



**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>	
Analiza danych wielowymiarowych		11.2.0082	
<b>Nazwa jednostki prowadzącej przedmiot</b>			
Faculty of Chemistry			
<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. UG, dr hab. Tomasz Puzyn; dr Agnieszka Gajewicz; dr Karolina Jagiełło			
<b>Formy zajęć, sposób ich realizacji i przypisana im liczba godzin</b>		<b>Liczba punktów ECTS</b>	
<b>Formy zajęć</b>		2 Przedmiot w wymiarze 15h wykładu i 15h ćwiczeń w laboratorium komputerowym + praca własna	
Wykład, Ćw. laboratoryjne			
<b>Sposób realizacji zajęć</b>			
zajęcia w sali dydaktycznej			
<b>Liczba godzin</b>			
Ćw. laboratoryjne: 15 godz., Wykład: 15 godz.			
<b>Cykl dydaktyczny</b>			
2019/2020 zimowy			
<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>- praca własna - przygotowanie się do zaliczenia</li> <li>- ćwiczenia laboratoryjne w pracowni komputerowej,</li> <li>praca własna - przygotowanie się do kolokwium</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 pisemny z pytaniami (zadaniami) otwartymi</li> <li>- egzamin pisemny testowy</li> <li>- kolokwium</li> <li>- Wykład:                             <ul style="list-style-type: none"> <li>•zaliczenie pisemne z pytaniami testowymi i otwartymi (zadaniami) oraz zaliczenie ustne (uzupełnienie egzaminu pisemnego).</li> </ul> </li> <li>Ćwiczenia laboratoryjne:                             <ul style="list-style-type: none"> <li>•wykonywanie zestawu ćwiczeń w laboratorium komputerowym oraz pisemna prezentacja uzyskanych wyników po każdym ćwiczeniu (sprawozdania),</li> <li>•pisemne kolokwium wejściowe przed każdym ćwiczeniem,</li> <li>•ustalenie oceny zaliczeniowej na podstawie ocen cząstkowych.</li> </ul> </li> </ul>	
		<b>Podstawowe kryteria oceny</b>	

**Wykład:**  
Zaliczenie pisemne składające 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 zaliczenie ustne. 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 zaliczenia (pisemnego i ustnego) musi być poprawiona podczas zaliczenia poprawkowego odbywającego się w oparciu o te same zasady co zaliczenie w pierwszym terminie.

**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 kształcenia**

zakładany efekt kształcenia	Egzamin	Zaliczenie ćwiczeń	mtd. dydakt 3	mtd. dydakt 4	mtd. dydakt 5	mtd. dydakt 6	mtd. dydakt 7	mtd. dydakt 8
<b>Wiedza</b>								
K_W02	+							
K_W07	+							
<b>Umiejętności</b>								
K_U01		+						
K_U05		+						
K_U08		+						
K_U09		+						

**Określenie przedmiotów wprowadzających wraz z wymogami wstępnymi**

**A. Wymagania formalne**

matematyka  
statystyka

**B. Wymagania wstępne**

Zanajomość podstaw matematyki i statystyki

**Cele kształcenia**

Zaprezentowanie studentom zakresu możliwości zastosowania metod analizy danych wielowymiarowych.  
Zdobycie przez studentów umiejętności posługiwania się najważniejszymi metodami chemometrycznymi (wykonywanie analiz i interpretacji uzyskanych wyników),  
Zapoznanie się przez studentów z dostępnym oprogramowaniem realizującym metody chemometryczne.

## Treści programowe

### A. Problematyka wykładu:

Wprowadzenie do metod chemometrycznych: specyfika danych wielowymiarowych; różnice pomiędzy statystyką chemiczną a chemometrią; obszar zainteresowań chemometrii; podział metod chemometrycznych; przegląd oprogramowania komputerowego realizującego metody chemometryczne (m.in. środowisko KNIME, MATLAB, Statistica, Origin, SPSS).

Metody analizy struktury wewnętrznej wielowymiarowych danych chemicznych: podobieństwo obiektów w wielowymiarowej przestrzeni cech; hierarchiczna analiza skupień (HCA) jako przykład metody analizy podobieństwa; analiza głównych składowych (PCA) jako przykład metody poszukiwania projekcji. Przykłady wykorzystania tej grupy metod w różnych obszarach chemii.

Modelowanie zjawisk i procesów z wykorzystaniem metod regresyjnych i klasyfikacyjnych: regresja liniowa jednej i wielu zmiennych (LR i MLR), regresja głównych składowych (PCR) oraz regresja metodą częściowych najmniejszych kwadratów (PLS); liniowa analiza dyskryminacyjna (LDA), nieliniowy klasyfikator k-najbliższych sąsiadów (kNN); wykorzystanie sztucznych sieci neuronowych (ANN) do rozwiązywania problemów regresyjnych i klasyfikacyjnych; metody wyboru optymalnego zestawu zmiennych w modelu (wybór krokowy, wybór przy użyciu algorytmu genetycznego); walidacja modeli regresyjnych i klasyfikacyjnych. Przykłady wykorzystania tej grupy metod w różnych obszarach chemii.

### B. Problematyka ćwiczeń laboratoryjnych:

Wprowadzenie do analiz chemometrycznych prowadzonych w środowisku KNIME. Zasady pracy z komputerem. Hierarchiczna analiza skupień (HCA).

Analiza głównych składowych (PCA).

Regresja liniowa jednej i wielu zmiennych (LR/MLR).

## Wykaz literatury

### A. Literatura wymagana do ostatecznego zaliczenia zajęć (zdania egzaminu):

J. Mazerski: Podstawy chemometrii. Gdańsk: Wydawnictwo Politechniki Gdańskiej, 2000

#### A.1. Literatura wykorzystywana podczas zajęć

Skrypt do ćwiczeń laboratoryjnych przygotowywany przez pracowników Pracowni Chemometrii Środowiska

### B. Literatura uzupełniająca:

S. D. Brown, R. Tauler, B. Walczak (red.): Comprehensive chemometrics: Chemical and biochemical data analysis. Amsterdam: Elsevier, 2009

R. Kramer: Chemometric techniques for quantitative analysis. New York: Marcel Dekker, Inc, 2005

D. Zuba, A. Parczewski (red.): Chemometria w analityce: wybrane zagadnienia. Kraków: Wydawnictwo Instytutu Ekspertyz Sądowych, 2008

JM. Dobosz: Wspomagana komputerowo statystyczna analiza danych. Warszawa: Akademicka Oficyna Wydawnicza EXIT, Warszawa 2004

## Efekty kształcenia (obszarowe i kierunkowe)

K\_W02 ma wiedzę z zakresu matematyki, biologii, chemii i fizyki w zakresie niezbędnym do opisu, interpretacji i modelowania podstawowych zjawisk i procesów biologicznych

K\_W07 zna podstawy analizy numerycznej, zna na poziomie podstawowym co najmniej jeden pakiet do obliczeń symbolicznych, zna podstawowe pakiety oprogramowania użytkowego do prezentacji wyników i analizy danych

K\_U01 potrafi zastosować wiedzę matematyczną i informatyczną do formułowania, analizowania i rozwiązywania prostych zadań związanych z bioinformatyką  
K\_U05 potrafi projektować wykorzystując podstawowe techniki algorytmiczne i struktury danych, analizować, pisać uruchamiać i testować programy w wybranym środowisku programistycznym

K\_U08 potrafi stosować podstawowe pakiety oprogramowania użytkowego do prezentacji wyników i analizy danych

K\_U09 stosuje wybrane techniki i narzędzia badawcze z dziedzin nauk przyrodniczych i ścisłych

## Wiedza

Student zna:

1. podstawowy podział metod chemometrycznych, wymieni zastosowania poszczególnych grup tych metod w chemii (analityce chemicznej, naukach sądowych, kosmologii i innych);
2. wie, jakie oprogramowanie komputerowe realizuje poszczególne metody;
3. podstawy teoretyczne (algorytm działania) najważniejszych metod chemometrycznych, m.in: HCA, PCA oraz LR/MLR.

## Umiejętności

Student potrafi:

1. wykorzystać środowisko KNIME do obliczeń chemometrycznych;
2. odpowiednio przygotować dane do analiz chemometrycznych;
3. przeprowadzić analizy struktury wewnętrznej zbioru danych metodami HCA i PCA oraz poprawnie zinterpretować uzyskane wyniki;
4. zbudować model regresyjny (metodą LR/MLR), poprawnie przeprowadzi proces walidacji oraz wykonać predykcję zmiennej zależnej w oparciu o zmienną niezależną (zmienne niezależne).

## Kompetencje społeczne (postawy)

## Kontakt

puzi@qsar.eu.org