagania formalne**

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

<b>B. Wymagania wstępne</b> rachunek macierzowy, zagadnienie własne, znajomość środowiska Linux, podstawy programowania w języku Python	
<b>Cele kształcenia</b>	
<ol style="list-style-type: none"> <li>1. Zaprezentowanie studentom zakresu możliwości zastosowania technik eksploracji danych wielowymiarowych w bioinformatyce, w szczególności w analizie danych genomicznych, transkryptomicznych, proteomicznych i metabolomicznych oraz chemomicznych.</li> <li>2. Zdobycie przez studentów umiejętności posługiwania się najważniejszymi technik eksploracji danych wielowymiarowych (poprawny dobór metody, wykonywanie analiz i interpretacja uzyskanych wyników).</li> <li>3. Zapoznanie się przez studentów z możliwościami bibliotek Python w zakresie technik eksploracji danych wielowymiarowych.</li> </ol>	
<b>Treści programowe</b>	
<p>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, metoda Warda); Problem redukcji wymiarowości zbioru danych; Redundancja informacji; Obiekty odbiegające w przestrzeni wielowymiarowej; Algorytmy uczenia maszynowego bez nadzoru: jedno- i dwukierunkowa hierarchiczna analiza skupień (HCA), analiza głównych składowych (PCA), grupowanie metodą k-średnich, samoorganizujące się mapy Kohonena. Idea uczenia głębokiego (deep learning); Metody wyboru zmiennych i obiektów reprezentatywnych.</p> <p>Na zajęciach szczególny nacisk położony zostanie na praktyczne zastosowanie technik eksploracji danych wielowymiarowych w analizie danych genomicznych, transkryptomicznych, 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 personalizowanej. Tworzenie skryptów w Python. Łączenie dostępnych narzędzi (R, Python, KNIME).</p>	
<b>Wykaz literatury</b>	
<p>A. Literatura wymagana do ostatecznego zaliczenia zajęć (zdania egzaminu):</p> <p>A.1. wykorzystywana podczas zajęć</p> <ul style="list-style-type: none"> <li>• Skrypt do ćwiczeń laboratoryjnych przygotowywany przez pracowników Zespołu Chemometrii Środowiska</li> </ul> <p>A.2. studiowana samodzielnie przez studenta</p> <ul style="list-style-type: none"> <li>• J. Mazerski: Podstawy chemometrii. Gdańsk: Wydawnictwo Politechniki Gdańskiej, 2000</li> <li>• M. Gałęwski: Programowanie w języku R. PWN, 2016</li> <li>• M. Lutz: Python. Wprowadzenie. Helion, 2002</li> <li>• S. Raschka: Python. Uczenie maszynowe. Helion, 2016</li> </ul>	
<b>Kierunkowe efekty uczenia się</b>  KW_03: Ma wiedzę z zakresu metod matematycznych i statystycznych pozwalającą na opis i modelowanie procesów i zjawisk biologicznych KU_03: Stosuje podstawowe metody matematyczne i statystyczne do opisu zjawisk i analizy danych; posiada umiejętność podstawowej analizy danych w profesjonalnych bazach danych wykorzystywanych w bioinformatyce	<b>Wiedza</b> <ol style="list-style-type: none"> <li>1. Student zna podstawy teoretyczne (algorytm działania) najważniejszych technik eksploracji danych wielowymiarowych.</li> <li>2. Student wskaże przykłady zastosowania technik eksploracji danych wielowymiarowych w bioinformatyce.</li> <li>3. Student zna biblioteki Python wykorzystywane w eksploracji danych wielowymiarowych.</li> </ol>
	<b>Umiejętności</b> <ol style="list-style-type: none"> <li>1. Student potrafi poprawnie sformułować problem badawczy (pytanie badawcze) i dobrać do niego odpowiednią technikę eksploracji danych wielowymiarowych.</li> <li>2. 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.</li> <li>3. Student potrafi poprawnie zaprezentować (w formie pisemnego sprawozdania) przeprowadzone analizy oraz przedyskutować uzyskane wyniki.</li> </ol>
	<b>Kompetencje społeczne (postawy)</b>  Student określa zakres wiedzy i umiejętności w zakresie posługiwania się technikami eksploracji danych wielowymiarowych, w których chciałby się dalej rozwijać, biorąc pod uwagę plany zawodowe.
<b>Kontakt</b>	
tomasz.puzyn@ug.edu.pl	