

KARTA KURSU (realizowanego w specjalności)

Inżynieria oprogramowania

(nazwa specjalności)

Nazwa	Analiza danych
Nazwa w j. ang.	<i>Data Analysis</i>

Koordinator	dr Roman Czapla	Zespół dydaktyczny
		mgr Katarzyna Marczak mgr Patryk Mazurek
Punktacja ECTS*	3	

Opis kursu (cele kształcenia)

Celem kursu jest rozwinięcie umiejętności inżynierskiej analizy danych z wykorzystaniem języka Python oraz bibliotek NumPy, Pandas, Matplotlib i Seaborn. Studenci poznają metody pozyskiwania, przetwarzania i przygotowania danych do dalszej analizy oraz uczą się interpretować uzyskane wyniki w kontekście rozwiązywanego problemu informatycznego.

Szczególny nacisk kładziony jest na jakość danych, poprawność przekształceń oraz świadomy dobór metod analitycznych i wizualizacyjnych. Kurs obejmuje zagadnienia związane z eksploracją danych, analizą zależności między zmiennymi oraz przygotowaniem danych do zastosowań predykcyjnych.

W ramach zajęć studenci realizują projekt analityczny oparty na rzeczywistym zbiorze danych, obejmujący podstawowe etapy pracy z danymi, zakończone krótkim opracowaniem wyników. Kurs stanowi podstawę do dalszych przedmiotów związanych z analizą danych, uczeniem maszynowym i inżynierią systemów przetwarzających dane.

Warunki wstępne

Wiedza	Student zna podstawy programowania w języku Python, w tym zmienne, instrukcje sterujące, funkcje i podstawowe typy danych.
Umiejętności	Student potrafi tworzyć proste programy w języku Python, uruchamiać skrypty, korzystać z bibliotek zewnętrznych oraz pracować w środowisku Jupyter Notebook lub podobnym.
Kursy	<u>Wymagane zaliczenie kursu</u> : Podstawy Programowania w języku Python

Efekty uczenia się

Wiedza	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
--------	-----------------------------	-------------------------------------

	Po zakończeniu kursu student:	
	W01: posiada wiedzę dotyczącą etapów procesu analizy danych oraz rozumie ich znaczenie w rozwiązywaniu problemów informatycznych.	S1_W01 S1_W04
	W02: ma uporządkowaną wiedzę na temat struktur danych i mechanizmów przetwarzania danych w bibliotekach NumPy i Pandas.	S1_W02 S1_W04
	W03: rozumie metody statystyki opisowej oraz techniki analizy zależności między zmiennymi wykorzystywane w eksploracji danych.	S1_W04
	W04: posiada wiedzę na temat metod przygotowania danych do analizy, w tym czyszczenia, transformacji i kodowania zmiennych, oraz rozumie ich wpływ na jakość wyników.	S1_W03 S1_W04
	W05: zna zasady wizualizacji danych i kryteria doboru form prezentacji wyników analizy.	S1_W04

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Umiejętności	Po zakończeniu kursu student:	
	U01: sprawnie pozyskuje dane z różnych źródeł oraz przygotowuje je do dalszej analizy z wykorzystaniem narzędzi programistycznych.	S1_U01 S1_U07
	U02: realizuje operacje przekształcania i agregacji danych (filtrowanie, sortowanie, grupowanie) oraz interpretuje ich rezultaty.	S1_U01 S1_U06 S1_U07
	U03: analizuje zależności między zmiennymi i interpretuje wyniki statystyczne w kontekście rozwiązywanego problemu.	S1_U07
	U04: przygotowuje dane do zastosowań predykcyjnych poprzez odpowiednie przekształcenia, selekcję cech i podział zbioru danych.	S1_U05 S1_U07
	U05: opracowuje czytelne wizualizacje oraz prezentuje wyniki analizy w formie uporządkowanego raportu technicznego.	S1_U07 S1_U10

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Kompetencje społeczne	Po zakończeniu kursu student:	
	K01: współpracuje przy realizacji zadań analitycznych, planując etapy pracy i dzieląc odpowiedzialność w zespole.	S1_K03
	K02: jest świadomy znaczenia rzetelności analizy danych oraz odpowiedzialności za interpretację i prezentację wyników.	S1_K02 S1_K04

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	15			15			

Studia niestacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	10			10			

Opis metod prowadzenia zajęć

Zajęcia realizowane są w formie wykładów oraz ćwiczeń laboratoryjnych.

Podczas wykładów omawiane są podstawowe koncepcje związane z analizą danych, w tym etapy pracy z danymi, metody przygotowania zbiorów do analizy, elementy statystyki opisowej oraz zasady wizualizacji wyników. Nacisk położony jest na zrozumienie procesu analizy danych oraz świadomego doboru metod do charakteru problemu.

Ćwiczenia mają charakter praktyczny i odbywają się w środowisku Python (np. Jupyter Notebook). Studenci wykonują zadania obejmujące wczytywanie danych, ich czyszczenie, przekształcanie, analizę zależności oraz wizualizację wyników.

W ramach zajęć realizowane są krótkie zadania projektowe oparte na rzeczywistych zbiorach danych, umożliwiające samodzielne zastosowanie poznanych narzędzi i metod.

Formy sprawdzania efektów uczenia się

	E – learning	Gry dydaktyczne	Ćwiczenia w szkole	Zajęcia terenowe	Praca laboratoryjna	Projekt indywidualny	Projekt grupowy	Udział w dyskusji	Referat	Test zaliczeniowy	Egzamin ustny	Egzamin pisemny	Zadania problemowe
W01					x			x		x			
W02					x			x		x			
W03					x			x		x			
W04					x			x		x			
W05					x			x		x			
U01					x	x		x					
U02					x	x		x					
U03					x	x		x					
U04					x	x		x					
U05					x	x		x					
K01						x		x					
K02						x		x					

Kryteria oceny	Osiągnięcie efektów kształcenia podanych powyżej uprawnia studentów do uzyskania oceny nie wyższej niż dostateczna, pod warunkiem spełnienia minimalnych wymagań określonych dla przedmiotu.
	Warunkiem uzyskania oceny dostatecznej (3.0) lub dobrej (4.0) jest: <ul style="list-style-type: none"> • zaliczenie testu sprawdzającego wiedzę i umiejętności z zakresu analizy danych, • aktywne wykonywanie zadań realizowanych podczas ćwiczeń laboratoryjnych.
	Prowadzący może uwzględnić stopień zaangażowania studenta w trakcie zajęć laboratoryjnych (systematyczność pracy, samodzielność rozwiązywania zadań, poprawność interpretacji wyników) przy ustalaniu oceny końcowej z zaliczenia.
	Ocena bardzo dobra (5.0) wymaga, oprócz spełnienia powyższych warunków, wykonania krótkiego projektu analitycznego obejmującego samodzielne zastosowanie poznanych metod do wybranego zbioru danych.

Obecność na wykładach jest warunkiem zaliczenia tej części kursu.

Uwagi

Treści merytoryczne (wykaz tematów)

1. Wprowadzenie do analizy danych: znaczenie analizy danych i jej zastosowania w różnych dziedzinach, środowisko pracy Python (Jupyter Notebook), przegląd bibliotek: NumPy, Pandas, Matplotlib, Seaborn.
2. Struktury i operacje na danych: tworzenie i manipulowanie strukturami danych w Pandas i NumPy, wczytywanie danych z plików (CSV, Excel, JSON), filtrowanie, sortowanie, grupowanie i agregacja danych, operacje warunkowe i tworzenie nowych kolumn.
3. Statystyka opisowa w praktyce: typy danych i ich charakterystyka, obliczanie i interpretacja miar tendencji centralnej i zmienności (średnia, mediana, kwartyle, wariancja), podstawy korelacji i analiza zależności między zmiennymi.
4. Czyszczenie i przygotowanie danych: identyfikacja i obsługa danych brakujących i błędnych, usuwanie duplikatów, konwersje typów danych, skalowanie, standaryzacja oraz kodowanie zmiennych kategoriycznych.
5. Wizualizacja danych: tworzenie podstawowych wykresów (słupkowe, liniowe, histogramy) za pomocą Matplotlib, oraz zaawansowanych wizualizacji (boxploty, heatmaps, pairploty) z użyciem Seaborn; dobór wykresów do rodzaju danych, estetyka i czytelność prezentacji.
6. Analiza danych w praktyce: praca z rzeczywistymi zbiorami danych (np. dane społeczne, gospodarcze, edukacyjne), identyfikacja problemów w danych, eksploracja zależności i prezentacja wyników w formie raportów z komentarzem i wizualizacjami.
7. Wprowadzenie do analizy danych dla uczenia maszynowego: przygotowanie danych do analizy predykcyjnej (czyszczenie, selekcja cech, podział na zbiory treningowe i testowe), podstawowe modele regresji i klasyfikacji jako ilustracja praktycznych zastosowań.

Wykaz literatury podstawowej

1. W. Ayd, M. Harrison, *Pandas. Receptury. Obliczenia naukowe, szeregi czasowe i eksploracyjna analiza danych w Pythonie*. Wydanie III, Gliwice, Helion 2025 (wybrane fragmenty);
2. W. McKinney, *Python w analizie danych. Przetwarzanie danych za pomocą pakietów Pandas i NumPy oraz środowiska Jupyter*. Wydanie III, Gliwice, Helion 2023 (wybrane fragmenty);
3. T. Nield, *Podstawy matematyki w data science. Algebra liniowa, rachunek prawdopodobieństwa i statystyka*, Helion, Gliwice, 2023 (wybrane fragmenty);
4. J. VanderPlas, *Python Data Science. Niezbędne narzędzia do pracy z danymi*. Wydanie II, Helion, Gliwice 2023 (wybrane fragmenty);

5. A. Navlani, A. Fandango, I. Idris, *Python i praca z danymi. Przetwarzanie, analiza, modelowanie i wizualizacja. Wydanie III*, Helion, Gliwice 2022 (wybrane fragmenty);
6. K. Gallatin, Ch. Albon, *Uczenie maszynowe w Pythonie. Receptury. Od przygotowania danych do deep learningu*. Wydanie II, Gliwice, Helion, 2023 (wybrane fragmenty)

Wykaz literatury uzupełniającej

1. M. Walker, *Czyszczenie danych w Pythonie. Receptury. Nowoczesne techniki i narzędzia Pythona do wykrywania i eliminacji zanieczyszczeń oraz wydobywania kluczowych cech z danych*, Helion, Gliwice 2021;
2. R. Johansson, *Matematyczny Python. Obliczenia naukowe i analiza danych z użyciem NumPy, SciPy i Matplotlib*, Helion, Gliwice 2021; (wybrane fragmenty);

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) – **studia stacjonarne**

Liczba godzin w kontakcie z prowadzącymi	Wykład	15
	Konwersatorium (ćwiczenia, laboratorium itd.)	15
	Pozostałe godziny kontaktu studenta z prowadzącym	5
Liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	10
	Realizacja zadań domowych (problemowych) po zapoznaniu się z niezbędną literaturą przedmiotu	10
	Przygotowanie projektu lub prezentacji na podany temat (praca indywidualna lub w grupie)	10
	Przygotowanie do egzaminu/zaliczenia	10
Ogółem bilans czasu pracy		75
Liczba punktów ECTS w zależności od przyjętego przelicznika		3

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) – **studia niestacjonarne**

Liczba godzin w kontakcie z prowadzącymi	Wykład	10
	Konwersatorium (ćwiczenia, laboratorium itd.)	10
	Pozostałe godziny kontaktu studenta z prowadzącym	5
Liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	15
	Realizacja zadań domowych (problemowych) po zapoznaniu się z niezbędną literaturą przedmiotu	15
	Przygotowanie projektu lub prezentacji na podany temat (praca indywidualna lub w grupie)	10
	Przygotowanie do egzaminu/zaliczenia	10
Ogółem bilans czasu pracy		75
Liczba punktów ECTS w zależności od przyjętego przelicznika		3