Informacje podstawowe

Nazwa przedmiotu	Analiza danych biologicznych z użyciem Pythona i AI
Prowadzący	dr Monika Fajfer-Jakubek
Wydział	Wydział Biologii i Nauk o Środowisku
Kierunek	Biologia
Poziom studiów	II stopnia
Profil studiów	Ogólnoakademicki
Forma studiów	Stacjonarna
Moduł specjalnościowy/ścieżka *(jeśli dotyczy)*	Nie dotyczy
Dyscyplina naukowa, do której odnoszą się efekty uczenia się	Nauki biologiczne
Przedmiot obowiązuje od roku akademickiego	2025/2026
Rok studiów	I/II
Semestr	I-IV
Język wykładowy	Polski
Status przedmiotu	Do wyboru
Odniesienie do kierunkowych efektów uczenia się	BI2_W09, BI2_U07, BI2_K02
Rodzaj zajęć	Ćwiczenia
Liczba godzin	Ćwiczenia: 30h
Liczba ECTS	3 ECTS
Wymagania wstępne	Podstawowa wiedza z zakresu statystyki na poziomie studiów I stopnia
Opis i cele przedmiotu	Przedmiot „Analiza danych biologicznych z użyciem Pythona i AI" ma na celu praktyczne przygotowanie studentów do pracy z danymi biologicznymi z wykorzystaniem narzędzi programistycznych oraz wybranych rozwiązań opartych na sztucznej inteligencji. W trakcie ćwiczeń studenci uczą się wczytywania, czyszczenia, porządkowania i wizualizacji danych biologicznych z użyciem języka Python oraz bibliotek takich jak pandas, matplotlib i seaborn, a także interpretowania wyników w kontekście biologicznym. Zajęcia obejmują również wykorzystanie narzędzi AI wspierających analizę danych, pracę z literaturą naukową, przygotowanie narracji naukowej oraz prezentowanie wyników analiz. Celem przedmiotu jest rozwinięcie umiejętności logicznego analizowania danych biologicznych, krytycznej oceny

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

	wyników, pracy zespołowej oraz świadomego i etycznego stosowania narzędzi AI w badaniach biologicznych.

Treści programowe- ćwiczenia

	Temat/blok zajęć: Ć	Liczba godzin
1.	C1. Wprowadzenie do AI i modeli jezykowych (LLM). Czym są i jak dzialaja duże modele językowe. Techniki efektywnego promptowania. Przegląd narzędzi AI przydatnych w biologii. Porównanie modeli komercyjnych vs. open-source. Lokalne modele jezykowe - teoria i praktyka z Ollama. Dlaczego lokalne LLM? Prywatność danych badawczych, niezależność od zewnętrznych API, praca offline, brak kosztów subskrypcji. Przeglad dostepnych modeli (Llama 3, Mistral, CodeLlama, BioMistral). Pierwsze uruchomienie modelu lokalnego. (online)	4
2.	C2 Wprowadzenie do analizy danych biologicznych z użyciem Pythona: charakterystyka danych (eksperymentalnych, środowiskowych, parazytologicznych). Praca w Google Colab . Wczytywanie danych (CSV, XLSX). Podstawy pandas: struktura DataFrame, przegląd danych (info(), describe()), selekcja i filtrowanie. Typy zmiennych biologicznych. (stacjonarnie)	4
3.	C3. Czyszczenie i przygotowanie danych biologicznych do analizy. Identyfikacja i obsługa braków danych (NaN) – strategie usuwania i uzupełniania w kontekście biologicznym. Wykrywanie błędów wprowadzania danych (wartości ujemne, literówki). Standaryzacja nazw zmiennych i kategorii. Zmiana typów danych. Tworzenie nowych zmiennych pochodnych (np. wskaźniki zarażenia). Agregacja danych i grupowanie (groupby). Porównywanie grup biologicznych (siedliska, płcie, grupy wiekowe). (online)	4
4.	C4 Wizualizacja danych biologicznych. Tworzenie podstawowych wykresów z biblioteką matplotlib (wykresy punktowe, liniowe, słupkowe). Wizualizacje statystyczne z biblioteką seaborn (boxplot, violinplot, heatmapa korelacji). Zasady poprawnej wizualizacji danych naukowych: czytelność, odpowiedni dobór typu wykresu, opisy osi, legendy. Eksport wykresów do formatów graficznych. Interpretacja wykresów w kontekście biologicznym – co mówią dane? Przegląd narzędzi do automatycznej wizualizacji wykresów. (stacjonarnie). Wprowadzenie wymagań projektu cząstkowego. (stacjonarnie)	4
5.	C5 Projekt cząstkowy. Praca w grupach . Interpretacja i porównywanie danych biologicznych. Analiza różnic między grupami na podstawie statystyk opisowych i wizualizacji. Identyfikacja trendów, wzorców i obserwacji odstających.	2

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

	Biologiczne znaczenie wartości odstających (błąd vs. zjawisko biologiczne). Formułowanie krótkich opisów wyników. Wnioskowanie biologiczne na podstawie danych i wykresów – od obserwacji do hipotezy. Krytyczna ocena własnych wyników (online)	
6.	C6 Jak ciekawie zaprezentować wyniki analiz biologicznych z wykorzystaniem AI?. Tworzenie prezentacji naukowych z użyciem narzędzia Gamma: promptowanie, struktura prezentacji naukowej, edycja i personalizacja. Import własnych wykresów z Pythona. Alternatywne narzędzia prezentacyjne wspierane przez AI (Canva, Beautiful.ai, SlidesGPT). Dobre praktyki w prezentowaniu danych biologicznych. Przygotowanie szablonu prezentacji do projektu końcowego. Prezentowanie analiz projektu cząstkowego (stacjonarnie)	2
7.	C7 Wykorzystanie narzędzi AI w pracy z literaturą naukową. Wprowadzenie do NotebookLM (Google): tworzenie notatnika, import artykułów PDF. Analiza artykułów biologicznych: identyfikacja problemu badawczego, hipotez, metodyki, rodzaju danych i zmiennych. Zadawanie pytań do artykułu w języku naturalnym. Porównywanie własnych wyników z danymi z literatury. Funkcje NotebookLM: podsumowania, FAQ, briefing doc. Praca w parach: analiza artykułu i tworzenie pierwszej wersji prezentacji w Gamma (on-line)	2
8.	C8 Tworzenie narracji naukowej i podstawy analizy statystycznej. Funkcja Audio Overview w NotebookLM – generowanie podcastu na podstawie artykułu. Przygotowanie treści popularnonaukowych: przekształcanie wyników badań w przystępny przekaz. Krytyczna ocena treści generowanych przez AI: weryfikacja faktów, identyfikacja halucynacji. Etyka wykorzystania AI w nauce. Wprowadzenie do podstawowej analizy statystycznej z wykorzystaniem JASP: testy t, ANOVA, korelacje – interpretacja wyników. Inne narzędzia AI do pracy z literaturą (np. ResearchRabbit, Elicit, SciSpace, Consensus etc.).(stacjonarnie)	4
9.	C9 Projekt zaliczeniowy – praca zespołowa (2–3 osoby). Finalizacja projektu: samodzielna analiza zbioru danych biologicznych z wykorzystaniem poznanych narzędzi (pandas, matplotlib/seaborn, NotebookLM, Gamma/JASP). Konsultacje z prowadzącym, korekty i dopracowanie formy końcowej projektu (prezentacja, podcast lub krótki film). (online)	2
10.	C10 Prezentacja wyników projektów końcowych (2 godz.): wystąpienia zespołów, dyskusja, wzajemna ocena koleżeńska (*peer review*), podsumowanie kursu (stacjonarnie)	2
	Łącznie godzin:	30

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

Opis przedmiotowych efektów uczenia się i sposoby ich weryfikacji

Symbol efektu	**<u>Kierunkowe</u> efekty uczenia się** *(zgodne z programem na BIPUKSW)* ***Absolwent…*** *(zna i rozumie/potrafi/jest gotów)*	**Opis <u>przedmiotowych</u> efektów uczenia się** *Student...* *(wyłącznie czasownikami operacyjnymi - czynności, które da się zweryfikować, mierzalne; w nawiasie należy podać numery tematów zajęć, które realizują dany efekt)* ***Student…***	**Sposoby weryfikacji efektów uczenia się** *(np.: kolokwium pisemne, egzamin ustny, egzamin pisemny, sprawozdanie, prezentacja na zajęciach, raport, projekt indywidualny, grupowy..)*
BI2_W09	Absolwent zna i rozumie fundamentalne dylematy współczesnej cywilizacji: ekonomiczne, społeczne, prawne i etyczne związane z naukami biologicznymi, zna formy pozyskiwania funduszy na badania i zasady tworzenia projektów badawczych	• opisuje prawne, etyczne i społeczne ograniczenia wykorzystania narzędzi AI oraz modeli językowych w analizie danych biologicznych (C1, C7–C8); • wyjaśnia znaczenie rzetelności, prywatności danych i weryfikacji treści generowanych przez AI w pracy badawczej (C1, C8); • omawia zasady przygotowania projektu analitycznego obejmującego dane biologiczne, narzędzia Python oraz wybrane narzędzia AI (C5, C9–C10).	wejściówki, projekt cząstkowy, projekt końcowy, prezentacja wyników.
BI2_U07	Absolwent potrafi na podstawie posiadanej wiedzy, dokonać analizy procesu lub zjawiska	• wczytuje, porządkuje i przygotowuje dane biologiczne do analizy z użyciem języka Python oraz biblioteki pandas (C2–C3); • wykonuje podstawową analizę danych biologicznych, tworzy zmienne pochodne, grupuje dane i interpretuje uzyskane wyniki w kontekście biologicznym (C3–C5); • tworzy poprawne wizualizacje danych biologicznych z użyciem bibliotek matplotlib i seaborn oraz formułuje na ich	wejściówki, projekt cząstkowy, projekt końcowy, prezentacja wyników.

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

		podstawie wnioski biologiczne (C4–C5); • wykorzystuje narzędzia AI do analizy literatury naukowej, przygotowania narracji naukowej oraz prezentacji wyników analiz biologicznych (C6–C9);	
BI2_K02,	Absolwent jest gotów zasięgania opinii ekspertów w przypadku trudności z samodzielnym rozwiązaniem problemu	• konsultuje trudności analityczne i interpretacyjne z prowadzącym oraz korzysta z informacji zwrotnej podczas realizacji projektu (C5, C9); • współpracuje w zespole projektowym, dzieląc się zadaniami związanymi z analizą danych, wizualizacją i prezentacją wyników (C5–C10); • krytycznie ocenia własne wyniki, treści generowane przez AI oraz jakość wykorzystanych źródeł naukowych (C7–C10).	projekt cząstkowy, projekt końcowy

Metody dydaktyczne

(*dostosowane do przedmiotowych efektów uczenia się*)

Ćwiczenia prowadzone w sali komputerowej lub online obejmują ćwiczenia praktyczne, pracę projektową (indywidualną i zespołową), analizę rzeczywistych zbiorów danych biologicznych, studia przypadków oraz dyskusję wyników. Wykorzystywane są prezentacje multimedialne, autorskie instrukcje i karty pracy, środowiska programistyczne (Python, Jupyter Notebook), biblioteki analityczne oraz narzędzia wizualizacji danych. Pomoce dydaktyczne obejmują: komputery, projektor multimedialny, dostęp do Internetu, bazy danych biologicznych oraz materiały udostępniane przez prowadzącego.

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

Opis nakładu pracy studenta w ECTS

Ćwiczenia

Kontakt z prowadzącym	Aktywność	Liczba godzin	Razem liczba godzin / ECTS
bezpośredni	udział w zajęciach	30	30/1,2 ETCS
	udział w zaliczeniach poza zajęciami	0	
	udział w konsultacjach	0	
praca własna	przygotowanie do zajęć *(czytanie, praca pisemna, tłumaczenie, ...)*	25	45/1,8 ECTS
	przygotowanie do zaliczenia *(np. czytanie, prezentacja, projekt, ...)*	20	

Kryteria oceny końcowej

(Opis składowych oceny końcowej zajęć, rozkład procentowy lub punktowy, informacja o dopuszczalnej liczbie nieobecności, inne kryteria)

Ocena końcowa z przedmiotu składa się z dwóch projektów realizowanych w trakcie ćwiczeń oraz wejściówki sprawdzającej przygotowanie studenta do zajęć:

- **projekt cząstkowy** – 30% oceny końcowej,
- **projekt zaliczeniowy (końcowy)** – 50% oceny końcowej.
- **wejściówka** (20%)

Punktacja wejściówek oraz projektów:
100–94% – 5
93–88% – 4,5
87–80% – 4
79–70% – 3,5
69–60% – 3
59% i mniej – 2

Projekt cząstkowy polega na zespołowej analizie zbioru danych biologicznych obejmującej ich opis, wizualizację oraz interpretację biologiczną wyników. Oceniana jest poprawność analizy (50%), sposób prezentacji danych zgodnie z udostępnionymi wytycznymi (20%) oraz umiejętność formułowania wniosków (30%). Prace studentów są archiwizowane na Moodle.

Projekt końcowy obejmuje samodzielną analizę danych biologicznych z wykorzystaniem poznanych narzędzi (Python, biblioteki analityczne, wybrane narzędzia AI) oraz przygotowanie formy prezentacji wyników (prezentacja, podcast lub krótki film). Ocenie podlega poprawność merytoryczna analiz (50%), logiczna struktura pracy (20%), krytyczne podejście do wyników (20%) oraz jakość prezentacji(10%). Prace studentów są archiwizowane na Moodle.

Warunkiem zaliczenia przedmiotu jest oddanie obu projektów oraz pozytywna ocena z wejściówek. Obecność na ćwiczeniach obowiązkowa.

Efekty uczenia się w zakresie wiedzy

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

Ocena 3,0
Student w podstawowym stopniu opisuje etyczne, społeczne i prawne aspekty wykorzystania narzędzi AI w analizie danych biologicznych. Rozpoznaje podstawowe ograniczenia modeli językowych i konieczność weryfikacji wyników, ale jego uzasadnienia są ogólne.
Ocena 4,0
Student poprawnie omawia etyczne, społeczne i prawne uwarunkowania stosowania AI w biologii. Wyjaśnia znaczenie prywatności danych, rzetelności analiz oraz krytycznej oceny treści generowanych przez AI.
Ocena 5,0
Student bardzo dobrze i szczegółowo omawia dylematy związane z wykorzystaniem AI w analizie danych biologicznych. Krytycznie odnosi się do ograniczeń narzędzi AI, zasad ochrony danych, rzetelności naukowej i odpowiedzialnego prowadzenia analiz.

Efekty uczenia się w zakresie umiejętności
Ocena 3,0
Student wykonuje podstawową analizę danych biologicznych z użyciem języka Python, ale często wymaga wsparcia prowadzącego. Potrafi wczytać dane, wykonać proste operacje w pandas i przygotować podstawowe wykresy, jednak interpretacja wyników jest fragmentaryczna.
Ocena 4,0
Student samodzielnie analizuje dane biologiczne, poprawnie je porządkuje, grupuje, wizualizuje i interpretuje. Potrafi zastosować biblioteki pandas, matplotlib i seaborn oraz wykorzystać narzędzia AI do pracy z literaturą i prezentacji wyników.
Ocena 5,0
Student bardzo dobrze i samodzielnie przeprowadza analizę danych biologicznych, świadomie dobiera metody, tworzy czytelne wizualizacje i formułuje pogłębione wnioski biologiczne. Umiejętnie wykorzystuje narzędzia AI, krytycznie oceniając ich wyniki i ograniczenia.

Efekty uczenia się w zakresie kompetencji społecznych
Ocena 3,0
Student w podstawowym zakresie korzysta z konsultacji i informacji zwrotnej przy rozwiązywaniu problemów analitycznych. Współpracuje w zespole, ale jego udział w organizacji pracy i krytycznej ocenie wyników jest ograniczony.
Ocena 4,0
Student aktywnie korzysta z konsultacji, współpracuje w zespole i odpowiedzialnie realizuje powierzone zadania. Potrafi krytycznie oceniać wyniki analiz oraz treści wygenerowane przez narzędzia AI.
Ocena 5,0
Student bardzo dobrze współpracuje w zespole, inicjuje konsultacje w sytuacjach problemowych i odpowiedzialnie wspiera proces rozwiązywania trudności analitycznych. Wykazuje wysoką świadomość ograniczeń własnej wiedzy, krytycznie ocenia wyniki i dba o rzetelność pracy projektowej.

Oceny połówkowe:
Ocena **3,5** jest wystawiana w przypadku pełnego zaliczenia efektów uczenia się na poziomie 3,0, jednak z niewystarczającym spełnieniem wymagań na ocenę 4,0.
Ocena **4,5** jest wystawiana w przypadku pełnego zaliczenia efektów uczenia się na poziomie 4,0, jednak bez pełnego osiągnięcia wymagań na ocenę 5,0.

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.

Literatura obowiązkowa

1.	Ekmekci, B., McAnany, C. E., & Mura, C. (2016). An introduction to programming for bioscientists: A Python primer. PLoS Computational Biology, 12(6), e1004867. https://doi.org/10.1371/journal.pcbi.1004867
2.	Allen B. Downey Myśl w języku Python : nauka programowania /; przekład: Piotr Pilch. 2025
3.	Li, Y., Wu, F. X., & Ngom, A. (2018). *A review on machine learning principles for multi-view biological data integration*. Briefings in Bioinformatics, 19(2), 325–340. https://doi.org/10.1093/bib/bbw113

Literatura uzupełniająca

1.	Zhou, Z., et al. (2023). Explainable artificial intelligence for biological data analysis. arXiv preprint arXiv:2312.06082. https://arxiv.org/abs/2312.06082
2.	Monika Fajfer-Jakubek; Bożena Sikora (2025) Population Structure and Climate Effects on *Geckobia* Infestation in *Ptyodactylus* Geckos from Israel and West Bank, with Descriptions of G. parva sp. nov. and G. inermis sp. Nov. *Animals 15*(23), 3461

*** lista rodzajów zajęć**

- ćwiczenia (audytoryjne, translatoryjne, terenowe, warsztatowe, projektowe)
- ćwiczenia laboratoryjne, komputerowe
- lektorat języka obcego nowożytnego/starożytnego
- wykład kierunkowy
- wykład monograficzny lub konwersatorium monograficzne
- seminarium dyplomowe
  (sem. magisterskie, licencjackie lub inżynierskie, na którym student pod kierunkiem opiekuna pracy przygotowuje pracę dyplomową, wykorzystując metody adekwatne do realizowanej tematyki badawczej)
- pracownia dyplomowa (programistyczna, chemiczna, fizyczna, biologiczna, inżynierska)
  (zajęcia laboratoryjne, na których student pod kierunkiem opiekuna pracy przygotowuje pracę dyplomową wykorzystując metody adekwatne do realizowanej tematyki badawczej)

Prawa autorskie do sylabusa przysługują UKSW. Sylabus może być wykorzystywany jedynie dla celów związanych z kształceniem na studiach odbywających się w UKSW, korzystanie z niego w innych celach wymaga zgody UKSW.