



Karta przedmiotu

Nazwa i kod przedmiotu	CHEMOINFORMATYKA, PG_00070339						
Kierunek studiów	InfoBioChem						
Data rozpoczęcia studiów	luty 2027 r.		Rok akademicki realizacji przedmiotu		2026/2027		
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		mieszane (blended-learning)		
Rok studiów	1		Język wykładowy		polski		
Semestr studiów	1		Liczba punktów ECTS		3.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Chemiczny -> Katedra Chemii Fizycznej						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr hab. Agnieszka Gajewicz-Skrętna				
	Prowadzący zajęcia z przedmiotu						
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	30.0	0.0	30.0	0.0	0.0	60
	W tym liczba godzin zajęć na odległość: 12.0						
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		5.0		10.0	75
Cel przedmiotu	Przedmiot obejmuje podstawowe zagadnienia chemometrii, w tym metody analizy danych eksperymentalnych wykorzystywanych w badaniach chemicznych, biologicznych i biotechnologicznych. Studenci poznają zasady projektowania eksperymentów oraz przetwarzania, analizy i interpretacji danych pomiarowych. Omawiane są metody analizy wielowymiarowej, redukcji wymiarów i walidacji modeli, a także przykłady zastosowania chemometrii w praktyce badawczej i przemysłowej.						

Efekty uczenia się przedmiotu	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[K7_W01] zna i rozumie metody, techniki i narzędzia stosowane przy rozwiązywaniu zadań bioinformatycznych w tym z zakresu modelowania molekularnego	Student posiada wiedzę z zakresu matematyki niezbędną dla właściwego opisu statystycznego zestawów danych. Student posiada wiedzę z zakresu fizyki niezbędną w procesie analizy danych dotyczących zagadnień technicznych oraz technologicznych.	[SW1] Ocena wiedzy faktograficznej [SW2] Ocena wiedzy zawartej w prezentacji [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym
	[K7_W03] zna i rozumie metody statystycznej analizy zjawisk biologicznych, procesów biotechnologicznych lub chemicznych oraz specjalistyczne narzędzia bioinformatyczne	Po ukończeniu kursu każdy student zna podstawy teoretyczne (algorytm działania) najważniejszych technik eksploracji danych oraz metod uczenia maszynowego. Student wskaże przykłady zastosowania metod chemoinformatycznych w farmacji, chemii i ochronie środowiska.	[SW1] Ocena wiedzy faktograficznej [SW2] Ocena wiedzy zawartej w prezentacji [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym
	[K7_U05] potrafi wybrać metody modelowania komputerowego i zastosować je do rozwiązywania problemów z zakresu funkcjonowania i regulacji złożonych układów	Po ukończeniu kursu każdy student potrafi samodzielnie sformułować problem badawczy a następnie w celu jego rozwiązania dobrać odpowiednie techniki eksploracji danych i/lub metody uczenia maszynowego. Student umie wykonać projekcję i wizualizację danych oraz analizę podobieństwa i grupowanie a także potrafi zinterpretować uzyskane wyniki. Student umie zbudować model regresyjny i klasyfikacyjny, poprawnie dobierając hiperparametry modelu oraz wykona predykcję zmiennej objaśnianej w oparciu o zbiór zmiennych objaśniających. Ma świadomość odpowiedzialności za podejmowane decyzje. Potrafi pracować w grupie jak i indywidualnie oraz jest świadomy konieczności dotrzymywania założonych terminów	[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU5] Ocena umiejętności zaprezentowania wyników realizacji zadania [SU3] Ocena umiejętności wykorzystania wiedzy uzyskanej w ramach przedmiotu

Treści przedmiotu	Treści przedmiotu - wykład		
	<p>Treści omawiane w ramach wykładu obejmują następujące obszary:</p> <ul style="list-style-type: none"> Wprowadzenie do chemoinformatyki: definicje, historia, znaczenie w farmakologii, chemii i ochronie środowiska. Kodowanie struktury chemicznej w chemoinformatyce: notacja SMILES, notacja daktyloskopowa (<i>fingerprints</i>), etc. Techniki eksploracji danych (<i>data mining</i>): techniki projekcji i wizualizacji wielu zmiennych; analiza podobieństwa, metody grupowania i klastrowania. Techniki uczenia maszynowego do rozwiązywania problemów klasyfikacyjnych i regresyjnych (np., regresja liniowa i jej warianty - regresja grzbietowa, technika Lasso, elastyczna sieć, regresja głównych składowych, regresja metodą częściowych najmniejszych kwadratów; algorytm K-najbliższych sąsiadów; technika drzew regresyjnych i klasyfikacyjnych; las losowy; metoda maszyny wektorów nośnych; sieci neuronowe, etc.). Wyzwania współczesnej chemoinformatyki (ilość i wiarygodność danych wejściowych; problem niezbalansowanych danych klasyfikacyjnych, brak przejrzystości algorytmów, etc.). Przykłady zastosowania chemoinformatyki w wspomaganym komputerowo projektowaniu leków (CADD), w chemii i ochronie środowiska. 		
Wymagania wstępne i dodatkowe	Treści przedmiotu - laboratoria		
	<p>Treści omawiane w ramach laboratorium obejmują następujące obszary:</p> <ul style="list-style-type: none"> Przegląd najpopularniejszych narzędzi chemoinformatycznych dostępnych w systemie otwartym, np., <i>ChemmineR</i>; <i>RDKit</i>; <i>KNIME</i>; <i>Enalos+ KNIME nodes</i>; <i>OEChem</i>; <i>CDK</i>; <i>Open Babel</i>, etc. Praktyczne wprowadzenie do technik projekcji i wizualizacji danych, analizy podobieństwa oraz metod grupowania i klastrowania (np., macierz korelacji, mapa ciepła (ang. <i>heatmap</i>); metoda k-średnich; metoda grupowania wokół centroidów (PAM); hierarchiczna analiza skupień; metoda głównych składowych, etc.) Praktyczne wprowadzenie do metod regresyjnych i klasyfikacyjnych (np., regresja liniowa i jej warianty - regresja grzbietowa, technika Lasso, elastyczna sieć, regresja głównych składowych, regresja metodą częściowych najmniejszych kwadratów; algorytm K-najbliższych sąsiadów; technika drzew regresyjnych i klasyfikacyjnych; las losowy; metoda maszyny wektorów nośnych; etc.). Projekt końcowy: opracowanie i przedstawienie wyników analiz chemoinformatycznych obejmujących wykorzystanie wybranych technik eksploracji danych oraz metod uczenia maszynowego do zdefiniowanego problemu badawczego. 		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	Wykład: test + pytania otwarte	50.0%	50.0%
	Laboratorium: projekt	50.0%	50.0%
Zalecana lista lektur	Podstawowa lista lektur	<ul style="list-style-type: none"> Bąk A, Polański J. Podstawy chemoinformatyki leków. Wydanie drugie rozszerzone. 2018, Wydawnictwo Uniwersytetu Śląskiego, ISBN:9788380128965. Morzy T, Eksploracja danych. Metody i algorytmy. 2013, Wydawnictwo Naukowe PWN. ISBN:9788301171759. Szeliga M. Data Science i uczenie maszynowe. 2017, Wydawnictwo Naukowe PWN. ISBN:9788301192327. 	
	Uzupełniająca lista lektur	<ul style="list-style-type: none"> Engel T, Gasteiger J. Chemoinformatics: Basic Concepts and Methods (1st Edition). 2018, Wiley-VCH. ISBN:9783527693788. Engel T, Gasteiger J. Applied Chemoinformatics: Achievements and Future Opportunities. 2018, Wiley-VCH. ISBN:9783527806546. 	
	Adresy eZasobów		

Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<p>Przykładowe zagadnienia teoretyczne:</p> <ul style="list-style-type: none"> • Metody redukcji wymiarowości danych. • Co oznacza optymalizacja hiperparametrów modelu? • Różnice między uczeniem nadzorowanym a uczeniem nienadzorowanym. • Różnice między regresją a klasyfikacją. • Właściwości głównych składowych. • Zasady działania algorytmu XGBoost. • Wymień i omów założenia trzech wybranych technik radzenia sobie z niezbalansowaniem danych w klasyfikacji. • Na czym polega analiza ADMET? • Zastosowanie metod eksploracji danych w projektowaniu leków. • Czym jest wirtualny screening? • Wyjaśnij zasady <i>FAIR data</i> w chemoinformatyce <p>Przykładowe zagadnienia obliczeniowe:</p> <ul style="list-style-type: none"> • Przeprowadź klasyfikację podanych związków chemicznych za pomocą modelu drzewa decyzyjnego i metody K-najbliższych sąsiadów. • Porównaj efektywność obu powyższych metod klasyfikacyjnych. • Analiza podobieństwa - narysuj dendrogram i przeprowadź jego interpretację. • Oblicz sumaryczny procent wariancji wyjaśnianej przez trzy pierwsze główne składowe. • Zakoduj cząsteczkę chemiczną w formacie SMILES. • Na podstawie podanych statystyk opisowych charakteryzujących poniższe modele regresyjne do przewidywania szybkości wchłaniania leku uszereguj je od modelu o najlepszych parametrach do modelu o najsłabszych parametrach.
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.