

Karta przedmiotu

Nazwa i kod przedmiotu	BIG DATA ANALYSIS - METODY PRZETWARZANIA DANYCH, PG_00070144						
Kierunek studiów	InfoBioChem						
Data rozpoczęcia studiów	luty 2027 r.		Rok akademicki realizacji przedmiotu		2027/2028		
Poziom kształcenia	II stopnia		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	1		Język wykładowy		polski		
Semestr studiów	2		Liczba punktów ECTS		3.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		egzamin		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Chemiczny -> Katedra Biotechnologii i Mikrobiologii						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr inż. Paweł Wityk				
	Prowadzący zajęcia z przedmiotu						
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	30.0	0.0	15.0	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		2.0		13.0	75
Cel przedmiotu	Celem przedmiotu jest zapoznanie studentów z pojęciem Big Data oraz zasadami prawidłowego pozyskiwania, organizowania, opisywania, przechowywania, przygotowania i analizy dużych zbiorów danych biologicznych. Na przykładzie danych biologicznych/biochemicznych student pozna strukturę poprawnego zbioru danych, znaczenie formatów plików i metadanych, metody kontroli jakości i filtrowania oraz podstawy statystyki opisowej i wnioskowania statystycznego. Student nabędzie praktyczne umiejętności porządkowania i wstępnego opracowania danych oraz wykonywania analiz statystycznych w JASP oraz przetwarzania i interpretacji danych w oprogramowaniu OpenSource. Celem jest również rozwinięcie umiejętności jasnego prezentowania i krytycznego omawiania danych uzyskanych podczas realizacji pracy magisterskiej.						

Efekty uczenia się przedmiotu	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[K7_K05] jest gotów do ustawicznego uczenia się, rozwijania wiedzy interdyscyplinarnej oraz aktualizowania umiejętności w dynamicznie zmieniającej się rzeczywistości naukowej i zawodowej	jest gotów do samodzielnego aktualizowania wiedzy z zakresu zarządzania i analizy danych biologicznych, krytycznej oceny jakości danych i stosowanych metod oraz rzetelnego prezentowania wyników własnych badań.	[SK1] Ocena umiejętności pracy w grupie [SK2] Ocena postępów pracy [SK4] Ocena umiejętności komunikacji, w tym poprawności językowej
	[K7_W07] zna i rozumie w pogłębionym stopniu budowę i zasady działania komponentów i systemów związanych z kierunkiem studiów, w tym teorie, metody i złożone zależności między nimi oraz wybrane zagadnienia szczegółowe – właściwe dla kierunku studiów InfoBioChem	zna strukturę typowego zbioru danych biologicznych, rozumie znaczenie metadanych, identyfikatorów, próbek kontrolnych i replikacji oraz wpływ braków danych, wartości odstających, filtrowania i przekształcania danych na wyniki analizy.	[SW2] Ocena wiedzy zawartej w prezentacji [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym
	[K7_W05] ma wiedzę obejmującą teoretyczne metody obliczeniowe i informatyczne stosowane do rozwiązywania problemów – właściwych dla kierunku studiów InfoBioChem	zna i rozumie podstawowe pojęcia związane z Big Data, zasady opisywania i przechowywania danych, najczęściej stosowane formaty plików oraz podstawy statystyki opisowej i wnioskowania statystycznego wykorzystywane w analizie danych biologicznych.	[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym [SW2] Ocena wiedzy zawartej w prezentacji
	[K7_U07] potrafi stosować metody uczenia maszynowego i sztucznej inteligencji do analizy dużych zbiorów danych biologicznych	potrafi przygotować i filtrować dane, obliczać i interpretować podstawowe statystyki, dobrać i wykonać odpowiednią analizę oraz czytelnie przedstawić jej wyniki, wykorzystując JASP lub oprogramowanie pokrewne. Potrafi omówić dane uzyskane podczas realizacji pracy magisterskiej i sformułować wnioski z ich analizy.	[SU1] Ocena realizacji zadania [SU2] Ocena umiejętności analizy informacji [SU3] Ocena umiejętności wykorzystania wiedzy uzyskanej w ramach przedmiotu [SU4] Ocena umiejętności korzystania z metod i narzędzi

Treści przedmiotu	<p>Treści przedmiotu - wykład</p> <ol style="list-style-type: none"> 1. Pojęcie Big Data: cechy dużych zbiorów danych, źródła danych oraz przykłady zastosowań w biologii, biotechnologii i chemii. 2. Dane jakościowe i ilościowe; obserwacje, zmienne, typy i skale pomiarowe. 3. Organizacja poprawnej tabeli danych: jeden wiersz jako obserwacja, jedna kolumna jako zmienna, nazwy zmiennych, identyfikatory i słowniki danych. 4. Metadane i dokumentowanie zbioru: pochodzenie danych, opis próbek, jednostki, warunki pomiaru, daty i wersje. 5. Formaty danych i ich zastosowanie: CSV, TSV, XLSX, TXT, JSON oraz formaty specjalistyczne; zasady importu, eksportu i konwersji. 6. Przechowywanie, porządkowanie i udostępnianie danych: struktura katalogów, nazewnictwo plików, wersjonowanie, kopie zapasowe, bezpieczeństwo i integralność danych. 7. Zasady FAIR, odtwarzalność analizy i rzetelne zarządzanie danymi badawczymi. 8. Kontrola jakości danych: kompletność, spójność, duplikaty, błędy zapisu, brakujące wartości i wartości odstające. 9. Filtrowanie i przygotowanie danych do analizy; wpływ czyszczenia, transformacji, normalizacji i skalowania na wynik. 10. Statystyka opisowa: liczebność, średnia, mediana, kwartyle, minimum, maksimum, wariancja, odchylenie standardowe i współczynnik zmienności. 11. Opisywanie rozkładu danych i dobór wizualizacji: tabele, histogramy, wykresy pudełkowe, punktowe i słupkowe. 12. Podstawy wnioskowania statystycznego: populacja i próba, hipotezy, wartość p, przedział ufności, wielkość efektu i założenia testów. 13. Porównywanie grup oraz analiza zależności: podstawowe testy parametryczne i nieparametryczne, korelacja i regresja. 14. Specyfika danych metabolomicznych jako przykładu danych wielowymiarowych; wielokrotne porównania, PCA i podstawy interpretacji biologicznej. 15. Poprawne raportowanie, prezentowanie i interpretowanie danych; najczęstsze błędy i nadużycia. <hr/> <p>Treści przedmiotu - laboratoria</p> <ol style="list-style-type: none"> 1. Przygotowanie poprawnej tabeli danych: nazwy zmiennych, identyfikatory, jednostki, metadane i kodowanie grup. 2. Zapisywanie i konwersja plików XLSX, CSV i TXT; separator dziesiętny, kodowanie znaków i kontrola poprawności importu. 3. Porządkowanie danych: sortowanie, filtrowanie, wyszukiwanie duplikatów i błędów, oznaczanie brakujących wartości oraz podstawowe formuły. 4. Obliczanie średniej, mediany, odchylenia standardowego, współczynnika zmienności i innych statystyk opisowych w JASP.
-------------------	---

5. Tworzenie tabel podsumowujących i podstawowych wykresów; dobór sposobu prezentacji do rodzaju danych.
6. Import danych do JASP, definiowanie typów zmiennych i stosowanie filtrów.
7. Statystyki opisowe i ocena rozkładu danych w JASP; identyfikacja wartości odstających.
8. Porównywanie dwóch i wielu grup w JASP; sprawdzanie założeń i wybór testu parametrycznego lub nieparametrycznego.
9. Analiza korelacji i podstawy regresji w JASP; interpretacja tabel, wykresów, wartości p, przedziałów ufności i wielkości efektu.
10. Przygotowanie danych biologicznych do importu: układ próbek, grup i metabolitów oraz kontrola brakujących wartości.
11. Filtrowanie, imputacja, normalizacja, transformacja i skalowanie danych w MetaboAnalyst.
12. Porównania jednowymiarowe, korekta wielokrotnych porównań i wizualizacja wyników w MetaboAnalyst.
13. PCA, analiza skupień i mapy cieplne w MetaboAnalyst; interpretacja podobieństw, różnic i obserwacji odstających.
14. Analiza wzbogacenia i ścieżek metabolicznych oraz formułowanie wniosków biologicznych.
15. Opracowanie kompletnego przykładu: od tabeli danych przez czyszczenie i analizę do eksportu wyników, wykresów i krótkiego raportu.

Treści przedmiotu - seminarium

1. Przygotowanie prezentacji danych uzyskanych podczas realizacji pracy magisterskiej.
2. Przedstawienie problemu badawczego, sposobu pozyskania danych, rodzaju i wielkości zbioru oraz zastosowanych formatów i metod przechowywania.
3. Charakterystyka struktury danych, zmiennych, metadanych, grup badawczych, replikacji i ewentualnych próbek kontrolnych.
4. Omówienie sposobu kontroli jakości, filtrowania, postępowania z brakującymi wartościami i przygotowania danych do analizy.
5. Prezentacja zastosowanych statystyk opisowych, testów, tabel i wykresów oraz uzasadnienie doboru metod.
6. Interpretacja uzyskanych wyników w kontekście pytania badawczego i aktualnej wiedzy dziedzinowej.
7. Omówienie ograniczeń zbioru i analizy, potencjalnych źródeł błędów oraz możliwych dalszych etapów pracy.
8. Dyskusja nad prezentowanymi danymi, informacja zwrotna od grupy i prowadzącego oraz doskonalenie sposobu komunikowania wyników naukowych.

Wymagania wstępne i dodatkowe	Podstawowa wiedza z biologii, biochemii i matematyki. Podstawowa umiejętność pracy z komputerem, plikami i arkuszem kalkulacyjnym. Umiejętność czytania anglojęzycznej literatury naukowej. Nie jest wymagana wcześniejsza znajomość programów JASP ani programowania. Studenci powinni mieć dostęp do danych uzyskanych podczas realizacji pracy magisterskiej; jeżeli zbiór danych nie jest jeszcze kompletny, można wykorzystać dane częściowe lub pilotażowe. Podczas zajęć mogą być wykorzystywane wyłącznie dane odpowiednio zanonimizowane i przeznaczone do celów dydaktycznych.		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	Egzamin (wykład)	60.0%	40.0%
	Laboratoria (Sprawozdanie)	60.0%	40.0%
	Seminarium (Prezentacja)	60.0%	20.0%
Zalecana lista lektur	Podstawowa lista lektur	James G., Witten D., Hastie T., Tibshirani R., Taylor J., An Introduction to Statistical Learning: with Applications in Python, Springer, 2023 Tosi, Davide, Redon Kokaj, and Marco Roccetti. "15 years of Big Data: a systematic literature review." <i>Journal of Big Data</i> 11.1 (2024): 73. https://jasp-stats.org/ https://www.metaboanalyst.ca/	
	Uzupełniająca lista lektur	Luo, Senlin, and Limin Pan. <i>Big Data Analysis: Theory and Technology</i> . Springer Nature, 2026.	
	Adresy eZasobów		
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<div>1. Przygotowanie poprawnej tabeli danych do analizy: identyfikatory próbek, zmienne, jednostki, metadane i formaty plików.</div> <div>2. Wykrywanie błędów, duplikatów, brakujących wartości i wartości odstających w zbiorze danych.</div> <div>3. Obliczanie i interpretacja średniej, mediany, odchylenia standardowego, współczynnika zmienności oraz podstawowych wykresów.</div> <div>4. Filtrowanie, normalizacja, transformacja i skalowanie danych metabolomicznych.</div> <div>5. Dobór i wykonanie podstawowych analiz statystycznych w programie JASP: porównanie grup, korelacja i regresja.</div> <div>6. Interpretacja wyników analizy PCA, map cieplnych oraz wyników porównań statystycznych.</div> <div>7. Prezentacja i omówienie danych uzyskanych podczas realizacji pracy magisterskiej: struktura danych, zastosowane metody, wyniki, ograniczenia i wnioski.</div>		
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy		

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.